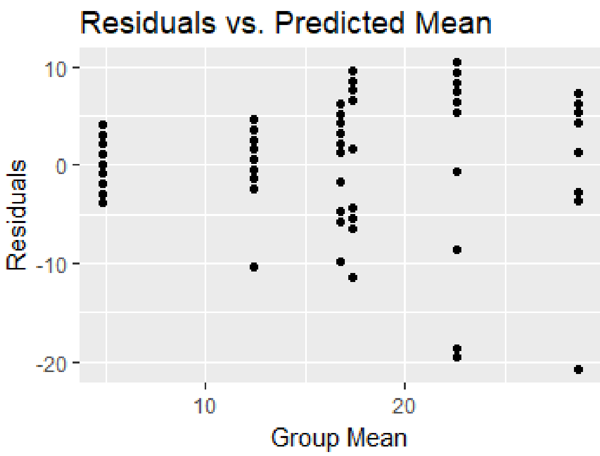
Chapter 14 and 15 Homework

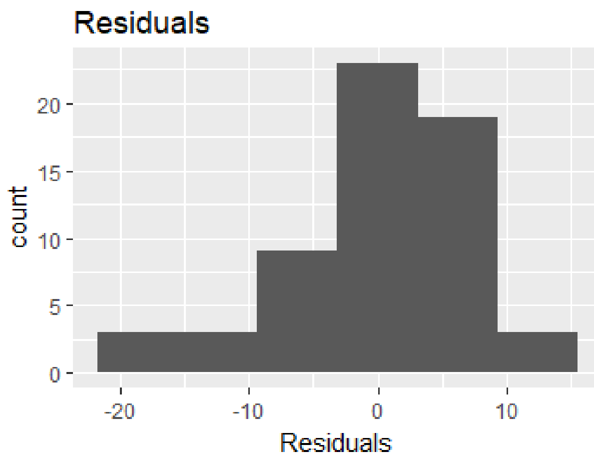
Due 4-1-2018

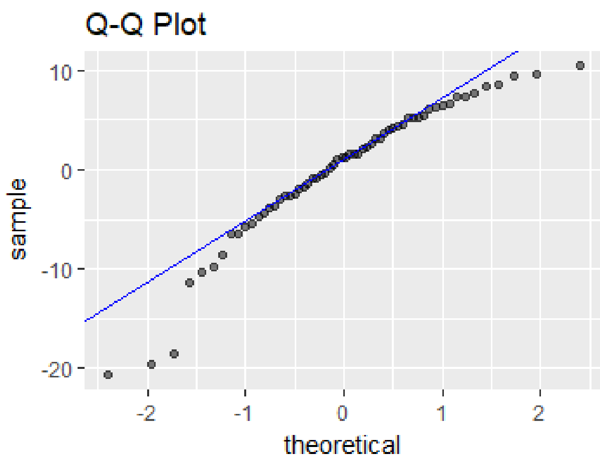
Problem 14.8

(a) Assess ANOVA assumptions using the graph from PROC MIXED.

The most concerning plot is the scatterplot with an apparent pattern as you increase the Group Mean. Though, with a wonky residual histogram and Q-Q Plot that is overpredicting at the extremes this data does not appear to fit a trustworthy ANOVA scenario.







(b) Perform an ANOVA to determine if there is an interaction between the age group and types of products; $\alpha = 0{:}05$

Table 1: ANOVA Table for Attention Spans

Row Names	SS	df	MS	F	P-Value
Main Effect					
Age Group	1349.6333	2	674.81665	12.868	2.67x10^{-5}
Products	1995.2667	1	1995.2667	38.048	9.15x10^{-8}
Interaction					
(Age Group:Products)	14.2333	2	7.1167	0.136	0.873
Error	2831.8000	54	52.4407		
Total	6190.9	59			

Check the Interaction; Reject if $F_{Interaction} > F_{\alpha,df_{Interaction},df_{Error}}$:

$F_{Interaction} = 0.873 < 3.1682 = F_{0.05,2,54}$
Thus, the interaction is not significant and can be removed from the model.

