

Directions: Type or clearly handwrite your solutions to each of the following exercises. Partial credit cannot be given unless all work is shown. You may work in groups provided that each person takes responsibility for understanding and writing out the solutions. Additionally, you must give proper credit to your collaborators by providing their names on the line below (if you worked alone, write "No Collaborators"):

COLLABORATORS:

1. [+10]: Some researchers were interested in studying the effects of different fertilizer amount (Low Nitrogen and High Nitrogen) and different genotypes (one energy line and one grain line) of sorghum on biomass. For each genotype, six pots of one-week-old seedlings were available, and each pot held one seedling. For each genotype, the researchers randomly assigned three pots to high nitrogen treatment (H) and the remaining three pots to low nitrogen (L) treatment. After 2 weeks, the fresh weight for each seedling was measured. In total, there are 12 observations.

1. Identify the experimental units.
2. Identify the observational units.
3. Identify the treatments.
4. Identify the response variable.
5. Does the experiment utilize replication? Answer yes/no and provide a brief justification.
6. Does the experiment utilize blocking? Answer yes/no and provide a brief justification.
7. Does the experiment utilize randomization? Answer yes/no and provide a brief justification.

2. [+10]: A statistics teacher wanted to determine if having business students use clickers to respond to questions posed in a business statistics class would improve student learning. The teacher decided to have students use clickers in one class of introductory business statistics 226. She did not have students use clickers in a second class of business statistics 226 that she taught during the same semester. She tossed a coin to select the class to use the clickers. She used the same book and the same lectures in both classes and gave the same assignments and same exams to both classes. There were 90 students in each class. At the end of the semester, she compared the final exam scores for the students in the class that used clickers to the final exam scores for the students in the class that did not use clickers.

1. Identify the experimental units.
2. Identify the observational units.
3. Identify the treatments.
4. Identify the response variable.
5. Does the experiment utilize replication? Answer yes/no and provide a brief justification.
6. Does the experiment utilize blocking? Answer yes/no and provide a brief justification.
7. Does the experiment utilize randomization? Answer yes/no and provide a brief justification.

3. [+30]: The file `guinea_pigs.csv` (available on Canvas) contains data on survival times (in days) of guinea pigs that were randomly assigned either to a control group or to a treatment group that received a dose of tubercle bacilli (Doksum, K. (1974), *Annals of Statistics*, pp 267-77).

```
# knitr::opts_chunk$set(echo = F)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

guinea_pigs <- read.csv("C:/Users/samue/Downloads/guinea_pigs.csv")
# summary(guinea_pigs)
controlData <- guinea_pigs %>%
  filter(Treatment == "Control") %>%
  na.omit()
treatmentData <- guinea_pigs %>%
  filter(Treatment == "Bacilli") %>%
  na.omit()

summary(controlData)
```

```
##      Pig      Time      Treatment
## Min.   : 1.00   Min.   : 18.0   Length:64
## 1st Qu.:16.75   1st Qu.:141.8   Class :character
## Median :32.50   Median :316.5   Mode  :character
## Mean   :32.50   Mean   :345.2
## 3rd Qu.:48.25   3rd Qu.:570.8
## Max.   :64.00   Max.   :735.0
```

```
summary(treatmentData)
```

```
##      Pig      Time      Treatment
## Min.   : 65.00   Min.   : 76.0   Length:58
## 1st Qu.: 79.25   1st Qu.:161.0   Class :character
## Median : 93.50   Median :214.5   Mode  :character
## Mean   : 93.50   Mean   :242.5
## 3rd Qu.:107.75   3rd Qu.:306.0
## Max.   :122.00   Max.   :598.0
```

1. Use R to compute the following summary statistics for each treatment group:

Statistic	Control	Bacilli
-----	-----	-----

Median

Q1

Q3

IQR

Sample Mean

Standard Deviation

2. Use R to construct side-by-side box plots of survival times for the two treatment groups and include it with this assignment.
3. Use the box plots from part (b) and the summary statistics from part (a) to describe and compare features of the distributions of survival times for the two treatment groups.
4. There is no function built into R to easily perform the randomization test to determine whether the average survival times of the guinea pigs in the two treatment groups is the same or different. Write your own function to perform the test (there is an example in the optional R lab). Interpret the results by providing

1. the null and alternative hypotheses;
2. observed test statistic;
3. randomization histogram;
4. p-value;
5. interpretation of the test results (stated in the context of the problem).

Total: 50 points **# correct:** %: