Problem 1

(a) Which of the following corresponds to a Type I error in hypothesis testing?

A : Rejecting H_0 when H_a is true

B : Failing to reject H_0 when H_0 is true

C : Failing to reject H_0 when H_a is true

D : Rejecting H_0 when H_0 is true

(b) Which of the following statements is false?

A : If we make a Type II error, we have missed detecting an event or effect when there actually was one.

B : If we increase the probability of making a Type II error, we decrease the probability of making a Type I error.

C : The 95% confidence interval around a given sample mean is wider than the 99% confidence interval around that mean.

D : If we reject the null hypothesis at the $\alpha = 0.05$ level, then we should also reject it at the $\alpha = 0.1$ level.

(c) Which of the following is NOT a linear contrast of population means?

A : $\mu_1 - \mu_2$

B : $\mu_2 - \frac{1}{2}\mu_3 - \frac{1}{2}\mu_5$

C : $\mu_1 + \mu_2 - \mu_3 - \mu_4$

D : $\mu_1 - \mu_2 + \frac{1}{3}\mu_3 - \frac{1}{4}\mu_4$

(d) The standard deviation of GRE Quantitative scores is 10 points. A researcher decides to take a random sample of 64 students' scores to estimate the average GRE Quantitative score. What is the standard deviation of the sample mean?

A : 2

B : 1.25

C : 10

D : Can't be determined without sample mean

(e) What is the minimum sample size needed in order to estimate μ such that the 95% CI to be 2 in width if $\sigma = 5$

A : 125

B : 55

C : 97

D : 961

(f) If the true means of the J populations are equal, then MSTr/MSE should be:

A : more than 10.00

B : close to 0.00

C : close to 1.00

Problem 2

A graduate school administrator would like to know the average TOEFL (Test of English as a Foreign Language) score for international applicants. She take a random sample of 100 international applicants. Use $\bar{x}_n = 80$, $s = 21$ to answer the following questions:

(a) Construct a 95% confidence interval (using $t_{0.025,df=99} = 1.984$) for the average TOEFL score for all international applicants.

i Point estimate: $\bar{x}_n = 80$

ii Standard error (SE): $\frac{s}{\sqrt{n}} = \frac{2.1}{\sqrt{100}} = 2.1$

iii Margin of error (ME): $t_{\alpha/2,df=n-1} \times \text{SE} = 1.984 \times 2.1 = 4.1664$

iv 95% confidence interval: Point estimate $\pm$ ME $= 80 \pm 4.1664 = (75.8336, 84.1664)$

(b) One of the graduate school administrators would like to know if the average TOEFL score is greater than 79, the university minimum requirement. Conduct a hypothesis test for this assessment (using $t_{0.05,df=99} = 1.660$) and $\alpha = 0.05$ for this test).

i $H_0 : \mu = 79$ vs. $H_a : \mu > 79$

ii $t_{obs} = \frac{80-79}{\frac{2.1}{\sqrt{100}}} = \frac{1}{2.1} = 0.4762$

iii Since $t_{obs} = 0.4762 < t_{0.05,df=99} = 1.660$. Fail to reject $H_0 : \mu = 79$.

iv We do not have enough statistical evidence that the average TOEFL score is greater than 79 at 0.05 level.

(c) Compute the minimum sample size needed such that the 95% CI for average TOEFL score to be 5 in width if $\sigma = 20$

$$
\begin{aligned}
n &= \left(\frac{\sigma \times Z_{\alpha/2}}{\text{ME}}\right)^2 \\
&= \left(\frac{20 \times 1.96}{5/2}\right)^2 = 245.8624 \\
&\Rightarrow n = 246.
\end{aligned}
$$

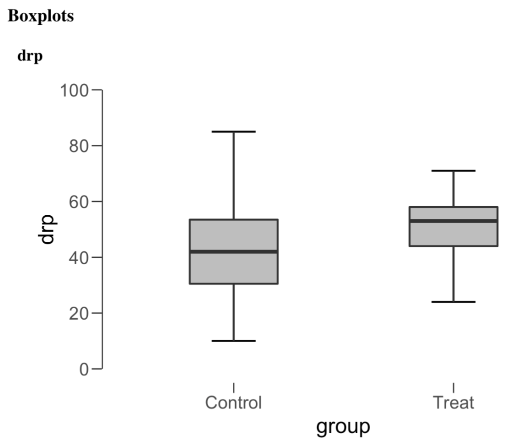
Problem 3

An educational researcher wants to evaluate the effectiveness of directed reading activities in terms of the performance of pupils on Degree of Reading Power test (DRP). She conducts an experiment with two groups of pupils, one control group and one group that was given Directed Reading Activities, and she records their reading performance. She wants to know if having directed reading activities will improve the average DRP score.

(a) What is the point estimate of the average difference (Control - Treatment)?