(c) If there is an interaction, create a probability plot for age group and product type and provide an interpretation for the interaction

Dealing with the Interaction Term

Since the Interaction term of AgeGroup:Products is not significant ($< F_{0.05,2,54} = 3.1682$), we will remove it from our ANOVA model and just look at the Main Effects.

(d) If there is not an interaction, remove the interaction term, recompute the ANOVA table, and determine if there are main effects of age group and product type; $\alpha = 0{:}05$

Table 2: ANOVA Table for Attention Spans without Interaction Term

Row Names	SS	df	MS	F	P-Value
Main Effect					
Age Group	1350	2	674.8	13.28	1.91x10^{-5}
Products	1995	1	1995.3	39.26	5.60x10^{-8}
Error	2846	56	50.8		
Total	6191	59			

Checking to see if the F Statistics Are Greater Than $F_{\alpha,df_{MainEffect},df_{Error}}$:

$$F_{AgeGroup} = 13.28 > 3.1619 = F_{\alpha,df_{MainEffect},df_{Error}}$$
$$F_{Products} = 13.28 > 4.0130 = F_{\alpha,df_{MainEffect},df_{Error}}$$

Both the Main Effects are statistically significant therefore, we should move onto Post Hoc testing to look at statistically different pairs.

(e) Report significantly different pairs using Tukey's W and $\alpha = 0{:}05$; REMEMBER: if there is an interaction present, we must account for that when performing post-hoc testing

To find which groups are different from each other, we will utilize Tukey's W to check which means are different from one another. We will verify this by checking which one's p-values are less than 0.05 after using the TukeyHSD test on our ANOVA model in R.

Conclusion/Interpretation:

With out Tukey W test being run, the differences lie in the following means:

- a_2 and a_3 differ from a_1
- p_2 differs from p_1

15.10

(a) Assess ANOVA assumptions using the graph from PROC MIXED.

With a Residual scatter plot that does not have a readily apparent pattern, a Residual histogram that is mound shaped, and a Q-Q plot that fits the theoretical Normal model reasonably well, I feel confident using ANOVA on this data.

(b) Determine if there is a difference between gasoline blends using ANOVA for Latin Square; $\alpha = 0{:}05$

Table 3: ANOVA Table for LSD Gasoline Brands

Row Names	SS	df	MS	F	P-Value
Treatment/Gasoline Brand	106.3	3	35.42	8.218	0.0151
Column/Car Model	8.3	3	2.78	0.644	0.6143
Row/Driver	755.4	3	251.79	58.412	7.84x10^{-5}
Error	25.9	6	4.31		
Total	895.9	15			

Hypotheses:

$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$
H_1 : at least one mean is different.

Test Statistic:

F = 8.218

Rejection Region:

Reject if $F_0 \geq F_{\alpha,df_{trt},df_{error}}; F_{\alpha,df_{trt},df_{error}} = 4.7571$

Conclusion/Interpretation:

Since $8.218 \geq 4.7571$, we have strong enough evidence to reject the hypothesis that the means are the same. There is evidence to suggest at least one mean is different from the others.

(c) If there is a difference per part (b), use Tukey's W to determine the significantly different pairs; $\alpha = 0{:}05$

To find which groups are different from each other, we will utilize Tukey's W to check which means are different from one another. We will verify this by checking which one's p-values are less than 0.05 after using the TukeyHSD test on our ANOVA model in R.

