Homework 6 Soln: Hypothesis Tests for 2 Independent Populations due Fri 11/9 at 10 am

*Submit your homework to your TA's mailbox (and Canvas if you wish) before the due date/time. The mailboxes are to the left as you enter the Medical Science Center (1300 University Ave.) from the main University Ave. entrance.

*No late homework will be accepted for credit.

*If a problem asks you to use R, include a copy of the code and output. Please edit your code and output to be only the relevant portions.

*If a problem does not specify how to compute the answer, you may use any appropriate method but show the work that is necessary for the TA to follow and evluate your reasoning. I may ask you to use R or manual calculations on exam, so practice accordingly.

*Staple multiple papers together; present solutions to all prompts clearly; and put the discussion you attend on top to get full "Completion Points".

Total Points: 45 Completion Points: 10; Accuracy Points: 35

1. An eel supply company has two large vats of adolescent eels (called 'elvers'). One vat contains American eels, and the other contains European eels. An eel researcher suspects that American eels have a larger number of scales than European eels. Let μ_{Am} be the population mean for American, and μ_{Eu} be the population mean for European. The species of eels can be considered independent. The researcher visits the company and uses an ingenious method to get random samples from each vat (details omitted) and counts the number of scales on each selected elver. Suppose the data from each sample supports the assumption that both species scale counts are approximately normal. The sample statistics are summarized below:

Type	Sample Size	Sample Mean	Sample Variance
American	8	220	21
European	5	204	23

(a) Explain why a two sample t test with an equal variance assumption is appropriate to use here. List the null and alternative hypothesis.

3 points; 1 for each assumption check *Answer: The two samples are independent, and come from normally distributed populations. Besides,* $S_{Eu}/S_{Am} = \sqrt{23}/\sqrt{21} = 1.046 < 2$. *So all the three assumptions are met for 2-sample t-test with equal variance assumption.*
$H_0 : \mu_{Am} - \mu_{Eu} = 0$ *vs.* $H_A : \mu_{Am} - \mu_{Eu} > 0$

(b) At $\alpha = 0.1$, find the rejection region in the scale of t and also in the scale of $\bar{X}_{Am} - \bar{X}_{Eu}$ and use them to make a reject or not reject decision based on the observed test statistic and observed difference in means, respectively. Then make a conclusion in the context of the problem.

6 points; 2 points for RR of t; 2 points for RR of $\bar{x}_{Am} - \bar{x}_{Eu}$***; 2 points for comparing an observed value to RR and rejecting the null*** *Answer: The rejection region in the scale of* t *is* $t > t_{5+8-2,0.1} = 1.363$ *since we have a 1 sided, greater than alternative. Based on* $t = \dfrac{\bar{x}_{Am} - \bar{x}_{Eu}}{s_p * \sqrt{1/n_{Am} + 1/n_{Eu}}}$*, to reverse standardize we need to find the estimated standard error.*

Start by finding $s_p^2 = \frac{(21*7+23*4)}{11} = 21.73$. *Then the SE is* $\sqrt{21.73(\frac{1}{8}+\frac{1}{5})} = 2.66$. *Reverse standardizing gives* $1.363 = \frac{\bar{x}_{Am}-\bar{x}_{Eu}}{2.66}$ *so* $\bar{x}_{Am} - \bar{x}_{Eu} > 1.363 * 2.66 = 3.62$. *Thus the rejection region is reject if* $\bar{X}_{Am} - \bar{X}_{Eu} > 3.62$. *The observed difference was* $220 - 204 = 16$. *The observed t test statistic is* $t_{obs} = \frac{16-0}{2.66} = 6.02$ *which is far outside our rejection region. So we would reject the null. The evidence suggests American eels have more scales.*

(c) Decide the same test based on p-value compared to $\alpha = 0.1$.

2 points; 1 for pvalue; 1 for reject *Answer:* $\bar{X}_{Am} - \bar{X}_{Eu} = 220-204 = 216$ $t_{obs} = \frac{16-0}{2.66} = 6.02$*;*
$p - value = p(t_{11} > 6.02) < 0.0005$*;*
The p-value is significant at $\alpha = .10$ *since it is smaller than .10, so we reject the null.*

(d) Suppose we instead wanted to test:

$$H_0 : \mu_{Am} - \mu_{Eu} = 10$$
$$\text{vs.}$$
$$H_A : \mu_{Am} - \mu_{Eu} > 10$$

If the same data is used for both tests, would the p-value for this test be larger or smaller than the p-value that would have been computed for the test in part (a)? Explain your answer. You do not need to actually compute either p-value to answer this question.

