Aashwin Vats

Homework#4

1. *The data are a random sample of 1000 mortgages from all 65,702 mortgages purchased by Federal Home Loan Banks in 2018 and reported to the Federal Housing Finance Agency. The data include the loan interest rate (\Rate"), the amount of the loan ("Amount"), and a yes/no variable to indicate if this was a borrower's first mortgage.*
Research Question: Are the amounts of first mortgages smaller?

We will do a Wilcoxon ran-sum test to answer this question.
We have two populations:

- Population 1 consists of first mortgage borrowers.
- Population 2 consists of borrowers who have taken mortgage at least once before, let's call them secondary borrowers.

a) Null Hypothesis:

Y_1 = random amount from population 1 (first time borrowers)
Y_2 = random amount from population 2 (secondary borrowers)

If you take random amount from each population, neither is more likely to have a smaller amount.

H_0: $P(Y_1 < Y_2) = 0.5$

P: Probability
This means that there are 50-50 chances which amount is smaller. We will assume $P(Y_1 = Y_2) = 0$.

b) Alternative Hypothesis:

$H_A = P(Y_1 < Y_2) > 0.5$

Amount for first time borrowers is more likely to be smaller than the amount of secondary borrowers. This is a one-sided test. H_A is determined by the research question, "Is the amount of first mortgages smaller?"

A two-sided alternative would be
$H_A : P(Y_1 < Y_2) \neq 0.5$

c) Wilcox test:

R Code and Output:

```
#Wilcox test
loans <- read.csv("Loans.csv")
View(loans)

with(loans,wilcox.test(x=Amount[First=="Yes"],
                       y=Amount[First=="No"],
                       alternative="less",
                       correct=FALSE,
                       exact=FALSE))
```

```
> with(loans,wilcox.test(x=Amount[First=="Yes"],
+                        y=Amount[First=="No"],
+                        alternative="less",
+                        correct=FALSE,
+                        exact=FALSE))

        Wilcoxon rank sum test

data:  Amount[First == "Yes"] and Amount[First == "No"]
W = 71036, p-value = 3.508e-06
alternative hypothesis: true location shift is less than 0
```

d) Statistical Conclusion:

There is convincing evidence that a randomly chosen amount from first time borrowers (pop'n 1) is more than 50% likely to be smaller than a randomly chosen amount from secondary borrowers (pop'n 2). (p-value ≈ 0.0000035, one –sided Wilcoxon rank-sum test).

2. *Using the diet study data (ex0112), conduct a two-sample Welch's t-test to answer the research question, "does including fish oil in diet have any effect on diastolic blood pressure?" Include your code and output.*

```
#Welch's Test:
t.test(BP~Diet, data= ex0112, alternative="greater")
```

```
> #Welch's Test:
> t.test(BP~Diet, data= ex0112, alternative="greater")

        Welch Two Sample t-test

data:  BP by Diet
t = 3.0621, df = 9.2643, p-value = 0.006542
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 3.111056      Inf
sample estimates:
   mean in group FishOil mean in group RegularOil
                6.571429                -1.142857
```

3. *The research question is, "Does timber harvest reduce shade on fish-bearing streams?"*

a) Yes, it is reasonable to assume that the two-shade measurements at a single site are independent. The shade measurements pre-harvest will not have any impact on the shade measurements post-harvest as we are comparing the impact on the same site under two different circumstances.

b) R code and Output:

```
shade <- read.csv("shade.csv")
View(shade)

#Sign test
difference<-with(shade,PRE-POST)
n <- length(difference)
K <- length(which(difference>0))
binom.test(K,n,alternative="greater")
```

```
> binom.test(K,n,alternative="greater")

        Exact binomial test

data:  K and n
number of successes = 17, number of trials = 26, p-value =
0.08432
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.4738376 1.0000000
sample estimates:
probability of success
             0.6538462
```

c) Statistical conclusion:

There is suggestive, but inconclusive evidence that the timber harvest reduces the shade on fish-bearing streams, meaning that more than 50% of median differences are negative (p-value 0.08432, one-sided sign test).