Conclusion/Interpretation:

With out Tukey W test being run, the differences lie in the following means:

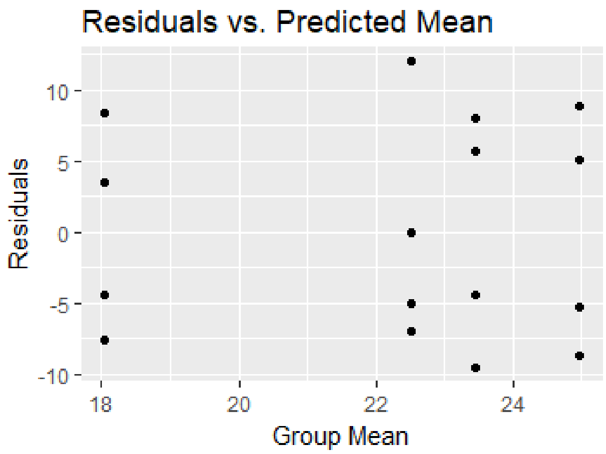
- Brand B is bigger than Brand C.
- Brand D is larger than Brand C.

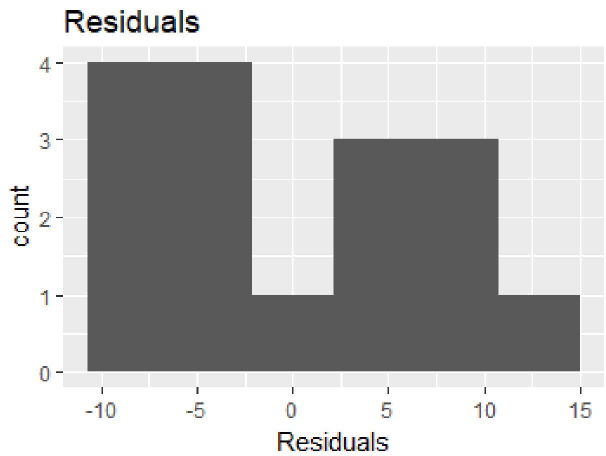
(d) From part (c), which blend (or blends) gives the highest gas mileage?

No hierarchy jumps out based on Tukey W's analysis. We can't tell which brand is on top. All we can be "sure of" ($\alpha = 0.05$) is that Brand B and Brand D are bigger than Brand C.

(e) Determine if there is a difference between gasoline blends using ANOVA for completely randomized designs (i.e., ignore the blocking factors); α = 0:05

With the following ANOVA graphs, I am hesitant to put faith in any results from the ANOVA statistical test.





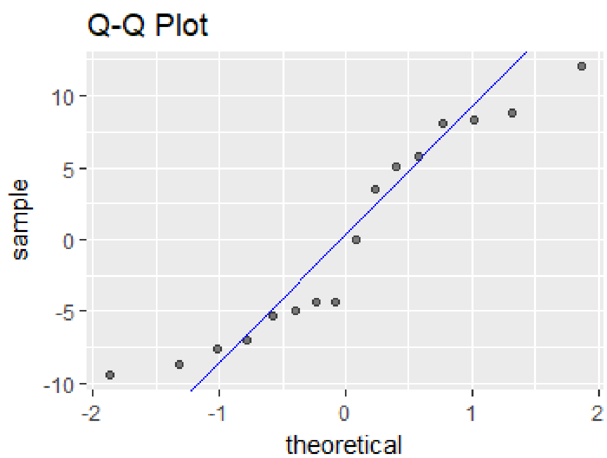


Table 4: ANOVA Table for Gasoline Brands ignoring Blocking Factors

Row Names	SS	df	MS	F	P-Value
Treatment/Gasoline Brand	106.3	3	35.4239	0.538	0.665
Error	789.6	12	65.80		
Total	895.9	15			

Hypotheses:

$H_0 : \mu_A = \mu_B = \mu_C = \mu_D$
H_1 : at least one mean is different than the rest

Test Statistic:

F = 0.538

Rejection Region:

Reject F_0 if $F_0 \geq F_{\alpha,df_{trt},df_{error}}$. $F_{\alpha,df_{trt},df_{error}} = 3.4903$

Conclusion/Interpretation:

Since our test statistic is less than the Rejection bound, we fail to reject the null hypothesis that the means are the same. We do not have strong enough evidence to suggest that there is a mean different from the rest.

(e) Compare the MSE from part (b) to the MSE from part (e) - what happens when we ignore the blocking factors?