2 points; 1 for larger; 1 point for explanation *Answer: The p-value would be larger for the test:* $H_0 : \mu_{Am} - \mu_{Eu} = 10$ *, since the observed difference of 16 is closer to the hypothesized difference for the test in (d). This makes the numerator smaller, making the t statistic closer to 0, which results in a larger p-value.*

(e) Suppose instead we choose to perform a Welch's t test. What does that mean for the assumptions we are making? Perform a Welch's T test for the same data. How much do the p values differ?

4 points; 1 for each assumption; 2 for p value; 1 similarity of p values *Answer: The Welch's t test doesn't assume equal variance. It is ok to let the populuation variances be different and do the welch's calculations. Because we are making fewer assumptions the Welch's t often gives slightly larger pvalues (lower evidence) but the p value will not be dramatically different in this case.*

Answer: For Welch's T test: $t = \frac{\bar{x_1}-\bar{x_2}}{\sqrt{s_1^2/n_1+s_2^2/n2}}$. $SE = \sqrt{s_1^2/n_1 + s_2^2/n_2} = \sqrt{21/8 + 23/5} = 2.687936$ *so* $t_{obs} = \frac{220-204-0}{2.688} = 5.952381$ *We compare this to a t distribution with* $df = \frac{[(s_1^2/n_1)+(s_2^2/n_2)]^2}{(s_1^2/n_1)^2/(n_1-1)+(s_2^2/n_2)^2/(n_2-1)}$ *We know* $s_1^2/n_1 = 21/8 = 2.625$ *and* $s_2^2/n_2 = 23/5 = 4.6$*, so* $df = \frac{(2.625+4.6)^2}{2.625^2/7+4.6^2/4} = 8.32$ *p value:* $P(t_8 > 5.9524) < .0005$

2. Data on household vehicle miles of travel (VMT) are compiled annually by the Federal Highway Administration. A researcher is interested in whether there is a difference in last year's mean MVT for midwestern and southern households. Independent random samples of 15 midwestern households and 14 southern households provided the following data on last year's VMT, in thousands of miles:

$$Midwest : 16.2, 12.9, 17.3, 14.6, 18.6, 10.8, 11.2, 16.6, 16.6, 24.4, 20.3, 20.9, 9.6, 15.1, 18.3$$

$$South : 22.2, 19.2, 9.3, 24.6, 20.2, 15.8, 18.0, 12.2, 20.1, 16.0, 17.5, 18.2, 22.8, 11.5$$

(a) Graph the data as you see fit. Why did you choose the graph(s) you did and what does it (do they) tell you? Also calculate summary statistics relevant to the research question.

ANSWER: I plotted a qqnorm plot for each distribution. From these, I can see that the population normality assumption is reasonably met. Because the research question is interested in the population means, I will compute the sample means and standard deviations $\bar{X_m} = 16.23$, $\bar{X_s} = 17.69$, $s_m^2 = 4.06^2$ *and* $s_s^2 = 4.422^2$, $n_m = 15$, $n_s = 14$. *I also plotted a side by side box plot to check the equal variance assumption.*

(b) Perform a two population t test for means at the 10% level assuming equal variance and justify why the assumptions of the test are resonably met (or describe what assumptions we are assuming are met). As part of this test, specify your hypotheses, calculate a p value and make a conclusion in the context of the question. Compute the p value just using your calculator and t table. Then find an exact p value by running the same test in R.

7 points; 2: hypheses; 2 p value by hand; 2 p value from R; 1: fail to reject the null *We can perform a two population t test for means assuming equal variance because we are assuming the two populations are independent and the observations within each population are independent. The standard deviations of the samples are also within a factor of 2, so equality of variance is reasonable. Now, performing the test:* $H_0 : \mu_M - \mu_S = 0$ *vs* $H_A : \mu_M - \mu_S \neq 0$ $t = \dfrac{\bar{x}_M - \bar{x}_S}{s_p * \sqrt{1/n_M + 1/n_S}}$. *First,* $s_p^2 = \frac{4.06^2*(14)+4.422^2*13}{15+14-2} = 484.9735/27 = 17.96$, *so* $s_p = \sqrt{17.96} = 4.238$. *So, SE* $= 4.238 * \sqrt{1/15 + 1/14} = 1.575$ *Also* $Obs_{diff} = \bar{x}_M - \bar{x}_S = 16.23 - 17.69 = -1.46$. *It follows that* $t_{obs} = \frac{-1.46-0}{1.575} = -0.9269$. *To calculate p value, we need df: df=15+14-2=27. so pvalue:* $0.30 < 2 * P(T_{27} < -0.9269) < 0.40$. *The pvalue calculated in R is 0.3622. This is insufficient evidence at the* 10% *level to reject the null; There is insufficient evidence of a difference.*

(c) Construct a 90% confidence interval for the true difference in means (by hand and then again in R). Describe how this confidence interval relates to your findings in part b.

