CHƯƠNG 8

CÁC KIỂM ĐỊNH NHỮNG GIẢ THUYẾT VỀ CÁC SỐ TRUNG BÌNH VÀ TỶ LỆ

***Về chương này*:**

Như chúng ta đã giải thích trong Chương 7, có hai phương pháp để thực hiện việc suy luận về các tham số tổng thể dựa trên dữ liệu mẫu. Phương pháp thứ nhất, ước lượng thống kê, là chủ đề của Chương 7. Mục đích của Chương 8 là trình bày phương pháp thứ hai về việc thực hiện các suy luận về những tham số tổng thể - kiểm định các giả thuyết về giá trị của chúng. Cũng giống như trường hợp trong Chương 7, chúng ta sẽ trình bày qui trình cho các tình huống mà ở đó những cỡ mẫu là đủ lớn để tạo ra sự xấp xỉ chuẩn trong các phân phối mẫu về các trị thống kê mẫu được sử dụng cho việc suy luận. Chúng ta cũng sẽ giải thích qui trình cho các tình huống mà ở đó những cỡ mẫu là nhỏ nhưng các tổng thể được chọn mẫu là chuẩn.

NGHIÊN CỨU ĐIỂN HÌNH

CÁC NHÀ THẦU CHO NGÀNH CÔNG NGHIỆP QUỐC PHÒNG THỜI HẬU CHIẾN TRANH LẠNH: LIỆU SỰ THAY ĐỔI LÀ KHẢ DĨ?

Với sự tan rã của khối Cộng sản tại Đông Âu, thì những sự giảm sút về chi tiêu cho quốc phòng tại Hoa Kỳ đã gây ra những vấn đề lớn tại các tiểu bang mà có nguồn thu phụ thuộc chủ yếu vào các ngành công nghiệp quốc phòng mang lại lợi nhuận cao. Những vấn đề này bị làm cho tồi tệ thêm bởi sự đóng cửa nhiều căn cứ quân sự và việc cắt giảm qui mô của các căn cứ khác. Hơn nữa, ngành công nghiệp ký kết các hợp đồng thầu về quốc phòng đang đối mặt với những thách thức về năng lực sản xuất dư thừa quá mức, những cấu trúc doanh nghiệp quá nặng nề, và những hoạt động không hiệu quả. Khi quan tâm đến những tình huống này, các nhà thầu của chính phủ phải trở nên ngày càng hiệu quả và hữu hiệu về mặt chi phí hơn.

Trong một nghiên cứu nhằm đánh giá các hệ thống hạch toán chi phí đang được sử dụng bởi các nhà thầu cho ngành công nghiệp quốc phòng, Rezaee và Elmore (1993) đã điều tra 112 công ty. Các bảng câu hỏi có thể sử dụng được hoàn tất bởi 50 người trả lời (các kế toán giám sát), mà vào lúc đó được phân thành hai nhóm: 25 nhà thầu cho ngành công nghiệp quốc phòng và 25 nhà thầu cho ngành công nghiệp không phải quốc phòng. Kết quả một phần của cuộc điều tra này liên quan đến những câu trả lời (được đo bằng Thang điểm Likert) đối với các câu hỏi về việc sử dụng các qui trình dự trù ngân sách và hoạch định được trình bày trong Bảng 8.1.

BẢNG **8.1**

	Quốc phòng	**Không phải quốc phòng**	**Giá trị *T***
Hoạch định chiến lược			
1 Ngân sách được kiểm tra về sự nhất quán với các mục tiêu dài hạn	4.0425	4.2000	-0.30
2 Xác nhận chính thức về các mục tiêu, chiến lược, v.v được sử dụng cho việc hoạch định phương hướng của công ty	4.1625	4.8800	-1.45
Ngân sách và Hoạch định			
1 Ngân sách được sử dụng trong việc đánh giá hiệu quả hoạt động của những thành viên riêng lẻ	3.1600	4.5200	-2.56*
2 So sánh giữa các chi phí thực tế so với chi phí được dự trù ngân sách	3.9891	5.1782	2.64*
3 Ngân sách của các phòng ban riêng lẻ	2.8800	4.6800	-3.238
4 Sự tham gia của quản lý cấp trung và thấp hơn trong việc dự trù ngân sách	3.6879	5.0800	-3.53*
5 Ngân sách linh hoạt	2.1861	3.6000	-2.64*

* Có ý nghĩa ở mức 0.01

Liệu có những sự khác biệt thực sự trong các câu trả lời đối với những phần của bảng câu hỏi cho các ngành công nghiệp quốc phòng so với các ngành không phải quốc phòng? Những khác biệt này có thể được đo lường theo tiêu chuẩn nào? Trong Chương 7, chúng ta có khả năng ước lượng những khác biệt về các số trung bình tổng thể bằng cách sử dụng hoặc sự phân phối chuẩn chuẩn hóa hoặc sự phân phối *t* như là sự phân phối có liên quan, tùy thuộc vào cỡ mẫu, phương pháp chọn mẫu, và bản chất của tổng thể được nhắc đến mà được chọn mẫu. Trong chương này, các bạn sẽ học hỏi về các phương pháp chính thức cho việc kiểm định những giả định về nhiều tham số tổng thể. Chúng ta sẽ sử dụng các kỹ thuật này trong Phần 8.12 để xác định liệu có những sự khác biệt thực sự hay không giữa các ngành công nghiệp quốc phòng so với các ngành không phải quốc phòng mà có liên quan đến các qui trình hạch toán và dự trù ngân sách.

8.1 KIỂM ĐỊNH NHỮNG GIẢ THUYẾT VỀ CÁC THAM SỐ TỔNG THỂ

Một số vấn đề thực tiễn đòi hỏi chúng ta phải ước lượng giá trị của một tham số tổng thể; các tình huống khác đòi hỏi chúng ta phải thực hiện các quyết định liên quan đến giá trị của tham số đó. Ví dụ, nếu một công ty dược phẩm đang lên men một bể chứa thuốc kháng sinh, thì công ty này ắt muốn kiểm định hiệu lực của các mẫu thuốc kháng sinh và sử dụng các mẫu này để ước lượng hiệu lực trung bình μ của thuốc kháng sinh chứa trong bể. Mặt khác, giả thuyết rằng không có quan ngại nào rằng hiệu lực của loại thuốc kháng sinh đó là quá cao; thì quan ngại duy nhất của công ty này là rằng hiệu lực trung bình liệu có vượt quá một mức tối thiểu nào đó do chính phủ qui định cụ thể nhằm để cho bể thuốc đó được tuyên bố là có thể chấp nhận được để bán. Trong trường hợp này, công ty đó ắt sẽ không mong muốn ước lượng hiệu lực trung bình này. Thay vào đó, công ty này muốn chứng tỏ rằng hiệu lực trung bình của thuốc kháng sinh trong bể đó đã vượt quá một mức tối thiểu được qui định bởi chính phủ. Như vậy, công ty này ắt sẽ muốn quyết định liệu hiệu lực trung bình có vượt quá hiệu lực tối thiểu được cho phép hay không. Vấn đề của công ty dược phẩm này minh họa một **sự kiểm định thống kê về giả thuyết.**

Lý do hợp lý được sử dụng trong việc kiểm định một giả thuyết có một sự tương đồng đáng kinh ngạc về qui trình được sử dụng tại một phiên tòa xét xử. Khi xét xử một người vì tội trộm cắp, thì tòa án cho rằng bị cáo là vô tội cho đến khi được chứng minh là có tội. Bên nguyên thu thập và trình bày tất cả các bằng chứng sẵn có trong một nỗ lực nhằm phủ nhận giả thuyết "không có tội" và vì thế đạt được một sự kết tội. Tuy nhiên, nếu bên nguyên thất bại trong việc bác bỏ giả thuyết "không có tội" này, thì điều này không chứng minh được rằng bị cáo là "vô tội" mà chỉ đơn thuần là chưa có đủ bằng chứng để kết luận rằng bị cáo là "có tội".

Vấn đề thống kê này miêu tả sinh động cho hiệu lực của thuốc kháng sinh đóng vai trò như là bị cáo. Giả thuyết cần được kiểm định này, được gọi là **giả thuyết không,** là rằng hiệu lực không vượt quá mức tiêu chuẩn tối thiểu của chính phủ. Bằng chứng trong tường hợp này được chứa đựng trong mẫu của các mẫu xét nghiệm được lấy ra từ cái bể đó. Công ty dược phẩm này, đóng vai trò như là công tố viên, tin rằng một **giả thuyết thay thế** là có thật - cụ thể là, rằng hiệu lực của thuốc kháng sinh đó thật sự vượt quá mức tiêu chuẩn tối thiểu. Do vậy, công ty này cố gắng sử dụng bằng chứng trong mẫu để bác bỏ giả thuyết không (hiệu lực không vượt quá mức tiêu chuẩn tối thiểu), qua đó ủng hộ cho giả thuyết thay thế (hiệu lực vượt qua mức tiêu chuẩn tối thiểu). Các bạn sẽ thừa nhận qui trình này như là một đặc trưng cốt yếu của phương pháp khoa học, mà trong đó tất cả các lý thuyết được đề xuất phải được so sánh với thực tế.

Trong chương này, chúng ta sẽ giải thích các khái niệm cơ bản về một sự kiểm định về một giả thuyết và chứng minh các khái niệm này bằng một số kiểm định thống kê hết sức hữu ích về các giá trị của một tham số tổng thể, một tỷ lệ tổng thể, sự khác biệt giữa một cặp số trung bình tổng thể, và sự khác biệt giữa hai tỷ lệ nhị thức. Chúng ta sẽ sử dụng bốn sự ước lượng điểm được thảo luận trong Chương 7, $\bar{x}$, $(\bar{x}_1 - \bar{x}_{2})$, $\hat{p}$, và $(\hat{p}_1 - \hat{p}_2)$, như là các trị thống kê kiểm định và, khi làm việc này, sẽ đạt được một sự thống nhất trong bốn sự kiểm định thống kê này. Tất cả bốn sự kiểm định thống kê này sẽ, đối với các mẫu lớn, có những phân phối mẫu là chuẩn hay có thể được ước lượng xấp xỉ bởi một phân phối chuẩn. Đối với các mẫu nhỏ, thì các bài kiểm định thống kê có liên quan đến số trung bình tổng thể có liên quan đến sự phân phối *t* Student.

8.2 MỘT KIỂM ĐỊNH THỐNG KÊ VỀ GIẢ THUYẾT

Một sự kiểm định thống kê về giả thuyết bao gồm bốn phần:

- một giả thuyết không, được ký hiệu bởi H_0
- một giả thuyết thay thế, được ký hiệu bởi H_a
- một trị thống kê kiểm định
- một vùng bác bỏ

Sự định rõ của bốn yếu tố này xác định một bài kiểm định cụ thể; việc thay đổi một hay nhiều hơn các bộ phận này tạo ra một bài kiểm định mới.

Giả thuyết thay thế là một giả thuyết mà nhà nghiên cứu mong muốn ủng hộ. **Giả thuyết không** là sự phủ nhận của giả thuyết thay thế; nghĩa là, nếu giả thuyết không là sai, thì giả thuyết thay thế phải là đúng. Vì những lý do mà các bạn sẽ thấy sau đây, việc chứng minh sự ủng hộ cho giả thuyết thay thế bằng cách trình bày bằng chứng (dữ liệu mẫu) mà chỉ ra rằng giả thuyết không là sai là việc làm dễ dàng hơn nhiều. Như thế chúng ta đang xây dựng một trường hợp nhằm ủng hộ giả thuyết thay thế bằng cách sử dụng một phương pháp mà tương tự như bằng chứng của sự trái ngược.

Mặc dù chúng ta muốn có được bằng chứng trong việc ủng hộ cho giả thuyết thay thế, thì giả thuyết không là giả thuyết cần phải được kiểm tra. Như vậy, H_0 sẽ xác định cụ thể các giá trị được giả thuyết cho một hay nhiều hơn các tham số tổng thể.

VÍ DỤ **8.1** Chúng ta muốn biết về tiền lương trung bình mỗi giờ của công nhân xây dựng tại tiểu bang California là khác với $14, đó là mức trung bình trên toàn quốc. Sau đây là giả thuyết thay thế, được biểu diễn bằng

$$H_a : \mu \neq 14$$

Giả thuyết không được viết như sau

$$H_0 : \mu = 14$$

Chúng ta sẽ muốn bác bỏ giả thuyết không, như vậy qua đó kết luận rằng số trung bình của bang California là không bằng với $14.

VÍ DỤ **8.2** Một qui trình nghiền hiện đang tạo ra một tỷ lệ bình quân là 3% sản phẩm có lỗi. Chúng ta quan tâm đến việc chứng minh rằng một sự điều chỉnh đơn giản đối với cái máy này sẽ làm giảm *p*, tỷ lệ của sản phẩm có lỗi được sản xuất ra trong qui trình nghiền này. Vì thế, chúng ta viết ra giả thuyết thay thế như sau:

$$H_a : p < 0.3$$

và giả thuyết không như sau

$$H_0 : p = 0.3$$

Nếu chúng ta có thể bác bỏ H_0, thì chúng ta có thể kết luận rằng qui trình được điều chỉnh này tạo ra ít sản phẩm có lỗi hơn.

Có một sự khác biệt về các loại hình của giả thuyết thay thế được cho trong các Ví dụ 8.1 và 8.2. Trong Ví dụ 8.1, không có sự khác biệt mang tính định hướng nào được đề nghị cho giá trị của μ; nghĩa là, μ có thể hoặc lớn hơn hoặc nhỏ hơn $14 nếu H_a là đúng. Tuy nhiên, trong Ví dụ 8.2, chúng ta quan tâm một cách cụ thể đến việc phát hiện một sự khác biệt mang tính định hướng trong giá trị của p; nghĩa là, nếu H_a là đúng, thì giá trị của p sẽ nhỏ hơn 0.03. Loại kiểm định này được gọi là **kiểm định một phía của giả thuyết.**

Quyết định bác bỏ hay chấp nhận giả thuyết không được căn cứ vào thông tin chứa trong một mẫu được lấy ra từ tổng thể quan tâm. Các giá trị của mẫu được sử dụng để tính toán một con số duy nhất, tương ứng với một điểm trên đường thẳng, mà hoạt động như là một vật quyết định. Giá trị quyết định này được gọi là **trị thống kê kiểm định.** Toàn bộ tập hợp các giá trị mà trị thống kê kiểm định này có thể có được chia thành hai tập hợp, hay hai vùng. Một tập hợp, bao gồm các giá trị mà ủng hộ cho giả thuyết thay thế, được gọi là **vùng bác bỏ**. Tập hợp kia, bao gồm các giá trị mà không mâu thuẫn với giả thuyết không, được gọi là **vùng chấp nhận.**

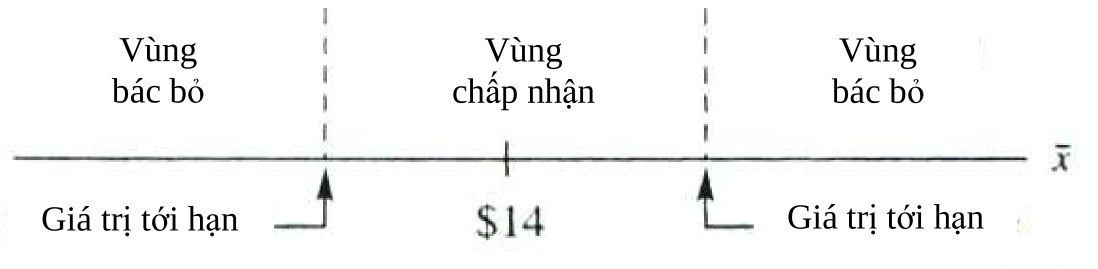
Các vùng chấp nhận và bác bỏ được phân cách bởi một **giá trị tới hạn** của trị thống kê kiểm định đó. Nếu trị thống kê kiểm định này được tính từ một mẫu cụ thể có một giá trị nằm trong vùng bác bỏ, thì giả thuyết không bị bác bỏ, và giả thuyết thay thế H_a được chấp nhận. Nếu trị thống kê đó rơi vào vùng chấp nhận, thì hoặc là giả thuyết không được chấp nhận hoặc trị kiểm định đó bị đánh giá là không thuyết phục. Trong bất cứ trường hợp nào, thì sự thất bại trong việc bác bỏ H_a hàm ý rằng dữ liệu này không đại diện đủ bằng chứng để hỗ trợ H_a. Các tình huống này dẫn đến quyết định mà sẽ được giải thích sau đây.

VÍ DỤ **8.3** Đối với sự kiểm định về giả định được cho trong Ví dụ 8.1, thì tiền lương bình quân $\bar{x}$ cho một mẫu ngẫu nhiên gồm 100 công nhân xây dựng tại tiểu bang California có lẽ cung cấp cho chúng ta một trị kiểm định thống kê tốt cho việc kiểm định

$$H_0 : \mu = 14 \quad \text{so với} \quad H_a : \mu \neq 14$$

Bởi vì trung bình của mẫu là số ước lượng tốt nhất cho trung bình tổng thể tương ứng, nên chúng ta sẽ thiên về việc bác bỏ H_0 và ủng hộ H_a nếu số trung bình mẫu $\bar{x}$ là hoặc nhỏ hơn $14 hoặc lớn hơn nhiều so với $14. Như vậy, vùng bác bỏ ắt chứa các giá trị cả lớn lẫn nhỏ của $\bar{x}$, như được thể hiện trong Hình 8.1

HÌNH **8.1** Các vùng bác bỏ và chấp nhận của Ví dụ 8.1



Qui trình quyết định được mô tả trên đây phụ thuộc vào hai loại hình sai lầm mà phổ biến trong các vấn đề quyết định hai chọn lựa.

ĐỊNH NGHĨA Một **sai lầm loại I** đối với một kiểm định thống kê là sai số được tạo ra từ việc bác bỏ giả thuyết không khi giả định này là đúng. Xác suất của việc tạo ra một sai lầm loại I được biểu thị bởi ký hiệu α.

Một **sai lầm loại II** đối với một kiểm định thống kê là sai lầm được tạo ra từ việc chấp nhận (không phải bác bỏ) giả thuyết không khi giả thuyết này là sai và một giả thuyết thay thế nào đó là đúng. Xác suất của việc tạo ra một sai lầm loại II được biểu thị bởi ký hiệu β.

Hai khả năng xảy ra này cho giả thuyết không - nghĩa là, đúng hoặc sai - cùng với hai quyết định mà người làm thí nghiệm có thể thực hiện, được chỉ ra trong bảng hai chiều này, Bảng 8.2. Những sự xảy ra của các sai lầm loại I và II được thể hiện trong các ô thích hợp.

BẢNG **8.2** Bảng quyết định

	Giả thuyết không	
Quyết định	**Đúng**	**Sai**
Bác bỏ H_0	Sai lầm loại I	Quyết định đúng
Chấp nhận H_0	Quyết định đúng	Sai lầm loại II

Mức độ thích hợp của một kiểm định thống kê đối với một giả thuyết được đo lường bởi các xác suất của việc tạo ra sai lầm loại I và II, được biểu thị bởi các ký hiệu α và β.

Các phần khác nhau của một kiểm định thống kê đối với một giả thuyết được tóm tắt trong phần trình bày sau đây.

Các Phần của một sự Kiểm định Thống kê

- **Giả thuyết không:** Giả thuyết mà được cho là đúng cho đến khi được chứng minh là sai; sự phủ nhận của giả thuyết thay thế
- **Giả thuyết thay thế:** Giả thuyết mà người nghiên cứu mong muốn ủng hộ hay chứng minh là đúng
- **Kiểm định một phía của giả thuyết:** Một kiểm định mà có một sự khác biệt một phía về tham số quan tâm nếu giả thuyết thay thế là đúng
- **Kiểm định hai phía của giả thuyết:** Một kiểm định mà có một sự khác biệt hai phía (hoặc lớn hơn hoặc nhỏ hơn) về tham số quan tâm nếu giả thuyết thay thế là đúng

- **Trị thống kê kiểm định:** Một trị thống kê được tính từ các đại lượng mẫu mà sẽ được sử dụng như là một giá trị quyết định
- **Vùng bác bỏ:** Các giá trị của trị thống kê kiểm định mà qua đó H_0 sẽ bị bác bỏ
- **Vùng chấp nhận:** Các giá trị của trị thống kê kiểm định mà qua đó H_0 sẽ được chấp nhận
- **Các giá trị tới hạn của trị thống kê kiểm định:** Các giá trị của trị thống kê kiểm định mà phân cách các vùng bác bỏ và chấp nhận
- **Kết luận:** Chuỗi hành động phải được thuân theo, được căn cứ vào giá trị quan sát được của trị thống kê kiểm định
- **Sai lầm loại I (với xác suất α):** Bác bỏ H_0 khi H_0 là đúng
- **Sai lầm loại II (với xác suất β):** Chấp nhận H_0 khi H_0 là sai

Các kiểm định cho mẫu lớn đối với giả thuyết có liên quan đến những số trung bình và tỷ lệ của tổng thể là tương tự nhau. Sự tương tự nhau này nằm trong sự kiện rằng tất cả các số ước lượng điểm được thảo luận trong Chương 7 là không bị lệch và có những phân phối xác suất mà, đối với các mẫu lớn, có thể tuân theo xấp xỉ một phân phối chuẩn. Vì vậy, chúng ta có thể sử dụng các số ước lượng điểm này như là các trị thống kê kiểm định để kiểm định các giả thuyết về những tham số liên quan. Đối với các mẫu nhỏ, các hình thức của những sự kiểm định thống kê về giả thiết là tương tự nhau, nhưng những phân phối mẫu của các trị thống kê kiểm định này có phân phối *t* Student. Chúng ta sẽ xem xét các kiểm định về giả thuyết có liên quan đến bốn tham số tổng thể $\mu, p, \mu_1 - \mu_2,$ và $p_1 - p_2$ một cách riêng biệt trong những phần sau.

8.3 MỘT KIỂM ĐỊNH GIẢ THUYẾT CHO MỘT SỐ TRUNG BÌNH CỦA TỔNG THỂ

Hãy xem xét một mẫu ngẫu nhiên gồm *n* đại lượng được lấy từ một tổng thể có trung bình μ và độ lệch chuẩn σ. Chúng ta muốn kiểm định một giả thuyết có dạng[+]

$$H_0 : \mu = \mu_0$$

trong đó μ_0 là một giá trị được giả thuyết nào đó của μ, so với một giả thuyết một phía khác

$$H_a : \mu > \mu_0$$

Chúng ta sẽ sử dụng một số zêrô in nhỏ dưới dòng để chỉ giá trị của tham số được xác định cụ thể bởi H_0. Lưu ý rằng H_0 cung cấp một giá trị về tham số sẽ được kiểm định, nghĩa là, μ bằng với μ_0, trong khi H_a cho ta một dãy các giá trị có thể có của μ. Số trung bình mẫu $\bar{x}$ là ước lượng tốt nhất cho giá trị thực tế của μ, mà đang là vấn đề bàn luận. Các giá trị nào của $\bar{x}$ sẽ dẫn chúng ta đến việc tin rằng H_0 là sai và μ trên thực tế là lớn hơn giá trị được giả

[+] Lưu ý rằng nếu kiểm định này bác bỏ giả thuyết không $\mu = \mu_0$ và ủng hộ cho giả thuyết thay thế $\mu > \mu_0$, thì kiểm định này chắc chắn sẽ bác bỏ giả thuyết không có dạng $\mu < \mu_0$, bởi vì giả thuyết này thậm chí còn mâu thuẫn hơn giả thuyết thay thế. Vì lý do này, trong cuốn sách này chúng tôi thừa nhận giả thuyết không cho một kiểm định một phía như là $\mu = \mu_0$ hơn là $\mu \le \mu_0$.

thuyết? Những giá trị này của $\bar{x}$ mà cực kỳ *lớn* ắt hàm ý rằng μ là lớn hơn so với giá trị được giả thuyết. Vì vậy, chúng ta sẽ bác bỏ H_0 nếu $\bar{x}$ là "quá lớn".

Vấn đề kế tiếp là định nghĩa điều mà chúng ta xem là "quá lớn". Các giá trị của $\bar{x}$ mà nằm quá xa độ lệch chuẩn về phía bên phải của trung bình rất ít có khả năng xảy ra. Do vậy, chúng ta có thể định nghĩa "quá lớn" là cách xa μ_0 bởi quá nhiều lần độ lệch chuẩn. Hãy nhớ rằng độ lệch chuẩn hay sai số chuẩn của $\bar{x}$ được tính bằng

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

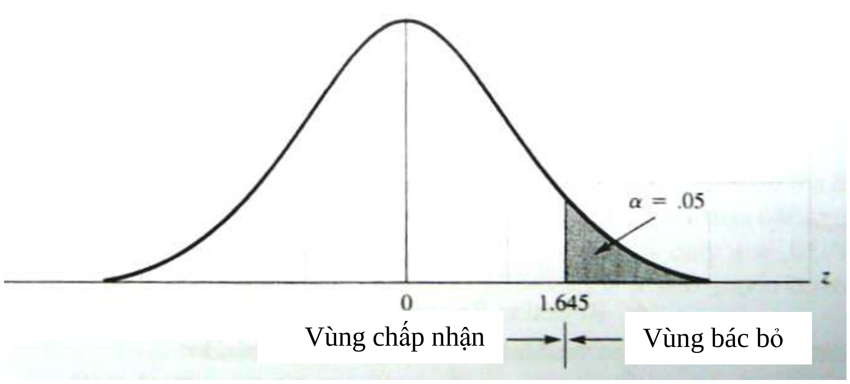
Kiểm định cho Mẫu lớn

Khi cỡ mẫu *n* là lớn, thì phân phối xác suất của $\bar{x}$ là xấp xỉ chuẩn, và chúng ta có thể đo lượng con số các độ lệch chuẩn để cho $\bar{x}$ nằm cách μ_0 qua việc sử dụng **trị thống kê kiểm định chuẩn hóa.**

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

mà có một phân phối chuẩn chuẩn hóa khi $H_0 : \mu = \mu_0$ là đúng. Nếu giả thuyết thay thế là $H_a : \mu > \mu_0$, thì xác suất của α về việc bác bỏ giả thuyết không, khi giả thuyết này là đúng, là bằng với diện tích dưới đường cong chuẩn nằm phía trên vùng bác bỏ. Như vậy, nếu chúng ta muốn α = 0.05, thì chúng ta ắt sẽ bác bỏ H_0 khi $\bar{x}$ là lớn hơn 1.645 lần độ lệch chuẩn về phía bên phải của μ_0. Tương tự như vậy, chúng ta ắt sẽ bác bỏ H_0 nếu trị thống kê kiểm định chuẩn hóa *z*, được định nghĩa ở trên, là lớn hơn 1.645 (Hình 8.2).

HÌNH **8.2** Sự phân phối của $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ khi H_0 là đúng

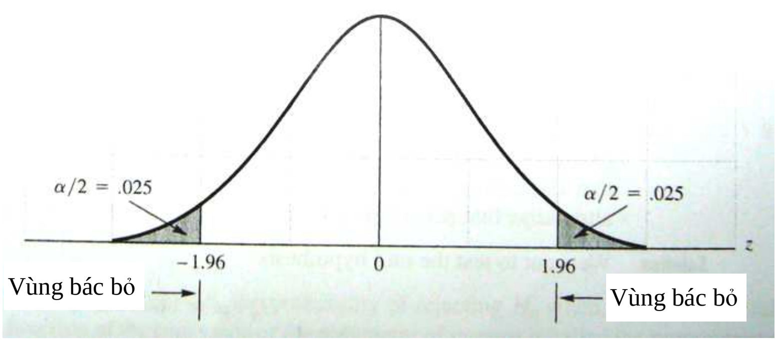


Nếu chúng ta mong muốn tìm ra những sự đi lệch hoặc lớn hơn hay nhỏ hơn μ_0, thì giả thuyết thay thế ắt sẽ là *hai phía,* được thể hiện như sau

$$H_a : \mu \neq \mu_0$$

mà hàm ý rằng hoặc $\mu > \mu_0$ hoặc $\mu < \mu_0$. Các giá trị của $\bar{x}$ mà hoặc "quá lớn" hoặc "quá nhỏ" nếu xét về khoảng cách của chúng với μ_0 sẽ bị đặt trong vùng bác bỏ. Bởi vì chúng ta vẫn muốn α = 0.05, cho nên diện tích trong vùng bác bỏ được chia đồng đều giữa hai phía của độ lệch chuẩn, như được thể hiện trong Hình 8.3. Bằng cách sử dụng trị thống kê kiểm định chuẩn hóa *z*, chúng ta sẽ bác bỏ H_0 nếu $z > 1.96$ hay $z < -1.96$. Với các giá trị khác nhau của α, thì các giá trị tới hạn của *z* mà phân cách các vùng bác bỏ và chấp nhận sẽ thay đổi.

HÌNH **8.3** Vùng bác bỏ của một kiểm định hai phía với α = 0.05



Kiểm định Thống kê cho Mẫu Lớn đối với μ

1. Giả thuyết không: $H_0 : \mu = \mu_0$
2. Giả thuyết thay thế:

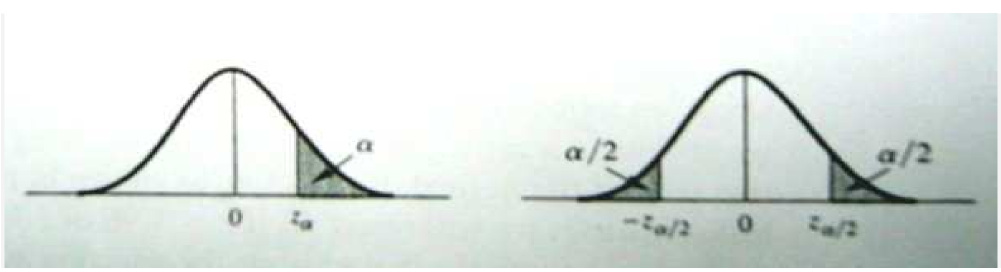
Kiểm định Một phía	**Kiểm định Hai phía**
$H_a : \mu > \mu_0$	$H_a : \mu \neq \mu_0$
(hay $H_a : \mu < \mu_0$)	

3. Trị thống kê kiểm định: $z = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Nếu σ là chưa được biết (thường là như vậy), thì thay thế độ lệch chuẩn của mẫu *s* cho σ.

4. Vùng bác bỏ:

Kiểm định Một phía	**Kiểm định Hai phía**
$z > z_\alpha$	$z > z_{\alpha/2}$ hay
(hay $z < -z_\alpha$ khi giả thuyết thay thế là $H_a : \mu < \mu_0$	$z < -z_{\alpha/2}$



Các giả thiết: Số *n* quan sát trong mẫu được chọn lựa ngẫu nhiên từ tổng thể, và *n* là lớn, ví dụ, *n* > 30.

VÍ DỤ **8.4** Sản lượng hàng ngày tại một nhà máy hóa chất, được ghi nhận cho $n = 50$ ngày, có một số trung bình và độ lệch chuẩn của mẫu là $\bar{x} = 871$ tấn và $s = 21$ tấn. Hãy kiểm định giả thuyết rằng sản lượng bình quân hàng ngày của nhà máy đó là $\mu = 880$ tấn mỗi ngày so với giả thuyết thay thế là μ hoặc lớn hơn hay nhỏ hơn 880 tấn mỗi ngày.

Lời giải Chúng ta muốn kiểm định giả thuyết không

$$H_0 : \mu = 880 \text{ tấn}$$

so với giả thuyết thay thế

$$H_a : \mu \neq 880 \text{ tấn}$$

Ước lượng điểm cho μ là $\bar{x}$. Do vậy, trị thống kê kiểm định là

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Đối với $\alpha = 0.05$, thì vùng bác bỏ là $z > 1.96$ hay $z < -1.96$ (được trình bày trong Hình 8.3). Bằng cách sử dụng s để ước lượng xấp xỉ σ, chúng ta có được

$$z = \frac{871 - 880}{21/\sqrt{50}} = -3.03$$

Bởi vì giá trị tính toán được của z rơi vào vùng bác bỏ, cho nên chúng ta bác bỏ giả thiết rằng $\mu = 880$ tấn. (Trên thực tế, có vẻ như sản lượng trung bình là nhỏ hơn 880 tấn mỗi ngày). Xác suất của việc bác bỏ H_0, qua việc giả định rằng điều này là đúng, chỉ là $\alpha = 0.05$. Vì vậy, chúng ta tin tưởng hợp lý rằng kết luận của chúng ta rằng $\mu \neq 880$ tấn là chính xác.

Kiểm định thống kê dựa trên trị thống kê kiểm định tuân theo phân phối chuẩn, với α đã biết, và khoảng tin cậy $(1-\alpha)100\%$ (Phần 7.4) là có liên quan với nhau một cách rõ rệt. Khoảng $\bar{x} \pm 1.96\sigma/\sqrt{n}$, hay xấp xỉ 871 ± 5.82 cho Ví dụ 8.4, được lập nên để cho trong việc chọn mẫu lặp lại $(1-\alpha)100\%$ của các khoảng sẽ bao quanh μ. Lưu ý rằng $\mu = 880$ không rơi vào khoảng này, nên chúng ta ắt sẽ nghiêng về việc bác bỏ $\mu = 880$ như là một giá trị có khả năng xảy ra và kết luận rằng sản lượng trung bình hàng ngày thật sự là khác với 880.

Có một sự tương đồng khác giữa sự kiểm định này với khoảng tin cậy trong Phần 7.4. Kiểm định này là "xấp xỉ" bởi vì chúng ta đã thay thế s, một giá trị xấp xỉ, cho σ. Nghĩa là, xác suất α của một sai lầm loại I được chọn lựa từ sự kiểm định này không phải *bằng đúng* 0.05, mà chỉ rất gần như vậy. Điều này sẽ đúng với nhiều kiểm định thống kê, bởi vì không phải tất cả các giả định sẽ được thỏa mãn *đúng như vậy*.

Bởi vì α là xác suất của việc bác bỏ H_0 khi giả thuyết này là đúng, cho nên đây là một đại lượng của cơ may *bác bỏ sai* H_0. Bởi vì β là xác suất của việc chấp nhận H_0 khi giả thuyết này là sai, cho nên phần bù của nó, $1-\beta$, là xác suất của việc bác bỏ H_0 khi giả thuyết này là sai và đo lường cơ may của việc *bác bỏ sai* H_0. Xác suất này, $1-\beta$, được gọi là **năng lực của sự kiểm định,** cơ may mà kiểm định này vận hành như nó phải như vậy.

ĐỊNH NGHĨA **Năng lực của sự kiểm định,** được cho bởi

$$1-\beta = P \text{ (bác bỏ } H_0 \text{ khi } H_0 \text{ là sai)}$$

đo lường khả năng của sự kiểm định có thể thực hiện như yêu cầu.

Một đồ thị của $1-\beta$, xác suất của việc bác bỏ H_0 khi trên thực tế H_0 là sai, là một hàm số của giá trị thực của tham số quan tâm được gọi là **đường cong năng lực** cho sự kiểm định thống kê này. Lý tưởng là chúng ta ắt sẽ muốn α nhỏ và **năng lực** ($1-\beta$) lớn. Người làm thí nghiệm phải có khả năng xác định cụ thể các giá trị của α và β, qua đó đo lường các rủi ro của những sai số liên quan mà anh/chị ta xem như có tầm quan trọng thiết thực và mong muốn tìm ra. Vùng bác bỏ cho sự kiểm định này sẽ được đặt ở vị trí phù hợp với giá trị được xác định cụ thể của α; cỡ mẫu sẽ được chọn lựa đủ lớn để đạt được một giá trị có thể chấp nhận được của β cho độ lệch được xác định cụ thể mà người làm thí nghiệm mong muốn tìm ra. Chọn lựa có thể được thực hiện bằng cách tham khảo các đường cong năng lực, tương ứng với các cỡ mẫu khác nhau, cho sự kiểm định được chọn.

VÍ DỤ **8.5** Tham khảo lại Ví dụ 8.4. Hãy tính xác suất β của việc chấp nhận H_0 nếu μ thật sự bằng với 870 tấn. Hãy tính năng lực của sự kiểm định, $1-\beta$.