The MSE_{Brand} for the LSD was 35.42. The MSE_{Brand} for the ANOVA not blocked on Car Model and Driver was 35.4239. Although their MSE are very similar, their difference lies in their df_{error}. Since we split on Car Model and Driver in the first group, it allows for less error to be absorbed by the error term.

(f) Which would you say is the appropriate analysis?

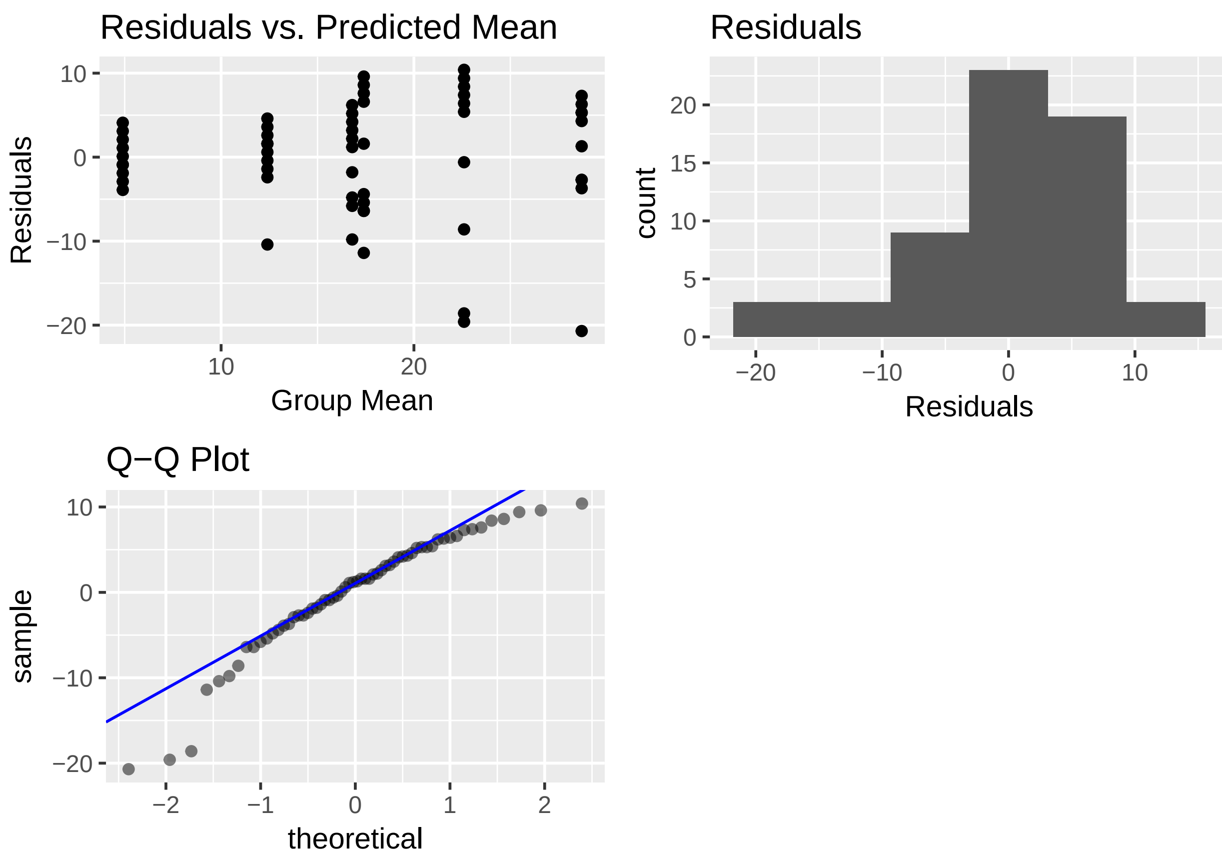
You could base this answer based solely on the PROC MIXED graphs for each group. The winner would be the ANOVA including the blocking factors. When we don't account for the blocking factors, the PROC MIXED graphs hardly represent a normal situation lending its resultant analysis null and void.

