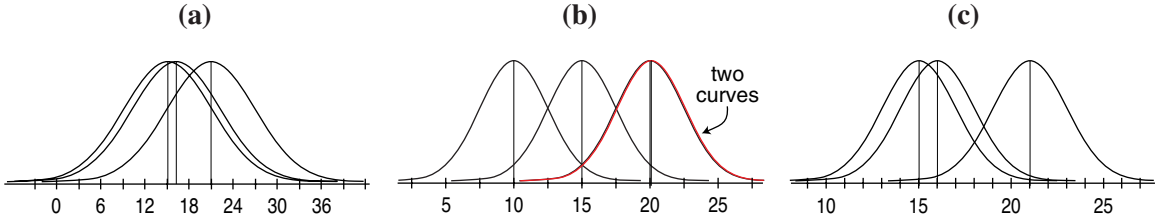
Chapter 12 Solutions

12.1. **(a)** H_0 says the *population* means are all equal. **(b)** Experiments are best for establishing causation. **(c)** ANOVA is used when the *explanatory* variable has two *or more* values.

12.2. **(a)** If we reject H_0, we conclude that *at least one* mean is different from the rest. **(b)** There is no theoretical limit to the number of means that can be compared. **(c)** Two-way ANOVA is used to examine the effect of two explanatory variables on a response variable.

12.3. We were given sample sizes $n_1 = 25$, $n_2 = 22$, and $n_3 = 19$ and standard deviations $s_1 = 22$, $s_2 = 20$, and $s_3 = 18$. **(a)** Yes: The guidelines for pooling standard deviations say that the ratio of largest to smallest should be less than 2; we have $\frac{22}{18} \doteq 1.22 < 2$. **(b)** Squaring the three standard deviations gives $s_1^2 = 484$, $s_2^2 = 400$, and $s_3^2 = 324$. **(c)** $s_p^2 = \dfrac{24s_1^2 + 21s_2^2 + 18s_3^2}{24 + 21 + 18} \doteq 410.2857$. **(d)** $s_p = \sqrt{s_p^2} \doteq 20.2555$.

12.4. These pictures are shown below. Note that for (b), there are four curves, but two coincide almost exactly (one centered at 20, the other at 20.1).



12.5. **(a)** This sentence describes *between*-group variation. Within-group variation is the variation that occurs by chance among members of the same group. **(b)** The *sum of* squares (not the mean squares) in an ANOVA table will add. **(c)** The common population standard deviation σ (not its estimate s_p) is a parameter.

12.6. The answers are found in Table E (or using software) with $p = 0.05$ and degrees of freedom $I - 1$ and $N - I$. **(a)** $I = 5$, $N = 15$, df 4 and 10: $F > 3.48$ (software: 3.4781). **(b)** $I = 5$, $N = 30$, df 4 and 25: $F > 2.76$ (software: 2.7587). **(c)** $I = 5$, $N = 45$, df 4 and 40: $F > 2.61$ (software: 2.6060). **(d)** As the degrees of freedom increase, values from an F distribution tend to be smaller (closer to 1), so smaller values of F are statistically significant. In terms of ANOVA conclusions, we have learned that with smaller samples (fewer observations per group), the F statistic needs to be fairly large in order to reject H_0.

12.7. Assuming the t (ANOVA) test establishes that the means are different, contrasts and multiple comparison provide no further useful information. (With two means, there is only one comparison to make, and it has already been made by the t test.)

12.8. **(a)** The stated hypothesis is $\mu_{50\%} < \frac{1}{2}(\mu_{0\%} + \mu_{100\%})$, so we use the contrast $\psi = \frac{1}{2}(\mu_{0\%} + \mu_{100\%}) - \mu_{50\%}$, with coefficients 0.5, -1, and 0.5. The hypotheses can then be stated H_0: $\psi = 0$ vs. H_a: $\psi > 0$. **(b)** The estimated contrast is $c = \frac{1}{2}(50 + 120) - 75 = 10$ cm^3, with standard error $\text{SE}_c = s_p\sqrt{\frac{0.25}{40} + \frac{1}{40} + \frac{0.25}{40}} \doteq 5.8095$, so

the test statistic is $t \doteq \frac{10}{5.8095} \doteq 1.7213$ with df $= 117$. The one-sided P-value is $P = 0.0439$, so this is significant at $\alpha = 0.05$, but not at $\alpha = 0.05$.

Note: *We wrote the contrast so that it would be positive when H_a is true (in keeping with the text's advice). We could also test this hypothesis using the contrast $\psi' = \mu_{50\%} - \frac{1}{2}(\mu_{0\%} + \mu_{100\%})$, or even $\psi'' = \mu_{0\%} + \mu_{100\%} - 2\mu_{50\%}$. The resulting t statistic is the same (except possibly in sign) regardless of the way the contrast is stated.*

12.9. **(a)** Response: egg cholesterol level. Populations: chickens with different diets or drugs. $I = 3$, $n_1 = n_2 = n_3 = 25$, $N = 75$. **(b)** Response: rating on five-point scale. Populations: the three groups of students. $I = 3$, $n_1 = 31$, $n_2 = 18$, $n_3 = 45$, $N = 94$. **(c)** Response: quiz score. Populations: students in each TA group. $I = 3$, $n_1 = n_2 = n_3 = 14$, $N = 42$.

12.10. **(a)** Response: time to complete VR path. Populations: children using different navigation methods. $I = 4$, $n_i = 10$ $(i = 1, 2, 3, 4)$, $N = 40$. **(b)** Response: calcium content of bone. Populations: chicks eating diets with differing pesticide levels. $I = 5$, $n_i = 13$ $(i = 1, 2, 3, 4, 5)$, $N = 65$. **(c)** Response: total sales between 11:00 a.m. and 2:00 p.m. Populations: customers responding to one of four sample offers. $I = 4$, $n_i = 5$ $(i = 1, 2, 3, 4)$ and $N = 20$.