7 points; by hand: 2 pts correct t mult; 1 pt correct SE; 2 pts final CI; 2 pts: 0 falls in so fail to reject null *ANSWER:* $(\bar{x}_M - \bar{x}_S) \pm t_{27,.05} * s\sqrt{1/15 + 1/14} = (16.23 - 17.69) \pm 1.703 * 1.575 = (-4.14, 1.22)$. *Since 0 falls within this interval, we would draw a similar reject the null conclusion.*

3. Several neurosurgeons wanted to determine whether a dynamic system (Z-plate) reduced operative time relative to a static system (ALPS plate). The operative times, in minutes, for 14 dynamic replications of the operataion and 6 static replications were obtained and are given below:

$$Dynamic : 370, 360, 510, 445, 295, 315, 490, 345, 450, 505, 335, 280, 325, 500$$

$$Static : 430, 445, 455, 455, 490, 535$$

(a) Graph the data as you see fit. Why did you choose the graph(s) you did and what does it (do they) tell you? Also calculate summary statistics relevant to the research question.

ANSWER: I plotted a qqnorm plot for each distribution. These are not great, but within possibility of what we could see sampling from a normal population by looking at qqnorm(rnorm(14)). I can see that the population normality assumption is reasonably met. Because the research question is interested in the population means, I will compute the sample means and standard deviations $\bar{X_D} = 394.64$, $\bar{X}_s = 468.33$, $s_D^2 = 84.75^2$ *and* $s_s^2 = 38.17^2$, $n_d = 14$, $n_s = 6$. *I also plotted side by side boxplots to visually examine the equal variance assumption.*

(b) Perform a two population t test for means at the 5% level not assuming equal variance and justify why the assumptions of the test are resonably met (or describe what assumptions we are assuming are met). As part of this test, specify your hypotheses, calculate a p value and make a conclusion in the context of the question. Compute the p value just using your calculator and t table. Then find an exact p value by running the same test in R.

We can perform a two population t test for means assuming unequal variance because we are assuming the two populations are independent and the observations within each population are independent. The standard deviations of the samples are also not within a factor of 2, so equality of variance is less reasonable. Now, performing the test: $H_0 : \mu_D - \mu_S = 0$ *vs* $H_A : \mu_D - \mu_S < 0$ $t = \dfrac{\bar{x}_D - \bar{x}_S}{\sqrt{s_D^2/n_D + s_S^2/n_S}}$. *First,* $SE = \sqrt{84.75^2/14 + 38.17^2/6} = 27.49$ *Also* $Obs_{diff} = \bar{x}_D - \bar{x}_S = 394.64 - 468.33 = -73.73$. *It follows that* $t_{obs} = \frac{-73.73-0}{27.49213} = -2.68$. *To calculate p value, we need df:* $df = \frac{[84.75^2/14+38.17^2/6]^2}{\frac{(84.75^2/14)^2}{13}+\frac{(38.17^2/6)^2}{5}} = 17.832 = 17$ *when rounded down. So pvalue:* $0.005 < P(T_{17} < -2.68) < 0.01$. *The pvalue calculated in R is 0.0077. This is sufficient evidence at the 5% level to reject the null; There is sufficient evidence to suggest the mean operative time is lower with the dynamic system.*

(c) Construct and interpret an 95% confidence interval for the true difference in means (by hand and then again in R) applying the same assumptions as in part b. Describe how this confidence interval relates to your findings in part b.

ANSWER: $(\bar{x}_D - \bar{x}_S) \pm t_{17,.025} * \sqrt{s_d^2/14 + s_s^2/6} = (394.6429 - 468.3333) \pm 2.110 * 27.49 = (-131.69, -15.68)$. *We are 95% confident that this interval covers the true difference in operator time. This confidence interval also suggets the mean time of the dynamic proceedure is lower than that of the static (since both enpoints are negative.)*

4. Two new mathematics learning techniques are being tested. Twenty students were randomly selected from a population. $n_A = 9$ of them were randomly assigned to use technique A, and $n_B = 11$ of them were randomly assigned to use technique B. Each student spent 30 minutes learning the technique to which they were assigned, and then were asked to complete a task. The time to complete the task was recorded, in seconds. A shorter time indicates better mastery of the task. The data are below:

$$TechniqueA : 23.1, 21.4, 31.6, 34.5, 21.9, 36.0, 30.2, 33.1, 39.5$$

$$TechniqueB : 32.7, 36.8, 39.1, 37.3, 40.3, 46.8, 41.4, 53.0, 55.6, 54.1, 28.3$$

We wish to test:

$$H_o : \mu_A - \mu_B = 0$$

vs.

$$H_A : \mu_A - \mu_B \neq 0$$

using $\alpha = 0.05$.