Chapter 14 and 15 HW

Kyle Ligon

March 30, 2018

14.8 (a) Assess ANOVA assumptions using the graph from PROC MIXED.



14.8 (b) Perform an ANOVA to determine if there is an interaction between the age group and types of products;

```
## [1] 3.168246
```

```
## [1] 3.161861
```

```
## [1] 4.012973
```

(e) Report significantly different pairs using Tukey's W and $\alpha = 0{:}05$; REMEMBER: if there is an interaction present, we must account for that when performing post-hoc testing

```
# if there was an interaction, we would do this.
frame_Tukey <- frame %>%
                mutate(Combination = paste0(ageGroup, "-", products))
new_anova_mod <- aov(data = frame_Tukey, formula = att_spans ~ Combination)

TukeyHSD(new_anova_mod)
```

```
##   Tukey multiple comparisons of means
```

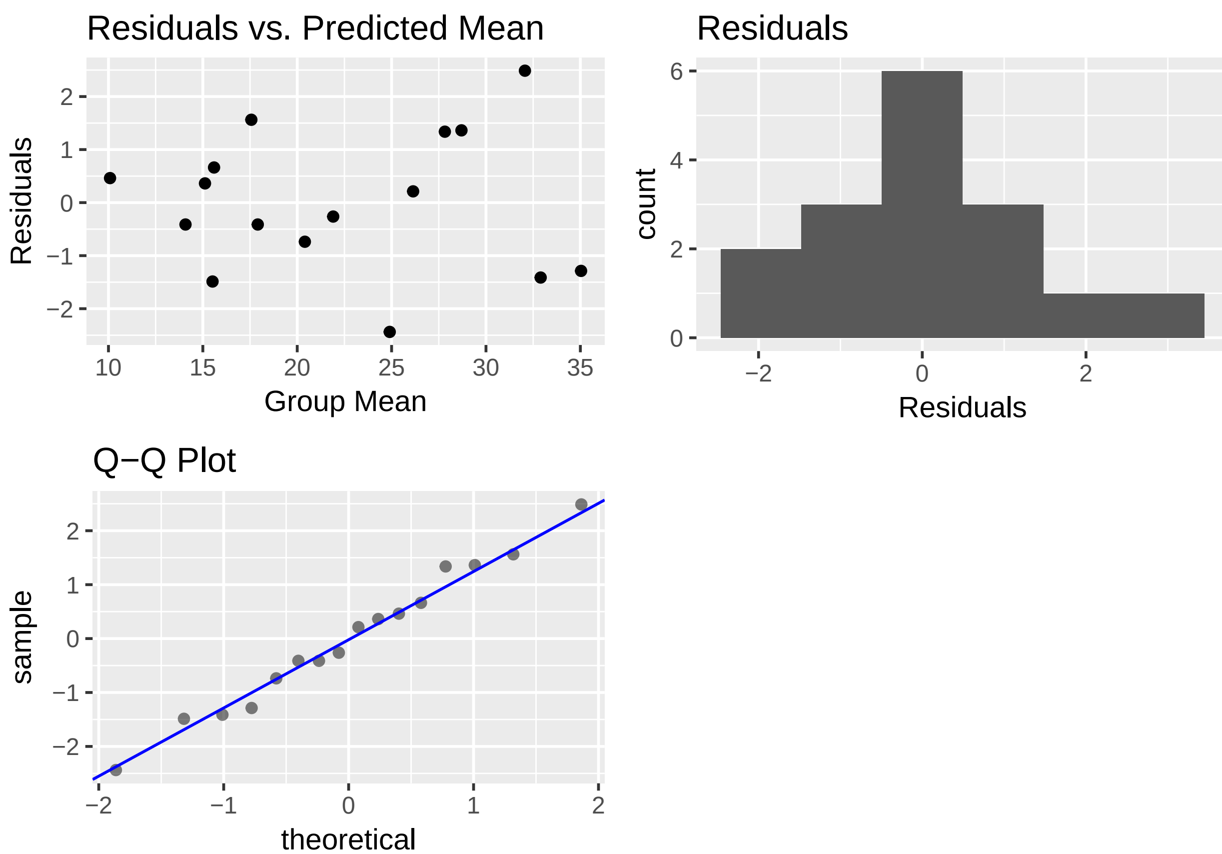
```
##     95% family-wise confidence level
##
## Fit: aov(formula = att_spans ~ Combination, data = frame_Tukey)
##
## $Combination
##              diff         lwr       upr     p adj
## a1-p2-a1-p1 12.5   2.9317963 22.068204 0.0039408
## a2-p1-a1-p1  7.5  -2.0682037 17.068204 0.2056830
## a2-p2-a1-p1 17.7   8.1317963 