

Quick note for any markers: my handwriting and photography sucks, so at any point you find difficulty reading my real answers, look at the Appendix as it includes typed answers (for the questions that are handwritten). These are identical to the handwritten ones, just easier to understand. If you have no difficulties understanding my writing, then please ignore this message.

MXB107 Assessment 2

Harry Wright, N11085614

Semester 2 2022

NOTE THIS ASSESSMENT IS DUE ON 30 October BY 11:59 PM.

Question 1

The geometric distribution describes a random variable X : the number of Bernoulli trials required before the first success, e.g. Let X be the number of coin tosses needed to obtain one head.

The probability mass function for the geometric distribution is

$$p(x) = (1 - p)^{x-1}p$$

where

$$E(X) = \frac{1}{p}$$

- a. Find the Method of Moments estimator of p . (2 points)

Type Your Answer Here:

Handwritten solution for Question 1a:

1. a) $p(x) = (1-p)^{x-1}p$

$E(X) = \frac{1}{p} \rightarrow \frac{1}{E(X)} = p$

$m_1 = \frac{1}{n} \sum_{i=1}^n x_i$

$\hat{p} = \frac{1}{m_1} = \frac{1}{\sum_{i=1}^n x_i}$

- b. Find the Maximum Likelihood estimator of p . (3 points)

Type Your Answer Here:

b)

$$f(x) = (1-p)^{x_i-1} p$$

We define likelihood function as:

$$L(p) = \prod_{i=1}^n (1-p)^{x_i-1} p$$

Then we define log-likelihood function as:

$$\begin{aligned} \log L(p) &= \log \left(\prod_{i=1}^n (1-p)^{x_i-1} p \right) \\ &= \sum_{i=1}^n \log [(1-p)^{x_i-1} p] \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^n \log ((1-p)^{x_i-1}) + \sum_{i=1}^n \log (p) \\ &= (x_i-1) \sum_{i=1}^n \log (1-p) + \sum_{i=1}^n \log p \end{aligned}$$

$$\text{We take } \frac{d \log L(p)}{dp} = 0$$

$$\text{Which is } \frac{d}{dp} \log L(p) = \frac{\sum_{i=1}^n x_i - n}{1-p} + \frac{n}{p} = 0$$

$$\text{Solve for } \hat{p}: \sum_{i=1}^n \frac{x_i - n}{1-p} = \frac{n}{p}$$

$$\sum x_i - n = \frac{n-p}{p}$$

$$\sum x_i - n = \frac{n}{p} - n$$

$$\hat{p} = \frac{n}{\sum x_i}$$

Question 2

30 students in a chemistry class each performed an experiment measuring the amount of copper (Cu) recovered from a solution of copper sulfate CuSO_4 . The sample mean and standard deviation of the results from each of the 30 experiments is $\bar{x} = 0.145$ mol and $s = 0.0051$ mol.

Find a 90% confidence interval for the mean amount of copper (Cu) recovered from the experiment.

(3 points)

Type Your Answer Here:

2.

Since the sample size is 30, we have a passable large sample size.

Given a confidence interval of 90% and also two sided, $Z_{\alpha/2} = 1.645$

$$\left(\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$
$$= \left(0.145 - 1.645 \frac{0.0051}{\sqrt{30}}, 0.145 + 1.645 \frac{0.0051}{\sqrt{30}} \right)$$
$$= (0.1435, 0.1465)$$

Question 3

A random sample of 130 individuals recorded an average body temperature of 36.8C with a standard deviation of 0.41C. Traditional data indicate that "normal" human body temperature is 37C.

- a. Does this experimental data provide sufficient evidence to reject the null hypothesis that "normal" human body temperature is 37C? Assume a Type I error rate of $\alpha = 0.01$. (3 points)

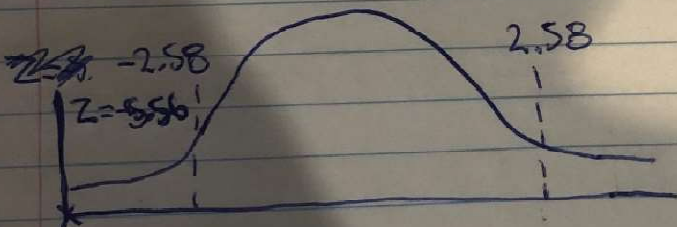
Type Your Answer Here:

3.

a) Since the sample size is 130, we have a large sample size. This means we can take the null hypothesis as:

$$H_0: \mu = 37 \quad H_A: \mu \neq 37$$

Where μ is the average of ^{human} ~~human~~ body temperature. Given a Type I error of ~~$\alpha = 0.01$~~ and two-sided $Z_{\alpha/2} = 2.58$

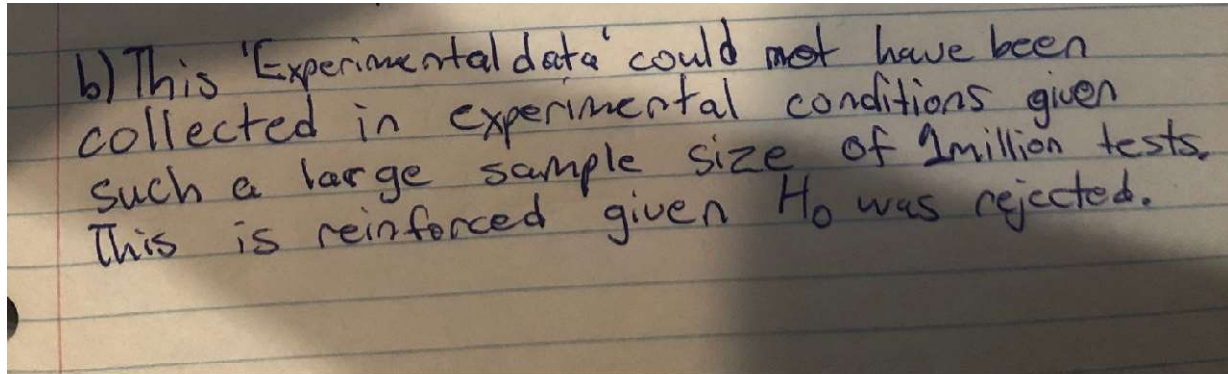
$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$
$$Z = \frac{36.8 - 37}{0.41 / \sqrt{130}} = -5.56$$


Since $Z = -5.56$ is within the reject region, from the plot above we reject H_0 .

- b. The 37C standard was derived in 1868 by a German doctor who claimed it was based on a sample of 1 million temperatures recorded throughout their research. What results can you draw about this research based on the results of your hypothesis test? (2 points)

Type Your Answer Here:

This 'experimental data' couldn't have been conducted in experimental conditions given such a huge sample size of 1 million tests. This can also be backed up given the rejected H_0 .



Question 4

Polling of the marginal state seats of Currumbin, Mansfield and Aspley by YouGov for the Australian Conservation Foundation shows a combined two-party preference, based on a survey of 600, individuals of 52%-48% for Labor versus the LNP, compared with an almost exact 50%-50% for these three seats in 2017.

Is there evidence to reject the null hypothesis that the proportion of voters preferring Labor over the LNP is less than or equal to 50%? Assume a Type I error rate of $\alpha = 0.05$. (4 points)

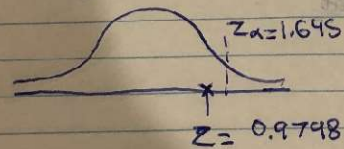
Type Your Answer Here:

4. Since the sample size is 600, we have a large sample size. This means we can take the null hypothesis as

$$H_0: p_0 \leq 0.5, H_A: p_0 > 0.5$$

Where p_0 is the proportion of voters who prefer ~~Labor~~ over ~~LNP~~ Labor LNP

Given a Type I error of $\alpha = 0.05$ and given its a one-sided test,

$$Z_{\alpha} = 1.645$$
$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.52 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{600}}}$$
$$Z = 0.9798$$


Given the plot above, we fail to reject the null hypothesis

Question 5

To compare the performance of two swimmers, their 100m freestyle times were recorded independently of one and other at random 10 times during both practice and competition.

Swimmer 1	Swimmer 2
59.62	59.81
59.74	59.41
59.48	59.32
59.43	59.63
59.65	59.76
59.72	59.50
59.50	59.64
59.63	59.83
60.01	59.86
59.68	59.51

- a. Compute a 95% confidence interval for the difference in the average times for Swimmer 1 and Swimmer 2. (3 points)

Type your Answer Here:

```
swimmer1 =  
c(59.62,59.74,59.48,59.43,59.65,59.72,59.5,59.63,60.01,59.68)  
swimmer2 =  
c(59.81,59.41,59.32,59.63,59.76,59.5,59.64,59.83,59.86,59.51)
```

```
xbar1 <- mean(swimmer1)  
xbar2 <- mean(swimmer2)
```

```
s1 <-sd(swimmer1)  
s2 <-sd(swimmer2)
```

```
vasw1 <- var(swimmer1)  
vasw2 <- var(swimmer2)
```

```
> xbar1  
[1] 59.646
```

```
> xbar2  
[1] 59.627
```

```
> s1  
[1] 0.1648029
```

```
> s2  
[1] 0.1879746
```

```
> vasw1  
[1] 0.02716
```

```
> vasw2  
[1] 0.03533444
```

5.

a)

First, check for unequal variance.
if $\frac{\text{larger } s^2}{\text{smaller } s^2} > 3$. $\frac{0.0533}{0.0276} = 1.93 < 3$

No case of unequal variance, use normal T distribution given a small sample size of $10 < 30$.