(a) Graph the data as you see fit. Why did you choose the graph(s) you did and what does it (do they) tell you? Also calculate summary statistics relevant to the research question.

The qqnorm plot for Technique B do not look terrible, but it and definitely the qqnorm plot for Technique B are not as linear as I would like. Also because the same sizes are small-medium sized, the CLT may not save us. Because we are interested in the mean performance, I calculated $\bar{X_A} = 30.14$, $\bar{X_B} = 42.31$, $s_A^2 = 6.57^2$, *and* $s_B^2 = 9.00^2$, $n_A = 9$, $n_B = 11$

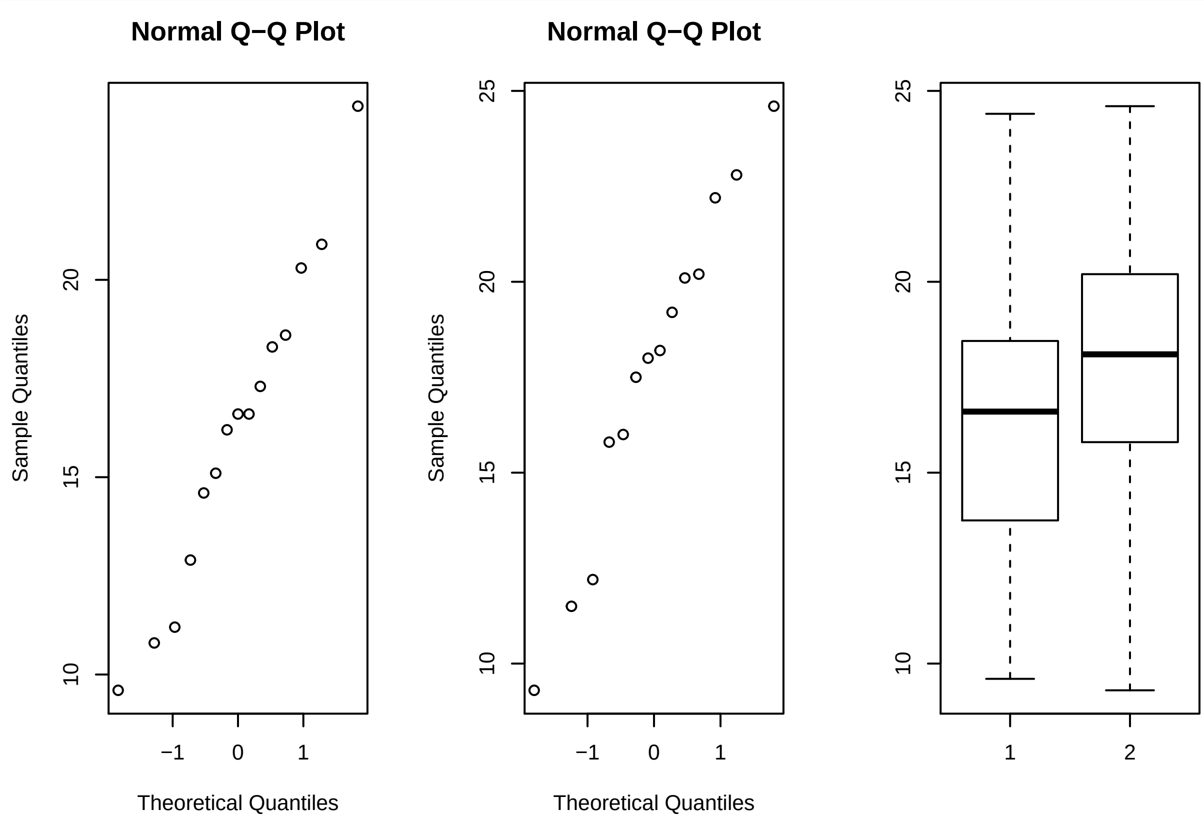
(b) Use the bootstrap to perform the test, using B = 10000 resamplings and set.seed(1). Display the histogram of your generated $\hat{t}$ values. Compute your t_{obs} and a p-value. Make a reject or not reject decision. Finally, state your conclusion in the context of the problem.