27.268204 0.0000174
## a3-p1-a1-p1 11.9   2.3317963 21.468204 0.0069038
## a3-p2-a1-p1 23.8  14.2317963 33.368204 0.0000000
## a2-p1-a1-p2 -5.0 -14.5682037  4.568204 0.6380006
## a2-p2-a1-p2  5.2  -4.3682037 14.768204 0.5984868
## a3-p1-a1-p2 -0.6 -10.1682037  8.968204 0.9999680
## a3-p2-a1-p2 11.3   1.7317963 20.868204 0.0118661
## a2-p2-a2-p1 10.2   0.6317963 19.768204 0.0302869
## a3-p1-a2-p1  4.4  -5.1682037 13.968204 0.7509637
## a3-p2-a2-p1 16.3   6.7317963 25.868204 0.0000807
## a3-p1-a2-p2 -5.8 -15.3682037  3.768204 0.4800463
## a3-p2-a2-p2  6.1  -3.4682037 15.668204 0.4231404
## a3-p2-a3-p1 11.9   2.3317963 21.468204 0.0069038
```

```
#but since there's not, we can just do the Tukey W on the old ANOVA model
TukeyHSD(anova_mod_noI)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = att_spans ~ ageGroup + products, data = frame)
##
## $ageGroup
##        diff        lwr      upr     p adj
## a2-a1  6.35  0.9224517 11.77755 0.0180817
## a3-a1 11.60  6.1724517 17.02755 0.0000105
## a3-a2  5.25 -0.1775483 10.67755 0.0599740
##
## $products
##           diff      lwr      upr p adj
## p2-p1 11.53333 7.845991 15.22068 1e-07
```