Use T-distribution to calculate the 95% confidence interval:

$$CI = (\bar{x}_1 - \bar{x}_2) \pm \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$= (59.646 - 59.627) \pm \sqrt{s_p^2 \left(\frac{1}{10} + \frac{1}{10} \right)}$$

$$s_p^2 = \frac{(10-1)s_1^2 + (10-1)s_2^2}{10+10-2} = \frac{9(0.1648029)^2 + 9(0.1879746)^2}{18}$$

$$s_p^2 \approx 0.03124$$

Using $v = n_1 + n_2 - 2$ to find the degrees of freedom, $v = 18$.

$$t_{18, 0.95} = 1.734$$

Therefore, the 95% confidence interval is

$$((\bar{x}_1 - \bar{x}_2) - \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, (\bar{x}_1 - \bar{x}_2) + \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)})$$

$$= (-0.1181, 0.1561)$$

- b. Perform a hypothesis test of the difference between the two swimmers average 100m freestyle times. Is there evidence that Swimmer 2 is, on average faster than Swimmer 1? Assume a Type I error rate of $\alpha = 0.05$. (2 points)

Type Your Answer Here:

(Uses same code for some variables from previous part)

b) Using t -distribution we define the null hypothesis as:

$$H_0: \mu_1 - \mu_2 \leq 0 \quad H_A: \mu_1 - \mu_2 > 0$$

where μ_1 is the average of swimmer 1 and μ_2 is the average of swimmer 2.

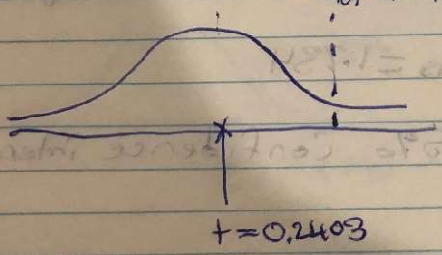
$$t = \frac{\mu_1 - \mu_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} t_1$$

$$s_p^2 = 0.03124$$

$$t = \frac{59.646 - 57.627}{\sqrt{0.03124 \left(\frac{1}{10} + \frac{1}{10} \right)}}$$

$$t \approx 0.2403$$

Given a Type I error of $\alpha = 0.05$,
Determine
Using $t_{18, 0.95} = 1.734$
 $t_{18, 0.95} = 1.734$



Since $t = 0.2403$ is not within the reject region, we fail to reject the null hypothesis that swimmer 2 is on average faster than swimmer 1.

Question 6

To test the effects of alcohol on reaction times seven individuals participated in an experiment where their reaction times were measured using the same means both before and after ingesting 90 millilitres of 40% alcohol.

Person	Before	After
1	4	7
2	5	8
3	5	3
4	4	5
5	3	4
6	6	5
7	2	5

Does mean reaction time increase after consuming alcohol? Use a Type I error rate of $\alpha = 0.05$. (3 points)

Type Your Answer Here:

```
Before = c(4,5,5,4,3,6,2)
```

```
After = c(7,8,3,5,4,5,5)
```

```
xbar1 <- mean(Before)
```

```
xbar2 <- mean(After)
```

```
s1 <- sd(Before)
```

```
s2 <- sd(After)
```



```
> xbar1  
[1] 4.142857
```

```
> xbar2  
[1] 5.285714
```

```
> s1  
[1] 1.345185
```

```
> s2  
[1] 1.704336
```

6.

Since the sample size is $7 < 30$ we have a small sample size. t -distribution.

We can take the null hypothesis as:

$$H_0: \mu_1 - \mu_2 = 0 \text{ and } H_A: \mu_1 - \mu_2 \neq 0$$

Where μ_1 is the mean reaction time before consuming alcohol and μ_2 is the mean reaction time after consuming alcohol.

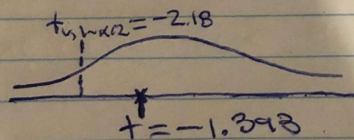
$$t = \frac{\mu_1 - \mu_2}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{4.142857 - 5.285714}{\sqrt{s_p^2(\frac{1}{10} + \frac{1}{10})}}$$

$$s_p^2 = \frac{(7-1)s_1^2 + (7-1)s_2^2}{7+7-2} = 2.35714$$

$$t = \frac{4.142857 - 5.285714}{\sqrt{2.35714(\frac{1}{7} + \frac{1}{7})}}$$

$$t = -1.393$$

Given Type I error of $\alpha = 0.05$, $t_{\alpha/2, 12} = q(0.975, 12) = 2.18$



Since $t = -1.393$ is not within the reject region, we fail to reject the null hypothesis that reaction time increases after consuming alcohol.

Question 7

The warpbreaks data set contains the breaks in yarn during weaving. We wish to examine factors that lead to breaks in the yarn.

- Perform a single factor ANVOA considering the type of wool, is there a significant difference in the number of breaks by type of wool? Use Tukey's Honest Significant Differences to identify which (if any pairs are different). (2 points)

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
wool	1	450.7	450.7	2.668	0.1084
Residuals	52	8782	168.9	NA	NA

- wool:**

	diff	lwr	upr	p adj
B-A	-5.778	-12.88	1.32	0.1084

Type Your Answer Here:

```
data("warpbreaks")

lmwarp <- lm(breaks~wool,warpbreaks)
anovawarp <- lmwarp%>%anova()
anovawarp

(lmwarp%>%aov())%>%TukeyHSD()
```

Analysis of Variance Table

```
Response: breaks
      Df Sum Sq Mean Sq F value Pr(>F)
wool    1  450.7   450.67   2.6684 0.1084
Residuals 52 8782.1   168.89
```

Fit: aov(formula = .)

```
$wool
      diff      lwr      upr      p adj
B-A -5.777778 -12.87523  1.319679 0.1083971
```

Given there is only one pair, there will not be any significant differences in the types of wool. Further, assuming a Type I error of $\alpha = 0.05$, given the model's p adj value of 0.1084, it means that the type of wool has no significant effect on the number of breaks statistically.

- b. Repeat the ANOVA but now block on the tension of the loom (L,M,H), does this change the results? (3 points)

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
wool	1	450.7	450.7	3.339	0.07361
tension	2	2034	1017	7.537	0.001378
Residuals	50	6748	135	NA	NA

- **wool:**

	diff	lwr	upr	p adj
B-A	-5.778	-12.13	0.5729	0.07361

Type Your Answer Here:

```
lmwarp2 <- lm(breaks~wool+tension,warpbreaks)
anovawarp2 <- lmwarp2%>%anova()
anovawarp2

(lmwarp2%>%aov()%>%TukeyHSD())
```

Analysis of variance Table

```
Response: breaks
      Df Sum Sq Mean Sq F value    Pr(>F)
wool    1  450.7   450.67    3.3393 0.073614 .
tension  2 2034.3  1017.13    7.5367 0.001378 **
Residuals 50 6747.9   134.96
```

```
Fit: aov(formula = .)
```

```
$wool
      diff      lwr      upr      p adj
B-A -5.777778 -12.12841  0.5728505 0.0736137
```

```
$tension
      diff      lwr      upr      p adj
M-L -10.000000 -19.35342 -0.6465793 0.0336262
H-L -14.722222 -24.07564 -5.3688015 0.0011218
H-M  -4.722222 -14.07564  4.6311985 0.4474210
```

The upper 95% confidence interval value has now decreased from 1.320 to 0.5729 and the lower value has increased from -12.88 to -12.13. Further, the p-value (for wool) has decreased from 0.1084 to 0.07361, but it is still not statistically significant at the Type I error rate of $\alpha = 0.05$ to effect the model.

- c. For both models, use Tukey's HSD to identify which pairs of types of wool (if any) are statistically significantly different. (2 points)

```
## [1] "Model 1 Tukey HSD"
```

- **wool:**

	diff	lwr	upr	p adj
B-	-	-	1.32	0.1084
A	5.778	12.88		

```
## [1] "Model 2 Tukey HSD"
```

- **wool:**

	diff	lwr	upr	p adj
B-	-	-	0.5729	0.07361
A	5.778	12.13		

Type Your Answer Here:

Statistically, neither model is statistically significant as their p-adj (p-value adjusted for pairs) are greater than the type I error of 0.05.