4 points; 2 for t obs; 2 for p value *ANSWER: using seed(1), I got* $t_{obs} = -3.488601$*, There were 13 that were less than or equal and 9987 that were greater than or equal, so we will take a 2 sided p value of* $2 * (13/10000) = 2 * .0013 = 0.0026$*. We have very strong evidence against the null. Evidence suggests there is a difference in the mean performance using Technique A and Technique B.*

(c) Use R to perform a two-group t test for means (i) assuming equal variance and then (ii) not assuming equal variance (Welch's T) and report the p values from each. Describe how the three p values are related. Explain how the histogram in part (b) and the summary statistics in part (a) hinted at this relationship.

ANSWER: Equal Var T p value: 0.0033; Welch's p value: 0.00266, bootstrap p value: 0.0026. We see that all of these p values are very similar. The histogram in part a closely resembled a t distribution with 18 degrees of freedom (centered at 0 and tails slightly longer than the normal). Also, because our sample variances were similar in part (a) we are not surprised that making that assumption (in equal var test) or not (in welch's and bootstrap) didn't change our results much.

```
#Question 2:
Midwest<-c(16.2, 12.9, 17.3, 14.6, 18.6, 10.8, 11.2, 16.6, 16.6, 24.4, 20.3, 20.9, 9.6, 15.1, 18.3)
South<-c(22.2, 19.2, 9.3, 24.6, 20.2, 15.8, 18.0, 12.2, 20.1, 16.0, 17.5, 18.2, 22.8, 11.5)
par(mfrow=c(1,3))
qqnorm(Midwest)
qqnorm(South)
boxplot(Midwest, South)
```




```
mean(Midwest); mean(South)
```

```
## [1] 16.22667

## [1] 17.68571
```

```r
sd(Midwest); sd(South)
```

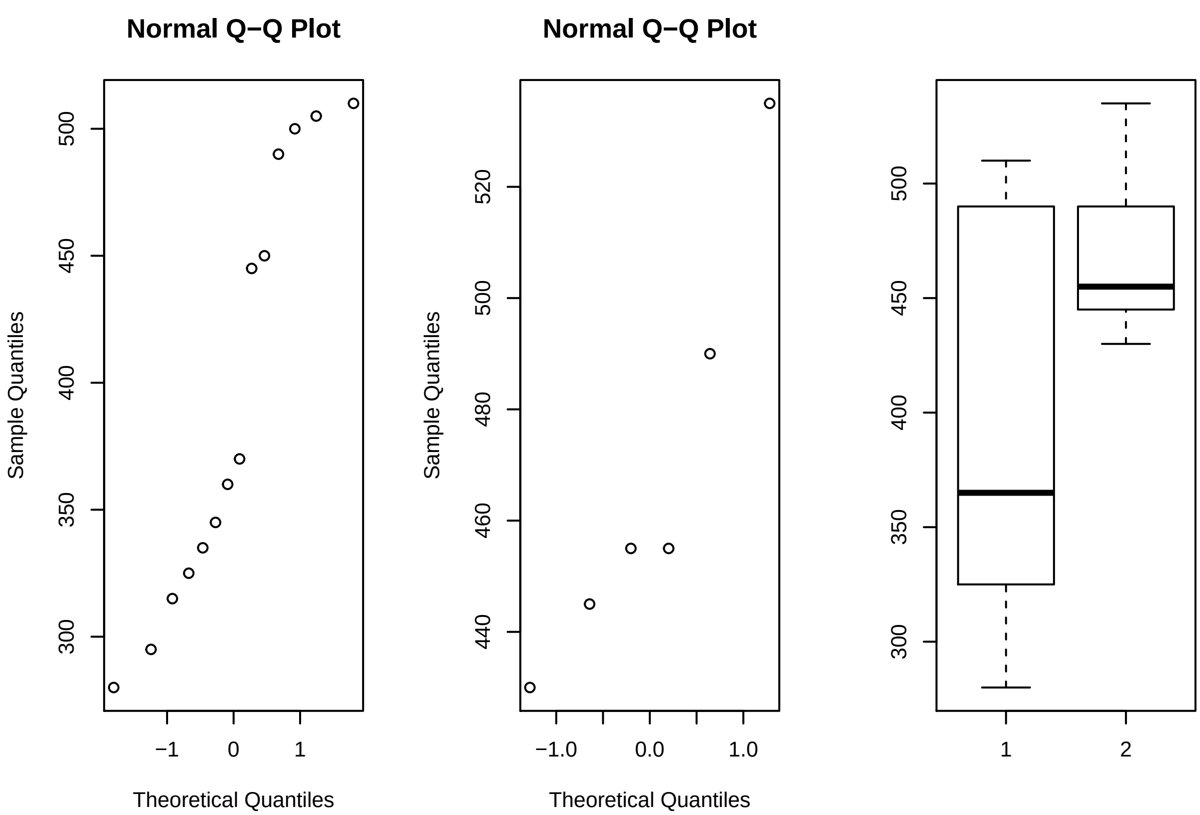
```
## [1] 4.055062

## [1] 4.42247
```

```r
t.test(Midwest, South, paired=FALSE, var.equal=TRUE, conf.level=.90)
```

```
##
##  Two Sample t-test
##
## data:  Midwest and South
## t = -0.92689, df = 27, p-value = 0.3622
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  -4.140237  1.222142
## sample estimates:
## mean of x mean of y
##  16.22667  17.68571
```

```r
#Question 3:
Dynamic<-c(370, 360, 510, 445, 295, 315, 490, 345, 450, 505, 335, 280, 325, 500)
Static<-c(430, 445, 455, 455, 490, 535)
qqnorm(Dynamic)
qqnorm(Static)
boxplot(Dynamic, Static)
```



```r
md<-mean(Dynamic); ms<-mean(Static)
md-ms
```

```
## [1] -73.69048
```

```r
sd(Dynamic); sd(Static)
```

```
## [1] 84.74996
```

```
## [1] 38.1663
```

