Publish | Preview | Edit

This quiz is unpublished

Only teachers can see the quiz until it is published.

Homework 3

This is a preview of the draft version of the quiz

After you have completed the reading assignments of Chapters 5 and 6 from *The Basic Practice of Statistics* (8th edition), please complete the following problems from the textbook and then answer the questions on Canvas that correspond to these textbook questions. NOTE: Some of the problems are located in the text of the chapter, and some are in a section of practice problems that follow the chapter; the last question concerns your DAP.

Chapter 5

- 5.23
- 5.25
- 5.26

Quiz Type	Graded Quiz
Points	100
Assignment Group	Homework
Shuffle Answers	No
Time Limit	No Time Limit

Multiple Attempts	No
View Responses	Always
Show Correct Answers	After Oct 2 at 11:59pm
One Question at a Time	No

Due	For	Available from	Until
Oct 2	Everyone	-	Dec 18 at 11:59pm

Preview

Correct answers will be available on Oct 2 at 11:59pm.

Score for this quiz: **100** out of 100

Submitted Oct 1 at 1:55am

This attempt took 405 minutes.

Question 1

2 / 2 pts

5.20) Figure 5.10 is a scatterplot of the price of a hot dog against the price of beer (per ounce) at 24 major-league ballparks in 2015.13 The line is the least-squares regression line for predicting the price of a hot dog from the price of beer. If another ballpark charges 0.60 dollars per ounce for beer, you predict the price of a hot dog to be close to...

