8.8 Since there are four groups, degrees of freedom for the groups is $4-1=3$. Since the total number of data values is 40, the total degrees of freedom is $40-1=39$. This leaves 36 degrees of freedom for the Error (or use $n - \#groups = 40-4=36$). The Mean Squares are found by dividing each SS by its df, so we compute

$$MSG = 960/3 = 320 \quad \text{(for Groups)} \qquad \text{and} \qquad MSE = 5760/36 = 160 \quad \text{(for Error)}$$

The F-statistic is the ratio of the mean square values, we have

$$F = \frac{MSG}{MSE} = \frac{320}{160} = 2.0$$

The completed table is shown below.

Source	df	SS	MS	F-statistic
Groups	3	960	320	2.0
Error	36	5760	160	
Total	39	6720		

8.10 Since there are four groups, degrees of freedom for the groups is $4-1=3$. Since the total number of data values is $5+8+7+5=25$, the total degrees of freedom is $25-1=24$. This leaves 21 degrees of freedom for the Error (or use $n - \#groups = 25-4=21$). We also find the missing sum of squares for Groups by subtraction, $SSG = SSTotal - SSE = 1400 - 800 = 600$. The Mean Squares are found by dividing each SS by its df, so we compute

$$MSG = 600/3 = 200 \quad \text{(for Groups)} \qquad \text{and} \qquad MSE = 800/21 = 38.095 \quad \text{(for Error)}$$

The F-statistic is the ratio of the mean square values, we have

$$F = \frac{MSG}{MSE} = \frac{200}{38.095} = 5.25$$

The completed table is shown below.

Source	df	SS	MS	F-statistic
Groups	3	600	200	5.25
Error	21	800	38.095	
Total	24	1400		

8.12 (a) Since the degrees of freedom for the groups is 4, the number of groups is 5.

(b) The hypotheses are

$$\begin{aligned} H_0 &: \quad \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \\ H_a &: \quad \text{Some } \mu_i \neq \mu_j \end{aligned}$$

(c) Using 4 and 35 for the degrees of freedom with the F-distribution, we see the upper-tail area beyond F=5.71 gives a p-value of 0.0012.

(d) We reject H_0. There is strong evidence that the population means are not all the same.

8.14 (a) Since the degrees of freedom for the groups is 3, the number of groups is 4.

(b) The hypotheses are

$$\begin{aligned} H_0: &\quad \mu_1 = \mu_2 = \mu_3 = \mu_4 \\ H_a: &\quad \text{Some } \mu_i \neq \mu_j \end{aligned}$$

(c) Using 3 and 16 for the degrees of freedom with the F-distribution, we see the upper-tail area beyond F=0.75 gives a p-value of 0.538.

(d) We do not reject H_0. We do not find convincing evidence of any differences between the population means.

8.16 (a) Using μ_r, μ_g, and μ_b to represent mean number of anagrams solved by someone with prior exposure to red, green, and black, respectively, the hypotheses are

$$\begin{aligned} H_0: &\quad \mu_r = \mu_g = \mu_b \\ H_a: &\quad \text{At least two of the means are different} \end{aligned}$$

(b) We subtract to see that $SSE = SStotal - SSG = 84.7 - 27.7 = 57.0$. Since there are three groups, df for color is $3 - 1 = 2$. Since the total sample size is $19 + 27 + 25 = 71$, total df is $71 - 1 = 70$. Thus, the error df is 68. We divide by the respective degrees of freedom to find MSG and MSE and then divide those to find the F-statistic. The analysis of variance table is shown, and the F-statistic is 16.5.

Source	DF	SS	MS	F
Groups	2	27.7	13.85	16.5
Error	68	57.0	0.84	
Total	70	84.7		

(c) We find the area above $F = 16.5$ in an F-distribution with numerator df equal to 2 and denominator df equal to 68, and see that the p-value is essentially zero.

(d) Reject H_0 and conclude that the means are not all the same. The color of prior instructions has an effect on how students perform on this anagram test. By looking at the sample group means, it appears that seeing red prior to the test may hinder students ability to solve the anagrams.

8.24 (a) The mice in bright light gained the most ($\overline{x}_{LL} = 11.01$), while the mice in the normal light/dark cycle gained the least ($\overline{x}_{LD} = 5.93$).

(b) Yes, the groups have similar variability since no standard deviation for a group is more than twice another standard deviation.

(c) We have

$$z\text{-score} = \frac{x - \overline{x}}{s} = \frac{17.4 - 11.01}{2.624} = 2.435$$

This value is 2.435 standard deviations above the mean, which causes some concern but allows us to proceed with the analysis.

(d) The cases are the mice. There are two variables. Which light group a mouse is assigned to is categorical while body mass gain over four weeks is quantitative.

8.25 (a) The null hypothesis is that the amount of light at night does not affect how much weight is gained. The alternative hypothesis is that the amount of light at night has some effect on mean weight gain.

(b) We see from the computer output that the F-statistic is 8.38 while the p-value is 0.002. This is a small p-value, so we reject H_0. There is evidence that mean weight gain in mice is influenced by the amount of light at night.