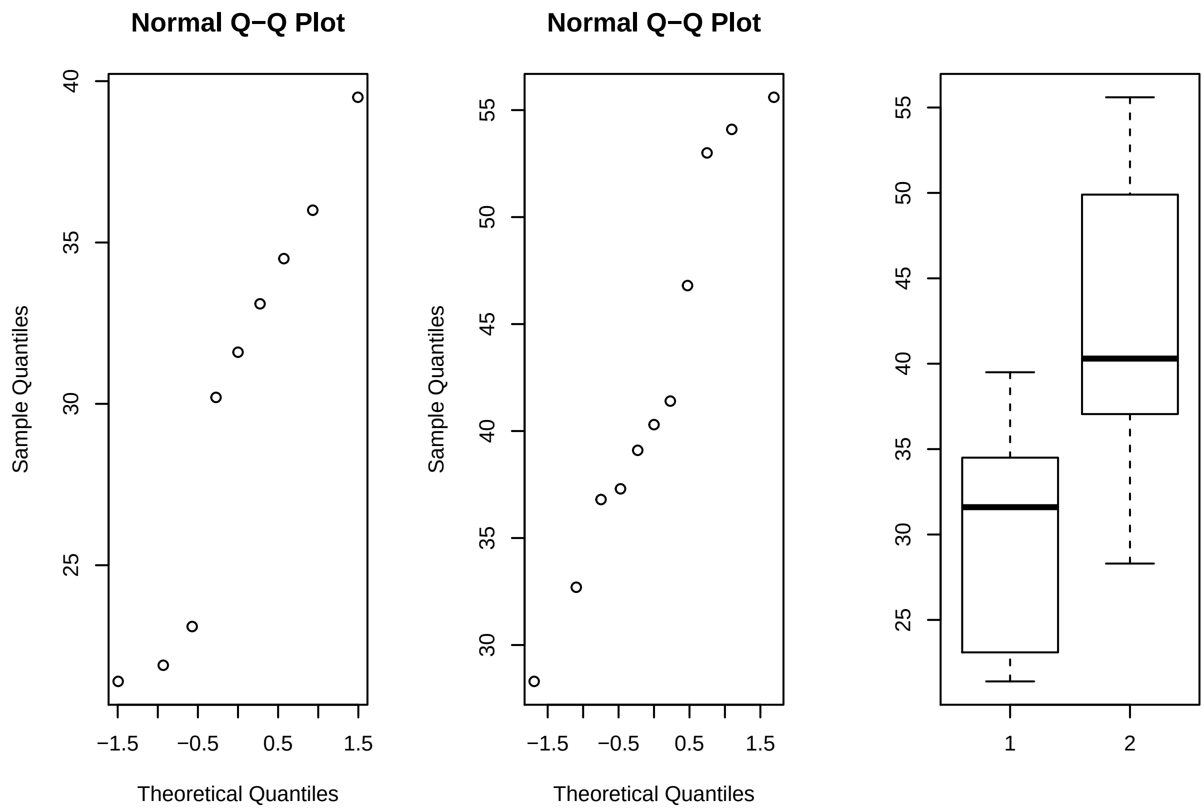
```r
t.test(Dynamic, Static, paired=FALSE, var.equal=FALSE, conf.level=.95, alternative="less")
```

```
##
##  Welch Two Sample t-test
##
## data:  Dynamic and Static
## t = -2.6804, df = 17.832, p-value = 0.007679
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -25.99307
## sample estimates:
## mean of x mean of y
##  394.6429  468.3333
```

```r
t.test(Dynamic, Static, paired=FALSE, var.equal=FALSE, conf.level=.95, alternative="two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  Dynamic and Static
## t = -2.6804, df = 17.832, p-value = 0.01536
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -131.48826  -15.89269
## sample estimates:
## mean of x mean of y
##  394.6429  468.3333
```

```r
#Question 4:
TechA<-c(23.1, 21.4, 31.6, 34.5, 21.9, 36.0, 30.2, 33.1, 39.5)
TechB<-c(32.7, 36.8, 39.1, 37.3, 40.3, 46.8, 41.4, 53.0, 55.6, 54.1, 28.3)
qqnorm(TechA)
qqnorm(TechB)
boxplot(TechA, TechB)
```




```
mean(TechA); mean(TechB)
```

```
## [1] 30.14444
## [1] 42.30909
```

```
sd(TechA); sd(TechB)
```

```
## [1] 6.573643
## [1] 8.996272
```

