Chapter 25 – Analysis of Variance

Section 25.1

1. Popcorn

a) H_0: The mean number of unpopped kernels is the same for all four brands of popcorn. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$
H_A: The mean number of unpopped kernels is not the same for all four brands of popcorn.

b) MS_T has $k-1 = 4-1 = 3$ degrees of freedom.
MS_E has $N-k = 16-4 = 12$ degrees of freedom.

c) The F-statistic is 13.56 with 3 and 12 degrees of freedom, resulting in a P-value of 0.00037. We reject the null hypothesis and conclude that there is strong evidence that the mean number of unpopped kernels is not the same for all four brands of popcorn.

d) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

2. Skating.

a) H_0: The mean score is the same for each focus. $(\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5)$
H_A: The mean scores for each focus are not all the same.

b) MS_T has $k-1 = 5-1 = 4$ degrees of freedom.
MS_E has $N-k = 30-5 = 25$ degrees of freedom.

c) The F-statistic is 7.43 with 4 and 25 degrees of freedom, resulting in a P-value of 0.00044. We reject the null hypothesis and conclude that there is strong evidence that the mean scores for each focus are not all the same.

d) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

3. Gas mileage.

a) H_0: The mean gas mileage is the same for each muffler. $(\mu_1 = \mu_2 = \mu_3)$
H_A: The mean gas mileages for each muffler are not all the same.

b) MS_T has $k-1 = 3-1 = 2$ degrees of freedom.
MS_E has $N-k = 24-3 = 21$ degrees of freedom.

c) The F-statistic is 2.35 with 2 and 21 degrees of freedom, resulting in a P-value of 0.1199. We fail to reject the null hypothesis and conclude that there is no evidence to suggest that gas mileage associated with any single muffler is different than the others.

d) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

e) By failing to notice that one of the mufflers resulted in significantly different gas mileage, you have committed a Type II error.

4. Darts.

a) H_0: The mean distance from dart to bull's-eye is the same for all four stances. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$

H_A: The mean distances from dart to bull's-eye are not all the same.

b) MS_T has $k-1=4-1=3$ degrees of freedom.

MS_E has $N-k=40-4=36$ degrees of freedom.

c) The F-statistic is 1.41 with 3 and 36 degrees of freedom, resulting in a P-value of 0.2557. We fail to reject the null hypothesis and conclude that there is no evidence that the mean distances from dart to target are not all the same.

d) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

e) By failing to notice that one of the stances had a mean distance from dart to bull's-eye significantly different than the others, you have made a Type II error.

Section 25.2

5. Activating baking yeast.

a) H_0: The mean activation time is the same for all four recipes. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$

H_A: The mean activation times for each recipe are not all the same.

b) The F-statistic is 44.7392 with 3 and 12 degrees of freedom, resulting in a P-value less than 0.0001. We reject the null hypothesis and conclude that there is strong evidence that the mean activation times for each recipe are not all the same.

c) Yes, it would be appropriate to follow up with multiple comparisons, because we have rejected the null hypothesis.

6. Frisbee throws.

a) H_0: The mean distance is the same for each grip. $(\mu_1 = \mu_2 = \mu_3)$

H_A: The mean distances for each grip are not all the same.

b) The F-statistic is 2.0453 with 2 and 21 degrees of freedom, resulting in a P-value equal to 0.1543. We fail to reject the null hypothesis and conclude that there is not enough evidence that the mean distances for each type of grip are not all the same.

c) No, it would not be appropriate to follow up with multiple comparisons, because we have failed to reject the null hypothesis.

Section 25.3

7. Cars.

a) H_0: The mean mileage is the same for each engine type (number of cylinders). $(\mu_4 = \mu_5 = \mu_6 = \mu_8)$

H_A: The mean mileages for each engine type are not all the same.

b) The Similar Variance condition is not met, because the boxplots show distributions with radically different spreads. A re-expression of the response variable may equalize the spreads, allowing us to proceed with an analysis of variance.

8. Wine production.

a) H_0: The mean bottle price is the same at all three locations. $(\mu_C = \mu_K = \mu_S)$
H_A: The mean bottle prices for each location are not all the same.

b) The spreads of *Bottle Price* do not appear to be similar, violating the Similar Variance Condition. There are also several outliers in the Seneca group. It would be a good idea to look at a plot of the residuals from an *ANOVA* vs. *the predicted values* to see if the variance increases with predicted value. If the analysis is performed, it would be a good idea to compute the analysis both with and without the Seneca outliers.

