Results

Xingkai Wu

15
Earned points

40:45
Time for this attempt

Your Answers:

1 0 points earned

The following regression equation models price of car (y; dependent variable, in $ 1000s) as a function of brands (independent variables). There are three brands, coded as **effect** variables (x1, x2, x3), where brand 3 is the reference brand.

$$y = 50 + 50\,x1 - 30\,x2 + e$$

Assuming equal sample sizes by brands, what would be the **estimated average price** of brand x3 car in the market?

- ○ $30,000
- ○ $60,000
- ✗ ◉ $50,000

Correct Answer: **$30,000**

- ○ $70,000
- ○ $23,333

2 1 point earned

If X (customer satisfaction) predicts Y (sales), then which of the following is most likely to be true, in absence of any other information given to you?

- ✓ ◉ X and Y are correlated
- ○

X directly causes Y

X indirectly causes Y, via another variable W

X causes Y, both directly and indirectly

X causes Y, and Y causes X

3 1 point earned

Which of the following statements is TRUE about reliability and validity?

If customer satisfaction is measured reliably, then it is valid

It is possible to measure customer satisfaction with validity, but not reliably

If customer satisfaction is measured with validity, then it is reliable

It is not possible to measure customer satisfaction with validity

✓ If customer satisfaction is measured reliably, then it is not necessarily valid

4 1 point earned

Suppose a linear regression model of sales y, measured in actual dollar $CAD,

$$y = a + b1\,x1 + b2\,x2 + e$$

where x1=quality (measured on a scale of 1-10, 1= poor, 10=excellent), x2 is % discount (continuous variable, 0-100).

Which of the following statement is TRUE?

Impact of x1 on y, as measured by b1, is not dependent on how (x1,x2) are correlated

✓ b1 measures partial correlation between (y,x1)

b1 is always going be less than b2 because the spread of x1 (i.e., 1-10) is much less than that of x2 (i.e., 0-100)

b1 measures zero order correlation between (y,x1)

b1 measures semi-partial correlation between (y,x1)

5 2 points earned

Which of the following scenarios can be modeled using linear regression?

- How price increase can lead to specific change in sales?
- How price and sales are correlated?
- How sales increase can lead to specific change in price?
- ✓ How average price can be predicted using sales?
- How specific price increase can lead to specific change in sales?

6 1 point earned

Suppose you placed three puppies (A1, A2, A3) under different diet programs for six weeks. At the end of six weeks, you measured their respective weights. What is the independent variable in this experiment?

- Time interval, i.e., six weeks
- Puppies
- Weights
- ✓ Diet programs
- Heights

7 1 point earned

Suppose a linear regression model of sales y, measured in actual dollar $CAD,

$$y = a + b1\,x1 + b2\,x2 + b12\,x1^{*}x2$$

where x1=gender (0=M, 1=F), x2 is % discount (continuous variable, 0-100).

Which of the following statements is TRUE?

- b12 is always greater than b2
- ✓ b12 measures the interaction effect of (x1,x2) on sales
- b12 is geometric average of (b1,b2)
- b12 is always greater than b1
- b12 is always less than min (b1,b2)

8 0 points earned

Suppose you have p variables, X (i), where i=1,2,3...p, measured on interval scale. The correlation matrix, Correlation (X), based on a sample of n=100k is an identity matrix, I (pxp), of order p.

Which of the following is the most likely to be TRUE?

- ◯ We can apply dependence multivariate models but not interdependence models using the original data
- ◯ For any subset of the given sample of n=100k, the correlation matrix, Correlation (X), is always going to be an identity matrix
- ◯ For any subset of the given sample of n=100k, if the correlation matrix is identity matrix, then it implies the data is not suitable for multivariate analytics
- ✗ ◉ For any (i,j), (X(i), X(j)) are independent

Correct Answer: **For any subset of the given sample of n=100k, if the correlation matrix is identity matrix, then it implies the data is not suitable for multivariate analytics**

- ◯ The data is not suitable for any multivariate analytics

9 1 point earned

Suppose you want to understand the impact of screen time on sleeping patterns among MBA students at U of T. You placed 33% participants in group A, where 2 hours of screen time is allowed in a day, 33% participants in group B, where 4 hours of screen time is allowed in a day, and the remaining 34% in group C, where no screen time is allowed in a day.

This is an experiment best characterized by

- ◯ Within group design: Independent variable is sleeping patterns, dependent variable is screen time
- ◯ Between group design: Independent variable is sleeping patterns, dependent variable is screen time
- ◯ Within group design: Independent variable is education (e.g., MBA), dependent variable is sleeping patterns
- ✓ ◉ Between group design: Independent variable is screen time, dependent variable is sleeping patterns
- ◯ Within group design: Independent variable is screen time, dependent variable is sleeping pattern

10 1 point earned

The following regression equation models price of car (y; dependent variable, in $ 1000s) as a function of brands (independent variables). There are three brands, coded as dummy variables (x1, x2, and x3), where brand 3 is the reference brand.

$$y = 50 + 10\,x1 + 20\,x2 + e$$

If the sample sizes by brands are n1=30, n2=20, n3=50

What would be the **estimated average price** of car in the market (average across all brands)?

