

Midterm 2 Question Bank

Stat 218

Your exam questions will be randomly selected from this bank of questions. There **will not** be a solution key posted. It is your responsibility to discuss your ideas with your group members and / or with Dr. Theobald during office hours prior to the exam.

Golden Ticket

Scenario	One Quantitative Response	Two Quantitative Variables	Quant. Response and Categ. Explanatory
Type of plot	Dot plot, Histogram, Boxplot	Scatterplot	Faceted Histograms, Side-by-side Boxplots
Summary measure	Mean	Slope or Correlation	Difference in Means
Parameter notation	μ	Slope: β_1 ; Correlation: ρ	$\mu_1 - \mu_2$
Statistic notation	\bar{x}	Slope: b_1 ; Correlation: r	$\bar{x}_1 - \bar{x}_2$

Provided Formulas

$$IQR = Q3 - Q1$$

1.5 IQR Rule: above $Q3 + (1.5 \times IQR)$ or below $Q1 - (1.5 \times IQR)$

$$\hat{y} = b_0 + b_1 \times x$$

$$Residual = y - \hat{y}$$

t-based confidence interval: point estimate $\pm t_{df}^* \times SE$

$$SE(\mu) = \frac{\sigma}{\sqrt{n}}$$

$$SE(\mu_1 - \mu_2) = \sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}$$

Q1[10 points]

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.

The researchers were interested in investigating if there was a difference between how far male and female butterflies travel, on average, between landing sites.

(a)[4 pts] Fill in each blank with one of the options in parentheses 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) _____.

(b)[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 Boxplot
- Histogram

(c) [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.

Q2[8 points] Is insomnia related to education level? Researchers at the University of Memphis, in Tennessee investigated this question in the Journal of Abnormal Psychology (February, 2005). Eight hundred (800) adults living in Tennessee were selected to participate in the study, using a random-digit telephone dialing procedure. Responses were collected from a total of 575 adults. Two variables measured were level of education and insomnia status (normal sleeper or chronic insomniac). The researchers discovered that the lower the level of education, the more likely the person was to have chronic insomnia.

(a)[2 points] What are the observational units in this study?

(b)[3 points] Fill in the blanks with one of the options provided in parentheses to identify and explain the study design.

This is a(n) (observational study / randomized experiment) _____
because (level of education / insomnia status) _____ (was /
was not) _____ randomly (assigned / sampled) \underline{ }
.

(c)[3 points] Which types of sampling bias may be present in this study? Select **all** that are present, or if there is no bias present, select “No bias.”

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

Q3 In 2004, the state of North Carolina released to the public a large dataset containing information on births recorded in this state. This data set has been of interest to medical researchers who are studying the relation between habits and practices of expectant mothers and the birth of their children. This analysis will focus on a random sample of 1,000 observations from the published dataset.

Our analysis will focus on the relationship between a baby's birth weight and the mother's smoking status.

```
## # A tibble: 1,000 x 13
##   fage  mage mature    weeks premie visits gained weight lowbirthweight sex
##   <int> <dbl> <chr>    <dbl> <chr>    <dbl> <dbl> <dbl> <chr>    <chr>
## 1    34    34 younger m~    37 full ~    14    28    6.96 not low    male
## 2    36    31 younger m~    41 full ~    12    41    8.86 not low    fema~
## 3    37    36 mature mom    37 full ~    10    28    7.51 not low    fema~
## 4    NA    16 younger m~    38 full ~    NA    29    6.19 not low    male
## 5    32    31 younger m~    36 premie    12    48    6.75 not low    fema~
## 6    32    26 younger m~    39 full ~    14    45    6.69 not low    fema~
## 7    37    36 mature mom    36 premie    10    20    6.13 not low    fema~
## 8    29    24 younger m~    40 full ~    13    65    6.74 not low    male
## 9    30    32 younger m~    39 full ~    15    25    8.94 not low    fema~
## 10   29    26 younger m~    39 full ~    11    22    9.12 not low    male
## # ... with 990 more rows, and 3 more variables: habit <chr>, marital <chr>,
## #   whitemom <chr>
```

(a)[2 points] Using the preview of the dataset above, what is the observational unit of this study?

(b)[2 points] Is this an observational study or an experiment? Circle one.

Observational Study

Experiment

(c)[3 points] Researchers are interested in the difference in the mean baby birth weight between mothers who smoke and mothers who do not smoke. Using the table below, report the observed statistic for this comparison. Indicate in your answer what notation should be used for this statistic.

Answer:

Notation:

```
##      habit  min      Q1 median      Q3      max      mean      sd  n missing
## 1 nonsmoker 1.19 6.5700   7.35 8.060 10.42 7.269873 1.232846 867      0
## 2 smoker   0.75 5.9525   7.03 7.805  9.25 6.677193 1.596645 114      0
```

(d)