

Chương Trình Giảng Dạy Kinh tế Fulbright

Học kỳ Thu năm 2010

Các Phương Pháp Phân Tích Định Lượng

Đáp án Bài tập 5

KIỂM ĐỊNH THỐNG KÊ

Bài 1:

Trong năm 1960 một cuộc điều tra dân số đã chỉ ra rằng độ tuổi mà người đàn ông Mỹ lập gia đình lần đầu có trung bình là 23,3. Tuy nhiên, có một đánh giá được chấp thuận khá rộng rãi là nam thanh niên ở Mỹ ngày nay lập gia đình trễ hơn. Chúng ta muốn kiểm định xem độ tuổi trung bình của lần lập gia đình đầu tiên của họ có tăng lên sau 40 năm hay không?

a) Hãy lập ra các giả thuyết hợp lý (H_0 và H_a)?

$$H_0: \mu = 23,3$$

$$H_a: \mu > 23,3$$

Lấy mẫu ngẫu nhiên 40 đàn ông ở Mỹ mới lập gia đình lần đầu tiên năm ngoái. Kết quả thống kê mô tả từ mẫu cho thấy độ tuổi trung bình của họ khi lập gia đình lần đầu tiên là 24,2 tuổi và độ lệch chuẩn là 5,3 năm.

b) Thực hiện việc kiểm định giả thuyết và tìm p_{value} ?

$$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{24,2 - 23,3}{5,3 / \sqrt{40}} = 1,07$$

$$p_{value} = P(z > 1,07) = 0,1432$$

c) Giải thích ý nghĩa của p_{value} trong trường hợp này?

Nếu tuổi trung bình của lần lập gia đình đầu tiên vẫn là 23,3 tuổi, vẫn có xác suất là 14,32% là chúng ta có được trung bình mẫu là 24,2 tuổi hoặc lớn hơn do sự biến thiên trong việc lấy mẫu tự nhiên

d) Các anh chị hãy cho biết kết luận của mình?

p_{value} là xác suất sai lầm loại I (bác bỏ H_0 khi nó đúng) được tính theo trị kiểm định của mẫu là 14,32%, lớn hơn sai lầm cho phép là 5%. Không bác bỏ H_0 .

Bài 2:

Một sinh viên điều tra những ảnh hưởng tiềm tàng của chất caffeine đối với việc học thi. Có 30 sinh viên tình nguyện tham gia quá trình điều tra này và được chia làm 2 nhóm. Đầu tiên, mỗi nhóm sẽ làm một bài kiểm tra về trí nhớ. Sau đó, một nhóm sẽ được cho uống hai ly nước ngọt có caffeine và nhóm còn lại sẽ uống hai ly không có caffeine. Ba mươi phút sau mỗi nhóm sẽ thi một bài kiểm tra khác có độ khó tương đương với bài kiểm tra đầu tiên. So sánh kết quả thi trong 2 lần cho thấy đối với nhóm 15 sinh viên uống nước ngọt có chất caffeine thì điểm trung bình giảm xuống -0,933 điểm với độ lệch chuẩn là 2,988 điểm. Đối với nhóm uống nước ngọt không có caffeine thì điểm trung bình tăng lên 1,429 điểm với độ lệch chuẩn là 2,441 điểm. Giả sử khác biệt về điểm thi tuân theo phân phối chuẩn.

- a) Điểm của nhóm sinh viên uống nước ngọt có chất caffeine có thay đổi một cách có ý nghĩa không? Kiểm định các giả thuyết và cho biết kết luận của các bạn?

$$H_0: \mu_d = 0$$

$$H_a: \mu_d \neq 0$$

$$t = \frac{\bar{d} - 0}{s_d \sqrt{n}} = \frac{-0,933}{2,988 / \sqrt{15}} = -1,21$$

$$-2,145 < t < 2,145 \rightarrow \text{Không bác bỏ } H_0$$

- b) Điểm của nhóm sinh viên uống nước ngọt không có chất caffeine có thay đổi một cách có ý nghĩa không? Kiểm định các giả thuyết và cho biết kết luận của các bạn?