Section 25.4

9. Tellers.

a) H_0: The mean time to serve a customer is the same for each teller. $(\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6)$
H_A: The mean times to serve a customer for each teller are not all the same.

b) The *F*-statistic is 1.508 with 5 and 134 degrees of freedom, resulting in a P-value equal to 0.1914. We fail to reject the null hypothesis and conclude that there is not enough evidence that the mean times to serve a customer for each teller are not all the same.

c) No, it would not be appropriate to follow up with multiple comparisons, because we have failed to reject the null hypothesis.

10. Hearing 4 lists.

a) H_0: The mean hearing score is the same for all four lists. $(\mu_1 = \mu_2 = \mu_3 = \mu_4)$
H_A: The mean hearing scores for each list are not all the same.

b) The *F*-statistic is 4.9192 with 3 and 92 degrees of freedom, resulting in a P-value equal to 0.0033. We reject the null hypothesis and conclude that there is strong evidence that the mean hearing scores for each list are not all the same.

c) Yes, it would be appropriate to follow up with multiple comparisons, because we have rejected the null hypothesis.

Section 25.5.

11. Eye and hair color.

An analysis of variance is not appropriate, because *eye color* is a categorical variable. The students could consider a chi-square test of independence.

12. Zip codes.

An analysis of variance is not appropriate, because *zip code* is a categorical variable. The students could consider categorizing zip codes by the first digit, corresponding to the region of the country (0 on the east coast to 9 on the west coast), and performing a chi-square test of independence.

Chapter Exercises.

13. Yogurt.

a) MS_T has $k-1=3-1=2$ degrees of freedom.

$$MS_T = \frac{SS_T}{df_T} = \frac{17.300}{2} = 8.65$$

MS_E has $N-k=9-3=6$ degrees of freedom.

$$MS_E = \frac{SS_E}{df_E} = \frac{0.4600}{6} \approx 0.0767$$

b) $F\text{-statistic} = \dfrac{MS_T}{MS_E} = \dfrac{8.65}{0.0767} \approx 112.78$

c) With a P-value equal to 0.000017, there is very strong evidence that the mean taste test scores for each method of preparation are not all the same.

d) We have assumed that the experimental runs were performed in random order, that the variances of the treatment groups are equal, and that the errors are Normal.

e) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

f) $s_p = \sqrt{MS_E} = \sqrt{0.0767} \approx 0.277$ points

14. Smokestack scrubbers.

a) MS_T has $k-1=4-1=3$ degrees of freedom.

$$MS_T = \frac{SS_T}{df_T} = \frac{81.2}{3} \approx 27.067$$

MS_E has $N-\text{k}=20-4=16$ degrees of freedom.

$$MS_E = \frac{SS_E}{df_E} = \frac{30.8}{16} = 1.925$$

b) $F\text{-statistic} = \dfrac{MS_T}{MS_E} = \dfrac{27.067}{1.925} \approx 14.0606$

c) With a P-value equal to 0.00000949, there is very strong evidence that the mean particulate emissions for each smokestack scrubber are not all the same.

d) We have assumed that the experimental runs were performed in random order, that the variances of the treatment groups are equal, and that the errors are Normal.

e) To check the Similar Variance condition, look at side-by-side boxplots of the treatment groups to see whether they have similar spreads. To check the Nearly Normal condition, look to see if a normal probability plot of the residuals is straight, look to see that a histogram of the residuals is nearly normal, and look to see if a residuals plot shows no pattern, and no systematic change in spread.

f) $s_p = \sqrt{MS_E} = \sqrt{1.925} \approx 1.387\,\text{ppb}$

15. Eggs.

a) H_0: The mean taste test scores are the same for both real and substitute eggs. $(\mu_R = \mu_S)$

H_A: The mean taste test scores are different. $(\mu_R \neq \mu_S)$

b) The F-statistic is 31.0712 with 1 and 6 degrees of freedom, resulting in a P-value equal to 0.0014. We reject the null hypothesis and conclude that there is strong evidence that the mean taste test scores for real and substitute eggs are different. The real eggs have a higher mean taste test score.

c) The Similar Variance assumption is a bit of a problem. The spread of the distribution of taste test scores for real eggs looks greater than the spread of the distribution of taste test scores for substitute eggs. Additionally, we cannot check the Nearly Normal condition because we don't have a residuals plot. Caution should be used in making any conclusions.

d) H_0: The mean taste test score of brownies made with real eggs is the same as the mean taste test score of brownies made with substitute eggs. $(\mu_R = \mu_S \text{ or } \mu_R - \mu_S = 0)$

H_A: The mean taste test score of brownies made with real eggs is different than the mean taste test score of brownies made with substitute eggs. $(\mu_R \neq \mu_S \text{ or } \mu_R - \mu_S \neq 0)$

Independent groups assumption: Taste test scores for each type of brownie should be independent.
Randomization condition: Brownies were tasted in random order.
Nearly Normal condition: We don't have the actual data, so we can't check the distribution of the sample. We will need to assume that the distribution of taste test scores is unimodal. The boxplots show distributions that are at least symmetric.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's t-model, with 6 degrees of freedom ($n_R + n_S - 2 = 6$ for a pooled t-test). We will perform a two-sample t-test.

