

Quiz #2

Dustin Leatherman

February 17, 2019

1

In Galileo's data on the motion of falling bodies (Case 10.1.1 on page 272 in The Statistical Sleuth), why is horizontal distance, rather than height, the response variable?

Galileo's experiment measured how horizontal distance during falling was affected by height. Height was a control variable while Horizontal distance was influenced by height.

2

Consider the mammal brain weight data from Section 9.1.2 in The Statistical Sleuth, the model

$$\mu\{\text{lbrain} \mid \text{lbody}, \text{lgest}, \text{llitter}\} = \beta_0 + \beta_1 \text{lbody} + \beta_2 \text{lgest} + \beta_3 \text{llitter}$$

and the hypothesis $H_0 : \beta_2 = \beta_3 = 0$.

(a) Why can this not be tested by the two t-tests reported in the standard output?

The t-test in the output is referring to each variable individually. The alternative hypothesis to the one described is that at least one of those variables are not equal to 0. Since there are multiple variables, they are affecting the other to some degree. The coefficient values change based on the parameters present, even if the parameters are insignificant. The T-Test that is output may give an indication whether or not those parameters are significant but the proper way to test is to use a Sum of Squares F-Test between a model with and without those parameters.

(b) Why can this not be tested by the two t-tests along with an adjustment for multiple comparisons?

Even after applying an adjustment such as the Bonferroni, the values will still be affected by the other as described above so it will not work.

3

In the bat echolocation case study, a possible statement in conclusion of the analysis is this: "It is estimated that the median in-flight energy expenditure for echolocating bats is 1.08 times as large as the median in-flight energy expenditure for non-echolocating bats of similar body size." Referring to Display 10.4 (p.275 in The Statistical Sleuth), explain why including the phrase, "of similar body size" in this statement is suspect. What alternative wording is available?

"of similar body size" is vaguely referring to Body Size which was used in the study. It would be better to say "after controlling for Body Size" instead as it implies that it was a variable that was part of the model and that it has been taken into consideration.

4

A regression of the number of crimes committed in a day on volume of ice cream sales in the same day showed that the coefficient of ice cream sales was positive and significantly differed from zero. Which of the following is the most likely explanation.

(a) The content of ice cream (probably the sugar) encourages people to commit crimes.

- (b) *Successful criminals celebrate by eating ice cream.*
- (c) *A pathological desire for ice cream is triggered in a certain percentage of individuals by certain environmental conditions (such as warm days), and these individuals will stop at nothing to satisfy their craving.*
- (d) Another variable, such as temperature, is associated with both crime and ice cream sales.**
- As much as I want C to be right, D is the correct answer.

5

The following display summarizes results from the regression of the log of the number of butterfly species observed on the log of the size of the reserve and the number of observations, from 16 reserves in the Amazon River Basin.

Table not shown

(a) What is the two sided p-value for the test of whether size of reserve has any effect on number of species, after accounting for the days of observation? What is the one-sided p-value if the alternative is that size has a positive effect? Does this imply that there is no evidence that the median number of species is related to reserve size? The researchers tended to spend more days searching for butterflies in the larger reserves. How might this affect the interpretation?

Effect of Reserve Size on # of species: 0.2443

$H_0: lsize \leq 0$

$H_A: lsize > 0$

```
df = 13
est.mean <- 0.0809
est.se <- 0.1131
t <- (est.mean - 0) / est.se
t
```

```
## [1] 0.7152962
```

```
# probability that a value lower than 0 will be present.
pt(t, df)
```

```
## [1] 0.7564626
```

One sided p-value = 0.2435

There is not enough evidence to suggest that the median number of species is positively impacted by reserve size (p-value = 0.2435). Since the researchers spent more time at larger reserves, it could mean that smaller reserves are under-represented and thus the results are more susceptible to biases for larger reserves.

(b) What is a two-sided p-value for the test that the coefficient of lsize is 1?

```
df = 13
est.mean <- 0.0809
est.se <- 0.1131
t <- (est.mean - 1) / est.se
t
```

```
## [1] -8.126437
```

```
# probability that a value greater than 1 or less than 1  
1 - pt(t, df) * 2
```

```
## [1] 0.9999981
```

(c) What is the 95% confidence interval for the coefficient of lsize?

```
ucl <- est.mean + qt(0.975, df) * est.se  
lcl <- est.mean - qt(0.975, df) * est.se
```

```
ucl
```

```
## [1] 0.3252377
```

```
lcl
```

```
## [1] -0.1634377
```

Lower Confidence Limit = -0.1634 Upper Confidence Limit = 0.3252

(d) What proportion of the variation in log number of species remains unexplained by log size and days of observations?

$1 - .1141 = \mathbf{0.8859}$