Sample Exam

September 21, 2017

- You have XX minutes to complete this exam.

- The exam packet is **9** pages: **1** cover sheet, **7** questions, and **2** pages of R code. Make sure you have everything.

- The exam is out of **0** points.

- A single 8x11 page of formulas and definitions (one sided) is allowed along with a calculator. No worked out problems. This page will be turned in with your exam.

- Show all your work. Partial credit can be awarded.

- All phones must be put away during the entire exam. If your phone is out you will be asked to turn in your exam and leave immediately.

- Use the back pages as scratch paper, more can be made available. All scratch paper is to be submitted with your exam. If you want me to review notes on your scratch paper specifically, tell me so and make sure it's findable.

- Be sure to write your answers as full sentences.

- There is R code handed out along with the exam. Not all code is used in this exam, but all code necessary to perform the calculations on the exam is included. This will not completely substitute for a calculator.

- Each piece of R code has a letter indicator. Use this to indicate what code you used to answer every question when you use it. This is not optional.

- The exam will not start until all materials are put away.

- Bathroom breaks are allowed, but you must leave all materials at your desk, and your exam with me.

Name: _____

Do not turn this page until instructed to do so.

1. Ages from a random sample of 20 dogs was stored in the R vector `dog_age`, and shown on the first page of the Appendix for this exam. Use this information to answer the following questions.

   (a) Report the following values.

Mean:	Variance:	SD:
Range:	Mode:	Median:
Q1:	Q3:	IQR:

   (b) Half of the dog ages lie between what two values?

   (c) Are there any outliers in this data set? Use the fences to decide. State the fences and if there are any outliers, identify them.

   (d) Construct a boxplot (modified if appropriate) for this data set. Label everything and include an axis.

2. The `survey` data set in `R` contains contains the responses of 237 Statistics I students at the University of Adelaide to a number of questions. The codebook for select variables in this data can be found in the Appendix.

(a) Define the data type for all variables listed in the codebook.

(b) What variable(s) would it be appropriate to create a scatterplot of?

(c) What variable(s) would it be appropriate to create a bar chart of?

(d) Describe one interesting relationship in this data and what type of graphic you would create to examine this relationship. You can sketch a picture if you think it would help.

3. Two way tables for the cross-combination of gender and exercise frequency are available in the Appendix. This data comes from the student survey data pre-built into R. Use these tables to answer the following questions. Show your work as needed. Single number answers will suffice.

   (a) What percent of the sample are female?

   (b) What percent of the sample exercise at least some of the time?

   (c) What percent of female students exercise frequently?

   (d) What percent of male students don't exercise at all?

   (e) What percent of students who don't exercise at all are female?

4. Rosiglitazone is the active ingredient in the controversial type 2 diabetes medicine Avandia and hasbeen linked to an increased risk of serious cardiovascular problems such as stroke, heart failure, and death. A common alternative treatment is pioglitazone, the active ingredient in a diabetes medicine called Actos. Data collected as part of a nationwide retrospective observational study of 227,571 Medicare beneficiaries aged 65 years or older are summarized in the contingency table below.

		Cardiovascular problems		
		Yes	No	Total
Treatment	Rosiglitazone	2,593	65,000	67,593
	Pioglitazone	5,386	154,592	159,978
	Total	7,979	219,592	227,571

Use the information in this table to answer the following questions.

(a) How many patients on Pioglitazone had cardiovascular problems? How many patients on Rosiglitazone had problems? Can we conclude from these numbers that the rate of cardiovascular problems for those on a Pioglitazone treatment is higher.

(b) Do the data suggest that diabetic patients who are taking rosiglitazone are more likely to have cardiovascular problems than those on pioglitazone? Justify your answer.

(c) What proportion of all patients had cardiovascular problems?