The pooled sample variance is $s_p^2 = \dfrac{(4-1)(0.651)^2 + (4-1)(0.395)^2}{(4-1)+(4-1)} \approx 0.2899$

The sampling distribution model has mean 0, with standard error

$$SE_{pooled}(\bar{y}_R - \bar{y}_S) = \sqrt{\frac{0.2899}{4} + \frac{0.2899}{4}} \approx 0.3807.$$

The observed difference between the mean scores is 6.78 – 4.66 = 2.12.

$$t = \frac{(\bar{y}_R - \bar{y}_T) - (0)}{SE_{pooled}(\bar{y}_R - \bar{y}_T)} \approx \frac{2.12 - 0}{0.3807} \approx 5.5687$$

With a $t = 5.5687$ and 6 degrees of freedom, the two-sided P-value is 0.0014. Since the P-value is low, we reject the null hypothesis. There is strong evidence that the mean taste test score for brownies made with real eggs is different from the mean taste test score for brownies made with substitute eggs. In fact, there is evidence that the brownies made with real eggs taste better.

The P-value for the 2-sample pooled t-test, 0.0014, which is the same as the P-value for the analysis of variance test. The F-statistic, 31.0712, was approximately the same as $t^2 = (5.5727)^2 \approx 31.05499$. Also, the pooled estimate of the variance, 0.2899, is approximately equal to MS_E , the mean squared error. This is because an analysis of variance for two samples is equivalent to a 2-sample pooled t-test.

16. Auto noise filters.

a) H_0: The mean noise level is the same for both types of filters. $(\mu_1 = \mu_2)$

H_A: The mean noise levels are different. $(\mu_1 \neq \mu_2)$

b) The *F*-statistic is 0.7673 with 1 and 33 degrees of freedom, resulting in a P-value equal to 0.3874. We fail to reject the null hypothesis and conclude that there is not enough evidence to suggest that the mean noise levels are different for each type of filter.

c) The Similar Variance assumption seems reasonable. The spreads of the two distributions look similar. We cannot check the Nearly Normal condition because we don't have a residuals plot.

d) H_0: The mean noise level is the same for both types of filters. $(\mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0)$

H_A: The mean noise level is different for each type of filters. $(\mu_1 \neq \mu_2 \text{ or } \mu_1 - \mu_2 \neq 0)$

Independent groups assumption: Noise levels for the different types of filters should be independent.
Randomization condition: Assume that this study was conducted using random allocation of the noise filters.
Nearly Normal condition: This might cause some difficulty. The distribution of the noise levels for the new device is skewed. However, the sample sizes are somewhat large, so the CLT will help to minimize any difficulty.

Since the conditions are satisfied, it is appropriate to model the sampling distribution of the difference in means with a Student's *t*-model, with 33 degrees of freedom ($n_1 + n_2 - 2 = 33$ for a pooled *t*-test). We will perform a two-sample *t*-test.

The pooled sample variance is $s_p^2 = \dfrac{(18-1)(3.2166)^2 + (17-1)(2.43708)^2}{(18-1)+(17-1)} \approx 8.2097$

The sampling distribution model has mean 0, with standard error:

$$SE_{pooled}(\bar{y}_1 - \bar{y}_2) = \sqrt{\frac{8.2097}{18} + \frac{8.2097}{17}} \approx 0.9690.$$

The observed difference between the mean scores is 81.5556 – 80.7059 = 0.8497.

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (0)}{SE_{pooled}(\bar{y}_1 - \bar{y}_2)} \approx \frac{0.8497 - 0}{0.9690} \approx 0.8769$$

With a $t = 0.877$ and 33 degrees of freedom, the two-sided P-value is 0.3874. Since the P-value is high, we fail to reject the null hypothesis. There is no evidence that the mean noise levels for the two filters are different.

The P-value for the 2-sample pooled *t*-test, 0.3874, which is the same as the P-value for the analysis of variance test. The *F*-statistic, 0.7673, was approximately the same as $t^2 = (0.8770)^2 \approx 0.7673$. Also, the pooled estimate of the variance, 8.2097, is approximately equal to MS_E, the mean squared error. This is because an analysis of variance for two samples is equivalent to a 2-sample pooled *t*-test.

