THEORETICAL PORTION

PROBLEM #1.

Type I Error is when the null hypothesis is rejected even though it shouldn't have been.
Type II Error is accepting the null hypothesis even though it was in fact false.

A Type I Error is more relevant to reduce in probability and can be reduced by having a lower significance level. By doing this, it broadens the span of certainty and reduces the change that the true data value is outside of that span.

A Type II Error probability is also important and can be reduced from increasing the sample size. A larger sample size will be more beneficial and accurate than a smaller sample size.

PROBLEM #2.

a. H_0: $p = p_0$, H_a: $p \neq p_0$, $\alpha = 0.01$, Z-score = 2.5758 (with $\alpha/2$), $p_0 = 39\%$
$\hat{p} = 72/150 = 0.48$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \qquad Z = \frac{0.48 - 0.39}{\sqrt{\frac{0.39*(1-0.39)}{150}}}$$

Z = 2.2599
Z is not greater than or equal to $Z_{\alpha/2}$ and Z is not less than or equal to $-Z_{\alpha/2}$
Conclusion: Fail to Reject H_0; There is not enough evidence to support that the actual percentage of the population has type A blood at the $\alpha = 0.01$ significant level.

b. H_0: $p = p_0$, H_a: $p \neq p_0$, $\alpha = 0.05$, Z-score = 1.96 (with $\alpha/2$)
Z = 2.2599
Z *is* greater than or equal to $Z_{\alpha/2}$ and Z is not less than or equal to $-Z_{\alpha/2}$
Conclusion: Reject H_0; there is sufficient evidence at the $\alpha = 0.05$ significant level to support that the actual percentage of the population has type A blood.

PROBLEM #3.

a. *n* = 20, *x* = 52 minutes, $\sigma = 7$, $\alpha = 0.05$
H_0: $\mu = \mu_0$, H_a: $\mu < \mu_0$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \qquad z = \frac{52 - 55}{\frac{7}{\sqrt{20}}}$$

z = -1.9166

Using $\Phi^{-1}(0.972) = 1.917$, gives the area up to the upper bound of the positive Z-score of 1.917. Thus 1 - 0.972 gives the p-value of 0.028. Since $p < \alpha$, we **reject H_0** because there is sufficient evidence to support the average drying time is not equal to the normal drying time.

b. The probability of type I error is equivalent to the significant level of .05, thus the probably of a **Type I Error is 5%**.

c. In order to find the type II error, the critical value of Z is calculated by Z_α when $\alpha = 0.05$ which is -1.645 ($\Phi^{-1}(0.95) = 1.645$) using multiplied by the standard deviation of 7 and adding the μ_0 of 55. This equates to $Z_{critical} = 43.4857$. The probability of a type two error is $P(Z > (Z_{critical} - \mu)/\sigma)$. Performing a CDF for $P(Z > -1.3592) = 0.08704$ [in R: pnorm(43.4857, 53, 7)]. The probably of a **Type II Error is 8.7%**.

PROBLEM #4.

a. H_0: $\lambda = 15$, H_a: $\lambda > 15$, to reject the null hypothesis, the value of c using a Poisson distribution and an type I error probability to 0.10 is 20. Concluding that $X > 20$.

b. The probability of a type II error using the cut off point of 20 from [#4,a] (since X must be 20 or greater)
<u>code:</u>

```
x_critical_.1 = (qnorm(.1) * (sqrt(20)) + 20)
type_II_.10 = ((x_critical_.1 - 25) / (sqrt(20)));pnorm(type_II_.10)
```

<u>outcome:</u>

Type II Error probability: 0.8207%

PROBLEM #5.

H_0: $\mu = 5$, H_a: $\mu > 5$

a. $n\bar{X} \sim N(n\mu, n\sigma^2)$
   $E(n\bar{X}) = nE(\bar{X}) = n\mu$
   $Var(n\bar{X}) = n^2 Var(\bar{X}) = n^2 \sigma^2/n = n\sigma^2$

b. For determining the if the null hypothesis should be rejected, a Z-score will be used since we are assuming $n > 30$. Thus:

$$Z = \frac{\sum X_i - n\mu}{n\sigma^2}$$

c. $\alpha = 0.10$, $n = 40$, $\Sigma_i X_i = 251.04$, and $s = 5.69$

$$Z = \frac{251.04 - 40(5)}{40(5.69)^2}$$

   $Z = 0.03941$

d. p-value = 0.4843
   Since the p-value is greater than the significant level (0.10), I **Fail to Reject H_0**. There is not sufficient evidence to support that the company's lightbulb lifetime is comparable to the leading brands 3 year lifetime, as at 10% significant level.