Lời giải Vùng chấp nhận cho sự kiểm định trong Ví dụ 8.4 được đặt trong khoảng $\mu_0 \pm 1.96\sigma_{\bar{x}}$. Thay thế bằng các giá trị bằng số, chúng ta có được

$$880 \pm 1.96\left(\frac{21}{\sqrt{50}}\right)$$

hay từ

$$874.18 \quad \text{đến} \quad 885.82$$

Xác suất của việc chấp nhận H_0 nếu trong thực tế $\mu = 870$ là bằng với diện tích nằm phía dưới phân phối mẫu của trị thống kê kiểm định $\bar{x}$ phía trên khoảng từ 874.18 đến 885.82. Bởi vì $\bar{x}$ sẽ được phân phối chuẩn với trung bình bằng với 870 và $\sigma_{\bar{x}} \approx 21/\sqrt{50} = 2.97$, nên β bằng với diện tích nằm bên dưới đường cong chuẩn giữa 874.18 và 885.82 (xem Hình 8.4). Tính toán các giá trị *z* tương ứng với 874.18 và 885.82, chúng ta có được

$$z_1 = \frac{\bar{x}_1 - \mu}{\sigma/\sqrt{n}} \approx \frac{874.18 - 870}{21/\sqrt{50}} = 1.41$$

$$z_2 = \frac{\bar{x}_2 - \mu}{\sigma/\sqrt{n}} \approx \frac{885.82 - 870}{21/\sqrt{n}} = 5.33$$

Sau đó

$$\beta = P \text{ (chấp nhận } H_0 \text{ khi } \mu = 870)$$

$$= P(874.18 < \bar{x} < 885.82 \text{ khi } \mu = 870)$$

$$= P(1.41 < z < 5.33)$$

Bạn có thể thấy từ Hình 8.4 rằng diện tích nằm bên dưới đường cong chuẩn phía trên $\bar{x} = 885.82$ (hay $z = 5.33$) là không đáng kể. Vì thế

$$\beta = P(z > 1.41) = 0.5 - 0.4207 = 0.0793$$

và năng lực của sự kiểm định là

$$1-\beta = 1 - 0.0793 = 0.9297$$

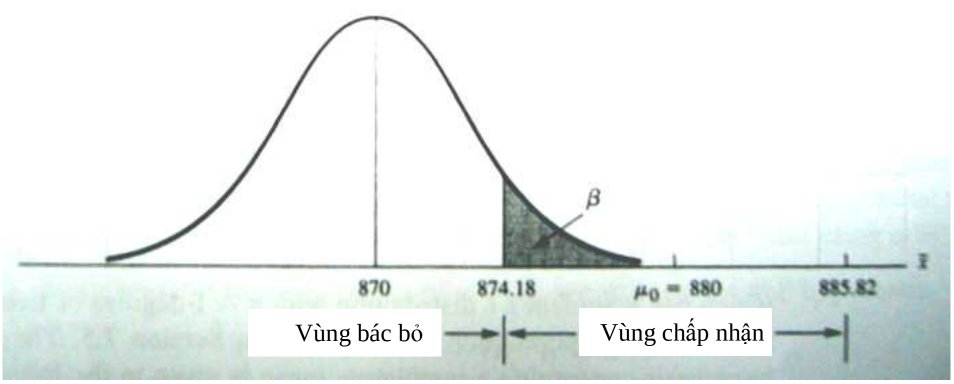
Xác suất của việc bác bỏ đúng H_0, khi đã biết rằng μ thật sự bằng với 870 là 0.9207, hay xấp xỉ 92 cơ may trên 100.

Các giá trị của $(1-\beta)$ có thể được tính cho những giá trị khác nhau của μ_a khác với $\mu_0 = 880.$ Ví dụ, nếu $\mu_0 = 885,$

$$\beta = P(874.18 < \bar{x} < 885.82 \text{ khi } \mu = 885)$$
$$= P(-3.64 < z < 0.28) = 0.5 + 0.1103 = 0.6103$$

và năng lực là $(1-\beta) = 0.3897.$ Bảng 8.3 cho thấy năng lực của sự kiểm định cho những giá trị khác nhau của μ_a, và một đường cong được vẽ trong Hình 8.5. Lưu ý rằng năng lực của sự kiểm định tăng lên khi khoảng cách giữa μ_a và μ_0 gia tăng. Kết quả là một đường cong hình chữ U cho sự kiểm định hai phía này.

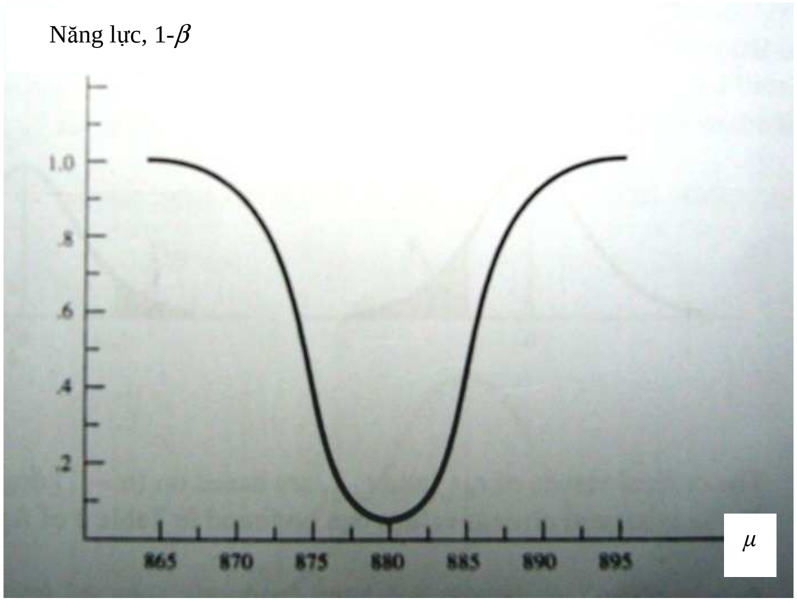
HÌNH **8.4** Tính toán β trong Ví dụ 8.5



BẢNG **8.3** Các giá trị của $(1-\beta)$ cho những giá trị khác nhau của μ_a, Ví dụ 8.5

μ_a	$(1-\beta)$	μ_a	$(1-\beta)$
865	0.9990	883	0.1726
870	0.9207	885	0.3897
872	0.7673	888	0.7673
875	0.3897	890	0.9207
877	0.1726	895	0.9990
880	0.0500		

HÌNH **8.5** Đường cong năng lực cho Ví dụ 8.5



Kiểm định cho Mẫu Nhỏ

Khi cỡ mẫu là nhỏ và độ lệch chuẩn tổng thể σ là chưa được biết, thì sự kiểm định giả thuyết về một số trung bình tổng thể μ được dựa trên trị thống kê kiểm định

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

mà có một phân phối *t* Student với *n* - 1 bậc tự do khi chọn mẫu từ một tổng thể phân phối chuẩn, như được mô tả trong Phần 7.5. Sự kiểm định thống kê của một giả thuyết liên quan đến một số trung bình tổng thể được thể hiện trong phần trình bày sau. Lưu ý rằng các vùng bác bỏ được tìm thấy khi sử dụng các giá trị tới hạn của *t* được cho trong Bảng 4 của Phụ lục II.

Kiểm định cho Mẫu Nhỏ về một Giả thuyết Liên quan đến Trung bình Tổng thể

Kiểm định Thống kê cho Mẫu Lớn của μ

1. Giả thuyết không: $H_0 : \mu = \mu_0$

2. Giả thuyết thay thế:

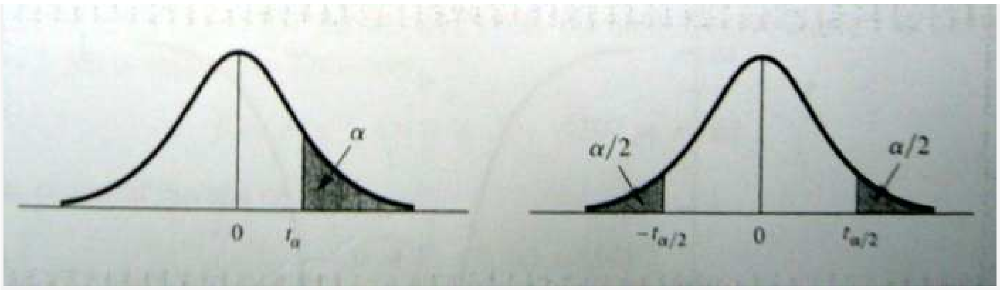
Kiểm định Một phía	**Kiểm định Hai phía**
$H_a : \mu > \mu_0$	$H_a : \mu \neq \mu_0$
(hay $H_a : \mu < \mu_0$)	

3. Trị thống kê kiểm định: $t = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

Nếu σ là chưa được biết (thường là như vậy), thì thay thế độ lệch chuẩn của mẫu *s* cho σ.

4. Vùng bác bỏ:

Kiểm định Một phía	**Kiểm định Hai phía**
$t > t_{\alpha}$	$t > t_{\alpha/2}$ hay
(hay $t < -t_{\alpha}$ khi giả thuyết thay thế là $H_a : \mu < \mu_0$	$t < -t_{\alpha/2}$



Các giá trị tới hạn của t, t_{α}, và $t_{\alpha/2}$, được căn cứ vào (*n* - 1) bậc tự do. Những giá trị tới hạn được lập trong bảng tính này có thể được tìm thấy trong Bảng 4 của Phụ lục II.

Các giả thiết: Mẫu này được chọn lựa ngẫu nhiên từ một tổng thể có phân phối chuẩn.

VÍ DỤ **8.6** Trong Ví dụ 7.4, chúng ta đã khảo sát một thí nghiệm được thiết ké nhằm đánh giá một qui trình mới cho sản xuất kim cương nhân tạo. Một nghiên cứu về các chi phí của qui trình này cho thấy rằng trọng lượng trung bình của các viên kim cương phải lớn hơn 0.5 cara nhằm để cho qui trình này hoạt động ở một mức có khả năng thu được lợi nhuận. Liệu trọng lượng của sáu viên kim cương tổng hợp, 0.46, 0.61, 0.52, 0.48, 0.57, và 0.54 cara, có cung cấp đủ bằng chứng để chỉ rà rằng trọng lượng trung bình của kim cương được sản xuất ra bởi qui trình này có vượt quá 0.5 cara? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

Lời giải Bởi vì chúng ta muốn phát hiện các giá trị của $\mu > 0.5$, nên chúng ta sẽ kiểm định giả thuyết không

$$H_0 : \mu = 0.5$$

so với giả thuyết thay thế

$$H_a : \mu > 0.5$$

Trị thống kê kiểm định là

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Bởi vì chúng ta mong muốn chỉ phát hiện các giá trị lớn của μ, nên chúng ta sẽ thực hiện một sự kiểm định một phía cao. Vùng bác bỏ cho kiểm định này đối với $\alpha = 0.05$ và (*n* - 1) = (6 - 1) = 5 bậc tự do là $t > 2.015$. Đây là giá trị của *t*, được cho trong Bảng 4 của Phụ lục II, mà thay thế $\alpha = 0.05$ ở phía lớn hơn của phân phối *t* (xem Hình 8.6). Trung bình và độ lệch chuẩn của mẫu cho trọng lượng của sáu viên kim cương là

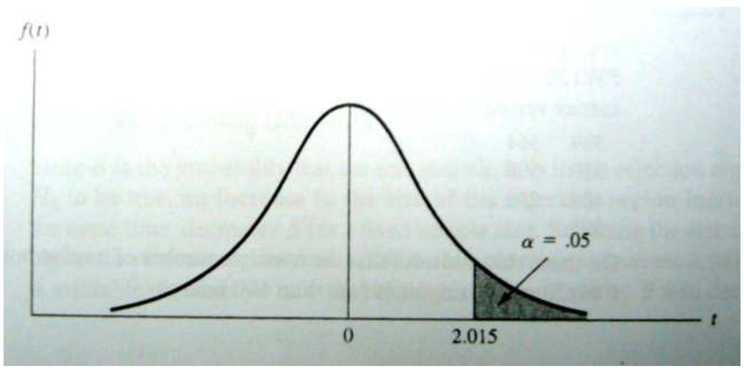
$$\bar{x} = 0.53 \quad \text{và} \quad s = 0.0559$$

Thay thế các giá trị này vào công thức tính trị thống kê kiểm định, chúng ta thu được

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{0.53 - 0.5}{0.0559\sqrt{6}} = 1.31$$

Bởi vì giá trị tính toán được của trị thống kê kiểm định không rơi vào trong vùng bác bỏ, cho nên chúng ta không bác bỏ H_0. Sự không bác bỏ của H_0 hàm ý rằng dữ liệu này chưa đại diện đủ bằng chứng để cho thấy trọng lượng trung bình của kim cương vượt quá 0.5 cara.

HÌNH **8.6** Vùng bác bỏ cho kiểm định trong Ví dụ 8.6



Sự tính toán xác suất của một sai lầm loại II, β, cho kiểm định *t* là rất khác biệt và vượt quá phạm vi của cuốn sách này. Tuy vậy, chúng ta có thể đạt được một ước lượng khoảng cho μ (xem Phần 7.5) để xác định một dãy các giá trị có thể có cho μ. Nếu người làm thí nghiệm quan tâm đến việc giảm giá trị của β, anh hay chị nên gia tăng cỡ mẫu.

Chuỗi lệnh Minitab, **Stat → Basic Statistics → 1-Sample t** được sử dụng cho sự kiểm định đối với mẫu nhỏ về một số trung bình tổng thể; chuỗi này cũng có thể được sử dụng nhằm tạo ra một khoảng tin cậy cho mẫu nhỏ. Người sử dụng phải bấm vào *nút* có đánh dấu "Test mean", nhập giá trị của số trung bình tổng thể cần được kiểm định, và giả thuyết thay thế phù hợp. Trong hộp **Variables,** đánh hay chọn cột mà trong đó lưu trữ dữ liệu này, và bấm **OK.**

Kết quả Minitab cho một kiểm định *t* cho một mẫu sử dụng dữ liệu trong Ví dụ 8.6 được trình bày trong Bảng 8.4. Ngoài giá trị quan sát được của trị thống kê kiểm định, *t* = 1.32, thì kết quả còn cho ra số trung bình của mẫu, độ lệch chuẩn của mẫu, và sai số chuẩn của số trung bình (SE MEAN = $s/\sqrt{n}$). So sánh với các kết quả của chúng ta trong Ví dụ 8.6, thì sự khác biệt duy nhất là trong độ chính xác thập phân được báo cáo cho các kết quả.

BẢNG **8.4** Kết quả Minitab cho dữ liệu cho dữ liệu của Ví dụ 8.6

```
KI•M ••NH V• MU = 0.5000 VS MY FT 0.500
            N          TRUNG      •Ô L•CH    TRUNG      T          GIÁ TR• P
                       BÌNH       CHU•N      BÌNH SAI
                                             S• CHU•N
C1          6          0.5300     0.0559     0.0228     1.32       0.12

MTB > KHO•NG TIN C•Y 95 C1
            N          TRUNG      •Ô L•CH    TRUNG      95.0 PH•N TR•M C.I.
```

```
                      BÌNH       CHU•N      BÌNH SAI
                                            S• CHU•N
          6           0.5300     0.0559     0.0228      ( 0.4714.   0.5886)
MTB >
```

Để lập nên một khoảng tin cậy cho số trung bình tổng thể, bấm vào *nút* có tựa đề “Confidence interval”. Đánh vào mức tin cậy mong muốn, và chọn cột có chứa dữ liệu đó. Khoảng tin cậy 95% cho μ qua việc sử dụng dữ liệu trong Ví dụ 8.6 cũng được trình bày trong Bảng 8.4, và các kết quả này nhất quán với những kế quả được cho trong Ví dụ 8.4.

VÍ DỤ **8.7** Ủy ban Quản trị Hàng không Liên bang (FAA) cung cấp một báo cáo hàng tháng về việc sử dụng máy bay và độ tin cậy của động cơ phản lực đối với các phi đội bay của Hoa Kỳ. Dữ liệu sau đây cho ta số giờ động cơ bình quân tính trên mỗi may bay được trang bị bằng động cơ tua-bin phản lực mẫu PW120 của Pratt và Whitney cho mỗi trong số n = 7 hãng hàng không. (Số lượng máy bay của mỗi hãng hàng không thay đổi từ 5 đến 15).

389	364
359	308
408	295
393	

Liệu những dữ liệu này có chỉ ra rằng số giờ động cơ bình quân của mỗi máy bay khi sử dụng mẫu động cơ PW120 là ít hơn 400 giờ không?

Lời giải Kiểm định giả thuyết không rằng μ = 400 giờ động cơ bình quân so với giả thuyết thay thế rằng μ là nhỏ hơn 400 giờ sẽ tạo ra một sự kiểm định một phía. Như vậy

$$H_0 : \mu = 400$$
$$H_a : \mu < 400$$

trong đó μ là số giờ động cơ bình quân của mỗi máy bay. Sử dụng $\alpha = 0.05$ và thay thế 0.05 vào phía thấp của phân phối *t*, chúng ta tìm thấy giá trị tới hạn của *t* đối với n = 7 thước đo (hay đối với n - 1 = 6 bậc tự do) là t = 1.943. Vì vậy, chúng ta sẽ bác bỏ H_0 nếu t < -1.943 (xem Hình 8.7).

Bạn có thể xác minh rằng số trung bình và độ lệch chuẩn của mẫu đối với n = 7 thước đo trong bảng này là

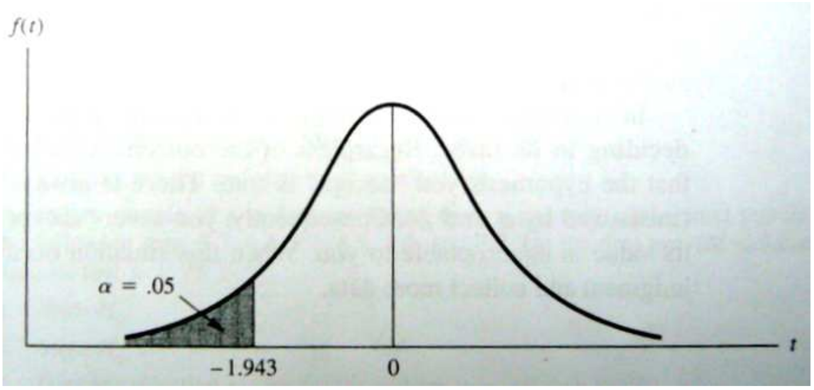
$$\bar{x} = 359.43 \qquad \text{và} \qquad s = 43.16$$

Thay thế các giá trị này vào công thức tính trị thống kê kiểm định cho ra

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{359.43 - 400}{43.16\sqrt{n}} = -2.487$$

Bởi vì giá trị quan sát được của *t* rơi vào vùng bác bỏ, cho nên có đủ bằng chứng để nói lên rằng số giờ động cơ bình quân là ít hơn 400. Hơn nữa, chúng ta sẽ tin tưởng hợp lý rằng chúng ta đã thực hiện quyết định chính xác. Khi sử dụng qui trình của mình, chúng ta phải bác bỏ nhầm H_0 chỉ với $\alpha = 0.05$ của thời gian trong các áp dụng lặp lại của bài kiểm định thống kê này.

HÌNH **8.7** Vùng bác bỏ trong Ví dụ 8.7



Các Bình luận Liên quan đến Những Tỷ lệ Sai lầm

Bởi vì α là xác suất để cho trị thống kê kiểm định này rơi vào vùng bác bỏ, nên khi giả thuyết H_0 là đúng, thì **một sự tăng lên trong kích cỡ của vùng bác bỏ làm gia tăng** α và, cùng lúc đó, làm giảm β đối với một cỡ mẫu cố định. Việc giảm bớt kích cỡ của vùng bác bỏ làm giảm α và làm tăng β. Nếu cỡ mẫu *n* được tăng lên, thì nhiều thông tin hơn là sẵn có về điều để làm căn cứ cho quyết định, và, đối với α cố định, thì β sẽ giảm đi.

Xác suất β cho việc tạo ra một sai lầm loại II thay đổi tùy thuộc vào giá trị thực sự của tham số tổng thể. Ví dụ, giả định rằng chúng ta mong muốn kiểm định giả thuyết không rằng tham số nhị thức *p* là bằng với $p_0 = 0.4$. Hơn nữa, giả định rằng H_0 là sai và rằng *p* thật sự bằng với một giá trị thay thế, ví dụ, p_a. Vậy thì đại lượng này sẽ dễ dàng được xác minh hơn, một $p_a = 0.4001$ hay một $p_a = 1.0$? Chắc chắn là, nếu *p* thật sự bằng với 1.0, thì tất cả các lần thử nghiệm duy nhất sẽ có kết quả là thành công và các kết quả mẫu sẽ tạo ra bằng chứng mạnh trong việc ủng hộ $H_0: p = 0.4$. Mặt khác, $p_a = 0.4001$ nằm rất gần với $p_0 = 0.4$ đến nỗi thật vô cùng khó khăn trong việc xác minh p_a mà không có một mẫu rất lớn. Nói cách khác, xác suất β của việc chấp nhận H_0 sẽ thay đổi tùy thuộc vào sự khác biệt giữa giá trị thực sụ của *p* và giá trị được giả định p_0. Lý tưởng là p_a càng nằm cách xa p_0, thì xác suất của việc bác bỏ H_0 càng cao. Xác suất này được đo bằng $1-\beta$, mà được gọi là **năng lực** của sự kiểm định.

Đối với các giá trị không đổi của *n* và α, **năng lực của sự kiểm định phải tăng lên khi khoảng cách giữa các giá trị thực sự và được giả định của tham số gia tăng.** Một sự gia tăng trong cỡ mẫu *n* sẽ làm tăng năng lực, $1-\beta$, đối với tất cả các giá trị khác của tham số đang được kiểm định. Như vậy, chúng ta có thể tạo ra một đường cong năng lực tương ứng cho từng cỡ mẫu.

Trên thực tế, β thường là chưa được biết, hoặc là bởi vì tham số này chưa bao giờ được tính trước khi thực hiện sự kiểm định hoặc bởi vì có lẽ việc tính toán cho sự kiểm định này là điều vô cùng khó khăn. Vì vậy, thay vì chấp nhận giả thuyết không khi trị thống kê kiểm định rơi vào vùng chấp nhận, thì các bạn nên từ chối việc đánh giá. Nghĩa là, bạn không nên chấp nhận giả thuyết không trừ phi bạn biết rủi ro (được đo bằng β) của việc tạo ra một quyết định không chính xác. Lưu ý rằng bạn sẽ không bao giờ bị đối mặt với tình huống "không có kết luận" khi trị thống kê kiểm định rơi vào vùng bác bỏ. Sau đó bạn có thể bác bỏ giả thuyết không (và chấp nhận giả thuyết thay thế), bởi vì bạn luôn luôn biết giá trị của α, xác suất của việc bác bỏ giả thuyết không khi giả thuyết này là đúng. Sự kiện rằng β thường chưa được biết giải thích lý do tại sao chúng ta cố gắng ủng hộ giả thuyết thay thế bằng cách bác bỏ giả thuyết không. Khi chúng ta đi đến quyết định này, thì xác suất α mà một quyết định như vậy là không chính xác đã được biết.

Kết luận là hãy ghi nhớ rằng **"chấp nhận" một giả thuyết thay thế có nghĩa là quyết định ủng hộ giả thuyết đó.** Bất luận kết quả của một sự kiểm định có thế nào, thì bạn không bao giờ *chắc chắn* rằng giả định mà bạn "chấp nhận" là đúng. **Luôn luôn có một rủi ro của sự sai lầm (được đo bằng α và β).** Vì vậy, bạn không bao giờ "chấp nhận" H_0 nếu β là chưa được biết hay giá trị của nó là không thể chấp nhận được với bạn. Khi tình huống này xảy ra, bạn nên từ chối việc đánh giá và thu thập thêm dữ liệu.

Bài tập

Các Kỹ thuật Cơ bản

8.1 Một mẫu ngẫu nhiên gồm $n = 35$ quan sát từ một tổng thể tạo ra một số trung bình $\bar{x} = 2.4$ và một độ lệch chuẩn bằng với 0.29. Giả định rằng bạn mong muốn chứng minh rằng số trung bình tổng thể μ vượt quá 2.3

a. Tìm giả thuyết thay thế cho sự kiểm định này.

b. Tìm giả thuyết không cho sự kiểm định này.

c. Nếu bạn mong muốn rằng xác suất của việc quyết định (sai lầm) rằng $\mu > 2.3$ trong khi trên thực tế $\mu = 2.3$, bằng với 0.05, thì giá trị của α cho kiểm định này là bao nhiêu?

d. Trước khi bạn thực hiện bài kiểm định này, hãy nhìn qua dữ liệu và sử dụng trực giác của bạn để quyết định liệu số trung bình mẫu $\bar{x} = 2.4$ có hàm ý rằng $\mu > 2.3$ hay không. Bây giờ hãy kiểm định giả thuyết không. Liệu dữ liệu có cung cấp bằng chứng đủ để chỉ ra rằng $\mu > 2.3$ hay không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

8.2 Tham khảo lại Ví dụ 8.1. Giả định rằng bạn mong muốn chứng minh rằng dữ liệu mẫu này ủng hộ cho giả thuyết rằng số trung bình của tổng thể này là nhỏ hơn 2.9. Tìm các giả thuyết không và thay thế cho sự kiểm định này. Liệu bài kiểm định này có phải là kiểm định một phía hoặc hai phía không? Hãy giải thích.

8.3 Tham khảo lại Ví dụ 8.1 và 8.2. Giả định rằng bạn mong muốn xác minh một giá trị của μ mà khác với 2.9, nghĩa là, một giá trị của μ hoặc lớn hơn hay nhỏ hơn 2.9. Hãy xác định các giả thuyết không và thay thế cho sự kiểm định này. Liệu giả thuyết thay thế có đơn giản hàm ý về một sự kiểm định một phía hay hai phía không?

8.4 Một mẫu ngẫu nhiên gồm $n = 40$ quan sát từ một tổng thể tạo ra trung bình $\bar{x} = 83.8$ và một độ lệch chuẩn bằng với 2.9. Giả định rằng bạn mong muốn chứng minh rằng số trung bình tổng thể μ là nhỏ hơn 84.

a. Tìm giả thuyết thay thế cho sự kiểm định này.

b. Tìm giả thuyết không cho sự kiểm định này.

c. Nếu bạn mong muốn xác suất của việc quyết định (sai lầm) rằng $\mu < 84$, trong khi trên thực tế $\mu > 84$, bằng với 0.05, thì giá trị của α cho kiểm định này là bao nhiêu?

d. Trước khi bạn thực hiện bài kiểm định này, hãy nhìn qua dữ liệu và sử dụng trực giác của bạn để quyết định liệu số trung bình mẫu $\bar{x} = 84$ có hàm ý rằng $\mu < 84$ hay không. Bây giờ hãy kiểm định giả thuyết không. Liệu dữ liệu có cung cấp bằng chứng đủ để chỉ ra rằng $\mu < 84$ hay không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

8.5 Tham khảo lại Bài tập 8.4, trong đó $H_0: \mu = 84$ được kiểm định so với $H_a: \mu < 84$.

a. Tìm giá trị tới hạn của $\bar{x}$ cần thiết cho việc bác bỏ H_0.

b. Tính toán $\beta = P$ [bác bỏ H_0 khi $\mu = 82.8$]. Lặp lại tính toán này cho $\mu = 82.4, 82.6,$ và 83.4.

c. Sử dụng các giá trị của β tính toán được trong câu (b) để vẽ đồ thị cho đường cong năng lực đối với kiểm định này.

8.6 Một mẫu ngẫu nhiên gồm n = 4 quan sát từ một tổng thể được phân phối chuẩn tạo ra dữ liệu sau đây: 9.4, 12.2, 10.7, và 11.6. Liệu dữ liệu này có cung cấp bằng chứng đủ để cho biết rằng $\mu > 10$?

a. Xác định H_a.

b. Xác định H_0.

c. Tìm vùng bác bỏ cho kiểm định của $\alpha = 0.10$.

d. Thực hiện bài kiểm định này và phát biểu các kết luận của bạn.

8.7 Một mẫu ngẫu nhiên gồm n = 6 quan sát từ một tổng thể tuân theo phân phối chuẩn tạo ra dữ liệu sau đây: 3.7, 6.4, 8.1, 8.8, 4.9, và 11.6. Liệu dữ liệu này có cung cấp bằng chứng đủ để cho biết rằng $\mu < 7$?

a. Xác định H_a.

b. Xác định H_0.

c. Tìm vùng bác bỏ cho kiểm định với $\alpha = 0.10$.

d. Thực hiện bài kiểm định này và phát biểu các kết luận của bạn.

8.8 Kiểm định giả thuyết không: $H_0: \mu = 3$ so với $H_a: \mu > 3$ với $\alpha = 0.05, n = 12, \bar{x} = 31.8,$ và $s^2 = 0.21$.

8.9 Kiểm định giả thuyết không: $H_0: \mu = 48$ so với $H_a: \mu \neq 48$ với $\alpha = 0.10, n = 25, \bar{x} = 47.1,$ và $s^2 = 4.7$.

8.10 Bản in Minitab sau đây tạo ra khi chuỗi lệnh **Stat → Basic Statistics → 1-Sample t** được sử dụng đối với một tập hợp dữ liệu được lưu trữ trong ô C1. Sử dụng bảng in này để xác định tất cả bốn phần của một sự kiểm định thống kê về giả thuyết và rút ra những kết luận phù hợp cho $\alpha = 0.01$.

```
KI•M ••NH MU = 5.0000 VS MU G.T 5.0000
                    N      TRUNG     •• L•CH      TRUNG          T  GIÁ TR• P
                           BÌNH       CHU•N   BÌNH SAI
                                              S• CHU•N
C1                 11      5.364      1.502      0.453       0.80       0.20
```

Các Ứng dụng

8.11 Một nhà quản lý tài sản khẳng định rằng sự lựa chọn cổ phiếu phổ thông của cô ta cho khoản đầu tư tính chung sẽ cao hơn sự thay đổi hàng năm trong mức bình quân cổ phiếu của Standard & Poor. Một chọn lựa ngẫu nhiên gồm ba sự lựa chọn cổ phiếu của nhà quản lý này cho thấy các khoản gia tăng hàng năm là 22%, 12% và 31% so với một sự gia tăng trong mức bình quân của Standard & Poor là 19%. Liệu mẫu gồm ba lựa chọn cổ phiếu này có cung cấp bằng chứng đủ để cho thấy rằng sự gia tăng trung bình trong tất cả chọn lựa cổ phiếu của nhà quản lý đó vượt quá mức 19% không?

a. Xác định H_a.

b. Xác định H_0.

c. Tìm vùng bác bỏ cho kiểm định với $\alpha = 0.05$.

d. Thực hiện bài kiểm định này và phát biểu các kết luận của bạn.

8.12 Một đại lý bán xe hơi mới đã tính toán rằng công ty phải đạt được mức lợi nhuận bình quân cao hơn 4.8% về doanh số bán hàng các xe hơi mới được phân bổ cho công ty. Một mẫu ngẫu nhiên gồm n = 80 chiếc xe cho ta một trung bình và độ lệch chuẩn của lợi tức tính theo tỷ lệ phần trăm mỗi chiếc xe là $\bar{x} = 4.87\%$ và s = 3.9%. Liệu dữ liệu này có cung cấp bằng chứng đủ để chỉ ra rằng chính sách của nhà quản lý bán hàng này trong việc cải thiện giá cả bán hàng đang đạt được một mức lợi tức trung bình vượt quá 4.8% mỗi năm?

a. Xác định giả thuyết thay thế rằng nhà quản lý bán hàng này mong muốn chứng tỏ là có thật.

b. Khảo cứu dữ liệu này. Chỉ từ trực giác của bạn, bạn có nghĩ rằng dữ liệu này ủng hộ cho giả thuyết thay thế của câu (a)?

c. Xác định giả thuyết không cần được kiểm tra.

d. Người chủ của công ty muốn chắc chắn một cách hợp lý rằng quyết định này là chính xác nếu như trên thực tế dữ liệu này chứng tỏ rằng công ty đang hoạt động ở một mức lợi nhận có thể chấp nhận được. Để đạt được điều này, người chủ công ty muốn kiểm định giả thuyết không bằng cách sử dụng $\alpha = 0.01$. Hãy giải thích sự chọn lựa này đối với α sẽ đạt được mục tiêu của người chủ này như thế nào?

e. Tìm vùng bác bỏ cho sự kiểm định này.

f. Thực hiện sự kiểm định này và phát biểu các kết luận của bạn theo cách thức có thể hiểu được đối với người chủ công ty này. So sánh câu trả lời của bạn với phần đoán trực quan trong câu (b).

8.13 Một công ty sản xuất móc khóa kim loại kỳ vọng việc giao hàng một mức bình quân là 1200 hộp móc khóa mỗi ngày. Một sự phân tích về các chuyến giao hàng này trong 30 ngày vừa qua cho ra một số trung bình $\bar{x} = 1186$ hộp mỗi ngày và một phương sai $\sigma^2 = 2480\,(\text{hộp})^2$ mỗi ngày. Liệu dữ liệu này có cung cấp bằng chứng đủ để cho thấy rằng nhu cầu trung bình hàng ngày đối với móc khóa đang giảm đi, nghĩa là, thấp hơn 1200 hộp mỗi ngày?

a. Tìm giả thuyết thay thế rằng công ty sản xuất này mong muốn xác minh.

b. Khảo cứu dữ liệu này. Từ trực giác của mình, bạn có nghĩ rằng dữ liệu này ủng hộ cho giả thuyết thay thế của câu (a)?

c. Tìm H_0 cho sự kiểm định này.

d. Tìm vùng bác bỏ cho kiểm định này với $\alpha = 0.10$.

e. Thực hiện sự kiểm định này và phát biểu các kết luận thực tiễn được rút ra từ sự kiểm định này. So sánh câu trả lời của riêng bạn với câu trả lời của bạn cho câu (b).

8.14 Sự tăng trưởng vô cùng mạnh mẽ của ngành công nghiệp tôm hùm (được gọi là tôm hùm có gai) Florida trong 20 năm qua đã làm cho ngành này trở thành ngành công nghiệp thủy sản đáng giá thứ hai của tiểu bang này. Nhiều năm trước đây, một tuyên bố của chính phủ Bahamas mà ngăn cấm những ngư dân săn tôm hùm của Mỹ không được đánh bắt trên vùng thềm lục địa của Bahamas được kỳ vọng là sẽ tạo ra một sự giảm sút nghiêm trọng về trọng lượng của những lần kéo vào bờ tính bằng pao mỗi con tôm hùm mỗi lồng bẫy. Theo hồ sơ lưu trữ, trọng lượng trung bình mỗi lần kéo vào bờ tính bằng pao là 30.31 pao. Một sự chọn mẫu ngẫu nhiên gồm 20 lồng bẫy tôm hùm kể từ khi sự hạn chế đánh bắt của Bahamas có hiệu lực đã cho ta kết quả sau đây (tính bằng pao):

17.4	18.9	39.6	34.4	19.6
33.7	37.2	43.4	41.7	27.5
24.1	39.6	12.2	25.5	22.1
29.3	21.1	23.8	43.2	24.4

Liệu những trọng lượng đưa vào bờ này có cung cấp đủ bằng chứng để ủng hộ cho luận điểm rằng trọng lượng đưa vào bờ trung bình mỗi lồng bẫy đã giảm đi sau khi có sự áp đặt lệnh cấm của chính phủ Bahamas không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

8.15 Trong Bài tập 2.13, chúng ta đã báo cáo rằng chiếc xe Tropica có một quãng đường đi được bình quân là 44.7 dặm, theo Ủy ban Tài nguyên Không khí California. Một sự kiểm tra độc lập bao gồm 30 lần thử tạo ra một mẫu các quãng đường đi được cho xe Tropica. Bản in Minitab sau đây có được từ việc sử dụng 30 quan sát này.

```
KI•M ••NH MU = 44.7000 VS MU N.E 44.7000
                    N      TRUNG     •• L•CH      TRUNG          T  GIÁ TR• P
                           BÌNH      CHU•N   BÌNH SAI
                                             S• CHU•N
CÁC QUÃNG          30     45.273      1.199      0.219       2.62       0.014
•••NG •I
```

a. Dựa vào bản in Minitab này, hãy giải thích qui trình kiểm định giả thuyết phù hợp để xác minh hay tranh luận về quãng đường đi được 44.7 dặm này. Hãy sử dụng $\alpha = 0.05$.

b. Tham khảo lại Bài tập 7.100. Liệu những kết quả trong câu (a) có nhất quán với những kết luận mà bạn có được trong bài tập đó? Hãy giải thích.

