Take Home Quiz 3

Robert Campbell

Due 1:00pm Friday, May 7

This quiz should take you approximately 25 minutes. Place your answers into this markdown document, knit it, and hand in the result as a PDF. Just answering is not enough - you need to include the R code that produces your answer.

You may use R, the internet, and any reference material, but do not work together and do not get help (except from Dr. Clair).

Problem 1

A company is receiving a shipment of parts. They randomly select a sample and check that the parts are made to specifications. If they find that the sample don't meet specifications, they return the whole shipment.

a. What is the null hypothesis for this test? That all of the parts in the shipment meets specifications.
b. What would a type I error be for this test? What problems could a type I error cause? That all parts in the shipment meet specification, but the test showed that they did not.
   This would cause good parts to be returned or scrapped and new parts need to be produced to fill the order, generating unnecesarry waste.
c. What would a type II error be for this test? What problems could a type II error cause? That not all parts in the shipment meet specification, but the test showed that they did.
   This would cause the shipment to go to market or production with bad parts. Depending on the part that could be a minor but acceptable risk (some cheap product, like a toy) or it could be some critical part and the risk would bwe unacceptable (faulty ignition in a car).

Problem 2

On episode 2 of the Nicolas Cage documentary "The History of Swear Words," six comedians took part in an experiment. Each was asked to submerge their hand in a bucket of ice water for as long as they could stand it. Four of the comedians were allowed to swear out loud while performing the experiment, the other two were forbidden from swearing. The two clean subjects spent 53 and 58 seconds in the water, while the four swearing comedians spent 69, 78, 87, and 140 seconds in the water.

```
test<- data.frame(time=c(53, 58, 69, 78, 87, 140), swear<- c(0, 0, 1, 1, 1, 1))
```

a. Test for a difference between the two groups using a Wilcoxon rank sum test and report your p-value.
   p-value = 0.1333

```r
wilcox.test(time~swear, data=test)
```

```
##
##  Wilcoxon rank sum exact test
##
## data:  time by swear
## W = 0, p-value = 0.1333
## alternative hypothesis: true location shift is not equal to 0
```

b. Repeat your analysis with a t-test.
   p-value = 0.09572

```r
t.test(time~swear, data=test)
```

```
##
##  Welch Two Sample t-test
##
## data:  time by swear
## t = -2.3567, df = 3.1439, p-value = 0.09572
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -88.01207  12.01207
## sample estimates:
## mean in group 0 mean in group 1
##            55.5            93.5
```

c. Were either of your analyses significant at the $\alpha = 0.05$ level? What conclusions would you draw from this experiment?
   No, neither were. The correlation between swearing and time submerged in water is not significant in this sample.

Problem 3

Suppose a population is normally distributed with mean 25 and standard deviation 7.

We want to test $H_0 : \mu = 20$ against $H_a : \mu \neq 20$ at the $\alpha = 0.05$ level of significance.

a. What power would a t-test have with a sample of size $n = 20$?
   Power = 0.8576

```r
power.t.test(n=20, delta=20-25, sd=7, type="one.sample")
```

```
##
##      One-sample t test power calculation
##
##               n = 20
##           delta = 5
##              sd = 7
##       sig.level = 0.05
##           power = 0.8575538
##     alternative = two.sided
```

b. What sample size would you need for a power of 95%? n is approximately 28 (assuming integers).

```
power.t.test(delta=20-25, sd=7, power=0.95, type="one.sample")
```

```
##
##      One-sample t test power calculation
##
##               n = 27.47122
##           delta = 5
##              sd = 7
##       sig.level = 0.05
##           power = 0.95
##     alternative = two.sided
```

Problem 4

Problems 4 and 5 both use data on oak tree acorns from Marcelo & Patterson, 1990. The authors suggest that larger acorns allow for greater success in ecological niches, leading to a wider geographic range.

Data is available on our course data page, with `Range` giving the geographic range in km^2 x 100 and `Acorn_size` giving the acorn size in cm^3.