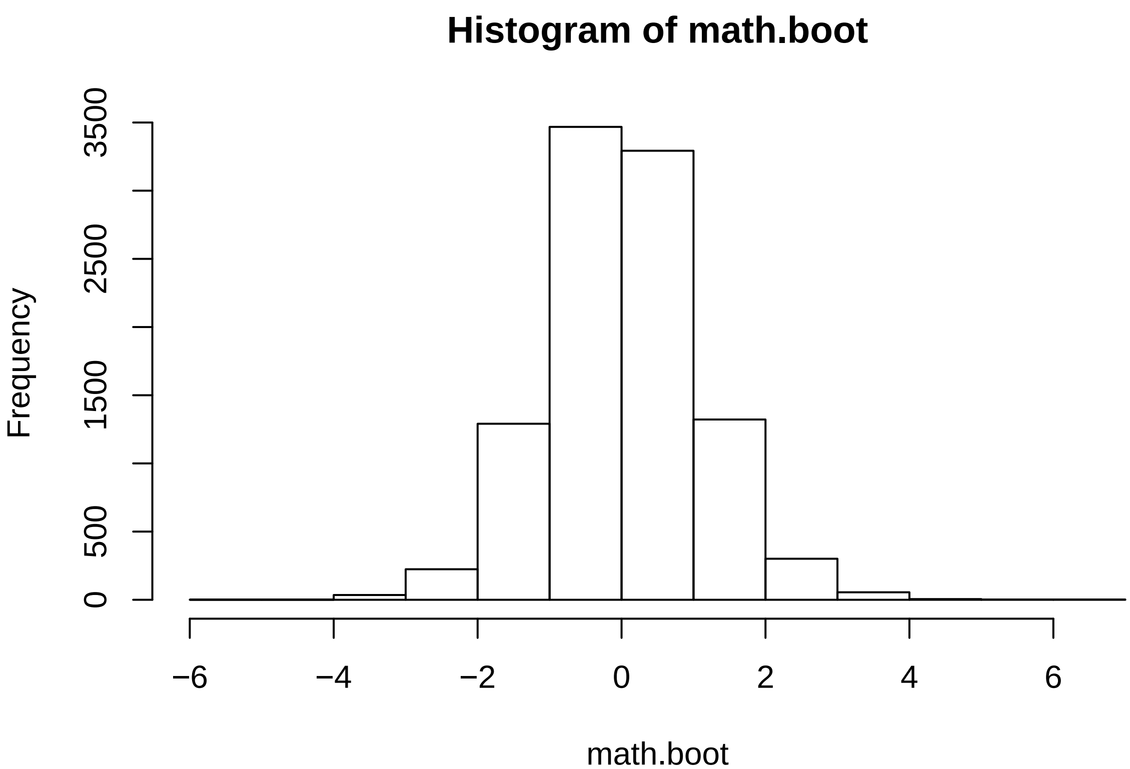
```
length(TechA); length(TechB)
```

```
## [1] 9
## [1] 11
```

