Assignment 3

James La Fontaine

12/10/2021

## Question 1

### 1ai)

H0: mX = mY

H1: mX ≠ mY

binom.test(12, 17)

##   
## Exact binomial test  
##   
## data: 12 and 17  
## number of successes = 12, number of trials = 17, p-value = 0.1435  
## alternative hypothesis: true probability of success is not equal to 0.5  
## 95 percent confidence interval:  
## 0.4404173 0.8968645  
## sample estimates:  
## probability of success   
## 0.7058824

The p-value > 0.05 so we cannot reject the null hypothesis that the median of X equals the median of Y at the significance level of 5%.

### 1aii)

H0: mX = mY

H1: mX ≠ mY

x = c(26.1, 26.6, 27.4, 27.5, 27.8, 28.1, 28.4, 29.5, 29.8, 30.4, 30.4, 31.2, 31.5, 32.9, 33.6, 34.1, 35.9)  
y = c(27.4, 28.1, 22.9, 31.3, 16.3, 50.1, 20.0, 24.6, 23.3, 19.3, 24.4, 24.4, 29.5, 27.6, 21.7, 25.4, 39.4)  
  
wilcox.test(x, y, paired = TRUE)

##   
## Wilcoxon signed rank exact test  
##   
## data: x and y  
## V = 124, p-value = 0.02322  
## alternative hypothesis: true location shift is not equal to 0

The p-value < 0.05 so we reject the null hypothesis at the median of X equals the median of Y at the significance level of 5% and can conclude that there is sufficient evidence to show that the location of X and Y differ.

### 1aiii)

H0: μX = μY

H1: μX ≠ μY

x = c(26.1, 26.6, 27.4, 27.5, 27.8, 28.1, 28.4, 29.5, 29.8, 30.4, 30.4, 31.2, 31.5, 32.9, 33.6, 34.1, 35.9)  
y = c(27.4, 28.1, 22.9, 31.3, 16.3, 50.1, 20.0, 24.6, 23.3, 19.3, 24.4, 24.4, 29.5, 27.6, 21.7, 25.4, 39.4)  
  
t.test(x, y, paired = TRUE)

##   
## Paired t-test  
##   
## data: x and y  
## t = 1.6402, df = 16, p-value = 0.1205  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.954790 7.484202  
## sample estimates:  
## mean of the differences   
## 3.264706

The p-value > 0.05 so we cannot reject the null hypothesis that the mean of X equals the mean of Y at the significance level of 5%.

## 1b)

The sign test has a larger type II error rate / smaller power and so it is plausible that the null hypothesis has incorrectly not been rejected in this test. For the t-test we have made the assumption that the differences between X and Y are normally distributed and since we have a small sample size of only 17, it is plausible that these differences do not follow a normal distribution. It would be more appropriate to give more consideration to the outcome of the Wilcoxon signed-rank test in this case, which simply assumes that the differences between X and Y are continuous and follow a symmetrical distribution, which is a reasonable assumption under the null hypothesis. Therefore, there is mild evidence that X and Y differ in location, however, further testing with a larger sample would be required to make stronger conclusions.

## 1c)

B = 20000  
n = 17  
numRejectionsSign = 0  
numRejectionsWilcoxon = 0  
numRejectionsT = 0  
   
for (i in 1:B) {  
 numSuccesses = 0  
 sampleDifference = rnorm(n, 3, 5)  
 for (number in sampleDifference) {  
 if (sign(number) == 1) {  
 numSuccesses = numSuccesses + 1  
 }  
 }  
 if (binom.test(numSuccesses, n)$p.value < 0.05) {  
 numRejectionsSign = numRejectionsSign + 1  
 }  
 if (wilcox.test(sampleDifference)$p.value < 0.05) {  
 numRejectionsWilcoxon = numRejectionsWilcoxon + 1  
 }  
 if (t.test(sampleDifference)$p.value < 0.05) {  
 numRejectionsT = numRejectionsT + 1  
 }  
}  
  
powerSign = numRejectionsSign / B  
powerWilcoxon = numRejectionsWilcoxon / B  
powerT = numRejectionsT / B  
  
cat("Simulated power of sign test: ", powerSign, "\n")

## Simulated power of sign test: 0.4859

cat("Simulated power of Wilcoxon test: ", powerWilcoxon, "\n")

## Simulated power of Wilcoxon test: 0.60815

cat("Simulated power of t-test: ", powerT, "\n")

## Simulated power of t-test: 0.64395

## Question 2

### 2a)

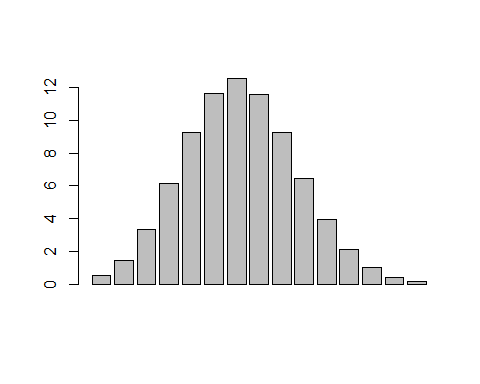
germinations = c(3, 4, 5, 6, 7, 8 ,9, 10, 11, 12, 13, 17)  
count = c(1, 2, 2, 4, 10, 16, 9, 11, 13, 4, 7, 1)  
  
experiments = data.frame(germinations, count)  
  
data = rep(experiments$germinations, experiments$count)  
  
p1 = prop.test(sum(data), 80\*30)  
  
p1

##   
## 1-sample proportions test with continuity correction  
##   
## data: sum(data) out of 80 \* 30, null probability 0.5  
## X-squared = 362.7, df = 1, p-value < 2.2e-16  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.2871084 0.3243539  
## sample estimates:  
## p   
## 0.3054167