12.11. For all three situations, the hypotheses are H_0: $\mu_1 = \mu_2 = \mu_3$ vs. H_a: at least one mean is different. The degrees of freedom are DFG = DFM $= I - 1$ ("model" or "between groups"), DFE = DFW $= N - I$ ("error" or "within groups"), and DFT $= N - 1$ ("total"). The degrees of freedom for the F test are DFG and DFE.

Situation	I	N	DFG	DFE	DFT	df for F statistic
Egg cholesterol level	3	75	2	72	74	$F(2, 72)$
Student opinions	3	94	2	91	93	$F(2, 91)$
Teaching assistants	3	42	2	39	41	$F(2, 39)$

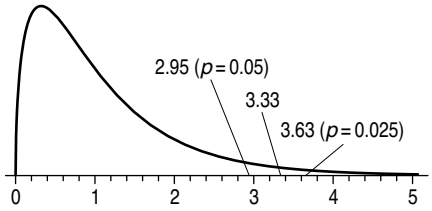
12.12. For all three situations, the hypotheses are H_0: $\mu_1 = \mu_2 = \cdots = \mu_I$ vs. H_a: at least one mean is different. The degrees of freedom are DFG = DFM $= I - 1$ ("model" or "between groups"), DFE = DFW $= N - I$ ("error" or "within groups"), and DFT $= N - 1$ ("total"). The degrees of freedom for the F test are DFG and DFE.

Situation	I	N	DFG	DFE	DFT	df for F statistic
VR navigation methods	4	40	3	36	39	$F(3, 36)$
Effect of pesticide on birds	5	65	4	60	64	$F(4, 60)$
Effect of free food on sales	4	20	3	16	19	$F(3, 16)$

12.13. **(a)** This sounds like a fairly well designed experiment, so the results should at least apply to this farmer's breed of chicken. **(b)** It would be good to know what proportion of the total student body falls in each of these groups—that is, is anyone overrepresented in this sample? **(c)** How well a TA teaches one topic (power calculations) might not reflect that TA's overall effectiveness.

12.14. **(a)** This sounds like a fairly well designed experiment, assuming the subjects come from a group which is representative of the population. (We assume that this teaching tool is intended for use with children and that the children used in the experiment were themselves deaf.) **(b)** This should at least give information about pesticide effect on bone calcium in chicks. It might not apply to adult chickens, or other species of birds. **(c)** The results *might* extend to similar sandwich shops, and perhaps to other times of day, or to weekend sales.

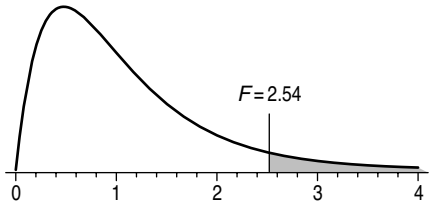
12.15. **(a)** With $I = 4$ groups and $N = 32$, we have df $I - 1 = 3$ and $N - I = 28$. In Table E, we see that $2.95 < F < 3.63$. **(b)** The sketch on the right shows the observed F value and the critical values from Table E. **(c)** $0.025 < P < 0.050$ (software gives 0.0337). **(d)** The alternative hypothesis states that at least one mean is different, not that all means are different.

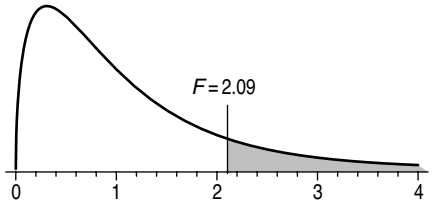


12.16. Compare each F statistic to an $F(I - 1, N - I)$ distribution:

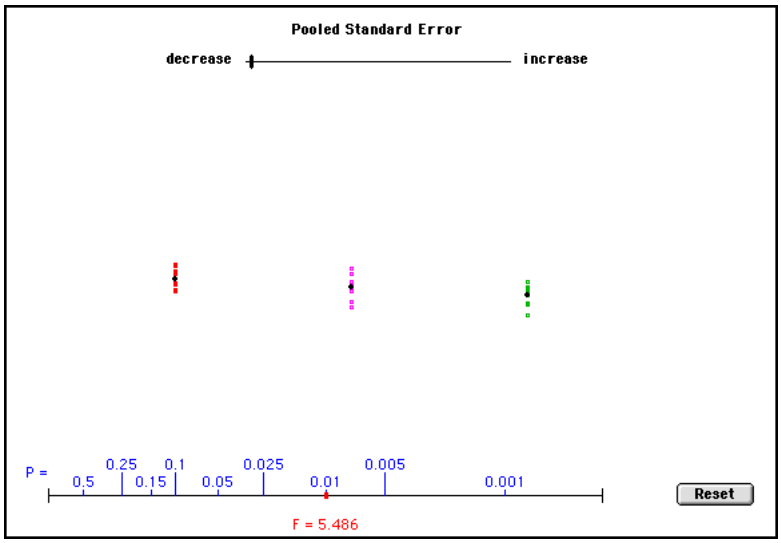
	F	I	N	DFG	DFE	Critical values	P-value (Table E)	P-value (software)
(a)	2.31	7	35	6	28	$2.00 < F < 2.45$	$0.050 < P < 0.100$	0.0615
(b)	2.83	5	55	4	50	$2.56 < F < 3.05$	$0.025 < P < 0.050$	0.0342
(c)	4.08	6	66	5	60	$3.34 < F < 4.76$	$0.001 < P < 0.010$	0.0030

12.17. **(a)** $I = 5$ and $N = 45$, so the degrees of freedom are 4 and 40. $F = \frac{127}{50} = 2.54$. Comparing to the $F(4, 40)$ distribution in Table E, we find $2.09 < F < 2.61$, so $0.050 < P < 0.100$. (Software gives $P \doteq 0.0546$.) **(b)** $I = 4$ and $N = 28$, so the degrees of freedom are 3 and 24. $F = \frac{40/3}{153/24} \doteq 2.0915$. Comparing to the $F(3, 24)$ distribution in Table E, we find $F < 2.33$, so $P > 0.100$. (Software gives $P \doteq 0.1279$.)

