HW3: Sections 2.3, 5.2, and 6.2

Your Name Here

The code below just loads some packages and makes it so that enough digits are printed that you won't get confused by rounding errors.

```
library(dplyr) # functions like summarize
```

```
## Warning: package 'dplyr' was built under R version 3.5.2
```

```
library(ggplot2) # for making plots
```

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
library(mosaic) # convenient interface to t.test function
```

```
## Warning: package 'mosaic' was built under R version 3.5.2
```

```
## Warning: package 'ggformula' was built under R version 3.5.2
```

```
library(readr)
library(gmodels)

options("pillar.sigfig" = 10) # print 10 significant digits in summarize output
```

Problem 1: Sleuth3 2.12 (Marijuana use during pregnancy)

For the birth weights of babies in two groups, one born to mothers who used marijuana during pregnancy and the other born to mothers who did not, the difference in sample averages (mothers who did not use marijuana minus mothers who used marijuana) was 280 grams, and the standard error of the difference was 46.66 grams with 1,095 degrees of freedom.

(a) From the information above, find a 95% confidence interval for $\mu_2 - \mu_1$ (the difference in means for mothers who did not use marijuana and for mothers who did use marijuana) and interpret the interval in context. As part of your answer, explain what the phrase "95% confident" means.### (a) From the information above, find a 95% confidence interval for $\mu_2 - \mu_1$ (the difference in means for mothers who did not use marijuana and for mothers who did use marijuana) and interpret the interval in context.

```
qt(0.975, df = 1095)
```

```
## [1] 1.962133
```

```
280 - 1.962 * 46.66
```

```
## [1] 188.4531
```

```
280 + 1.962 * 46.66
```

```
## [1] 371.5469
```

A 95% confidence interval for the difference in mean birth weights in the population of babies born to mothers who did not use marijuana during pregnancy is and the population of babies born to mothers who used marijuana during pregnancy is [188.5, 371.6] grams. In 95% percent of samples, an interval calculated in this way would contain the difference in population means.

It's worth noting that this is almost certainly an observational study, so we cannot claim from this analysis alone that smoking marijuana during pregnancy leads to reduced birth weights. That said, in combination with other knowledge, the evidence of a connection is strong.

(b) Find a 90% confidence for $\mu_2 - \mu_1$. No need to interpret the interval in context.

```
qt(0.95, df = 1095)
```

```
## [1] 1.646246
```

```
280 - 1.646 * 46.66
```

```
## [1] 203.1976
```

```
280 + 1.646 * 46.66
```

```
## [1] 356.8024
```

An approximate 90% confidence interval for the difference in means is [203.2, 356.8] grams.

(c) Calculate the t statistic for a test of the claim that there is no difference in the birth weights for the two groups. You don't need to find the p-value.