- ○ $32,000
- ○ $180,000
- ○ $80,000
- ✓ ◉ $57,000
- ○ $23,333

11 0 points earned

Which measurement scale is used for the following variables X1 and X2?

X1= Reaction time of two chemicals in a lab (in seconds)

X2= Air quality of the lab (measured on a scale of 1-5, 1= extremely poor, 5= excellent)

- ✗ ◉ X1=interval, X2= ordinal

Correct Answer: **X1= ratio, X2=ordinal**

- ○ X1= ratio, X2=ordinal
- ○ X1=categorical, X2=ordinal
- ○ X1=ratio, X2=categorical
- ○ X1=nominal, X2=interval

12 1 point earned

Which of the following is the most likely to be TRUE for variables measured on nominal scale?

- ○ We can use only dependence multivariate techniques but not interdependence using nominal variables
- ○

Nominal variables can be easily converted into ratio scale variables

Such variables are not suitable for multivariate analytics

Design of experiment or A/B testing can be performed only with nominal variable

✓ We can use nominal variables as valid independent variables in linear regression but we need to interpret them appropriately and correctly

13 0 points earned

Which of the following statements is true regarding hypothesis testing?

× A null hypothesis can be rejected or accepted based on some analysis from the data

Correct Answer: **A null hypothesis can only be rejected but can never be accepted**

A null hypothesis can only be rejected but can never be accepted

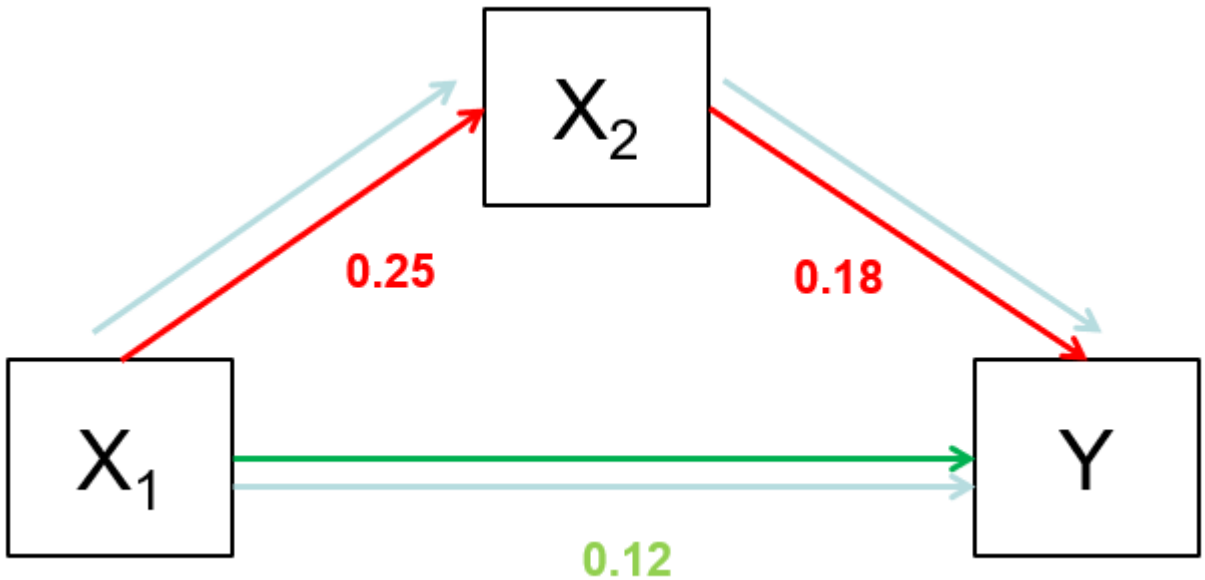
When a null hypothesis is rejected, it is always rejected correctly

It is only when significance level is at 5%, i.e., alpha=5%, we have reasons to believe in the outcome of our hypothesis testing

False negative probability is exactly the same as false positive probability

14 1 point earned

What would be the total estimated impact of x1 on y based on the following figure



0.310

- 0.550
- 0.370
- 0.050
- ✓ 0.165

15 0 points earned

Suppose you want to study the impact of specific diet on academic performance among high school students using an experimental design. Which of the following types of experiment is the most suitable for studying such impact?

- Lab experiment
- Either lab or field experiment
- ✗ Natural experiment

Correct Answer: **Field experiment**

- Either lab or natural experiment
- Field experiment

16 2 points earned

Suppose you decide to use multiple independent t-tests instead of an ANOVA to test whether **four** random samples are from the same population using $\alpha = 0.06$, i.e., 6%. What would be the joint confidence level of your multiple t-tests?

- 76%
- ✓ 78%
- 94%
- 6.0%
- 88%

17 2 points earned

The following regression equation $Y = 50 + 10\,x1 + 20\,x2 + e$ is derived using **dummy coding**. Assuming there are three independent variables, x1,x2,x3, with x3 being the omitted variable.

What would be the transformed equation using **effect coding**, assuming equal sample size across groups?

- [x] Y = 60 +10 x2 + e
- [ ] Y = 60 -10 x3 + e
- [ ] Y = 80 - 20 x1 - 10 x2 + e
- [ ] Y = 26.7 + 10 x1 + 20 x2 + e
- [ ] Y = 180 + 120 x1 + 110 x2 + e