8.4 MỘT CÁCH THỨC KHÁC ĐỂ BÁO CÁO CÁC KẾT QUẢ VỀ NHỮNG KIỂM ĐỊNH THỐNG KÊ: CÁC GIÁ TRỊ *p*

Xác suất α của việc tạo ra một sai lầm loại I thường được gọi là mức ý nghĩa của một sự kiểm định thống kê, bởi vì chúng ta tuyên bố một sự khác biệt có ý nghĩa nếu như giá trị quan sát được của trị thống kê kiểm định rơi vào vùng bác bỏ được xác định bởi H_a và giá trị của α. Một số người làm thí nghiệm ưa thích việc sử dụng một mức ý nghĩa thay đổi. Ví dụ, nếu trong một sự kiểm định hai phía mà giá trị quan sát được của *z* trở thành $z_{0.025} = 1.96$, nhưng không thể bị bác bỏ nếu như $\alpha = 0.01$, bởi vì z = 2.03 là lớn hơn $z_{0.005} = 2.58$. Điều này ắt sẽ được báo cáo bằng cách nói rằng các kết quả kiểm định là có ý nghĩa ở mức ý nghĩa 5% nhưng không có ý nghĩa ở mức ý nghĩa 1%. Các nhà làm thí nghiệm khác ưa thích việc báo cáo các kết quả của mình bằng cách cung cấp giá trị nhỏ nhất của α mà qua đó những kết luận kiểm định là có ý nghĩa. Nếu chúng ta đã sử dụng các giá trị tới hạn của z bằng với ± 2.03, thì chúng ta ắt đã bác bỏ H_0, và giá trị của α mà chúng ta sử dụng ắt là

$$\begin{aligned} P(z \le -2.03) + P(z \ge 2.03) &= 2P(z \ge 2.03) \\ &= 2(0.5 - 0.4788) \\ &= 2(0.0212) = 0.0424 \end{aligned}$$

Giá trị này được gọi là **giá trị *p*[+]** hay **mức ý nghĩa quan sát được** của sự kiểm định.

ĐỊNH NGHĨA Giá trị *p* hay **mức ý nghĩa quan sát được** là giá trị nhỏ nhất của α mà qua đó các kết quả kiểm định là có ý nghĩa về mặt thống kê.

Một số chương trình máy tính thống kê tính toán các giá trị *p* cho những kiểm định thống kê chính xác đến bốn hay năm chữ số thập phân. Nhưng nếu bạn sử dụng các bảng thống kê để xác định một giá trị *p*, thì bạn sẽ chỉ có thể ước lượng xấp xỉ giá trị của nó, bởi vì phần lớn các bảng thống kê cho ta những giá trị tới hạn chỉ với các giá trị khác biệt rất lớn của α (ví dụ, 0.01, 0.025, 0.05, 0.10, v.v). Vì thế, giá trị *p* được báo cáo bởi hầu hết các nhà làm thí nghiệm là giá trị được lập bảng nhỏ nhất của α mà từ đó sự kiểm định vẫn **có ý nghĩa về mặt thống kê.** Ví dụ, nếu một kết quả kiểm định là có ý nghĩa về mặt thống kê đối với $\alpha = 0.10$ nhưng không có ý nghĩa khi $\alpha = 0.05$, thì giá trị *p* cho sự kiểm định này ắt sẽ được cho một giá trị *p* = 0.10, hay chính xác hơn là

$$0.05 < \text{giá trị } p < 0.10$$

Một cách thức khác để sử dụng giá trị *p* trong việc ra quyết định là bác bỏ H_0 nếu giá trị *p* là *ít hơn* giá trị của α, bởi vì điều này chỉ xảy ra khi giá trị quan sát được của trị thống kê kiểm định rơi vào vùng bác bỏ. Ví dụ, nếu trong một kiểm định một phía bên phải với $\alpha = 0.05$, giá trị quan sát được của một trị thống kê z là 2.04, thì giá trị *p* cho sự kiểm định này ắt sẽ là $P(z) > 2.04) = 0.0207$. Bởi vì giá trị *p* của 0.0207 là nhỏ hơn 0.05, cho nên chúng ta có thể bác bỏ H_0, khi biết rằng giá trị z = 2.04 nằm trong vùng bác bỏ khi $\alpha = 0.05$.

Nhiều tạp chí khoa học yêu cầu các nhà nghiên cứu phải báo cáo các giá trị *p* đi cùng với những kiểm định thống kê bởi vì các giá trị này sẽ cung cấp cho độc giả *nhiều thông tin* hơn là đơn giản khẳng định rằng một giả thuyết không sẽ bị hay không bị bác bỏ đối với một giá trị nào đó của α mà nhà làm thí nghiệm chọn lựa. Về một ý nghĩa nào đó, điều này cho phép một người đọc bài nghiên cứu được xuất bản đánh giá mức độ mà qua đó dữ liệu không phù hợp với giả thuyết không. Cụ thể, nó cho phép mỗi độc giả có thể chọn lựa giá trị riêng của anh/chị ta đối với α và sau đó quyết định liệu có dẫn đến việc bác bỏ giả thuyết không hay không.

Qui trình cho việc tìm kiếm giá trị *p* của một kiểm định được minh họa trong các ví dụ sau đây.

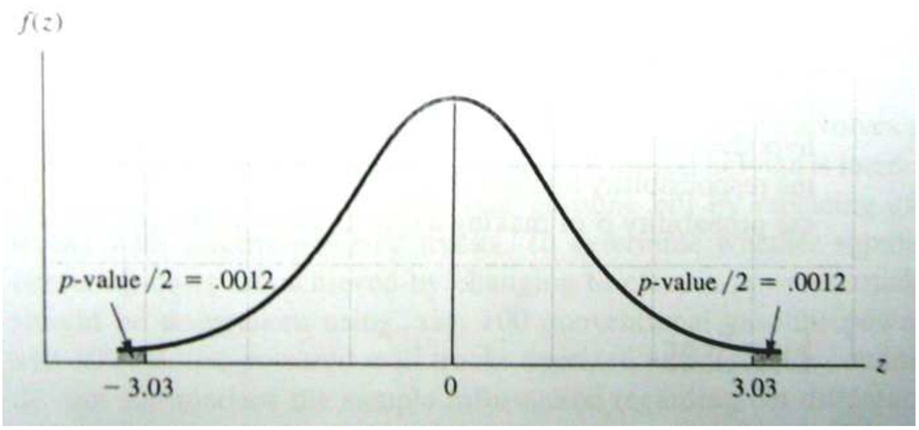
[+] Những người sử dụng các trị thống kê gọi mức ý nghĩa quan sát được là một "giá trị xác suất hay giá trị *p*". Ký hiệu *p* trong thành ngữ này không có liên hệ nào với tham số nhị thức *p*.

VÍ DỤ **8.8** Tìm giá trị *p* cho kiểm định thống kê trong Ví dụ 8.4. Giải thích kết quả của mình.

Lời giải Ví dụ 8.4 trình bày một bài kiểm định cho giả thuyết không $H_0 : \mu = 880$ so với giả thuyết thay thế $H_a : \mu \neq 880$. Giá trị của trị thống kê kiểm định này, được tính từ dữ liệu mẫu, là $z = -3.03$. Vì vậy, giá trị *p* cho sự kiểm định hai phía này là xác suất để cho $z \leq -3.03$ hay $z \geq 3.03$ (xem Hình 8.8).

Từ Bảng 3 trong Phụ lục II, bạn có thể thấy rằng diện tích tính bảng bên dưới đường cong chuẩn giữa *z* = 0 và *z* = 3.03 là 0.4988, và diện tích về phía bên phải của *z* = 3.03 là 0.5 - 0.4988 = 0.0012. Sau đó, bởi vì đây là một bài kiểm định hai phía, cho nên giá trị của α tương ứng với một vùng bác bỏ *z* > 3.03 hay *z* < -3.03 là 2(0.0012) = 0.0024. Vì thế, chúng ta báo cáo giá trị *p* cho kiểm định này là giá trị *p* = 0.0024.

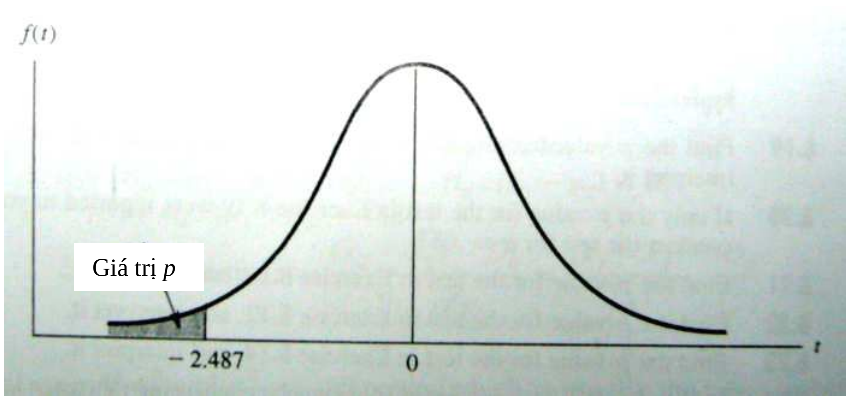
HÌNH **8.8** Xác định giá trị *p* cho kiểm định trong Ví dụ 8.8



VÍ DỤ **8.9** Nếu bạn đã hoạch định việc báo cáo các kết quả của kiểm định thống kê này trong Ví dụ 8.7, thì giá trị *p* nào mà bạn ắt báo cáo?

Lời giải Giá trị *p* cho kiểm định này là xác suất của việc quan sát một giá trị của trị thống kê *t* ít nhất trái ngược với giả thuyết không khi H_0 là đúng như giá trị quan sát được cho bộ dữ liệu này, cụ thể là một giá trị của $t \leq -2.487$ (xem Hình 8.9).

HÌNH **8.9** Giá trị *p* cho kiểm định trong Ví dụ 8.9



Không giống như bảng các diện tích nằm bên dưới đường cong chuẩn (Bảng 3 của Phụ lục II), thì Bảng 4 trong Phụ lục II không cho chúng ta các diện tích tương ứng với những giá trị khác nhau của *t*. Thay vào đó, bảng này cho chúng ta các giá trị của *t* tương ứng với những diện tích ở phía trên tương ứng 0.10, 0.05, 0.025, 0.010, và 0.005. Bởi vì phân phối *t* là đối xứng qua số trung bình của nó, nên chúng ta có thể sử dụng các diện tích ở phía trên này để ước lượng xấp xỉ xác suất để cho $t < -2.487$. Trị thống kê *t* cho sự kiểm định này là dựa trên 6 bậc tử do, vì vậy chúng ta tra d.f. = hàng 6 của Bảng 4 và tìm ra 2.487 rơi vào giữa $t_{0.025} = 2.447$ và $t_{0.010} = 3.143$. Bởi vì giá trị quan sát được của *t*, -2.487, là nhỏ hơn $-t_{0.025} = -2.447$ nhưng không nhỏ hơn $-t_{0.01} = -3.143$, cho nên chúng ta bác bỏ H_0 đối với $\alpha = 0.025$ nhưng không phải đối với $\alpha = 0.01$. Vì vậy, giá trị *p* cho kiểm định này được báo cáo là $0.01 \le$ giá trị $p \le 0.025$.

Để ủng hộ rằng một nhà nghiên cứu ủng hộ giá trị *p* cho một sự kiểm định và để sự giải thích của giá trị này cho độc giả không vi phạm qui trình kiểm định thống kê truyền thống được mô tả trong các phần trước, ta chỉ đơn giản để quyết định liệu bác bỏ giả thuyết không (với khả năng xảy ra một sai lầm loại I hay loại II) hay không cho độc giả. Như vậy, việc này làm dịch chuyển trách nhiệm từ việc chọn lựa một giá trị của α, và có thể là vấn đề đánh giá xác suất β của việc tạo ra một sai lầm loại II, sang một người đọc báo cáo này.

Bài tập

Các Kỹ thuật Cơ bản

8.16 Giả định rằng bạn đã kiểm định giả thuyết không $H_0 : \mu = 94$ so với giả thuyết thay thế $H_a : \mu < 94$. Đối với một mẫu ngẫu nhiên gồm $n = 52$ quan sát, $\bar{x} = 92.9$ và $s = 4.1$.

a. Tìm mức ý nghĩa quan sát được cho sự kiểm định này.

b. Nếu bạn mong muốn thực hiện kiểm định này bằng cách sử dụng $\alpha = 0.05$, thì các kết luận của bạn về kiểm định này ắt sẽ là như thế nào?

8.17 Giả định rằng bạn đã kiểm định giả thuyết không $H_0 : \mu = 94$ so với giả thuyết thay thế $H_a : \mu \ne 94$. Đối với một mẫu ngẫu nhiên gồm $n = 52$ quan sát, $\bar{x} = 92.1$ và $s = 4.1$.

a. Tìm mức ý nghĩa quan sát được cho sự kiểm định này.

b. Nếu bạn mong muốn thực hiện kiểm định này bằng cách sử dụng $\alpha = 0.05$, thì các kết luận của bạn về kiểm định này ắt sẽ là như thế nào?

8.18 Giả định rằng bạn đã kiểm định giả thuyết không $H_0 : \mu = 15$ so với giả thuyết thay thế $H_a : \mu \ne 15$. Đối với một mẫu ngẫu nhiên gồm $n = 18$ quan sát, $\bar{x} = 15.7$ và $s = 2.4$.

c. Tìm mức ý nghĩa quan sát được xấp xỉ cho sự kiểm định này.

d. Nếu bạn mong muốn thực hiện kiểm định này bằng cách sử dụng $\alpha = 0.05$, thì các kết luận của bạn về kiểm định này ắt sẽ là như thế nào?

Các Ứng dụng

8.19 Tìm giá trị *p* cho kiểm định về nhu cầu trung bình cho các móc khóa kim loại trong Bài tập 8.13, và giải thích giá trị này.

8.20 Nếu chỉ có giá trị *p* cho kiểm định trong Bài tập 8.19 được báo cáo với bạn, thì bạn có thể sử dụng giá trị này như thế nào để thực hiện kiểm định với $\alpha = 0.05$?

8.21 Tìm giá trị *p* cho kiểm định trong Bài tập 8.11, và giải thích giá trị này.

8.22 Tìm giá trị *p* cho kiểm định trong Bài tập 8.12, và giải thích giá trị này.

8.23 Tìm giá trị *p* cho kiểm định trong Bài tập 8.14, và giải thích giá trị này.

8.24 Trong Bài tập 7.20, chúng ta đã trình bày một số kết quả có liên quan đến các lãi suất cho vay thế chấp mà trong đó trung bình của mười kỳ vọng lãi suất là 8.5% và một độ lệch chuẩn bằng với 0.23%.

- **a.** Kiểm định giả thuyết $H_0: \mu = 8.7$ so với $H_a: \mu < 8.7$ khi sử dụng $\alpha = 0.05$.
- **b.** Tìm giá trị *p* cho kiểm định này và giải thích giá trị của nó.

8.25 Trong Bài tập 7.11, 40 yêu cầu thanh toán y tế nhận được trong tháng có một trung bình mẫu là \$930.

- **a.** Nếu độ lệch chuẩn của tổng thể là $\sigma = \$2000$, hãy kiểm định giả thuyết $H_0: \mu = \$800$ so với $H_a: \mu > \$800$ với $\alpha = 0.05$.
- **b.** Tìm giá trị *p* cho kiểm định này và giải thích giá trị của nó.

8.5 CÁC KIỂM ĐỊNH LIÊN QUAN ĐẾN SỰ KHÁC BIỆT GIỮA HAI SỐ TRUNG BÌNH TỔNG THỂ

Trong nhiều tình huống, câu hỏi thống kê cần được trả lời có liên quan đến một sự so sánh giữa hai số trung bình tổng thể. Ví dụ, Dịch vụ Bưu điện Hoa Kỳ quan tâm đến việc giảm bớt hóa đơn tiền xăng khổng lồ 350 triệu galông/năm bằng cách thay thế các xe tải chạy xăng bằng những xe tải chạy điện. Để xác định liệu khoản tiết kiệm đáng kể trong chi phí hoạt động có đạt được bằng cách thay đổi sang các xe tải chạy điện hay không, một nghiên cứu thí điểm nên được tiến hành bằng cách sử dụng, ví dụ, 100 xe chở thư chạy xăng như thường lệ và 100 xe chở thư chạy điện cùng hoạt động trong những điều kiện tương tự nhau. Trị thống kê mà tóm tắt thông tin mẫu có liên quan đến sự khác biệt về các số trung bình tổng thể $\mu_1 - \mu_2$ là sự khác biệt trong các số trung bình mẫu $\bar{x}_1 - \bar{x}_2$. Vì vậy, khi kiểm định liệu sự khác biệt trong các số trung bình mẫu có cho thấy sự khác biệt thực sự trong các số trung bình tổng thể có khác với một giá trị được xác định cụ thể, $\mu_1 - \mu_2 = D_0$ hay không, thì chúng ta ắt sẽ sử dụng số lượng độ lệch chuẩn mà $\bar{x}_1 - \bar{x}_2$ nằm cách với sự khác biệt được giả định D_0. Qui trình kiểm định chính thức **khi các cỡ mẫu là lớn** được thể hiện trong phần trình bày sau.

Kiểm định Thống kê cho Mẫu Lớn đối với (μ_1 - μ_2)

1. Giả thuyết không: $H_0: (\mu_1 - \mu_2) = D_0$ trong đó D_0 là một sự khác biệt nào đó được xác định cụ thể mà bạn mong muốn kiểm định. Đối với nhiều kiểm định, bạn sẽ mong muốn giả định rằng không có sự khác biệt nào giữa μ_1 và μ_2 - nghĩa là, $D_0 = 0$.

2. Giả thuyết thay thế:

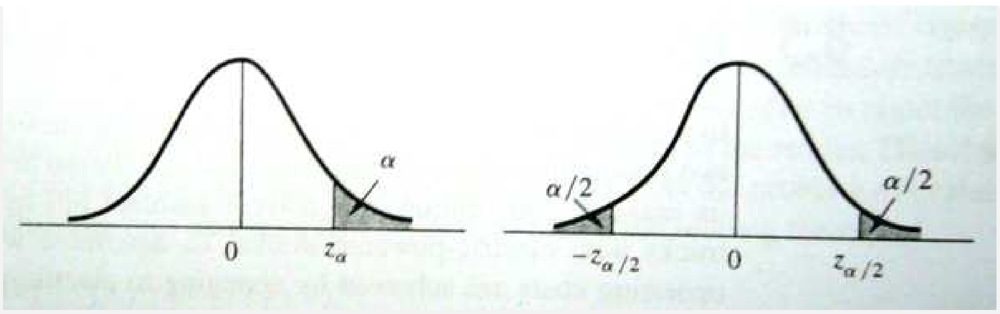
Kiểm định Một phía	**Kiểm định Hai phía**
$H_a:(\mu_1-\mu_2)>D_0$ [hay $H_a:(\mu_1-\mu_2)<D_0$]	$H_a:(\mu_1-\mu_2)\neq D_0$

3. Trị thống kê kiểm định: $z=\dfrac{(\bar{x}_1-\bar{x}_2)-D_0}{\sigma_{(\bar{x}_1-\bar{x}_2)}}=\dfrac{(\bar{x}_1-\bar{x}_2)-D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$

Nếu σ_1^2 và σ_2^2 là chưa được biết (thường là như vậy), thì thay thế các phương sai của mẫu s_1^2 và s_2^2 lần lượt cho σ_1^2 và σ_2^2.

4. Vùng bác bỏ:

Kiểm định Một phía	**Kiểm định Hai phía**
$z>z_\alpha$ [hay $z<-z_\alpha$ khi giả thuyết thay thế là $H_a:(\mu_1-\mu_2)<D_0$]	$z>z_{\alpha/2}$ hay $z<-z_{\alpha/2}$



Các giả thiết: Các mẫu này được chọn lựa ngẫu nhiên và độc lập từ hai tổng thể, và $n_1\geq 30$ và $n_2\geq 30$.

VÍ DỤ **8.10** Một công ty đang sử dụng một phương án trả lương mới là doanh số bán hàng cộng hoa hồng cho nhân sự bán hàng của mình muốn so sánh các kỳ vọng lương hàng năm của các nhân viên bán hàng nam và nữ của mình theo kế hoạch mới này. Các mẫu ngẫu nhiên gồm $n_1=40$ đại diện bán hàng nữ và $n_2=40$ đại diện bán hàng nam được yêu cầu dự báo về thu nhập hàng năm của mình theo kế hoạch mới này. Các số trung bình và độ lệch chuẩn của mẫu là

$$\bar{x}_1=\$31{,}083 \quad \bar{x}_2=\$29.745$$
$$s_1=\$2322 \qquad s_2=\$2569$$

Liệu dữ liệu này có cung cấp đủ bằng chứng để cho thấy rằng có một sự khác biệt về thu nhập trung bình hàng năm được kỳ vọng giữa các đại diện bán hàng nam và nữ? Hãy kiểm định bằng cách sử dụng $\alpha=0.05$.

Lời giải Bởi vì chúng ta mong muốn xác minh một sự khác biệt trong thu nhập trung bình hàng năm giữa các đại diện bán hàng nam và nữ, hoặc $\mu_1>\mu_2$ hoặc $\mu_1<\mu_2$, cho nên chúng ta muốn kiểm định giả thuyết không

$$H_0: \mu_1 = \mu_2, \quad \text{nghĩa là, } \mu_1 - \mu_2 = D_0 = 0$$

so với giả thuyết thay thế

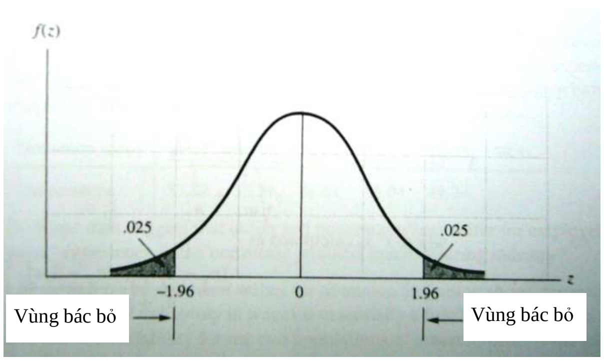
$$H_a: \mu_1 \neq \mu_2, \quad \text{nghĩa là, } \mu_1 - \mu_2 \neq 0$$

Chúng ta sử dụng s_1^2 và s_2^2 để ước lượng xấp xỉ lần lượt σ_1^2 và σ_2^2. Thay thế bằng các giá trị này, cùng với $\bar{x}_1$ và $\bar{x}_2$, vào công thức tính trị thống kê kiểm định *z*, chúng ta có được

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \approx \frac{(31{,}083 - 29{,}475) - 0}{\sqrt{\dfrac{(2132)^2}{40} + \dfrac{(2569)^2}{40}}} = 2.45$$

Sử dụng một kiểm định hai phía với $\alpha = 0.05$, chúng ta thay thế $\alpha/2 = 0.025$ vào mỗi phía của phân phối *z* và bác bỏ H_0 nếu $z > 1.96$ hay $z < -1.96$ (xem Hình 8.10). Bởi vì giá trị quan sát được của *z* = 2.45 vượt quá 1.96, nên trị thống kê kiểm định này rơi vào vùng bác bỏ. Chúng ta bác bỏ H_0 và kết luận rằng có một sự khác biệt trong các kỳ vọng tiền lương trung bình hàng năm giữa những đại diện bán hàng nam và nữ. Chúng ta nên cảm giác rất tin tưởng rằng chúng ta đã thực hiện một quyết định chính xác. Xác suất mà kiểm định của chúng ta ắt sẽ dẫn chúng ta đến việc bác bỏ H_0 khi trong thực tế giả thuyết này là đúng, chỉ là $\alpha = 0.05$.

HÌNH **8.10** Vị trí của vùng bác bỏ trong Ví dụ 8.10



Cách khác, chúng ta có thể tính toán giá trị *p* cho kiểm định này như sau

$$\text{Giá trị } p = P(z \geq 2.45 \text{ hay } z \leq -2.45) = 2P(z \geq 2.45)$$
$$= 2(0.5 - 0.4929) = 0.0142$$

Bởi vì giá trị *p* là *nhỏ hơn* giá trị mong muốn của $\alpha = 0.05$, H_0 bị bác bỏ và các kết luận là như nhau.

Kiểm định cho mẫu nhỏ đối với một sự khác biệt giữa các số trung bình tổng thể được căn cứ vào giả định rằng cả hai tổng thể này đều tuân theo phân phối chuẩn, và, thêm nữa, rằng cả hai tổng thể này đều có các phương sai bằng nhau - nghĩa là, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Trong trường hợp đặc biệt này, kiểm định cho mẫu nhỏ đối với giả định $H_0: \mu_1 - \mu_2 = D_0$ sử dụng trị thống kê kiểm định này

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

với s^2, ước lượng gộp của σ^2, được cho trong Phần 7.6. Trị thống kê kiểm định này có một phân phối *t* Student với $n_1 + n_2 - 2$ bậc tự do, và qui trình kiểm định là giống với phần được tóm tắt trong trình bày sau.

Kiểm định Thống kê cho Mẫu Nhỏ đối với (μ_1 - μ_2)

1. Giả thuyết không: $H_0 : (\mu_1 - \mu_2) = D_0$ trong đó D_0 là một sự khác biệt nào đó được xác định cụ thể mà bạn mong muốn kiểm định. Đối với nhiều kiểm định, bạn sẽ mong muốn giả định rằng không có sự khác biệt nào giữa μ_1 và μ_2 - nghĩa là, $D_0 = 0$.

2. Giả thuyết thay thế:

Kiểm định Một phía	**Kiểm định Hai phía**
$H_a : (\mu_1 - \mu_2) > D_0$	$H_a : (\mu_1 - \mu_2) \neq D_0$
[hay $H_a : (\mu_1 - \mu_2) < D_0$]	

3. Trị thống kê kiểm định: $t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

trong đó s^2 được tính như sau

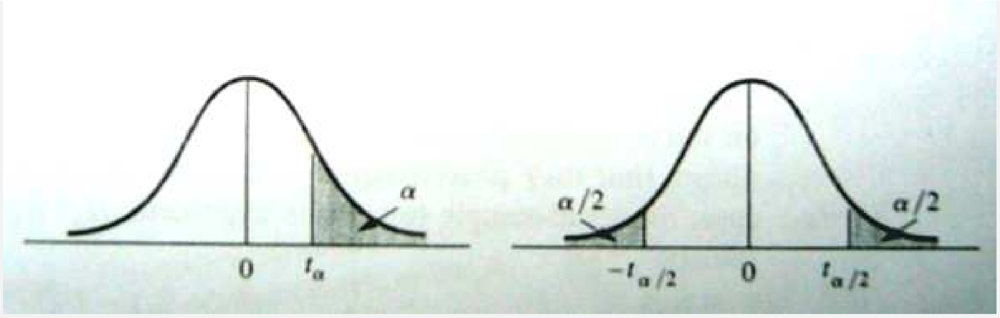
$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

hay

$$s^2 = \frac{\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

4. Vùng bác bỏ:

Kiểm định Một phía	**Kiểm định Hai phía**
$t > t_\alpha$	$t > t_{\alpha/2}$ hay $t < -t_{\alpha/2}$
[hay $t < -t_\alpha$ khi giả thuyết thay thế là $H_a : (\mu_1 - \mu_2) < D_0$]	



Các giá trị tới hạn của t, t_α, và $t_{\alpha/2}$, sẽ được căn cứ vào (n_1+n_2-2) bậc tự do. Các giá trị được lập bảng có thể được tìm thấy trong Bảng 4 của Phụ lục II.

Các giả thiết: Các mẫu này được chọn lựa ngẫu nhiên và độc lập từ các tổng thể được phân phối chuẩn. Các phương sai của những tổng thể này, σ_1^2 và σ_2^2 là bằng nhau.

VÍ DỤ **8.11** Mặc dù các mức tiền lương thuộc hiệp hội và không thuộc hiệp hội có xu hướng gia tăng với cùng tốc độ trong dài hạn, thì tiền lương thuộc hiệp hội thường nhanh hơn trong các thời kỳ khủng hoảng và sớm hơn trong các giai đoạn phục hồi kinh tế, và tiền lương không thuộc hiệp hội có xu hướng tăng nhanh hơn vào thời điểm muộn hơn trong chu kỳ kinh doanh khi các thị trường lao động là căng thẳng. Để nghiên cứu vấn đề này, một nhà kinh tế đã ghi nhận các mức tiền công bình quân mỗi giờ (bao gồm cả các khoản phúc lợi của nhân viên) của các nhân viên có hai năm kinh nghiệm cho 11 công ty sản xuất hàng tiêu dùng được lựa chọn ngẫu nhiên, 6 trong số này có các cửa hàng không thuộc hiệp hội và 5 trong số này có những cửa hàng thuộc hiệp hội. Dữ liệu là như sau:

Cửa hàng không thuộc hiệp hội	\$8.26	\$8.17	\$8.45	\$9.09	\$8.85	\$8.31
Cửa hàng thuộc hiệp hội	\$7.92	\$8.39	\$8.64	\$8.04	\$8.24	

Liệu những dữ liệu này có gợi ý rằng các mức tiền công thuộc hiệp hội và không thuộc hiệp hội có khác nhau đối với những nhân viên có hai năm kinh nghiệm trong ngành chế biến sản phẩm tiêu dùng không?

Lời giải Đặt μ_1 và μ_2 lần lượt là các mức tiền công trung bình cho những cửa hàng không thuộc hiệp hội và thuộc hiệp hội. Ngoài ra, giả định rằng độ biến thiên trong các mức tiền công về thực chất là một hàm số của những sự khác biệt riêng lẻ và rằng độ biến thiên cho hai tổng thể của các thước đo này là giống nhau.

Các số trung bình và độ lệch chuẩn của mẫu là

$$\bar{x}_1 = 8.522 \qquad s_1 = 0.3668$$
$$\bar{x}_2 = 8.246 \qquad s_2 = 0.2849$$

Sau đó ước lượng gộp của phương sai chung là

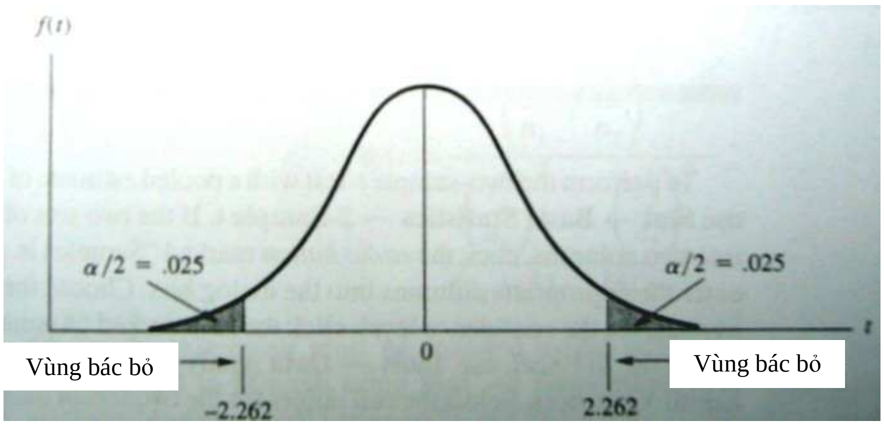
$$s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{0.67271+0.32467}{6+5-2} = 0.1108$$

Giả thuyết này cần được kiểm định

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{so với} \quad H_a : \mu_1 - \mu_2 \neq 0$$

Giả thuyết thay thế này hàm ý rằng chúng ta nên sử dụng một sự kiểm định thống kê hai phía và rằng vùng bác bỏ cho sự kiểm định này sẽ nằm ở cả hai phía của phân phối *t.* Tham khảo Bảng 4 trong Phụ lục II và lưu ý rằng giá trị tới hạn của *t* đối với $\alpha/2 = 0.025$ và $n_1 + n_2 - 2 = 9$ bậc tự do là 2.262. Vì vậy, chúng ta sẽ bác bỏ H_0 khi $t > 2.262$ hay $t < -2.262$ (xem Hình 8.11).

HÌNH **8.11** Vùng bác bỏ cho Ví dụ 8.11



Giá trị tính toán được của trị thống kê kiểm định này là

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{8.522 - 8.246}{\sqrt{0.1108\left(\dfrac{1}{6} + \dfrac{1}{5}\right)}} = 1.369$$

So sánh giá trị này với giá trị tới hạn, chúng ta thấy rằng giá trị tính toán được này không rơi vào vùng bác bỏ. Do vậy, chúng ta không thể kết luận rằng có một sự khác biệt về các mức tiền công trung bình giữa những cửa hàng thuộc hiệp hội và không thuộc hiệp hội.

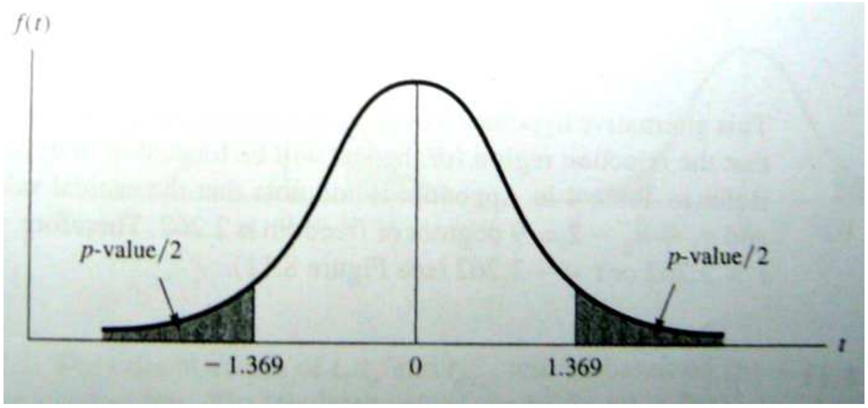
VÍ DỤ **8.12** Tìm giá trị *p* mà ắt đã được báo cáo cho kiểm định thống kê trong Ví dụ 8.11

Lời giải Giá trị quan sát được của *t* cho bài kiểm định hai phía này là *t* = 1.369. Vì thế, giá trị *p* cho bài kiểm định này ắt sẽ là hai lần của xác suất để cho *t* > 1.369 (xem Hình 8.12). Bởi vì chúng ta không thể có được xác suất này từ Bảng 4 của Phụ lục II, cho nên chúng ta ắt báo cáo rằng giá trị *p* cho bài kiểm định này là giá trị được lập bảng nhỏ nhất của α mà dẫn đến việc bác bỏ của H_0. Tra hàng tương ứng với 9 bậc tự do trong Bảng 4, chúng ta tìm thấy rằng giá trị quan sát được của *t* = 1.369 là nhỏ hơn $t_{0.10} = 1.383.$ Vì thế, chúng ta ắt sẽ báo cáo rằng

giá trị *p* > 2 (0.10) = 0.20

Kết luận của chúng ta không bác bỏ H_0 được khẳng định, bởi vì giá trị *p* là lớn hơn giá trị của $\alpha = 0.05$.

HÌNH **8.12** Giá trị *p* cho kiểm định trong Ví dụ 8.11



Để thực hiện sự kiểm định *t* cho hai mẫu với một ước lượng gộp về phương sai bằng cách sử dụng Minitab, hãy dùng chuỗi lệnh sau **Stat → Basic Statistics → 2-Sample t.** Nếu hai bộ dữ liệu này đã được nhập vào hai cột, thì bấm vào *nút* có tựa đề "Samples in different columns", và nhập các cột phù hợp vào hộp hội thoại. Chọn giả thuyết thay thế thích hợp, mức tin cậy, bấm vào hộp có tựa đề "Assume equal variances". và bấm **OK.** Trong Excel, sử dụng **Tools → Data Analysis → t-Test: Two Sample Assuming Equal Variances.** Chọn các dãy ô cho hai tập hợp dữ liệu này, nhập vào sự khác biệt được giả định về các số trung bình (thông thường $D_0 = 0$), nhập vào giá trị của α, và chọn ra một ô mà trong đó sẽ xuất hiện kết quả.

Kết quả Minitab cho một kiểm định *t* hai mẫu bằng cách sử dụng dữ liệu trong Ví dụ 8.11 được trình bày trong Bảng 8.5. Chú ý rằng mục SE MEAN (sai số chuẩn của trung bình), được cho trong từng cột, được tính bằng $s/\sqrt{n}$. Các mục còn lại là có tính tự giải thích và có thể được so sánh với những kết quả trong các Ví dụ 8.11 và 8.12.