(c) Yes, there is an association between weight gain and light conditions at night. From the means of the groups, it appears that mice with more light at night tend to gain more weight.

(d) Yes, we can conclude that light at night causes weight gain (in mice), since the results are significant and come from a randomized experiment.

8.26 The null hypothesis is that mean activity level is not related to light condition, while the alternative hypothesis is that the mean activity level is related to light condition. The F-statistic is 0.09 and the p-value is 0.910. This is a very large p-value, so we do not reject H_0. There is no convincing evidence at all that mean activity level of mice is different depending on the amount of light at night. The differences in weight gain in the previous exercise are not caused by different activity levels.

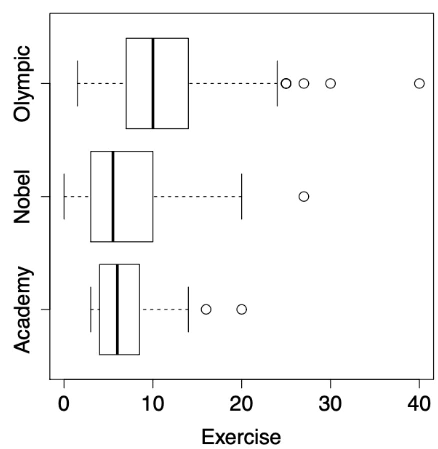
8.27 (a) The standard deviations are very different. In particular, the standard deviation for the LL sample ($s_{LL} = 1.31$) is more than double the standard deviation for the LD sample ($s_{LD} = 0.43$). This indicates that an ANOVA test may not be appropriate in this situation.

(b) A p-value of 0.652 from the randomization distribution is not small, so we would not reject a null hypothesis that the means are equal. There is not sufficient evidence to conclude that the mean amount consumed is different depending on the amount of light at night. Mice in different light at night conditions appear to eat the roughly similar amounts on average. Weight gain in mice with light at night is not a result of eating more food.

8.31 We start by finding the mean and standard deviations for *Exercise* within each of the award groups and for the whole sample.

	Academy	Nobel	Olympic	Total
Mean	6.81	6.96	11.14	9.05
Std. Dev.	4.09	4.74	5.98	5.74
Sample size	31	148	182	361

The standard deviations are not too different, so the equal variance condition is reasonable. Side-by-side boxplots of exercise between the award groups show some right skewness, but the sample sizes are large so this shouldn't be a problem.



To test $H_0 : \mu_1 = \mu_2 = \mu_3$ vs H_a : Some $\mu_i \neq \mu_j$ we use technology to produce an ANOVA table.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Award	2	1597.9	798.96	27.861	0.0000
Residuals	358	10266.3	28.68		
Total	360	11864.2			

The p-value from the ANOVA (0.0000) is very small, so we have strong evidence that at least one of the award groups has a mean exercise rate that differs from at least one of the other groups.

Exercise 8.46: We already have a significance *F*-score (the *P*-value is reported as 0.001). We achieve comparisons as required in parts (a)–(c) using the `TukeyHSD()` command:

```
TukeyHSD( aov(Pulse ~ Award, data=StudentSurvey) )
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Pulse ~ Award, data = StudentSurvey)

$Award
                    diff       lwr       upr     p adj
Nobel-Academy    1.698636 -3.878263  7.275535 0.7536936
Olympic-Academy -3.263382 -8.752515  2.225752 0.3423983
Olympic-Nobel   -4.962018 -8.083176 -1.840860 0.0006222
```

We do not see convincing evidence of different mean pulse rates in the pairing of part (a) (Nobel–Academy) since the proposed difference $\mu_{\text{Nobel}} - \mu_{\text{Academy}} = 0$ is inside the confidence interval $(-3.878, 7.276)$. Nor do we see a significant difference for the pairing of (b) (Olympic–Academy). However, for (c) (Olympic–Nobel), 0 is *not* inside the confidence interval (and the corresponding P-value is 0.00062. So, there is convincing evidence that the mean pulse rates are different in those two populations.

Exercise 8.52: We (re-)run the ANOVA:

```
anova(lm(GillRate ~ Calcium, data=FishGills3))

Analysis of Variance Table

Response: GillRate
          Df  Sum Sq Mean Sq F value  Pr(>F)
Calcium    2  2037.2 1018.61  4.6484 0.01208 *
Residuals 87 19064.3  219.13
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

and see that, indeed, the P-value suggests we reject the null hypothesis that $\mu_{\text{High}} = \mu_{\text{Low}} = \mu_{\text{Medium}}$ at the 5% level. To probe for a group-wise difference, we again use `TukeyHSD()`:

```
TukeyHSD(lm(GillRate ~ Calcium, data=FishGills3))

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = x)

$Calcium
                diff        lwr        upr     p adj
Low-High    10.333333   1.219540 19.4471264 0.0222533
Medium-High  0.500000  -8.613793  9.6137931 0.9906108
Medium-Low  -9.833333 -18.947126 -0.7195402 0.0313247
```

This time, two pairings reveal differences of means significant at the 5% level: Low–High, and Medium–Low.