```
280/46.66
```

```
## [1] 6.000857
```

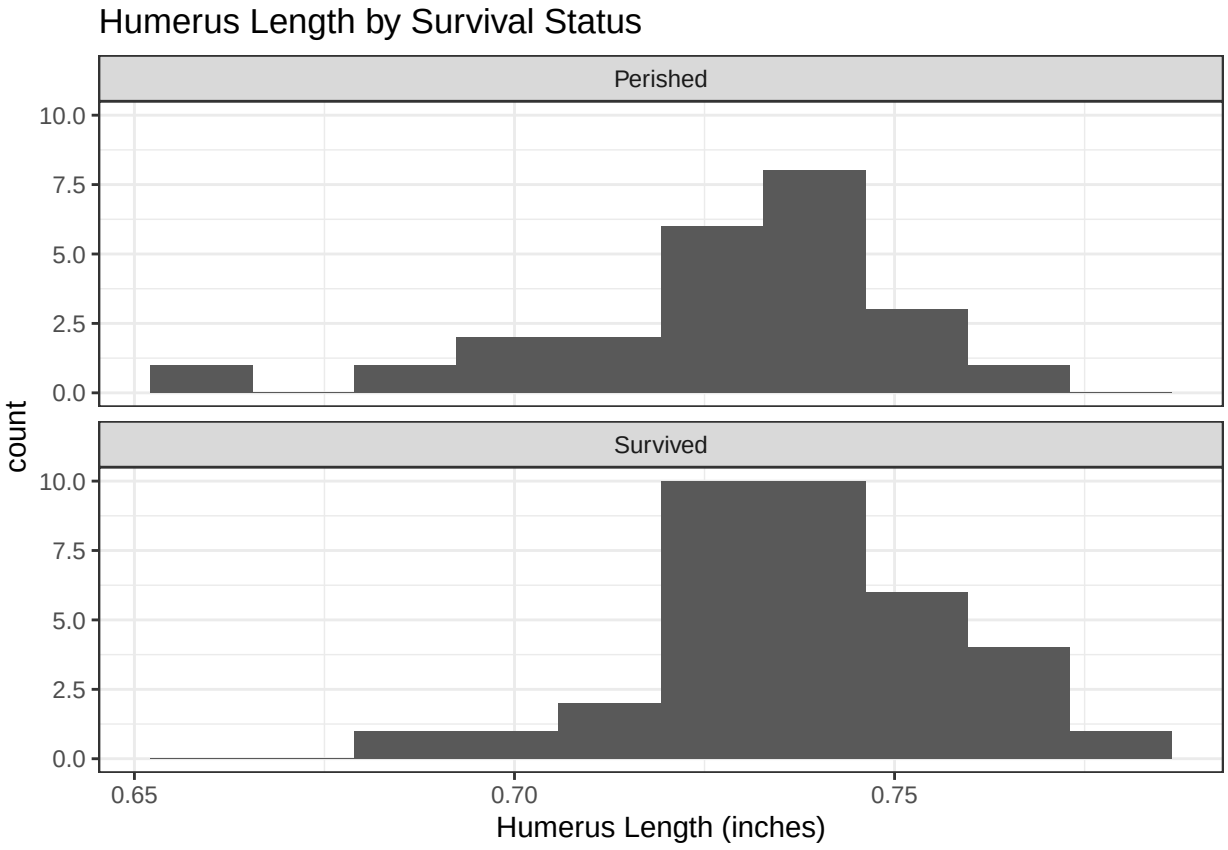
Problem 2: Adapted from Sleuth3 2.21

In 1899, biologist Hermon Bumpus presented as evidence of natural selection a comparison of numerical characteristics of moribund house sparrows that were collected after an uncommonly severe winter storm and which had either perished as a result of their injuries or survived. The following R code reads in a data set with the length of the humerus (arm bone) in inches for 59 of these sparrows, grouped according to whether they survived or perished. Analyze these data to summarize the evidence that the mean humerus length is different in the two populations.

```
sparrows <- read.csv("http://www.evanlray.com/data/sleuth3/ex0221_sparrows.csv")
```

(a) Make an appropriately labelled plot of the data.

```
ggplot(data = sparrows, mapping = aes(x = Humerus)) +
  geom_histogram(, bins = 10) +
  facet_wrap( ~ Status, ncol = 1) +
  xlab("Humerus Length (inches)") +
  ggtitle("Humerus Length by Survival Status") +
  theme_bw()
```



(b) Fit a linear model to the data and print out the model summary.

```
mf <- lm(Humerus ~ Status, data = sparrows)
summary(mf)
```

```
##
## Call:
## lm(formula = Humerus ~ Status, data = sparrows)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.068917 -0.010000  0.001083  0.014000  0.042000
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.727917   0.004370 166.555   <2e-16 ***
## StatusSurvived 0.010083   0.005674   1.777   0.0809 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02141 on 57 degrees of freedom
## Multiple R-squared:  0.05249,    Adjusted R-squared:  0.03587
## F-statistic: 3.158 on 1 and 57 DF,  p-value: 0.0809
```

(c) Write down the equation for the mean in terms of the linear model coefficients β_0 and β_1.

$\mu = \beta_0 + \beta_1 StatusSurvived$

(d) What is the interpretation of each of the parameters β_0 and β_1?

β_0 is the mean humerus length in the population of birds who died in the storm.

β_1 is the difference between the mean humerus length in the populations of birds who survived the storm and the mean humerus length in the population of birds who died in in the storm.

(e) Using the output from the model summary in part (b), conduct a test of the claim that there is no difference in the mean length of the humerus in the populations of sparrows before and after the storm.

The p-value for this test is 0.0809, providing weak evidence against the null hypothesis of no difference in the mean length of the humerus in sparrows who survived or died in the storm.

(f) Find a confidence interval for the difference in the mean length of the humerus in the populations of sparrows before and after the storm and discuss what it means in context. You can do this either using the `fit.contrast` function as we've done in class, or by calling the `confint` function directly on your model fit from part (b) to obtain confidence intervals for the linear model coefficients.

