Module 3 Assesment

BTRY 6020

Question 1 (2 pts)

Suppose I believe my data is generated by the following model:

$$Y_i = b_0 + b_1 X_{i,1} + b_2 X_{i,2} + b_3 X_{i,3} + b_4 X_{i,4} + \varepsilon_i.$$

I want to test the null hypothesis that $X_{i,2}$ is not associated with Y_i after adjusting for $X_{i,1}$, $X_{i,3}$, and $X_{i,4}$. The alternative hypothesis is that there is some association between $X_{i,2}$ and Y_i, even after adjusting for the other covariates. What is the null hypothesis and the alternative hypothesis:

Answer

Question 2 (2 pts)

Suppose I gather 35 observations and fit the model specified above. Given the output below, calculate the t-statistic for testing the hypothesis. Round this answer to two digits after the decimal.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.971	0.171	11.533	0
X1	0.861	0.212	4.063	0
X2	0.373	0.194	???	???
X3	1.078	0.188	5.738	0
X4	-0.057	0.221	-0.259	0.798

Answer

Question 3 (2 pts)

Suppose you are interested in testing the null hypothesis $H_0 : b_2 = .5$. Given the table above, calculate the t-statistic for testing this null hypothesis. Round this answer to two digits after the decimal.

Answer

Question 4 (2 pts)

You can use the `qt` function to get the cut-off from a T distribution. Specifically, the code below gets the cutoff so that the area to right of that cut-off is alpha / 2 for a T distribution with z degrees of freedom. Calculate the p-value for the table in Question 2, and also for the null hypothesis in Question 3.

```
qt(alpha / 2, df = z, lower = F)
```

Question 5 (1 pt)

Calculate a 90% confidence interval for the coefficient of X_1.

Answer

Question 6 (2 pts)

Consider two worlds. In both, you are interested in testing the null hypothesis that $H_0 : b_1 = 0$ vs $H_A : b_1 \neq 0$. In the first setting $b_1 = 1$ and in the second setting $b_1 = 2$. If all other things are equal, in which setting do you have more power to reject the null hypothesis. Give a brief explanation of why?

Question 7 (2 pts)

Suppose you are interested in testing the null hypothesis that $H_0 : b_1 = 0$ vs $H_A : b_1 \neq 0$. However, the true $b_1 = 1$. Suppose you are deciding to test the null hypothesis with either a $\alpha = .05$ or $\alpha = .1$ level test. All other things are equal, in which test would have more power to reject the null hypothesis. Give a brief explanation of why?

Housing Data

Recall the housing data that we've been considering in lecture. We can load the data using the following code:

```r
fileName <- url("https://raw.githubusercontent.com/ysamwang/btry6020_sp22/main/lectureData/estate.csv")
housing_data <- read.csv(fileName)

head(housing_data)
```

```
##   id  price area bed bath  ac garage pool year quality style   lot highway
## 1  1 360000 3032   4    4 yes      2   no 1972  medium     1 22221      no
## 2  2 340000 2058   4    2 yes      2   no 1976  medium     1 22912      no
## 3  3 250000 1780   4    3 yes      2   no 1980  medium     1 21345      no
## 4  4 205500 1638   4    2 yes      2   no 1963  medium     1 17342      no
## 5  5 275500 2196   4    3 yes      2   no 1968  medium     7 21786      no
## 6  6 248000 1966   4    3 yes      5  yes 1972  medium     1 18902      no
```

There are 522 observations with the following variables:

- price: in 2002 dollars
- area: Square footage
- bed: number of bedrooms
- bath: number of bathrooms
- ac: central AC (yes/no)
- garage: number of garage spaces
- pool: yes/no
- year: year of construction
- quality: high/medium/low
- home style: coded 1 through 7
- lot size: sq ft
- highway: near a highway (yes/no)

There is no age data in the table, but we can compute it on our own from the year variable

```r
housing_data$age <- 2002 - housing_data$year
```

Question 8 (3 pts)

Let $\log(price)$ be the dependent variable. Suppose we are interested in the association of $\log(price)$ with the lot size, after conditioning for the area, age, and number of bedrooms. Estimate the linear coefficient of interest and give an interpretation of the point estimate. Form a 95% confidence interval for the coefficient of interest.

Answer

Question 9 (3 pts)

Let $\log(price)$ be the dependent variable. Suppose we are interested in the association of $\log(price)$ with the number of bedrooms, after conditioning for the $\log(area)$, $\log(lot)$, and age. Conduct a hypothesis test with level $\alpha = .05$ for the null hypothesis that bedrooms is not associated with $\log(price)$ after conditioning for $\log(area)$, $\log(lot)$, and age. What is the resulting t statistic? What is the result of the hypothesis test?

Answer

Question 10 (3 pts)

Let $\log(price)$ be the dependent variable. Suppose we are interested in the association of $\log(price)$ with quality of the house, after conditioning for the $\log(area)$, age, and number of bedrooms. Conduct a hypothesis test with level $\alpha = .05$ for the null hypothesis that quality is not associated with $\log(price)$ after conditioning for the area, age, and number of bedrooms. What is the resulting statistic? What is the result of the hypothesis test?

Answer

Question 11 (3 pts)

Describe a hypothesis testing problem in your field where you might need to carefully choose the Type I error rate, α. Describe why you might use something other than the standard .05 when weighing the costs of a Type I error and Type II error.

Answer