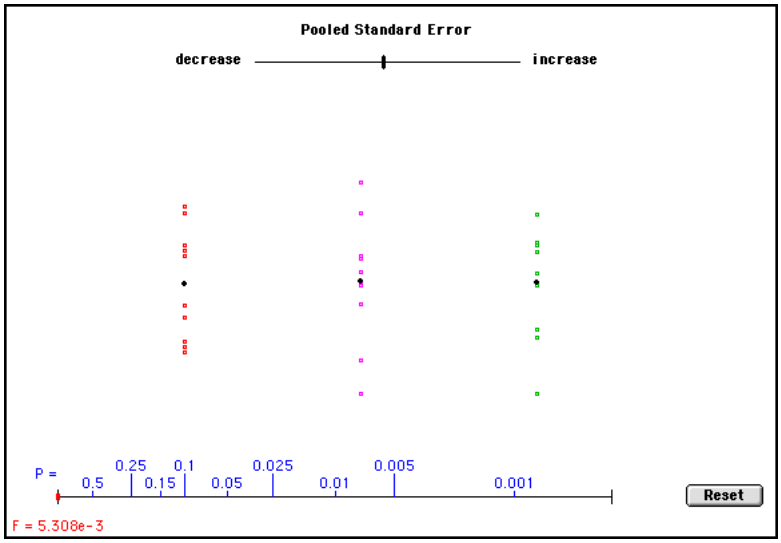




12.18. **(a)** P-values close to 0.01 occur when F is close to 5.483. (This is the value for df 2 and 27; this applet seems to have three samples with 10 observations each.) How close students can get to this depends on how much they play around with the applet, and the pooled standard error setting. **(b)** As variation increases, F decreases and P increases.



12.19. **(a)** F can be made very small (close to 0), and P close to 1. **(b)** F increases, and P decreases. Moving the means farther apart means that (even with moderate spread) it is easier to see that the three groups represent three different populations (that is, populations having different means). Therefore, the evidence against H_0 becomes stronger.



12.20. **(a)** No: The guidelines for pooling standard deviations say that the ratio of largest to smallest should be less than 2; we have $\frac{42}{20} = 2.1$. **(b)** Squaring the four standard deviations gives $s_1^2 = 1764$, $s_2^2 = 1444$, $s_3^2 = 400$, and $s_4^2 = 2025$. **(c)** $s_p^2 = \dfrac{24s_1^2 + 27s_2^2 + 149s_3^2 + 20s_4^2}{24 + 27 + 149 + 20} \doteq 824.65$. **(d)** $s_p = \sqrt{s_p^2} \doteq 28.72$. **(e)** Because the third sample was over twice as large as the other three put together, the pooled standard deviation is closest to s_3.

12.21. **(a)** Use matched pairs t methods; we examine the change in reaction time for each subject. **(b)** No: We cannot use ANOVA methods because we do not have four independent samples. (The same group of subjects performed each of the four tasks.)

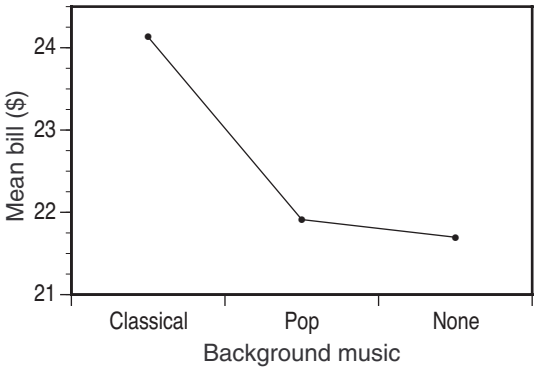
12.22. We have $\bar{x}_1 = 96.38$, $s_1 = 29.78$, $n_1 = 71$, $\bar{x}_2 = 109.44$, $s_2 = 31.05$, and $n_2 = 37$. For the pooled t procedure, we find $s_p \doteq 30.22$ and $t = -2.13$ (df $= 106$, $P \doteq 0.035$). The Minitab output below shows that $F = 4.54$ (t^2, up to rounding error).

Minitab output

```
Analysis of Variance
Source     DF        SS        MS        F        p
Factor      1      4149      4149     4.54    0.035
Error     106     96787       913
Total     107    100936
```

12.23. **(a)** With $I = 4$ and $N = 2290$, the degrees of freedom are DFG $= I - 1 = 3$ and DFE $= N - I = 2286$. **(b)** MSE $= s_p^2 = 4.6656$, so $F = \frac{\text{MSG}}{\text{MSE}} = \frac{11.806}{4.6656} \doteq 2.5304$. **(c)** The $F(3, 1000)$ entry in Table E gives $0.05 < P < 0.10$; software give $P \doteq 0.0555$.

12.24. **(a)** The plot of means suggests that spending is higher for classical music, while pop and no music appear to have the same effect. **(b)** Yes: The guidelines for pooling standard deviations say that the ratio of largest to smallest should be less than 2; we have $\frac{3.332}{2.243} \doteq 1.49 < 2$. **(c)** The degrees of freedom are DFG $= I - 1 = 2$ and DFE $= N - I = 138$. Comparing to an $F(2, 100)$ distribution in Table E, we see that $P < 0.001$; software gives $P \doteq 0.00005$.