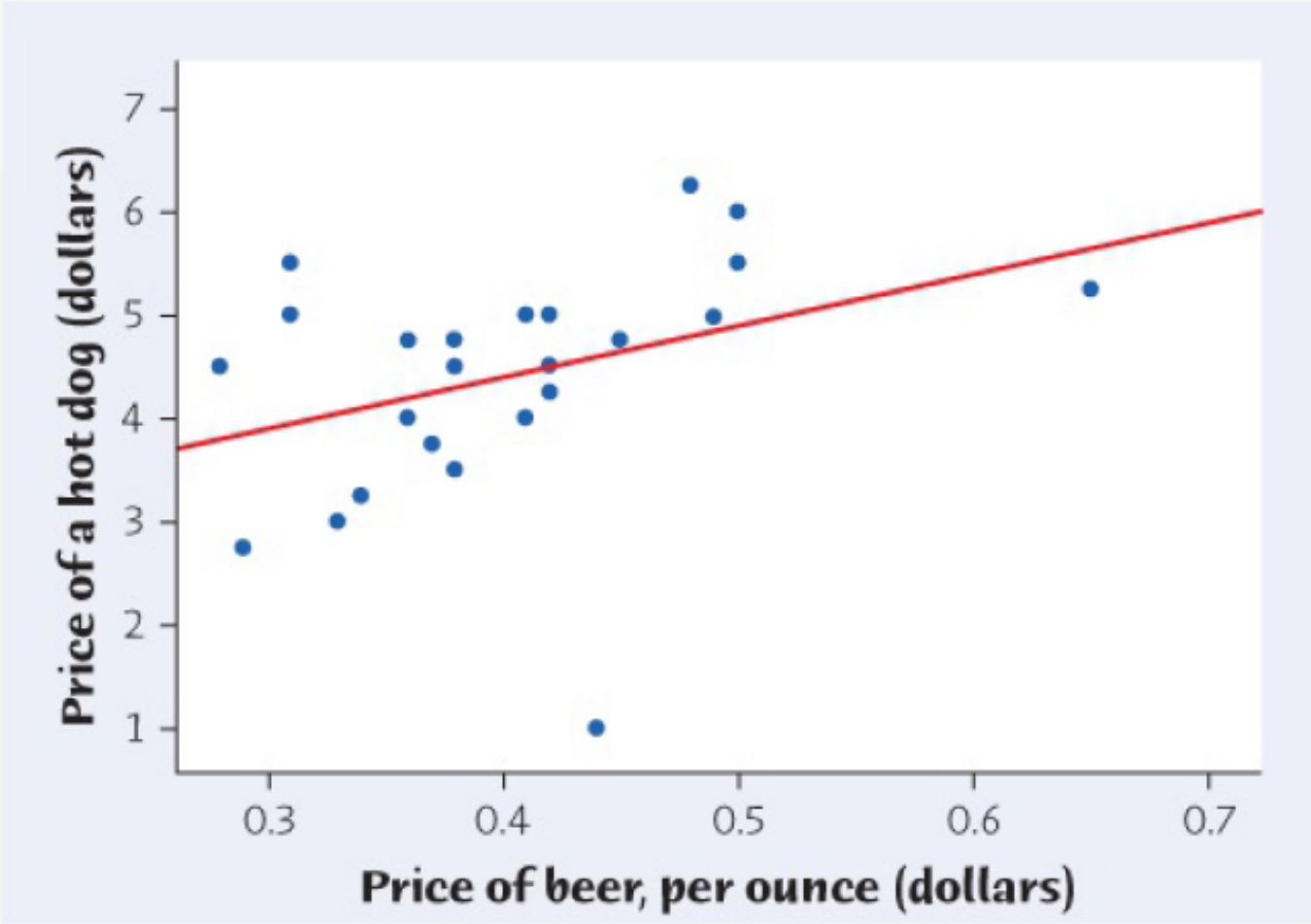

Figure 5.10
Moore/Notz/Fligner, *The Basic Practice of Statistics*, 8e,
© 2018 W.H. Freeman and Company

- $3.75
- $5.50

- $7.00

Question 2

2 / 2 pts

5.21) The slope of the line in Figure 5.10 is closest to... (hint: pay very close attention to the units of the X-axis!)

- -2.4
- 0.2
- 5.0

Question 3

7 / 7 pts

5.4 a), part I. Regression slopes.

Exercises 4.2 and 4.10 discuss a study in which scientists examined data on mean sea surface temperatures (in degrees Celsius) and mean coral growth (in centimeters per year)

over a several-year period at locations in the Gulf of Mexico and the Caribbean. Here are the data for the Gulf of Mexico:

Sea surface temperature	26.7	26.6	26.6	26.5	26.3	26.1
Growth	0.85	0.85	0.79	0.86	0.89	0.92

Using Stata or hand calculation, find the regression ***SLOPE*** for this set of data. Assume that sea surface is the predictor and growth is the outcome. I am accepting a range of answers between the hand calculation answer and the Stata answer. Round your final answer to two decimal places.

For the sake of not "double-penalizing" wrong answers, I am going to give you the exact values (rounded to the hundredth) *for the means and standard deviations of these two variables from the last time we used them, as well as* r. ***Use these exactly as they are if you do a hand calculation***.

Sea surface mean: 26.47

Sea surface *s*: 0.23

Coral growth mean: 0.86

Coral growth *s*: 0.04.

r = -0.81

-0.14

Question 4

7 / 7 pts

5.4 a), part II. Regression intercepts.

Exercises 4.2 and 4.10 discuss a study in which scientists examined data on mean sea surface temperatures (in degrees Celsius) and mean coral growth (in centimeters per year) over a several-year period at locations in the Gulf of Mexico and the Caribbean. Here are the data for the Gulf of Mexico:

Sea surface temperature	26.7	26.6	26.6	26.5	26.3	26.1
Growth	0.85	0.85	0.79	0.86	0.89	0.92

Using Stata or hand calculation, find the regression ***intercept*** for this set of data. Assume that sea surface is the predictor and growth is the outcome. I am accepting a range of answers between the hand calculation answer and the Stata answer. Round your final answer to two decimal places.

For the sake of not "double-penalizing" wrong answers, I am going to give you the exact values (rounded to the hundredth) *for the means and standard deviations of these two variables from the last time we used them, as well as* r. ***Use these exactly as they are if you do a hand calculation***.

Sea surface mean: 26.47

Sea surface *s*: 0.23

Coral growth mean: 0.86

Coral growth *s*: 0.04.

r = -0.81

4.57

Question 5

10 / 10 pts

5.4 c), part I. Regression interpretation. ***Note that I ask about the entire regression equation, going a bit beyond the book here.***

The equation of the line we got above says that for each degree change (Celsius), we expect a change in mean growth of coral (cm/yr) of roughly -0.16 (pick the closest answer to what you got if you worked by hand).

The intercept represents the value that we'd expect if the value of sea-surface temperature were 0 , whether that is realistic or not.

Our linear model of coral growth isn't too shabby here: about 66 percent of variation in the outcome is explained by variation in the predictor.

Answer 1:

-0.16

Answer 2:

0

Answer 3:

66

Question 6

4 / 4 pts

5.9) a) and b).

Exercise 4.8 (page 109) gives data on the fuel consumption *Y* of a car at various speeds *X*. Fuel consumption is measured in mpg, and speed is measured in miles per hour. Software tells us that the equation of the least-squares regression line is ....

Y-hat = 55.3286 - 0.02286*X*.

Using Stata or graph paper, obtain at least a rough scatterplot of the data (this will be slightly tedious by hand but should take only about two or three minutes).

Although samples can be misleading, using this sample as a rough guess, how appropriate is the model in this case?

- Inappropriate: Y has a clear *linear* relationship to X.

- Appropriate: Y has a clear *linear* relationship to X.
- Appropriate: Y has a clear *NON-linear* relationship to X.
- (•) Inappropriate: Y has a clear *NON-linear* relationship to X.

Question 7

6 / 6 pts

5.9d) Now, make a plot of the residuals against the values of x. Again, you can do this by hand or in Stata. To do this by hand, calculate the predicted values for all 10 observed *X*-values, and then take the difference from the actual Y_is to those predictions (AKA the Ŷs). Then, plot the coordinate pairs formed by the Xs and the residuals (Y_i - Ŷ).

To get the residuals in Stata, first run a regression before telling Stata to store results from that command in a new variable called whatever you like. In other words, the following four commands will get you what you want... (make sure to run "predict" right after "reg"; it, like return-values, is a post-analysis command that has to follow the analysis it refers to, much like when you used "ANS" on your calculators in high school).

```
reg Y X

predict uhat, residuals

corr uhat X

scatter uhat X  || lfit uhat X
```

However you make it, this scatterplot of residuals vs. predictor shows exactly zero linear relationship, by design (we actually derive the OLS estimators by this assumption).

But, that's not the end of the story: there is evidence of a nonlinear pattern between the residuals and the predictor, which might make linear regression inappropriate .

Answer 1:

exactly zero

Answer 2:

a nonlinear

Answer 3:

inappropriate

Question 8

6 / 6 pts

5.34(a) The mean height of American women in their twenties is about 64.3 inches, and the standard deviation is about 2.7 inches. The mean height of men the same age is about 69.9 inches, with standard deviation about 3.1 inches. Suppose that the correlation between the heights of husbands and wives is about $r = 0.5$. Please round to two digits past the decimal.

PART 1

What is the slope of the regression line of the husband's height on the wife's height in young couples?

0.57

Question 9

6 / 6 pts

5.34(a) The mean height of American women in their twenties is about 64.3 inches, and the standard deviation is about 2.7 inches. The mean height of men the same age is about 69.9 inches, with standard deviation about 3.1 inches. Suppose that the correlation between the heights of husbands and wives is about $r = 0.5$. Please round to two digits past the decimal.

PART 2

What is the intercept of the regression line of the husband's height on the wife's height in young couples?

33.25

Question 10

6 / 6 pts

For this question, Stata will be by far the most efficient tool. Ask for help loading the data if you have difficulty (or hand-enter them; but, I would not take this correlation coefficient by hand).

5.43b)

People with diabetes must manage their blood sugar levels carefully. They measure their fasting plasma glucose (FPG) several times a day with a glucose meter. Another measurement, made at regular medical checkups, is called HbA. This is roughly the percent of red blood cells that have a glucose molecule attached. It measures average exposure to glucose over a period of several months. Table 5.2 gives data on both HbA and FPG for 18 diabetics five months after they had completed a diabetes education class.

First, put up a scatterplot and find the correlation coefficient. Then, find the correlation 1) overall, 2) without subject 15, and then 3) without subject 18.

The Stata .dta file has a variable called "obs" which indicates a subject's number. To omit (without deleting), say, subject 15 from a correlation coefficient, you could run the following conditional command (then, adjust as needed for subject 18).

```
corr fpg hba if obs != 15
```

When you omit subject 15, the correlation increases . When you omit subject 18, the correlation decreases .