```
fit.contrast(mf, "Status", c(-1, 1), conf = 0.95)
```

```
##                     Estimate  Std. Error  t value  Pr(>|t|)     lower CI
## Status c=( -1 1 ) 0.01008333 0.005674365 1.776998 0.0809045 -0.001279386
##                     upper CI
## Status c=( -1 1 ) 0.02144605
## attr(,"class")
## [1] "fit_contrast"
```

We are 95% confident that the difference in mean humerus lengths in the populations of sparrows that survived in the storm and sparrows that died in the storm is between -0.001 and 0.021 inches.

Problem 3: Adapted from Sleuth3 6.22

Was *Tyrannosaurus rex* warm-blooded? The R code below reads in data with several measurements of the oxygen isotopic composition of bone phosphate in each of 123 bone specimens from a single *Tyrannosaurus rex* skeleton. It is known that the oxygen isotopic composition of vertebrate bone phosphate is related to the body temperature at which the bone forms. Differences in means at different bone sites would indicate nonconstant temperatures throughout the body. Minor temperature differences would be expected in warm-blooded animals.

The following R code reads in a data set with measurements of oxygen isotopic composition of vertebrate bone phosphate (per mil deviations from SMOW) in 12 bones of a singe Tyrannosaurus rex specimen. For each bone sample, multiple measurements were taken.

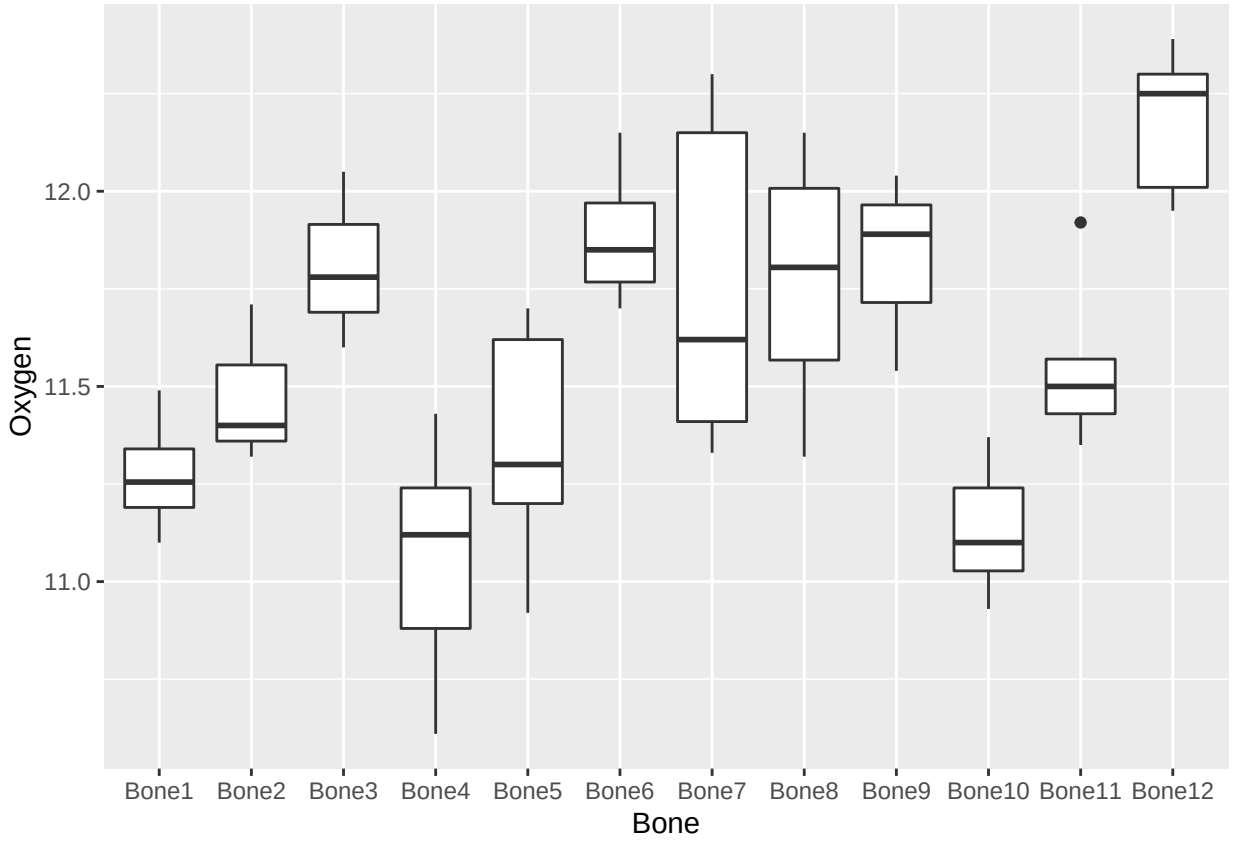
```
trex <- read_csv("http://www.evanlray.com/data/sleuth3/ex0523_trex.csv") %>%
  mutate(
    Bone = factor(Bone, levels = paste0("Bone", 1:12))
  )
```