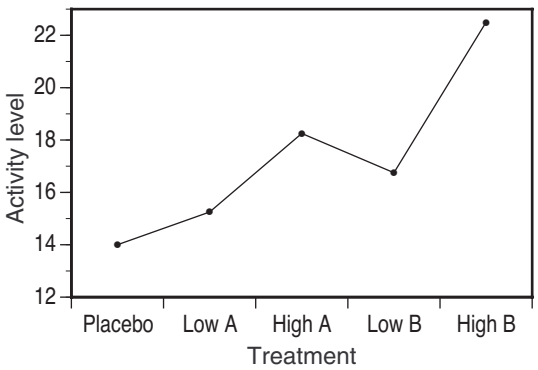


We have strong evidence that the means are not all the same. **(d)** The higher average bill for classical music led to this conclusion; the difference between pop music and no background music is negligible. **(e)** The setting of this experiment ("a single high-end restaurant in England") might limit how much this conclusion can be generalized. It *might* extend to other high-end restaurants, but perhaps not to "family-style" restaurants, and almost certainly not to fast-food restaurants.

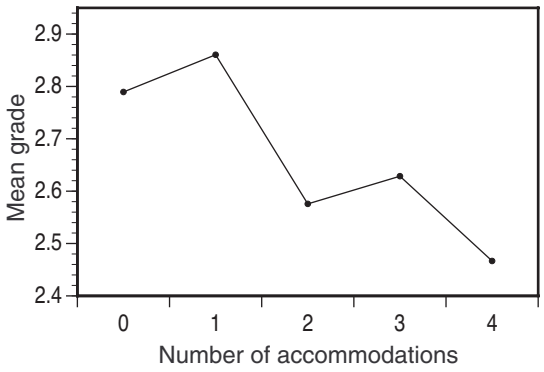
12.25. **(a)** The plot suggests that both drugs cause an increase in activity level, and Drug B appears to have a greater effect. **(b)** Yes: The guidelines for pooling standard deviations say that the ratio of largest to smallest should be less than 2; we have $\sqrt{\frac{12.25}{6.25}} = 1.4 < 2$. The pooled variance is

$$s_p^2 = \frac{3(s_1^2 + s_2^2 + \cdots + s_5^2)}{3+3+3+3+3} = \frac{149.25}{15} = 9.95$$

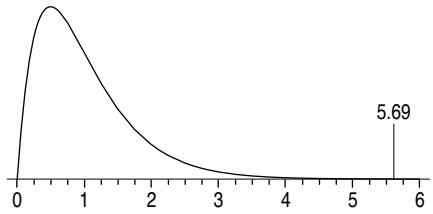
and $s_p = \sqrt{9.95} \doteq 3.1544$. **(c)** The degrees of freedom are DFG $= I - 1 = 4$ and DFE $= N - I = 15$. **(d)** Comparing to an $F(4, 15)$ distribution in Table E, we see that $0.010 < P < 0.025$; software gives $P \doteq 0.0156$. We have significant evidence that the means are not all the same.



12.26. **(a)** It is useful to connect the points on the plot, to make the pattern (or lack thereof) more evident. There is some suggestion that average grade decreases as the number of accommodations increases. **(b)** Having too many decimal points is distracting; in this situation, no useful information is gained by having more than one or two digits after the decimal point. For example, the first mean and standard deviation would be more effectively presented as 2.79 and 0.85. **(c)** The largest-to-smallest SD ratio is slightly over 2 (about 2.009), so pooling is not advisable. (If we pool in spite of this, we find $s_p \doteq 0.8589$.) **(d)** Eliminating data points (without a legitimate reason) is always risky, although we could run the analysis with and without them. Combining the last three groups would be a bad idea if the data suggested that grades rebounded after 2 accommodations (i.e., if the average grades were higher for 3 and 4 accommodations), but as that is not the case, lumping 2, 3, and 4 accommodations seems reasonable. **(e)** ANOVA is not appropriate for these data, chiefly because we do not have 245 independent observations. **(f)** There may be a number of local factors (for example, student demographics or teachers' attitudes toward accommodations) which affected grades; these effects might not be the same elsewhere. **(g)** One weakness is that we do not have a control group for comparison; that is, we cannot tell what grades these students (or a similar group) would have had without accommodations.



12.27. **(a)** The variation in sample size is some cause for concern, but there can be no extreme outliers in a 1-to-7 scale, so ANOVA is probably reliable. **(b)** Pooling is reasonable: $\frac{1.26}{1.03} \doteq 1.22 < 2$. **(c)** With $I = 5$ groups and total sample size $N = 410$, we use an $F(4, 405)$ distribution. We can compare 5.69 to an $F(4, 200)$ distribution in Table E and conclude that $P < 0.001$, or with software determine that $P \doteq 0.0002$. **(d)** Hispanic Americans have the highest emotion scores, Japanese are in the middle, and the other three cultures are the lowest (and very similar).