You may want to verify this with a scatter plot. Rather than run multiple scatterplots and regression lines with and without these subjects, when you have a small data-set, simply

labeling the observations with their observation number is not a bad idea. Try...

```
scatter fpg hba, mlabel(obs)
```

...and see if the scatterplot fits your intuition.

Answer 1:

increases

Answer 2:

decreases

Question 11

7 / 7 pts

Use this Table for Questions 6.12 and 6.14

Use Social Networking on Phone	Yes	No
Age 18-29	228	112
Age 30-49	281	281
Age 50-64	106	481
Age 65+	21	408

6.12) What percent of individuals who use social networking on their phone are age 18-29?

- a) about 12%
- b) about 36%
- c) about 67%

Question 12

7 / 7 pts

6.14) What percent of individuals age 18-29 use social networking on their phone?

- a) about 12%
- b) about 36%
- c) about 67%

Question 13

12 / 12 pts

For the following question, please refer to this table. ***For the sake of simplicity, please just round your answer to the nearest integer and input it as an integer with nothing else***, e.g. "9", "10", etc.

TABLE 6.2 Marital status and salary level (thousands of men)

Income	Marital Status				Total
	Single (Never Married)	Married	Divorced	Widowed	
No income	6,013	2,016	702	112	8,843
$1–$49,999	24,381	33,867	6,830	2,375	67,453
$50,000–$99,999	4,718	19,100	2,415	533	26,766
$100,000 and over	1,478	11,262	998	248	13,986
Total	36,590	66,245	10,945	3,268	117,048

6.20) Give (in percentages) the two marginal distributions, for marital status and for income. Enter your answers in the tables below. ***For the sake of simplicity, please just round your answer to the nearest integer and input it as an integer with nothing else***, e.g. "9", "10", etc.

Marginal Distribution of Marital Status

Single	Married	Divorced	Widowed

31 %	57 %	9 %	3 %

Marginal Distribution of Income

No income	$1-49,999	$50,000-99,999	$100,000 and over
8 %	58 %	23 %	12 %

Answer 1:

31

Answer 2:

57

Answer 3:

9

Answer 4:

3

Answer 5:

8

Answer 6:

58

Answer 7:

23

Answer 8:

12

Question 14

7 / 7 pts

Refer to the table above.

6.22) Give (in percents) the conditional distribution of income level among single men. ***For the sake of simplicity, please just round your answer to the nearest integer and input it as an integer with nothing else***, e.g. "9", "10", etc.

Income	None	$1-49,999	$50,000-99,999	$100,000 +
Percent of single men	16 %	67 %	13 %	4 %

Answer 1:

16

Answer 2:

67

Answer 3:

13

Answer 4:

4

Question 15

7 / 7 pts

6.28) "It is right to use animals for medical testing if it might save human lives." The General Social Survey asked 1152 adults to react to this statement. Here is the two-way table of their responses:

Response	Male	Female
Strongly agree	76	59
Agree	270	247
Neither agree nor disagree	87	139
Disagree	61	123

Strongly disagree	22	68

How do the distributions of opinion differ between men and women? Answer the following questions based on your work.

1) What percent of MEN *strongly agree* that animal testing is justified if it might save human lives? [Select]

2) What percent of WOMEN *disagree or strongly disagree* (ie. combine these categories) that animal testing is justified? [Select]

3) Which group is more likely to *agree or strongly agree* that animal testing is justified if it might save human lives? [Select]

Answer 1:

14.7 percent

Answer 2:

30.0 percent

Answer 3:

Men

Question 16

4 / 4 pts

This question is not in the book. It deals with a common problem interpreting two-way tables.

Based on the table below (drawn from the CPS 2019), please answer the following true/false questions.

Agriculture	Central city		
	no	yes	Total
no	97.49	99.55	98.02
yes	2.51	0.45	1.98
Total	100.00	100.00	100.00

Assuming that these sample data give a good guide to the population...

If you live outside of a central city, your probability of working in agriculture is low, say, much less than 10 percent: True

However, if you live outside of a central city, your conditional probability of working in agriculture is still more than double that of a city-dweller: True

---

Answer 1:

True

Answer 2:

True

Quiz Score: **100** out of 100