5. Data on the counts of insects in agricultural experimental units treated with different insecticide spray was analyzed in the article Beall, G., (1942) *The Transformation of data from entomological field experiments*, Biometrika, 29, 243-262.
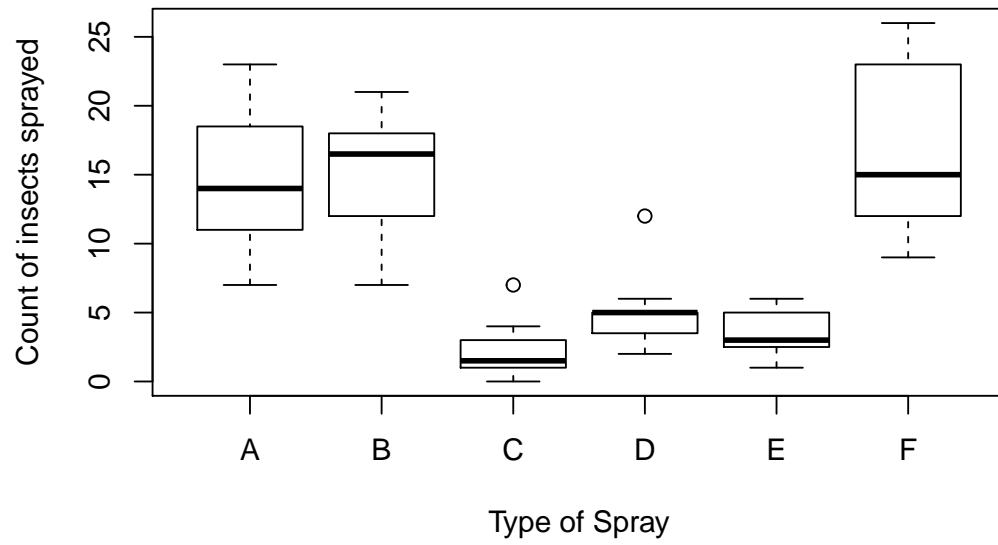

What are three pieces of information you get from this plot.

6. The following histograms display the heights of males and females from the student survey data. The Appendix contains summary statistics that you will need to use to answer the following questions.
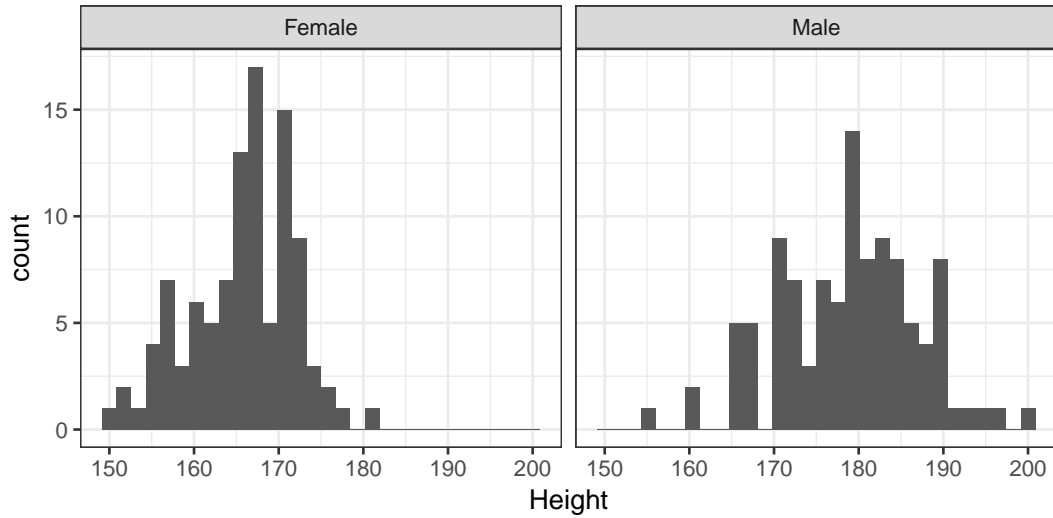
(a) What does the height of the bars represent? Be specific.

(b) On average which gender is taller? Which gender is more variable in their reported heights? Provide numerical justification for your answers.