BẢNG **8.5** Kết quả Minitab cho dữ liệu trong Ví dụ 8.11

```
T HAI M•U CHO C1 SO V•I C2
               N   TRUNG BÌNH   •• L•CH CHU•N  SAI S• CHU•N C•A TRUNG BÌNH
C1             6     8.522         0.367                 0.15
C2             5     8.246         0.285                 0.13

95 PCT C1 FOR MU C1 - MU C2: ( -0.18.  0.73)

KI•M ••NH T MU C1 = MU C2 (SO V•I GI• THUY•T KHÔNG): T=1.37  P=0.29  DF=9

•• L•CH CHU•N CHUNG =    0.333
```

Trước khi kết luận thảo luận của chúng ta, chúng ta nên nhớ lại rằng **những sự lệch hướng vừa phải so với giả thiết rằng các tổng thể có phân phối xác suất chuẩn không ảnh hưởng nghiêm trọng đến phân phối của trị thống kê kiểm định và hệ số tin cậy cho khoảng tin cậy tương ứng. Mặt khác, các phương sai của tổng thể phải nên gần bằng nhau nhằm đảm bảo rằng các qui trình được trình bày trên đây là có giá trị.**

Nếu có một lý do để tin rằng các phương sai của tổng thể là rất khác nhau, thì hai sự thay đổi phải được thực hiện trong qui trình kiểm định và ước lượng. Bởi vì số ước lượng gộp s^2 không còn phù hợp nữa, cho nên các phương sai của mẫu s_1^2 và s_2^2 được sử dụng như những số ước lượng cho σ_1^2 và σ_2^2. Trị thống kê kiểm định tạo ra là

$$\frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Khi các cỡ mẫu là *nhỏ*, thì các giá trị tới hạn cho trị thống kê này được tìm thấy trong Bảng 4 của Phụ lục II, bằng cách sử dụng bậc tự do được ước lượng xấp xỉ bởi

$$\text{d.f.} \approx \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}}$$

Rõ ràng là, kết quả này phải được làm tròn đến gần số nguyên gần nhất. Qui trình này được thực hiện trong Minitab bằng cách sử dụng **Stat → Basic Statistics → 2-Sample t** (không bấm vào "Assume equal variances"), và trong Excel sử dụng **Tools → Data Analysis → t-Test: Two Sample Assuming Unequal Variances.**

Trong Phần 8.9, chúng ta trình bày một qui trình nhằm kiểm định một giả thuyết có liên quan đến sự bằng nhau của hai phương sai của tổng thể mà có thể được sử dụng để xác định liệu các phương sai của mẫu đang nhắc đến có bằng nhau hay không.

Nếu như có một lý do để tin rằng các giả định về tính chuẩn đã bị vi phạm, thì bạn có thể kiểm định về một sự dịch chuyển trong vị trí của hai phân phối tổng thể bằng cách sử dụng kiểm định *U* Mann-Whitney không tham số. Qui trình kiểm định này, mà yêu cầu ít các giả định hơn, hầu như nhạy cảm trong việc xác minh một sự khác biệt về các số trung bình của mẫu khi các điều kiện cần thiết cho kiểm định *t* được đáp ứng. Qui trình này có lẽ còn nhạy cảm hơn khi các giả định này không được đáp ứng. (Tham kháo Mendenhall, Beaver, và Beaver, *A Course in Business Statistics,* xuất bản lần thứ 4. 1996, Duxbury Press.)

Bài tập

Các Kỹ thuật Cơ bản

8.26 Các mẫu ngẫu nhiên độc lập gồm $n_1 = 80$ và $n_2 = 80$ được chọn lần lượt từ các tổng thể 1 và 2. Các tham số của tổng thể, những số trung bình và phương sai của mẫu được thể hiện trong bảng đi kèm sau đây.

	Tổng thể	
Các tham số và trị thống kê	**1**	**2**
Trung bình tổng thể	μ_1	μ_2
Phương sai tổng thể	σ_1^2	σ_2^2
Cỡ mẫu	80	80
Trung bình mẫu	11.6	9.7
Phương sai mẫu	27.9	38.4

a. Nếu mục tiêu nghiên cứu của bạn là nhằm chứng minh rằng μ_1 lớn hơn μ_2, hãy xác định các giả thuyết không và thay thế mà bạn ắt chọn cho một kiểm định thống kê.

b. Liệu kiểm định trong câu (a) là kiểm định một phía hay hai phía?

c. Tìm trị thống kê kiểm định mà bạn ắt sử dụng cho kiểm định trong các câu (a) và (b), và vùng bác bỏ với $\alpha = 0.10$.

d. Nhìn vào dữ liệu này. Từ trực giác của mình, bạn có nghĩ rằng dữ liệu này cung cấp bằng chứng đủ để cho thấy rằng μ_1 là lớn hơn μ_2 không?

8.27 Tham khảo lại Bài tập 8.26. Tìm mức ý nghĩa quan sát được đối với sự kiểm định này.

8.28 Tham khảo lại Bài tập 8.26.

a. Giải thích các điều kiện thực tiễn mà ắt sẽ kích thích bạn muốn thực hiện một kiểm định *z* hai phía.

b. Tìm các giả thuyết không và thay thế.

c. Sử dụng dữ liệu của Bài tập 8.26 để thực hiện sự kiểm định này. Liệu dữ liệu này có cung cấp bằng chứng đủ để bác bỏ H_0 và chấp nhận H_a? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

d. Các kết luận thực tiễn có thể được rút ra từ sự kiểm định trong câu (c) là gì?

8.29 Giả định rằng bạn mong muốn xác mình sự khác biệt giữa μ_1 và μ_2 (hoặc $\mu_1 > \mu_2$, hoặc $\mu_1 < \mu_2$) và rằng thay vì chạy một kiểm định hai phía bằng cách sử dụng $\alpha = 0.10$, thì bạn sử dụng qui trình kiểm định sau đây: Bạn chờ đợi cho đến khi đã thu thập dữ liệu mẫu và đã tính toán được $\bar{x}_1$ và $\bar{x}_2$. Nếu $\bar{x}_1$ là lớn hơn $\bar{x}_2$, thì bạn chọn giả thuyết thay thế $H_a : \mu_1 > \mu_2$ và chạy một kiểm định một phía, qua việc đặt $\alpha_1 = 0.10$ ở phía trên của phân phối *z*. Mặt khác, nếu $\bar{x}_2$ là lớn hơn $\bar{x}_1$, thì bạn đảo ngược qui trình này và chạy một kiểm định một phía, qua việc đặt $\alpha_2 = 0.10$ ở phía thấp của phân phối *z*. Nếu bạn sử dụng qui trình này và nếu μ_1 là thật sự bằng với μ_2, thì xác suất α mà bạn sẽ kết luận rằng μ_1 không bằng μ_2 là bao nhiêu (nghĩa là, xác suất α mà bạn sẽ từ chối sai lầm H_0 khi H_0 là đúng là bao nhiêu?) Bài tập này chứng minh lý do tại sao các kiểm định thống kê phải được thực hiện *trước khi* quan sát dữ liệu.

8.30 Các mẫu ngẫu nhiên độc lập gồm $n_1 = n_2 = 4$ quan sát được chọn từ hai tổng thể chuẩn với các phương sai bằng nhau. Dữ liệu được thể hiện dưới đây.

Mẫu 1	12	9	14	14
Mẫu 2	11	9	10	8

a. Giả định rằng bạn mong muốn xác định liệu μ_1 có lớn hơn μ_2 hay không. Tìm giả thuyết thay thế cho sự kiểm định này.

b. Phát biểu H_0.

c. Tìm vùng bác bỏ cho kiểm định này với $\alpha = 0.10$.

d. Thực hiện kiểm định này và phát biểu các kết luận của bạn.

e. Tìm giá trị *p* xấp xỉ cho kiểm định này, và giải thích giá trị này.

Các Ứng dụng

8.31 Để so sánh các khả năng chọn lựa chứng khoán của hai công ty môi giới, chúng ta đã so sánh khoản lợi tức hàng năm (đã loại trừ phí môi giới) cho khoản đầu tư trị giá $1000 vào mỗi trong số 30 chứng khoán niêm yết tại mỗi trong số danh mục cổ phiếu "được đề nghị nhiều nhất" của hai công ty này. Các số trung bình và độ lệch chuẩn (tính bằng đôla) cho mỗi trong số hai mẫu này được trình bày trong bảng đi kèm sau đây. Chúng ta muốn xác định liệu dữ liệu này có cung cấp đủ bằng chứng để cho thấy rằng liệu có một sự khác biệt giữa hai công ty môi giới này về lợi tức trung bình tính trên mỗi cổ phiếu được đề nghị hay không.

	Công ty	
Các trị thống kê mẫu	**1**	**2**
Cỡ mẫu	30	30
Trung bình	264	199
Độ lệch chuẩn	157	111

a. Phát biểu giả thuyết thay thế mà sẽ trả lời tốt nhất cho câu hỏi này.

b. Phát biểu H_0.

c. Tìm vùng bác bỏ cho kiểm định này với $\alpha = 0.01$.

d. Thực hiện kiểm định này và phát biểu các kết luận của bạn.

8.32 Trong Bài tập 7.34, chúng ta đã mô tả một sự so sánh về mức trung bình của các hồ sơ xin vay mua nhà được chấp thuận của một ngân hàng từ tháng Tư đến tháng Năm. Các cỡ, số trung bình, và độ lệch chuẩn của mẫu cho hai tháng này được chép lại trong bảng sau đây.

Các trị thống kê mẫu	**tháng Tư**	**tháng Năm**
Cỡ mẫu	57	66
Trung bình	$78,000	$82,700
Độ lệch chuẩn	$6,300	$7,100

a. Liệu dữ liệu này có cung cấp đủ bằng chứng để cho thấy rằng liệu có một sự khác biệt trong giá trị trung bình của các hồ sơ xin vay mua nhà từ tháng Tư đến tháng Năm? Hãy kiểm định bằng cách sử dụng $\alpha = 0.10$.

b. Khác biệt nào được tạo ra khi bạn thực hiện kiểm định này trong câu (a) khi sử dụng $\alpha = 0.10$ so với $\alpha = 0.05$? Hãy giải thích.

8.33 Một chuỗi siêu thị đã chọn mẫu những ý kiến khách hàng về dịch vụ do các siêu thị của chuỗi này cung cấp cả trước lẫn sau khi đội ngũ nhân viên của một cửa hàng được tham gia vào ba kỳ huấn luyện hàng tuần qua băng video mà nhằm mục đích cải thiện các quan hệ khách hàng. Các mẫu ngẫu nhiên độc lập gồm mười khách hàng, từng người một được phỏng vấn trước và sau các kỳ huấn luyện, và mỗi người được yêu cầu xếp hạng dịch vụ của cửa hàng đó trên thang điểm từ 1 (nghèo nàn) đến 10 (tuyệt vời). Trung bình và độ lệch chuẩn cho từng mẫu được thể hiện trong bảng đi kèm sau đây. Chúng ta muốn xác định liệu dữ liệu này có trình bày bằng chứng đủ để chỉ ra rằng khóa huấn luyện này là có hiệu quả trong việc gia tăng điểm số dịch vụ khách hàng.

Trước	**Sau**
$\bar{x}_1 = 6.82$	$\bar{x}_2 = 8.17$
$s_1 = 0.95$	$s_2 = 0.56$

a. Phát biểu giả thuyết thay thế mà sẽ trả lời tốt nhất cho câu hỏi này.

b. Phát biểu H_0.

c. Tìm vùng bác bỏ cho kiểm định này với $\alpha = 0.05$.

d. Thực hiện kiểm định này và phát biểu các kết luận thực tiễn được rút ra từ sự kiểm định này.

e. Mô tả rủi ro mà bạn phải chịu khi đi đến một kết luận không chính xác trong câu (d).

8.34 Tham khảo lại Bài tập 7.101. Dữ liệu này phản ảnh thời gian (tính bằng giây) để tải chương trình Ami Pro 2.0 trên một máy tính cá nhân IBM PS/2 Model 90 486DX/33 khi sử dụng các hệ điều hành Standard Windows và Enhanced Windows được chép lại ở đây.

Standard		**Enhanced**	
1.56	1.20	1.59	0.96
1.41	1.38	1.68	1.09
1.48	1.54	1.17	1.26
1.37	1.41	0.94	1.23
1.39	1.16	1.56	1.30

Nếu bạn *không thể* thừa nhận rằng các phương sai này là bằng nhau, hãy sử dụng một kiểm định phù hợp đối với giả định nhằm xác định liệu thực sự có một sự khác biệt trong thời gian bình quân để tại chương trình Ami Pro 2.0 khi sử dụng các hệ điều hành Standard Windows và Enhanced Windows. Hãy sử dụng $\alpha = 0.01$.

8.35 Bảng in Minitab trình bày dưới đây được chạy ra khi sử dụng dữ liệu trong Bài tập 8.34.

```
T HAI M•U CHO C1 VS C2
                           N   TRUNG BÌNH   •• L•CH CHU•N  SAI S• CHU•N C•A
                                                                 TRUNG BÌNH
C1                        10        1.390           0.129             0.041
C2                        10        1.278           0.259             0.082
95% C1 CHO MU C1 - MU C2: ( -0.086, 0.310)
KI•M ••NH T CHO MU C1 = MU C2 (SO V•I GI• THUY•T KHÔNG): T=1.22 P=0.24
DF=13
```

a. Các giả thuyết không và thay thế cho kiểm định được chạy trong Minitab là như thế nào?

b. Giá trị *p* của kiểm định trong câu (a) là bao nhiêu?

c. Căn cứ vào giá trị *p*, liệu bạn có suy luận rằng có một sự khác biệt đáng kể trong các thời gian trung bình để tải chương trình Ami Pro 2.0 khi sử dụng các hệ điều hành Standard Windows so với Enhanced Windows ? Hãy giải thích.

8.6 MỘT KIỂM ĐỊNH VỀ KHÁC BIỆT CẶP

Một nhà sản xuất muốn so sánh chất lượng độ bền của hai loại vỏ xe khác nhau, *A* và *B*. Trong sự so sánh này, một vỏ xe thuộc loại *A* và một vỏ xe thuộc loại *B* được chỉ định ngẫu nhiên và lắp vào các bánh sau của mỗi trong số năm chiếc xe hơi. Các chiếc xe này sau đó được lái đi trong quãng đường tính bằng dặm được xác định cụ thể, và lượng hao mòn được ghi nhận cho từng chiếc vỏ xe. Những đại lượng này được thể hiện trong Bảng 8.6. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng có một sự khác biệt trong khối lượng hao mòn bình quân cho hai loại vỏ xe này không?

BẢNG **8.6** Dữ liệu về độ hao mòn vỏ xe

Chiếc xe	**Vỏ xe loại *A***	**Vỏ xe loại *B***
1	10.6	10.2
2	9.8	9.4
3	12.3	11.8
4	9.7	9.1
5	8.8	8.3
	$\bar{x}_1 = 10.24$	$\bar{x}_2 = 9.76$

Khi phân tích dữ liệu này, chúng ta lưu ý rằng sự khác biệt giữa hai số trung bình của mẫu là $(\bar{x}_1 - \bar{x}_2) = 0.48$, là một lượng khá nhỏ nếu xét đến độ biến thên của dữ liệu này và số lượng nhỏ của các thước đo có liên quan. Thoạt tiên nhìn, dường như là có ít bằng chứng để chỉ ra một sự khác biệt giữa các số trung bình mẫu, một sự phỏng đoán mà chúng ta có thể kiểm tra bằng phương pháp đã được trình bày trong Phần 8.5.

Ước lượng gộp của phương sai chung σ^2 là

$$s^2 = \frac{\sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

$$= \frac{6.932 + 7.052}{5 + 5 - 2}$$

$$= 1.748$$

và

$$s = 1.32$$

Giá trị tính toán được này của *t* được sử dụng để kiểm định giả định rằng $\mu_1 = \mu_2$ là:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{10.24 - 9.76}{1.32\sqrt{\frac{1}{5} + \frac{1}{5}}} = 0.57$$

một giá trị mà gần như không đủ lớn để bác bỏ giả định rằng $\mu_1 = \mu_2$. Khoảng tin cậy 95% tương ứng là

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2} s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (10.24 - 9.76) \pm (2.306)(1.32)\sqrt{\frac{1}{5} + \frac{1}{5}}$$

hay -1.45 đến 2.41. Lưu ý rằng khoảng này là khá rộng nếu xem xét đến sự khác biệt nhỏ giữa các số trung bình của mẫu.

Cái nhìn thứ hai vào dữ liệu này bộc lộ một sự không nhất quán đáng lưu ý với kết luận này. Chúng ta lưu ý rằng đại lượng độ hao mòn cho vỏ xe loại *A* là lớn hơn so với giá trị tương ứng cho loại *B* đối với *mỗi* trong số năm chiếc xe này. Những khác biệt này, được ghi nhận bằng $d = A - B$, là như sau:

Chiếc xe	$d = A - B$
1	0.4
2	0.4
3	0.5
4	0.6
5	0.5
	$\bar{d} = 0.48$

Nếu không có sự khác biệt nào trong độ mòn trung bình của vỏ xe đối với hai loại vỏ xe này, thì xác suất để cho vỏ xe A cho thấy bị mòn nhiều hơn vỏ xe B là bằng với $p = 0.5$, và năm chiếc xe này tương ứng với $n = 5$ lần thử nhị thức độc lập. Đặt x đại diện cho số lần mà đại lượng độ mòn cho vỏ xe loại A là lớn hơn đại lượng đối với vỏ xe loại B. Một kiểm định hai phía của giả thuyết không $p = 0.5$ ắt bao gồm một vùng bác bỏ gồm có $x = 0$ và $x = 5$ và $\alpha = P(x = 0) + P(x = 5) = 2(1/2)^5 = 1/16 = 0.025.$ Bởi vì năm của những khác biệt này là dương ($x = 5$), cho nên chúng ta có bằng chứng để chỉ ra rằng một sự khác biệt hiện hữu trong độ hao mòn trung bình của hai loại vỏ xe này.

Bạn sẽ lưu ý rằng chúng ta đã sử dụng hai kiểm định thống kê khác nhau để kiểm định cùng một giả thuyết. Liệu có khác thường không khi kiểm định *t*, mà sử dụng nhiều thông tin (số đại lượng mẫu thực tế) hơn kiểm định nhị thức, thất bại trong việc cung cấp đủ bằng chứng cho việc bác bỏ giả định $\mu_1 = \mu_2$?

Có một sự giải thích cho sự không nhất quán này. Sự kiểm định *t* được mô tả trong Phần 8.5 không phải là kiểm định thống kê phù hợp phải được sử dụng cho ví dụ của chúng ta. Qui trình kiểm định thống kê này của Phần 8.5 yêu cầu rằng hai mẫu này phải *độc lập và ngẫu nhiên*. Chắc chắn rằng, yêu cầu về sự độc lập đã bị vi phạm bởi cách thức mà qua đó thí nghiệm này được tiến hành. (Cặp) các thước đo này, một vỏ xe *A* và một vỏ xe *B*, đối với một chiếc xe cụ thể rõ ràng là có liên hệ với nhau. Một sự xem qua dữ liệu cho thấy rằng các đại lượng này xấp xỉ có cùng độ lớn đối với một chiếc xe nhưng thay đổi rất đáng lưu ý từ chiếc xe này sang chiếc xe khác. Dĩ nhiên đây chính xác là điều mà chúng ta có lẽ kỳ vọng. Độ mòn vỏ xe được tạo ra phần lớn bởi các thói quen của người lái xe, sự cân bằng của các bánh xe, và bề mặt con đường. Bởi vì mỗi chiếc xe có một người lái xe khác nhau, nên chúng ta ắt kỳ vọng một lượng biến thiên lớn trong dữ liệu của chiếc xe này so với chiếc xe khác. Khi các mẫu được lấy ra theo một cách thức để cho một quan sát trong mẫu thứ hai có liên quan với một quan sát trong mẫu thứ nhất, thì các mẫu này được cho là *phụ thuộc*. Ví dụ, ghi nhận huyết áp của một người trước và sau khi uống thuốc cao huyết áp sẽ tạo ra các quan sát phụ thuộc, cũng giống như cách thức mà hai người thẩm định khi sử dụng các tiêu chuẩn giống nhau đi đến các giá trị được thẩm định tương tự nhau và vì vậy phụ thuộc nhau cho cùng một đặc trưng.

Sự tương đồng mà chúng ta đạt được với ước lượng khoảng đã chứng tỏ cho chúng ta thấy rằng bề rộng của các khoảng tin cậy đối với mẫu lớn phụ thuộc vào độ lớn của độ lệch chuẩn của số ước lượng điểm của tham số. Giá trị của nó càng nhỏ, thì sự ước lượng đó là càng tốt và có nhiều khả năng xảy ra hơn rằng chính trị thống kê kiểm định này sẽ cung cấp bằng chứng để bác bỏ giả thuyết không nếu như trên thực tế giả thuyết này là sai. Kiến thức về hiện tượng này được sử dụng trong việc *thiết kế* thí nghiệm độ mòn vỏ xe. Người làm thí nghiệm nhận thức được rằng các đại lượng độ mòn ắt thay đổi rất lớn từ chiếc xe này sang chiếc xe khác và rằng sự biến thiên này không thể được tách rời khỏi dữ liệu nếu các vỏ xe này được chỉ định gắn vào cho mười cái bánh xe theo một cách thức ngẫu nhiên. (Một sự chỉ định ngẫu nhiên của các vỏ xe ắt đã hàm ý rằng dữ liệu này nên được phân tích theo qui trình của Phần 8.5). Thay vào đó, một sự so sánh về độ mòn giữa hai loại vỏ xe *A* và *B* được thực hiện cho từng chiếc xe đã tạo ra năm đại lượng về sự khác biệt. Sự thiết kế này, được gọi là **kiểm định khác biệt cặp,** loại trừ được ảnh hưởng của sự thay đổi từ chiếc xe này qua chiếc xe khác và tạo ra nhiều thông tin hơn về sự khác biệt trung bình trong độ bền của hai loại vỏ xe này.

Một sự phân tích phù hợp đối với dữ liệu này ắt sẽ sử dụng năm thước đo khác nhau này để kiểm định giả thiết rằng sự khác biệt bình quân μ_d là bằng với 0 hay, tương tự như vậy, để kiểm định giả thuyết không $H_0 : \mu_d = \mu_1 - \mu_2 = 0$ so với giả thuyết thay thế $H_a : \mu_d = (\mu_1 - \mu_2) \neq 0$.

Kiểm định Khác biệt Cặp cho $(\mu_1 - \mu_2) = \mu_d$

1. Giả thuyết không: $H_0 : \mu_d = 0$

2. Giả thuyết thay thế:

Kiểm định Một phía	**Kiểm định Hai phía**
$H_a : \mu_d > 0$	$H_a : \mu_d \neq 0$
(hay $H_a : \mu_d < 0$)	

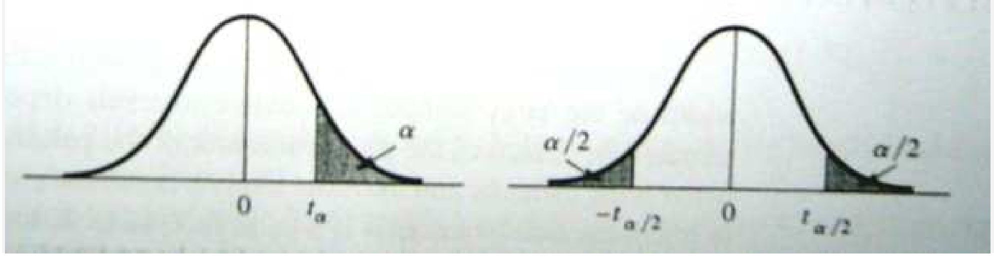
3. Trị thống kê kiểm định: $t = \dfrac{\bar{d} - 0}{s_d \sqrt{n}} = \dfrac{\bar{d}}{s_d \sqrt{n}}$

trong đó *n* = số lượng các khác biệt cặp

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n - 1}}$$

4. Vùng bác bỏ:

Kiểm định Một phía	**Kiểm định Hai phía**
$t > t_\alpha$	$t > t_{\alpha/2}$ hay $t < -t_{\alpha/2}$
[hay $t < -t_\alpha$ khi giả thuyết thay thế là $H_a : \mu_d < 0$]	



Các giá trị tới hạn của t, t_α, và $t_{\alpha/2}$, sẽ được căn cứ vào (*n* - 1) bậc tự do. Các giá trị được lập bảng có thể được tìm thấy trong Bảng 4 của Phụ lục II.

Các giả thiết: n khác biệt cặp được chọn lựa ngẫu nhiên từ một tổng thể tuân theo phân phối chuẩn.

VÍ DỤ **8.13** Liệu dữ liệu trong Bảng 8.6 có cung cấp đủ bằng chứng để cho thấy một sự khác biệt về độ mòn trung bình của các vỏ xe loại *A* và *B* không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

Lời giải Bạn có thể kiểm tra rằng trung bình và độ lệch chuẩn của năm thước đo khác biệt này là

$$\bar{d} = 0.48 \quad \text{và} \quad s_d = 0.0837$$

Sau đó

$$H_0 : \mu_d = 0 \quad \text{và} \quad H_a : \mu_d \neq 0$$

và

$$t = \frac{\bar{d} - 0}{s_d \sqrt{n}} = \frac{0.48}{0.0837/\sqrt{5}} = 12.8$$

Giá trị tới hạn này của *t* đối với một kiểm định thống kê hai phía, với $\alpha = 0.05$ và bốn bậc tự do, là 2.776. Bởi vì giá trị quan sát được này của *t* = 12.8 rơi xa khỏi vùng bác bỏ, nên có một bằng chứng khá thuyết phục về một sự khác biệt trong lượng hao mòn bình quân cho các vỏ xe loại *A* và *B*.

Bạn cũng có thể lập nên một khoảng tin cậy $(1-\alpha)100\%$ cho $(\mu_1 - \mu_2)$ dựa trên dữ liệu thu thập được từ một nghiên cứu về khác biệt cặp. Qui trình này được thể hiện trong phần trình bày sau đây.

Khoảng Tin cậy (1 - α) 100% cho Mẫu Nhỏ đối với $(\mu_1 - \mu_2) = \mu_d$ (Dựa trên Một Thí nghiệm Khác biệt Cặp)

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$$

trong đó

n = Số lượng các khác biệt cặp

và

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$$

Giả thiết: n khác biệt cặp được chọn ngẫu nhiên từ một tổng thể tuân theo phân phối chuẩn.

VÍ DỤ **8.14** Tìm khoảng tin cậy 95% cho $(\mu_1 - \mu_2) = \mu_d$ bằng cách sử dụng dữ liệu trong Bảng 8.6.

Lời giải Một khoảng tin cậy 95% cho khác biệt giữa độ mòn trung bình là

$$\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} = 0.48 \pm (2.776) \frac{0.0837}{\sqrt{5}}$$

hay 0.48 ± 0.10.

Khi các đơn vị được sử dụng để so sánh hai hay nhiều hơn các qui trình thể hiện độ biến thiên đáng lưu ý trước khi bất cứ qui trình thí nghiệm nào được tiến hành, thì tác động của sự biến thiên này có thể được giảm thiểu bằng cách so sánh các qui trình *trong nội bộ* các nhóm có những đơn vị tương đối đồng nhất được gọi là **các khối.** Theo cách nào, những tác động của các qui trình này không bị che mờ bởi độ biến thiên ban đầu trong số các đơn vị trong thí nghiệm. Một thí nghiệm được tiến hành theo cách thức này được gọi là một **thiết kế khối ngẫu nhiên hóa.** Trong một thí nghiệm có liên quan đến doanh số bán hàng hàng ngày, thì các khối có thể tượng trưng cho những ngày trong tuần; trong một thí nghiệm có liên quan đến việc tiếp thị sản phẩm, thì các khối có thể đại diện cho những khu vực địa lý. (Các thiết kế khối ngẫu nhiên hóa được thảo luận chi tiết hơn trong Phần 9.6).

Thiết kế thống kê của thí nghiệm về vỏ xe là một ví dụ đơn giản về một thiết kế khối ngẫu nhiên hóa, và trị thống kê kiểm định tạo ra thường được gọi là một kiểm định khác biệt cặp. Bạn sẽ lưu ý rằng sự kết cặp xảy ra khi thí nghiệm được lên kế hoạch và không phải sau khi dữ liệu được thu thập. Những sự so sánh về độ mòn vỏ xe được thực hiện trong nội bộ các khối tương đối đồng nhất (những chiếc xe), với những loại vỏ xe được chỉ định gắn ngẫu nhiên vào hai chiếc bánh xe.

Khối lượng thông tin có được từ việc tạo khối cho thí nghiệm về vỏ xe có thể được đo lường bởi việc so sánh khoảng tin cậy tính được cho sự phân tích không kết cặp (và sai lầm) với khoảng có được cho phân tích về khác biệt cặp. Khoảng tin cậy cho $(\mu_1 - \mu_2)$ mà ắt có lẽ đã được tính nếu như các vỏ xe này được chỉ định ngẫu nhiên cho mười bánh xe (không kết cặp) là chưa được biết hay có khả năng ắt có cùng độ lớn giống như khoảng từ -1.45 đến 2.41, mà được tính bằng cách phân tích dữ liệu quan sát được theo một cách thức không kết cặp. Việc kết cặp các loại vỏ xe trên các chiếc xe (việc tạo khối) và phân tích những khác biệt tạo ra cho ta ước lượng khoảng từ 0.38 đến 0.58. Lưu ý sự khác biệt trong những bề rộng của những khoảng này, mà chỉ ra sự gia tăng có qui mô rất đáng kể trong thông tin có được từ việc tạo khối trong thí nghiệm này.

Mặc dù việc tạo khối đã tỏ ra là rất có ích trong thí nghiệm về vỏ xe, thì có lẽ không phải lúc nào cũng xảy ra việc này. Chúng ta quan sát thấy rằng các bậc tự do sẵn có cho việc ước lượng σ^2 là ít hơn cho thí nghiệm có kết cặp so với thí nghiệm không kết cặp tương ứng. Nếu thật sự không có khác biệt nào trong số các khối, thì sự giảm sút trong các bậc tự do ắt tạo ra một sự gia tăng vừa phải trong giá trị của $t_{\alpha/2}$ được dùng trong khoảng tin cậy và vì vậy ắt gia tăng bề rộng của khoảng này. Dĩ nhiên điều này không xảy ra trong thí nghiệm về vỏ xe bởi vì sự suy giảm lớn trong sai số chuẩn của $\bar{d}$ nhiều hơn phần được bù đắp bởi sự mất mát trong các bậc tự do.

Để thực hiện sự kiểm định khác biệt cặp bằng cách sử dụng Minitab, sử dụng **Stat → Basic Statistics → Paired t.** Trong Excel, sử dụng **Tools → Data Analysis → t-Test: Paired Two Samples for Means.** Trong bất cứ chương trình nào, phải xác định các dãy dữ liệu, giả thuyết thay thế, các mức tin cậy, và sự khác biệt được giả định (thường là zêrô).

Trước khi kết luận, chúng tôi muốn nhấn mạnh lại một điểm. **Một khi bạn đã sử dụng một thiết kế kết cặp cho một thí nghiệm, thì bạn không còn chọn lựa về việc sử dụng sự phân tích không kết cặp của Phần 8.5. Những giả định mà qua đó kiểm định được căn cứ vào đã bị vi phạm. Lựa chọn thay thế duy nhất của bạn là sử dụng phương pháp đúng cho việc phân tích, kiểm định khác biệt có kết cặp (và khoảng tin cậy đi kèm) của phần này.**

Bài tập

Các Kỹ thuật Cơ bản

8.36 Một thí nghiệm về khác biệt có kết cặp được thực hiện qua việc sử dụng $n = 10$ cặp quan sát. Hãy kiểm định giả thuyết $H_0: \mu_1 - \mu_2 = 0$ so với $H_a: \mu_1 - \mu_2 \neq 0$ cho $\alpha = 0.05, \bar{d} = 0.3$, và $s_d^2 = 0.16$. Tìm giá trị *p* xấp xỉ cho kiểm định này.

8.37 Tìm khoảng tin cậy 95% cho $(\mu_1 - \mu_2)$ trong Bài tập 8.36.

8.38 Có bao nhiêu cặp quan sát mà bạn ắt cần có nếu bạn muốn ước lượng $(\mu_1 - \mu_2)$ trong Bài tập 8.36 chính xác trong giới hạn 0.1 với xác suất bằng với 95%?

8.39 Đối với một thí nghiệm về khác biệt cặp bao gồm $n = 18$ cặp, $\bar{d} = 5.7$ và $s_d^2 = 256.$ Chúng ta mong muốn xác minh $\mu_d > 0.$

a. Tìm các giả thuyết không và thay thế cho kiểm định này.

b. Thực hiện kiểm định này và phát biểu các kết luận của bạn.

8.40 Đối với một thí nghiệm về khác biệt cặp bao gồm $n = 12$ cặp, $\bar{d} = 0.13$ và $s_d^2 = 0.001.$ Tìm khoảng tin cậy 90% cho $(\mu_1 - \mu_2)$.

8.41 Một thí nghiệm về khác biệt cặp được tiến hành nhằm so sánh các số trung bình của hai tổng thể. Dữ liệu được thể hiện trong bảng đi kèm sau đây.

	Các cặp				
Tổng thể	**1**	**2**	**3**	**4**	**5**
1	1.3	1.6	1.1	1.4	1.7
2	1.2	1.5	1.1	1.2	1.8

a. Liệu dữ liệu trên có cung cấp đủ bằng chứng để chỉ ra rằng μ_1 khác với μ_2 không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

b. Tìm mức ý nghĩa xấp xỉ quan sát được cho kiểm định này, và giải thích giá trị của nó.

c. Tìm khoảng tin cậy 95% cho $(\mu_1 - \mu_2)$. So sánh sự giải thích của bạn về khoảng tin cậy này với các kết quả kiểm định trong câu (a).

d. Giả định nào mà bạn phải thực hiện để cho các suy luận của mình là có giá trị?

8.42 Một thí nghiệm về khác biệt cặp được tiến hành nhằm so sánh các số trung bình của hai tổng thể. Dữ liệu được thể hiện trong bảng đi kèm sau đây.

	Các cặp						
Tổng thể	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	8.9	8.1	9.3	7.7	10.4	8.3	7.4
2	8.8	7.4	9.0	7.8	9.9	8.1	6.9

a. Liệu dữ liệu trên có cung cấp đủ bằng chứng để chỉ ra rằng μ_1 khác với μ_2 không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.01$.

b. Tìm mức ý nghĩa xấp xỉ quan sát được cho kiểm định này, và giải thích giá trị của nó.

c. Tìm khoảng tin cậy 95% cho $(\mu_1 - \mu_2)$. So sánh sự giải thích của bạn về khoảng tin cậy này với các kết quả kiểm định trong câu (a).

d. Giả định nào mà bạn phải thực hiện để cho các suy luận của mình là có giá trị?

Các Ứng dụng

8.43 Để phản ứng lại lời than phiền rằng có một người định giá tài sản để đánh thuế cụ thể (A) là thiên lệch, một thí nghiệm được tiến hành nhằm so sánh người định giá tài sản để đánh thuế có tên trong lời than phiền này với một người định giá tài sản để đánh thuế khác (B) từ cùng một văn phòng. Tám tài sản được chọn, và mỗi tài sản được thẩm định bởi cả hai người định giá tài sản này. Các đánh giá (tính bằng ngàn đôla) được thể hiện trong bảng sau.