Question 8

The Loblolly dataset contains information on the growth of Loblolly pine trees.

- a. Fit a one-way ANOVA model of the height as a function of Seed, does the type of seed have a statistically significant effect on height? (2 points)

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Seed	13	163.7	12.59	0.02496	1
Residuals	70	35310	504.4	NA	NA

Type Your Answer Here:

```
data("Loblolly")
lmLob <- lm(height~Seed,Loblolly)
anovaLob <- lmLob%>%anova()
```

Analysis of variance Table

```
Response: height
      Df Sum Sq Mean Sq F value Pr(>F)
Seed   13    164   12.59   0.025    1
Residuals 70 35310  504.43
```

Consider that if $\text{Pr}(>F)$ (also known as p-value) is *LESS* than 0.05 for this One-way ANOVA, it means the value in the model is significant. Consider the $\text{Pr}(>F)$ for Seed is equal to 1, which means the type of seed does not have a statistically significant effect on height. This makes sense given that all the seeds are from the same breed of tree.

- b. Use Tukey's HSD to identify which (if any) pairs of seeds differ significantly (2 points)

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = .)
##
## $Seed
```

	diff	lwr	upr	p adj
## 327-329	0.2933333	-44.75780	45.34447	1
## 325-329	1.6100000	-43.44114	46.66114	1
## 307-329	1.0150000	-44.03614	46.06614	1
## 331-329	0.7066667	-44.34447	45.75780	1
## 311-329	1.3400000	-43.71114	46.39114	1
## 315-329	2.0466667	-43.00447	47.09780	1
## 321-329	0.8983333	-44.15280	45.94947	1
## 319-329	2.5683333	-42.48280	47.61947	1
## 301-329	2.9333333	-42.11780	47.98447	1
## 323-329	3.2400000	-41.81114	48.29114	1
## 309-329	3.4683333	-41.58280	48.51947	1
## 303-329	3.7933333	-41.25780	48.84447	1
## 305-329	4.8016667	-40.24947	49.85280	1
## 325-327	1.3166667	-43.73447	46.36780	1
## 307-327	0.7216667	-44.32947	45.77280	1
## 331-327	0.4133333	-44.63780	45.46447	1
## 311-327	1.0466667	-44.00447	46.09780	1
## 315-327	1.7533333	-43.29780	46.80447	1
## 321-327	0.6050000	-44.44614	45.65614	1
## 319-327	2.2750000	-42.77614	47.32614	1
## 301-327	2.6400000	-42.41114	47.69114	1
## 323-327	2.9466667	-42.10447	47.99780	1
## 309-327	3.1750000	-41.87614	48.22614	1
## 303-327	3.5000000	-41.55114	48.55114	1
## 305-327	4.5083333	-40.54280	49.55947	1
## 307-325	-0.5950000	-45.64614	44.45614	1
## 331-325	-0.9033333	-45.95447	44.14780	1
## 311-325	-0.2700000	-45.32114	44.78114	1
## 315-325	0.4366667	-44.61447	45.48780	1
## 321-325	-0.7116667	-45.76280	44.33947	1
## 319-325	0.9583333	-44.09280	46.00947	1
## 301-325	1.3233333	-43.72780	46.37447	1
## 323-325	1.6300000	-43.42114	46.68114	1
## 309-325	1.8583333	-43.19280	46.90947	1
## 303-325	2.1833333	-42.86780	47.23447	1
## 305-325	3.1916667	-41.85947	48.24280	1
## 331-307	-0.3083333	-45.35947	44.74280	1
## 311-307	0.3250000	-44.72614	45.37614	1
## 315-307	1.0316667	-44.01947	46.08280	1
## 321-307	-0.1166667	-45.16780	44.93447	1

## 319-307	1.5533333	-43.49780	46.60447	1
## 301-307	1.9183333	-43.13280	46.96947	1
## 323-307	2.2250000	-42.82614	47.27614	1
## 309-307	2.4533333	-42.59780	47.50447	1
## 303-307	2.7783333	-42.27280	47.82947	1
## 305-307	3.7866667	-41.26447	48.83780	1
## 311-331	0.6333333	-44.41780	45.68447	1
## 315-331	1.3400000	-43.71114	46.39114	1
## 321-331	0.1916667	-44.85947	45.24280	1
## 319-331	1.8616667	-43.18947	46.91280	1
## 301-331	2.2266667	-42.82447	47.27780	1
## 323-331	2.5333333	-42.51780	47.58447	1
## 309-331	2.7616667	-42.28947	47.81280	1
## 303-331	3.0866667	-41.96447	48.13780	1
## 305-331	4.0950000	-40.95614	49.14614	1
## 315-311	0.7066667	-44.34447	45.75780	1
## 321-311	-0.4416667	-45.49280	44.60947	1
## 319-311	1.2283333	-43.82280	46.27947	1
## 301-311	1.5933333	-43.45780	46.64447	1
## 323-311	1.9000000	-43.15114	46.95114	1
## 309-311	2.1283333	-42.92280	47.17947	1
## 303-311	2.4533333	-42.59780	47.50447	1
## 305-311	3.4616667	-41.58947	48.51280	1
## 321-315	-1.1483333	-46.19947	43.90280	1
## 319-315	0.5216667	-44.52947	45.57280	1
## 301-315	0.8866667	-44.16447	45.93780	1
## 323-315	1.1933333	-43.85780	46.24447	1
## 309-315	1.4216667	-43.62947	46.47280	1
## 303-315	1.7466667	-43.30447	46.79780	1
## 305-315	2.7550000	-42.29614	47.80614	1
## 319-321	1.6700000	-43.38114	46.72114	1
## 301-321	2.0350000	-43.01614	47.08614	1
## 323-321	2.3416667	-42.70947	47.39280	1
## 309-321	2.5700000	-42.48114	47.62114	1
## 303-321	2.8950000	-42.15614	47.94614	1
## 305-321	3.9033333	-41.14780	48.95447	1
## 301-319	0.3650000	-44.68614	45.41614	1
## 323-319	0.6716667	-44.37947	45.72280	1
## 309-319	0.9000000	-44.15114	45.95114	1
## 303-319	1.2250000	-43.82614	46.27614	1
## 305-319	2.2333333	-42.81780	47.28447	1
## 323-301	0.3066667	-44.74447	45.35780	1
## 309-301	0.5350000	-44.51614	45.58614	1
## 303-301	0.8600000	-44.19114	45.91114	1
## 305-301	1.8683333	-43.18280	46.91947	1
## 309-323	0.2283333	-44.82280	45.27947	1
## 303-323	0.5533333	-44.49780	45.60447	1
## 305-323	1.5616667	-43.48947	46.61280	1
## 303-309	0.3250000	-44.72614	45.37614	1

## 305-309	1.3333333	-43.71780	46.38447	1
## 305-303	1.0083333	-44.04280	46.05947	1

Type Your Answer Here:

```
(lmLob%>%aov())%>%TukeyHSD())
```

Fit: aov(formula = .)

\$Seed

	diff	lwr	upr	p adj	
327-329	0.2933333	-44.75780	45.34447	1	
325-329	1.6100000	-43.44114	46.66114	1	
307-329	1.0150000	-44.03614	46.06614	1	
331-329	0.7066667	-44.34447	45.75780	1	
311-329	1.3400000	-43.71114	46.39114	1	
315-329	2.0466667	-43.00447	47.09780	1	
321-329	0.8983333	-44.15280	45.94947	1	
319-329	2.5683333	-42.48280	47.61947	1	
301-329	2.9333333	-42.11780	47.98447	1	
323-329	3.2400000	-41.81114	48.29114	1	
309-329	3.4683333	-41.58280	48.51947	1	
303-329	3.7933333	-41.25780	48.84447	1	
305-329	4.8016667	-40.24947	49.85280	1	
325-327	1.3166667	-43.73447	46.36780	1	
307-327	0.7216667	-44.32947	45.77280	1	
331-327	0.4133333	-44.63780	45.46447	1	
311-327	1.0466667	-44.00447	46.09780	1	
315-327	1.7533333	-43.29780	46.80447	1	
321-327	0.6050000	-44.44614	45.65614	1	
319-327	2.2750000	-42.77614	47.32614	1	
301-327	2.6400000	-42.41114	47.69114	1	
323-327	2.9466667	-42.10447	47.99780	1	
309-327	3.1750000	-41.87614	48.22614	1	
303-327	3.5000000	-41.55114	48.55114	1	
305-327	4.5083333	-40.54280	49.55947	1	
307-325	-0.5950000	-45.64614	44.45614	1	
331-325	-0.9033333	-45.95447	44.14780	1	
311-325	-0.2700000	-45.32114	44.78114	1	
315-325	0.4366667	-44.61447	45.48780	1	
321-325	-0.7116667	-45.76280	44.33947	1	
319-325	0.9583333	-44.09280	46.00947	1	
301-325	1.3233333	-43.72780	46.37447	1	
323-325	1.6300000	-43.42114	46.68114	1	
309-325	1.8583333	-43.19280	46.90947	1	
303-325	2.1833333	-42.86780	47.23447	1	

305-325	3.1916667	-41.85947	48.24280	1
331-307	-0.3083333	-45.35947	44.74280	1
311-307	0.3250000	-44.72614	45.37614	1
315-307	1.0316667	-44.01947	46.08280	1
321-307	-0.1166667	-45.16780	44.93447	1
319-307	1.5533333	-43.49780	46.60447	1
301-307	1.9183333	-43.13280	46.96947	1
323-307	2.2250000	-42.82614	47.27614	1
309-307	2.4533333	-42.59780	47.50447	1
303-307	2.7783333	-42.27280	47.82947	1
305-307	3.7866667	-41.26447	48.83780	1
311-331	0.6333333	-44.41780	45.68447	1
315-331	1.3400000	-43.71114	46.39114	1
321-331	0.1916667	-44.85947	45.24280	1
319-331	1.8616667	-43.18947	46.91280	1
301-331	2.2266667	-42.82447	47.27780	1
323-331	2.5333333	-42.51780	47.58447	1
309-331	2.7616667	-42.28947	47.81280	1
303-331	3.0866667	-41.96447	48.13780	1
305-331	4.0950000	-40.95614	49.14614	1
315-311	0.7066667	-44.34447	45.75780	1
321-311	-0.4416667	-45.49280	44.60947	1
319-311	1.2283333	-43.82280	46.27947	1
301-311	1.5933333	-43.45780	46.64447	1
323-311	1.9000000	-43.15114	46.95114	1
309-311	2.1283333	-42.92280	47.17947	1
303-311	2.4533333	-42.59780	47.50447	1
305-311	3.4616667	-41.58947	48.51280	1
321-315	-1.1483333	-46.19947	43.90280	1
319-315	0.5216667	-44.52947	45.57280	1
301-315	0.8866667	-44.16447	45.93780	1
323-315	1.1933333	-43.85780	46.24447	1
309-315	1.4216667	-43.62947	46.47280	1
303-315	1.7466667	-43.30447	46.79780	1
305-315	2.7550000	-42.29614	47.80614	1
319-321	1.6700000	-43.38114	46.72114	1
301-321	2.0350000	-43.01614	47.08614	1
323-321	2.3416667	-42.70947	47.39280	1
309-321	2.5700000	-42.48114	47.62114	1
303-321	2.8950000	-42.15614	47.94614	1
305-321	3.9033333	-41.14780	48.95447	1
301-319	0.3650000	-44.68614	45.41614	1
323-319	0.6716667	-44.37947	45.72280	1
309-319	0.9000000	-44.15114	45.95114	1
303-319	1.2250000	-43.82614	46.27614	1
305-319	2.2333333	-42.81780	47.28447	1

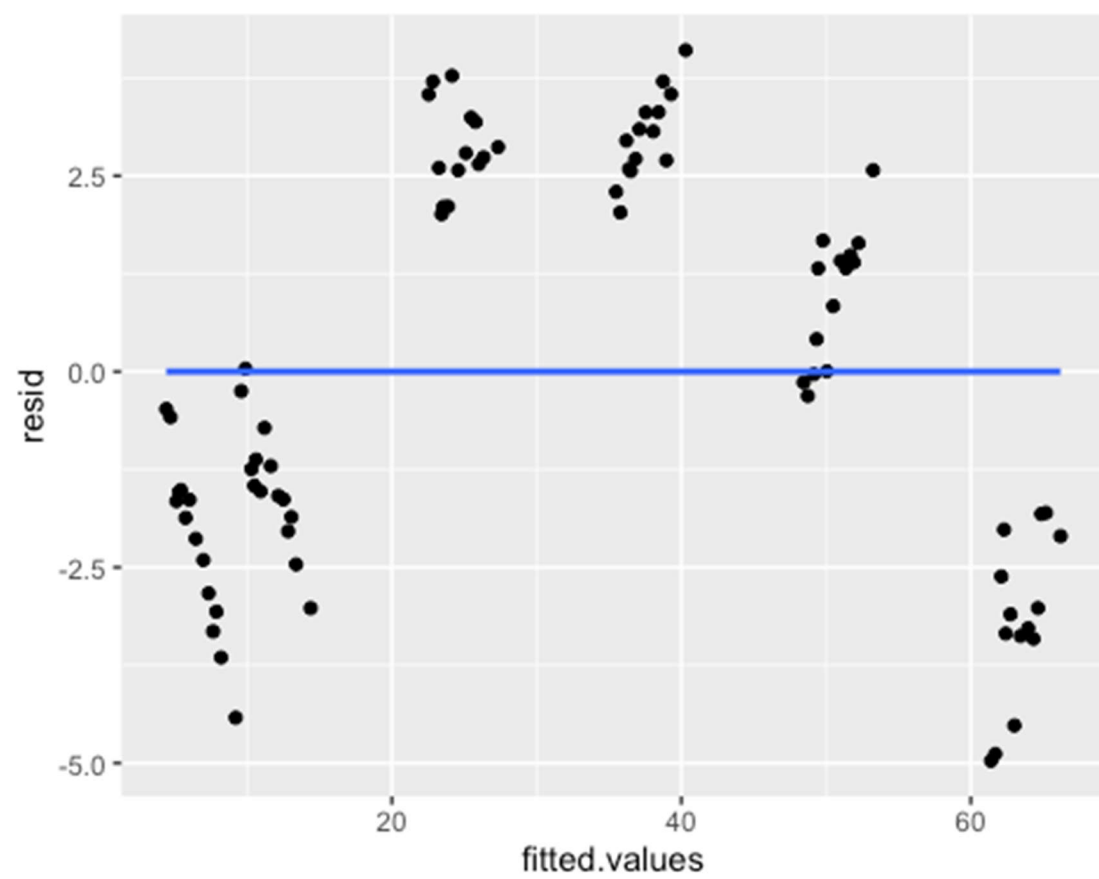
323-301	0.3066667	-44.74447	45.35780	1
309-301	0.5350000	-44.51614	45.58614	1
303-301	0.8600000	-44.19114	45.91114	1
305-301	1.8683333	-43.18280	46.91947	1
309-323	0.2283333	-44.82280	45.27947	1
303-323	0.5533333	-44.49780	45.60447	1
305-323	1.5616667	-43.48947	46.61280	1
303-309	0.3250000	-44.72614	45.37614	1
305-309	1.3333333	-43.71780	46.38447	1
305-303	1.0083333	-44.04280	46.05947	1

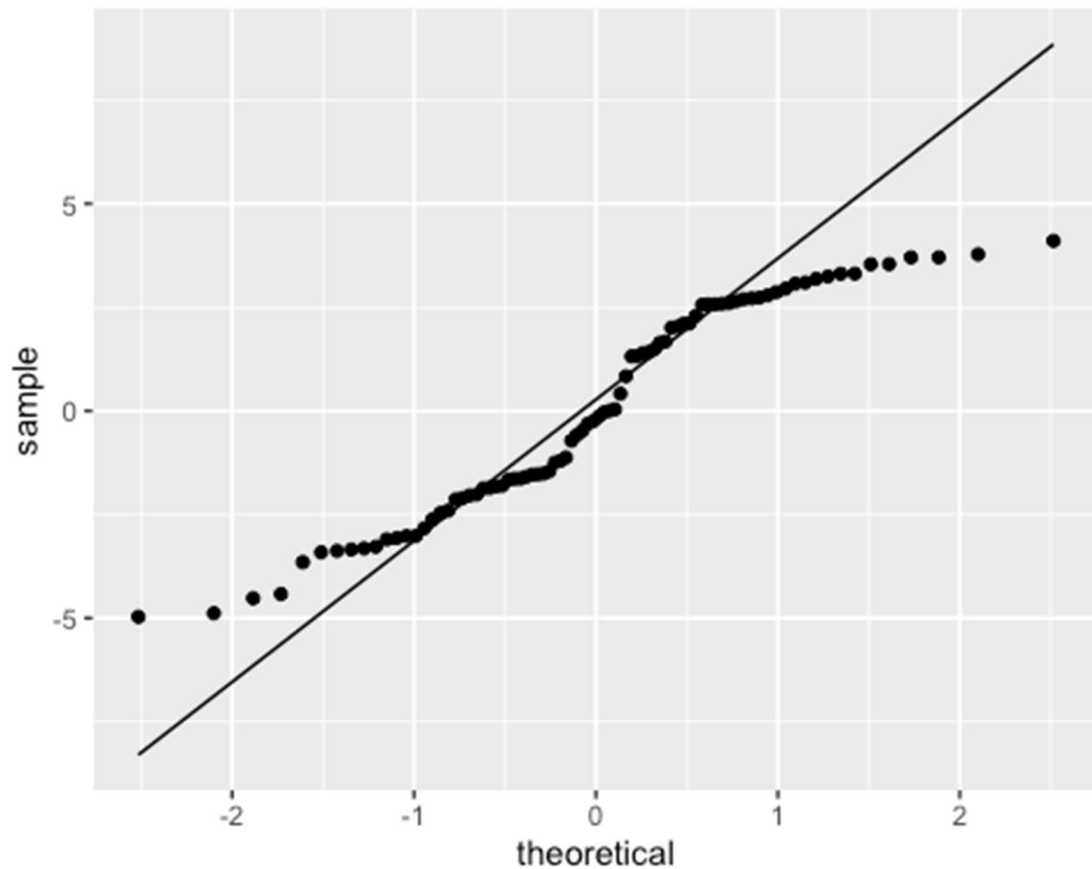
Considering a Type I error of $\alpha = 0.05$, since no pair has a p-value of less than 0.05, it means that no pairs differ significantly. This can also be confirmed given the fact all p-values in each pair are 1. Again it makes sense given that these seeds are all from the same breed of tree.

- c. Now perform an ANCOVA for height as a function of Seed, but control for the trees age. Does this change the results from the one-way ANOVA? If so, how? Based on the residuals, is it reasonable to accept that the assumptions for linear regression are met? (5 points)

Analysis of Variance Model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
age	1	34762	34762	4374	3.8e-64
Seed	13	163.7	12.59	1.584	0.1112
Residuals	69	548.4	7.947	NA	NA





Type Your Answer Here:

```
lmLob2 <- lm(height~age+Seed,Loblolly)
anovaLob2 <- lmLob2>%anova()
anovaLob2 #Print for Analysis of Variance Model

ggplot(lmLob2, aes(x=fitted(lmLob),y=resid(lmLob)))+ #Plot lmLob for
fitted and resid
  geom_point()+ #Make data appear as dots on plot
  geom_hline(yintercept=0,color="blue")+ #add horizontal blue line
across the middle
  xlab("fitted.values")+ #Make the x axis label the same as one in
word document
  ylab("resid") #Make the y axis label the same as one in word
document

qqplot.data <- function(vec) #argument: vector of numbers
{
  y <- quantile(x=vec[!is.na(vec)],probs=c(0.25,0.75)) #Get the
quantiles of the vector AS LONG AS THEY ARE NOT NA. Using this with
```

```

probs=c(0.25,0.75) allows y to return a numeric vector of Q1 and Q3,
these being the quantiles of the data. This is for the slope.
  x <- qnorm(c(0.25,0.75)) #Return 0.25 and 0.75 as normal quantiles,
which are the quantiles we want the slope to be. This is for the
slope.
  slope <- diff(y)/diff(x) #The difference of y over x, also known as
the quantiles of the data vs the quantiles we desire.
  int <- y[1L] - slope * x[1L] #The result of our datasets' lower
quantils vs the slope mutliplied by the desired lower quantile.
  d <- data.frame(resids = vec) #Create a local dataframe from the
input vector (vec) of numbers and name it resids
  ggplot(d, aes(sample = resids))+ #Generate the ggplot, using resids
as the sample.
  stat_qq()+ #Make the ggplot a QQplot
  geom_abline(slope=slope, intercept=int)+ #Generate the line to run
labs( #Give the plot approrpiate labels.
  y="sample",
  x="theoretical")+
  coord_cartesian(xlim=c(-2.5,2.5),ylim=c(-9,9)) #Just fix the x and
y limits to make it look like graph in word document
}
qqplot.data(lmLob2$residuals)

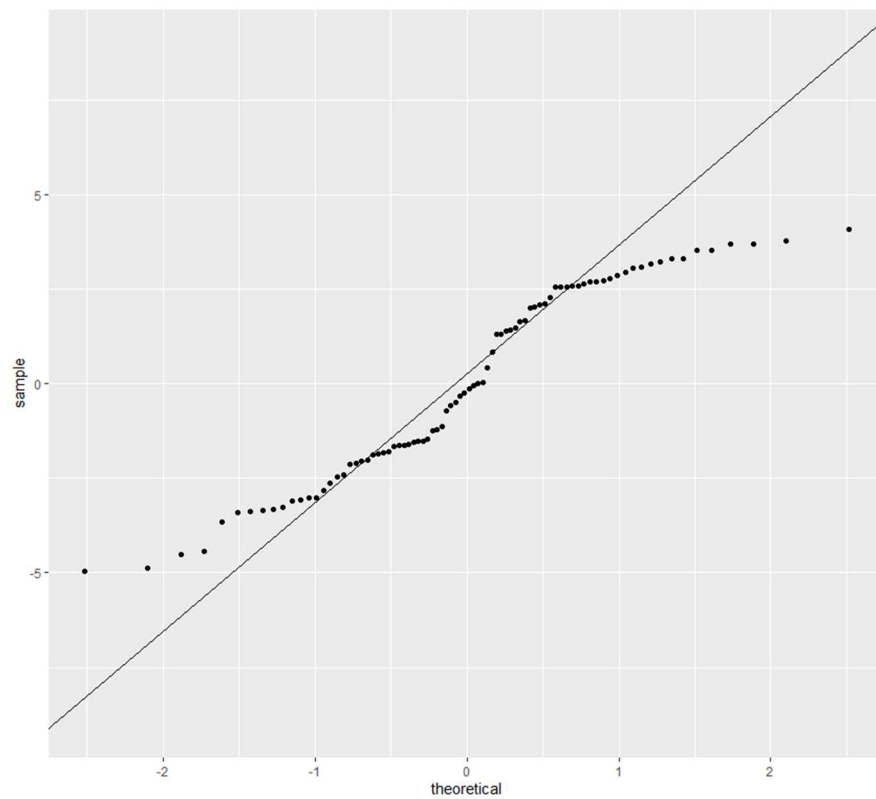
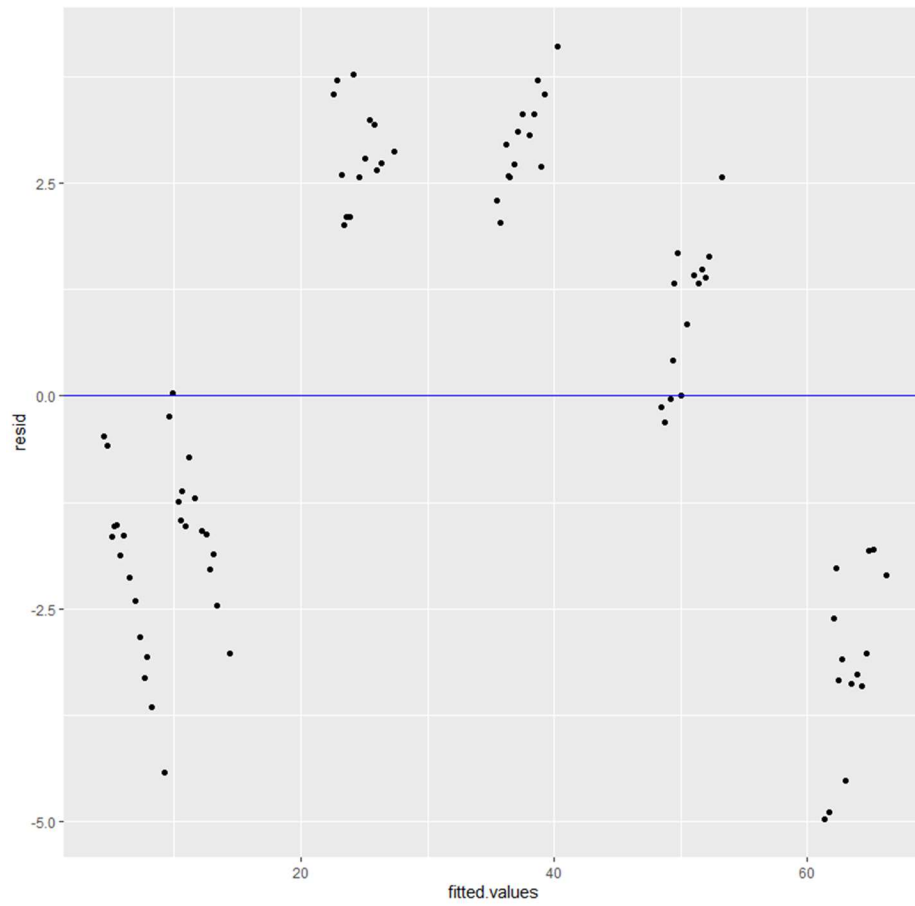
```

```
> anovaLob2
```

```
Analysis of Variance Table
```

```
Response: height
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
age	1	34762	34762	4373.961	<2e-16	***
Seed	13	164	13	1.584	0.1112	
Residuals	69	548	8			



There is a change in the results between the one-way ANOVA and the new model. The new model's F-value and p-value have changed significantly, and it shows that age is statistically significant to the model. Seed is still not statistically significant to the model however given a p-value of 0.1112 is greater than the typical Type I error rate of $\alpha = 0.05$.