7. Using the `survey` data set the following plot comparing the hand spans for each student was created. Use this plot and additional code in the Appendix to answer the following questions.
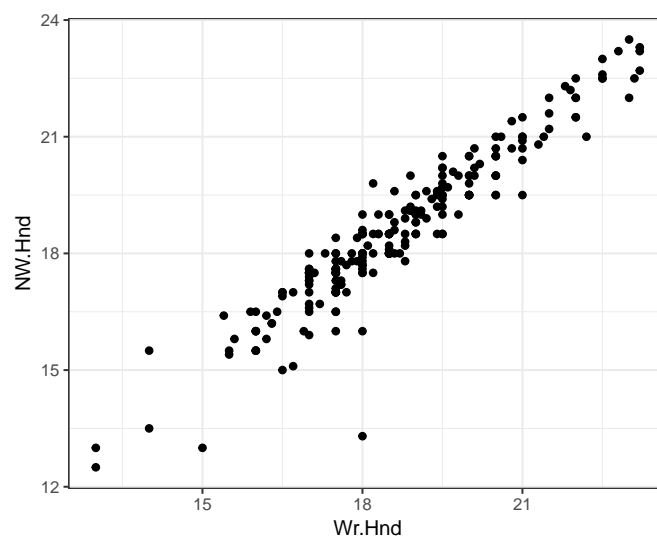
(a) Describe the relationship observed here. Be sure to talk about the direction, strength, and form of the relationship. Additional results found in the Appendix may be useful here.

R code and subsequent output for Exam 1

```
library(openintro)
```

Dog ages

```
dog_age

## [1] 98 90 92 70 83 45 92 90 76 72 74 63 88 75 84 80 80 68 77 92

summary(dog_age)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   45.00   73.50   80.00   79.45   90.00   98.00

var(dog_age)

## [1] 155.1026
```

Student survey data

- **Sex** The sex of the student. (Factor with levels "Male" and "Female".)

- **Wr.Hnd** span (distance from tip of thumb to tip of little finger of spread hand) of writing hand, in centimetres.

- **W.Hnd** writing hand of student. (Factor, with levels "Left" and "Right".)

- **Exer** how often the student exercises. (Factor, with levels "Freq" (frequently), "Some", "None".)

- **Smoke** how much the student smokes. (Factor, levels "Heavy", "Regul" (regularly), "Occas" (occasionally), "Never".)

- **Height** height of the student in centimetres.

- **Age** age of the student in years.

Summary Statistics of height by gender

```
tapply(survey$Height, survey$Sex, FUN=summary) # A

## $Female
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   150.0   162.6   166.8   165.7   170.0   180.3      16
##
## $Male
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   154.9   172.8   180.0   178.8   185.0   200.0      12

tapply(survey$Height, survey$Sex, FUN=var, na.rm=TRUE) # B

##   Female     Male
## 37.84436 70.22862

table(survey$Sex) # C

##
## Female    Male
##    118     118
```

Correlation

```r
cor(survey$Wr.Hnd, survey$NW.Hnd, use="pairwise.complete.obs") # D
```

```
## [1] 0.9483103
```

Frequency and proportion tables

```r
table(survey$Exer) # E
```

```
##
## Freq None Some
##  115   24   98
```

```r
round(prop.table(table(survey$Exer)),2) # F
```

```
##
## Freq None Some
## 0.49 0.10 0.41
```

```r
table(survey$Sex,survey$Exer) # G
```

```
##
##          Freq None Some
##   Female   49   11   58
##   Male     65   13   40
```

```r
round(prop.table(table(survey$Sex,survey$Exer)),2) # H
```

```
##
##          Freq None Some
##   Female 0.21 0.05 0.25
##   Male   0.28 0.06 0.17
```

```r
round(prop.table(table(survey$Sex,survey$Exer), margin=1),2) # I
```

```
##
##          Freq None Some
##   Female 0.42 0.09 0.49
##   Male   0.55 0.11 0.34
```

```r
round(prop.table(table(survey$Sex,survey$Exer), margin=2),2) # J
```

```
##
##          Freq None Some
##   Female 0.43 0.46 0.59
##   Male   0.57 0.54 0.41
```