12.28. (a) The largest-to-smallest SD ratios are 2.84, 1.23, and 1.14, so the text's guidelines are satisfied for intensity and recall, but not for frequency. **(b)** As in the previous exercise, $I = 5$ and $N = 410$, so we use an $F(4, 405)$ distribution. From the $F(4, 200)$ distribution in Table E, we can conclude that $P < 0.001$ for all three response variables. With software, we find that the P-values are much smaller; all are less than 0.00002. We conclude that, for each variable, we have strong evidence that some group mean is different. (This conclusion is cautious in the case of frequency because of our concern about the standard deviations.) **(c)** The table below shows one way of summarizing the means. For each variable, it attempts to identify low (underlined), medium, and high (boldface) values of that variable. Hispanic Americans were higher than other groups for all four variables. Asian Americans were low for all variables (the lowest in all but global score). Japanese were low on all but global score, while European Americans and Indians were in the middle for all but global score. **(d)** The results might not generalize to, for example, subjects who are from different parts of their countries or not in a college or university community. **(e)** Create a two-way table with counts of men and women in each cultural group. The Minitab output on the right gives $X^2 = 11.353$, df $= 4$, and $P = 0.023$, so we have evidence (significant at $\alpha = 0.05$) that the gender mix was not the same for all cultures. Specifically, Hispanic Americans and European Americans had higher percentages of women, which might further affect how much we can generalize the results.

Minitab output

```
             Women       Men     Total
   1            38         8        46
             31.64     14.36

   2            22        11        33
             22.70     10.30

   3            57        34        91
             62.59     28.41

   4           102        58       160
            110.05     49.95

   5            63        17        80
             55.02     24.98

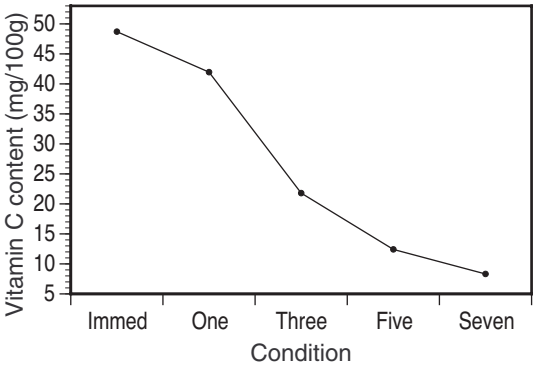
Total          282       128       410

ChiSq =  1.279 +  2.817 +
         0.021 +  0.047 +
         0.499 +  1.100 +
         0.589 +  1.297 +
         1.156 +  2.547 = 11.353
df = 4, p = 0.023
```

	Score	Frequency	Intensity	Recall
European Amer.	<u>4.39</u>	82.87	2.79	49.12
Asian Amer.	<u>4.35</u>	<u>72.68</u>	<u>2.37</u>	<u>39.77</u>
Japanese	4.72	<u>73.36</u>	<u>2.53</u>	<u>43.98</u>
Indian	<u>4.34</u>	82.71	2.87	49.86
Hispanic Amer.	**5.04**	**92.25**	**3.21**	**59.99**

12.29. **(a)** Table below ($\bar{x}$, s, $s_{\bar{x}}$ in mg/100 g). **(b)** To test H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ vs. H_a: not all μ_i are equal, we have $F = 367.74$ with df 4 and 5, and $P < 0.0005$, so we reject the null hypothesis. Minitab output below. **(c)** Plot below. We conclude that vitamin C content decreases over time.

Condition	n	$\bar{x}$	s	$s_{\bar{x}}$
Immediate	2	48.705	1.534	1.085
One day	2	41.955	2.128	1.505
Three days	2	21.795	0.771	0.545
Five days	2	12.415	1.082	0.765
Seven days	2	8.320	0.269	0.190



Minitab output

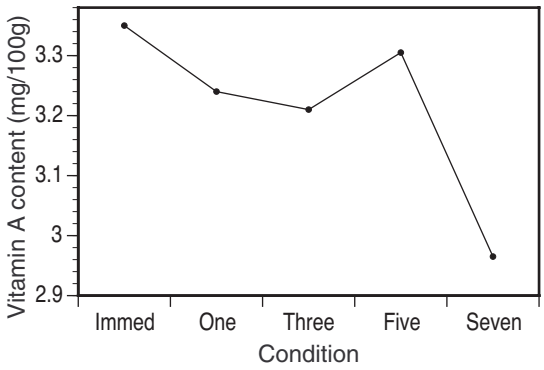
```
Analysis of Variance on VitC
Source     DF         SS         MS         F        p
Days        4    2565.72     641.43    367.74    0.000
Error       5       8.72       1.74
Total       9    2574.44
```

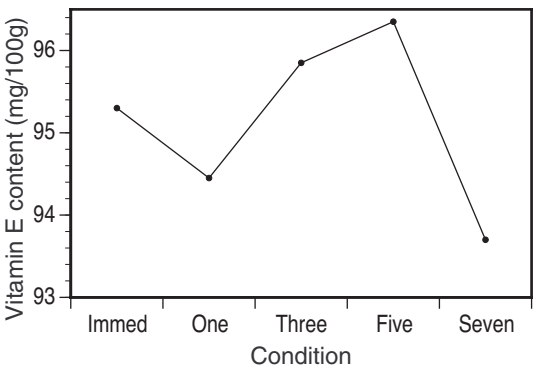
12.30. We have 10 comparisons to make, and df = 5, so the Bonferroni critical value with $\alpha = 0.05$ is $t^{**} = 4.7733$. The pooled standard deviation is $s_p \doteq 1.3207$, and the standard error of each difference is $s_p\sqrt{1/2 + 1/2} = s_p$, so two means are significantly different if they differ by $t^{**}s_p \doteq 6.3041$. All differences are significant *except* the five-day/seven-day difference.

12.31. Means, etc. ($\bar{x}$, s, $s_{\bar{x}}$ in mg/100 g) below, along with plots of means. The hypotheses are H_0: $\mu_1 = \cdots = \mu_5$ vs. H_a: not all μ_i are equal. For vitamin A, $F = 12.09$ (df 4 and 5), so $P = 0.009$—we reject H_0 and conclude that vitamin A content changes over time (it appears to decrease, except for the rise at "Five days"). For vitamin E, $F = 0.69$ (df 4 and 5), so $P = 0.630$—we cannot reject the null hypothesis.

Vitamin A	n	$\bar{x}$	s	$s_{\bar{x}}$
Immediate	2	3.350	0.01414	0.010
One day	2	3.240	0.05657	0.040
Three days	2	3.210	0.07071	0.050
Five days	2	3.305	0.07778	0.055
Seven days	2	2.965	0.06364	0.045

Vitamin E	n	$\bar{x}$	s	$s_{\bar{x}}$
Immediate	2	95.30	0.9900	0.700
One day	2	94.45	1.7678	1.250
Three days	2	95.85	2.1920	1.550
Five days	2	96.35	1.9092	1.350
Seven days	2	93.70	1.9799	1.400





Minitab output

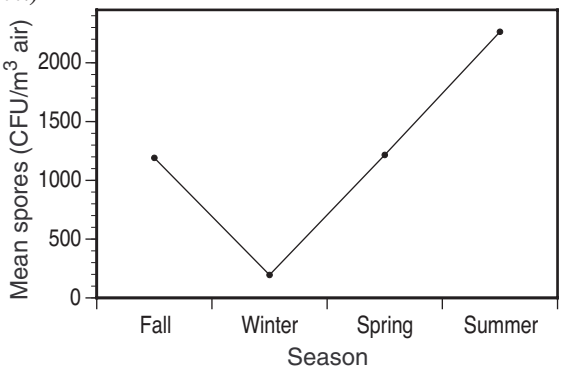
```
Analysis of Variance on VitA
Source     DF        SS        MS        F        p
Days        4   0.17894   0.04473    12.09    0.009
Error       5   0.01850   0.00370
Total       9   0.19744
--------------------------------------------------
Analysis of Variance on VitE
Source     DF        SS        MS        F        p
Days        4      9.09      2.27     0.69    0.630
Error       5     16.47      3.29
Total       9     25.56
```

12.32. (a) We did not reject H_0 for vitamin E, so we have no reason to believe that any means are different. **(b)** As in Exercise 12.30, $t^{**} = 4.7733$. The pooled standard deviation is $s_p \doteq 0.06083$ and the standard error of each difference is $s_p\sqrt{1/2 + 1/2} = s_p$, so two means are significantly different if they differ by $t^{**}s_p \doteq 0.2903$. By this standard, only the immediate/seven-day and five-day/seven-day differences are significant.

12.33. There is no significant evidence of change in vitamin E content. Vitamin C content decreases over time, with significant change occurring over the first five days. For vitamin A, the fluctuations over the first five days are not significant, but the drop from day 5 to day 7 is.

12.34. (a) Means, standard deviations, and standard errors of the mean are below, along with a plot of the means. Note that the standard deviations grossly violate our guideline for pooling: $\frac{1066.97}{89.94} \doteq 11.86$. As we might expect, spores are highest in the summer, and lowest in the winter. **(b)** ANOVA gives $F = 4.83$ (df 3 and 8) and $P = 0.033$—the differences are significant at the 5% level. (The small samples and large variation make the seemingly large differences not as significant as we might expect.)

	n	$\bar{x}$	s	$s_{\bar{x}}$
Fall	3	1191	89.94	51.926
Winter	3	195	163.72	94.522
Spring	3	1216	773.12	446.360
Summer	3	$2263.\overline{6}$	1066.97	616.014

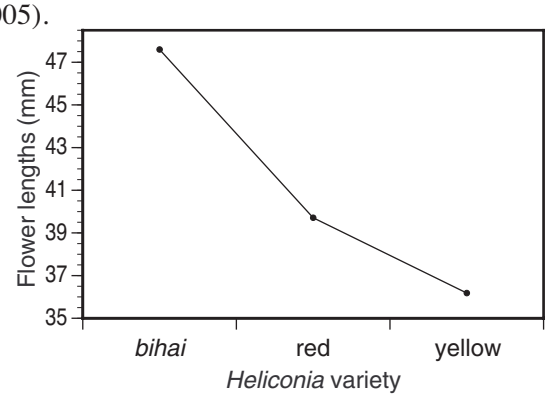


Minitab output

```
Analysis of Variance
Source     DF        SS        MS        F        p
Factor      3   6422013   2140671     4.83    0.033
Error       8   3542047    442756
Total      11   9964060
```

12.35. See also the solution to Exercise 1.79. The means, standard deviations, and standard errors (all in millimeters) are given below. We reject H_0 and conclude that at least one mean is different ($F = 259.12$, df 2 and 51, $P < 0.0005$).

Variety	n	$\bar{x}$	s	$s_{\bar{x}}$
bihai	16	47.5975	1.2129	0.3032
red	23	39.7113	1.7988	0.3751
yellow	15	36.1800	0.9753	0.2518

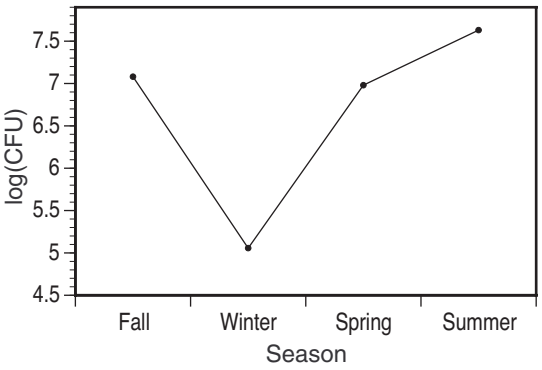


Minitab output

```
Analysis of Variance on length
Source     DF        SS        MS        F        p
Factor      2   1082.87    541.44   259.12    0.000
Error      51    106.57      2.09
Total      53   1189.44
```