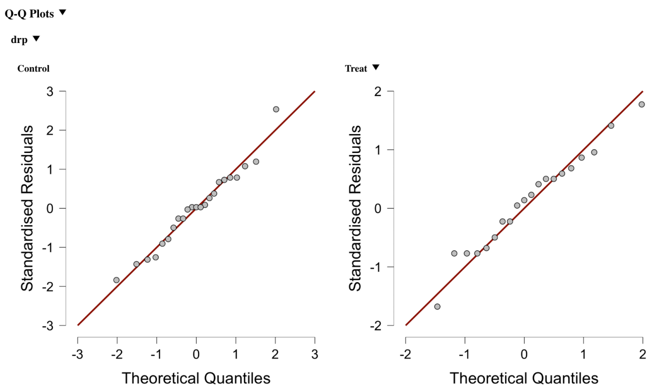
Descriptive Statistics

	drp	
	Control	Treat
Valid	23	21
Missing	0	0
Mean	41.522	51.476
Std. Deviation	17.149	11.007
Minimum	10.000	24.000
Maximum	85.000	71.000



$$\hat{\mu}_{control} - \hat{\mu}_{trt} = 41.522 - 51.476 = -9.954$$

(b) Use the QQ-plots below to comment on the normality assumption



According to the QQ-plots, all the data points are relatively close to their 1-1 line. Therefore, the normality assumption is reasonable for both control and treatment group.

(c) State the null and alternative hypotheses

$H_0 : \mu_{control} - \mu_{trt} = 0$ vs. $H_a : \mu_{control} - \mu_{trt} < 0$

(d) Draw the conclusion at 0.05 level using the JASP output below:

Independent Samples T-Test ▼

Independent Samples T-Test

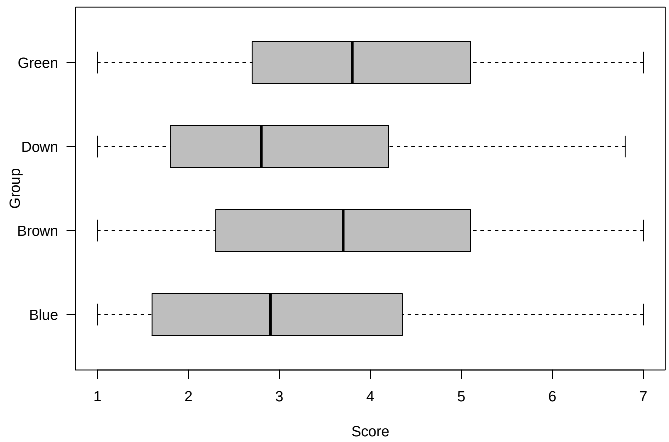
	Test	Statistic	df	p	Mean Difference	SE Difference	95% CI for Mean Difference	
							Lower	Upper
drp	Student	−2.267	42.000	0.014	−9.954	4.392	-∞	−2.567
	Welch	−2.311	37.855	0.013	−9.954	4.308	-∞	−2.691

Note. For all tests, the alternative hypothesis specifies that group *Control* is less than group *Treat*.

The P-value for both tests are less than $\alpha = 0.05$. Therefore, we do have enough statistical evidence that having directed reading activities improves the average DRP score at 0.05 level.

Problem 4

An advertising manager wants to investigate post–advertisement attitude towards a brand expressed by four different groups - each group saw the same advertisement except that the eye-color ("Blue" = blue eyes, "Brown" = brown eyes, "Green"= green eyes, "Down" = eye color cannot be seen) of the model was manipulated. The advertising manager would like to know if model's eye color affects the attitude towards the brand. Use the R output to answer the following questions:



Group <fctr>	mean <dbl>	sd <dbl>	n <int>
Blue	3.194030	1.754724	67
Brown	3.724324	1.715356	37
Down	3.107317	1.525351	41
Green	3.859740	1.665933	77

(a) State the null and alternative hypotheses

$H_0 : \mu_{Blue} = \mu_{Brown} = \mu_{Down} = \mu_{Green}$ vs. H_a: At least one mean is different than others.

(b) Fill in the missing elements in the ANOVA table below:

Source	df	SS	MS	F statistic
Group	3	SSTr = 24.4197	MSTr = 8.1399	F_{obs} = 2.8940
Error	218	SSE = 613.1387	MSE = 2.8126	
Total	221	SSTo = 637.5584		

(c) Perform an overall F-test for part (a) at 0.05 level
(using $F_{0.95,df_1=3,df_2=218} = 2.646$).

Since $F_{obs} = 2.8940 > F_{0.95,df_1=3,df_2=218} = 2.646$. We do have enough evidence that the model's eye color affects the attitude towards to the brand.

