STAT641 - Homework 1

Marcel Gietzmann-Sanders

1. General Questions

a. Marcel Gietzmann-Sanders - 31308957
b.
c. No
d. Just wanting a broad introduction to the subject.
e. I'm studying fisheries and frequentists stats is everywhere. I'd like to get a better handle on the Bayesian side of things.
f. Yes
g. This will be the first time.
h. Yes
i. I've been a software engineer working primarily in Python for the past 7 years. All of my classes here have used R. So at this point I probably have over 10,000 hours of Python and several hundred hours of R experience.
j. I'm on a macbook but I typically run a Linux VM for my work.
k. No

2. Opening Problems

(a) Use R to simulate the following; state the values it generates.

```
rbinom(5, size=12, prob=0.4)
```

```
## [1] 3 5 7 5 9
```

n is the number of trials per experiment and p is the probability of a success (or equivalently, of a 1) per trial.

```
rnorm(5, mean=0.0, sd=0.75)
```

```
## [1]  0.03922513  0.47911105 -1.59215559 -1.22548798  1.81404411
```

(b) Confidence Intervals

> I purchase a random sample of n = 49 avocados at Fred Meyer. The sample mean weight is 4.3 ounces with a standard deviation of 0.36 ounces. Calculate a 95% confidence interval for the mean weight of all such avocados, and interpret your confidence interval. (Be sure to state the formula you're using, and fill in the various values in the formula.) What, exactly, is it that happens with 95% probability? What does the "95%" refer to when constructing a 95% confidence interval.

The formula for a 95% confidence interval is:

$$\bar{\mu} \pm Z_{0.975} \frac{s}{\sqrt{n}}$$

where $\bar{\mu}$ is the sample mean, s is the sample standard deviation, and $Z_{0.975}$ is the z-score at which the cumulative probability of a normal distribution is 97.5%. Because this is two tailed, choosing the 97.5% ensures that we are covering 95% of the distribution.

In our case this becomes:

$$4.3 \pm 1.96 \frac{0.36}{\sqrt{49}} = 4.3 \pm 0.1008$$

So this is saying (informally) that we are 95% confident that the true mean weight of avocados at Fred Meyer lies somewhere between 4.1992 and 4.4008 ounces. Formally if we were to repeat this experiment over and over 95% of the time the true mean weight would lie within the confidence interval derived in that experiment.

(c) Hypothesis Testing

> I conducted a poll consisting of a random sample of n = 1000 sane individuals. 790 of those polled answered “Yes” to the question, “Is the world going insane?” Test the hypotheses, H0 : p = 0.75 versus Ha : p > 0.75, where p is the proportion of all sane individuals who believe the world is going insane. (It’s okay to conduct the test by hand or to use the R function prop.test.) At level α = 0.05, do we reject H0 or fail to reject H0?
>
> It turns out that the p-value for this test is 0.00174. At level α = 0.01, should we reject H0? Why or why not?

```
prop.test(790, 1000, p=0.75, alternative="greater", correct=FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  790 out of 1000, null probability 0.75
## X-squared = 8.5333, df = 1, p-value = 0.001744
## alternative hypothesis: true p is greater than 0.75
## 95 percent confidence interval:
##  0.7680456 1.0000000
## sample estimates:
##    p
## 0.79
```

In this case the p-value is 0.00174 which is far below our $\alpha = 0.05$ threshold. Therefore we can go ahead and reject the null hypothesis.

We should also reject the null hypothesis given an $\alpha = 0.01$ because we are still below the threshold.