12.36. Means, standard deviations, and standard errors of the mean are below, along with a plot of the means. ANOVA gives $F = 11.93$ (df 3 and 8) and $P = 0.003$—the differences are significant at the 1% level. Taking logarithms made the evidence against H_0 stronger, in part because the standard deviations were greatly reduced. (They are now in line with the text's guidelines for pooling.)

	n	$\bar{x}$	s	$s_{\bar{x}}$
Fall	3	7.0806	0.0772	0.0446
Winter	3	5.0566	0.7751	0.4475
Spring	3	6.9801	0.5891	0.3401
Summer	3	7.6305	0.5630	0.3250

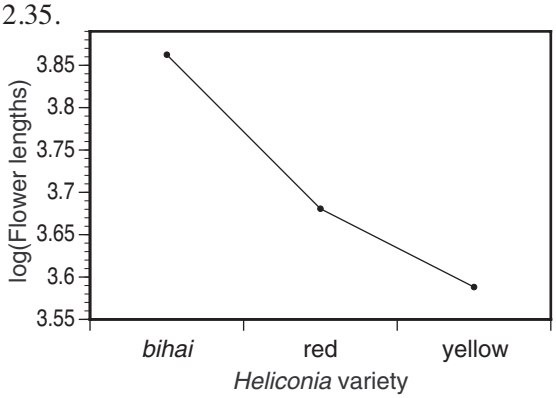


Minitab output

```
Analysis of Variance
Source     DF        SS        MS        F        p
Factor      3    11.368     3.789    11.93    0.003
Error       8     2.541     0.318
Total      11    13.909
```

12.37. The means, standard deviations, and standard errors are given below. We reject H_0 and conclude that at least one mean is different ($F = 244.27$, df 2 and 51, $P < 0.0005$). These results are essentially the same as in Exercise 12.35.

Variety	n	$\bar{x}$	s	$s_{\bar{x}}$
bihai	16	3.8625	0.02515	0.006286
red	23	3.6807	0.04496	0.009374
yellow	15	3.5882	0.02698	0.006966



Minitab output

```
Analysis of Variance on loglen
Source     DF        SS        MS        F        p
Factor      2   0.61438   0.30719   244.27    0.000
Error      51   0.06414   0.00126
Total      53   0.67852
```

12.38. **(a)** Statistics and plots are below. **(b)** The standard deviations satisfy the text's guidelines for pooling. One concern is that all three distributions are slightly left-skewed and the youngest nonfiction death is an outlier. **(c)** ANOVA gives $F = 6.56$ (df 2 and 120) and $P = 0.002$, so we conclude that at least one mean is different. **(d)** The appropriate contrast is $\psi_1 = \frac{1}{2}(\mu_{\text{nov}} + \mu_{\text{nf}}) - \mu_{\text{p}}$. (This is defined so that the $\psi_1 > 0$ if poets die younger. This is not absolutely necessary but is in keeping with the text's advice.) The null hypothesis is H_0: $\psi_1 = 0$; the Yeats quote hardly seems like an adequate reason to choose a one-sided alternative, but students may have other opinions. For the test, we compute $c \doteq 10.9739$, $\text{SE}_c \doteq 3.0808$, and $t \doteq 3.56$ with df $= 120$. The P-value is very small regardless of whether H_a is one- or two-sided, so we conclude that the contrast is positive (and poets die young). **(e)** For this comparison, the contrast is $\psi_2 = \mu_{\text{nov}} - \mu_{\text{nf}}$, and the hypotheses are H_0: $\psi_2 = 0$ vs. H_a: $\psi_2 \neq 0$. (Because the alternative is two-sided, the subtraction in this contrast can go either way.) For the test, we compute $c \doteq -5.4272$, $\text{SE}_c \doteq 3.4397$, and $t \doteq -1.58$ with df $= 120$. This gives $P = 0.1172$; the difference between novelists and nonfiction writers is not significant. **(f)** With three comparisons and df $= 120$, the Bonferroni critical value is $t^{**} = 2.4280$. The pooled standard deviation is $s_p \doteq 14.4592$, so the differences, standard errors, and t values are:

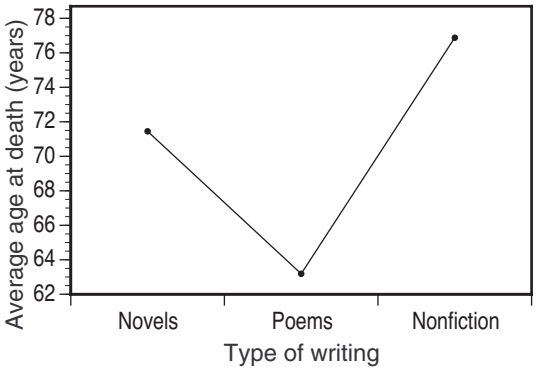
$$\bar{x}_{\text{nov}} - \bar{x}_{\text{p}} \doteq 8.2603, \quad \text{SE}_{\text{nov-p}} = s_p\sqrt{\tfrac{1}{67} + \tfrac{1}{32}} \doteq 3.1071, \quad t \doteq 2.66$$

$$\bar{x}_{\text{nov}} - \bar{x}_{\text{nf}} \doteq -5.4272, \quad \text{SE}_{\text{nov-nf}} = s_p\sqrt{\tfrac{1}{67} + \tfrac{1}{24}} \doteq 3.4397, \quad t \doteq -1.58$$

$$\bar{x}_{\text{p}} - \bar{x}_{\text{nf}} \doteq -13.6875, \quad \text{SE}_{\text{p-nf}} = s_p\sqrt{\tfrac{1}{32} + \tfrac{1}{24}} \doteq 3.9044, \quad t \doteq -3.51$$

