HW 22

For Problems 1-2, consider the following table represents sales of beer in Pennsylvania where it must be sold only by the case. A friend suggests that this means beer is likely less expensive in PA than elsewhere. To test this theory, consider the price per case for $n = 12$ popular beers at *Rite Buy* in PA and at *Total Wine* in DE, where beer is also sold in 6-packs and 12-packs and isn't taxed.

```
mydata <- read.csv("https://raw.githubusercontent.com/ProfSuzy/Stat61/main/beer_data_2005.csv",
                   header=T)
mydata <- mydata[,1:4]
mydata
```

```
##              Beer DE.price..Total.Wine. PA.price..Rite.Buy. Difference..PA...DE.
## 1       Yuengling                 13.99               18.59                 4.60
## 2    Rolling Rock                 13.99               19.19                 5.20
## 3           Coors                 16.38               18.99                 2.61
## 4       Budweiser                 18.99               18.59                -0.40
## 5         Fosters                 19.98               23.99                 4.01
## 6          Pete’s                 21.98               22.19                 0.21
## 7     Dock Streed                 22.99               21.99                -1.00
## 8       Dos Equis                 23.98               26.59                 2.61
## 9       Newcastle                 21.48               29.99                 8.51
## 10 Sierra Nevada                  22.98               29.99                 7.01
## 11         Corona                 23.98               30.19                 6.21
## 12        Victory                 27.99               28.29                 0.30
```

We are interested in estimating μ_{diff}, the mean difference in price for a case of beer in PA vs DE. The matched pairs t-test is useful since it seems reasonable to treat the price differences as a set of $n = 12$ independent values with mean μ_{diff} and some standard deviation σ_{diff}.

Problem 1

Suppose the distribution of differences in price is close enough to a Normal distribution to invoke the CLT even though $n = 12$.

(a) Compare the standard error for the difference in averages you would get if you treated these as two independent samples compared to the (more appropriate) standard error for matched pairs. Explain how this shows the value of matching when it is reasonable to do so.

(b) Find an approximate 99% CI for μ_{diff} and interpret what this interval suggests about the price differences.

```
## replace this with code for Problem 1
```

Problem 2

Now let's not suppose we can justify the use of the CLT. The sign test is a non-parametric alternative that only assumes the data of $n = 12$ differences are IID with some constant probability θ of being positive. Thus we can treat the number of positive differences, Y, as a Binomial(n, θ) RV and test $H_0 : \theta = 0.5$ vs $H_1 : \theta \neq 0.5$.

(a) Interpret the null and alternative hypotheses in the context of this problem.

(b) Use the binomial probability function ($dbinom()$ in R) to find the exact p-value for this test and explain its meaning in the context of this problem.

```
## replace this with code for Problem 1
```

Problem 3

A 3-year study with 72 chronic cocaine user considered an antidepressant drug called desipramine as a possible treatment for addiction. A clinical trial compared outcomes for subjects randomly assigned to take either desipramine, lithium, or a placebo. The counts of subjects who relapsed within the 3 years are reported in the following table.

```
coc_dat = data.frame( drug = c("Desipramine", "Lithium", "Placebo"),
                      n = c(24, 24, 24),
                      relapse = c(10, 18, 20))
coc_dat
```

```
##          drug  n relapse
## 1 Desipramine 24      10
## 2     Lithium 24      18
## 3     Placebo 24      20
```

(a) Construct a 95% CI for the difference in relapse rates for the placebo compared to desipramine.

(b) The lithium group was included because this had been show previously to be an effective treatment for addiction. To argue for desipramine, it should be shown to be at least as effective as lithium. Define parameters, state hypotheses and carry out the test using Fisher's exact test. Give the exact p-value and explain your conclusion at an $\alpha = 0.01$ significance level.

(c) State hypotheses and carry out a Chi-square test of independence for all three groups. Give an approximate p-value for the test and explain what it represents in the context of this problem.

```
## replace this with code for Problem 1
```