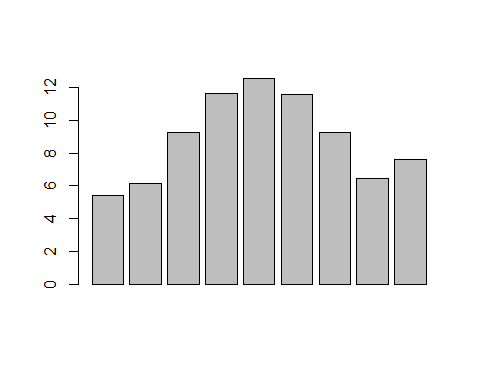
prop.estimate = as.numeric(p1$estimate)

### 2b)

barplot(dbinom(3:17, 30, prop.estimate) \* 80)



X1 <- cut(data, breaks = c(0, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, Inf))  
T1 <- table(X1)  
grouped.data <- as.numeric(T1)  
  
p <- c(pbinom(5, 30, prop.estimate), dbinom(6:12, 30, prop.estimate), 1 - pbinom(12, 30, prop.estimate))  
  
barplot(p \* 80)



### 2c)

chi1 = chisq.test(x=grouped.data, p = p)  
  
chi1

##   
## Chi-squared test for given probabilities  
##   
## data: grouped.data  
## X-squared = 5.924, df = 8, p-value = 0.6557

X.squared = unname(chi1$statistic)  
  
# recalculate the p-value using the correct degrees of freedom  
  
1 - pchisq(unname(X.squared), length(grouped.data) - 2)

## [1] 0.5486503

The p-value 0.5487 > 0.05 and so there is insufficient evidence to conclude that there is a difference between a Binomial distribution and the distribution of the number of germinations of seeds of the tested plant.

## Question 4

### 4a)

Angle = c(rep(seq(0,30,10), each=10))  
  
Panel = c(rep(rep(1:5, each = 2), 4))  
  
Power = c(42.3, 41.4, 42.2, 40.3, 37.6, 35.7, 36.8, 34.9, 45.8, 43.7, 42.1, 40.2, 42.1, 40.3, 38.4, 36.5, 38.0, 37.1, 45.2, 43.1, 42.6, 40.8, 42.7, 40.8, 38.6, 36.7, 40.2, 38.3, 46.9, 44.8, 43.6, 41.5, 43.8, 41.9, 41.9, 39.8, 42.9, 40.8, 45.4, 43.5)  
  
data = data.frame(Angle, Panel, Power)  
  
model1 = lm(Power ~ factor(Angle) + factor(Panel), data = data)  
  
anova(model1)

## Analysis of Variance Table  
##   
## Response: Power  
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(Angle) 3 36.890 12.297 5.8632 0.002614 \*\*   
## factor(Panel) 4 235.522 58.880 28.0748 4.602e-10 \*\*\*  
## Residuals 32 67.113 2.097   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Let α = the effect of angle elevation on power output.

Let β = the effect of panel type on power output.

μij = μ + αi + βj

H0A: α1 = α2 = α3 = α4 = 0

H1A: H ̅0

F value = 5.8632

p-value = 0.002614

Have assumed that: - There is no statistical interaction between the factors and thus factor effects are additive. - We have random samples drawn independently of each other from the different populations, each having a normal distribution - All populations have the same variance, σ2

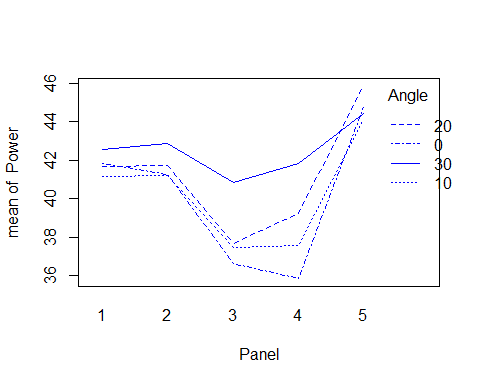
0.002614 < 0.05. Therefore, there is sufficient evidence to conclude at the 5% level of significance that the mean power output of solar panels varies between the different angles of elevation and thus that they have an effect on mean power output.

### 4b)

model2 = lm(Power ~ factor(Angle) \* factor(Panel), data = data)  
  
anova(model2)

## Analysis of Variance Table  
##   
## Response: Power  
## Df Sum Sq Mean Sq F value Pr(>F)   
## factor(Angle) 3 36.890 12.297 6.9610 0.002163 \*\*   
## factor(Panel) 4 235.522 58.880 33.3317 1.383e-08 \*\*\*  
## factor(Angle):factor(Panel) 12 31.782 2.649 1.4993 0.204458   
## Residuals 20 35.330 1.767   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

with(data, interaction.plot(Panel, Angle, Power, col = "blue"))

 0.2045 > 0.05. Therefore, there is insufficient evidence to conclude at the the 5% level of significance that there is interaction between panel type and the angle of elevation.