```
## Parsed with column specification:
## cols(
##   Oxygen = col_double(),
```

```
##   Bone = col_character()
## )
```

(a) Make side by side box plots of the data

```
ggplot(data = trex, mapping = aes(x = Bone, y = Oxygen)) +
  geom_boxplot()
```



(b) Bones 2 and 3 were both gastralia (roughly similar to human ribs), and bones 4 and 5 were both dorsal vertebra (part of the backbone). Is there evidence that the mean oxygen isotopic composition is different for the gastralia than for the dorsal vertebra? Specify a linear combination of means for the 12 bones that could be used to address this question, and conduct a relevant hypothesis test and find a confidence interval. Interpret all of your results in context.

Let μ_2 be the mean oxygen isotopic composition in bone 2, and similar for μ_3, μ_4, and μ_5.

We are interested in whether or not $0.5(\mu_2 + \mu_3) = 0.5(\mu_4 + \mu_5)$, or equivalently, whether $0.5\mu_2 + 0.5\mu_3 + (-0.5)\mu_4 + (-0.5)\mu_5 = 0$. Formally, in terms of all 12 means, our hypotheses could be stated as:

$H_0 : 0\mu_1 + 0.5\mu_2 + 0.5\mu_3 + (-0.5)\mu_4 + (-0.5)\mu_5 + 0\mu_6 + 0\mu_7 + 0\mu_8 + 0\mu_9 + 0\mu_{10} + 0\mu_{11} + 0\mu_{12} = 0$

$H_A : 0\mu_1 + 0.5\mu_2 + 0.5\mu_3 + (-0.5)\mu_4 + (-0.5)\mu_5 + 0\mu_6 + 0\mu_7 + 0\mu_8 + 0\mu_9 + 0\mu_{10} + 0\mu_{11} + 0\mu_{12} \neq 0$

```
dinosaur_fit <- lm(Oxygen ~ Bone, data = trex)
fit.contrast(dinosaur_fit, "Bone", c(0, 0.5, 0.5, -0.5, -0.5, 0, 0, 0, 0, 0, 0, 0), conf.int = 0.95)
```

```
##                                              Estimate Std. Error  t value
## Bone c=( 0 0.5 0.5 -0.5 -0.5 0 0 0 0 0 0 0 ) 0.4413333  0.1407312 3.136001
```

```
##                                                    Pr(>|t|)  lower CI
## Bone c=( 0 0.5 0.5 -0.5 -0.5 0 0 0 0 0 0 0 ) 0.003206777 0.1569049
##                                                   upper CI
## Bone c=( 0 0.5 0.5 -0.5 -0.5 0 0 0 0 0 0 0 ) 0.7257617
## attr(,"class")
## [1] "fit_contrast"
```

We estimate that mean oxygen isotopic composition is about 0.44 units (not sure what the units are) higher for the gastralia than for the dorsal vertebra, with a 95% confidence interval of about 0.16 to 0.73; in about 95% percent of samples, a confidence interval calculated in this way would contain the difference in mean oxygen isotopic concentrations for these bones. Additionally, the p-value for a test of whether the difference is 0 is about 0.003. These data provide very strong evidence against the null hypothesis of no difference. Altogether, the data indicate that the mean oxygen isotopic concentration is different in these bone groups, suggesting that there were differences in body temperatures in those regions of the dinosaur's body and lending support to the theory that the T. rex was warm-blooded.