e. Power of test:
   <u>code:</u>

```
x_critical_.1 = (qnorm(.1) * (5.69/sqrt(40))) + 5
type_II_.10 = ((x_critical_.1 - 6) / (5.69));1-pnorm(type_II_.10)
x_critical_.1 = (qnorm(.1) * (5.69/sqrt(40))) + 5
type_II_.10 = ((x_critical_.1 - 7) / (5.69));1-pnorm(type_II_.10)
x_critical_.1 = (qnorm(.1) * (5.69/sqrt(40))) + 5
type_II_.10 = ((x_critical_.1 - 8) / (5.69));1-pnorm(type_II_.10)
```

   <u>outcome:</u>

   True mean 6, power = 0.6474
   True mean 7, power = 0.7103
   True mean 8, power = 0.7673

COMPUTATIONAL PORTION

Code and graphs included with each part, not indexed.

PROBLEM #1 - COMPUTATIONAL

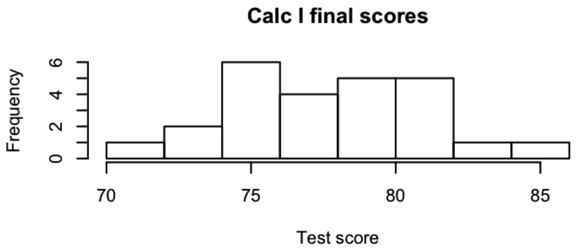
#1, part *a.*

code:

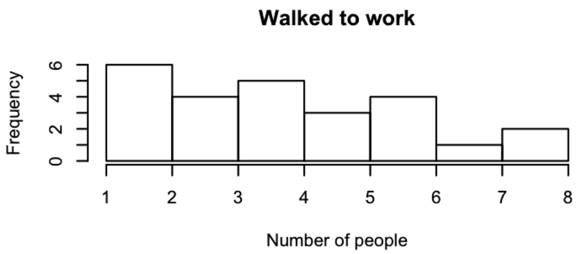
```
IIN <- read.table("isitnormal", header=T)

###FOR HISTOGRAMS###
{
par(mfrow=c(3,2))
hist(IIN[,1], main="Calc I final scores", xlab="Test score")
hist(IIN[,2], main="Store: amount of people before first purchase", xlab="Number of people")
hist(IIN[,3], main="Walked to work", xlab="Number of people")
hist(IIN[,4], main="Amount of time to finish first problem on test",xlab="Time (minutes)")
hist(IIN[,5], main="Error measurements on a well calibrated device", xlab="Error amount (mm)")
hist(IIN[,6], main="Rainfall in September", xlab="Amount of rainfall (inches)")
}

##FOR BOXPLOTS###
{
par(mfrow=c(1,6))
boxplot(IIN[,1], ylab="Calc I final scores - Test score")
boxplot(IIN[,2], ylab="Store: amount of people before first purchase - Number of people")
boxplot(IIN[,3], ylab="Walked to work - Number of people")
boxplot(IIN[,4], ylab="Amount of time to finish first problem on test - Time (minutes)")
boxplot(IIN[,5], ylab="Error measurements on a well calibrated device - Error amount (mm)")
boxplot(IIN[,6], ylab="Rainfall in September - Amount of rainfall (inches)")
}

##FOR NORMALITY###
{
par(mfrow=c(3,2))
ggplot() + qqnorm(IIN[,1], main="Calc I final scores", xlab="Normal")#; qqline(IIN[,1])
qqnorm(IIN[,2], main="Store: amount of people before first purchase", xlab="Normal"); qqline(IIN[,2])
qqnorm(IIN[,3], main="Walked to work", xlab="Normal"); qqline(IIN[,3])
qqnorm(IIN[,4], main="Amount of time to finish first problem on test", xlab="Normal"); qqline(IIN[,4])
qqnorm(IIN[,5], main="Error measurements on a well calibrated device", xlab="Normal"); qqline(IIN[,5])
qqnorm(IIN[,6], main="Rainfall in September", xlab="Normal"); qqline(IIN[,6])
}
```

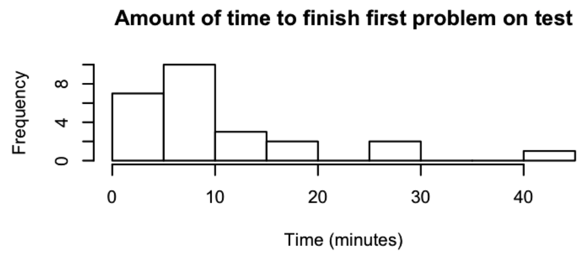
output:

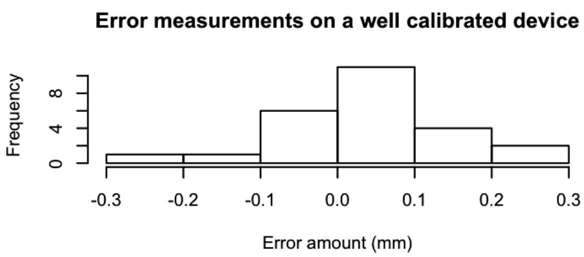
(ON NEXT PAGE)

