

Weekly Quiz 8

The **due date** for this quiz is **Mon 18 Mar 2013 3:30 AM GMT**.

Question 1

Suppose this is the result of 85 hypothesis tests:

	$\beta = 0$	$\beta \neq 0$	CLAIMS TOTALS
Claim $\beta = 0$	50	10	60
Claim $\beta \neq 0$	5	20	25
Hypothesis Totals	55	30	85

What is the (observed) rate of false discoveries? What is the (observed) rate of false positives?

- ☒ False discovery rate = 0.20 False positive rate = 0.09
- ☐ False discovery rate = 0.25 False positive rate = 0.10
- ☐ False discovery rate = 0.09 False positive rate = 0.20
- ☐ False discovery rate = 0.17 False positive rate = 0.33

Question 2

Generate P-values according to the following code:

```
set.seed(3343)
pValues = rep(NA,100)
for(i in 1:100){
  z = rnorm(20)
  x = rnorm(20)
  y = rnorm(20,mean=0.5*x)
  pValues[i] = summary(lm(y ~ x))$coef[2,4]
}
```

How many are significant at the $\alpha = 0.1$ level when controlling the family wise error rate using the methods described in the lectures? When controlling the false discovery rate at the $\alpha = 0.1$ level as described in the lectures?

- ☐ FWER = 61 FDR = 7
- ☒ FWER = 7 FDR = 61
- ☐ FWER = 5 FDR = 32
- ☐ FWER = 3 FDR = 5

Question 3

Suppose I want to generate data from the following model with a simulation:

$$y = b_0 + b_1x + b_2z + e$$

where $b_0=1$, $b_1=2$, $b_2=3$ and x , z , and e are normally distributed. Which one of the following is not a step in the simulation process?

- ☐ Generate x , z , and e using `rnorm()`
- ☐ Generate the fitted values by adding $y_{\text{fit}} = 1 + 2x + 3z$
- ☐ Generate the y -values by adding $y_{\text{fit}}+e$
- ☒ Generate a random sample of values for b_0 , b_1 , and b_2

Question 4

Suppose data are generated from a model:

$$y = b_0 + b_1x + e$$

where $b_0=1$, $b_1=2$ and x and e both have a normal distribution with mean zero and variance one. After the data are created, some data are lost. Use the `lm()` function in base R for model fitting. **Case 1:** Build a simulation where all values of y are observed but higher values of x are likely to be missing. Does the estimate of b_1 change on average? If so how? **Case 2** Build a simulation where all values of x are observed but higher values of y are likely to be missing. Does the estimate of b_1 change on average? If so how?

- ☐ Case 1: b_1 is overestimated Case 2: b_1 is overestimated

- ☐ Case 1: b1 is overestimated Case 2: b1 is estimated correctly
- ☒ Case 1: b1 is estimated correctly Case 2: b1 is underestimated
- ☐ Case 1: b1 is underestimated Case 2: b1 is underestimated

Question 5

Exactly as in the last question, suppose data are generated from a model: $y = b_0 + b_1x + e$ where $b_0=1$, $b_1=2$ and x and e both have a normal distribution with mean zero and variance one. After the data are created, some data are lost. Answer the same questions below, but this time, use the `rlm()` function in the MASS package to fit the linear model instead of the `lm()` function in base R. Case 1: Build a simulation where all values of y are observed but higher values of x are likely to be missing. Does the estimate of b_1 change on average? If so how? Case 2 Build a simulation where all values of x are observed but higher values of y are likely to be missing. Does the estimate of b_1 change on average? If so how?

- ☐ Case 1: b1 is overestimated Case 2: b1 is estimated correctly
- ☐ Case 1: b1 is underestimated Case 2: b1 is underestimated
- ☐ Case 1: b1 is overestimated Case 2: b1 is overestimated
- ☒ Case 1: b1 is estimated correctly Case 2: b1 is underestimated

☒ In accordance with the Honor Code, I certify that my answers here are my own work.

Submit Answers

Save Answers