$$H_0: \mu_d = 0$$

$$H_a: \mu_d \neq 0$$

$$t = \frac{\bar{d} - 0}{s_d \sqrt{n}} = \frac{1,429}{2,441 / \sqrt{15}} = 2,27$$

$$t > 2,145 \rightarrow \text{Bác bỏ } H_0$$

Bài 3:

Một Trung tâm Quốc gia về Thống kê Giáo dục thường xuyên giám sát các khía cạnh khác nhau của giáo dục phổ thông. Các số liệu thống kê của họ trong năm 1996 được xem là cơ sở để đánh giá những thay đổi trong giáo dục phổ thông. Trong năm 1996 có 34% học sinh đã không nghỉ học một ngày nào. Một cuộc điều tra trong năm 2000 cho thấy khi khảo sát 8368 học sinh phổ thông thì tỷ lệ này giảm xuống còn 33%. Các quan chức giáo dục rất quan tâm đến vấn đề này và cho rằng tỷ lệ học sinh tham gia lớp đã có sự thay đổi.

- a) Hãy lập ra các giả thuyết hợp lý (H_0 và H_a)?

$$H_0: p_{2000} = 0,34$$

$$H_a: p_{2000} \neq 0,34$$

b) Thực hiện việc kiểm định giả thuyết và tìm giá trị p_{value} ?

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0,33 - 0,34}{\sqrt{\frac{0,34 * 0,66}{8368}}} = -1,93$$

$$p_{value} = 2 * P(z > 1,93) = 2 * 0,0268 = 0,0536$$

c) Các bạn có kết luận gì về tình trạng tham gia lớp của các học sinh?

Với p_{value} tương đối nhỏ chúng ta có thể bác bỏ giả thuyết H_0 (tỷ lệ học sinh tham gia lớp trong năm 2000 một cách có ý nghĩa thống kê khác với số liệu năm 1996 là 34%). Tuy nhiên, nếu bạn một cách máy móc so sánh p_{value} với mức ý nghĩa 5% và kết luận là không bác bỏ H_0 thì kết luận đó cũng được chấp nhận.

d) Bạn có nghĩ rằng sự khác biệt sự khác biệt giữa tỷ lệ học sinh tham gia lớp trong hai năm 1996 và 2000 là có ý nghĩa không? Giải thích?

Sự khác biệt giữa tỷ lệ học sinh tham gia lớp trong hai năm 1996 và 2000 là có ý nghĩa thống kê nhưng lại không có ý nghĩa lớn về mặt thực tế. Chúng ta có thể chờ một vài năm nữa để lấy mẫu và kiểm định lại giả thuyết.

Bài 4:

Theo báo cáo của Hiệp hội Ung thư Hoa Kỳ (“Hô 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. Trong các mẫu ngẫu nhiên gồm 200 đàn ông và 200 phụ nữ được điều tra thì có 62 đàn ông và 54 phụ nữ là những người hút thuốc.

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?

Tỷ lệ đàn ông và phụ nữ hút thuốc là:

$$p_{nam}^{\wedge} = 62/200 = 0,31; q_{nam}^{\wedge} = 1 - 0,31 = 0,69$$

$$p_{nu}^{\wedge} = 54/200 = 0,27; q_{nu}^{\wedge} = 1 - 0,27 = 0,73$$

$$H_0: p_{nam} - p_{nu} = 0$$

$$H_a: p_{nam} - p_{nu} \neq 0$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sigma_{(\hat{p}_1 - \hat{p}_2)}} = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\frac{\hat{p}_{nam} \hat{q}_{nam}}{n_1} + \frac{\hat{p}_{nu} \hat{q}_{nu}}{n_2}}} = \frac{0,31 - 0,27}{\sqrt{\frac{0,31 * 0,69}{200} + \frac{0,27 * 0,73}{200}}} = 0,88$$

$-1,96 < z < 1,96 \rightarrow$ Không bác bỏ giả thuyết H_0 với mức ý nghĩa 5%