17. School system.

a) H_0: The mean math test score is the same at each of the 15 schools. $(\mu_A = \mu_B = \ldots = \mu_O)$

H_A: The mean math test scores are not all the same at each of the 15 schools.

b) The *F*-statistic is 1.0735 with 14 and 105 degrees of freedom, resulting in a P-value equal to 0.3899. We fail to reject the null hypothesis and conclude that there is not enough evidence to suggest that the mean math scores are not all the same for the 15 schools.

c) This does not match our findings in part (b). Because the intern performed so many *t*-tests, he may have committed several Type I errors. (Type I error is the probability of rejecting the null hypothesis when there is actually no difference between the schools.) Some tests would be expected to result in Type I error due to chance alone. The overall Type I error rate is higher for multiple *t*-tests than for an analysis of variance, or a multiple comparisons method. Because we failed to reject the null hypothesis in the analysis of variance, we should not do multiple comparison tests.

18. Fertilizers.

a) H_0: The mean height of beans is the same for each of the 10 fertilizers. $(\mu_A = \mu_B = \ldots = \mu_J)$

H_A: The mean heights of beans are not all the same for each of the 10 fertilizers.

b) The *F*-statistic is 1.1882 with 9 and 110 degrees of freedom, resulting in a P-value equal to 0.3097. We fail to reject the null hypothesis and conclude that there is not enough evidence to suggest that the mean bean heights are not all the same for the 10 fertilizers.

c) This does not match our findings in part (b). Because the lab partner performed so many *t*-tests, he may have committed several Type I errors. (Type I error is the probability of rejecting the null hypothesis when there is actually no difference between the fertilizers.) Some tests would be expected to result in Type I error due to chance alone. The overall Type I error rate is higher for multiple *t*-tests than for an analysis of variance, or a multiple comparisons method. Because we failed to reject the null hypothesis in the analysis of variance, we should not do multiple comparison tests.

19. Cereals.

a) H_0: The mean sugar content is the same for each of the 3 shelves. $(\mu_1 = \mu_2 = \mu_3)$

H_A: The mean sugar contents are not all the same for each of the 3 shelves.

b) The *F*-statistic is 7.2721 with 2 and 74 degrees of freedom, resulting in a P-value equal to 0.0013. We reject the null hypothesis and conclude that there is strong evidence to suggest that the mean sugar content of cereal on at least one shelf differs from that on other shelves.

c) We cannot conclude that cereals on shelf two have a higher mean sugar content than cereals on shelf three, or that cereals on shelf two have a higher mean sugar content than cereals on shelf one. We can conclude only that the mean sugar contents are not all equal.

d) We can conclude that the mean sugar content on shelf two is significantly different from the mean sugar contents on shelves one and three. In fact, there is evidence that the mean sugar content on shelf two is greater than that of shelf one and shelf three.

20. Cereals protein by shelf.

a) H_0: The mean protein content is the same for each of the 3 shelves. $(\mu_1 = \mu_2 = \mu_3)$
H_A: The mean protein contents are not all the same for each of the 3 shelves.

b) The *F*-statistic is 5.8445 with 2 and 74 degrees of freedom, resulting in a P-value equal to 0.0044. We reject the null hypothesis and conclude that there is strong evidence to suggest that the mean protein content of cereal on at least one shelf differs from that on other shelves.

c) We cannot conclude that cereals on shelf two have a higher mean protein content than cereals on shelf three, or that cereals on shelf two have a higher mean protein content than cereals on shelf one. We can conclude only that the mean protein contents are not all equal.

d) We can conclude that the mean protein content on shelf two is significantly different from the mean protein contents on shelf three. In fact, there is evidence that the mean protein content on shelf three is greater than that of shelf two. The other pairwise comparisons are not significant at $\alpha = 0.05$.

21. Downloading.

a) H_0: The mean download time is the same for each of the 3 times of day. $(\mu_{Early} = \mu_{Evening} = \mu_{Late})$
H_A: The mean download times are not all the same for each of the 3 times of day.