Tài sản	Người định giá A	Người định giá B
1	36.3	35.1
2	48.4	46.8
3	40.2	37.3
4	54.7	50.6
5	28.7	29.1
6	42.8	41.0
7	36.1	35.3
8	39.0	39.1

a. Liệu dữ liệu trên có cung cấp đủ bằng chứng để chỉ ra rằng người định giá tài sản A có xu hướng tạo ra những định giá cao hơn so với người định giá tài sản B không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

b. Ước lượng sự khác biệt trong các định giá trung bình đối với hai người định giá tài sản này.

c. Giả định nào mà bạn cần thực hiện nhằm làm cho những suy luận trong các câu (a) và (b) có giá trị?

d. Giả định rằng người định giá tài sản A đã được so sánh với một tiêu chuẩn ổn định hơn, ví dụ, mức bình quân $\bar{x}$ của các thẩm định được thực hiện bởi bốn người định giá được chọn từ văn phòng đó. Như vậy, mỗi tài sản ắt sẽ được thẩm định bởi A và còn bởi mỗi trong số bốn người định giá khác, và $x_A - \bar{x}$ ắt sẽ được tính toán. Nếu sự thẩm định trong câu (a) là có giá trị, liệu bạn có thể sử dụng kiểm định *t* về sự khác biệt cặp để kiểm định giả thuyết cho rằng sự thiên lệch, khác biệt trung bình giữa các thẩm định của A và mức trung bình của các thẩm định do bốn người thẩm định khác thực hiện, là bằng với zêrô. Hãy giải thích.

8.44 Một sự sụt giảm gần đây trong giá trị của đồng đôla so với các ngoại tệ khác được kỳ vọng sẽ làm gia tăng giá trị hàng xuất khẩu của Hoa Kỳ. Một sự so sánh về các chuyến gửi hàng của năm hiện hành so với năm trước đó (tính bằng ngàn thùng hàng) cho mỗi trong số sáu công ty xuất khẩu của Hoa Kỳ được thể hiện trong bảng đi kèm sau đây. Giả định rằng các công ty xuất khẩu này đại diện cho một mẫu ngẫu nhiên được chọn từ trong số tất cả các công ty xuất khẩu của Hoa Kỳ.

	Năm	
Công ty Xuất khẩu	Hiện hành	Năm trước
1	4.81	4.27
2	5.03	5.97

3	2.38	2.61
4	4.26	3.96
5	5.14	4.86
6	3.93	3.17

a. Liệu dữ liệu trên có cung cấp đủ bằng chứng để chỉ ra một sự gia tăng trong số thùng hàng trung bình được xuất khẩu từ năm trước so với năm nay không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

b. Tìm giá trị *p* xấp xỉ cho kiểm định này.

c. Tìm khoảng tin cậy 95% cho sự gia tăng trung bình trong số lượng thùng hàng được gởi đi.

8.45 Nỗ lực khuyến khích khách hàng thực hiện thanh toán sớm cho các hóa đơn, một nhà quản lý của một công ty tư vấn đề nghị khách hàng một tỷ lệ chiết khấu 2% trên các hóa đơn được thanh toán trong vòng 30 ngày kể từ ngày xuất hóa đơn đó. Nhằm đánh giá ảnh hưởng của chính sách mới này đối với việc thanh toán đúng hạn, nhà quản lý này đã chọn mẫu ngẫu nhiên 15 khách hàng và ghi nhận số ngày để thanh toán cho hóa đơn cuối cùng được xuất theo hệ thống cũ và hóa đơn đầu tiên được xuất theo hệ thống khuyến khích. Dữ liệu, tính bằng ngày, được trình bày trong bảng đi kèm dưới đây.

	Các cặp														
Hóa đơn	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Hệ thống cũ	92	88	65	85	96	64	65	62	90	89	65	75	84	90	80
Hệ thống khuyến khích	28	30	29	85	29	28	26	29	88	30	70	30	27	92	29

a. Liệu dữ liệu trên có cung cấp đủ bằng chứng để chỉ ra rằng thời gian thanh toán bình quân được giảm bớt trong hệ thống khuyến khích không? Giá trị *p* của kiểm định này là bao nhiêu?

b. Tìm khoảng tin cậy 95% cho sự giảm đi về thời gian thanh toán trung bình mỗi khách hàng sau khi hệ thống khuyến khích được áp dụng. Giải thích khoảng này.

8.46 Liệu nhạc nền có ảnh hưởng thái độ của người đi mua hàng tại siêu thị? Một thí nghiệm được thiết kế để trả lời câu hỏi này được tiến hành tại một siêu thị trong suốt các tháng mua hàng trong mùa hè tương đối ổn định. Hai ngày được chọn vào giữa tuần. Một ngày được ngẫu nhiên chỉ định để không mở nhạc nền. Trong suốt ngày thứ hài, nhạc nền có tiết tấu chậm được chơi. Doanh số bán hàng hàng ngày (tính bằng đôla) cho 12 tuần được thể hiện trong bảng kèm theo sau đây. Hãy sử dụng bản in Minitab để trả lời cho câu hỏi sau đây.

	Tuần											
Nhạc	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Không có	14,172	15.485	13,922	12,204	15,501	15,106	14,608	13,946	15,002	14,670	16,202	13,286
Tiết tấu chậm	15,917	16,110	14,818	14,709	13,982	16,416	14,727	14,823	14,825	15,949	15,488	14,955

```
K•T C•P T CHO KHÔNG CÓ - CH•M
                              N       TRUNG BÌNH    •• L•CH CHU•N    SAI S• CHU•N
                                                                    C•A TRUNG BÌNH
KHÔNG                        12            14509             1088               314
CH•M                         12            15143              833               240
KHÁC BI•T                    12             -635             1154               333
90% kho•ng tin c•y cho s• khác bi•t trung bình: (-1233, -36)
```

```
Ki•m ••nh T cho s• khác bi•t trung bình = 0 (so v•i không ph•i = 0) = -1.90
Giá tr• p = 0.083
```

a. Liệu dữ liệu này có cung cấp bằng chứng đủ để chỉ ra một sự khác biệt về doanh số bán hàng trung bình hàng ngày cho những ngày khi mà không có nhạc nền được chơi so với những ngày mà nhạc nền có tiết tấu chậm được chơi không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.10$.

b. Tìm khoảng tin cậy 90% cho sự khác biệt về doanh số bán hàng trung bình hàng ngày cho những ngày không có nhạc so với các ngày có nhạc tiết tấu chậm.

8.7 KIỂM ĐỊNH GIẢ THUYẾT CHO MẪU LỚN ĐỐI VỚI MỘT TỶ LỆ NHỊ THỨC

Khi một mẫu ngẫu nhiên gồm *n* lần thử nghiệm giống nhau được rút ra từ một tổng thể nhị thức, thì tỷ lệ mẫu $\hat{p}$ có một phân phối xấp xỉ chuẩn khi *n* là lớn, với trung bình *p* và độ lệch chuẩn

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

Để kiểm định một giả thuyết thuộc dạng

$$H_0 : p = p_0$$

so với một giả thuyết thay thế một phía hay hai phía

$$H_a : p \neq p_0 \quad \text{hay} \quad H_a : p > p_0 \quad \text{hay} \quad H_a : p < p_0$$

trị thống kê này được lập nên bằng cách sử dụng $\hat{p}$, số ước lượng tốt nhất của tỷ lệ tổng thể thực sự *p*. Kiểm định cho mẫu lớn này được tóm tắt trong phần trình bày sau đây.

Kiểm định cho Mẫu Lớn đối với một Tỷ lệ Tổng thể *p*

1. Giả thuyết không: $H_0 : p = p_0$

2. Giả thuyết thay thế:

Kiểm định Một phía	**Kiểm định Hai phía**
$H_a : p > p_0$	$H_a : p \neq p_0$
[hay $H_a : p < p_0$]	

3. Trị thống kê kiểm định: $z = \dfrac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$, với $\hat{p} = \dfrac{x}{n}$

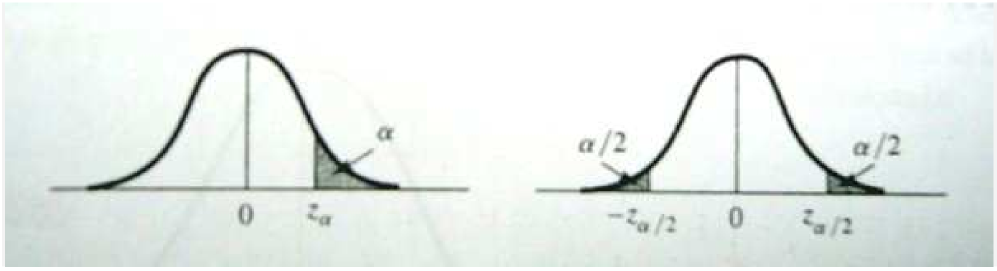
trong đó *x* là số lượng các thành công trong *n* lần thử nhị thức[+]

4. Vùng bác bỏ:

[+] Một trị thống kê tương đương được tìm ra bằng cách nhân tử số và mẫu số của *z* với *n* để có được

$$z = \frac{x - np_0}{\sqrt{np_0 q_0}}$$

Kiểm định Một phía	**Kiểm định Hai phía**
$z > z_\alpha$ (hay $z < -z_\alpha$ khi giả thuyết thay thế là $H_a: p < p_0$)	$z > z_{\alpha/2}$ hay $z < -z_{\alpha/2}$



Các giả thiết: Việc chọn mẫu này thỏa mãn các giả định về một thí nghiệm nhị thức (Phần 4.2), và *n* là đủ lớn để cho phân phối mẫu của $\hat{p}$ có thể được ước lượng xấp xỉ bởi một phân phối chuẩn. Khoảng $p \pm 2\sigma_{\hat{p}}$ phải được chứa trong khoảng từ 0 đến 1.

VÍ DỤ **8.15** Xấp xỉ 1 trên 10 người tiêu dùng ưa thích nhãn hiệu cola A hơn. Sau một chiến dịch quảng cáo tại một khu vực bán hàng đã biết, 200 người uống cola được chọn ngẫu nhiên từ các khách hàng tại một khu vực chợ và được phỏng vấn để xác định tính hiệu quả của chiến dịch này. Kết quả của cuộc điều tra đã cho thấy rằng một tổng số gồm 26 người biểu lộ một sự ưa thích đối với nhãn hiệu cola A. Liệu những dữ liệu này có cung cấp đủ bằng chứng để cho thấy một sự gia tăng trong mức độ chấp nhận nhãn hiệu A tại khu vực đó?

Lời giải Chúng ta giả định rằng mẫu này đáp ứng các yêu cầu của một thí nghiệm nhị thức. Câu hỏi này có thể được trả lời bằng cách kiểm định giả thiết

$$H_0: p = 0.10$$

so với giả thiết

$$H_a: p > 0.10$$

Một kiểm định một phía được sử dụng bởi vì chúng ta muốn xác minh liệu giá trị của *p* có lớn hơn 0.10. (Chiến dịch quảng cáo này đã tạo ra một *sự gia tăng* trong giá trị của *p*.)

Số ước lượng điểm của *p* là $\hat{p} = x/n$, và trị thống kê kiểm định là

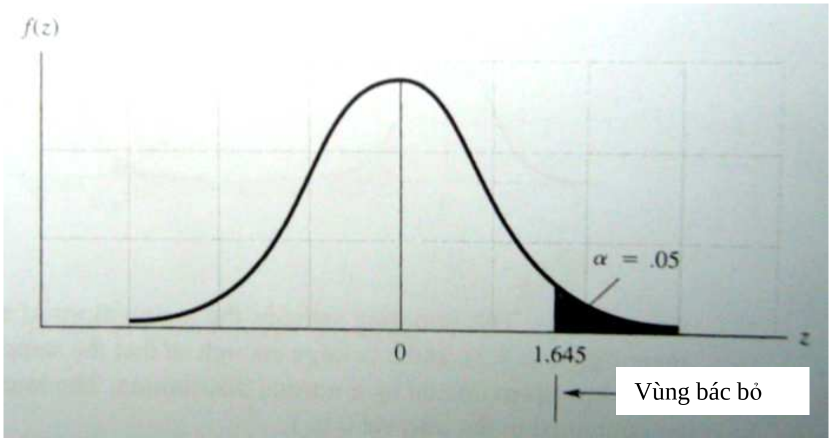
$$z = \frac{\hat{p} - p_0}{\sqrt{p_o q_o / n}}$$

Khi H_0 là đúng, thì giá trị của *p* là $p_0 = 0.1$, và phân phối mẫu của $\hat{p}$ có một trung bình bằng với p_0 và một độ lệch chuẩn là $\sqrt{p_o q_o / n}$. **Vì thế,** $\sqrt{\hat{p}\hat{q}/n}$ **không được sử dụng để ước lượng sai số chuẩn của** $\hat{p}$ **trong trường hợp này bởi vì trị thống kê này được tính toán theo giả định rằng** H_0 **là đúng.** (Khi ước lượng giá trị của *p* bằng cách sử dụng số ước lượng $\hat{p}$, thì sai số chuẩn của $\hat{p}$ là không được biết và được *ước lượng* bởi $\sqrt{\hat{p}\hat{q}/n}$ **.**)

Với α = 0.05, chúng ta ắt bác bỏ H_0 khi *z* > 1.645 (Hình 8.13). Với $\hat{p} = 26/200 = 0.13$, thì giá trị của trị thống kê này là

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \frac{0.13 - 0.10}{\sqrt{\dfrac{(0.10)(0.90)}{200}}} = 1.41$$

HÌNH **8.13** Vị trí của vùng bác bỏ trong Ví dụ 8.15



Giá trị *p* đi cùng với sự kiểm định này là

giá trị *p* = *P* (*z* > 1.41) = 0.5000 - 0.4207 = 0.0793

Giá trị tính toán được này của trị thống kê kiểm định không rơi vào trong vùng bác bỏ: giá trị *p* là 0.0793, và vì vậy *chúng ta không bác bỏ* H_0.

Liệu chúng ta có chấp nhận H_0? Không, cho đến khi chúng ta đã xác định các giá trị thay thế của *p* khác với $p_0 = 0.1$ mà *có ý nghĩa thực tiễn.* Xác suất của một sai lầm loại II nên được tính toán bằng cách sử dụng những giá trị thay thế này. Nếu β là đủ nhỏ, thì chúng ta ắt chấp nhận H_0 với sự quan tâm đầy đủ đến rủi ro của một quyết định sai lầm.

Các gợi ý khi giải toán

Khi kiểm định một giả thuyết có liên quan đến *p*, hãy sử dụng p_0 (chứ không phải $\hat{p}$) để tính toán $\sigma_{\hat{p}}$ ở mẫu số của trị thống kê *z*. Lý do cho việc này là rằng vùng bác bỏ được xác định bởi phân phối của $\hat{p}$ khi giả thuyết không là đúng, nghĩa là, khi $p = p_0$.

Các Ví dụ 8.4 và 8.15 minh họa cho một điểm quan trọng. **Nếu dữ liệu cung cấp đủ bằng chứng để bác bỏ H_0, thì xác suất của một quyết định sai lầm α được biết trước bởi vì α được sử dụng trong việc định vị vùng bác bỏ. Bởi vì α thường là nhỏ, nên chúng ta khá chắc chắn rằng chúng ta đã thực hiện một quyết định chính xác.** Mặt khác, nếu dữ liệu không cung cấp đủ bằng chứng để bác bỏ H_0, thì các kết luận là không rõ ràng. Lý tưởng là theo qui trình kiểm định thống kê được trình bày trong Phần 8.2, chúng ta ắt đã xác định cụ thể trước một p_a thay thế có ý nghĩa quan trọng và chọn *n* sao cho β ắt sẽ nhỏ. Thật không may là nhiều thí nghiệm không đuợc tiến hành theo cách thức lý tưởng này. Một người nào đó chọn một cỡ mẫu, và người làm thí nghiệm hay nhà thống kê bị buộc phải đánh giá bằng chứng.

Việc tính toán β là không quá khó khăn đối với thủ tục kiểm định thống kê được trình bày trong phần nay nhưng có lẽ là cực kỳ khó khăn, nếu không muốn nói là vượt quá khả năng của một người mới bắt đầu, trong các tình huống kiểm định khác. **Một qui trình đơn giản hơn nhiều là *không bác bỏ H_0* hơn là chấp nhận nó, và sau đó ước lượng bằng cách sử dụng một khoảng tin cậy.** Khoảng này sẽ cho bạn một dãy các giá trị đáng tin cậy của *p*.

Bài tập

Các Kỹ thuật Cơ bản

8.47 Một mẫu ngẫu nhiên gồm $n = 1000$ quan sát từ một tổng thể nhị thức tạo ra $x = 279$.

a. Nếu giả thuyết nghiên cứu của bạn là rằng *p* nhỏ hơn 0.3, thì bạn nên chọn giả thuyết thay thế nào? giả thuyết không của bạn?

b. Liệu giả thuyết thay thế của bạn trong câu (a) có hàm ý một kiểm định thống kê một phía hay hai phía?

c. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng *p* là nhỏ hơn 0.3 không? Hãy kiểm định qua việc sử dụng $\alpha = 0.05$.

8.48 Một mẫu ngẫu nhiên gồm $n = 2000$ quan sát từ một tổng thể nhị thức tạo ra $x = 1238$.

a. Nếu giả thuyết nghiên cứu của bạn là rằng *p* lớn hơn 0.6, thì bạn nên chọn giả thuyết thay thế nào? giả thuyết không của bạn?

b. Liệu giả thuyết thay thế của bạn trong câu (a) có hàm ý một kiểm định thống kê một phía hay hai phía?

c. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng *p* là lớn hơn 0.6 không? Hãy kiểm định qua việc sử dụng $\alpha = 0.05$.

8.49 Một mẫu ngẫu nhiên gồm 120 quan sát được chọn từ một tổng thể nhị thức, và 72 lần thành công được quan sát. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng *p* là lớn hơn 0.5 không? Hãy kiểm định qua việc sử dụng $\alpha = 0.05$.

8.50 Tham khảo lại Bài tập 8.49. Giá trị *p* cho kiểm định này là bao nhiêu? Liệu giá trị *p* này có khẳng định kết luận của bạn trong Bài tập 8.49 không? Hãy giải thích.

Các Ứng dụng

8.51 Trong cuộc điều tra năm 1993 về lứa sinh viên năm thứ nhất đại học được báo cáo trong Bài tập 6.43 ("Thống kê về sinh viên năm nhất", 1994). 86% sinh viên đại học năm thứ nhất trong cuộc điều tra này đã nhận được sự hỗ trợ tài chính từ cha mẹ hay gia đình họ. Trong một cuộc điều tra tương tự, một mẫu gồm $n = 1000$ sinh viên đại học năm thứ nhất được chọn ngẫu nhiên từ lứa sinh viên năm thứ nhất đó đã thể hiện cho thấy rằng 89% các sinh viên trong cuộc điều tra này nhận được sự hỗ trợ tài chính từ cha mẹ hay gia đình họ.

a. Giả định rằng tỷ lệ phần trăm của năm 1993, 86% trong thực tế là giá trị của tổng thể và sử dụng một kiểm định cho giả thuyết với mức ý nghĩa 0.05 để xác định liệu tỷ lệ phần trăm của năm nay có thể hiện cho thấy rằng con số này đã thay đổi theo thời gian và không còn chính xác nữa.

b. Giá trị *p* đi cùng với kiểm định này là bao nhiêu? Liệu kết luận của bạn khi sử dụng giá trị *p* này có nhất quán với các kết quả của kiểm định này trong câu (a) không?

8.52 Một dịch vụ chuyển séc thành tiền mặt đã tìm thấy rằng xấp xỉ 5% của tất cả các tấm séc đưa đến dịch vụ này để đổi thành tiền mặt là giả. Sau khi thực hiện một hệ thống kiểm tra séc nhằm giảm thiểu tổn thất của mình, dịch vụ này nhận thấy chỉ có 45 tấm séc là giả mạo trong tổng số 1124 tấm séc được chuyển thành tiền mặt.

a. Nếu bạn mong muốn thực hiện một kiểm định thống kê để xác định liệu hệ thống kiểm tra séc này có làm giảm xác suất để cho một tấm séc giả mạo được chuyển thành tiền mặt không, thì bạn nên chọn lựa giả thuyết thay thế nào? giả thuyết không nào?

b. Liệu giả thuyết thay thế của bạn trong câu (a) có hàm ý một kiểm định một phía hay hai phía không? Hãy giải thích.

c. Khi lưu ý dữ liệu này, trực giác của bạn nói với bạn điều gì? Liệu bạn có nghĩ rằng hệ thống kiểm tra séc này là hiệu quả trong việc giảm bớt tỷ lệ các tấm séc giả mạo mà đã được chuyển thành tiền mặt không?

d. Thực hiện một kiểm định thống kê cho giả thuyết không trong câu (a), và phát biểu các kết luận của bạn. Hãy kiểm định qua việc sử dụng α = 0.05. Liệu các kết luận của kiểm định này có nhất quán với trực giác của bạn trong câu (c) không?

8.53 Từ kinh nghiệm trong quá khứ, một đại lý bán hàng gia dụng đã tìm thấy rằng 10% khách hàng của cô ta mua hàng trả góp đã thanh toán hết hóa đơn của mình trước khi đến kỳ hạn thanh toán của lần trả góp hàng tháng cuối cùng (lần thứ 24). Nghi ngờ một sự gia tăng trong tỷ lệ phần trăm này, người chủ đại lý này đã điều tra 200 người mua hàng trả góp về ý định của họ. Trong số này, 33 người khẳng định rằng họ có kế hoạch thanh toán hết khoản nợ của mình trước kỳ trả góp cuối cùng. Liệu dữ liệu này có cung cấp bằng chứng đủ để cho thấy rằng tỷ lệ phần trăm những người mua hàng trả góp mà sẽ thanh toán hết khoản nợ của mình trước kỳ trả chậm cuối cùng có vượt quá 10% không?

a. Xác định giả thuyết thay thế cho kiểm định này.

b. Phát biểu giả thuyết không.

c. Tìm vùng bác bỏ cho α = 0.05.

d. Tiến hành kiểm định này và phát biểu các kết luận của bạn.

8.54 Một nhà xuất bản một tạp chí tin tức đã tìm thấy từ kinh nghiệm trong quá khứ rằng 60% những người đặt mua dài hạn tạp chí của họ đã gia hạn việc đặt mua dài hạn của mình. Bởi vì đang hướng đến một cuộc khủng hoảng kinh doanh, nên công ty này quyết định chọn ngẫu nhiên một mẫu nhỏ những người đặt mua dài hạn và, thông qua việc đặt câu hỏi qua điện thoại, xác định liệu họ có kế hoạch gia hạn việc đặt mua dài hạn của mình hay không. Một trăm lẻ tám trong số 200 đã xác nhận rằng họ có kế hoạch gia hạn việc đặt mua dài hạn của mình.

a. Nếu bạn muốn xác minh liệu dữ liệu này có cung cấp đủ bằng chứng về một sự sụt giảm trong tỷ lệ p của tất cả những người đặt mua báo dài hạn mà sẽ gia hạn, thì bạn sẽ chọn lựa giả thuyết thay thế của mình như thế nào? giả thuyết không của bạn như thế nào?

b. Tiến hành kiểm định này với việc sử dụng α = 0.05. Phát biểu các kết quả.

c. Tìm khoảng tin cậy 95% cho p.

d. Có bao nhiêu người đặt mua báo dài hạn ắt đã được đưa vào trong mẫu của nhà xuất bản này nhằm ước lượng p nằm trong giới hạn 0.01, với mức tin cậy 95%?

8.8 MỘT KIỂM ĐỊNH CHO MẪU LỚN ĐỐI VỚI GIẢ THUYẾT VỀ SỰ KHÁC BIỆT GIỮA HAI TỶ LỆ NHỊ THỨC

Khi trọng tâm của một thí nghiệm hay một nghiên cứu là ở sự khác biệt trong tỷ lệ của các cá nhân hay vật phẩm sở hữu một đặc tính cụ thể, thì trị thống kê vô cùng quan trọng cho việc kiểm định các giả thuyết về $p_1 - p_2$ là sự khác biệ trong các tỷ lệ của mẫu $\hat{p}_1 - \hat{p}_2$. Qui trình kiểm định chính thức có liên quan đến sự khác biệt trong các tỷ lệ tổng thể được thể hiện trong phần trình bày sau đây.

Một Kiểm định Thống kê cho Mẫu Lớn đối với (p_1 - p_2)

1. Giả thuyết không: $H_0 : (p_1 - p_2) = D_0$ trong đó D_0 là một sự khác biệt đã được xác định cụ thể mà bạn mong muốn kiểm định. Đối với nhiều kiểm định, bạn sẽ mong muốn giả thuyết rằng không có khác biệt nào giữa p_1 và p_2; nghĩa là,

$$D_0 = 0$$

2. Giả thuyết thay thế:

Kiểm định Một phía	**Kiểm định Hai phía**
$H_a : (p_1 - p_2) > D_0$	$H_a : (p_1 - p_2) \neq D_0$
[hay $H_a : (p_1 - p_2) < D_0$]	

3. Trong việc chọn lựa trị thống kê kiểm định thích hợp, chúng ta bắt đầu với:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sigma_{(\hat{p}_1 - \hat{p}_2)}} = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}}$$

trong đó $\hat{p}_1 = x_1 / n_1$ và $\hat{p}_2 = x_2 / n_2$. Bởi vì p_1 và p_2 là chưa được biết, nên chúng ta cần phải ước lượng xấp xỉ các giá trị của chúng nhằm tính toán độ lệch chuẩn $\hat{p}_1 - \hat{p}_2$ mà xuất hiện trong mẫu số của trị thống kê *z*. Các ước lượng xấp xỉ này là sẵn có cho hai trường hợp.

Trường hợp I: Nếu chúng ta giả thuyết rằng p_1 bằng với p_2, nghĩa là

$$H_0 : p_1 = p_2$$

hay tương đương rằng

$$p_1 - p_2 = 0$$

thì $p_1 = p_2 = p$ và ước lượng tốt nhất của *p* đạt được bằng cách gộp lại dữ liệu từ cả hai mẫu. Vì thế, nếu x_1 và x_2 là các số lượng thành công có được từ hai mẫu này, thì

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Trị thống kê kiểm định ắt sẽ là

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}} \text{ hay } z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Trường hợp II: Mặt khác, nếu ta giả thuyết rằng D_0 là *không* bằng với zêrô, nghĩa là,

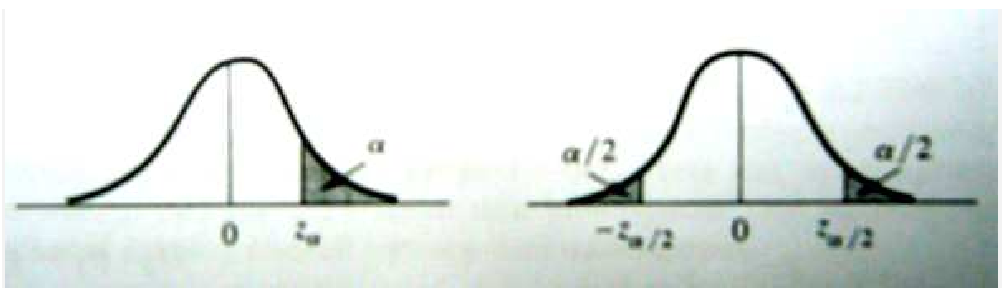
$$H_0 : (p_1 - p_2) = D_0$$

trong đó $D_0 \neq 0,$ thì các ước lượng tốt nhất của p_1 và p_2 lần lượt là $\hat{p}_1$ và $\hat{p}_2$. Trị thống kê kiểm định ắt sẽ là

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}$$

4. Vùng bác bỏ:

Kiểm định Một phía	**Kiểm định Hai phía**
$z > z_\alpha$	$z > z_{\alpha/2}$ hay $z < -z_{\alpha/2}$
[hay $z < -z_\alpha$ khi giả thuyết thay thế là $H_a : (p_1 - p_2) < D_0$]	



Các giả thiết: Các mẫu được chọn theo một cách thức ngẫu nhiên và độc lập từ hai tổng thể nhị thức, và n_1 và n_2 là đủ lớn để cho phân phối mẫu của $\hat{p}_1 - \hat{p}_2$ có thể được ước lượng xấp xỉ bởi một phân phối chuẩn. Khoảng $(p_1 - p_2) \pm 2\sigma_{(\hat{p}_1 - \hat{p}_2)}$ phải được chứa trong khoảng từ -1 đến 1.

VÍ DỤ **8.16** Một người quản lý bệnh viện nghi ngờ rằng trễ hạn trong việc thanh toánh các hóa đơn viện phí đã gia tăng trong năm vừa qua. Hồ sơ lưu trữ của bệnh viện cho thấy rằng các hóa đơn của 48 trong số 1284 người nhập viện trong tháng Tư đã trễ hạn trong hơn 90 ngày. Con số này so với 34 trong số 1002 người nhập viện trong cùng tháng này năm trước đó. Liệu những dữ liệu này có cung cấp đủ bằng chứng để cho thấy có một sự gia tăng trong tỷ lệ trễ hạn thanh toán vượ quá 90 ngày không? Hãy kiểm định qua việc sử dụng α = 0.10.

Lời giải Đặt p_1 và p_2 lần lượt tượng trưng cho các tỷ lệ của tất cả các ca nhập viện tiềm tàng vào tháng Tư của năm này và năm trước đó, mà ắt đã cho phép các khách hàng của mình trễ hạn trong một giai đoạn vượt quá 90 ngày, và đặt $n_1 = 1284$ ca nhập viện trong năm nay và $n_2 = 1002$ ca nhập viện năm trước đó tượng trưng cho các mẫu ngẫu nhiên độc lập từ các tổng thể này. Bởi vì chúng ta muốn xác mình một sự gia tăng trong tỷ lệ trễ hạn, nếu có hiện hữu, cho nên chúng ta sẽ kiểm định giả thuyết không

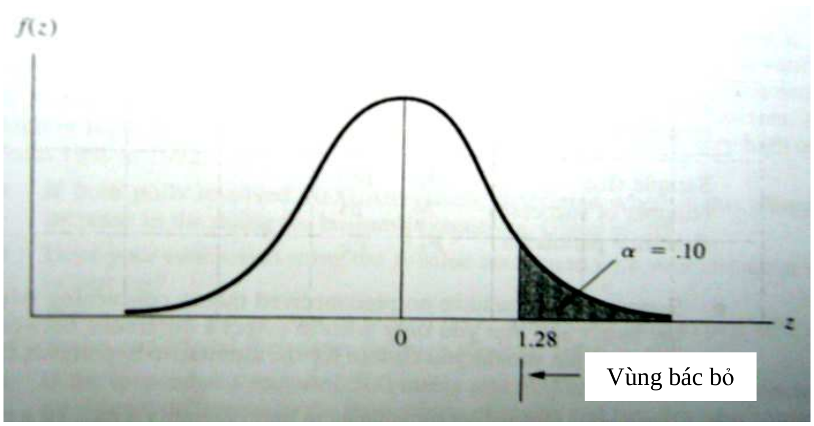
$$H_0 : p_1 = p_2, \text{ nghĩa là, } p_1 - p_2 = D_0 = 0$$

so với

$$H_a : p_1 > p_2, \text{ nghĩa là, } p_1 - p_2 > 0$$

Để tiến hành kiểm định này, chúng ta sẽ sử dụng trị thống kê *z* và ước lượng xấp xỉ giá trị của $\sigma_{(\hat{p}_1 - \hat{p}_2)}$ bằng cách sử dụng ước lượng gộp của *p* được mô tả trong Trường hợp I. Bởi vì H_a hàm ý một kiểm định một phía, cho nên chúng ta sẽ bác bỏ H_0 chỉ cho những giá trị lớn của *z*. Vì vậy, với $\alpha = 0.10$ thì chúng ta bác bỏ H_0 nếu *z* > 1.28 (xem Hình 8.14).

HÌNH **8.14** Vị trí của vùng bác bỏ trong Ví dụ 8.16



Các ước lượng của p_1 và p_2 là

$$\hat{p}_1 = \frac{48}{1284} = 0.0374 \quad \text{và} \quad \hat{p}_2 = \frac{34}{1002} = 0.0339$$

Ước lượng gộp của *p* cần có cho $\sigma_{(\hat{p}_1 - \hat{p}_2)}$ là

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{48 + 34}{1284 + 1002} = 0.0359$$

Trị thống kê kiểm định là

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.0374 - 0.0339}{\sqrt{(0.0359)(0.0961)\left(\frac{1}{1284} + \frac{1}{1002}\right)}} = 0.45$$

Giá trị *p* của kiểm định này là

giá trị *p* = *P*(*z* > 0.45) = 0.5000 - 0.1736 = 0.3264

Bởi vì giá trị tính toán được này của *z* không rơi vào vùng bác bỏ và giá trị *p* là khá lớn, cho nên chúng ta không thể bác bỏ giả thuyết không rằng $p_1 = p_2$. Dữ liệu này cung cấp không đủ bằng chứng để chỉ ra rằng tỷ lệ các khách hàng trễ hạn thanh toán vào thán Tư của năm nay vượt quá tỷ lệ tương ứng của năm trước đó.

Bài tập

Các Kỹ thuật Cơ bản

8.55 Các mẫu ngẫu nhiên độc lập gồm $n_1 = 140$ và $n_2 = 140$ quan sát được chọn ngẫu nhiên lần lượt từ hai tổng thể nhị thức 1 và 2. Số lượng thành công trong các mẫu này và những tham số của tổng thể được cho trong bảng đi kèm sau đây.

	Tổng thể	
Các Trị thống kê và Tham số	**1**	**2**
Cỡ mẫu	140	140
Số lượng thành công	74	81
Tham số nhị thức	p_1	p_2

a. Giả định rằng bạn không có lý thuyết nào được thừa nhận trước đó liên quan đến tham số nào, p_1 hay p_2, là lớn hơn và rằng bạn chỉ mong muốn xác minh một sự khác biệt giữa hai tham số này, nếu có hiện hữu. Bạn nên chọn giả thuyết thay thế nào cho một kiểm định thống kê? giả thuyết không nào?

b. Liệu giả thuyết thay thế của bạn trong câu (a) có hàm ý một kiểm định một phía hay hai phía không?

c. Tiến hành kiểm định này và phát biểu các kết luận của bạn. Kiểm định bằng cách sử dụng $\alpha = 0.05$.

8.56 Tham khảo lại Bài tập 8.55. Giả định rằng vì các lý do thực tiễn mà bạn biết rằng p_1 không thể lớn hơn p_2.

a. Với kiến thức đã biết này, bạn nên chọn giả thuyết thay thế nào cho kiểm định thống kê của mình? giả thuyết không của bạn là gì?

b. Liệu giả thuyết thay thế của bạn trong câu (a) có hàm ý một kiểm định một phía hay hai phía không?

c. Tiến hành kiểm định này và phát biểu các kết luận của bạn. Kiểm định bằng cách sử dụng $\alpha = 0.10$.

8.57 Các mẫu ngẫu nhiên độc lập gồm $n_1 = 280$ và $n_2 = 350$ quan sát được chọn ngẫu nhiên lần lượt từ các tổng thể nhị thức 1 và 2. Số lượng thành công trong mẫu và các tham số của tổng thể được thể hiện trong bảng đi kèm sau đây.

	Tổng thể	
Các Trị thống kê và Tham số	**1**	**2**
Cỡ mẫu	280	350
Số lượng thành công	132	178
Tham số nhị thức	p_1	p_2

a. Giả định rằng bạn biết rằng p_1 không bao giờ có thể lớn hơn p_2, và bạn muốn biết rằng liệu p_1 có nhỏ hơn p_2 không. Bạn nên chọn giả thuyết thay thế và giả thuyết không như thế nào?

b. Liệu giả thuyết thay thế của bạn trong câu (a) có hàm ý một kiểm định một phía hay hai phía không?

c. Tiến hành kiểm định này và phát biểu các kết luận của bạn. Kiểm định bằng cách sử dụng $\alpha = 0.05$.

Các Ứng dụng

8.58 Một nhà sản xuất điều chỉnh một dây chuyền sản xuất nhằm giảm bớt tỷ lệ trung bình sản phẩm mắc lỗi. Để xác định liệu sự điều chỉnh này có hiệu quả không, nhà sản xuất này đã chọn mẫu ngẫu nhiên 400 sản phẩm trước khi điều chỉnh dây chuyền sản xuất này và 400 sản phẩm sau khi điều chỉnh. Tỷ lệ phần trăm các sản phẩm mắc lỗi trong những mẫu này là

trước: 5.25%
sau: 3.5%

a. Nếu sự điều chỉnh này không thể có khả năng gia tăng tỷ lệ sản phẩm mắc lỗi, thì nhà sản xuất này nên chọn giả thuyết thay thế như thế nào? giả thuyết không ra sao?

b. Tiến hành kiểm định này bằng cách sử dụng $\alpha = 0.05$. Giải thích các kết quả này.