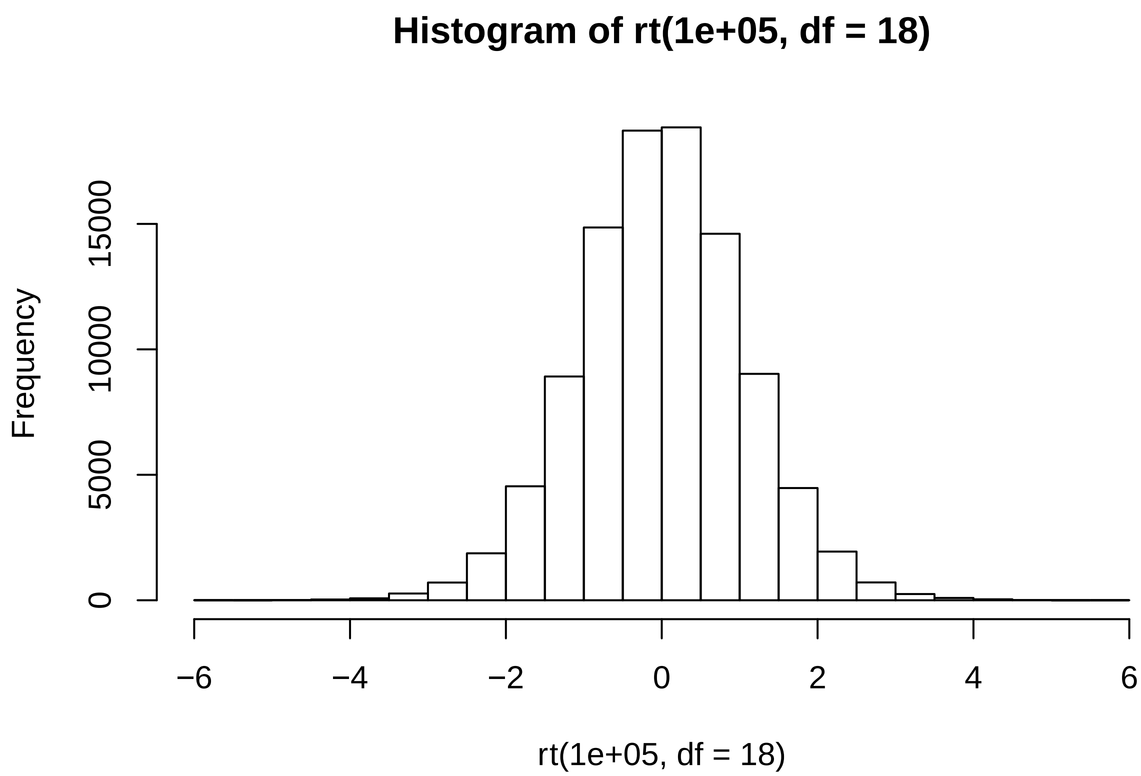
```
#Bootstrap Code:
# Here's one way to do the bootstrap for a difference of two means in R:
# dat1 and dat2 are data from the two groups. nboot is the number of resamples.
#Notice obsdiff is computed as mean(dat1)-mean(dat2) - order matters!
boottwo = function(dat1, dat2, nboot) {
  bootstat = numeric(nboot)
  obsdiff = mean(dat1) - mean(dat2)
  n1 = length(dat1)
  n2 = length(dat2)
  for(i in 1:nboot) {
    samp1 = sample(dat1, size = n1, replace = T)
    samp2 = sample(dat2, size = n2, replace = T)
    bootmean1 = mean(samp1)
    bootmean2 = mean(samp2)
```

```
    bootvar1 = var(samp1)
    bootvar2 = var(samp2)
    bootstat[i] = ((bootmean1 - bootmean2) - obsdiff)/sqrt((bootvar1/n1) + (bootvar2/n2))
}
  return(bootstat)
}

B = 10000
set.seed(1)
math.boot = boottwo(TechA, TechB, B)  #Notice TechA put in first for bootstrap and t. obs computation
par(mfrow=c(1,1))
hist(math.boot)
```




```
hist(rt(100000, df=18)) #This is giving a histrogram of T_18
```




```
(t.obs = (mean(TechA) - mean(TechB)) /
         sqrt(var(TechA) / length(TechA)  + var(TechB) / length(TechB)))  #-3.488601
```

```
## [1] -3.488601
```

```
(m.low  = sum(math.boot <= t.obs)) #13
```

```
## [1] 13
```

```
(m.high = sum(math.boot >=t.obs)) #9987
```

```
## [1] 9987
```

```
summary(math.boot>t.obs)
```

```
##    Mode   FALSE    TRUE
## logical      13    9987
```

```
(p.val = 2*m.low/ B)
```

```
## [1] 0.0026
```

```
t.test(TechA, TechB, var.equal=TRUE, alternative="two.sided")
```

```
##
##  Two Sample t-test
##
## data:  TechA and TechB
## t = -3.3786, df = 18, p-value = 0.003346
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -19.72893  -4.60036
## sample estimates:
```

```
## mean of x mean of y
##  30.14444  42.30909
```

```r
t.test(TechA, TechB, var.equal=FALSE, alternative="two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  TechA and TechB
## t = -3.4886, df = 17.823, p-value = 0.002654
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -19.495724  -4.833569
## sample estimates:
## mean of x mean of y
##  30.14444  42.30909
```