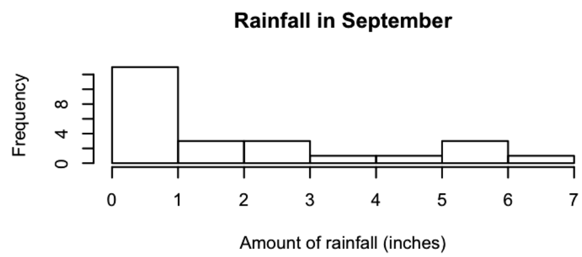

Calc I final scores
Frequency
0
2
4
6
70
75
80
85
Test score

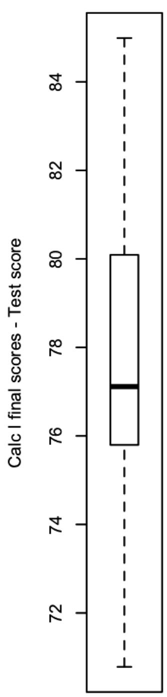

Store: amount of people before first purchase
Frequency
0
5
10
0
1
2
3
4
5
6
7
Number of people

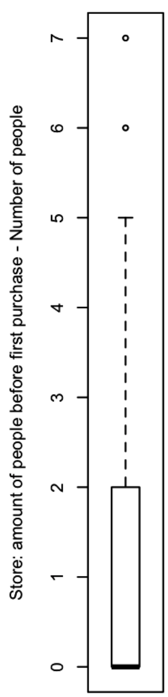

Walked to work
Frequency
0
2
4
6
1
2
3
4
5
6
7
8
Number of people

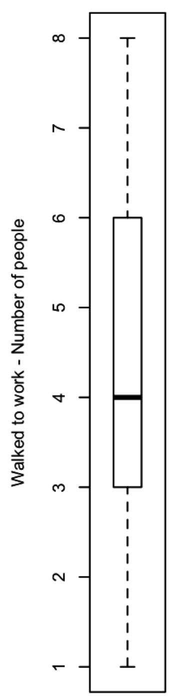

Amount of time to finish first problem on test
Frequency
0
4
8
0
10
20
30
40
Time (minutes)

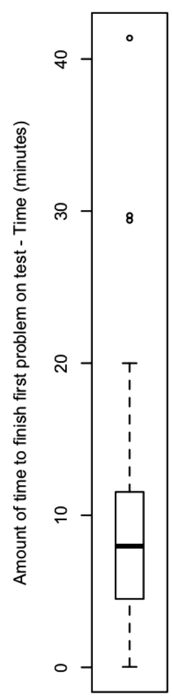

Error measurements on a well calibrated device
Frequency
0
4
8
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
Error amount (mm)

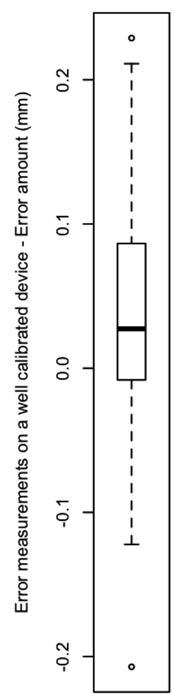

Rainfall in September
Frequency
0
4
8
0
1
2
3
4
5
6
7
Amount of rainfall (inches)

