Section 14.1: One-Way ANOVA: Comparing Several Means

14.1 Restaurant satisfaction

a) The response variable is the performance gap, the factor is which restaurant the guest visited to, and the categories are the six restaurants.

b) H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$; H_a: at least two of the population means are unequal.

c) $df_1 = 5$ because there are six groups and $df_1 = g - 1$; $df_2 = 144$ because there are 150 people in the study and six groups, and $df_2 = N - g$.

d) From a table or technology: $F = 2.28$ and higher.

14.2 Satisfaction with banking

a) H_0: $\mu_1 = \mu_2 = \mu_3$; H_a: At least two of the population means are unequal. μ_1 denotes the population mean level of satisfaction for Group 1, those who interact with a teller at the bank the most; μ_2 denotes the population mean level of satisfaction for Group 2, those who use ATMs the most, and μ_3 indicates the population mean level of satisfaction for Group 3, those who use the bank's Internet banking service the most.

b) $df_1 = g - 1 = 3 - 1 = 2$; $df_2 = N - g = 400 - 3 = 397$; From technology: $F = 3.02$ and higher.

c) If the null hypothesis were true, the probability would be 0.63 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct. We cannot conclude that at least two of the population means are unequal.

d) The assumptions are (1) that the population distributions of the response variable for the g groups are normal, (2) those distributions have the same standard deviation σ, and (3) the data resulted from randomization. The third assumption is the most important.

14.3 What's the best way to learn French?

a) (i) Assumptions: Independent random samples, normal population distributions with equal standard deviations

(ii) Hypotheses: H_0: $\mu_1 = \mu_2 = \mu_3$; H_a: At least two of the population means are unequal.

(iii) Test statistic: $F = 2.50$ ($df_1 = 2$, $df_2 = 5$)

(iv) P-value: 0.18

(v) Conclusion: If the null hypothesis were true, the probability would be 0.18 of getting a test statistic at least as extreme as the value observed. There is not much evidence against the null. It is plausible that the null hypothesis is correct and that there is no difference among the population mean quiz scores of the three types of students.

b) The P-value is not small likely because of the small sample sizes. Because the numerator of the between-groups variance estimate involves multiplying by the sample size n for each group, a smaller n leads to a smaller overall between-groups estimate of variance and a smaller test statistic value.

c) This was an observational study because students were not assigned randomly to groups. There could have been a lurking variable, such as school GPA, that was associated with students' group membership, but also with the response variable. Perhaps higher GPA students are more likely to have previously studied a language and higher GPA students also tend to do better on quizzes than other students.

14.4 What affects the *F* value?

a) The F test statistic would be smaller because the between-groups estimate of the variance would be smaller, while the within-groups estimate would stay the same.

b) The F statistic would be larger because the within-subjects estimate of variance would be smaller.

c) The F statistic would be larger because the numerator of the between-groups variance estimate would be larger; this occurs because part of the calculation of the numerator of the between-groups estimate of variance involves multiplying by n.

d) The P-value in (a) would be larger because the F statistic is smaller, whereas the P-values in (b) and (c) would be smaller because the F statistics are larger. A larger F statistic corresponds with a smaller P-value.

14.5 Outsourcing

a) H_0: $\mu_1 = \mu_2 = \mu_3$; μ_1 represents the population mean satisfaction rating for San Jose, μ_2 for Toronto, and μ_3 for Bangalore.

b) $F = 13.00/0.47 = 27.6$, calculated by dividing the between-groups variance estimate, 13.00, by the within-groups variance estimate, 0.47. The degrees of freedom are: $df_1 = 2$ and $df_2 = 297$.

c) If the null hypothesis were true, the probability would be close to 0 of a test statistic at least as extreme as the value observed. We have very strong evidence that customer satisfaction ratings are different for at least two of the populations. With a 0.05 significance level, we would reject the null hypothesis.

14.6 ANOVA and box plots

a) Study 2 will more likely lead to a rejection of the null hypothesis. Judging from the box plots, the sample medians for the three groups seem to be fairly equal in Study 1 and rather different in Study 2. Since distribution looks symmetric, this implies that sample means in Study 2 are rather different and more likely to lead to rejection of null hypothesis.

b) The variability within each group seems to be the same for both Study 1 and Study 2, whereas the variability between the group medians is much larger for Study 2. Since distribution is fairly symmetric, this implies larger variability between group means than variability within a group for Study 2, leading to a larger *F* test statistic.

c) No, the small P-value only implies that at least two population means are different, but not necessarily all three.

14.7 Years of education

a) μ_1 represents the population mean ideal number of years of education for the residents of inner city;
μ_2 represents the population mean ideal number of years of education for the residents of suburbia;
μ_3 represents the population mean ideal number of years of education for the residents of countryside;
H_0: $\mu_1 = \mu_2 = \mu_3$.

b) The assumptions are that there are independent random samples, and normal population distributions with equal standard deviations.

c) The *F* statistic is 5.96, and the P-value is 0.003. If the null hypothesis were true, the probability would be 0.003 of getting a test statistic at least as extreme as the value observed. We have strong evidence that a difference exists between at least two of the population means for ideal numbers of years of educations.

d) We cannot conclude that every pair of places of stay has different population means. ANOVA tests only whether at least two population means are different.

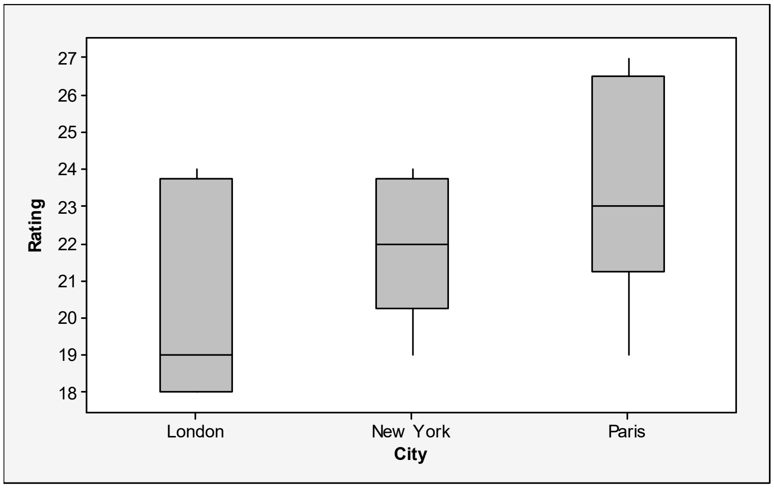
14.8 Smoking and personality

a) An *F* statistic of 3.00 is needed to get P-value = 0.05 in an ANOVA with 2 and 1635 degrees of freedom. The *F* statistic for the extraversion scale, 0.24, is not larger than the *F* statistic required to reject the null hypothesis. Therefore, we must fail to reject the null hypothesis.

b) This does not mean that the population means are necessarily equal. It is possible that this is a Type II error. Confidence intervals would show plausible differences other than 0.

14.9 French cuisine

a)



From Minitab:

Variable	City	N	Mean	StDev
Rating	London	8	20.375	2.774
	New York	8	21.875	1.885
	Paris	8	23.375	2.825

b) Hypotheses: H_0: $\mu_1 = \mu_2 = \mu_3$; H_a: At least two of the population means are unequal. From technology, $F = 2.8$ and the P-value is 0.083. If the null hypothesis were true, the probability would be 0.083 of getting a test statistic at least as extreme as the value observed. At a significance level of 0.05, we do not have sufficient evidence to reject H_0. It is plausible that the population mean ratings of French restaurants in New York, London, and Paris are equal.

14.10 Software and French ANOVA

a) From Minitab:

Variable	group	N	Mean	StDev
quiz score	1	3	6.00	2.00
	2	2	3.00	2.83
	3	3	8.00	2.65

b) From Minitab:

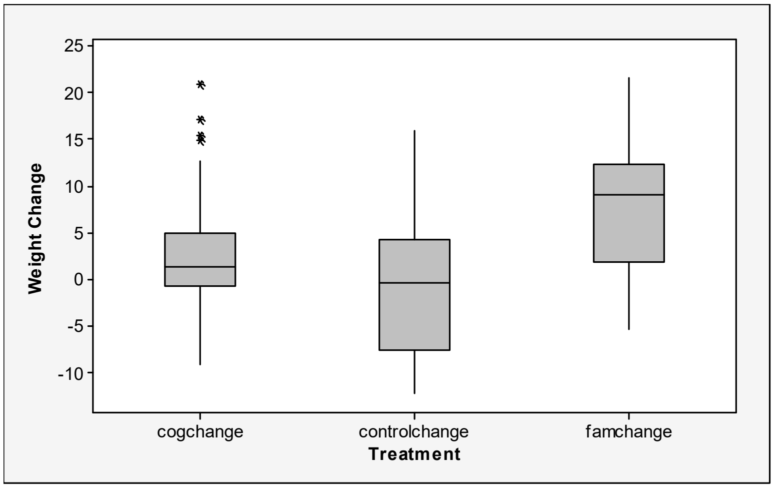
Source	DF	SS	MS	F	P
group	2	30.00	15.00	2.50	0.177
Error	5	30.00	6.00		
Total	7	60.00			

If the null hypothesis were true, the probability would be 0.18 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that there is no difference among means.

c) There are several ways in which this could be accomplished. For example, we could change the 5 to a 2. This change makes the P-value smaller because it increases the difference among means (between-groups variance estimate) while decreasing the difference within means (within-groups variance estimate).

14.11 Comparing therapies for anorexia

a)



From Minitab:

```
Level       N    Mean  StDev
famchange  17   7.265  7.157
cogchange  29   3.007  7.309
conchange  26  -0.450  7.989
```

The box plots and descriptive statistics suggest that the means of these groups are somewhat different. The standard deviations are similar.

b) $F = 5.42$ and the P-value is 0.006. If the null hypothesis were true, the probability would be 0.006 of getting a test statistic at least as extreme as the value observed. We have strong evidence that at least two population means are different.

c) The assumptions are that there are independent random samples and normal population distributions with equal standard deviations. There is evidence of skew, but the test is robust with respect to this assumption. The subjects were randomly assigned to treatments but are not a random sample of subjects suffering from anorexia, so results are highly tentative.

Section 14.2: Estimating Differences in Groups for a Single Factor

14.12 House prices and age

a) The 95% CI is $(\bar{y}_1 - \bar{y}_2) \pm t_{.025}(s)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = (305.8 - 242.8) \pm 1.972(110.38)\sqrt{\frac{1}{78}+\frac{1}{72}}$, or (26.3, 99.7).

b) Because 0 does not fall in this confidence interval, we can infer at the 95% confidence level that the population means are different (higher for new homes than for medium-aged homes).

14.13 Time on Facebook

The 95% CI is $(\bar{y}_1 - \bar{y}_2) \pm t_{.025}(s)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = (63.7 - 49.0) \pm 1.96(70.81)\sqrt{\frac{1}{440}+\frac{1}{407}}$, or (5.2, 24.2).

Because 0 does not fall in this confidence interval, we can infer at the 95% confidence level that the population means are different (higher for Freshman compared to Seniors).

14.14 Comparing telephone holding times

a) We are 95% confident that the difference in the population mean times that callers are willing to remain on hold for classical music versus Muzak is between 2.9 and 12.3 minutes. Since 0 does not fall in the interval, we conclude that the hold time is greater for classical music than for Muzak.

b) The margin of error would be the same for each pair of means because $t_{.025}$, s, and both values of n would be the same for each calculation.

14.14 (continued)

c) 0 does not fall in the confidence interval for the difference between classical music and advertising; therefore, we can infer that the population means are different. 0 does, however, fall in the confidence interval for the difference between advertising and Muzak; therefore, we cannot infer that the population means are different. Together with (a), the airline learned that their best bet for keeping customers on hold is to play classical music.

d) The sample sizes could be increased.

14.15 Tukey holding time comparisons

a) The only significant difference now is that between classical music and Muzak.

b) The margins of error are larger than with the separate 95% intervals because the Tukey method uses an overall confidence level of 95% for the entire set of intervals.

14.16 Hamburger sales

a) If the null hypothesis were true, the probability would be 0.389 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is true. We cannot conclude that at least two population means are different.

b) Margin of error $= t_{0.025,df_2}(s)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$, where $s = \sqrt{MSE} = \sqrt{40.08} = 6.331$ and n_1 and n_2 are the number of observations in Group 1 and Group 2, respectively, here $n_1 = n_2 = 5$, as there are five restaurants in each group, hence Margin of error $= t_{0.025,df_2}(6.331)\sqrt{\frac{1}{5}+\frac{1}{5}} =$ $2.179 \times 6.331 \times 0.707 = 8.72$. It will be same for all the pairs of means as different groups have same number of restaurants, hence n_1 and n_2 will remain same as 5 for all pairs.

c) The margin of error would be smaller because the 95% confidence level is applied to *each* interval rather than to the overall set.

14.17 Hamburger sales regression

a) $x_1 = 1$ for observations from the first group and 0 otherwise; $x_2 = 1$ for observations from the second group and 0 otherwise.

b) H_0: $\mu_1 = \mu_2 = \mu_3$; H_0: $\beta_1 = \beta_2 = 0$.

c) For group 1, $x_1 = 1$, $x_2 = 0$, so predicted mean response = 1200 + 600(1) + 3(0) = 1800. For group 2: 1200 + 600(0) + 300(1) = 1500. For group 3: 1200 + 600(0) + 300(1) = 1500.

14.18 Outsourcing satisfaction

a) The margin of error is the same because the sample sizes for the three service centers are the same. The margin of error is $t_{.025}(s)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 1.968(0.686)\sqrt{\frac{1}{100}+\frac{1}{100}} = 0.19$.

b) The 95% confidence intervals for the difference in population means are:
San Jose/Toronto: –0.2 ± 0.19, or (–0.39, –0.01)
San Jose/Bangalore: 0.5 ± 0.19, or (0.31, 0.69)
Toronto/Bangalore: 0.7 ± 0.19, or (0.51, 0.89)
Because 0 does not fall in any of these intervals, we can infer that all three pairs of population means are different. Toronto is higher than both San Jose and Bangalore. San Jose is higher than Bangalore.

c) The Tukey 95% multiple comparison confidence intervals are:
San Jose/Toronto: –0.2 ± 0.23, or (–0.43, 0.03)
San Jose/Bangalore: 0.5 ± 0.23, or (0.27, 0.73)
Toronto/Bangalore: 0.7 ± 0.23, or (0.47, 0.93)
Because 0 does not fall in the intervals for the difference between San Jose and Bangalore and for the difference between Toronto and Bangalore, we can infer that these two pairs of population means are different. Toronto and San Jose are each higher than Bangalore. 0 does fall, however, in the interval for the difference between San Jose and Toronto, and so we cannot infer that these population means are different.

14.18 (continued)

d) The intervals in (b) and (c) are different because (b) uses a 95% confidence level for *each* interval, whereas (c) uses a 95% confidence level for the overall set of intervals. The advantage of using the Tukey intervals is that it ensures that we achieve the overall confidence level of 95% for the entire set of intervals.

14.19 Regression for outsourcing

a) $x_1 = 1$ for observations from San Jose and $x_1 = 0$ otherwise; $x_2 = 1$ for observations from Toronto and $x_2 = 0$ otherwise.

b) (i) This is the estimate for β_1, which equals 0.5. The estimated difference between the population means for San Jose and Bangalore is 0.5.
(ii) This is the estimate for β_2, which equals 0.7.

14.20 Advertising effect on sales

a) $\mu_y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$, where $x_1 = 1$ for observations for radio and $x_1 = 0$ otherwise, $x_2 = 1$ for observations for TV and $x_2 = 0$ otherwise, and $x_3 = 1$ for observations for newspaper and $x_3 = 0$ otherwise.

b) Because $\mu_1 - \mu_4 = \beta_1$, $\mu_2 - \mu_4 = \beta_2$, and $\mu_3 - \mu_4 = \beta_3$, the ANOVA null hypothesis H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ is equivalent to H_0: $\beta_1 = \beta_2 = \beta_3 = 0$.

c) (i) A and D: 5
(ii) A and B: 15 (Because A must be 40, and B must be 25.)

14.21 French ANOVA

a) Group 2 – Group 1: (–8.7, 2.7)
Group 3 – Group 1: (–3.1, 7.1)
Group 3 – Group 2: (–0.7, 10.7)
Because 0 falls in all three confidence intervals, we cannot infer that any of the pairs of population means are different.

b) Note: Statistical software such as MINITAB is needed to complete this solution.
Group 2 – Group 1: (–10.3, 4.3)
Group 3 – Group 1: (–4.5, 8.5)
Group 3 – Group 2: (–2.3, 12.3)
Again, 0 falls in all three confidence intervals; we cannot infer that any of the pairs of population means are different. The intervals are wider than those in (a) because we are now using a 95% confidence level for the overall set of intervals rather than for each separate interval.

14.22 Multiple comparison for time on Facebook

a) There will be 6 comparisons.

b) The Tukey 95% multiple comparison confidence intervals are:
Freshman versus Sophomore: (–20, 6)
Freshman versus Junior: (–6, 19)
Freshman versus Senior: (2, 27)
Sophomore versus Junior: (–27, –0.5)
Sophomore versus Senior: (–6, 21)
Junior versus Senior: (–34, –9)

c) It is partially true. Seniors spent significantly less time on Facebook than Freshmen (by at least 2 minutes and at most 27 minutes) and Juniors (by at least 9 minutes and at most 43 minutes). However, there is no significant difference between seniors and sophomores.

Section 14.3: Two-Way ANOVA

14.23 Effect of fertilizers

a) The response variable is change in harvest. The factors are dosage level and type of fertilizers.

b) The four treatments are low-dose nitrogen, high-dose nitrogen, low-dose potassium, and high-dose potassium.

c) When we control dose level, we can compare change in harvest for the two types of fertilizers.

14.24 Fertilizer main effects

Hypothetical sets of population means will be different for each plot. The means provided are examples of possible answers.

a)

	Nitrogen	Potassium
Low	10	10
High	20	20

b)

	Nitrogen	Potassium
Low	10	20
High	10	20

c)

	Nitrogen	Potassium
Low	10	20
High	20	30

d)

	Nitrogen	Potassium
Low	10	10
High	10	10

14.25 Political ideology in 2014

a) H_0: The mean political ideology in the adult U.S. population is identical for blacks and whites, for each of the two sexes.

b) $F = 36.5/2.1 = 17.4$ and the P-value is approximately 0. If the null hypothesis were true, it is extremely unlikely to observe such a value for the F test statistic. We have strong evidence that the mean political ideology in the United States depends on race, for each sex.

c) With a P-value of 0.229, there is no evidence that mean political ideology in the United States differs by gender, for blacks and for whites.

14.26 House prices, bedrooms and age

a) $F = 81{,}021/10{,}095 = 8.03$

b) The small P-value of 0.000 provides strong evidence that the population mean house selling price depends on the age of the home. If the null hypothesis were true, the probability would be approximately 0 of getting a test statistic at least as extreme as the value observed.

14.27 Corn and manure

a) When we input the indicators (0 versus 1) into the regression equation, we get the following equations for the four treatments. The only difference between the equations on the top for the low manure groups and the ones on the bottom for the high manure groups is the addition of 1.96.

	Fertilizer	
Manure	**Low**	**High**
Low	11.65	11.65 + 1.88 = 13.53
High	11.65 + 1.96 = 13.61	11.65 + 1.88 + 1.96 = 15.49

b) For 17 degrees of freedom, $t_{.025} = 02.11$ is the value, the standard error of the estimate of the manure effect is 0.747, and the estimate of the manure effect is 1.96.

14.28 Hang up if message repeated?

a) H_0: The population mean holding time is equal for the three types of messages, for each fixed level of repeat time.

b) $F = 74.60/10.52 = 7.09$; the small P-value of 0.011 provides strong evidence that the population mean holding time depends on the type of message. If the null hypothesis were true, the probability would be 0.011 of getting a test statistic at least as extreme as the value observed.

c) The assumptions for two-way ANOVA are that the population distribution for each group is normal, the population standard deviations are identical, and the data result from a random sample or randomized experiment.

14.29 Regression for telephone holding times

a) The population regression model is $\mu_y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$.

Type	Length	x_1	x_2	x_3	Mean of *y*
Advert	10 min	1	0	1	$\alpha + \beta_1 + \beta_3$
Muzak	10 min	0	1	1	$\alpha + \beta_2 + \beta_3$
Classical	10 min	0	0	1	$\alpha + \beta_3$
Advert	5 min	1	0	0	$\alpha + \beta_1$
Muzak	5 min	0	1	0	$\alpha + \beta_2$
Classical	5 min	0	0	0	α

b) $\hat{y} = 8.867 - 5.000x_1 - 7.600x_2 + 2.556x_3$; 8.867 (rounds to 8.87) is the estimated holding time when a customer listens to classical music repeating every 5 minutes (all *x* values = 0); –5.0 represents the decrease in estimated mean when an advertisement is played, and –7.6 represents the decrease in estimated mean when Muzak is played – both at each level of repeat time; 2.556 (rounds to 2.56) represents the increase in estimated mean when the repeat time is 10 minutes, at all levels of type of message.

c) Note: Numbers in parentheses represent values on x_1 and x_2 for type of message, and on x_3 for repeating time. All numbers are rounded to two decimal places.

	10 minutes (1)	**5 minutes (0)**
Advertisement (1,0)	6.42	3.87
Muzak (0,1)	3.82	1.27
Classical music (0,0)	11.42	8.87

d) For a fixed message type, the estimated difference between mean holding times for 10-minute and 5-minute repeats is 2.56. This estimate is the coefficient for x_3, the indicator variable for repeat time.

e) The 95% confidence interval for β_3 is: $b_3 \pm t_{.025}se = 2.556 \pm 2.201(1.709)$, or (–1.2, 6.3).

14.30 Interaction between message and repeat time?

a) H_0: no interaction; $F = 0.67$; P-value = 0.535 (rounds to 0.54)

b) If the null hypothesis were true, the probability would be 0.54 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that there is no interaction. This lends validity to the previous analyses that assumed a lack of interaction.

14.31 Income by gender and job type

a) The response variable is hourly wage, and the two factors are degree and gender.

b)

	High School	**College**	**Advanced**
Male	17	33	43
Female	14	24	32

c) (i) For high school graduates, males make on average \$3 more per hour than females.
(ii) For college graduates, males make on average \$9 more per hour than females.
There is an interaction because the difference between males and females is not the same for the two types of degrees. In particular, the wage gap is much larger for graduates with a college degree compared to a high school degree.

d) There are several possible hypothetical mean hourly wages. This is but one example.

	High School	**College**	**Advanced**
Male	17	33	43
Female	14	30	40

14.32 Ideology by gender and race

a) The means for women are almost the same (black: 4.164; white: 4.2675), whereas the mean for white males is about 0.6 higher than the mean for black males (black: 3.819; white: 4.4443).

b) When race is not considered, the overall means for females and males could be very similar so that an ANOVA with gender as the only factor would not be significant. However, when both factors are considered, there is a clear effect due to gender.

c) From a two-way ANOVA, we learn that the effect of gender differs based on race (as described above), but we do not learn this from the one-way ANOVA.

14.33 Attractiveness and getting dates

a) The response variable is the number of dates in the last three months. The factors are gender and attractiveness.

b) The data suggest that the effect of attractiveness differs according to the level of gender. There appears to be a large effect of attractiveness among women, and little to no effect among men.

c) The population standard deviations likely differ, based on the sample standard deviations. ANOVA is typically robust with respect to violations of this assumption.

14.34 Diet and weight gain

a) $F = 13.90$ and the P-value is approximately 0. The small P-value provides very strong evidence that the mean weight gain depends on the protein level.

b) H_0: no interaction; $F = 2.75$ and the P-value is 0.07. The P-value is not smaller than 0.05, so we cannot reject the null hypothesis. It is plausible that there is no interaction.

c) $se = s\sqrt{\dfrac{1}{n \text{ for low protein beef}} + \dfrac{1}{n \text{ for high protein beef}}} = 14.648\sqrt{\dfrac{1}{10} + \dfrac{1}{10}} = 6.551$

The 95% CI is $\left(\bar{y}_{\text{high beef}} - \bar{y}_{\text{low beef}}\right) = t_{.025}se = (100.0 - 79.20) \pm 2.005(6.551)$, or (7.7, 33.9).

14.35 Regression of weight gain on diet

a) Let $x_1 = 1$ for beef and 0 otherwise, $x_2 = 1$ for cereal and 0 otherwise, and $x_1 = x_2 = 0$ for pork. Likewise, let $x_3 = 1$ for high-protein and $x_3 = 0$ for low protein. $\mu_y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

b) From technology: weight_gain = $81.8 + 0.50x_1 - 4.20x_2 + 14.5x_3$
The parameter estimate for x_3, 14.5, is the difference in the estimate of the weight gain between low and high protein diets.

c) The null hypothesis of no effect of protein source is H_0: $\beta_1 = \beta_2 = 0$.

d)

	High	**Low**
Beef	81.8 + 0.50(1) – 4.20(0) + 14.5(1) = 96.8	81.8 + 0.50(1) – 4.20(0) + 14.5(0) = 82.3
Cereal	81.8 + 0.50(0) – 4.20(1) + 14.5(1) = 92.1	81.8 + 0.50(0) – 4.20(1) + 14.5(0) = 77.6
Pork	81.8 + 0.50(0) – 4.20(0) + 14.5(1) = 96.3	81.8 + 0.50(0) – 4.20(0) + 14.5(0) = 81.8

It means that we assume that the difference between means for the two (or three) categories for one factor is the same in each category of the other factor. In this example, we assume that the difference between the high and low protein levels is the same for each source of protein.

Chapter Problems: Practicing the Basics

14.36 Good friends and marital status

a) We can denote the number of good friends means for the population that these five samples represent by μ_1 for married, μ_2 for widowed, μ_3 for divorced, μ_4 for separated, and μ_5 for never married. The null hypothesis is H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$. The alternative hypothesis is H_a: At least two of the population means are different.

14.36 (continued)

b) No; large values of F contradict the null, and when the null is true the expected value of the F statistic is approximately 1.

c) If the null hypothesis were true, the probability would be 0.53 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that no difference exists among the five marital status groups in the population mean number of good friends.

14.37 Going to bars and having friends

a) (i) H_0: $\mu_1 = \mu_2 = \mu_3$; H_a: At least two of the population means are unequal.
   (ii) $F = 3.03$
   (iii) P-value = 0.049; if the null hypothesis were true, the probability would be 0.049 of getting a test statistic at least as extreme as the value observed. We have evidence that a difference exists among the three frequencies of bar-going in the population mean number of friends.

b) Yes; the sample standard deviations suggest that the population standard deviations might not be equal and the distributions are probably not normal.

14.38 Singles watch more TV

a) The assumptions are an approximately normal distribution of number of hours of TV watching (not so important here because of large sample size), equal standard deviation of number of hours of TV watching in each status group and data gathered through randomization. $F = 54.8/6.0 = 9.1$, which is far out in the tail of the F distribution with $df_1 = 2$ and $df_2 = 1472$. The P-value is 0.0001. There is strong evidence of a difference between the mean hours of TV watching.

b) The means for single and married subjects and single and divorced subjects are significantly different. Compared to married subjects, single subjects watch on average at least 0.3 more hours of TV. The lower bound for the interval comparing single to divorced subjects is close to zero, so the difference between these two groups may be very small.

c) The 95% CI is $(\bar{y}_s - \bar{y}_m) \pm t_{.025}(s)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = (3.27-2.65) \pm 1.961(2.45)\sqrt{\frac{1}{459}+\frac{1}{731}}$, or (0.33, 0.91).

d) The corresponding interval formed with the Tukey method would be wider because it uses a 95% confidence level for the overall set of intervals, rather than for each one separately.

14.39 Comparing auto bumpers

a) The assumptions are that there are independent random samples, and normal population distributions with equal standard deviations.

b) H_0: $\mu_1 = \mu_2 = \mu_3$; H_a: At least two of the population means are unequal.

c) $F = 18.50$; $df_1 = 2$, $df_2 = 3$

d) The P-value is 0.02.

e) If the null hypothesis were true, the probability would be 0.02 of getting a test statistic at least as extreme as the value observed. We have strong evidence that a difference exists among the three types of bumpers in the population mean cost to repair damage.

14.40 Compare bumpers

a) The margin of error is $t_{.025}(s)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 3.182(2.00)\sqrt{\frac{1}{2}+\frac{1}{2}} = 6.4$.

b) The confidence interval formed using the Tukey 95% multiple comparison uses a 95% confidence level for the overall set of intervals, rather than a 95% confidence level for each separate interval.

c) Let $x_1 = 1$ for Bumper A and 0 otherwise, $x_2 = 1$ for Bumper B and 0 otherwise, and $x_1 = x_2 = 0$ for Bumper C.

d) 13 is the estimated mean damage cost for Bumper C, –11 is the difference between the estimated mean damage costs between Bumpers A and C, and –10 is the difference between the estimated mean damage costs between Bumpers B and C.

14.41 Segregation by region

a)

	Mean	Standard Deviation
NE	70.75	5.56
NC	73.75	11.90
S	62.00	3.65
W	61.25	7.18

b) We can denote the segregation index means for the population that these four samples represent by μ_1 for Northeast, μ_2 for North Central, μ_3 for South, and μ_4 for West. The null hypothesis is H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$. The alternative hypothesis is that at least two of the population means are different.

c) From technology: $F = 2.64$ and the P-value is 0.097. There is some evidence, but not very strong, that a difference exists among the four regions in the population mean segregation index.

d) The ANOVA would not be valid because this would not be a randomly selected sample.

14.42 Compare segregation means

a) From technology, the margin of error is 16.179.

b) All the intervals contain 0; therefore, no pair is significantly different.

14.43 Georgia political ideology

a) The ANOVA assumption of equal population standard deviations does seem plausible given the similar sample standard deviations.

b) Dot plots generated by software indicate the possibility of some skew, particularly among the sample of Republicans which appears to be skewed to the left. One outlier, however, contributes to the appearance of being skewed to the left, and so this inference must be treated with caution. Regardless, the normality assumption is not as important as the assumption that the groups are random samples from the population of interest.

c) The confidence interval does not include 0; we can conclude at the 95% confidence level that on the average Republicans are higher in conservative political ideology than are Democrats.

d) We can conclude that a difference exists among the three political parties in their political ideologies; however, we can only conclude that there is a difference between Republicans and Democrats and between Independents and Republicans. We cannot draw conclusions about differences between Independents and Democrats.

14.44 Comparing therapies for anorexia

a) From Minitab:

```
Variable        N   Mean  StDev
cogchange      29   3.01   7.31
controlchange  26  -0.45   7.99
famchange      17   7.26   7.16

Source      DF      SS     MS     F      P
Treatment    2   614.6  307.3  5.42  0.006
Error       69  3910.7   56.7
Total       71  4525.4
```

14.44 (continued)

For a 95% confidence interval with $df = 69$, technology gives $t_{.025} = 1.99$ and $s = \sqrt{56.7} = 7.53$.
Using Fisher's method (confirmed by technology):

Control – Cog: $(\bar{y}_1 - \bar{y}_2) \pm t_{.025}(s)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = (-0.45-3.01) \pm 1.99(7.53)\sqrt{\frac{1}{26}+\frac{1}{29}} = (-7.5, 0.6)$

Cog – Family: $(3.01-7.26) \pm 1.99(7.53)\sqrt{\frac{1}{17}+\frac{1}{29}} = (-8.8, 0.3)$

Control – Family: $(-0.45-7.26) \pm 1.99(7.53)\sqrt{\frac{1}{26}+\frac{1}{17}} = (-12.4,\ -3.0)$

The differences between the control and cognitive treatments and the family and cognitive treatments include 0; it is plausible that there are no population mean differences in these two cases. The other interval (i.e., family and control) does not include 0. We can infer that the population mean weight change is greater among those who receive family therapy than among those in the control group.

b) Technology gives the following Tukey 95% multiple comparison confidence intervals:
Difference between control and cognitive: (–8.3, 1.4)
Difference between family and cognitive: (–9.8, 1.3)
Difference between family and control: (–13.3, –2.1)
The interpretations are the same as for the intervals in (a). The intervals are wider because the 95% confidence level is for the entire set of intervals rather than for each interval separately.

14.45 Years of education varies by region?

Years of education by quadrant of city

Source	DF	SS	MS	F	P
Quadrant	3	29.5	29.5/3 = 9.83	9.83/2.32 = 4.25	0.01
Error	199 – 3 = 196	483.5 – 29.5 = 454	454/196 = 2.32		
Total	199	483.5			

14.46 House with garage

a) Let $x_1 = 1$ for houses with a garage and 0 otherwise.
From technology: HP in thousands = 247 + 26.3x_1. The intercept, 247, is the estimated mean selling price in thousands when a house does not have a garage, and 26.3 is the difference in the estimated mean selling price in thousands between a house with a garage and a house without a garage.

b) From Minitab:

```
Predictor    Coef  SE Coef     T      P
Garage       26.28   19.27  1.36  0.174
```

If the null hypothesis were true, the probability would be 0.174 of getting a test statistic at least as extreme as the value observed. We do not have sufficient evidence to conclude that the population mean house selling price in thousands is significantly higher for houses with a garage than without a garage.

c) From Minitab:

```
Source   DF       SS     MS     F      P
Garage    1    24830  24830  1.86  0.174
Error   198  2644040  13354
Total   199  2668870
```

As with the regression model, the P-value is 0.174; this is the same result.

d) The value of t in (b) is the square root of the value of F in (c).

14.47 Ideal number of kids by gender and race

a) The response variable is ideal number of kids. The factors are gender and race.

b) If there were no interaction between gender and race in their effects, it would mean that the difference between population means for the two genders is the same for each race. There are many possible sets of population means that would show a strong race effect and a weak gender effect and no interaction. This is one hypothetical set.

	Female	**Male**
Black	3.5	3.3
White	1.5	1.3

c) H_0: no interaction; H_a: There is an interaction. $F = 1.36$ and the P-value is 0.24. If the null hypothesis were true, the probability would be 0.24 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is true and that there is no interaction.

14.48 Regress kids on gender and race

a) The coefficient of f (0.04) is the difference between the estimated means for men and women for each level of race. The fact that this estimate is close to 0 indicates that there is very little difference between the means for men and women, given race.

b)

	Female (1)	**Male (0)**
Black (1)	2.83	2.79
White (0)	2.46	2.42

c) The data suggest that there is an effect only of race. If the null hypothesis were true, the probability would be close to 0 of getting a test statistic at least as extreme as the value observed. We have strong evidence that on the average blacks report a higher ideal number of children than whites do (by 0.37).