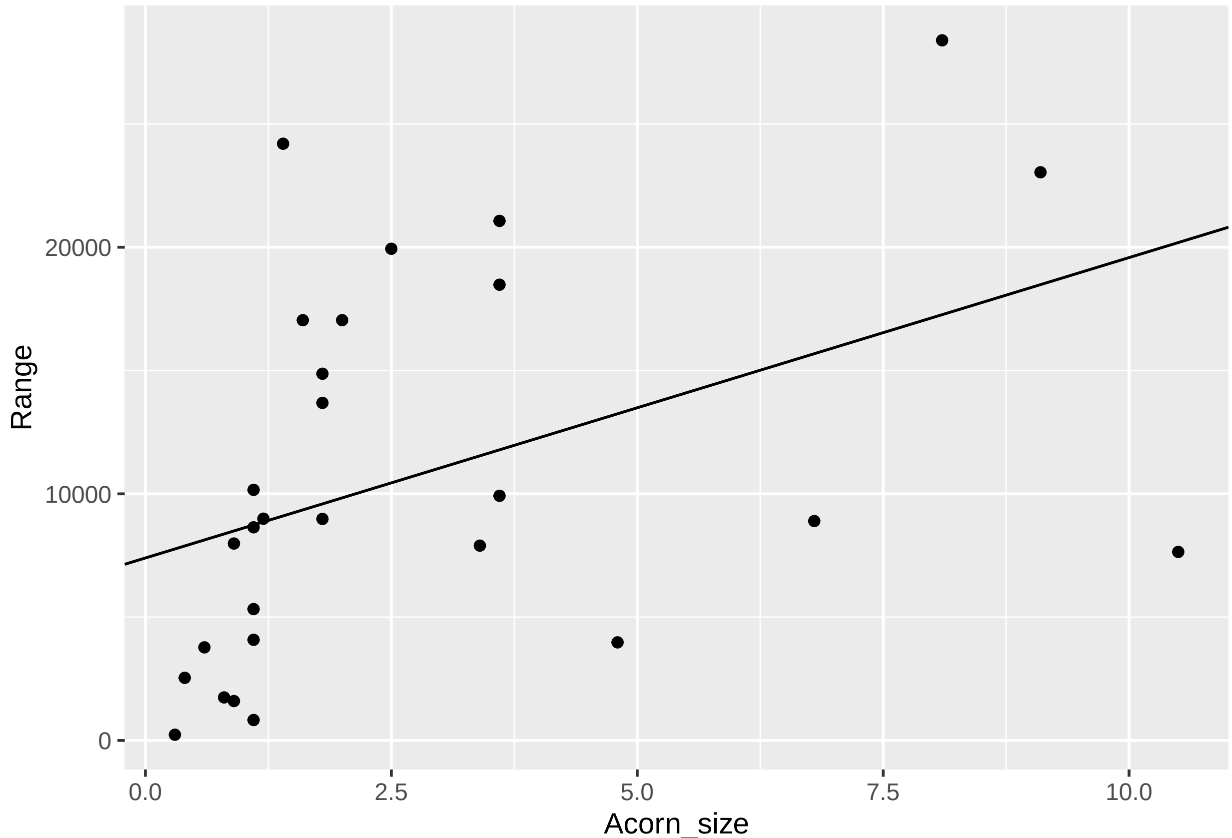
```
acorns <- read.csv("https://mathstat.slu.edu/~clair/stat/data/acorns.csv")
```

Make a scatterplot showing Range as explained by Acorn_size, and add the regression line to your plot.

```
range_mod<-lm(Range ~ Acorn_size, data=acorns)
range_mod
```

```
##
## Call:
## lm(formula = Range ~ Acorn_size, data = acorns)
##
## Coefficients:
## (Intercept)   Acorn_size
##        7399         1218
```

```
acorns %>% ggplot(aes(x=Acorn_size, y=Range)) + geom_point() +
  geom_abline(slope=1218, intercept=7399)
```



What is the equation of the regression line? y: Range, x: Acorn_size
y = 7399 + 1218x

Problem 5

Continue working with acorns.

a. Predict the geographic range for an oak tree with acorn size 3.6 cm^3.

```
predict(range_mod, newdata = data.frame(Acorn_size=3.6))
```

```
##        1
## 11784.92
```

b. Test for a linear relationship between acorn size and geographic range. Report a p-value with your conclusion.
   There is a linear relationship between acorn size and geographic range with a p-value of 0.02384.

```
summary(range_mod)
```

```
##
## Call:
## lm(formula = Range ~ Acorn_size, data = acorns)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12544.7  -5667.2   -560.4   5632.2  15091.2
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7399.3     1949.8   3.795 0.000796 ***
## Acorn_size    1218.2      507.5   2.400 0.023836 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7204 on 26 degrees of freedom
## Multiple R-squared:  0.1814, Adjusted R-squared:  0.1499
## F-statistic: 5.761 on 1 and 26 DF,  p-value: 0.02384
```