8.59 Trong một cuộc trưng cầu ý kiến (Gallup, 1994, trang 135) về tính trung thực và các tiêu chuẩn đạo đức, $n = 1000$ người Mỹ được hỏi để xếp hạng về tính trung thức và các tiêu chuẩn đạo đức của những người trong các lĩnh vực khác nhau là rất cao, cao, trung bình, thấp, và rất thấp. Mặc dù khoảng 50% những người được chọn mẫu đã xếp hạng những người trong phần lớn các lĩnh vực là trung bình, thì hơn 50% xếp hạng những người bán dược phẩm, dược sĩ, bác sĩ và thư ký là rất cao hay cao. Mặt khác, chỉ có 20% xếp hạng những chuyên viên kinh doanh như là rất cao hay cao vào năm 1993, tăng từ mức 18% trong năm 1992.

a. Nếu cả hai cuộc trưng cầu ý kiến này có liên quan đến 1000 người Mỹ, hãy xác định liệu sự thay đổi này có phản ảnh một sự gia tăng cơ bản trong vị trí xếp hạng của những chuyên viên kinh doanh không. Hãy sử dụng $\alpha = 0.05$.

b. Liệu kết luận của bạn khi sử dụng giá trị *p* đi cùng với kiểm định có nhất quán với các kết luận của bạn trong câu (a) không?

8.60 Trong một cuộc trưng cầu ý kiến liên quan đến số lượng các khoản thuế liên bang mà người Mỹ phải trả, 49% đàn ông và 60% phụ nữ đã thừa nhận các khoản thuế mình phải trả là quá cao (Gallup, 1994, trang 69).

a. Nếu những người trả lời bao gồm 500 đàn ông và 500 phụ nữ, liệu chúng ta có thể kết luận rằng có một sự khác biệt đáng kể trong các tỷ lệ đàn ông và phụ nữ mà xem các khoản thuế họ phải trả là quá cao không? Hãy sử dụng $\alpha = 0.01$.

b. Liệu giá trị *p* có ủng hộ cho các kết luận của bạn trong câu (a) không?

8.61 Mặc dù sự kiện rằng phần lớn những lữ khách kinh doanh là các khách hàng lớn nhất của những hãng hàng không, họ trả những giá cao nhất. Các lữ khách kinh doanh thường lên kế hoạch các chuyến đi rất gấp gáp và nhiều khả năng dàn xếp lại những kế hoạch di chuyển của họ vào phút cuối. Kết quả là, họ thường không thể tận dụng được các mức giá chiết khấu. Trong nghiên cứu về chi phí di chuyển bằng đường hàng không của doanh nghiệp, RichardJ. Fox và Frederick J. Stephenson (1990) báo cáo rằng 45% trong số 56 công ty là người sử dụng lớn (có chi phí di chuyển hàng năm tối thiểu là $10 triệu) và 76% trong số 145 công ty là người sử dụng nhỏ (có chi phí di chuyển hàng năm thấp hơn $10 triệu) đã thiết lập các đại lý lữ hành tại công ty của mình.

a. Hãy kiểm định một sự khác biệt có ý nghĩa trong tỷ lệ những người sử dụng lớn và nhỏ mà đã thiết lập các đại lý lữ hành tại công ty mình. Hãy sử dụng $\alpha = 0.01$.

b. Lập một ước lượng khoảng tin cậy 90% cho sự khác biệt trong tỷ lệ của những người sử dụng lớn và nhỏ mà đã hình thành các đại lý lữ hành tại công ty mình. Liệu sự ước lượng này có khẳng định kết quả trong câu (a) không?

8.62 Theo một báo cáo của Hiệp hội Ung thư Hoa Kỳ ("Hồ sơ những người hút thuốc", 1990), có nhiều đàn ông hơn phụ nữ hút thuốc và số người hút thuốc chết sớm cao gấp hai lần những người không hút thuốc. Trong các mẫu ngẫu nhiên gồm 200 đàn ông và 200 phụ nữ, 62 đàn ông và 54 phụ nữ được điều tra là những người hút thuốc.

a. Liệu có đủ bằng chứng để kết luận rằng tỷ lệ của những người đàn ông hút thuốc khác với tỷ lệ của những người hút thuốc là phụ nữ không?

b. Giá trị *p* của kiểm định trong câu (a) là bao nhiêu?

8.9 CÁC SUY LUẬN LIÊN QUAN ĐẾN PHƯƠNG SAI TỔNG THỂ

Chúng ta đã thấy trong các phần trước rằng một sự ước lượng về phương sai của tổng thể σ^2 đóng vai trò nền tảng cho những qui trình thực hiện các suy luận về những số trung bình tổng thể. Hơn nữa, có nhiều tình huống thực tiễn mà ở đó σ^2 là mục tiêu chủ yếu của một sự điều tra thực nghiệm; vì vậy đại lượng này có lẽ đóng một vị trí có tầm quan trọng lớn hơn nhiều so với số trung bình tổng thể.

Các công cụ đo lường khoa học phải cung cấp được những kết quả không bị lệch với một sai số đo lường rất nhỏ. Một đồng hồ đo độ cao của máy bay mà đo lường độ cao chính xác *nhìn chung* ắt sẽ có ít giá trị nếu độ lệch chuẩn của sai số đo lường là 5000 bộ. Trên thực tế, sự lệch trong một công cụ đo lường thường có thể được điều chỉnh cho chính xác, được đo bằng độ lệch chuẩn của sai số đo lường, thường là một hàm số của sự thiết kế bản thân công cụ đó và không thể kiểm soát được.

Các bộ phận cơ khí trong một qui trình sản xuất phải được chế tạo với độ biến thiên tối thiểu nhằm giảm bớt các sản phẩm không đúng kích cỡ và vì vậy là các sản phẩm mắc lỗi. Và, nhìn chung, việc duy trì một phương sai tối thiểu trong các thước đo về những đặc tính chất lượng của một sản phẩm công nghiệp nhằm đạt được sự kiểm soát qui trình và vì vậy giảm thiểu tỷ lệ phần trăm của sản phẩm chất lượng kém là điều đáng mong ước.

Phương sai của mẫu

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{n-1}$$

là một số ước lượng không bị lệch của phương sai tổng thể σ^2. Vì vậy phân phối của các phương sai mẫu được tạo ra bằng việc chọn mẫu lặp lại sẽ có một phân phối xác suất mà bắt đầu tại $s^2 = 0$ (bởi vì s^2 không thể là số âm) với một trung bình bằng với σ^2. Không giống như phân phối của $\bar{x}$, phân phối của s^2 là không đối xứng, hình dạng chính xác phụ thuộc vào phân phối xác suất của tổng thể.

Đối với phương pháp sau đây, chúng ta sẽ giả định rằng mẫu được rút ra từ một tổng thể chuẩn và rằng s^2 được căn cứ trên một mẫu ngẫu nhiên gồm *n* thước đo và sở hữu $(n-1)$ bậc tự do. Bước tiếp theo và rất rõ ràng là xem xét phân phối của s^2 trong việc chọn mẫu lặp lại từ một phân phối chuẩn đã được xác định cụ thể - một phân phối với trung bình và phương sai cụ thể - và lập bảng cho giá trị tới hạn của s^2 đối với một số các diện tích một phía được sử dụng phổ biến. Nếu điều này được thực hiện, thì chúng ta sẽ tìm thấy rằng phân phối mẫu của s^2 là độc lập với trung bình tổng thể μ nhưng sở hữu một phân phối khác cho từng cỡ mẫu và từng giá trị của σ^2. Nhiệm vụ này là rất mất công, nhưng may mắn là có thể được đơn giản hóa bằng *việc chuẩn hóa*, như đã được thực hiện bằng cách sử dụng *z* trong các bảng chuẩn.

Đại lượng

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

được các nhà thống kê gọi là **biến số chi-bình-phương** (χ là mẫu tự Hy Lạp chi), phù hợp một cách đáng ngưỡng mộ cho các mục đích của chúng ta. Phân phối của đại lượng này trong việc chọn mẫu lặp lại được gọi là, như chúng ta có lẽ ngờ là vậy, một **phân phối xác suất chi-bình-phương.** Phương trình của hàm mật độ cho phân phối chi-bình-phương là rất nổi tiếng với các nhà thống kê mà đã lập bảng cho các giá trị tới hạn tương ứng với các diện tích một phía khác nhau của phân phối này. Những giá trị này được trình bày trong Bảng 5 của Phụ lục II.

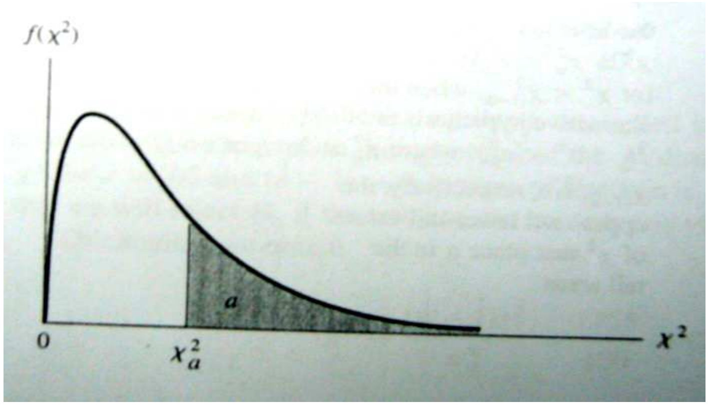
Hình dạng của phân phối chi-bình-phương, cũng giống như hình dạng của phân phối *t*, sẽ thay đổi theo cỡ mẫu hay, tương tự như vậy, theo các bậc tự do đi cùng với s^2. Vì thế Bảng 5 trong Phụ lục II được lập nên chính xác theo cùng cách thức với bảng *t*, với bậc tự do được thể hiện trong các cột đầu tiên và sau cùng. Một sự chép lại một phần Bảng 5 trong Phụ lục II được trình bày trong Bảng 8.7. Ký hiệu χ^2_a cho thấy rằng giá trị được lập bảng của χ^2 là để cho một diện tích *a* nằm về phía bên phải của nó. (Xem Hình 8.15). Diễn đạt theo thuật ngữ xác suất,

$$P(\chi^2 > \chi^2_a) = a$$

BẢNG **8.7** Định dạng của bảng chi-bình-phương, Bảng 5 trong Phụ lục II

d.f.	$\chi^2_{0.995}$	...	$\chi^2_{0.950}$	$\chi^2_{0.900}$	$\chi^2_{0.100}$	$\chi^2_{0.050}$	...	$\chi^2_{0.005}$	d.f.
1	0.0000393		0.0039321	0.015708	2.70554	3,83146		7.87944	1
2	0.0100251		0.102587	0.210720	4.60517	5.99147		10.5966	2
3	0.0717212		0.351846	0.584375	6.25139	7.81473		12.8381	3
4	0.206990		0.710721	1.063623	7.77944	9.48773		14.8602	4
5	0.411740		$\boxed{1.145476}$	1.61031	9.23635	11.0705		16.7496	5
6	0.675727		1.63539	2.20413	10.6446	12.5916		18.5476	6
...	...	...	...	...	...	...	...	...	...
15	4.60094		7.26094	8.54675	22.3072	$\boxed{24.9958}$		21.8013	15
16	5.14224		7.96164	9.31223	23.5418	26.2962		34.2672	16
17	5.69724		8.67176	10.0852	24.7690	27.5871		35.7185	17
18	6.26481		9.39046	10.8649	25.9894	28.8693		37.1564	18
19	6.84398		10.1170	11.6509	27.2036	30.1435		38.5822	19
...	...	...	...	...	...	...	...	...	...

HÌNH **8.15** Một phân phối chi-bình-phương



Như vậy 99% của diện tích nằm dưới phân phối χ^2 nằm về phía bên phải của $\chi^2_{0.99}$. Lưu ý rằng các giá trị quá mức của χ^2 phải được lập bảng cho cả hai phía cao và thấp của phân phối này bởi vì phân phối này là không đối xứng.

Bạn có thể kiểm tra khả năng của mình trong việc sử dụng bảng này bằng cách kiểm nghiệm các phát biểu sau đây. Xác suất để cho χ^2, căn cứ vào *n* = 16 đại lượng (d.f. = 15) sẽ vượt quá 24.9958 là 0.05. Đối với một mẫu gồm *n* = 6 đại lượng (d.f. = 5), 95% của diện tích nằm bên dưới phân phối χ^2 sẽ nằm về phía bên phải của $\chi^2 = 1.145476$. Những giá trị này của χ^2 được đóng khung trong Bảng 8.7.

Kiểm định thống kê về một giả thuyết không có liên quan đến một phương sai tổng thể,

$$H_0 : \sigma^2 = \sigma_0^2$$

sẽ sử dụng trị thống kê kiểm định

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Nếu σ^2 thật sự lớn hơn giá trị được giả định σ_0^2, thì trị thống kê kiểm định này sẽ có xu hướng là lớn và nhiều khả năng rơi về phía cao của phân phối này. Nếu $\sigma^2 < \sigma_0^2$, thì trị thống kê kiểm định này sẽ có xu hướng là nhỏ và nhiều khả năng rơi vào phía thấp của phân phối χ^2. Cũng giống như trong các kiểm định thống kê khác, chúng ta có thể sử dụng hoặc một kiểm định thống kê một hay hai phía, tùy thuộc vào giả thuyết thay thế mà ta lựa chọn. Qui trình cho kiểm định chi-bình-phương được thể hiện trong phần trình bày sau đây.

Kiểm định Giả thuyết có Liên quan đến một Phương sai Tổng thể

1. Giả thuyết không: $H_0 : \sigma^2 = \sigma_0^2$

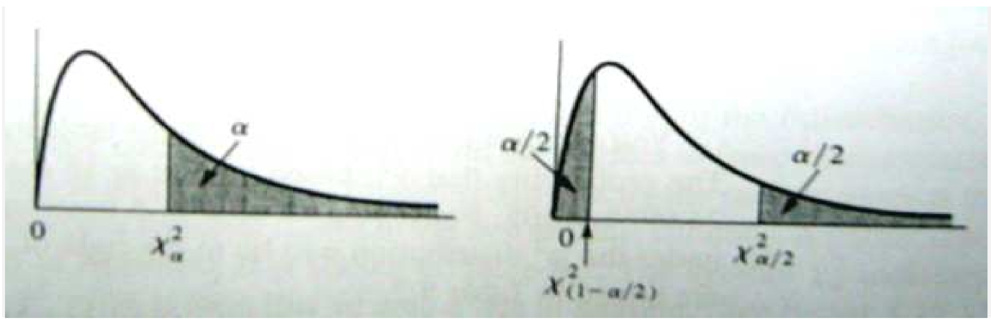
2. Giả thuyết thay thế:

Kiểm định Một phía	**Kiểm định Hai phía**
$H_a : \sigma^2 > \sigma_0^2$	$H_a : \sigma^2 \neq \sigma_0^2$
[hay $H_a : \sigma^2 < \sigma_0^2$]	

3. Trị thống kê kiểm định: $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$, với $\hat{p} = \dfrac{x}{n}$

4. Vùng bác bỏ:

Kiểm định Một phía	**Kiểm định Hai phía**
$\chi^2 > \chi^2_\alpha$ (hay $\chi^2 < \chi^2_{(1-\alpha)}$ khi giả thuyết thay thế là $H_a : \sigma^2 > \sigma_0^2$), trong đó χ^2_α và $\chi^2_{(1-\alpha)}$ lần lượt là các giá trị cao và thấp của χ^2 mà đặt α vào các diện tích một phía.	$\chi^2 > \chi^2_\alpha$ hoặc $\chi^2 < \chi^2_{(1-\alpha)}$, trong đó χ^2_α và $\chi^2_{(1-\alpha)}$ lần lượt là các giá trị cao và thấp của χ^2 mà đặt $\alpha/2$ vào các diện tích một phía.



Các giá trị tới hạn của χ^2 được dựa trên (n - 1) bậc tự do. Những giá trị được lập bảng này được thể hiện trong Bảng 5 của Phụ lục II.

Các giả thiết: Mẫu này được chọn lựa ngẫu nhiên từ một tổng thể chuẩn.

VÍ DỤ **8.17** Một nhà sản xuất xi măng đã xác nhận rằng bê tông được chuẩn bị từ sản phẩm của ông ta ắt sở hữu một sức chịu nén khá ổn định và rằng các đại lượng sức bền này ắt bằng với 40 kg trên mỗi xăngtimét bình phương. Một mẫu gồm n = 10 thước đo tạo ra một số trung bình và phương sai lần lượt bằng với

$$\bar{x} = 312 \quad \text{và} \quad s^2 = 195$$

Liệu những dữ liệu này có cung cấp đủ bằng chứng để bác bỏ sự xác nhận của nhà sản xuất này không?

Lời giải Như đã nói rõ, nhà sản xuất này xác nhận rằng dãy các đại lượng sức bền ắt bằng với 40 kg mỗi xăngtimét bình phương. Chúng ta sẽ giả định rằng ý ông ta muốn nói rằng các thước đo này sẽ nằm bên trong dãy này trong 95% thời gian và, vì vậy, rằng dãy này ắt xấp xỉ bằng với 4σ và rằng $\sigma = 10$. Chúng ta sau đó muốn kiểm định giả thuyết không

$$H_0 : \sigma^2 = (10)^2 = 100$$

so với giả thuyết thay thế

$$H_a : \sigma^2 > 100$$

Giả thuyết thay thế này đòi hỏi một sự kiểm định thống kê một phía, với toàn bộ vùng bác bỏ nằm ở phía cao của phân phối χ^2. Giá trị tới hạn của χ^2 đối với $\alpha = 0.05$ và

$(n-1)=9$ bậc tự do là $\chi^2 = 16.9190,$ mà hàm ý rằng chúng ta sẽ bác bỏ H_0 nếu trị thống kê kiểm định vượt quá giá trị này.

Tính toán, chúng ta có được

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{1755}{100} = 17.55$$

Bởi vì giá trị này của trị thống kê kiểm định rơi vào vùng bác bỏ, cho nên chúng ta kết luận rằng giả thuyết không là sai và rằng dãy của các đại lượng độ bền bêtông sẽ vượt quá điều mà nhà sản xuất xác nhận.

VÍ DỤ **8.18** Tìm mức ý nghĩa xấp xỉ quan sát được cho kiểm định trong Ví dụ 8.17.

Lời giải Xem xét hàng tương ứng với 9 bậc tự do trong Bảng 5 của Phụ lục II, bạn sẽ thấy rằng giá trị quan sát được của chi-bình-phương, $\chi^2 = 17.55,$ là lớn hơn giá trị được lập bảng, $\chi^2_{0.05} = 16.9190,$ và ít hơn $\chi^2_{0.025} = 19.0228.$ Vì vậy, mức ý nghĩa quan sát được (giá trị *p*) cho kiểm định nàhy nằm giữa 0.025 và 0.05. Chúng ta ắt báo cáo mức ý nghĩa quan sát được đối với kiểm định này là 0.025 < giá trị *p* < 0.05. Điều này nói với chúng ta rằng chúng ta ắt bác bỏ giả thuyết không đối với bất cứ giá trị nào của α bằng với 0.05 hay lớn hơn.

Một khoảng tin cậy cho σ^2 với hệ số tin cậy $(1-\alpha)100\%$ được thể hiện trong phần trình này sau đây:

Một Khoảng Tin cậy (1 - α)100% đối với σ^2

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

trong đó $\chi^2_{\alpha/2}$ và $\chi^2_{(1-\alpha/2)}$ lần lượt là các giá trị cao và thấp của χ^2 mà ắt nằm ở mức một nửa của α trong từng phía của phân phối chi-bình-phương.

Giả định: Mẫu này được chọn ngẫu nhiên từ một tổng thể chuẩn.

VÍ DỤ **8.19** Tìm khoảng tin cậy 90% cho σ^2 trong Ví dụ 8.17.

Lời giải Các giá trị được lập bảng của $\chi^2_{0.95}$ và $\chi^2_{0.05}$ tương ứng với (*n* - 1) bậc tự do là

$$\chi^2_{(1-\alpha/2)} = \chi^2_{0.95} = 3.32511 \quad \text{và} \quad \chi^2_{\alpha/2} = \chi^2_{0.05} = 16.9190$$

Thay thế các giá trị này và $s^2 = 195$ vào công thức tính khoảng tin cậy,

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

cho ta ước lượng khoảng cho σ^2.

$$\frac{9(195)}{16.9190} < \sigma^2 < \frac{9(195)}{3.32511} \quad \text{hay} \quad 103.73 < \sigma^2 < 527.80$$

So sánh Hai Phương sai của Tổng thể

Nhu cầu về các phương pháp thống kê để so sánh hai phương sai tổng thể là hết sức rõ ràng và dễ hiểu từ sự thảo luận này. Chúng ta có lẽ thường xuyên mong ước so sánh sự chính xác của một thiết bị đo lường với độ chính xác của một thiết bị khác, độ ổn định của một qui trình sản xuất này với tính chính xác của qui trình sản xuất kia, hay thậm chí sự biến thiên trong quá trình xếp hạng của một giáo sư đại học này với xếp hạng của giáo sư nọ.

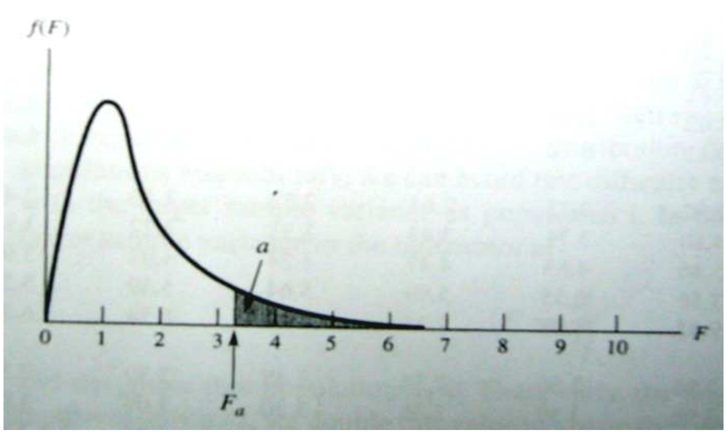
Về mặt trực giác, thì chúng ta có thể so sánh hai phương sai tổng thể, σ_1^2 và σ_2^2 bằng cách sử dụng tỷ lệ của các phương sai mẫu s_1^2/s_2^2. Nếu s_1^2/s_2^2 là gần bằng với 1, chúng ta ắt tìm được ít bằng chứng để cho thấy rằng σ_1^2 và σ_2^2 là không bằng nhau. Mặt khác, một giá trị rất lớn hay rất nhỏ của s_1^2/s_2^2 ắt cung cấp bằng chứng về một sự khác biệt trong các phương sai của tổng thể.

s_1^2/s_2^2 phải lớn hay nhỏ như thế nào để cho có đủ bằng chứng nhằm bác bỏ giả thuyết không $H_0 : \sigma_1^2 = \sigma_2^2$? Câu trả lời cho câu hỏi này có thể đạt được bằng cách nghiên cứu phân phối mẫu của s_1^2/s_2^2.

Khi các mẫu ngẫu nhiên độc lập được rút ra từ hai tổng thể chuẩn với các phương sai bằng nhau - nghĩa là, $\sigma_1^2 = \sigma_2^2$ – thì s_1^2/s_2^2 sở hữu một phân phối mẫu mà được các nhà thống kê biết đến như là **phân phối *F*.** Chúng ta không cần tự mình quan tâm đến phương trình của phân phối xác suất của *F*. Phân phối này là nổi tiếng, và các giá trị tới hạn đã được lập bảng. Những giá trị này xuất hiện trong Bảng 6 của Phụ lục II.

Hình dạng của phân phối *F* là không đối xứng và sẽ tùy thuộc vào số lượng bậc tự do đi cùng với tử số và mẫu số của $F = s_1^2/s_2^2$. Chúng ta sẽ đại diện các đại lượng này lần lượt là $v_1 = n_1 - 1$ và $v_2 = n_2 - 1$. Sự kiện này làm phức tạp thêm việc lập bảng của các giá trị tới hạn cho phân phối *F* và đòi hỏi phải xây dựng nên một bảng nhằm điều chỉnh các giá trị khác nhau của v_1, v_2 và *a*. (Xem Bảng 8.16).

HÌNH **8.16** Một phân phối *F* với $v_1 = 10$ và $v_2 = 10$



Trong Bảng 6 của Phụ lục II, các giá trị tới hạn của *F* đối với những diện tích nhằm phía bên phải tương ứng với *a* = 0.10, 0.05, 0.025, 0.010, và 0.005 được lập bảng cho những sự kết hợp khác nhau của bậc tự do tử số v_1 và bậc tự do mẫu số v_2. Một phần của Bảng 6 được chép lại trong Bảng 8.8. Các bậc tự do tử số v_1 được liệt kê theo hàng ngang ở trên cùng, và bậc tự do mẫu số v_2 được liệt kê theo chiều dọc ở phía ngoài cùng bên trái. Các giá trị của *a* được liệt kê trong cột thứ hai từ bên trái cũng như cột thứ hai từ bên phải. Đối với một sự kết hợp cố định của v_1 và v_2, thì các giá trị tới hạn thích hợp của *F* được tìm thấy trong dòng được chú dẫn bởi giá trị của *a* được yêu cầu.

Tham khảo Bảng 8.8, chúng ta lưu ý rằng $F_{0.05}$ cho các cỡ mẫu n_1 = 7 và n_2 = 10 (nghĩa là $v_1 = 6, v_2 = 9$) là 3.37. Tương tự, giá trị tới hạn $F_{0.05}$ cho các cỡ mẫu n_1 = 9 và n_2 = 12 (nghĩa là $v_1 = 8, v_2 = 11$) là 2.95. Những giá trị này của *F* được đóng khung trong Hình 8.8.

Theo cách thức tương tự, các giá trị tới hạn cho một diện tích một phía, *a* = 0.01, được trình bày trong Bảng 6 của Phụ lục II. Vì thế, nếu $v_1 = 6$ và $v_2 = 9,$ thì

$$P(F > F_0) = P(F > 5.80) = 0.01$$

Kiểm định thống kê cho giả thuyết không

$$H_0 : \sigma_1^2 = \sigma_2^2$$

sử dụng trị thống kê kiểm định

$$F = \frac{s_1^2}{s_2^2}$$

Khi giả thuyết thay thế này hàm ý một kiểm định một phía, nghĩa là,

$$H_a : \sigma_1^2 > \sigma_2^2$$

thì chúng ta có thể sử dụng trực tiếp các bảng này. Tuy nhiên, khi giả thuyết thay thế đòi hỏi một kiểm định hai phía, thì

$$H_a : \sigma_1^2 \neq \sigma_2^2$$

chúng ta lưu ý rằng vùng bác bỏ sẽ được phân thành các phía cao và thấp của phân phối *F* và rằng các bảng giá trị tới hạn cho phía thấp là thiếu một cách rõ rệt. Lý do cho sự vắng mặt của chúng được giải thích như sau: Chúng ta có quyền tự do khi xác định một trong hai tổng thể làm tổng thể I. Nếu tổng thể với phương sai mẫu lớn hơn được chỉ định làm tổng thể II, thì $s_2^2 > s_1^2$ và chúng ta sẽ quan tâm đến việc bác bỏ phía thấp của phân phối *F*. Bởi vì việc xác định các tổng thể là tùy ý, cho nên chúng ta có thể tránh được khó khăn này bằng cách chỉ định tổng thể có phương sai mẫu lớn hơn làm tổng thể I. Nói cách khác, luôn luôn đặt phương sai mẫu lớn hơn làm tử số trong công thức của

$$F = \frac{s_1^2}{s_2^2}$$

và chỉ định tổng thể đó là I. Sau đó, bởi vì diện tích ở phía bên phải sẽ chỉ tượng trưng cho $\alpha/2$, nên chúng ta nhân đôi giá trị này để có được giá trị chính xác cho xác suất của một sai lầm loại I α.

BẢNG **8.8** Định dạng của bảng F trong Phụ lục II

		ν_1								
ν_2	a	1	2	3	4	5	6	7	8	9
1	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
	.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3
	.010	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
	.005	16211	20000	21615	22500	23056	23437	23715	23925	24091
2	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.005	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4
3	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.005	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
	⋮	⋮			⋮			⋮		⋮
9	.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.050	5.12	4.26	3.86	3.63	3.48	$\boxed{3.37}$	3.29	3.23	3.18
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	.010	10.56	8.02	6.99	6.42	6.06	$\boxed{5.80}$	5.61	5.47	5.35
	.005	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
10	.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	$\boxed{2.95}$	2.90
	.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
	.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
	.005	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.005	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20

Kiểm định một Giả thuyết liên quan đến sự Bằng nhau của Hai Phương sai Mẫu

1. Giả thuyết không: $H_0 : \sigma_1^2 = \sigma_2^2$

2. Giả thuyết thay thế:

Kiểm định Một phía	**Kiểm định Hai phía**
$H_a : \sigma_1^2 > \sigma_2^2$	$H_a : \sigma_1^2 \neq \sigma_2^2$
[hay $H_a : \sigma_2^2 > \sigma_1^2$]	

3. Trị thống kê kiểm định:

Kiểm định Một phía	**Kiểm định Hai phía**

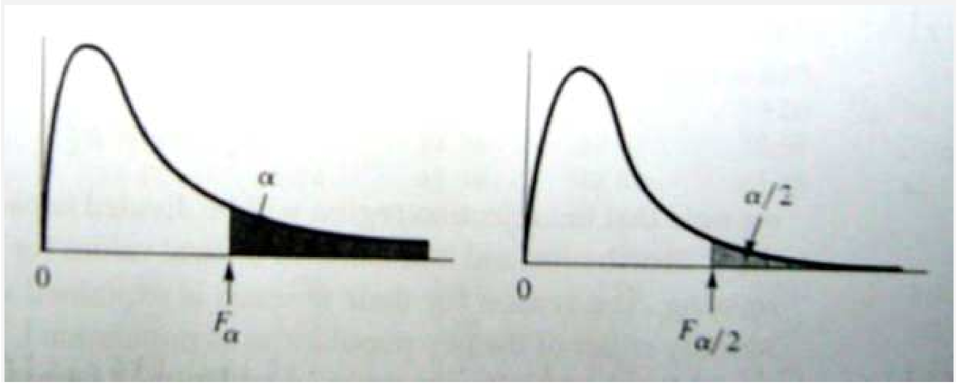
$F = \frac{s_1^2}{s_2^2}$	$F = \frac{s_1^2}{s_2^2}$

$\left(\text{hay } F = \frac{s_2^2}{s_1^2} \text{ cho } H_a : \sigma_2^2 > \sigma_1^2\right)$ trong đó s_1^2 là phương sai mẫu lớn hơn

4. Vùng bác bỏ:

Kiểm định Một phía	**Kiểm định Hai phía**
$F > F_\alpha$	$F > F_{\alpha/2}$

Khi $F = s_1^2 / s_2^2$, thì các giá trị tới hạn của F, F_α và $F_{\alpha/2}$ được căn cứ trên $v_1 = n_1 - 1$ và $v_2 = n_2 - 2$ d.f. Những giá trị được lập bảng này đối với α = 0.10, 0.05, 0.025, 0.01 và 0.005, có thể được tìm thấy trong Bảng 6 của Phụ lục II.



Các giả thiết: Các mẫu này được chọn lựa ngẫu nhiên và độc lập từ những tổng thể được phân phối chuẩn.

VÍ DỤ **8.20** Hai mẫu gồm 10 và 8 đại lượng được quan sát để sở hữu các phương sai mẫu lần lượt là $s_1^2 = 7.14$ và $s_2^2 = 3.21$. Liệu các phương sai mẫu này có đại diện đủ bằng chứng để cho thấy rằng những phương sai tổng thể là không bằng nhau?

Lời giải Giả định rằng các tổng thể này sở hữu các phân phối xác suất mà có hình dạng gò một cách hợp lý và vì vậy, cho tất cả các mục đích thực tiễn, sẽ thỏa mãn giả định rằng các tổng thể này là chuẩn.

Chúng ta muốn kiểm định giả thuyết không

$$H_0 : \sigma_1^2 = \sigma_2^2$$

so với giả thuyết thay thế

$$H_a : \sigma_1^2 \neq \sigma_2^2$$

Sử dụng Bảng 6 trong Phụ lục II và nhân đôi diện tích một phía này, chúng ta sẽ bác bỏ H_0 khi F > 3.68 với α = 0.10.

Giá trị tính toán được của trị thống kê kiểm định này là

$$F = \frac{s_1^2}{s_2^2} = \frac{7.14}{3.21} = 2.22$$

Lưu ý rằng trị thống kê kiểm định này không rơi vào vùng bác bỏ, cho nên chúng ta không bác bỏ $H_0: \sigma_1^2 = \sigma_2^2$. Vì thế, không có đủ bằng chứng cho thấy một sự khác biệt trong các phương sai mẫu.

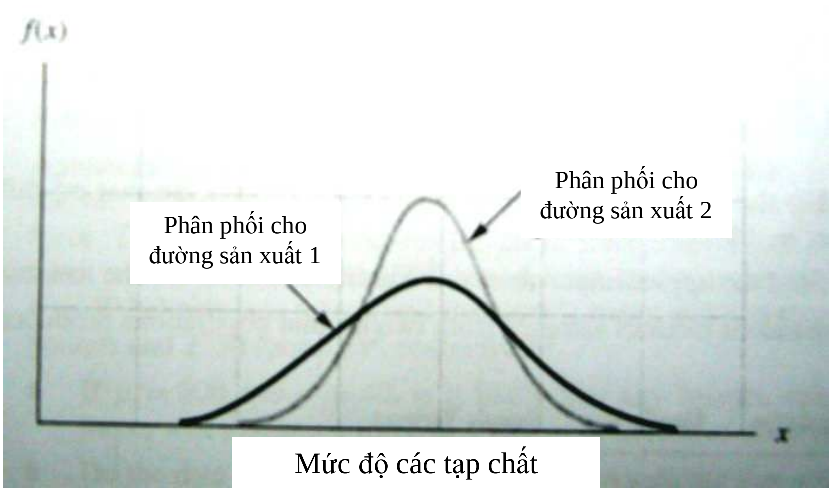
VÍ DỤ **8.21** Độ biến thiên trong khối lượng các tạp chất có trong một mẻ hóa chất đượ sử dụng cho một qui trình cụ thể tùy thuộc vào độ dài thời gian mà qui trình đó hoạt động. Một nhà sản xuất sử dụng hai dây chuyền sản xuất, 1 và 2, đã tạo ra một sự điều chỉnh nhẹ đối với qui trình 2, qua đó hy vọng giảm bớt sự biến thiên cũng như khối lượng trung bình các tạp chất trong hóa chất đó. Các mẫu gồm $n_1 = 25$ và $n_2 = 25$ đại lượng từ hai mẻ hóa chất tạo ra các số trung bình và phương sai như sau:

$$\bar{x}_1 = 3.2 \qquad s_1^2 = 1.04$$
$$\bar{x}_2 = 3.0 \qquad s_2^2 = 0.51$$

Liệu dữ liệu này có đại diện đủ bằng chứng để chỉ ra rằng sự biến thiên qui trình là ít hơn trong qui trình 2 không? Hãy kiểm định giả thuyết không $H_0 = \sigma_1^2 = \sigma_2^2$.

Lời giải Các ý nghĩa thực tiễn của ví dụ này được minh họa trong Hình 8.17. Chúng ta tin rằng các mức trung bình của các tạp chất trong hai dây chuyền sản xuất này là gần bằng nhau (trên thực tế, chúng có lẽ là bằng nhau) nhưng có một xác suất để cho sự thay đổi trong mức độ các tạp chất là nhỏ hơn nhiều trong dây chuyền 2. Sau đó các phân phối của những thước đo về tạp chất đối với hai dây chuyền sản xuất này ắt gần như có cùng mức trung bình, nhưng chúng ắt sẽ khác nhau trong mức độ biến thiên. Một phương sai lớn hơn của mức độ các tạp chất làm gia tăng xác suất của việc sản xuất ra các lô hàng hóa chất với mức độ các tạp chất cao không thể chấp nhận được. Do vậy, chúng ta hy vọng chứng minh rằng sự thay đổi qui trình này trong dây chuyền 2 đã làm cho σ_2^2 nhỏ hơn σ_1^2.

HÌNH **8.17** Các phân phối về những thước đo các tạp chất cho hai dây chuyền sản xuất trong Ví dụ 8.21



Kiểm định giả thuyết không

$$H_0 : \sigma_1^2 = \sigma_2^2$$

so với giả thuyết thay thế

$$H_a : \sigma_1^2 > \sigma_2^2$$

tại một mức ý nghĩa α = 0.05, chúng ta sẽ bác bỏ H_0 khi F là lớn hơn $F_{0.05} = 1.98$; nghĩa là, chúng ta sẽ áp dụng một kiểm định thống kê một phía.