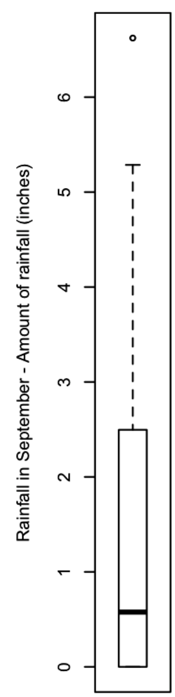

Calc I final scores - Test score
72
74
76
78
80
82
84

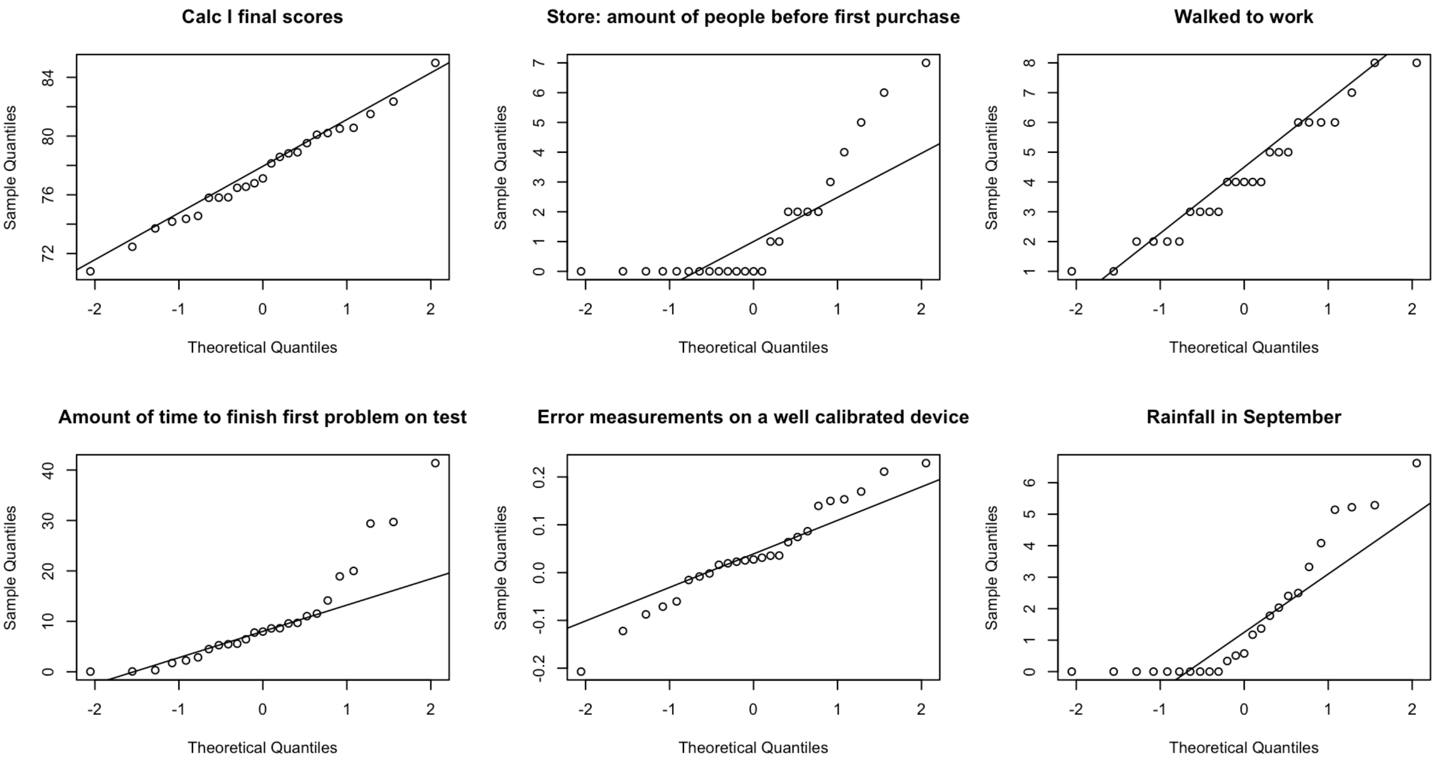

Store: amount of people before first purchase - Number of people
0
1
2
3
4
5
6
7


Walked to work - Number of people
1
2
3
4
5
6
7
8


Amount of time to finish first problem on test - Time (minutes)
0
10
20
30
40


Error measurements on a well calibrated device - Error amount (mm)
-0.2
-0.1
0.0
0.1
0.2


Rainfall in September - Amount of rainfall (inches)
0
1
2
3
4
5
6



<u>Conclusion:</u> The only column that appears to my normally distributed is the first column from comparing all three graphs above. Column 2 and 6 appear to come from an Geometric distribution OR Exponential distribution. Though column 6 could also come from a Weibull distribution. Columns 3 and 4 look to be from a Poisson distribution while the remaining column 5 looks to come from a Binomial distribution. These assumptions are all based on above graphical analysis.

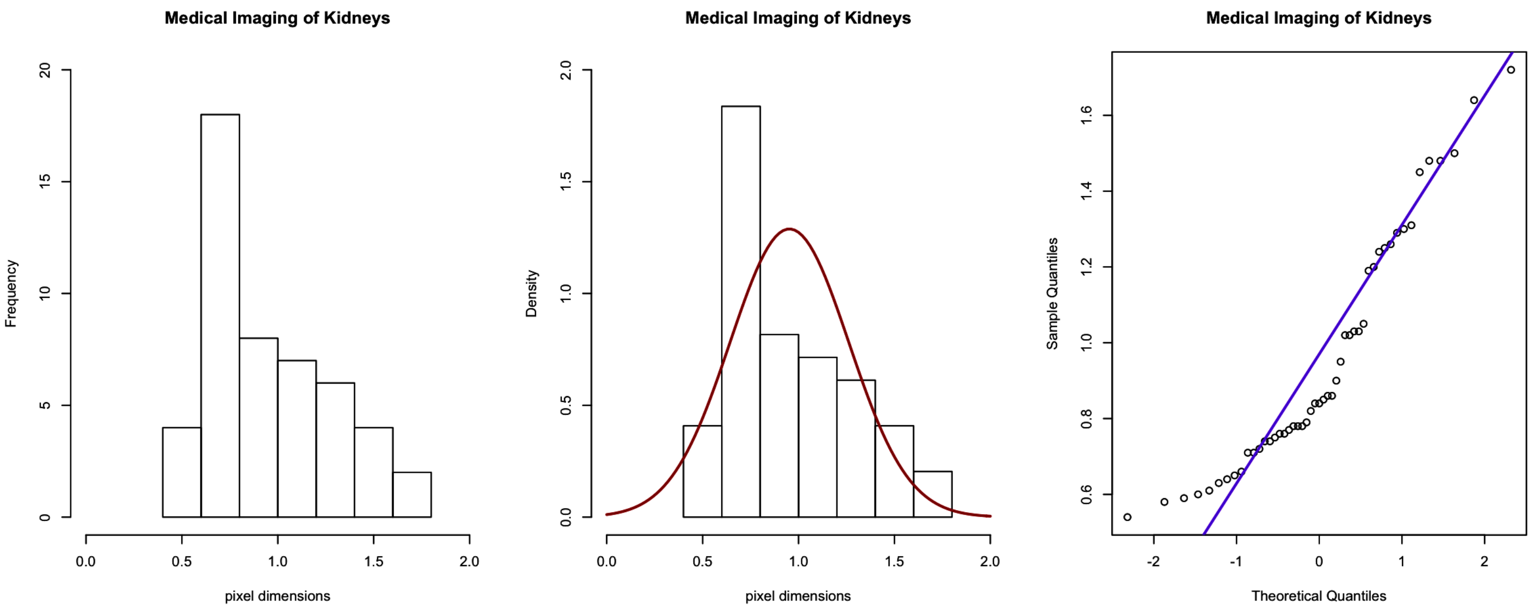
PROBLEM #2 - COMPUTATIONAL

#2, part a.

code:

```
ALD <- read.table("ALD Data", header=FALSE)
par(mfrow=c(1,3))
hist(ALD[,1], xlim=c(0,2), ylim=c(0,20),xlab="pixel dimensions", main="Medical Imaging of Kidneys")
hist(ALD[,1], xlim=c(0,2), ylim=c(0,2), xlab="pixel dimensions", main="Medical Imaging of Kidneys", freq=FALSE)
curve(dnorm(x,mean(ALD[,1]),sd(ALD[,1])), lwd=2,col="firebrick", add=TRUE)
qqnorm(ALD[,1], main="Medical Imaging of Kidneys"); qqline(ALD[,1], col="blueviolet", lwd=2)
```

output:



Conclusion: Based on the third graph, the ALD data does not seem plausible to represent a normal distribution. Normality must be assumed in order to accurately calculate the confidence interval for the true mean ALD.

#2, part b.

H_0: $\mu = 1.0$, H_a: $\mu \neq 1.0$, $\alpha = 0.05$

code:

```
z_value=(mean(ALD[,1])-1)/(sd(ALD[,1])/(sqrt(length(ALD[,1]))))
z_value
ALD_LB <- qnorm(.025,mean(ALD[,1],sd(ALD[,1])))
ALD_UB <- qnorm(.975,mean(ALD[,1],sd(ALD[,1])))
ALD_LB
ALD_UB
```

output:

```
> z_value
        [1] -1.075003
> ALD_LB
        [1] -1.094701
> ALD_UB
        [1] 2.825227
```

Conclusion: Since the z-value is greater than the lower bound, I **Fail to Reject H_0**. There is insufficient evidence to support the true average mean ALD is different than 1.0 at a 5% significance level.

#2, part c.

code:

```
p_value = 2*pnorm(-abs(z_value))
p_value
```

output:

```
> p_value
        [1] 0.2823734
```

Conclusion:

Since the p-value is greater than the significant level, it supports the conclusive statement from part b.

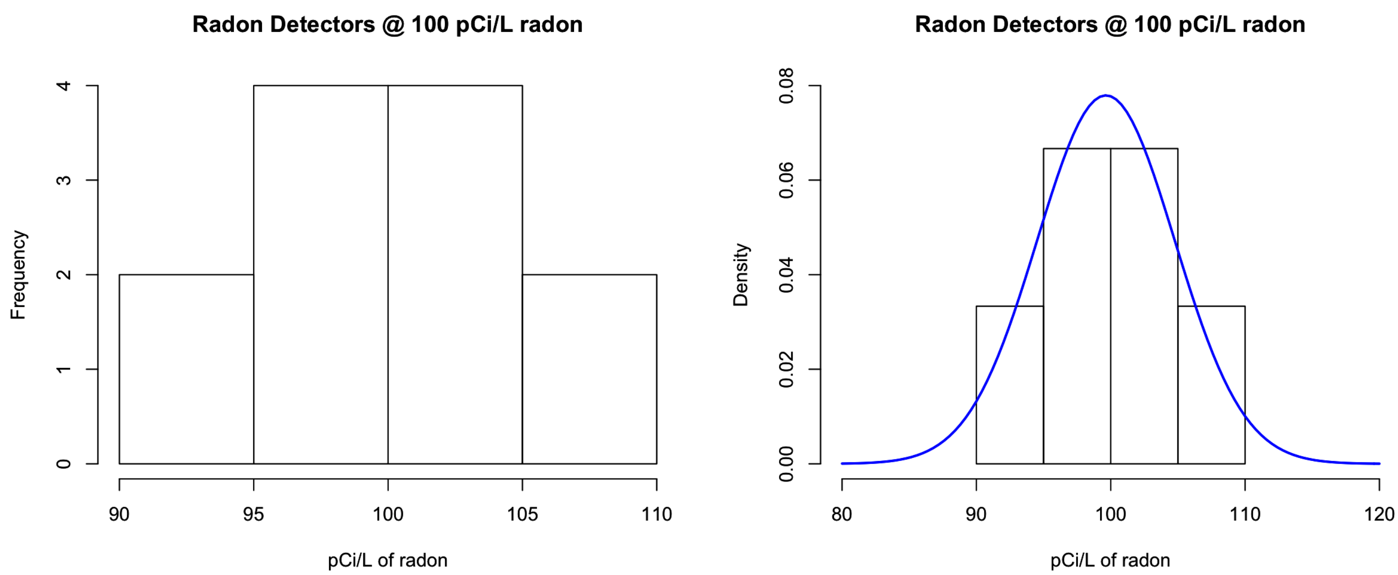
PROBLEM #3 - COMPUTATIONAL

#3, part a.

code:

```
rad_dat <- scan("radon data")
par(mfrow=c(1,2))
hist(rad_dat, xlim=c(90,110), xlab="pCi/L of radon", main="Radon Detectors @ 100 pCi/L radon")
hist(rad_dat, xlim=c(80,120), ylim=c(0,.08), xlab="pCi/L of radon", main="Radon Detectors @ 100 pCi/L radon", freq=FALSE)
curve(dnorm(x,mean(rad_dat),sd(rad_dat)), lwd=2,col="blue", add=TRUE)
```

output:



#3, part b.

H_0: $\mu = 100$, H_a: $\mu \neq 100$, $\alpha = 0.05$, degrees of freedom = 11

The data appears to be normal with the average of means closely resembling the population mean of 100 pCi/L. However, changing the number of breaks of the histogram to 6 or greater shows that it isn't as normally distributed. Also the sample size is less than 31 and the standard deviation of the population isn't know.

#3, part c.

Carrying out a t-test for calculating the p-value of the hypotheses with R coding:

code:

```
qt(c(.025,.975),df=11)
t_value=(mean(rad_dat)-100)/(sd(rad_dat)/(sqrt(12)))
t_value
p_value=2*pt(-abs(t_value), df=11)
p_value
```

output:

```
> qt(c(.025,.975),df=11)
        [1] -2.200985  2.200985
> t_value
        [1] -0.2321323
> p_value
        [1] 0.8206974
```

Conclusion: The first values show the rejection regions, the second as the test statistic, and the 3rd as the p-value. Since the p-value (0.8207) is greater than the significant level (.05), then there isn't significant evidence that the population mean reading under these conditions differs from 100 pCi/L, after being exposed to 100 pCi/L. Thus, **fail to reject H_0.**