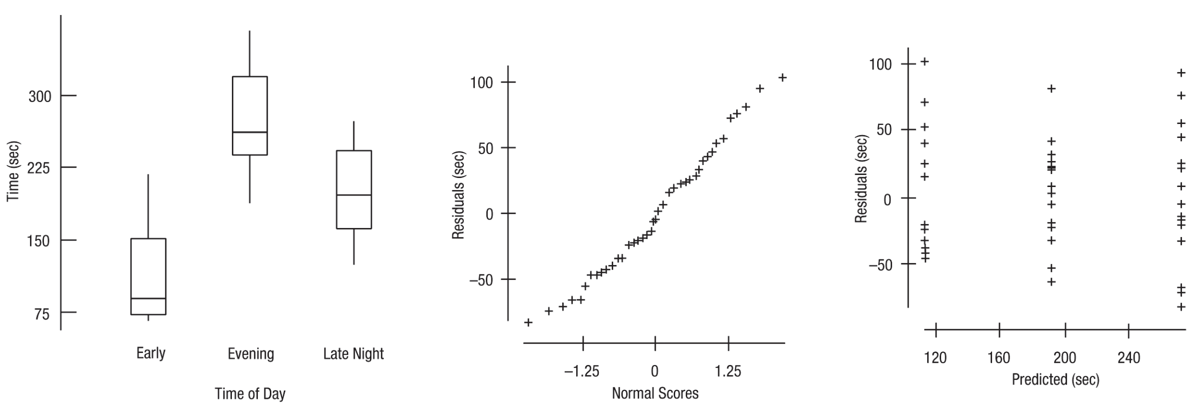
b)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F-ratio	P-Value
Time of Day	2	204,641	102,320	46.035	< 0.0001
Error	45	100,020	2222.67		
Total	47	304,661			

The *F*-statistic is 46.035 with 2 and 45 degrees of freedom, resulting in a P-value less than 0.0001. We reject the null hypothesis and conclude that there is strong evidence to suggest that the mean download time is different in at least one of the 3 times of day.

c) **Randomization condition:** The runs were not randomized, but it is likely that they are representative of all download times at these times of day.
Similar Variance condition: The boxplots show similar spreads for the distributions of download times for the different times of day.
Nearly Normal condition: The normal probability plot of residuals is reasonably straight.



d) A Bonferroni test shows that all three pairs are different from each other at $\alpha = 0.05$.

22. Analgesics.

a) H_0: The mean pain level reported is the same for each of the three drugs. $(\mu_A = \mu_B = \mu_C)$

H_A: The mean pain levels reported are not all the same for the three drugs.

b)

Analysis of Variance

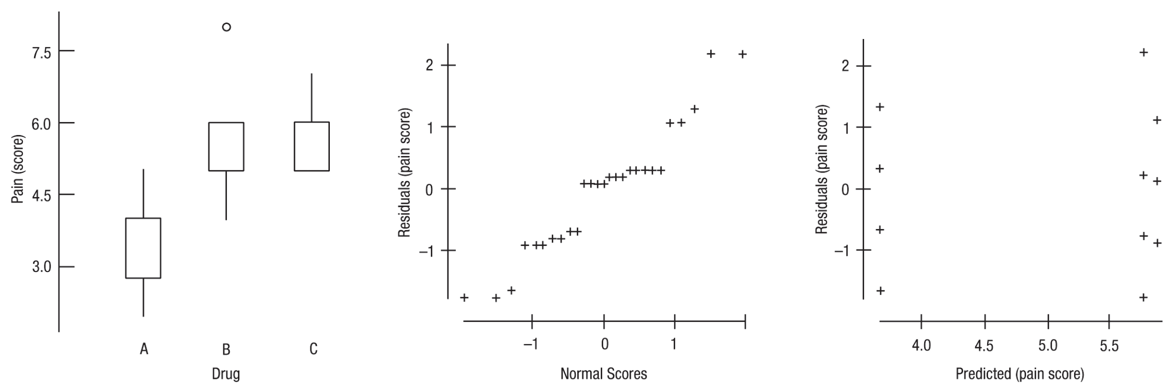
Source	DF	Sum of Squares	Mean Square	F-ratio	P-Value
Drug	2	28.2222	14.1111	11.9062	0.0003
Error	24	28.4444	1.1852		
Total	26	56.6666			

The F-statistic is 11.9062 with 2 and 24 degrees of freedom, resulting in a P-value equal to 0.0003. We reject the null hypothesis and conclude that there is strong evidence that the mean pain level reported is different in at least one of the three drugs.

Means and Standard Deviations

Level	n	Mean	Standard Deviation
A	9	3.667	0.866
B	9	5.778	1.481
C	9	5.889	0.782

c) **Randomization condition:** The volunteers were randomly allocated to treatment groups.
Similar Variance condition: The boxplots show similar spreads for the distributions of pain levels for the different drugs, although the outlier for drug B may cause some concern.
Nearly Normal condition: The normal probability plot of residuals is reasonably straight.



d) A Bonferroni test shows that drug A's mean pain level as reported is significantly below the other two, but that drug B's and C's means are indistinguishable at the $\alpha = 0.05$ level.