As for whether the assumptions for linear regression are met, first look at the plot of the fitted values vs residuals. As the plot is useful for checking that the variance is not a function, the variance should fall either side of the plot with no evidence of a trend or increase of magnitude as \hat{y} increases. The fitted vs residual values plot does not confirm this theory and shows that the resid and fitted values group like a parabola. See how they group up like a parabolic trend which means there is evidence of Heteroskedasticity.

The second plot, which is a plot of the quartiles against each other (theoretical vs calculated) indicates if a model is Gaussian in nature, which can be confirmed on a plot if the values align at a ~ 45 degree angle. The values in the qq-plot above (sample vs theoretical) do not align with the line closely and diverges on the upper and lower regions, which means it forms a uniform pattern.

With this in mind, it is we must deny that the assumptions for a linear regression are met.

d. Repeat the analysis using Tukey's HSD in part b.). What are the results? (2 points)

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = .)
##
## $Seed
##
```

	diff	lwr	upr	p adj
## 327-329	0.2933333	-5.3643615	5.951028	1.0000000
## 325-329	1.6100000	-4.0476948	7.267695	0.9992048
## 307-329	1.0150000	-4.6426948	6.672695	0.9999956
## 331-329	0.7066667	-4.9510281	6.364361	0.9999999
## 311-329	1.3400000	-4.3176948	6.997695	0.9998905
## 315-329	2.0466667	-3.6110281	7.704361	0.9916855
## 321-329	0.8983333	-4.7593615	6.556028	0.9999990
## 319-329	2.5683333	-3.0893615	8.226028	0.9459546
## 301-329	2.9333333	-2.7243615	8.591028	0.8675251
## 323-329	3.2400000	-2.4176948	8.897695	0.7678031
## 309-329	3.4683333	-2.1893615	9.126028	0.6779336
## 303-329	3.7933333	-1.8643615	9.451028	0.5391471
## 305-329	4.8016667	-0.8560281	10.459361	0.1824627
## 325-327	1.3166667	-4.3410281	6.974361	0.9999100
## 307-327	0.7216667	-4.9360281	6.379361	0.9999999
## 331-327	0.4133333	-5.2443615	6.071028	1.0000000
## 311-327	1.0466667	-4.6110281	6.704361	0.9999936
## 315-327	1.7533333	-3.9043615	7.411028	0.9981012
## 321-327	0.6050000	-5.0526948	6.262695	1.0000000
## 319-327	2.2750000	-3.3826948	7.932695	0.9791227
## 301-327	2.6400000	-3.0176948	8.297695	0.9340692
## 323-327	2.9466667	-2.7110281	8.604361	0.8638039
## 309-327	3.1750000	-2.4826948	8.832695	0.7912451
## 303-327	3.5000000	-2.1576948	9.157695	0.6647598
## 305-327	4.5083333	-1.1493615	10.166028	0.2640820
## 307-325	-0.5950000	-6.2526948	5.062695	1.0000000
## 331-325	-0.9033333	-6.5610281	4.754361	0.9999989
## 311-325	-0.2700000	-5.9276948	5.387695	1.0000000
## 315-325	0.4366667	-5.2210281	6.094361	1.0000000
## 321-325	-0.7116667	-6.3693615	4.946028	0.9999999
## 319-325	0.9583333	-4.6993615	6.616028	0.9999978
## 301-325	1.3233333	-4.3343615	6.981028	0.9999048
## 323-325	1.6300000	-4.0276948	7.287695	0.9990960
## 309-325	1.8583333	-3.7993615	7.516028	0.9966365
## 303-325	2.1833333	-3.4743615	7.841028	0.9852667
## 305-325	3.1916667	-2.4660281	8.849361	0.7853410
## 331-307	-0.3083333	-5.9660281	5.349361	1.0000000
## 311-307	0.3250000	-5.3326948	5.982695	1.0000000
## 315-307	1.0316667	-4.6260281	6.689361	0.9999946
## 321-307	-0.1166667	-5.7743615	5.541028	1.0000000
## 319-307	1.5533333	-4.1043615	7.211028	0.9994540
## 301-307	1.9183333	-3.7393615	7.576028	0.9954406

## 323-307	2.2250000	-3.4326948	7.882695	0.9826799
## 309-307	2.4533333	-3.2043615	8.111028	0.9617329
## 303-307	2.7783333	-2.8793615	8.436028	0.9063427
## 305-307	3.7866667	-1.8710281	9.444361	0.5420196
## 311-331	0.6333333	-5.0243615	6.291028	1.0000000
## 315-331	1.3400000	-4.3176948	6.997695	0.9998905
## 321-331	0.1916667	-5.4660281	5.849361	1.0000000
## 319-331	1.8616667	-3.7960281	7.519361	0.9965778
## 301-331	2.2266667	-3.4310281	7.884361	0.9825695
## 323-331	2.5333333	-3.1243615	8.191028	0.9511734
## 309-331	2.7616667	-2.8960281	8.419361	0.9100263
## 303-331	3.0866667	-2.5710281	8.744361	0.8212209
## 305-331	4.0950000	-1.5626948	9.752695	0.4128234
## 315-311	0.7066667	-4.9510281	6.364361	0.9999999
## 321-311	-0.4416667	-6.0993615	5.216028	1.0000000
## 319-311	1.2283333	-4.4293615	6.886028	0.9999590
## 301-311	1.5933333	-4.0643615	7.251028	0.9992866
## 323-311	1.9000000	-3.7576948	7.557695	0.9958384
## 309-311	2.1283333	-3.5293615	7.786028	0.9882042
## 303-311	2.4533333	-3.2043615	8.111028	0.9617329
## 305-311	3.4616667	-2.1960281	9.119361	0.6806898
## 321-315	-1.1483333	-6.8060281	4.509361	0.9999811
## 319-315	0.5216667	-5.1360281	6.179361	1.0000000
## 301-315	0.8866667	-4.7710281	6.544361	0.9999991
## 323-315	1.1933333	-4.4643615	6.851028	0.9999706
## 309-315	1.4216667	-4.2360281	7.079361	0.9997895
## 303-315	1.7466667	-3.9110281	7.404361	0.9981722
## 305-315	2.7550000	-2.9026948	8.412695	0.9114732
## 319-321	1.6700000	-3.9876948	7.327695	0.9988397
## 301-321	2.0350000	-3.6226948	7.692695	0.9921065
## 323-321	2.3416667	-3.3160281	7.999361	0.9735344
## 309-321	2.5700000	-3.0876948	8.227695	0.9456967
## 303-321	2.8950000	-2.7626948	8.552695	0.8778892
## 305-321	3.9033333	-1.7543615	9.561028	0.4920713
## 301-319	0.3650000	-5.2926948	6.022695	1.0000000
## 323-319	0.6716667	-4.9860281	6.329361	1.0000000
## 309-319	0.9000000	-4.7576948	6.557695	0.9999990
## 303-319	1.2250000	-4.4326948	6.882695	0.9999602
## 305-319	2.2333333	-3.4243615	7.891028	0.9821224
## 323-301	0.3066667	-5.3510281	5.964361	1.0000000
## 309-301	0.5350000	-5.1226948	6.192695	1.0000000
## 303-301	0.8600000	-4.7976948	6.517695	0.9999994
## 305-301	1.8683333	-3.7893615	7.526028	0.9964577
## 309-323	0.2283333	-5.4293615	5.886028	1.0000000
## 303-323	0.5533333	-5.1043615	6.211028	1.0000000
## 305-323	1.5616667	-4.0960281	7.219361	0.9994223
## 303-309	0.3250000	-5.3326948	5.982695	1.0000000
## 305-309	1.3333333	-4.3243615	6.991028	0.9998964
## 305-303	1.0083333	-4.6493615	6.666028	0.9999959

Type Your Answer Here:

```
(lmLob2>%aov( )>%TukeyHSD( ))
```

Fit: aov(formula = .)

\$Seed

	diff	lwr	upr	p adj
327-329	0.2933333	-5.3643615	5.951028	1.0000000
325-329	1.6100000	-4.0476948	7.267695	0.9992048
307-329	1.0150000	-4.6426948	6.672695	0.9999956
331-329	0.7066667	-4.9510281	6.364361	0.9999999
311-329	1.3400000	-4.3176948	6.997695	0.9998905
315-329	2.0466667	-3.6110281	7.704361	0.9916855
321-329	0.8983333	-4.7593615	6.556028	0.9999990
319-329	2.5683333	-3.0893615	8.226028	0.9459546
301-329	2.9333333	-2.7243615	8.591028	0.8675251
323-329	3.2400000	-2.4176948	8.897695	0.7678031
309-329	3.4683333	-2.1893615	9.126028	0.6779336
303-329	3.7933333	-1.8643615	9.451028	0.5391471
305-329	4.8016667	-0.8560281	10.459361	0.1824627
325-327	1.3166667	-4.3410281	6.974361	0.9999100
307-327	0.7216667	-4.9360281	6.379361	0.9999999
331-327	0.4133333	-5.2443615	6.071028	1.0000000
311-327	1.0466667	-4.6110281	6.704361	0.9999936
315-327	1.7533333	-3.9043615	7.411028	0.9981012
321-327	0.6050000	-5.0526948	6.262695	1.0000000
319-327	2.2750000	-3.3826948	7.932695	0.9791227
301-327	2.6400000	-3.0176948	8.297695	0.9340692
323-327	2.9466667	-2.7110281	8.604361	0.8638039
309-327	3.1750000	-2.4826948	8.832695	0.7912451
303-327	3.5000000	-2.1576948	9.157695	0.6647598
305-327	4.5083333	-1.1493615	10.166028	0.2640820
307-325	-0.5950000	-6.2526948	5.062695	1.0000000
331-325	-0.9033333	-6.5610281	4.754361	0.9999989
311-325	-0.2700000	-5.9276948	5.387695	1.0000000
315-325	0.4366667	-5.2210281	6.094361	1.0000000
321-325	-0.7116667	-6.3693615	4.946028	0.9999999
319-325	0.9583333	-4.6993615	6.616028	0.9999978
301-325	1.3233333	-4.3343615	6.981028	0.9999048
323-325	1.6300000	-4.0276948	7.287695	0.9990960
309-325	1.8583333	-3.7993615	7.516028	0.9966365
303-325	2.1833333	-3.4743615	7.841028	0.9852667
305-325	3.1916667	-2.4660281	8.849361	0.7853410
331-307	-0.3083333	-5.9660281	5.349361	1.0000000

311-307 0.3250000 -5.3326948 5.982695 1.0000000
315-307 1.0316667 -4.6260281 6.689361 0.9999946
321-307 -0.1166667 -5.7743615 5.541028 1.0000000
319-307 1.5533333 -4.1043615 7.211028 0.9994540
301-307 1.9183333 -3.7393615 7.576028 0.9954406
323-307 2.2250000 -3.4326948 7.882695 0.9826799
309-307 2.4533333 -3.2043615 8.111028 0.9617329
303-307 2.7783333 -2.8793615 8.436028 0.9063427
305-307 3.7866667 -1.8710281 9.444361 0.5420196
311-331 0.6333333 -5.0243615 6.291028 1.0000000
315-331 1.3400000 -4.3176948 6.997695 0.9998905
321-331 0.1916667 -5.4660281 5.849361 1.0000000
319-331 1.8616667 -3.7960281 7.519361 0.9965778
301-331 2.2266667 -3.4310281 7.884361 0.9825695
323-331 2.5333333 -3.1243615 8.191028 0.9511734
309-331 2.7616667 -2.8960281 8.419361 0.9100263
303-331 3.0866667 -2.5710281 8.744361 0.8212209
305-331 4.0950000 -1.5626948 9.752695 0.4128234
315-311 0.7066667 -4.9510281 6.364361 0.9999999
321-311 -0.4416667 -6.0993615 5.216028 1.0000000
319-311 1.2283333 -4.4293615 6.886028 0.9999590
301-311 1.5933333 -4.0643615 7.251028 0.9992866
323-311 1.9000000 -3.7576948 7.557695 0.9958384
309-311 2.1283333 -3.5293615 7.786028 0.9882042
303-311 2.4533333 -3.2043615 8.111028 0.9617329
305-311 3.4616667 -2.1960281 9.119361 0.6806898
321-315 -1.1483333 -6.8060281 4.509361 0.9999811
319-315 0.5216667 -5.1360281 6.179361 1.0000000
301-315 0.8866667 -4.7710281 6.544361 0.9999991
323-315 1.1933333 -4.4643615 6.851028 0.9999706
309-315 1.4216667 -4.2360281 7.079361 0.9997895
303-315 1.7466667 -3.9110281 7.404361 0.9981722
305-315 2.7550000 -2.9026948 8.412695 0.9114732
319-321 1.6700000 -3.9876948 7.327695 0.9988397
301-321 2.0350000 -3.6226948 7.692695 0.9921065
323-321 2.3416667 -3.3160281 7.999361 0.9735344
309-321 2.5700000 -3.0876948 8.227695 0.9456967
303-321 2.8950000 -2.7626948 8.552695 0.8778892
305-321 3.9033333 -1.7543615 9.561028 0.4920713
301-319 0.3650000 -5.2926948 6.022695 1.0000000
323-319 0.6716667 -4.9860281 6.329361 1.0000000
309-319 0.9000000 -4.7576948 6.557695 0.9999990
303-319 1.2250000 -4.4326948 6.882695 0.9999602
305-319 2.2333333 -3.4243615 7.891028 0.9821224
323-301 0.3066667 -5.3510281 5.964361 1.0000000
309-301 0.5350000 -5.1226948 6.192695 1.0000000

303-301	0.8600000	-4.7976948	6.517695	0.9999994
305-301	1.8683333	-3.7893615	7.526028	0.9964577
309-323	0.2283333	-5.4293615	5.886028	1.0000000
303-323	0.5533333	-5.1043615	6.211028	1.0000000
305-323	1.5616667	-4.0960281	7.219361	0.9994223
303-309	0.3250000	-5.3326948	5.982695	1.0000000
305-309	1.3333333	-4.3243615	6.991028	0.9998964
305-303	1.0083333	-4.6493615	6.666028	0.9999959

With the newer model, results indicate a significant difference with the significance of the model. The p-adj values have changed for most of the pairs in the Tukey's HSD, which indicates an increase of statistical significance, though none go below the alpha level of 0.05 which means ALL are still statistically insignificant.

Question 9

Grades in an elementary statistics class were classified by the students' majors. Is there any relationship between grade and major? (4 points)

	Biology	Other	Psychology
A	15	13	8
B	19	15	14
C	4	7	15
D-F	1	4	3

Type Your Answer Here:

9. first, compute tabular totals.

	Bio	Other	Psychology	Total
A	15	13	8	36
B	19	15	14	48
C	4	7	15	26
D-F	1	4	3	8
Total	39	39	40	118

With this, we take the null hypothesis as:

$$H_0: \pi_{i1} = \pi_{i2} = \dots = \pi_{ij}, \forall i \quad H_A: \pi_{i1} \neq \pi_{i2} \neq \dots \neq \pi_{ij}, \forall i$$

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

$$E_{ij} = \frac{n_i n_j}{N}$$

$$\begin{aligned} \chi^2 = & \frac{(n_{11} - E_{11})^2}{E_{11}} + \frac{(n_{12} - E_{12})^2}{E_{12}} + \frac{(n_{13} - E_{13})^2}{E_{13}} + \frac{(n_{14} - E_{14})^2}{E_{14}} + \frac{(n_{15} - E_{15})^2}{E_{15}} \\ & + \frac{(n_{21} - E_{21})^2}{E_{21}} + \frac{(n_{22} - E_{22})^2}{E_{22}} + \frac{(n_{23} - E_{23})^2}{E_{23}} + \frac{(n_{24} - E_{24})^2}{E_{24}} + \frac{(n_{25} - E_{25})^2}{E_{25}} \\ & + \frac{(n_{31} - E_{31})^2}{E_{31}} + \frac{(n_{32} - E_{32})^2}{E_{32}} + \frac{(n_{33} - E_{33})^2}{E_{33}} + \frac{(n_{34} - E_{34})^2}{E_{34}} + \frac{(n_{35} - E_{35})^2}{E_{35}} \\ & + \frac{(n_{41} - E_{41})^2}{E_{41}} + \frac{(n_{42} - E_{42})^2}{E_{42}} + \frac{(n_{43} - E_{43})^2}{E_{43}} \end{aligned}$$

$$\begin{aligned} = & 0.80856 + 0.102 + 1.44783 + 0.61974 + 0.04709 + 0.31701 + \\ & 0.45515 + 0.29838 + 4.3424 + 1.02227 + 0.69534 \\ & + 0.03061 = 12.1829 \end{aligned}$$

Determine $\chi_{\alpha/2, u}^2$, $u = (i-1)(j-1) = 6$, given type I error of $\alpha = 0.05$,
 $\chi_{0.05, 6}^2 = 12.592$

Given that $\chi^2 = 12.1829$ is less than $\chi_{0.05, 6}^2 = 12.592$ we fail to reject the null hypothesis

Appendix:

Question 1 A:

$$p(x) = p(1 - p)^{x-1}$$

$$E(X) = \frac{1}{p} \rightarrow \frac{1}{E(X)} = p$$