14.49 Energy drink

a) At both 11 P.M. and 8 A.M., the mean time Williams College students need to complete the task is 0.39 minutes shorter when consuming an energy drink.

b) The mean time Williams College students need to complete the task is 2 minutes shorter at night (11 P.M.) compared to in the morning (9 A.M.), whether or not consuming an energy drink.

c) The model is $\mu_Y = \alpha + \beta_1 e + \beta_2 d$. β_1 needs to equal zero.

d) For the energy drink, $t = -0.58$ and the P-value is 0.57. There is no evidence of a difference in means.

14.50 Income, gender, and education

a) The response variable is income and the factors are gender and education.

b)

	HS graduate	**Bachelor Degree**
Female	31,666	60,293
Male	43,493	94,206

c) (i) Among high school graduates, men are 43,493 – 31,666 = 11,827 above women.

(ii) Among college graduates, men are 94,206 – 60,293 = 33,913 above women.

If these are close estimates of the population, there is an interaction because the effect of gender is different among high school graduates than among college graduates. There is a bigger mean difference between genders among college graduates than among high school graduates.

14.51 Birth weight, age of mother, and smoking

This suggests an interaction since smoking status has a different impact at different fixed levels of age. There is a bigger mean difference between smokers and non-smokers among older women than among younger women.

14.52 TV watching by gender and race

a) The response variable is hours of TV watched per day and the factors are gender and race.

b) $F = 0.21$ and the P-value is 0.649. If the null hypothesis were true, the probability would be 0.649 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that there is no interaction.

c) (i) There is not a significant gender effect. The P-value of 0.259 is not less than a typical significance level such as 0.05.

(ii) There is a significant race effect. The small P-value of 0.000 is less than a typical significance level such as 0.05.

d) At each level of race, there is little difference between genders. At each level of gender, however, there is a large difference between races.

14.53 Salary and gender

a) The coefficient for gender, –13, indicates that at fixed levels of rank, men have higher estimated mean salaries than women by 13 (thousands of dollars).

b) $96.2 = \hat{\alpha} - 13(1) - 40(0) \Rightarrow 96.2 = \hat{\alpha} - 13 \Rightarrow \hat{\alpha} = 96.2 + 13 = 109.2$, so $\hat{y} = 109.2 - 13x_1 - 40x_2$.

(i) $\hat{y} = 109.2 - 13(0) - 40(0) = 109.2$

(ii) $\hat{y} = 109.2 - 13(1) - 40(1) = 56.2$

14.54 Political ideology interaction

a) For blacks, the difference in the mean political ideology is 3.82 – 3.81 = 0.01 between females and males. For whites, this difference is 4.13 – 1.22 = –0.09. Both differences are small and about the same. This indicates that the effect of gender on the mean may be the same for blacks and whites (i.e., no interaction) and that the effect is essentially zero; i.e., there is no difference between females and males, for either blacks or whites.

b) With a P-value of 0.54, there is no evidence of an interaction between gender and race. The effect of gender is the same for blacks and whites and, vice versa, the effect of race is the same for females and males.

Chapter Problems: Concepts and Investigations

14.55 Regress TV watching on gender and marital status

a) $\mu_y = \alpha + \beta_1 g + \beta_2 m_1 + \beta_3 m_2$

Gender	Status	g	m_1	M_2	Mean of y
Male	Single	1	1	0	$\alpha + \beta_1 + \beta_2$
Male	Married	1	0	1	$\alpha + \beta_1 + \beta_3$
Male	Divorced	1	0	0	$\alpha + \beta_1$
Female	Single	0	1	0	$\alpha + \beta_2$
Female	Married	0	0	1	$\alpha + \beta_3$
Female	Divorced	0	0	0	α

b) TV hours = 2.78 + 0.16g + 0.41m_1 – 0.21m_2. The mean hours of TV watching are estimated to be 0.16 hours higher for males, for each marital status.

c) The difference for both males and females is β_2.

For males, the difference is $(\alpha + \beta_1 + \beta_2) - (\alpha + \beta_1) = \beta_2$.

For females, the difference is $(\alpha + \beta_2) - (\alpha) = \beta_2$.

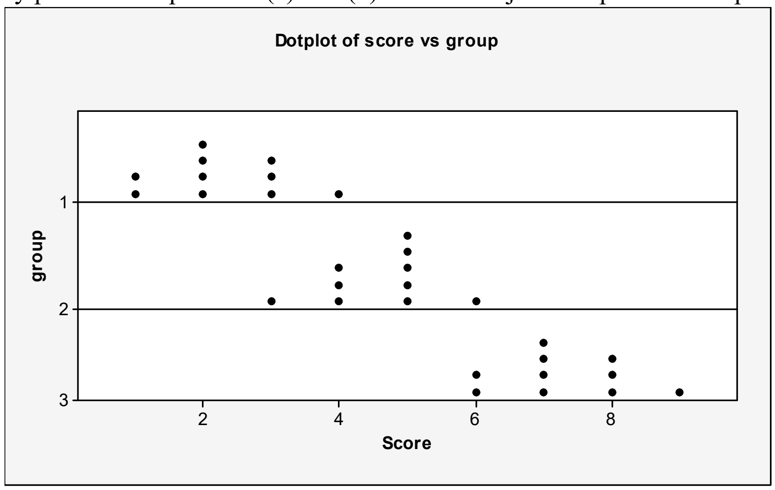
d) The 95% confidence interval for β_2 is: $b_2 \pm t_{.025} se = 0.41 \pm 1.962(0.184)$, or (0.05, 0.77). With 95% confidence, the mean number of hours watching TV for singles is between 0.05 hours and 0.77 hours larger than the mean for divorced subjects.