(d) Use the R output of Tukey Honest Significant Differences (HSD) test to identify the color pairs that are significantly different from each other at FWER $\bar{\alpha} = 0.1$

```
                    diff          lwr       upr      p adj
Brown-Blue    0.53029447 -0.261510938 1.3220999 0.41321572
Down-Blue    -0.08671278 -0.853231218 0.6798057 0.99377199
Green-Blue    0.66571041  0.019850181 1.3115706 0.08478926
Down-Brown   -0.61700725 -1.493593332 0.2595788 0.36810918
Green-Brown   0.13541594 -0.637880949 0.9087128 0.97770956
Green-Down    0.75242319  0.005039228 1.4998071 0.09648953
```

Blue-Green pair and Down-Green pair becuase their (adjusted) p-values are less 0.1, the $\bar{\alpha}$.

(e) Construct a 99% confidence interval (using t(0.005, 218) = 2.5986) for $L = \frac{1}{2}\mu_{\text{Green}} + \frac{1}{2}\mu_{\text{Brown}} - \frac{1}{2}\mu_{\text{Blue}} - \frac{1}{2}\mu_{\text{Down}}$

i Point estimate: $\hat{L} = \frac{1}{2}\hat{x}_{\text{Green}} + \frac{1}{2}\hat{x}_{\text{Brown}} - \frac{1}{2}\hat{x}_{\text{Blue}} - \frac{1}{2}\hat{x}_{\text{Down}} = \frac{1}{2} \times (3.859740 + 3.724324) - \frac{1}{2} \times (3.194030 + 3.107317) = 0.6414$

ii ME: $t_{\alpha/2,df} \times \sqrt{MSE \times \left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \frac{c_3^2}{n_3} + \frac{c_4^2}{n_4}\right)} = 2.5986 \times \sqrt{2.8126 \times \left(\frac{1/4}{67} + \frac{1/4}{37} + \frac{1/4}{41} + \frac{1/4}{77}\right)}$ $= 0.6336$

iii 99% confidence interval: Point estimate $\pm$ ME $= 0.6414 \pm 0.6336 = (0.0077, 1.2750)$

Problem 5

Suppose concrete cylinders for bridge supports. There are three ways of drying green concrete (A, B, and C), and you want to find the one that gives you the best compressive strength. The concrete is mixed in batches that are large enough to produce exactly three cylinders, and your production engineer believes that there is substantial variation in the quality of the concrete from batch to batch. You have data from 5 batches on each of the 3 drying processes. Use the R output below to answer the following questions:

```
Analysis of Variance Table

Response: x
          Df Sum Sq Mean Sq F value    Pr(>F)
trt        2   89.2   44.60  7.6239 0.0140226 *
blk        4  363.6   90.90 15.5385 0.0007684 ***
Residuals  8   46.8    5.85
```

(a) Test, at the 5% level of significance, whether these data provide sufficient evidence that at least one of the three treatments (A, B, and C) affects the average compressive strength

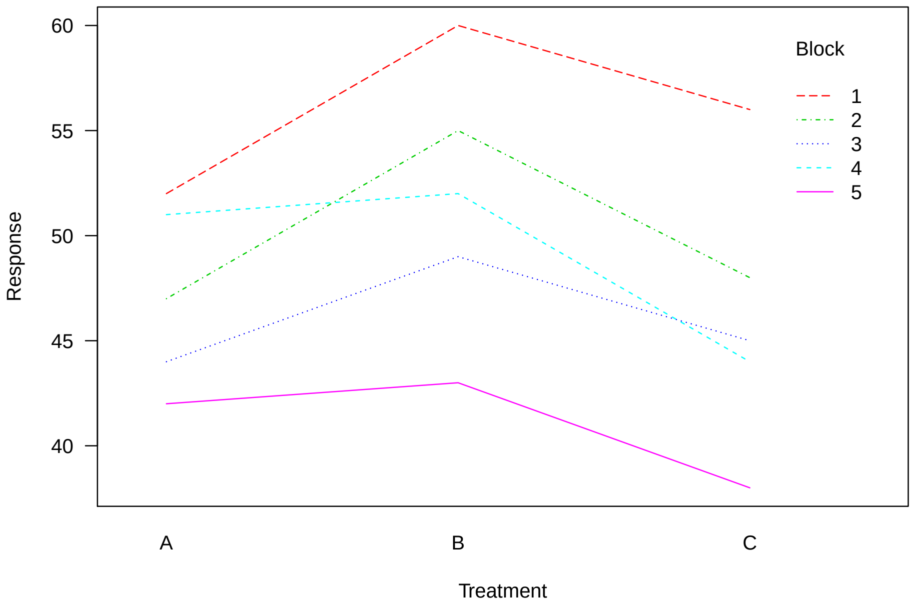
i $H_0 : \mu_A = \mu_B = \mu_C$ vs. H_a : at least one mean is different than the others.

ii $F_{obs} = 7.6239$, P-value $= 0.0140 < \alpha = 0.05$

iii Reject H_0.

iv We do have enough statistical evidence that at least one way of drying green concrete gives different average compressive strength than others at 0.05 level.

(b) Use the interaction plot below to assess the appropriateness of the assumption that treatments have the same effect in every block



Based on the interaction plot above, it is reasonable to assume that the treatment effects are consistent across blocks.