**STAT 216 Exam 1 – Spring 2022**

**PRACTICE EXAM**

Formula sheet:

Golden Ticket:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Scenario** | **One Categorical Response** | **Two Categorical Variables** | **One Quantitative Response** | **Two Quantitative Variables** | **Quant. Response and Categ. Explanatory** |
| Type of plot | Bar plot | Segmented bar plot,  Mosaic plot | Dot plot, Histogram, Boxplot | Scatterplot | Side-by-side boxplots, Stacked histograms or dotplots |
| Summary measure | Proportion | Difference in proportions | Mean | Slope or correlation | Difference in means |
| Parameter notation |  |  |  | Slope:  Correlation: |  |
| Statistic notation |  |  |  | Slope:  Correlation: |  |

Formulas

1. In 2015, Alcohol and Drug Services of Gallatin County randomly sampled 789 DUI offenders (181 current MSU students, the rest were non-MSU students).  The average age of those sampled was 29.3 years. Eighty-four of the MSU student DUI offenders were under 21 years old and the average age of MSU student DUI offenders was 21.7 years.  Among non-MSU student DUI offenders, 80 were under age 21 and the average age was 31.7 years.  Researchers would like to use these data to determine if whether a DUI offender is an MSU student can predict whether the DUI offender is under the age of 21 among all DUI offenders in Gallatin County.
2. [2 pts] Identify the observational units.
3. [2 pts] Is the value 31.7 a statistic or a parameter? Give the proper notation.

Circle one: statistic parameter

Notation:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. [4 pts] Using the information above, completely fill in the two-way table summarizing these data.

|  |  |  |  |
| --- | --- | --- | --- |
|  | MSU Student | Non-MSU Student | Total |
| Underage |  |  |  |
| Not Underage |  |  |  |
| Total |  |  |  |

1. [3 pts] Calculate the proportion of MSU student DUI offenders who are under age 21. What is the appropriate notation for this value?

Work:

Answer: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Notation:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. [3 pts] Calculate the proportion of non-MSU student DUI offenders who are under age 21. What is the appropriate notation for this value?

Work:

Answer: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Notation:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A screenshot of a cell phone

Description automatically generated

1. [3 pts] Based on the segmented bar chart above or your answers to parts (d) and (e), is there an association between whether a DUI offender is an MSU student and whether they are underage? Explain using the plot.

Is there an association? (Circle one) Yes No

Explain:

1. [3 pts] Gallatin County also recorded whether each DUI offender was a Montana resident. The Alcohol and Drug Services task force found that DUI offenders who were from Montana were more likely to be underage than non-Montana residents. What other information would need to be true in order for ‘whether the DUI offender is a Montana resident’ to meet the definition of a confounding variable? Select one.

* DUI offenders who were from Montana are more likely to be MSU students than non-Montana residents.
* DUI offenders who were from Montana are less likely to be MSU students than non-Montana residents.
* DUI offenders who were from Montana are just as likely to be MSU students than non-Montana residents.
* None of the above, because confounding variables would not occur in this study.

1. It has been well documented that women tend to prefer warmer temperatures compared to men (this phenomenon has been widely referred to as “the battle of the thermostat” in popular culture). However, researchers are interested in exploring whether there is a relationship between gender and cognitive performance as it relates to temperature. The 542 participants of this study were university students in Berlin, Germany, recruited using ORSEE (the Online Recruiting System for Economic Experiments), and randomly selected from the ORSEE subject pool. Gender was self-reported. Among the tests performed, participants were given a five minute, 50-question math task, where each question consisted of adding two 5-digit numbers together without the use of a calculator.  Participants were divided into 24 groups ranging from 23-25 participants in each and room temperatures (ranging from 16.19°C and 32.57°C) were randomly assigned to each group. The researchers would like to know if there is an association between room temperature and performance on the math assessment.
2. [3 pts] Identify the type of study design. Explain your choice in the context of the problem.

Circle one: Randomized Experiment Observational Study

Explanation:

The linear model output for the relationship between room temperature and performance on the math assessment from R is below.

Estimate Std. Error t value Pr(>|t|)

(Intercept) 10.26375470 1.07995837 9.5038429 6.561708e-20

temp 0.01982745 0.04398715 0.4507556 6.523467e-01

1. [3 pts] Use the linear model output above to write the least squares line in the context of the problem.
2. [3 pts] Interpret the value of slope in context of the problem. Select one.
   * + The predicted room temperature will increase by 10.264°C for every 1 additional math problem answered correctly.
     + The predicted number of correct math problems will increase by 10.264 problems for every increase in room temperature by 0.020°C.
     + For every increase in room temperature of 1°C the predicted number of correct math problems will increase by 0.020 problems.
     + For every 1 additional math problem answered correctly, the predicted temperature will increase by 0.020°C.
3. [3 pts] Using the least squares line in part b, predict the number of correct math problems for a participant in a room with a temperature of 19.10°C. *Be sure to include units on your answer. If you did not find an answer to part b, use .*

Work:

Answer: \_\_\_\_\_\_\_\_\_\_\_\_

Units: \_\_\_\_\_\_\_\_\_\_\_\_\_

1. [2 pts] Calculate the residual for a participant in a room with a temperature of 19.10°C and who had 15 correct math problems. *Be sure to include units on your answer.*

Work:

Answer: \_\_\_\_\_\_\_\_\_\_\_

Units:\_\_\_\_\_\_\_\_\_\_\_

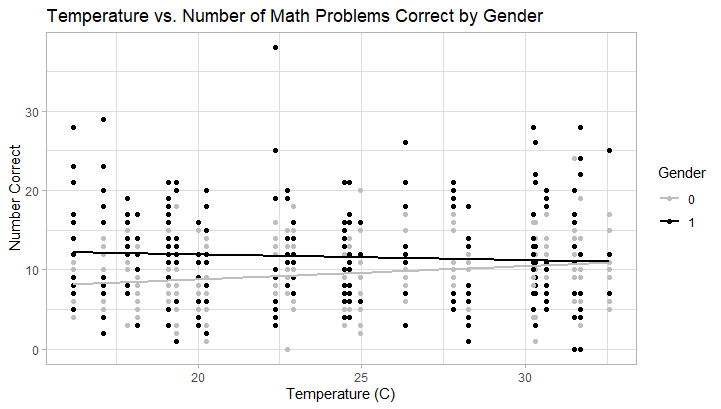
1. [3 pts] The value of the coefficient of determination was calculated to be 0.04%. Interpret this value in the context of the problem.
2. [3 pts] Using the fact that the coefficient of determination was calculated to be 0.04%, calculate the value of the correlation coefficient. What is the appropriate notation for this value?

Work:

Answer: \_\_\_\_\_\_\_\_\_\_\_

Notation: \_\_\_\_\_\_\_\_\_\_

Consider the scatterplot below.



1. [3 pts] Do gender (1 = male, 0 = nonmale) and temperature appear to interact to explain the number of correct math problems? Select one.
   * + Yes, since the slope of the line for male participants is different than the slope of the line for nonmale participants.
     + No, since the scatter of the gray points and the scatter of the black points are similar.
     + Yes, since the mean number of correct math problems for male participants is greater than the mean number of correct math problems for nonmale participants.
     + No, since at 32.5°C, the predicted number of correct math problems is the same for male and nonmale participants.
2. Data were collected by the Planet Money podcast to test a theory about crowd-sourcing.  Planet Money had a post on their website with pictures of Penelope, the cow, and asked people to guess how much she weighed (in pounds).  Over 17,000 people gave responses.  Penelope’s actual weight was 1,355 pounds.

The following summary statistics and plots were created with the data in R.

> favstats(weight)

 min       Q1   median     Q3    max       mean         sd       n    1   907.5     1245   1542   14555   1287.083   622.2028   17184

1. [2 pts] What is the sample size? Give proper notation.

Sample size:\_\_\_\_\_\_\_\_\_\_\_\_

Notation:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A close up of a mans face

Description automatically generatedA screenshot of a cell phone

Description automatically generated

1. [3 pts] Which is the better measure of center for this data? Select one.
   * + The median, as the shape of the distribution is strongly skewed.
     + The mean, as the shape of the distribution is symmetric
     + The mean, as the shape of the distribution is strongly skewed.
     + The median, as the shape of the distribution is symmetric.
2. [3 pts] Interpret the standard deviation in context of the problem. Select one.
   * + Each person’s guess for the weight of Penelope is 622.2028 lbs from the mean weight guess of 1287.083 lbs, on average.
     + Fifty percent of people’s weight guesses for Penelope are between 664.88 and 1909.29 lbs.
     + Fifty percent of people’s weight guesses for Penelope are between 907.5 and 1542 lbs.
     + The average person guesses a weight for Penelope within 622.2028 lbs of 1287.083 lbs.
3. [3 pts] Interpret the value for the third quartile in context of the problem by filling in the blanks:

Approximately \_\_\_\_\_\_\_\_\_ % of people’s weight guesses for Penelope are equal to (value of Q3)

\_\_\_\_\_\_ and (lower/higher) \_\_\_\_\_\_\_\_\_\_.

1. [3 pts] Which types of sampling bias may be present in this study? Select **all** that are true, or if no bias is present, select option (D) No bias.

* Selection bias
* Non-response bias
* Response bias
* No bias

1. Researchers in Southern England collected data on grassland butterflies. They were interested in whether movement patterns varied across species and between male and female butterflies. Researchers observed 164 butterflies over the three-year length of this study, of which 28 were female and 136 were male. These 164 butterflies are considered to be representative of all grassland butterflies. The butterfly movements were observed by measuring how far they flew (in meters) from one landing site to the next, called step distance. This was done by placing a flag at each landing site and measuring the distance between the flags using a mapping software. Is there a difference between how far male and female butterflies travel, on average, between landing sites?
   1. [4 pts] Fill in each blank with one of the options in paratheses to best describe the variables collected.

Step distance is the (explanatory/response) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and it is

(categorical/quantitative) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

Sex is the (explanatory/response) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and it is (categorical/quantitative) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

* 1. [3 pts] Which type of plot is most appropriate to display the relationship between step distance and sex of the butterfly? Select one.
* Segmented bar plot
* Scatterplot
* Side-by-side box plot
* Histogram
  1. [3 pts] Assuming a statistical difference in step distance is found between the males and females in the sample, what is the scope of inference for this study? Select one.
* Sex causes a difference in average step distance for all grassland butterflies.
* Sex is associated with a difference in average step distance for grassland butterflies similar to those sampled.
* Sex causes a difference in average step distance for grassland butterflies similar to those sampled.
* Sex is associated with a difference in average step distance for all grassland butterflies.