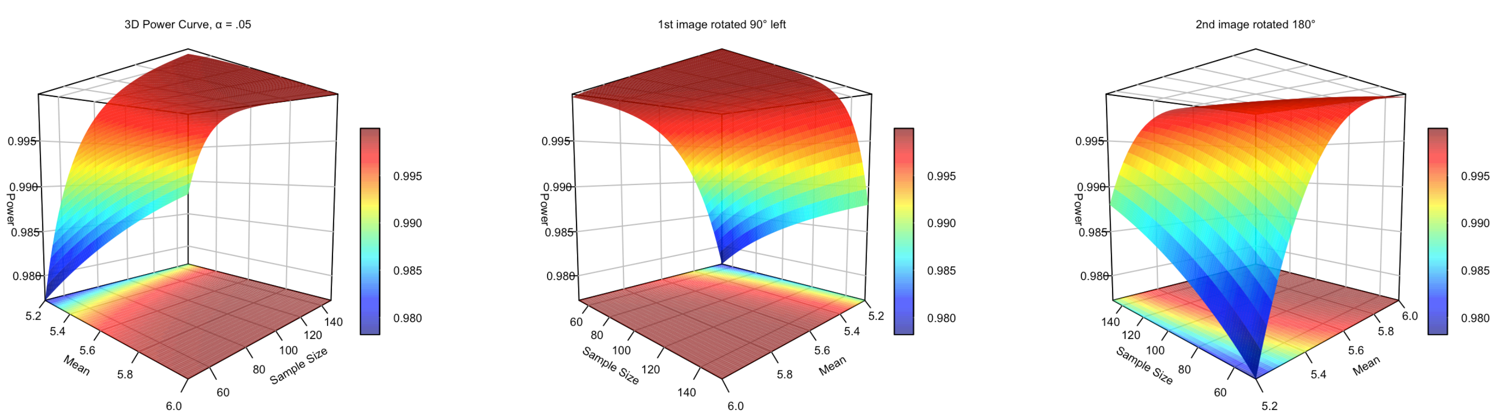
PROBLEM #4 - COMPUTATIONAL

#4, part *a combined with b.*

code:

```
n = seq(50,150,10)
glueMeans = seq(5.2,6,.2)
x_critical_.05 = (qnorm(.05) * (4/sqrt(n))) + 5
type_II_.05=matrix(0,nrow=5,ncol=11)
for (i in 1:5){
   for (j in 1:11){
      type_II_.05[i,j] = ((x_critical_.05[i] - glueMeans[i]) / (4/sqrt(n[j])))
   }
}
tAngle=45
pAngle=5
par(mfrow=c(1,3))
persp3D(glueMeans,n,1-pnorm(type_II_.05), theta=tAngle, phi=pAngle, xlab="Mean", ylab="Sample Size", zlab="Power",
         axes=TRUE, alpha=.7,bty="b2",colkey = list(length = 0.3,shift=-.02,col.ticks="white", width=1.2), image=TRUE,
         ticktype="detailed")
text(0,.6,"3D Power Curve, α = .05")
persp3D(glueMeans,n,1-pnorm(type_II_.05), theta=tAngle+90, phi=pAngle, xlab="Mean", ylab="Sample Size", zlab="Power",
         axes=TRUE, alpha=.7,bty="b2",colkey = list(length = 0.3,shift=-.02,col.ticks="white", width=1.2), image=TRUE,
         ticktype="detailed")
text(0,.6,"1st image rotated 90° left")
persp3D(glueMeans,n,1-pnorm(type_II_.05), theta=tAngle+270, phi=pAngle, xlab="Mean", ylab="Sample Size", zlab="Power",
         axes=TRUE, alpha=.7, bty="b2",colkey = list(length = 0.3,shift=-.02,col.ticks="white", width=1.2), image=TRUE,
         ticktype="detailed")
text(0,.6,"2nd image rotated 180°")
```

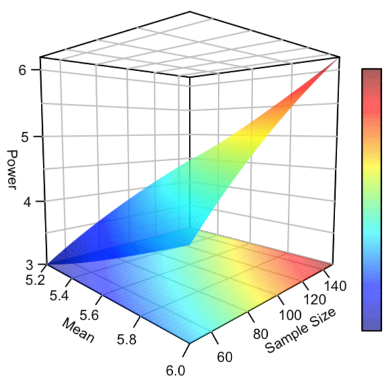
output:



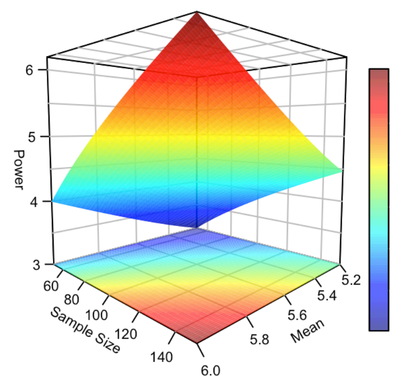
Josh Miltier

HW4- Stats

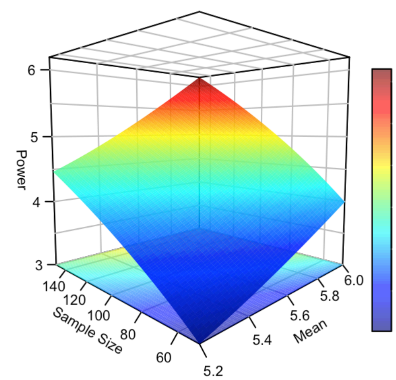
3D Power Curve, α = .05



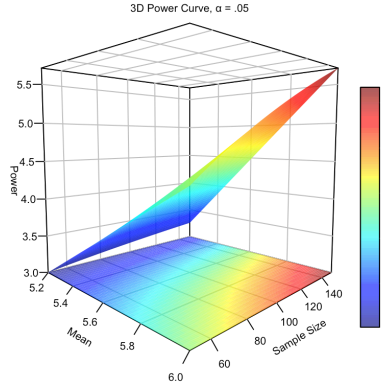
1st image rotated 90° left



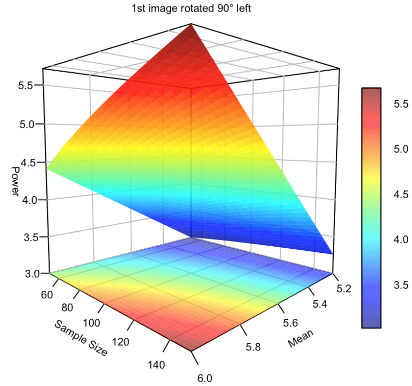
2nd image rotated 180°



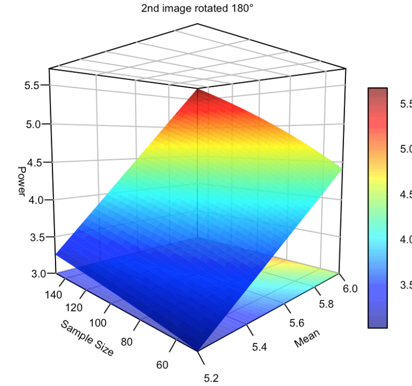
3D Power Curve, α = .05



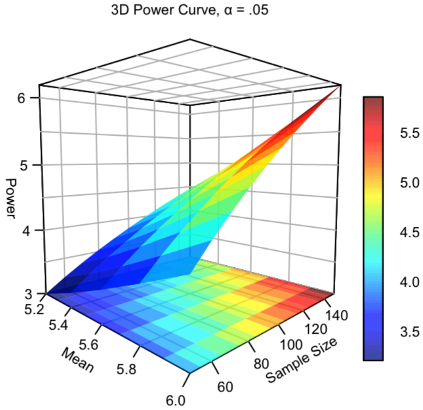
1st image rotated 90° left



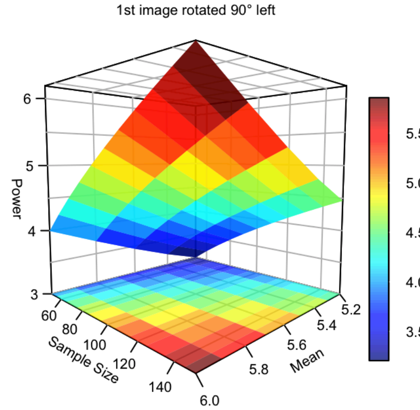
2nd image rotated 180°



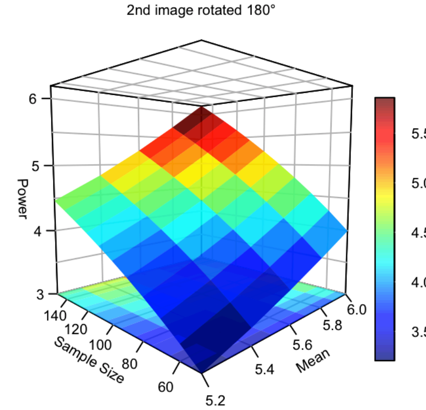
3D Power Curve, α = .05



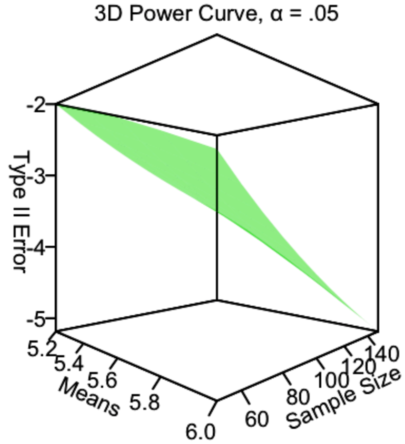
1st image rotated 90° left



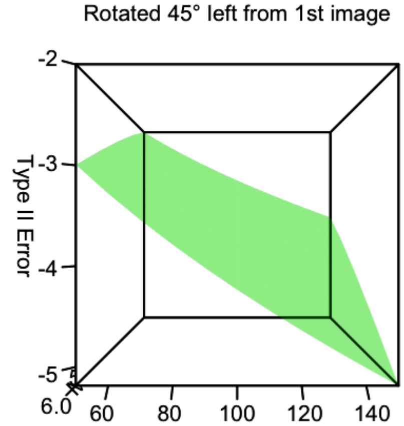
2nd image rotated 180°



3D Power Curve, α = .05



Rotated 45° left from 1st image



Rotated 90° left from 1st image