14.56 Number of friends and degree

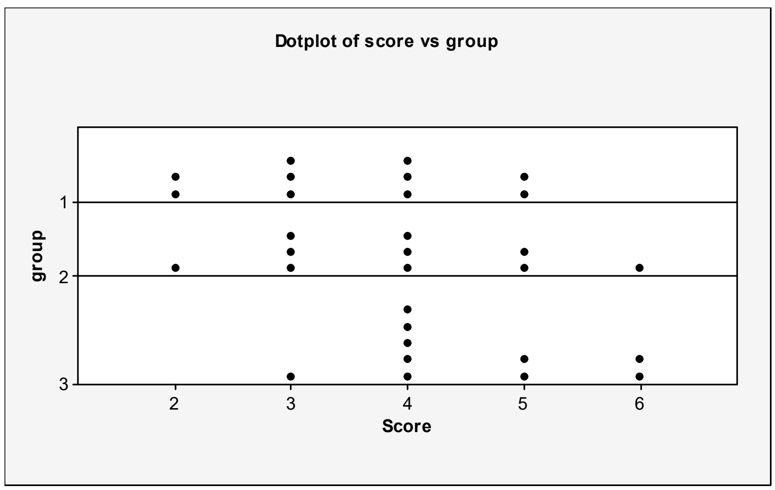
The short report will be different for each student, but should include and interpret the F test statistic of 0.81 and the associated P-value of approximately 0.52.

14.57 Sketch within- and between-groups variability

a) There are many possible dot plots for (a) and (b). These are just two possible dot plots.



b)



14.58 A = B and B = C, but A ≠ C?

There are many possible means that would meet these requirements. An example is means 10, 13, 17 for A, B, C.

14.59 Multiple comparison confidence

A confidence interval of 0.95 for a single comparison gives 95% confidence for only the one interval. A confidence level of 0.95 for a multiple comparison of all six pairs of means provides a confidence level for the whole set of intervals. In this case, the confidence level for each of the individual intervals would be a good deal higher than for the overall set. That way the error probability of 0.05 applies to the whole set, and not to *each* comparison.

14.60 Another Simpson paradox

a)

	Female	Male
Humanities	65,000	64,000
Science	72,000	71,000

The overall mean for women is $\dfrac{25(65,000)+5(72,000)}{30} = 66,167.$

The overall mean for men is $\dfrac{20(64,000)+30(71,000)}{50} = 68,200.$

14.60 (continued)

b) A one-way comparison of mean income by gender would reveal that men have a higher mean income than women do. A two-way comparison of mean incomes by gender, however, would show that at fixed levels of university divisions, women have the higher mean.

14.61 Multiple choice: ANOVA/regression similarities

The best answer is (d).

14.62 Multiple choice: ANOVA variability

The best answer is (c).

14.63 Multiple choice: Multiple comparisons

The best answer is (c).

14.64 Multiple choice: Interaction

The best answer is (c).

14.65 True or false: Interaction

True

♦♦14.66 What causes large or small *F*?

a) The four sample means would have to be identical.

b) There would have to be no variability (standard deviation of 0) within each sample. That is, all five individuals in each sample would have to have the same score.

♦♦14.67 Between-subjects estimate

a) $\frac{\sum(\bar{y}_i - \bar{y})^2}{g-1}$ estimates the variance, σ^2/n, of the distribution of the $\{\bar{y}_i\}$ values, because $\frac{\sum(\bar{y}_i - \bar{y})^2}{g-1}$ is essentially the formula for variance, in which each observation is a sample mean.

b) If $\frac{\sum(\bar{y}_i - \bar{y})^2}{g-1}$ estimates σ^2/n, then *n* times $\frac{\sum(\bar{y}_i - \bar{y})^2}{g-1}$ estimates σ^2.

♦♦14.68 Bonferroni multiple comparisons

a) Fertilizer: $1.88 \pm 2.46(0.7471)$, or (0.04, 3.7)
Manure: $1.96 \pm 2.46(0.7471)$, or (0.1, 3.8)

b) The P-value should be 0.05/35 = 0.0014.

♦♦14.69 Independent confidence intervals

a) (0.95)(0.95)(0.95)(0.95)(0.95) = 0.77

b) $(0.9898)^5 = 0.95$

♦♦14.70 Regression or ANOVA?

a) (i) In an ANOVA *F* test the number of bathrooms would be treated as a categorical variable. Three would not be considered "more" than one; it would simply be treated as a different category.
(ii) In a regression *t*-test, the number of bathrooms would be treated as a quantitative variable and we would be assuming a linear trend.

b) The straight-line regression approach would allow us to know whether increasing numbers of bathrooms led to an increasing mean house selling price. The ANOVA would only let us know that the mean house price was different at each of the three categories. Moreover, we could not use the ANOVA as a prediction tool for other numbers of bathrooms (although even with regression, we must be careful when we interpolate or extrapolate).

c) Mean selling price \$150,000 for 1 bathroom, \$100,000 for 2 bathrooms, and \$200,000 for 3 bathrooms. (There is not an increasing or decreasing overall trend.)

♦♦14.71 Three factors

a) 8 groups. Let N_L = low level of nitrogen, N_H = high level of nitrogen, PH_L = low level of phosphate, PH_H = high level of phosphate, PO_L = low level of potash and PO_H = high level of potash. Then the groups are $N_LPH_LPO_L$, $N_HPH_LPO_L$, $N_LPH_HPO_L$, $N_LPH_LPO_H$, $N_HPH_HPO_L$, $N_HPH_LPO_H$, $N_LPH_HPO_H$, $N_HPH_HPO_H$.

b) Let $x_1 = 1$ for the high level of nitrogen and 0 for the low level of nitrogen, $x_2 = 1$ for the high level of phosphate and 0 for the low level of phosphate, and $x_3 = 1$ for the high level of potash and 0 for the low level of potash. Then $\mu_Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$.

c) One possibility is $\hat{\alpha} = 6$, $\hat{\beta}_1 = 2$, $\hat{\beta}_2 = 2$, and $\hat{\beta}_3 = -2$.

Chapter Problems: Student Activities

14.72 Student survey data

The short reports will be different for each class.