From this we can tell that a2 and a3 differes from a1. Additionally, p2 is different from p1.

15.10 (a) Assess ANOVA assumptions using the graph from PROC MIXED.

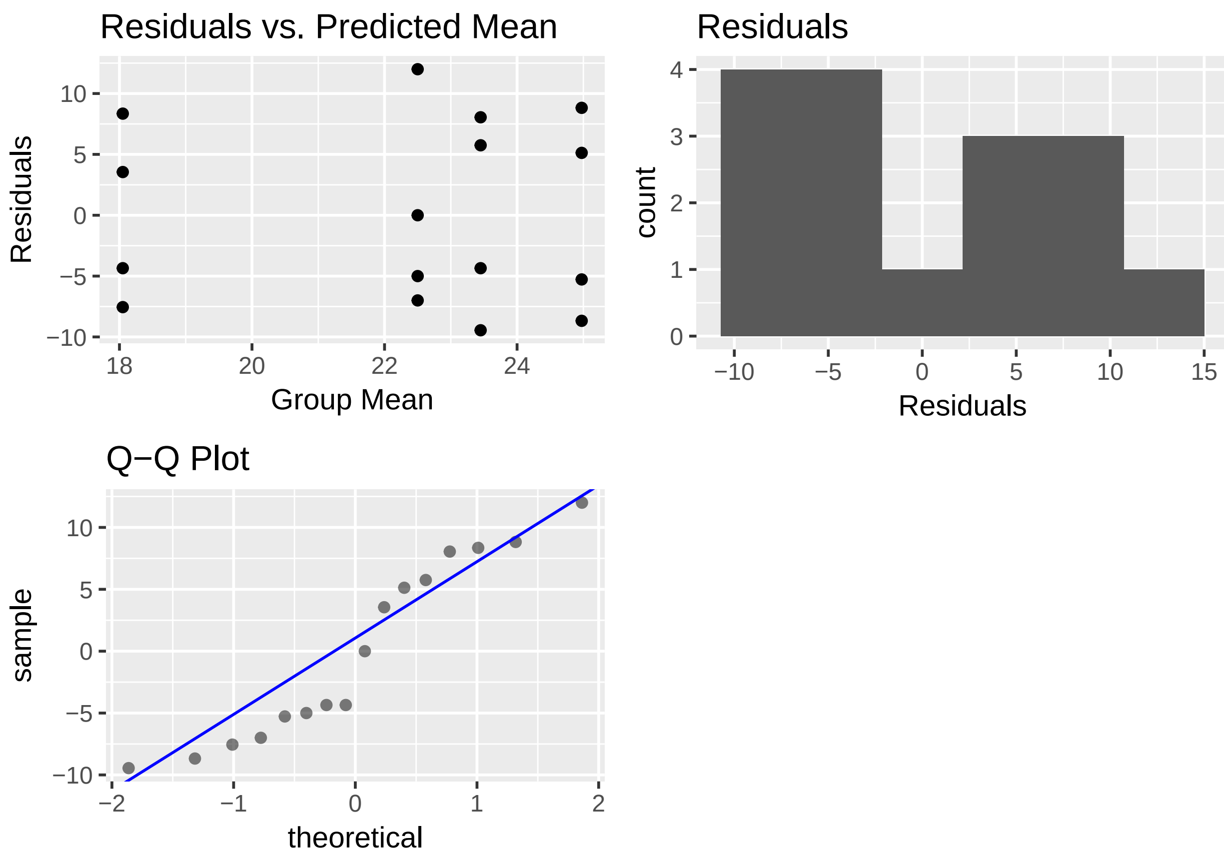


(c) If there is a difference per part (b), use Tukey's W to determine the significantly different pairs; $\alpha = 0{:}05$

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = mileage ~ car_model + driver + seed, data = petrol)
##
## $car_model
##           diff       lwr      upr     p adj
## cm2-cm1 -1.975 -7.057132 3.107132 0.5709528
## cm3-cm1 -0.650 -5.732132 4.432132 0.9686626
## cm4-cm1 -0.600 -5.682132 4.482132 0.9749663
## cm3-cm2  1.325 -3.757132 6.407132 0.8045388
## cm4-cm2  1.375 -3.707132 6.457132 0.7877151
## cm4-cm3  0.050 -5.032132 5.132132 0.9999841
##
## $driver
##          diff        lwr        upr     p adj
## d2-d1  17.475  12.392868 22.5571325 0.0000896
## d3-d1   3.425  -1.657132  8.5071325 0.1921707
## d4-d1  11.775   6.692868 16.8571325 0.0008306
## d3-d2 -14.050 -19.132132 -8.9678675 0.0003106
## d4-d2  -5.700 -10.782132 -0.6178675 0.0310697
## d4-d3   8.350   3.267868 13.4321325 0.0051400
##
## $seed
##       diff         lwr        upr     p adj
```

```
## B-A  2.475  -2.6071325  7.5571325 0.4053971
## C-A -4.450  -9.5321325  0.6321325 0.0828995
## D-A  0.950  -4.1321325  6.0321325 0.9128431
## C-B -6.925 -12.0071325 -1.8428675 0.0128824
## D-B -1.525  -6.6071325  3.5571325 0.7351363
## D-C  5.400   0.3178675 10.4821325 0.0390458
```

e) Determine if there is a difference between gasoline blends using ANOVA for completely randomized designs (i.e., ignore the blocking factors); $\alpha = 0{:}05\}$




```
##             Df Sum Sq Mean Sq F value Pr(>F)
## seed         3  106.3   35.42   0.538  0.665
## Residuals   12  789.6   65.80
```

```
## [1] 3.490295
```