$$m_1 = \frac{1}{n} \sum_i x_i$$

$$\hat{p} = \frac{1}{m_1} = \frac{n}{\sum_i x_i}$$

Question 1 B:

$$f(x) = (1 - p)^{x_i - 1} p$$

We define likelihood function as:

$$L(p) = \prod_{i=1}^n (1 - p)^{x_i - 1} p$$

Then define log-likelihood function as:

$$\begin{aligned} \log L(p) &= \log \left(\prod_{i=1}^n (1 - p)^{x_i - 1} p \right) \\ &= \sum_{i=1}^n \log[(1 - p)^{x_i - 1} p] \\ &= \sum_{i=1}^n \log((1 - p)^{x_i - 1}) + \sum_{i=1}^n \log p \\ &= (x_i - 1) \sum_{i=1}^n \log(1 - p) + \sum_{i=1}^n \log p \end{aligned}$$

We take $\frac{d \log L(p)}{dp} = 0$

Which is $\frac{d}{dp} \log L(p) = \frac{\sum_{i=1}^n x_i - n}{1 - p} + \frac{n}{p} = 0$

$\hat{p} \rightarrow \text{Solve(?)}$

$$\begin{aligned} \sum_{i=1}^n \frac{x_i - \mu}{1 - p} &= \frac{n}{p} \\ \sum x_i - n &= \frac{n - p}{p} \\ \sum x_i - n &= \frac{\mu}{p} - n \\ \hat{p} &= \frac{n}{\sum x_i} \end{aligned}$$

Question 2:

Since the sample size is 30, we have a passable large sample. Since this is a 90% confidence interval and we have two-sided, $Z_{\alpha/2} = 1.645$.

$$\begin{aligned} & \left(\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \\ &= \left(0.145 - 1.645 \frac{0.0051}{\sqrt{30}}, 0.145 + 1.645 \frac{0.0051}{\sqrt{30}} \right) \\ &= (0.1435, 0.1465) \end{aligned}$$

Question 3 A:

Since the sample size is 130, we have a large sample. This means we can make the null hypothesis as:

$$H_0: \mu = 37 \quad H_A: \mu \neq 37$$

Where μ is the average of human body temp, given a Type I error of $\alpha = 0.01$ and we have two-sided, $Z_{\alpha/2} = 2.58$.

$$\begin{aligned} Z &= \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \\ Z &= \frac{36.8 - 37}{0.41/\sqrt{130}} = -5.56 \end{aligned}$$

(Insert two-sided reject plot)

Since $Z = -5.56$ is within the rejection region, from the plot above we reject H_0

Question 3 B:

This 'experimental data' couldn't have been conducted in experimental conditions given such a huge sample size of 1 million tests. This can also be backed up given the rejected H_0 .

Question 4:

Since the sample size is 600, we have a large sample. This means we can make the null hypothesis as:

$$H_0: p_0 \leq 0.5 \quad H_A: p_0 > 0.5$$

Where p_0 is the proportion of voters who prefer Labor over LNP and \hat{p} is the sample statistic.

Given a Type I error of $\alpha = 0.05$ and we have one-sided, $Z_\alpha = 1.645$.

$$\begin{aligned} Z &= \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \\ &= \frac{0.52 - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{600}}} \\ Z &= 0.9798 \end{aligned}$$

(Insert right-handed reject plot)

Since $Z = 0.9798$ is not within the reject region, we fail to reject the null hypothesis.

Question 5 A:

There might be a case of unequal variance, so test if $\frac{Large\ s^2}{Small\ s^2} > 3 \rightarrow \frac{0.03533}{0.02716} = 1.3 < 3$.

No case of unequal variance, use normal T distribution given a small sample size of $10 < 30$ to calculate the 95% confidence interval for swimmer 1 and 2.

Where \bar{x}_1 = swimmer 1 and \bar{x}_2 = swimmer 2

$$\begin{aligned} CI &= (\bar{x}_1 - \bar{x}_2) \pm \sqrt{s_p^2 \left(\frac{1}{n_{x_1}} + \frac{1}{n_{x_2}} \right)} * t_{v, \alpha/2} \\ &= (59.646 - 59.627) \pm \sqrt{s_p^2 \left(\frac{1}{10} + \frac{1}{10} \right)} * t_{v, \alpha/2} \\ s_p^2 &= \frac{(10 - 1)s_1^2 + (10 - 1)s_2^2}{10 + 10 - 2} \\ &= \frac{9(0.1648)^2 + 9(0.1880)^2}{18} \\ &\approx 0.03124 \end{aligned}$$

We can find the degrees of freedom: $v = n_{x_1} - n_{x_2} - 2 = 18$, hence $t_{v, 1-\alpha} = 1.734$

\therefore The 95% confidence interval for the difference of average times between swimmer 1 and swimmer 2 is:

$$(-0.1181, 0.1561)$$

Question 5 B:

Since the sample size is 10, we have a small sample which means we must use T-distribution.

We define our hypothesis as:

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_A: \mu_1 - \mu_2 > 0$$

μ_1 is average of swimmer 1

μ_2 is average of swimmer 2

$$\begin{aligned} t &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \\ &= \frac{59.646 - 59.627}{\sqrt{0.03124 \left(\frac{1}{10} + \frac{1}{10} \right)}} \\ t &\approx 0.2403 \end{aligned}$$

Using the same $t_{v,1-\alpha} = t_{18,0.95} = 1.734$

(Insert right-handed reject plot)

Since $t \approx 0.2403$ is not within the reject region, we fail to reject the null hypothesis that Swimmer 2 is on average faster than Swimmer 1.

Question 6:

Since the sample size is 7, we have a small sample. This means we can make the null hypothesis as:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

μ_1 is the mean reaction time before consuming alcohol

μ_2 is the mean reaction time after consuming alcohol

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{4.142857 - 5.285714}{\sqrt{s_p^2 \left(\frac{1}{7} + \frac{1}{7} \right)}}$$

$$s_p^2 = \frac{(7-1)s_1^2 + (7-1)s_2^2}{7+7-2} = \frac{6(1.345185)^2 + 6(1.704336)^2}{12} = 2.35714$$

$$t = \frac{4.142857 - 5.285714}{\sqrt{2.35714 \left(\frac{1}{7} + \frac{1}{7} \right)}}$$

$$t \approx -1.393$$

Given a Type I error of $\alpha = 0.05$, $\frac{\alpha}{2} = 0.025$. Determine $t_{v, 1-\alpha/2} = qt(0.975, n_1 + n_2 - 2) = qt(0.975, 12) = 2.18$

(Insert left-handed reject plot)

Since $t \approx -1.393$ is not within the reject region, we fail to reject the null hypothesis that reaction time increases after consuming alcohol.

Question 9:

First, compute the tabular totals.

	Biology	Other	Psychology	Total
A	15	13	8	36
B	19	15	14	48
C	4	7	15	26
D-F	1	4	3	8
Total	39	39	40	118

With this, we take the null hypothesis as:

$$H_0: \pi_{i1} = \pi_{i2} = \cdots \pi_{ij}, \forall_i$$

$$H_A: \pi_{i1} \neq \pi_{i2} \neq \cdots \pi_{ij}, \forall_i$$

$$X^2 = \sum_i \sum_j \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

$$E_{ij} = \frac{n_i n_j}{N}$$

$$\begin{aligned} X^2 = & \frac{(n_{11} - E_{11})^2}{E_{11}} + \frac{(n_{12} - E_{12})^2}{E_{12}} + \frac{(n_{13} - E_{13})^2}{E_{13}} + \\ & + \frac{(n_{21} - E_{21})^2}{E_{21}} + \frac{(n_{22} - E_{22})^2}{E_{22}} + \frac{(n_{23} - E_{23})^2}{E_{23}} + \\ & + \frac{(n_{31} - E_{31})^2}{E_{31}} + \frac{(n_{32} - E_{32})^2}{E_{32}} + \frac{(n_{33} - E_{33})^2}{E_{33}} + \\ & + \frac{(n_{41} - E_{41})^2}{E_{41}} + \frac{(n_{42} - E_{42})^2}{E_{42}} + \frac{(n_{43} - E_{43})^2}{E_{43}} \end{aligned}$$

$$= 0.80856 + 0.102 + 1.44783 + 0.61974 + 0.04709 + 0.31701 + 2.45515 + 0.29538 \\ + 4.3424 + 1.02227 + 0.69534 + 0.69534 + 0.03061 \approx 12.1829$$

Determine $\chi_{v,1-\alpha}$, $v = (i - 1)(j - 1) = (4 - 1)(3 - 1) = 6$

Given a Type I error of $\alpha = 0.05$

$$\chi_{6,0.95} = 12.592$$

Given that $X^2 = 12.1829$ is less than $\chi_{6,0.95} = 12.592$, we fail to reject the null hypothesis.

END OF APPENDIX