Giá trị tính toán được của trị thống kê kiểm định,

$$F = \frac{s_1^2}{s_2^2} = \frac{1.04}{0.51} = 2.04$$

rơi vào vùng bác bỏ. Vì vậy, chúng ta kết luận rằng độ biến thiên của qui trình 2 là ít hơn sự biến thiên của qui trình 1.

Bài tập

Các Kỹ thuật Cơ bản

8.63 Một mẫu ngẫu nhiên gồm n = 25 quan sát từ một tổng thể chuẩn tạo ra một phương sai mẫu bằng với 21.4. Liệu dữ liệu này có cung cấp bằng chứng đủ để chỉ ra rằng $\sigma^2 > 15$ không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

8.64 Một mẫu ngẫu nhiên gồm n = 15 quan sát được chọn từ một tổng thể chuẩn. Trung bình và phương sai của mẫu là $\bar{x} = 3.91$ và $s^2 = 0.3214$. Tìm khoảng tin cậy 90% cho phương sai của tổng thể σ^2.

8.65 Một mẫu ngẫu nhiên gồm n = 22 quan sát được chọn từ một tổng thể chuẩn. Trung bình và phương sai của mẫu là $\bar{x} = 41.3$ và $s^2 = 14.14$. Liệu dữ liệu này có cung cấp bằng chứng đủ để chỉ ra rằng $\sigma^2 < 25$ không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

8.66 Tìm khoảng tin cậy 90% cho phương sai của tổng thể trong Bài tập 8.65.

8.67 Các mẫu ngẫu nhiên độc lập từ hai tổng thể chuẩn tạo ra các phương sai sau đây:

Cỡ Mẫu	Phương sai của Mẫu
16	55.7
20	31.4

a. Liệu dữ liệu này có cung cấp đủ bằng chứng để cho thấy rằng σ_1^2 khác với σ_2^2 không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

b. Tìm mức ý nghĩa xấp xỉ quan sát được cho kiểm định này, và giải thích giá trị của nó.

8.68 Các mẫu ngẫu nhiên độc lập từ hai tổng thể chuẩn tạo ra các phương sai sau đây:

Tổng thể	Cỡ Mẫu	Phương sai của Mẫu
1	13	18.3

2	13	7.9

a. Liệu dữ liệu này có cung cấp đủ bằng chứng để cho thấy rằng $\sigma_1^2 > \sigma_2^2$ không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

b. Tìm mức ý nghĩa xấp xỉ quan sát được cho kiểm định này, và giải thích giá trị của nó.

Các Ứng dụng

8.69 Tính ổn định của các đại lượng về những đặc tính của một sản phẩm sản xuất là quan trọng trong việc duy trì chất lượng sản phẩm. Trên thực tế, đôi lúc việc sở hữu những sự biến đổi nhỏ trong giá trị đo lường được của một đặc tính quan trọng nào đó của một sản phẩm và có số trung bình của qui trình chệch khỏi mục tiêu chút ít thì tốt hơn là phải chịu đựng sự biến thiên lớn với một giá trị trung bình phù hợp tuyệt vời với các yêu cầu. Tình huống thứ hai này có thể tạo ra một tỷ lệ phần trăm cao hơn của sản phẩm mắc lỗi so với tình huống đầu tiên. Một nhà sản xuất bóng đèn nghi ngờ rằng một trong các dây chuyền sản xuất của bà ta đang sản xuất ra các bóng với độ biến thiên cao về tuổi thọ sản phẩm. Để kiểm định giả thiết này, bà ta so sánh các tuổi thọ của *n* = 50 bóng đèn được chọn ngẫu nhiên từ dây chuyền bị ngi ngờ và *n* = 50 bóng từ một dây chuyền mà có vẻ "nằm trong tầm kiểm soát". Các số trung bình và phương sai cho hai mẫu này là như sau:

Dây chuyền bị nghi ngờ	Dây chuyền kiểm soát được
$\bar{x}_1 = 1520$	$\bar{x}_2 = 1476$
$s_1^2 = 92{,}000$	$s_2^2 = 37{,}000$

a. Liệu dữ liệu này có cung cấp đủ bằng chứng để cho thấy rằng các bóng đèn được sản xuất từ dây chuyền bị nghi ngờ sở hữu một phương sai về tuổi thọ sản phẩm lớn hơn các bóng đèn được làm ra từ dây chuyền mà được giả định là nằm trong tầm kiểm soát không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

b. Tìm mức ý nghĩa xấp xỉ quan sát được cho kiểm định này, và giải thích giá trị của nó.

8.70 Cơ quan Bảo vệ Môi trường (EPA) đã đặt ra một mức tiếng ồn tối đa cho xe tải hạng nặng là 83 đềxiben. Giới hạn này có thể được giải thích theo nhiều cách. Một cách để áp dụng giới hạn này ắt sẽ phải yêu cầu tất cả xe tải phải tuân thủ giới hạn tiếng ồn này. Phương pháp thứ hai ít thỏa mãn hơn ắt sẽ là phải yêu cầu mức độ tiếng ồn trung bình của một đoàn xe tải thấp hơn giới hạn này. Nếu phương pháp thứ hai là qui tắc, thì sự thay đổi trong mức độ tiếng ồn từ xe tải này sang xe tải khác ắt là quan trọng, bởi vì một giá trị lớn của σ^2 ắt hàm ý rằng nhiều xe tải đang vượt quá giới hạn này, ngay cả khi mức ồn trung bình của đoàn xe là 83 đềxiben. Dữ liệu cho sáu xe tải, tính bằng đềxibel là:

82.4, 83.8, 83.1, 82.3, 81.8, 83.0

Hãy sử dụng dữ liệu này để xây dựng nên khoảng tin cậy 95% cho σ^2, phương sai của các đại lượng về độ ồn của xe tải. Giải thích các kết quả của bạn.

8.71 Một công cụ chính xác được đảm bảo là có thể đọc một cách chính xác trong giới hạn 2 đơn vị. Một mẫu gồm bốn đại lượng dụng cụ đọc trên cùng một vật thể cho ra các đại lượng sau 353, 351, 351 và 355. Hãy kiểm định giả thuyết không rằng $\sigma = 0.7$ so với giả thuyết thay thế $\sigma > 0.7$. Tiến hành kiểm định này ở mức ý nghĩa $\alpha = 0.05$.

8.72 Tìm khoảng tin cậy 90% cho phương sai tổng thể trong Bài tập 8.71.

8.73 Một nhà sản xuất mũ bảo hộ cứng cho công nhân xây dựng quan tâm đến trung bình và phương sai của lực mà các mũ bảo hộ truyền sang người đội mũ khi chịu một lực chuẩn từ bên ngoài. Nhà sản xuất này mong muốn lực trung bình được truyền bởi mũ bảo hộ là 800 pao (hay ít hơn), thấp hơn khá nhiều so với giới hạn pháp lý là 1000 pao, và σ là ít hơn 40. Một mẫu ngẫu nhiên gồm n = 40 mũ bảo hộ được kiểm định, và trung bình và phương sai của mẫu được tìm thấy là lần lượt bằng 825 và 2350 pao.

a. Nếu $\mu = 800$ và $\sigma = 40$, thì liệu có khả năng xảy ra rằng bất cứ mũ bảo hộ nào, khi chịu một lực chuẩn từ bên ngoài, sẽ chuyển một lực sang người đội mũ vượt quá 1000 pao không? Hãy giải thích.

b. Liệu dữ liệu này có cung cấp bằng chứng đủ để chỉ ra rằng khi các mũ bảo hộ này chịu một lực tiêu chuẩn từ bên ngoài, thì lực trung bình được chuyển từ các chiếc mũ này vượt quá 800 pao?

c. Liệu dữ liệu này có cung cấp bằng chứng đủ để cho thấy rằng σ vượt quá 40 không?

8.74 Một nhà quản lý nhân sự hoạch định việc sử dụng kiểm định *t* Student để so sánh số lượng vắng mặt trung bình hàng tháng đối với hai loại nhân viên đã thông báo về một khó khăn có thể xảy ra. Sự biến thiên trong số lượng vắng mặt mỗi tháng tỏ ra khác nhau trong hai nhóm này. Để kiểm tra, nhà quản lý nhân sự này đã chọn ngẫu nhiên năm tháng và đếm số lượng vắng mặt cho từng nhóm. Dữ liệu được trình bày trong bảng sau đây.

Loại A	20	14	19	22	25
Loại B	37	29	51	40	26

a. Nhà quản lý nhân sự này quan tâm đến giả định nào là cần thiết cho việc sử dụng kiểm định *t*?

b. Liệu dữ liệu này có cung cấp bằng chứng đủ để cho thấy rằng những phương sai khác nhau cho các tổng thể của những người vắng mặt cho hai nhóm nhân viên này không? Hãy kiểm định với $\alpha = 0.10$, và giải thích các kết quả của kiểm định này.

8.75 Một nhà sản xuất dược phẩm mua một nguyên liệu cụ thể từ hai nhà cung cấp khác nhau. Mức trung bình của các tạp chất trong nguyên liệu này là xấp xỉ bằng nhau cho cả hai nhà cung cấp, nhưng nhà sản xuất này quan ngại về độ biến thiên của các tạp chất từ lô hàng này sang lô hàng khác. Nếu mức độ của các tạp chất này có xu hướng thay đổi một cách quá mức từ một nguồn cung ứng, thì điều này có thể ảnh hưởng đến chất lượng của sản phẩm dược phẩm. Để so sánh sự thay đổi về tỷ lệ phần trăm các tạp chất của hai nhà cung cấp này, nhà sản xuất nọ chọn ra mười lô hàng từ mỗi trong số hai nhà cung cấp này và đo tỷ lệ phàn trăm các tạp chất trong nguyên liệu cho từng lô hàng. Các số trung bình và phương sai của mẫu được cho trong bảng sau.

Nhà cung cấp A	**Nhà cung cấp B**
$\bar{x}_1 = 1.89$	$\bar{x}_2 = 1.85$
$s_1^2 = 0.273$	$s_2^2 = 0.094$
$n_1 = 10$	$n_2 = 10$

Liệu dữ liệu này có cung cấp bằng chứng đủ để cho thấy một sự khác biệt trong độ biến thiên của các mức tạp chất của lô hàng đối với hai nhà cung cấp này không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.10$. Dựa vào các kết quả của kiểm định này, bạn sẽ đưa ra các đề nghị nào cho nhà sản xuất dược phẩm đó?

8.10 MỘT SỐ BÌNH LUẬN VỀ LÝ THUYẾT CÁC KIỂM ĐỊNH GIẢ THUYẾT

Như đã được trình bày trong Phần 8.2, lý thuyết của một kiểm định thống kê về một giả định thật sự là một qui trình hết sức rõ ràng, qua đó cho phép người làm thí nghiệm hoặc bác bỏ hoặc chấp nhận giả thuyết không với các đại lượng rủi ro α và β. Thật không may là như chúng ta đã lưu ý, khung lý thuyết này không đáp ứng được cho tất cả các tình huống thực tiễn.

Điểm then chốt của lý thuyết đòi hỏi rằng chúng ta có khả năng xác định cụ thể một giả thuyết thay thế có nghĩa mà cho phép sự tính toán xác suất β cho một sai lầm loại II đối với tất cả các giá trị thay thế của (các) tham số này. Sự tính toán này có thể được thực hiện đối với nhiều kiểm định thống kê, bao gồm kiểm định cho mẫu lớn trong Phần 8.3, mặc dù việc tính toán β cho các kiểm định và cỡ mẫu khác có lẽ là điều khó khăn trong một số trường hợp. Mặt khác, trong một số tình huống kiểm định, việc xác định rõ ràng các giả thuyết thay thế cho H_0 mà có ý nghĩa thực tiễn là điều khó thực hiện. Điều này có thể xảy ra khi chúng ta muốn kiểm định một giả thuyết liên quan đến các giá trị của một tập hợp các tham số, một tình huống mà chúng ta sẽ gặp phải trong Chương 13 khi phân tích dữ liệu có tính liệt kê.

Trở ngại mà chúng ta đề cập này không làm mất giá trị việc sử dụng các kiểm định thống kê. Thay vào đó, trở ngại này thúc đẩy sự cẩn trọng trong việc rút ra các kết luận khi bằng chứng không đủ là sẵn có để bác bỏ giả thuyết không. Khó khăn trong việc xác định cụ thể các giả thuyết thay thế có ý nghĩa khác đối với giả thuyết không này, cùng với khó khăn gặp phải trong tính toán và lập bảng cho β cho các kiểm định khác ngoài những kiểm định thống kê đơn giản nhất, minh chứng cho việc nói vòng vo vấn đề này ở một đoạn giới thiệu. Vì thế chúng ta có thể áp dụng một trong hai qui trình. Chúng ta có thể trình bày giá trị *p* đi cùng với một kiểm định thống kê và để việc giải thích lại cho độc giả. Hoặc chúng ta có thể đồng ý áp dụng qui trình được mô tả trong Ví dụ 8.15 khi những giá trị được lập bảng của β (đường cong đặc trưng hoạt động) là không sẵn có cho kiểm định này. Khi trị thống kê kiểm định này rơi vào vùng bác bỏ, chúng ta sẽ "không bác bỏ" hơn là "chấp nhận" giả thuyết không. Những kết luận xa hơn có thể được thực hiện bằng cách tính toán một ước lượng khoảng cho tham số này hay qua việc tham khảo một trong nhiều sách giáo khoa thống kê đã xuất bản để biết các giá trị được lập bảng của β. Chúng ta sẽ không quá ngạc nhiên khi biết rằng những bảng lập ra này là không tiếp cận được, nếu không phải toàn không sẵn có, đối với một số trong những kiểm định thống kê phức tạp hơn.

Sự lựa chọn giữa một kiểm định một phía hay hai phía cho một tình huống đã biết được quyết định bởi các khía cạnh thực tiễn của vấn đề đó và sẽ tùy thuộc vào giá trị thay thế của tham số mà người làm thí nghiệm đang cố gắng xác minh. Nếu chúng ta phải chịu một sự tổn thất tài chính lớn nếu μ là lớn hơn μ_0 nhưng không phải chịu như vậy nếu nhỏ hơn, thì chúng ta ắt tập trung sự chú ý của mình vào việc xác minh các giá trị của μ lớn hơn μ_0. Vì thế, chúng ta ắt bác bỏ ở phía cao của phân phối này. Mặt khác, nếu chúng ta quan tâm ở mức độ ngang nhau trong việc xác minh các giá trị của μ mà hoặc nhỏ hơn hoặc lớn hơn μ_0, thì chúng ta ắt sử dụng một kiểm định hai phía.

8.11 CÁC GIẢ ĐỊNH

Như đã được lưu ý trước đây, các kiểm định và khoảng tin cậy dựa trên trị thống kê *t* Student, chi-bình-phương, và trị thống kê *F* đòi hỏi rằng dữ liệu phải thỏa mãn các giả định cụ thể để cho các xác suất sai số (đối với những kiểm định) và những hệ số tin cậy (cho các khoảng tin cậy) bằng với những giá trị mà chúng ta đã xác định cụ thể. Ví dụ, nếu các giả định bị vi phạm bởi việc lựa chọn một mẫu từ một tổng thể không phải nhị thức và dữ liệu này được sử dụng để lập nên khoảng tin cậy 95% cho μ, thì hệ số tin cậy thực sự có lẽ (mà chúng ta chưa biết) chỉ bằng với 0.85 thay vì là 0.95. Các giả định được tóm tắt trong phần trình bày sau đây để bạn thuận tiện theo dõi.

Các Giả định cho các trị thống kê *t*, χ^2, và *F*

1. Đối với các kiểm định và khoảng tin cậy cho mẫu nhỏ được mô tả trong chương này, chúng ta giả định rằng các mẫu này được chọn lựa ngẫu nhiên từ những tổng thể có phân phối chuẩn.
2. Khi hai mẫu được chọn, chúng ta giả định rằng các mẫu này được lựa chọn theo một cách thức độc lập, ngoại trừ trong trường hợp của một thí nghiệm có khác biệt cặp.
3. Đối với các kiểm định và khoảng tin cậy liên quan đến sự khác biệt giữa hai số trung bình của tổng thể μ_1 và μ_2, thì căn cứ vào các mẫu ngẫu nhiên độc lập, chúng ta giả định rằng $\sigma_1^2 = \sigma_2^2$.

Trong một tình huống chọn mẫu thực tiễn, bạn không bao giờ biết mọi thứ về phân phối xác suất của tổng thể được chọn mẫu. Nếu bạn biết, thì ắt đã không cần việc chọn mẫu hay các trị thống kê. Hơn nữa, rất ít có khả năng xảy ra rằng một tổng thể ắt sở hữu một cách chính xác các đặc trưng được mô tả ở trên. Vì thế, để cho hữu ích, thì các phương pháp suy luận được mô tả trong chương này phải dẫn đến các suy luận tốt về những sự lệch vừa phải khỏi các giả định đang hiện diện. Ví dụ, nếu tổng thể sở bữu một phân phối có hình dạng gò mà gần như chuẩn, thì chúng ta ắt muốn một khoảng tin cậy 95% được lập ra cho μ bằng với một với một hệ số tin cậy gần bằng 0.95. Tương tự như vậy, nếu chúng ta tiến hành một kiểm định *t* về giả thuyết không $\mu_1 = \mu_2$, dựa trên các mẫu ngẫu nhiên độc lập từ các tổng thể chuẩn, mà ở đó σ_1^2 và σ_2^2 không bằng đúng nhau, thì chúng ta muốn xác suất của việc bác bỏ một cách sai lầm giả thuyết không, α, xấp xỉ bằng với giá trị mà chúng ta đã sử dụng trong việc định vị vùng bác bỏ.

Một phương pháp thống kê mà không nhạy cảm với những sự lệch khỏi các giả định mà qua đó phương pháp này được căn cứ vào được cho là không *vững chắc*. Các kiểm định *t* là rất vững chắc đối với những sự lệch vừa phải so với tính chuẩn tắc. Mặt khác, các kiểm định chi-bình-phương và *F* là nhạy cảm với những sự lệch khỏi tính chuẩn. Kiểm định *t* cho việc so sánh hai số trung bình là có tính vững chắc vừa phải với những sự lệch khỏi giả định $\sigma_1^2 = \sigma_2^2$ khi $n_1 = n_2$. Tuy nhiên, kiểm định này trở nên nhạy cảm với những sự lệch khỏi giả định này khi n_1 trở nên lớn so với n_2 (hay ngược lại).

Nếu bạn quan ngại rằng dữ liệu của mình không thỏa mãn các giả định được qui định cho một trong các phương pháp thống kê được mô tả trong chương này, thì bạn có thể sử dụng một phương pháp thống kê không đối xứng để thực hiện sự suy luận. Những phương pháp này, mà đòi hỏi rất ít hay không có giả định nào về bản chất của các phân phối xác suất của tổng thể, là đặc biệt hữu dụng cho việc kiểm định các giả thiết, và một số phương pháp không tham số đã được phát triển cho việc ước lượng các tham số tổng thể. Những kiểm định các giả thiết có liên quan đến vị trí của một phân phối tổng thể hay một kiểm định cho sự tương đồng của hai phân phối tổng thển được trình bày trong Mendenhall, Beaver, và Beaver, *A Course in*

Business Statistics, xuất bản lần thứ 4, 1996, Duxbury Press. Nếu bạn có thể lựa chọn các mẫu tương đối lớn, thì bạn có thể sử dụng ước lượng cho mẫu lớn hay các qui trình kiểm định được trình bày trong chương này.

QUAY LẠI NGHIÊN CỨU ĐIỂN HÌNH

8.12 THÊM NỮA VỀ CÁC NHÀ THẦU CHO NGÀNH CÔNG NGHIỆP QUỐC PHÒNG THỜI HẬU CHIẾN TRANH LẠNH

Trong nghiên cứu của mình về các hệ thống hạch toán chi phí hiện đang được sử dụng bởi các nhà thầu cho ngành công nghiệp quốc phòng so với những nhà thầu cho ngành công nghiệp không phải quốc phòng, Rezaee và Elmore đã tóm tắt sự trả lời của 25 nhà thầu cho ngành công nghiệp quốc phòng và 25 nhà thầu cho ngành công nghiệp không phải quốc phòng. Thông tin trong Bảng 8.1 được lấy từ Bảng 2. Một Thang điểm Likert là thang điểm mà qua đó các trả lời được phân hạng theo sự đo lường về niềm tin vào sự xác nhận của người trả lời. Ví dụ, câu trả lời cho khẳng định "những sự kiểm soát quản lý tài chính truyền thống là quan trọng" có thể là như sau: 1, không đồng ý mạnh mẽ; 2, không đồng ý; 3, không có ý kiến; 4, đồng ý; hay 5, đồng ý mạnh mẽ. Bởi vì những câu trả lời này không được phân phối chuẩn, nên chúng ta cần khảo cứu các cỡ mẫu và phân phối mẫu của những trị thống kê mà ắt được sử dụng trong việc kiểm định những sự khác biệt đáng kể giữa các cặp số trung bình. Dữ liệu trong Bảng 8.1 được sao chép lại trong Bảng 8.9.

BẢNG **8.9**

	Quốc phòng	**Không phải quốc phòng**	**Giá trị *T***
Hoạch định chiến lược			
1 Ngân sách được kiểm tra về sự nhất quán với các mục tiêu dài hạn	4.0425	4.2000	-0.30
2 Xác nhận chính thức về các mục tiêu, chiến lược, v.v được sử dụng cho việc hoạch định phương hướng của công ty	4.1625	4.8800	-1.45
Ngân sách và Hoạch định			
1 Ngân sách được sử dụng trong việc đánh giá hiệu quả hoạt động của những thành viên riêng lẻ	3.1600	4.5200	-2.56*
2 So sánh giữa các chi phí thực tế so với chi phí được dự trù ngân sách	3.9891	5.1782	2.64*
3 Ngân sách của các phòng ban riêng lẻ	2.8800	4.6800	-3.238
4 Sự tham gia của quản lý cấp trung và thấp hơn trong việc dự trù ngân sách	3.6879	5.0800	-3.53*
5 Ngân sách linh hoạt	2.1861	3.6000	-2.64*

* Có ý nghĩa ở mức 0.01

Đối với mỗi trong số các loại được liệt kê trong bảng này, số trung bình được căn cứ vào một mẫu gồm 25 câu trả lời. Cỡ mẫu này là đủ lớn để đảm bảo sự chuẩn hóa xấp xỉ trong phân phối của $\bar{x}$ cho từng phân loại. Liệu chúng ta có thể giả định các phương sai bằng nhau cho hai tổng thể này được so sánh? Nếu chúng ta có thể, thì bậc tự do với s^2, ước lượng gộp của phương sai chung, sẽ là $n_1 + n_2 - 2 = 25 + 25 - 2 = 48$, mà chỉ ra rằng sự chuẩn chuẩn hóa hay các bảng giá trị tới hạn của *t* Student có thể được sử dụng trong việc xác định các giá trị tới hạn cho sự kiểm định.

Nếu chúng ta mong muốn khẳng định mức ý nghĩa 0.01, thì chúng ta sẽ tham khảo các giá trị tới hạn với một diện tích nằm ở phía bên phải của $\alpha/2 = 0.005$ để tìm ra $t_{0.005} = 2.576$ và $z_{0.005} = 2.58.$ Mặt khác, nếu chúng ta không thể giả định rằng các phương sai này là bằng nhau, thì các cỡ mẫu, mặc dù thuộc hai nhóm khác nhau, có thể được xem như đủ lớn để đảm bảo tính chuẩn xấp xỉ của trị thống kê kiểm định, và giá trị tới hạn một lần nữa ắt sẽ được lấy là $z_{0.005} = 2.58$.

Kiểm định cho từng cặp số trung bình sẽ tiến hành như sau. Các giả định cần được kiểm định là

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{so với} \quad H_a: \mu_1 - \mu_2 \neq 0$$

Trị thống kê kiểm định là

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{s^2\left(\frac{1}{25} + \frac{1}{25}\right)}}$$

Với d.f. = 48, và $\alpha = 0.05,$ giá trị tới hạn là $t_{0.005} = 2.576.$ Vì thế, chúng ta sẽ bác bỏ H_0, và kết luận rằng các số trung bình này là khác nhau có ý nghĩa nếu giá trị quan sát được của *t* là lớn hơn 2.576 hay nhỏ hơn - 2.576.

Nếu những tính toán về trị thống kê *t* trong bảy loại được cho trong Bảng 8.9 là chính xác, thì hai sự khác biệt đầu tiên là không có ý nghĩa, như được chỉ ra, và năm sự khác biệt tiếp sau là có ý nghĩa. Tuy nhiên, sự khác biệt tương ứng với loại 1 trong nhóm Ngân sách và Hoạch định là có ý nghĩa ở mức 5%, không phải mức 1% như được chỉ ra, các khác biệt còn lại có ý nghĩa ở mức 1%.

Tất cả những điều này có nghĩa gì? Chúng ta đã tìm ra rằng, mặc dù những khác biệt trung bình là không có ý nghĩa, thì cả các ngành công nghiệp quốc phòng lẫn không phải quốc phòng đều xem việc hoạch định chiến lược là rất quan trọng. Mặt khác, các nhà thầu cho ngành công nghiệp quốc phòng đánh giá tầm quan trọng ít hơn nhiều một cách đáng kể đối với các lĩnh vực liên quan đến ngân sách và hoạch định, qua đó chỉ ra rằng họ có một cơ cấu kiểm soát tổ chức tập trung hóa cao độ hơn so với các nhà thầu của các ngành công nghiệp không phải quốc phòng.

8.13 TÓM TẮT

Trong chương này, chúng ta đã trình bày các khái niệm cơ bản về một kiểm định thống kê cho một giả thuyết và đã chứng minh qui trình này cho các mẫu lớn và nhỏ. Một số các kiểm định được mô tả trong chương này là dựa vào Định lý Giới hạn Trung tâm và vì vậy áp dụng cho các mẫu lớn. Khi *n* là lớn, thì mỗi trong số các trị thống kê kiểm định có liên quan sở hữu một phân phối mẫu mà có thể được ước lượng xấp xỉ bởi phân phối chuẩn. Kết quả này, cùng với các đặc trưng của phân phối mẫu đã được nghiên cứu trong Chương 5, cho phép việc tính toán α, β, và các giá trị *p* cho những kiểm định thống kê.

Việc lưu ý rằng các trị thống kê t, χ^2, và *F* sử dụng trong các phương pháp thống kê cho mẫu nhỏ được căn cứ trên giả định rằng các tổng thể được chọn mẫu có một phân phối xác suất chuẩn là rất quan trọng. Yêu cầu này sẽ được đáp ứng một cách phù hợp cho nhiều loại hình của các đại lượng thí nghiệm khác nhau.

Bạn sẽ quan sát mối liên hệ rất chặt chẽ giữa các trị thống kê *t* Student và *z*, và vì vậy là sự tương đồng của các phương pháp cho việc kiểm định những giả thuyết và việc lập nên các khoảng tin cậy. Các trị thống kê χ^2 và F sử dụng trong việc tạo ra những suy luận có liên quan đến các phương sai tổng thể dĩ nhiên không đi theo kiểu hình này, nhưng lý do hợp lý được áp dụng trong việc xây dựng nên các kiểm định thống kê và khoảng tin cậy là giống nhau cho tất cả các phương pháp mà chúng ta đã trình bày.

Các gợi ý cho việc giải toán

Khi tiến hành một kiểm định thống kê cho một giả thuyết, việc đi theo cùng một qui trình cơ bản cho từng vấn đề là rất quan trọng.

1. **Xác định loại dữ liệu** có liên quan (định lượng hay nhị thức) và số lượng các mẫu có liên quan (một hay hai). Bạn quan tâm đến một số trung bình, một tỷ lệ, hay một phương sai? Điều này sẽ cho phép bạn xác định tham số quan tâm trong thí nghiệm.
2. **Kiểm tra các điều kiện** được yêu cầu cho phân phối mẫu của tham số phải được ước lượng xấp xỉ bởi một phân phối chuẩn. Đối với dữ liệu định lượng, thì cỡ (hay các cỡ) mẫu phải là 30 hay lớn hơn. Đối với dữ liệu nhị thức, thì một cỡ mẫu lớn sẽ đảm bảo rằng $p \pm 2\sigma_{\hat{p}}$ được chứa trong khoảng từ 0 đến 1 [$(p_1 - p_2) \pm 2\sigma_{(\hat{p}_1 - \hat{p}_2)}$ chứa trong khoảng từ - 1 đến 1 cho trường hợp hai mẫu].
3. **Phát biểu các giả thuyết không và thay thế (H_0 và H_a).** Giả thuyết thay thế là giả thuyết mà nhà nghiên cứu mong muốn ủng hộ; giả thuyết không là một sự ngược lại của giả thuyết thay thế.
4. **Phát biểu trị thống kê kiểm định** được sử dụng trong kiểm định về giả thuyết.
5. **Định vị vùng bác bỏ** cho kiểm định. Trong chương này, vùng bác bỏ sẽ được tìm thấy trong các diện tích một phía của sự phân phối chuẩn chuẩn hóa (*z*), phân phối *t*, phân phối χ^2 hay phân phối *F*. Vùng bác bỏ chính xác sẽ được xác định bởi giá trị mong muốn của α và loại giả thuyết thay thế (một phía hay hai phía).
6. **Tiến hành kiểm định,** qua việc tính toán giá trị quan sát được của trị thống kê kiểm định dựa vào dữ liệu mẫu.
7. **Rút ra các kết luận** căn cứ trên giá trị quan sát được của trị thống kê kiểm định. Nếu trị thống kê kiểm định rơi vào vùng bác bỏ, thì giả thuyết không sẽ bị bác bỏ và ủng hộ giả thuyết thay thế. Xác suất của một quyết định không chính xác là α. Tuy nhiên, nếu trị thống kê kiểm định không rơi vào vùng bác bỏ, thì chúng ta không thể bác bỏ giả thuyết không. Có không đủ bằng chứng để chứng tỏ rằng giả thuyết thay thế là đúng. Sự đánh giá được giữ chưa đưa ra cho đến khi thu thập được thêm nhiều dữ liệu.

Bài tập thêm

Bài tập thêm

8.76 Định nghĩa α và β cho một sự kiểm định thống kê của một giả định.

8.77 Mức ý nghĩa quan sát được của một sự kiểm định là gì?

8.78 Các điều kiện nào phải được đáp ứng để cho kiểm định *z* có thể được sử dụng để kiểm định một giả thiết liên quan đến một số trung bình tổng thể μ?

8.79 Các giả định nào được thực hiện khi sử dụng một kiểm định *t* Student để kiểm định một giả thuyết có liên quan đến một số trung bình tổng thể duy nhất?

8.80 Các giả định nào được thực hiện khi sử dụng một kiểm định *t* Student để kiểm định một giả thuyết có liên quan đến sự khác biệt giữa hai số trung bình tổng thể?

8.81 Tiền công hàng ngày trong một ngành cụ thể có phân phối chuẩn với một giá trị trung bình là \$23.20 và một độ lệch chuẩn là \$4.50. Một công ty trong ngành công nghiệp này thuê mướn 40 công nhân và trả cho những người này một mức tiền công bình quân là \$21.20 hàng ngày. Căn cứ vào số trung bình của mẫu này, liệu những công nhân này có được xem như là một mẫu ngẫu nhiên từ số tất cả công nhân trong ngành này không?

a. Tìm mức ý nghĩa quan sát được cho sự kiểm định này.

b. Nếu bạn có kế hoạch tiến hành kiểm định này bằng cách sử dụng $\alpha = 0.01$, thì kết luận về kiểm định của bạn ắt sẽ như thế nào?

8.82 Tỷ lệ lấp đầy chỗ cao của hãng hàng không trên các chuyến bay có lên kế hoạch là hết sức quan trọng đối với khả năng sinh lợi. Giả sử rằng một chuyến bay có lên kế hoạch phải đạt mức bình quân tối thiểu là lấp đầy 60% chỗ để sinh lợi và rằng một sự khảo cứu các tỷ lệ lấp đầy cho 120 chuyến bay khởi hành lúc 10h sáng từ Atlanta đến Dallas cho thấy một tỷ lệ lấp đầy trung bình mỗi chuyến bay là 58% và một độ lệch chuẩn là 11%.

a. Nếu μ là tỷ lệ lấp đầy chỗ trung bình mỗi chuyến bay và nếu công ty này mong muốn xác định liệu chuyến bay có lên kế hoạch này là không sinh lợi hay không, hãy tìm các giả thuyết không và thay thế cho kiểm định này.

b. Liệu giả thuyết thay thế trong câu (a) có hàm ý một kiểm định một phía hay hai phía không? Hãy giải thích.

c. Liệu dữ liệu về tỷ lệ lấp đầy chỗ cho 120 chuyến bay này có gợi ý rằng chuyến bay có lên kế hoạch này là không sinh lợi hay không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.10$.

8.83 Một nhà sản xuất các máy rửa xe hơi cung cấp một mẫu máy cụ thể với một trong ba màu A, B hay C. Trong số 1000 máy rửa xe đầu tiên được bán ra, thì 400 máy là có màu A. Liệu bạn ắt kết luận rằng có nhiều hơn 1/3 trong số tất cả khách hàng có một sự ưa thích hơn dành cho màu A không?

a. Tìm mức ý nghĩa quan sát được cho kiểm định này.

b. Nếu bạn có kế hoạch thực hiện kiểm định của mình qua việc sử dụng $\alpha = 0.05$, thì kết luận về kiểm định của bạn ắt là gì?

8.84 Một nhà sản xuất xác nhận rằng có ít nhất 20% dân chúng ưa thích sản phẩm của mình hơn. Một mẫu gồm 100 người được chọn để kiểm tra sự xác nhận này. Với $\alpha = 0.05$, thì tỷ lệ phần trăm của mẫu ắt phải nhỏ bao nhiêu trước khi xác nhận này có thể bị từ chối về mặt thống kê? (Lưu ý rằng việc này ắt đòi hỏi một kiểm định một phía cho giả định).

8.85 Tham khảo lại Bài tập 8.84. Mười sáu người trong mẫu gồm 100 khách hàng đã biểu lộ sự ưa thích hơn cho sản phẩm của nhà sản xuất này. Liệu kết quả này có trình bày đủ bằng chứng để bác bỏ sự xác nhận của nhà sản xuất này không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.10$.

8.86 Một nhà sản xuất có thể chịu đựng một lượng nhỏ (0.05 mg/l) các tạp chất trong một nguyên liệu cần thiết cho việc chế tạo sản phẩm của mình. Bởi vì sự kiểm nghiệm của phòng thí nghiệm về các tạp chất này phụ thuộc vào sai số thí nghiệm, nên nhà sản xuất này kiểm tra mỗi mẻ mười lần. Giả sử rằng giá trị trung bình của sai số thí nghiệm là 0 và vì vậy rằng giá trị trung bình của kết quả của mười lần kiểm tra này là một sự ước lượng không bị lệch của khối lượng thực tế của các tạp chất trong mẻ hàng đó. Đối với một mẻ hàng cụ thể của nguyên liệu đó, giá trị trung bình của kết quả mười lần kiểm tra này là 0.058 mg/l và độ lệch chuẩn là 0.012 mg/l. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng khối lượng tạp chất trong mẻ này vượt quá 0.05 mg/l không? Tìm giá trị xấp xỉ *p* cho kiểm định này, và giải thích giá trị của nó.

8.87 Nhiệt độ khi hoạt động của hai lò sấy sơn đi cùng với hai dây chuyền sản xuất được ghi nhận cho 20 ngày. (Sự kết cặp bị bỏ qua). Các số trung bình và phương sai của hai mẫu này là

$$\bar{x}_1 = 164 \qquad \bar{x}_2 = 168$$
$$s_1^2 = 81 \qquad s_2^2 = 172$$

Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra một sự khác biệt trong sự thay đổi nhiệt độ đối với hai lò này không? Hãy kiểm định giả thuyết rằng $\sigma_1^2 = \sigma_2^2$ ở mức ý nghĩa $\alpha = 0.10$.

