Version Dr. Seuss   Dr. Dre

REGRESSION PART

Oath: *"I will not discuss the exam contents with anyone on planet Earth until the answer key is posted to Blackboard."*

Sign Name: _____Key_____

The penalty for cheating on this Exam is a grade of 0% for Math 127 Exam 2.

Student Instructions

1. This test is graded out of 50 points and counts for 10% of your Math 127 grade. Points are in parentheses for each question.

2. You can use a calculator, but you cannot use your phone. You can use the calculator on the computers if you wish.

3. You will need to use www.statcrunch.com. This is the only permitted webpage.

4. You are permitted to use one 8.5" by 11" sheet of notes, front and back. You will submit it with your test.

   You may not use the pink sheet or copies of the pink sheet.

   You must produce (handwritten or typed up) your own sheet of notes.

   You may not use copies or scans of any instructor-created Math 127 content or answer keys.

5. Show work or points will be deducted. If you only report an answer and it is wrong, you will receive no credit.

4. Use the **"Marvel vs. DC at the Box Office"** dataset for #4.

"Foreign" is the Box Office Revenue reported in millions of U.S. dollars.

Example: *"Avengers"* made $895.237 million = $895,237,000 at the box office in foreign countries.

"Domestic" is the box office revenue reported in millions of U.S. dollars for movie theatres in the U.S.A.

4a. (1) Cook up the linear regression equation if the explanatory variable is *"Domestic"* and the response variable is *"Foreign"*.

$$\widehat{Foreign} = -15.78 + 1.334 (Domestic)$$

4b. (3) Interpret the value of R^2 with two sentences in context. You need both parts for full credit.

78.87% of the Variation in Foreign Revenue is explained by knowing Domestic Revenue. 21.13% is still unexplained.

5. Use the **"Neighborhood"** dataset for this question.

"Lot" is the size of the property, measured in square feet.

"Zillow Value" is how much the website Zillow thinks a home is worth

5a. (1) Give the linear regression equation. We will use *"Lot"* size to predict the *"Zillow Value"*.

$$\widehat{Zillow} = 173,571.97 + 2.15 (Lot)$$

5b. (3) There is a vacant 1-acre lot. How much might Zillow value this house at if a typical neighborhood house is built there? PS: One acre is 43,560 square feet. Technology only answer is OK.

Answer: $\widehat{Zillow} = \$267,247$

6. Use the **"Roller Coasters"** dataset on this one.

"Height" is the ground-to-top-of-first-hill measurement, in feet.

"Drop" is the top-of-first-hill-to-bottom-of-first-hill measurement, in feet.

6a. (1) Give the equation of the regression line. We will predict the *"Drop"* based on the *"Height"*.

$$\widehat{Drop} = 3.46 + 0.934 (Height)$$

6b. (3) *"Orient Express"* at Worlds of Fun has a residual of 2.25 feet. Does this coaster have an unusually large *"Drop"* for its *"Height"*? Yes / no and explain clearly for full credit.

No! The Studentized Residual is 0.107. (Hardly unusual!)

4

"ZZZ Retired

7. Use our **"Calendar Year 2017 Food Bank"** dataset for this problem.

X Y

Run the linear regression, using **"Carb Grams"** to predict the **"Calories"**.

7a. (2) Linear Equation: $\widehat{Calories} = \cancel{19.46 + 5.22}(Carbs)$

~~15.39 + 5.17 (Carbs)~~

7b. (2) Using the correct units for both *x* and *y*, interpret the slope with a sentence in context:

For each extra gram of Carbs, we expect an extra ~~5.22~~ 5.17 Calories.

7c. (2) The *y*-intercept has meaning in the context of the problem. Interpret this point in context.

For a food with x = 0 g. of Carbs, we expect Calories to be ~~19.46~~. 15.39

7d. (2) The point on the red line ~~(33, 191.78)~~ (33, 186.0) has real meaning in the context of this problem. Interpret in context.

A food with 33 Carb. g. is expected to have ~~191.78~~ 186.0 Calories.

7e. (2) One product in the whole dataset is missing its **"Carb Grams"**. Find that product and predict the **"Carb Grams"**. Show calculation here:

~~Row 149 Chicken Noodle Soup~~

~~Ca~~

oops

−44.30

7f. (2) The **"Mandarin Orange Segments"** in row 48 has a residual of ~~-49.56~~ 44.30. Interpret the value of this residual with a sentence in context:

This product has ~~49.56~~ 44.30 fewer Calories than we would expect for its x = 23 Carb grams.

5

7g. (2) After the survey was closed and the dataset was posted, a student brought in a can of "*StarKist Chunk Light*" tuna, which had "*Carb Grams*" = 0 and "*Calories*" = 45.

Show the calculation to arrive at the residual for this product. Do not add this data point and recalculate your regression. Just use the regression equation as is:

$$\text{Residual} = \text{Actual } y - \text{Predicted } y$$
$$= \text{Actual Cal} - \text{Predicted Cal}$$
$$= 45 - \cancel{19.46} = \cancel{25.54}$$

(red) 15.39   29.61

7h. (2) Interpret the value of S_e with a sentence in context: ___ On average, our Predicted Calories are off by ~~42.87~~ when using x = Carb Grams

(red) 41.05

7i. (2) Interpret the value of R^2 with a sentence in context: ___ ~~74.10%~~ (red 70.63%) of the Variation in Calories is explained by x = Carbs, and ~~25.90%~~ (red 29.37%) is still unexplained.

7j. (2) Describe with bullet points the relationship between "*Carb Grams*" and "*Calories*".

Form: Linear   Direction: Positive

Strength: r = ~~0.861~~ (red 0.8404) (Strong)

Outliers → Refer to below

7k. (1) How many products have large positive residuals? ~~10~~ (red 34)

7l. (1) How many products have large negative residuals? ~~1~~ (red 3)

7m. (1) How many products are outliers (high or low, give the total) for "*Carb Grams*"? (red) ✗ 27

7n. (1) How many products are outliers (high or low, give the total) for "*Calories*"? (red) ✗ 19

7o. (1) How many products have large Cook's Distances? (red) ✗ 41   4/180 = ~~0.02222~~   640   0.0625

7p. (1) Why is it unwise to predict "*Calories*" for "*Carb Grams*" = 99? Give the statistical reason. ___ This x-value is extrapolation (outside the scope of our data).