Name: Feedback

SM339 · Applied Statistics Spring 2023 · Uhan

# Quiz 4 – 2/8/2023

**Instructions.** You have 15 minutes to complete this quiz. You may use your plebe-issue TI-36X Pro calculator. You may not use any other materials.

Show all your work. To receive full credit, your solutions must be completely correct, sufficiently justified, and easy to follow.

| Problem | Weight | Score |
|---|---|---|
| 1 | 1 | |
| 2a | 1 | |
| 2b | 1 | |
| 2c | 1 | |
| Total | | / 40 |

**Problem 1.** A capacitor was charged with a 9-volt battery and then a voltmeter recorded the voltage as the capacitor was discharged. Measurements were taken every 0.02 seconds. Your data consists of two variables: *Voltage* (in volts) and *Time* (in seconds). You are interested in predicting *Voltage* based on *Time*.

After exploring the data, you decide that applying a log transformation to *Voltage* is appropriate. You fit a simple linear regression model with $log(Voltage)$ as the response variable, and *Time* as the explanatory variable (assume *log* is the natural logarithm). Your fitted model is

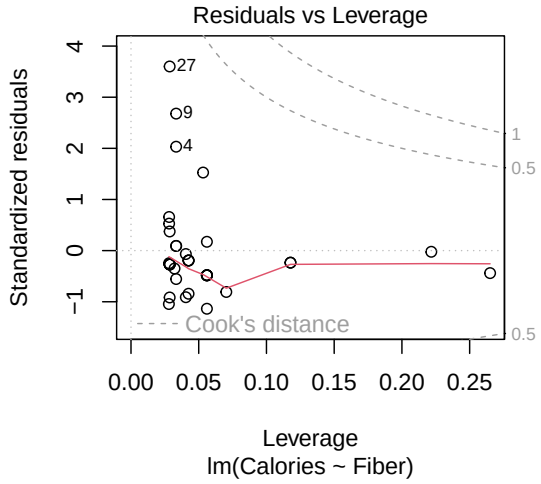
$$\widehat{log(Voltage)} = 2.19 - 2.06 Time$$

Use your fitted model to predict *Voltage* when *Time* = 0.05 seconds.

See Example 1e in Lesson 8 for a similar problem. Note that the problem specifies that *log* is the natural logarithm!

**Problem 2.** You are working with data for 36 breakfast cereals. Your data consists of two variables: *Calories* and *Fiber* (in grams). You are interested in predicting *Calories* based on *Fiber*. You fit a simple linear regression model with *Calories* as the response variable, and *Fiber* as the explanatory variable.

a. Using R, you generate the diagnostic plot to the right. Based on the rules of thumb we covered in class, circle the points that are classified with "very unusual" leverage. Briefly explain your reasoning below.




See Lesson 9 for the rules of thumb for identifying unusual points in simple linear regression. See Example 2 in Lesson 9 for a similar problem.

Below is output from `summary()` for your simple linear regression model.

```
Call:
lm(formula = Calories ~ Fiber, data = Cereal)

Residuals:
    Min      1Q  Median      3Q     Max
-17.363  -7.363  -4.005   1.413  55.801

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 117.3635     3.7216  31.536  < 2e-16 ***
Fiber        -4.3881     0.7358  -5.964  9.6e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15.72 on 34 degrees of freedom
Multiple R-squared:  0.5112,  Adjusted R-squared:  0.4969
F-statistic: 35.56 on 1 and 34 DF,  p-value: 9.603e-07
```

Suppose you want to perform a $t$-test for the slope of your simple linear regression model, with hypotheses

$$H_0 : \beta_1 = 0 \qquad H_A : \beta_1 \neq 0$$

b. Based on the output, do you reject or fail to reject the null hypothesis $H_0$? Briefly explain why. Use a significance level of $\alpha = 0.05$.

See Example 1c in Lesson 10 for a similar problem.

c. Based on your decision, state your conclusion about the slope of your model in the context of the problem. Be brief.

See Example 1c in Lesson 10 for a similar problem.

Be careful: never make a conclusion with absolute certainty! The $t$-test for simple linear regression slope (and other hypothesis tests) only give you evidence that the null or alternative hypotheses are true. As we discussed in class, you can say things like:

- "There is (statistically) significant evidence that..."
- "There is a (statistically) significant relationship between..."
- "$X$ is a (statistically) significant predictor..."
- "The slope is (statistically) signficantly different from zero..."