8.88 Một nhà máy sản xuất có hai hệ thống chế tạo cực kỳ phức tạp, trong đó một hệ thống có tuổi đời gấp đôi so với hệ thống kia. Cả hai hệ thống này được kiểm tra, bôi trơn, và bảo trì hai tuần một lần. Số lượng các sản phẩm cuối cùng được sản xuất ra hàng ngày bởi mỗi trong số hai hệ thống này được ghi nhận cho 30 ngày làm việc. Các kết quả được cho trong bảng sau. Liệu dữ liệu này có cung cấp đủ bằng chứng để kết luận rằng rằng độ biến thiên trong các lệnh sản xuất hàng ngày đã làm tăng sự bảo trì cho hệ thống sản xuất có tuổi đời cao hơn không? Hãy sử dụng mức ý nghĩa 5%.

Hệ thống Mới	**Hệ thống Cũ**
$\bar{x}_1 = 246$	$\bar{x}_2 = 240$
$s_1 = 15.6$	$s_2 = 28.2$

8.89 Một nhà sản xuất xác nhận rằng có ít nhất 95% thiết bị mà họ cung cấp cho một nhà máy là phù hợp với các qui chuẩn. Một sự khảo cứu một mẫu gồm 700 mảnh rời của thiết bị đã bộc lộ cho thấy rằng có 53 mảnh là có lỗi. Hãy kiểm định sự xác nhận của nhà sản xuất này bằng cách sử dụng $\alpha = 0.05$.

8.90 Khi quyết định nơi đặt trọng tâm quảng cáo, một phòng nghiên cứu thị trường cho một nhà sản xuất xe hơi lớn đã mong muốn so sánh số lượng xe hơi bình quân của mỗi gia đình tại hai khu vực của Hoa Kỳ. Giả sử rằng một nghiên cứu sơ khởi về số lượng xe hơi của mỗi gia đình cho *n* = 200 gia đình từ mỗi trong số hai khu vực trên cho ta các số trung bình và phương sai cho hai mẫu này được thể hiện trong bảng đi kèm sau đây.

	Khu vực 1	**Khu vực 2**
Cỡ mẫu	200	200
Trung bình của mẫu	1.30	1.37
Phương sai của mẫu	0.53	0.64

a. Lưu ý rằng một sự gia tăng nhỏ trong số lượng xe hơi trung bình của mỗi gia đình có thể tượng trưng cho một con số rất lớn xe hơi cho một khu vực. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra một sự khác biệt trong số lượng xe hơi trung bình của mỗi gia đình cho hai khu vực này không?

b. Hãy hình dung dữ liệu với bất cứ tổng thể nào trong số hai tổng thể này. *x* sẽ có những giá trị nào? Hãy tưởng tượng các phân phối xác suất cho hai tổng thể này. Liệu bản chất của chúng có vi phạm các điều kiện cần thiết để cho sự kiểm định trong câu (a) là có giá trị? Hãy giải thích.

c. Tìm khoảng tin cậy 95% cho số lượng xe hơi trung bình của mỗi gia đình cho khu vực 2. Hãy giải thích khoảng này.

d. Tìm khoảng tin cậy 90% cho sự khác biệt về số lượng xe hơi trung bình của mỗi gia đình cho hai khu vực này. Hãy giải thích khoảng này.

8.91 Tuổi thọ trung bình của một mẫu gồm 100 bóng đèn huỳnh quang do một công ty sản xuất ra đươc tính toán là 1570 giờ, và độ lệch chuẩn là 120 giờ. Nếu μ là tuổi thọ trung bình của tất cả các bóng đèn do công ty này sản xuất, hãy kiểm định giả thuyết $\mu = 1600$ giờ so với giả thuyết thay thế $\mu < 1600$.

a. Tìm mức ý nghĩa quan sát được cho kiểm định này.

b. Nếu bạn có kế hoạch tiến hành kiểm định của mình bằng cách sử dụng α = 0.05, thì kết luận về kiểm định của bạn là gì?

8.92 Hiện tại thì 20% các khách hàng tiềm năng mua một nhãn hiệu xà phòng cụ thể, ví dụ, nhãn hiệu A. Để gia tăng doanh số bán hàng, một công ty lên kế hoạch cho một chiến dịch quảng cáo rộng khắp. Vào cuối chiến dịch này, một mẫu gồm 400 khách hàng tiềm năng sẽ được phỏng vấn nhằm xác định liệu chiến dịch này có thành công không.

a. Phát biểu rõ H_0 và H_a xét theo *p*, xác suất để cho một khách hàng ưa thích xà phòng nhãn hiệu A hơn.

b. Công ty này sẽ kết luận rằng chiến dịch quảng cáo này là một thành công nếu như có ít nhất 92% trong số 400 khách hàng được phỏng vấn ưa thích nhãn hiệu A hơn. Hãy tìm α. (Sử dụng ước lượng xấp xỉ chuẩn cho phân phối nhị thức để đánh giá xác suất mong muốn).

8.93 Một thí nghiệm được thực hiện nhằm so sánh các độ dài thời gian trung bình cần thiết cho hai nhân viên ngân hàng, A và B, hoàn tất công việc giấy tờ cho các tài khoản tiết kiệm cá nhân khách hàng mới. Mười khách hàng được chỉ định ngẫu nhiên cho từng nhân viên, và độ dài thời gian phục vụ được ghi lại tính bằng phút cho từng khách hàng. Các số trung bình và phương sai cho hai mẫu này được thể hiện trong bảng đi kèm sau đây.

Nhân viên A	**Nhân viên B**
$\bar{x}_1 = 22.6$	$\bar{x}_2 = 28.5$
$s_1^2 = 16.36$	$s_2^2 = 18.92$

a. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra một sự khác biệt trong các thời gian trung bình cần thiết để hoàn tất công việc giấy tờ cho một tài khoản tiết kiệm khách hàng mới không? Hãy kiểm định bằng cách sử dụng α = 0.10.

b. Tìm mức ý nghĩa xấp xỉ quan sát được cho sự kiểm định này, và giải thích giá trị của nó.

8.94 Tham khảo lại Bài tập 8.93. Tìm khoảng tin cậy 95% cho sự khác biệt về các thời gian phục vụ trung bình.

8.95 Tham khảo lại Bài tập 8.93. Giả định rằng bạn muốn ước lượng sự khác biệt về các thời gian phục vụ trung bình chính xác trong giới hạn 1 phút với xác suất xấp xỉ bằng 0.95. Một mẫu lớn xấp xỉ bao nhiêu ắt sẽ được yêu cầu cho từng nhân viên ngân hàng (giả định rằng các cỡ mẫu sẽ là bằng nhau)? [*Gợi ý:* Để giải bài này, hãy sử dụng phương pháp của Phần 7.9].

8.96 Giả sử rằng một thí nghiệm được thiết kế nhằm ước lượng sự khác biệt giữa hai số trung bình của tổng thể $(\mu_1 - \mu_2)$. Các mẫu ngẫu nhiên độc lập có cỡ n_1 và n_2 đã được chọn lựa từ hai tổng thể, và trị thống kê $(\bar{x}_1 - \bar{x}_2)$ được sử dụng làm số ước lượng. Liệu khối lượng thông tin được trích ra từ dữ liệu này được tăng lên bằng cách kết cặp các quan sát kế tiếp nhau và phân tích các khác biệt không? Liệu đây có phải là một phương pháp phân tích thích hợp không?

8.97 Khi nào thì một người nên áp dụng một sự phân tích khác biệt cặp để thực hiện các suy luận có liên quan đến sự khác biệt giữa hai số trung bình?

8.98 Một công ty công ích đã thu thập dữ liệu để so sánh độ dài thời gian cần thiết để xử lý một hóa đơn sử dụng điện bằng cách sử dụng hai phương pháp xử lý khác nhau. Tám nhân viên ghi hóa đơn, từng người một, được cho một hóa đơn sử dụng duy nhất và được yêu cầu xử lý hóa đơn này bằng cách sử dụng cả hai qui trình 1 và 2. Các thời gian xử lý (tính bằng giây) được thể hiện trong bảng đi kèm sau đây. Liệu dữ liệu này có đưa ra đủ bằng chứng để chỉ ra một sự khác biệt trong thời gian xử lý trung bình cho hai phương pháp xử lý này không?

	Qui trình	
Người xử lý	**1**	**2**
1	3	4
2	1	2
3	1	3
4	2	1
5	1	2
6	2	3
7	3	3
8	1	3

a. Tìm mức ý nghĩa xấp xỉ quan sát được cho sự kiểm định này.

b. Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$.

8.99 Đặt một khoảng tin cậy 95% cho sự khác biệt về các thời gian xử lý trung bình cho hai phương pháp xử lý trong Bài tập 8.98.

8.100 Mức tiêu thụ nước hàng tháng của một căn hộ chung cư bình quân là khoảng 48,000 galông trong vòng năm năm qua. Trung bình và độ lệch chuẩn của mức tiêu thụ hàng tháng cho 12 tháng hiện nay là $\bar{x} = 51{,}102$ galông và $s = 5127$ galông. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng một yếu tố bất thường nào đó đang gây ra một mức tiêu thụ nước lớn hơn kỳ vọng cho căn hộ này - nghĩa là, một mức tiêu thụ vượt quá giá trị trung bình là 48,000 galông mỗi tháng không? Hãy sử dụng $\alpha = 0.05$.

8.101 Một nhà tâm lý học công nghiệp muốn so sánh hai phương pháp, A và B, cho việc truyền thụ cho nhân viên mới về các chính sách nhân sự của một công ty. Hai mươi nhân viên mới được cho một bài kiểm tra trí thông minh chung và sau đó được kết hợp, tùy theo điểm số của bài kiểm tra, thành mười cặp. Từ mỗi cặp, một nhân viên được chỉ định ngẫu nhiên cho phương pháp truyền thụ A và người thứ hai cho phương pháp truyền thụ B. Mỗi nhân viên được kiểm tra vào cuối một thời kỳ kéo dài bốn tuần. Các điểm số thành tích được thể hiện trong bảng sau được ghi nhận.

Cặp	**Phương pháp A**	**Phương pháp B**
1	36	35
2	37	35

3	41	40
4	42	41
5	36	36
6	35	34
7	52	40
8	33	31
9	40	39
10	38	37

a. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra các điểm số thành tích trung bình khác nhau giữa hai phương pháp truyền thụ này không? (Hãy sử dụng $\alpha = 0.05$).

b. Ước lượng sự khác biệt trung bình trong các điểm số thành tích bằng cách sử dụng khoảng tin cậy 98%.

8.102 Một nhà sản xuất một chiếc máy để đóng gói bột xà phòng khẳng định rằng chiếc máy này có thể nhồi vào các thùng cáctông ở một trọng lượng đã biết với một dãy không lớn hơn hai phần năm của một aoxơ. Trung bình và phương sai của một mẫu gồm tám hộp nặng 4 pao được tìm thấy lần lượt bằng với 3.1 và 0.018. Hãy kiểm định giả thuyết rằng phương sai của một tổng thể gồm 8 đại lượng này là $\sigma^2 = 0.01$so với giả thuyết thay thế rằng $\sigma^2 > 0.01$. Hãy sử dụng mức ý nghĩa $\alpha = 0.05$.

8.103 Tìm khoảng tin cậy 90% cho σ^2 trong Bài tập 8.102.

8.104 Trong những giả định nào thì phân phối *F* có thể được sử dụng để tạo ra suy luận về tỷ lệ của các phương sai của tổng thể?

8.105 Giá đóng cửa của hai cổ phiếu phổ thông được ghi nhận cho một giai đoạn 15 ngày. Các số trung bình và phương sai là

$$\bar{x}_1 = 40.33 \qquad \bar{x}_2 = 42.54$$
$$s_1^2 = 1.54 \qquad s_2^2 = 2.96$$

Liệu những dữ liệu này có cung cấp đủ bằng chứng để cho thấy một sự khác biệt trong độ biến thiên của các giá đóng cửa của hai loại cổ phiếu này không? Hãy tìm giá trị *p* xấp xỉ cho kiểm định này, và giải thích giá trị của nó.

8.106 Một nhà sản xuất hóa chất khẳng định rằng độ nguyên chất trong sản phẩm của ông ta không bao giờ thay đổi lớn hơn 2%. Năm mẻ hóa chất được kiểm nghiệm và cho ra kết quả độ nguyên chất là 98.2%, 97.1%, 98.95, 97.7% và 97.9%. Liệu dữ liệu này có cung cấp đủ bằng chứng để phủ nhận sự khẳng định của nhà sản xuất đó không? [*Gợi ý:* Để dễ tính toán, ta đặt dãy 2% bằng 4σ].

8.107 Tham khảo lại Bài tập 8.106. Tìm khoảng tin cậy 90% cho σ^2.

8.108 Một công ty in dòng chữ "trọng lượng 16 aoxơ" trên nhãn hàng của mình. Người giám sát kiểm soát chất lượng chọn chín lon một cách ngẫu nhiên và cân chúng. Bà ta nhận thấy $\bar{x} = 15.7$ và $s = 0.5$. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng trọng lượng trung bình này là thấp hơn trọng lượng ghi trên nhãn hàng của công ty đó không? (Hãy sử dụng $\alpha = 0.05$).

8.109 Một đại lý bán xe hơi quyết định so sánh doanh số bán hàng trung bình hàng tháng của hai người bán hàng, A và B. Bởi vì độ lớn của doanh số bán hàng thay đổi theo mùa và các ý kiến của dân chúng về nền kinh tế, nên đại lý bán xe này quyết định thực hiện một sự so sánh trên cơ sở hàng tháng. Dữ liệu có trong bảng sau đây cho chúng ta doanh số bán hàng hàng tháng (được làm tròn đến ngàn đôla) cho hai nhân viên bán hàng này. Hãy sử dụng bản in Minitab để trả lời các câu hỏi sau đây.

Tháng	**Người bán hàng A**	**Người bán hàng B**
Giêng	130	105
Hai	141	109
Ba	163	147
Tư	176	159
Năm	147	150
Sáu	160	134
Bảy	145	124
Tám	129	140
Chín	104	91
Mười	139	124
Mười Một	163	141
Mười Hai	151	147

```
K•T C•P T CHO A - B
              N    TRUNG BÌNH   •• L•CH CHU•N  SAI S• CHU•N C•A TRUNG BÌNH
A            12     145.67         19.27                  5.56
B            12     130.00         20.51                  5.92
KHÁC BI•T    12      15.67         10.92                  3.15
Kho•ng tin c•y 95% cho khác bi•t trung bình: (8.73, 22.61)
Ki•m ••nh t cho khác bi•t trung bình = 0 (so v•i không = 0):
Giá tr• t = 4.97 Giá tr• p =
```

a. Loại hình thiết kế thí nghiệm nào đã được sử dụng?

b. Liệu dữ liệu này có cung cấp đủ bằng chứng để cho thấy một sự khác biệt trong các doanh số bán hàng trung bình cho hai nhân viên bán hàng này không? Hãy kiểm định với α = 0.05.

c. Tìm khoảng tin cậy 95% cho $(\mu_A - \mu_B)$, và giải thích các kết quả của bạn.

8.110 Trong quá khứ, một nhà máy hóa chất đã sản xuất ra một mức bình quân là 1100 pao hóa chất mỗi ngày. Hồ sơ lưu trữ cho năm qua, dựa vào 260 ngày hoạt động, cho thấy một trung bình và độ lệch chuẩn là $\bar{x} = 1060$ pao và s = 340 pao mỗi ngày. Người quản lý nhà máy này muốn kiểm định liệu sản lượng bình quân hàng ngày có giảm đáng kể trong năm vừa qua không.

a. Tìm các giả thuyết không và thay thế thích hợp.

b. Nếu z được sử dụng như là một trị thống kê kiểm định, hãy xác định vùng bác bỏ tương ứng với α = 0.05.

c. Liệu dữ liệu này có cung cấp đủ bằng chứng để cho thấy một sự sụt giảm trong sản lượng bình quân hàng ngày không?

8.111 Cả công đoàn lẫn ban quản lý đều tiến hành các cuộc điều tra về ý kiến công nhân trước khi bỏ phiếu ủng hộ hay phản đối việc lập công đoàn cho một nhà máy công nghiệp lớn. Cuộc điều tra của công đoàn, bao gồm một mẫu gồm 500 công nhân, được báo cáo là cho thấy 54% số công nhân ủng hộ việc lập công đoàn. Một cuộc điều tra tương ứng của ban quản lý về 400 công nhân chỉ nhận thấy có 46% ủng hộ việc lập công đoàn. Liệu có khả năng rằng các cuộc điều tra có liên quan đến các cỡ mẫu này ắt tạo ra các tỷ lệ phần trăm ủng hộ việc lập công đoàn mà có khác biệt lớn với các tỷ lệ được trình bày ở đây? Hoặc liệu có thể xảy ra rằng có một điều gì đó bị sai về phương pháp điều tra hay phân tích dữ liệu cho hoặc công đoàn, hoặc ban quản lý, hay cả hai chăng? Tìm giá trị *p* đi cùng với sự kiểm định thích hợp và sử dụng giá trị này để thực hiện quyết định của mình.

8.112 Một khách sạn cần một tỷ lệ lấp đầy là 60% nhằm đạt được lợi nhuận. Chọn mẫu ngẫu nhiên 50 ngày tạo ra một tỷ lệ lấp phòng đầy trung bình là 62% và một độ lệch chuẩn là 8%. Liệu những dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng tỷ lệ lấp đầy phòng trung bình (cho tổng thể các ngày đại diện cho những ngày trong mẫu này) vượt quá 60% không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.10$.

8.113 Tham khảo lại Bài tập 8.110. Sử dụng qui trình được mô tả trong Ví dụ 8.5 để tính toán β cho nhiều giá trị khác nhau của μ. (Ví dụ, $\mu = 1040, 1030, 1020$). Sử dụng các giá trị tính toán được của β để xây dựng nên đường cong năng lực cho kiểm định thống kê này.

8.114 Hãng hàng không All-Nippon Airways (ANA) của Nhật Bản đã tìm thấy rằng các con mắt đầy vẻ đe dọa được sơn trên các ống hút gió cho động cơ phản lực máy bay của họ đã làm các con chim sợ hãi tránh xa và tiết kiệm được tiền bảo trì (*Gainesville Sun*, 16 tháng Mười Một, 1986). Một nghiên cứu về máy bay chở khách nhiều động cơ trong khoảng thời gian một năm đã tìm thấy một mức bình quân là một con chim đâm vào động cơ được sơn của máy bay so với mức trung bình chín con chim va chạm với các động cơ không được sơn.

- **a.** Loại dữ liệu nào mà bạn ắt kỳ vọng về số lượng chim đâm vào động cơ máy bay? Hãy giải thích.
- **b.** Giải thích tại sao số lượng các vụ đâm vào động cơ trên cùng một máy bay có lẽ là phụ thuộc và vì vậy vi phạm giả định rằng mẫu này được chọn ngẫu nhiên.

8.115 Tham khảo lại Bài tập 8.114. Giả sử rằng tất cả máy bay trong nghiên cứu này chứa đựng số lượng động cơ phản lực như nhau và rằng *x* tượng trưng cho số lượng chim đâm vào các động cơ tính trên mỗi máy bay.

- **a.** Nếu $n = 40$ máy bay được chọn ngẫu nhiên để sơn các động cơ, thì liệu 40 giá trị của *x* này có đại diện cho một mẫu ngẫu nhiên không?
- **b.** Giả sử rằng giá trị trung bình của các vụ chim va chạm vào động cơ không được sơn tính trên mỗi máy bay là $\mu = 9$. Nếu số trung bình của mẫu về các vụ va chạm của chim tính trên mỗi động cơ của mỗi máy bay là $\bar{x} = 1$, liệu bạn có đủ bằng chứng để chỉ ra rằng các con mắt đầy vẻ đe dọa được sơn trên những động cơ đã tạo ra một sự giảm sút trong con số trung bình các vụ va chạm của chim tính trên mỗi động cơ của mỗi máy bay không? Hãy kiểm định bằng cách sử dụng $\alpha = 0.05$. [*Gợi ý:* Số lượng các vụ va chạm của chim tính trên mỗi động cơ có khả năng là một biến số ngẫu nhiên Poisson. Bởi vì độ lệch chuẩn σ của một biến số ngẫu nhiên Poisson là bằng với căn bậc hai của trung bình μ của nó - nghĩa là, $\sigma = \sqrt{\mu}$ - cho nên độ lệch chuẩn này có thể được ước lượng bằng cách thay thế $\bar{x}$ cho μ, nghĩa là, $\hat{\sigma} = \sqrt{\bar{x}}$.]

8.116 Tham khảo lại nghiên cứu của hãng hàng không All-Nippon Airways được mô tả trong các Bài tập 8.114 và 8.115. All-Nippon tìm thấy rằng các con mắt đầy vẻ đe dọa được sơn trên các ống hút gió cho động cơ phản lực máy bay của họ dường như đã tạo ra một sự giảm đi trong số lượng trung bình các vụ va chạm của chim với mỗi động cơ phản lực (Họ ước tính một sự giảm bớt của chi phí bảo trì khoảng $200,000 cho một số lượng nhỏ các máy bay trong nghiên cứu của mình). Giả định rằng nghiên cứu này đã liên quan đến các mẫu ngẫu nhiên độc lập các máy bay, 40 máy bay có ống hút gió được sơn và 40 máy bay không có, và rằng số trung bình của mẫu về các vụ va chạm của chim tính trên mỗi máy bay là $\bar{x}_1 = 1$ cho các máy bay có ống hút gió được sơn và $\bar{x}_2 = 9$ cho những máy bay có ống hút gió không được sơn. Liệu những dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng con số trung bình các vụ va chạm của chim trên mỗi động cơ của mỗi máy bay là ít hơn đối với các máy bay có ống hút gió được sơn không? Hãy kiểm định bằng cách sử dụng α = 0.05. [*Gợi ý:* Số lượng các vụ va chạm của chim vào mỗi động cơ của mỗi máy bay có thể là một biến số ngẫu nhiên Poisson. Bởi vì độ lệch chuẩn σ của một biến số ngẫu nhiên Poisson là bằng với căn bậc hai của trung bình μ của nó - nghĩa là, $\sigma = \sqrt{\mu}$ - cho nên độ lệch chuẩn này có thể được ước lượng bằng cách thay thế $\bar{x}$ cho μ, nghĩa là, $\hat{\sigma} = \sqrt{\bar{x}}$.]

8.117 Một cách để so sánh giá cả tương đối của hai loại chứng khoán là so sánh các hệ số giá-thu nhập của chúng, hệ số của giá cả mỗi cổ phiếu của một loại chứng khoán với lượng tiền kiếm được bởi công ty đó tính trên mỗi cổ phiếu hàng năm. Bảng phía dưới đây (trong Bài tập 8.118) liệt kê các hệ số giá-thu nhập (hệ số P/E) cho mười công ty điện lực được chọn ngẫu nhiên so với tám loại chứng khoán blue-chip vào tháng Bảy năm 1994.

a. Liệu dữ liệu trên có cung cấp đủ bằng chứng để chỉ ra một sự khác biệt trong độ biến thiên của hệ số P/E giữa các cổ phiếu phổ thông ngành điện lực và cổ phiếu blue-chip của các ngành không phải dịch vụ công ích không? Hãy kiểm định bằng cách sử dụng α = 0.10.

b. Tìm giá trị *p* xấp xỉ cho kiểm định này, và giải thích giá trị của nó.

8.118 Tham khảo lại Bài tập 8.117. Căn cứ vào các kết quả của câu (a), liệu dữ liệu này có cung cấp bằng chứng đủ để chỉ ra những sự khác biệt trong các hệ số P/E giữa các cổ phiếu phổ thông ngành điện lực và cổ phiếu blue-chip của các ngành không phải dịch vụ công ích không?

a. Hãy kiểm định bằng cách sử dụng α = 0.05.

Bảng cho Bài tập 8.117

Cổ phiếu Điện lực	**P/E**	**Cổ phiếu Blue-chip**	**P/E**
Carolina Power & Light	13	IBM	14
Minnesota Power & Light	14	Abbot Labs	16
TECE Energy	14	Minnesota Mining	18
Duke Power	13	Safeway Stores	16
Wisconsin Energy Corp.	17	Reynolds Metals	18
Pacific Gas & Electric	10	Monsanto	15
Montana Power	12	Hilton Hotels	28
Houston Industries	11	Textron	12
Illinois Power	13		
Pensylvania Power & Light	10		

Nguồn: Dữ liệu từ *Press-Enterprise*, Riverside, California, 29/07/1994.

b. Tìm giá trị *p* xấp xỉ cho kiểm định này, và giải thích giá trị của nó.

8.119 Do có sự biến thiên trong các khoản chiết khấu từ việc đổi xe, mà lợi nhuận tính trên mỗi xe hơi mới bán ra bởi một đại lý bán xe thay đổi từ xe này sang xe khác. Lợi nhuận mỗi thương vụ bán hàng (tính bằng trăm đôla), được ghi vào bảng cho tuần vừa qua, là như sau:

6.3	9.4
6.2	7.7
4.4	8.3

Liệu những dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng lợi nhuận bình quân mỗi thương vụ bán xe là thấp hơn $780 không? Hãy kiểm định ở mức ý nghĩa $\alpha = 0.05$.

8.120 Một nhà sản xuất tivi khẳng định rằng sản phẩm của ông ta có một tuổi thọ trung bình không bị trục trặc là 3 năm. Ba hộ gia đình trong một cộng đồng đã mua các chiếc tivi này, và tất cả ba chiếc tivi được quan sát thấy là đã bị hỏng trước 3 năm, với các tuổi thọ đến lúc bị hỏng lần lượt là 2.5, 1.9 và 2.9 năm.

a. Liệu những dữ liệu trên có cung cấp đủ bằng chứng để phủ nhận sự khẳng định của nhà sản suất đó không? Hãy kiểm định bằng cách sử dụng mức ý nghĩa $\alpha = 0.05$.

b. Tính toán khoảng tin cậy 90% cho tuổi thọ trung bình của các chiếc tivi này.

8.121 Tham khảo lại Bài tập 8.120. Xấp xỉ có bao nhiêu quan sát ắt được đòi hỏi để ước lượng tuổi thọ trung bình của các chiếc tivi đó chính xác trong giới hạn 0.2 năm với xác suất bằng 0.90?

8.122 Một thí nghiệm được thực hiện nhằm so sánh hai thiết kế xe hơi. Hai mươi người được chọn ngẫu nhiên, và mỗi người được hỏi để xếp hạng mỗi thiết kế trên một thang từ 1 (nghèo nàn) đến 10 (tuyệt hảo). Các số điểm xếp hạng tạo ra sẽ được sử dụng để kiểm định giả thuyết không rằng mức chấp nhận trung bình là giống nhau cho cả hai thiết kế so với giả thuyết thay thế là rằng một trong hai thiết kế xe hơi này được ưa thích hơn. Liệu những dữ liệu này có thỏa mãn các giả định cần thiết cho một kiểm định *t* Student của Phần 8.5 không? Hãy giải thích.

8.123 Dữ liệu trong bảng đi kèm sau đây về các sự cố tổn thất thời gian (tổn thất số giờ sản xuất trung bình mỗi tháng trong một giai đoạn một năm) được thu thập cho cả thời gian trước lẫn sau khi một chương trình an toàn công nghiệp có hiệu lực. Dữ liệu được ghi nhận cho sáu nhà máy công nghiệp. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra liệu chương trình an toàn này có hiệu quả trong việc giảm bớt các sự cố tổn thất thời gian không? (Hãy sử dụng $\alpha = 0.10$).

	Số Nhà máy					
Dữ liệu thu thập được	**1**	**2**	**3**	**4**	**5**	**6**
Trước chương trình	38	64	42	70	58	30
Sau chương trình	31	58	43	65	52	29

8.124 Để so sánh nhu cầu cho hai món khai vị khác nhau, nhà quản lý một quán ăn tự phục vụ đã ghi nhận số lượng mua hàng cho từng món khai vị trong bảy ngày liên tiếp. Dữ liệu được thể hiện trong bảng sau. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra một nhu cầu trung bình lớn hơn cho một trong hai món khai vị không?

	Món khai vị	
Ngày	**A**	**B**
Thứ Hai	420	391
Thứ Ba	374	343
Thứ Tư	434	469
Thứ Năm	395	412
Thứ Sáu	637	538
Thứ Bảy	594	521
Chủ Nhật	679	625

8.125 Giới hạn của Cơ quan Bảo vệ Môi trường (EPA) về lượng thải ra được phép của các chất rắn lơ lửng vào các con sông và dòng suối là 60 miligam trên mỗi lít (mg/l) mỗi ngày. Một nghiên cứu về các mẫu nước được lấy từ cống xả tại một mỏ phốtphát cho thấy rằng trong một khoảng thời gian dài thì lượng thải trung bình hàng ngày của các chất rắn lơ lửng là 48 mg/l, nhưng lượng chất thải đo được mỗi ngày là rất thay đổi. Những người thanh tra của tiểu bang đã đo lường mức thải của các chất rắn lơ lửng cho n = 20 ngày và nhận thấy rằng s^2 = 39 (mg/l)^2. Tìm khoảng tin cậy 90% cho σ^2. Giải thích các kết quả của bạn.

8.126 Một nhà sản xuất các động cơ điện đã so sánh năng suất của những công nhân lắp ráp về hai loại kế hoạch làm việc 40 giờ hàng tuần, bốn ngày làm việc mười tiếng (kế hoạch 1) và năm ngày làm việc tám tiếng tiêu chuẩn (kế hoạch 2). Hai mươi công nhân được chỉ định cho từng kế hoạch làm việc, và số lượng các đơn vị sản phẩm lắp ráp được ghi nhận cho một giai đoạn là một tuần. Các số trung bình (tính bằng trăm đơn vị sản phẩm) và phương sai của mẫu cho hai kế hoạch làm việc này được trình bày trong bảng đi kèm sau đây.

	Kế hoạch Làm việc	
Trị thống kê	**1**	**2**
Trung bình của mẫu	43.1	44.6
Phương sai của mẫu	4.28	3.89

a. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra một sự khác biệt trong năng suất trung bình cho hai kế hoạch làm việc này không? Hãy kiểm định bằng cách sử dụng α = 0.05.

b. Tìm giá trị *p* xấp xỉ cho kiểm định này, và giải thích giá trị này.

c. Tìm khoảng tin cậy 95% cho sự khác biệt về năng suất trung bình cho hai kế hoạch làm việc này, và giải thích khoảng này.

8.127 Giả sử rằng nhà sản xuất trong Bài tập 8.126 mong muốn ước lượng sự khác biệt về năng suất trung bình hàng tuần cho hai kế hoạch làm việc này chính xác trong giới hạn một đơn vị sản phẩm. Có bao nhiêu công nhân ắt đã phải được đưa vào từng mẫu? [*Gợi ý:* Để giải bài này, hãy sử dụng phương pháp trong Phần 7.9.]

8.128 Tỷ lệ phần trăm (hay đôla) lợi nhuận trung bình mỗi dự án không phải là mối quan tâm duy nhất của một nhà phát triển bất động sản (hay bất cứ loại hình nhà đầu tư nào). Nhà phát triển này phải quan tâm đến một sự thay đổi lớn trong khoản lãi, bởi vì một khoản lãi có giá trị âm lớn (khoản lỗ) có thể khiến nhà phát triển này phải dừng cuộc chơi. Một nhà phát triển cụ thể đã lên kế hoạch cho các dự án để đạt được một tỷ suất lợi nhuận trung bình mỗi dự án à 12% với một dãy không lớn hơn 25%. Việc chọn mẫu các tỷ lệ phần trăm lợi nhuận mỗi dự án cho 25 dự án sau cùng trong số các dự án của nhà phát triển này tạo ra một trung bình và độ lệch chuẩn của mẫu lần lượt bằng 11.1% và 5.2%.

a. Giả sử rằng nhà phát triển này muốn khá chắc chắn rằng dãy của tỷ lệ phần trăm lợi nhuận mỗi dự án là không lớn hơn 25%. Giá trị nào của σ sẽ đạt được cho mục tiêu này? [*Gợi ý:* Hầu như tất cả các quan sát trong một tổng thể rơi vào bên trong 3 σ của trung bình tổng thể μ.]

b. Liệu dữ liệu này có cung cấp đủ bằng chứng để chỉ ra rằng sự thay đổi trong tỷ lệ phần trăm lợi nhuận mỗi dự án là lón hơn giá trị của σ đã được xác định cụ thể trong câu (a)? Hãy kiểm định bằng cách sử dụng α = 0.05.

8.129 Tham khảo lại Bài tập 8.128. Tìm khoảng tin cậy 95% cho phương sai của tỷ lệ phần trăm lợi nhuận mỗi dự án cho nhà phát triển này, và giải thích khoảng này.

Các Bài tập Sử dụng cácBộ Dữ liệu tại địa chỉ http://swlearning.com

8.130 Tham khảo bộ dữ liệu A. Đối với một vị trí xếp hạng cố định cho một phân khoa (khác với vị trí xếp hạng mà bạn đã chọn trong Bài tập 7.102), hãy chọn một mẫu có cỡ $n_1 = 10$ trong số 246 tiền lương của là nam giới, và, một cách độc lập, chọn một mẫu khác có cỡ $n_2 = 10$ trong số 246 tiền lương của nữ giới.

a. Sử dụng các kết quả mẫu của bạn (giả định một phương sai cơ bản chung) để kiểm định $H_0 : \mu_M - \mu_F = 0$ so với $H_a : \mu_M - \mu_F \neq 0$ với $\alpha = 0.05$.

b. Lập ước lượng khoảng tin cậy 95% của $\mu_M - \mu_F$. Liệu ước lượng của bạn có nhất quán với các kết quả trong câu (a) không?

8.131 Tham khảo lại bộ dữ liệu A. Một lần nữa, hãy chọn một vị trí xếp hạng khoa (vị trí xếp hạng giống như vị trí được chọn trong Bài tập 8.130 là tốt nhất) và bây giờ chọn lựa một mẫu kết cặp có cỡ n = 10 mà trong đó bạn ngẫu nhiên lựa chọn n = 10 trường đại học và ghi nhận cả tiền lương cho nam lẫn nữ cho mỗi trong số n = 10 trường này.

a. Sử dụng các mẫu kết cặp của mình, hãy kiểm định giả thuyết không $H_0 : \mu_M - \mu_F = \mu_d = 0$ so với $H_a : \mu_M - \mu_F = \mu_d \neq 0$ với $\alpha = 0.05$. Liệu kết luận của bạn có thay đổi nếu giả thuyết thay thế là $H_a : \mu_M - \mu_F = \mu_d > 0$ không?

b. Lập ước lượng khoảng tin cậy 95% của $\mu_M - \mu_F$.

c. So sánh giá trị của $s_{\bar{d}}$ được tìm ra trong câu (a) với giá trị của $s_{\bar{x}_M - \bar{x}_F}$ được tìm ra trong Bài tập 8.130. Liệu thiết kế khác biệt cặp này có cung cấp cho ta nhiều sự chính xác hơn trong các suy luận có liên quan đến $\mu_M - \mu_F = \mu_d$ không?

8.132 Tham khảo dữ liệu về kỳ hạn chi trả bình quân (bộ dữ liệu C). Hãy chọn ra một mẫu ngẫu nhiên có cỡ n = 30 (bạn có thể sử dụng một trong các mẫu mà bạn đã sử dụng trong Bài tập 7.104).

a. Bởi vì trung bình của tổng thể mà từ đó bạn đang chọn mẫu trên thực tế là $\mu = 38.56$, cho nên một kiểm định giả thuyết $H_0 : \mu = 38.56$ không nên bị bác bỏ. Hãy tiến hành kiểm định này bằng cách sử dụng một giả thuyết thay thế hai phía, với α = 0.01. Liệu bạn có đi đến một quyết định chính xác không?

b. Một kiểm định giả thuyết $H_0 : \mu = 20.00$ nên bị bác bỏ (bởi vì chúng ta biết rằng $\mu = 38.56$ cho tổng thể này). Hãy tiến hành kiểm định này bằng cách sử dụng thông tin mẫu và giả thuyết thay thế hai phía của bạn, với α = 0.01. Liệu bạn có đi đến một quyết định chính xác không?

c. Liệu có khả năng xảy ra là bạn sẽ có thể bác bỏ một kiểm định về giả thuyết $H_0 : \mu = 37.00$ mặc dù chúng ta biết rằng đây không phải là số trung bình thực sự của tổng thể không? Hãy giải thích.