The first and last differences are greater (in absolute value) than t^{**}, so those differences are significant. The second difference is the same one tested in the contrast of part (e); the standard error and the conclusion are the same.

	n	$\bar{x}$	s	$s_{\bar{x}}$
Novels	67	71.4478	13.0515	1.5945
Poems	32	63.1875	17.2971	3.0577
Nonfiction	24	76.8750	14.0969	2.8775

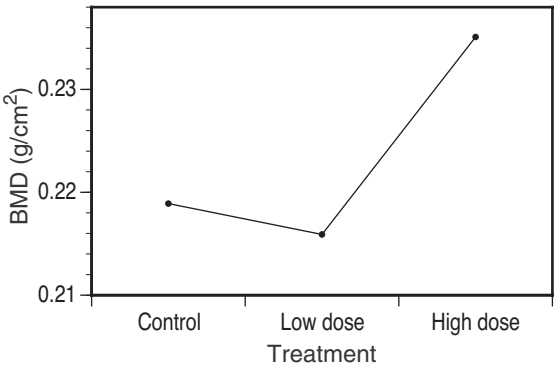


Minitab output

```
Analysis of Variance on AgeDeath
Source     DF         SS         MS        F        p
Writer      2       2744       1372     6.56    0.002
Error     120      25088        209
Total     122      27832
```

12.39. **(a)** The means, standard deviations, and standard errors are given below (all in grams per cm^2). **(b)** All three distributions appear to reasonably close to Normal, and the standard deviations are suitable for pooling. **(c)** ANOVA gives $F = 7.72$ (df 2 and 42) and $P = 0.001$, so we conclude that the means are not all the same. **(d)** With df $= 42$, 3 comparisons, and $\alpha = 0.05$, the Bonferroni critical value is $t^{**} = 2.4937$. The pooled standard deviation is $s_p \doteq 0.01437$ and the standard error of each difference is $\text{SE}_D = s_p\sqrt{1/15 + 1/15} \doteq 0.005246$, so two means are significantly different if they differ by $t^{**}\text{SE}_D \doteq 0.01308$. The high-dose mean is significantly different from the other two. **(e)** Briefly: High doses of kudzu isoflavones increase BMD.

	n	$\bar{x}$	s	$s_{\bar{x}}$
Control	15	0.2189	0.01159	0.002992
Low dose	15	0.2159	0.01151	0.002972
High dose	15	0.2351	0.01877	0.004847

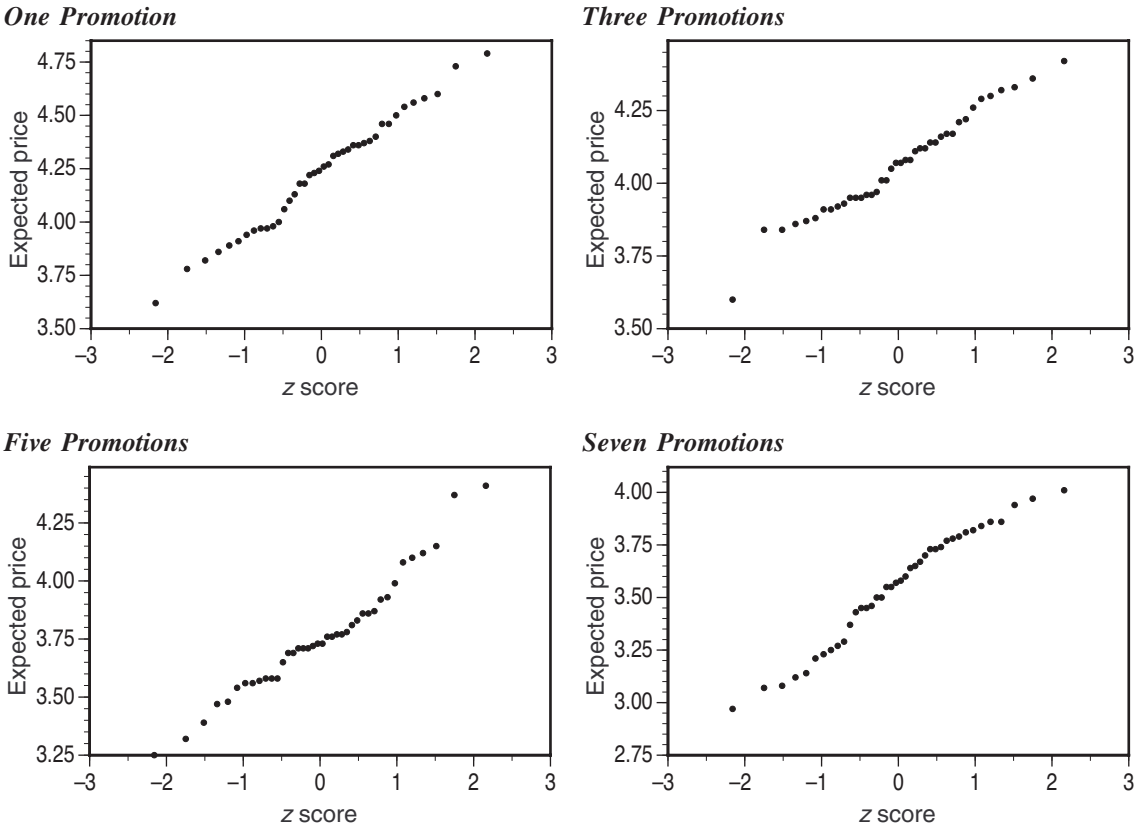


Minitab output

```
Analysis of Variance on BMD
Source     DF        SS        MS        F        p
Factor      2  0.003186  0.001593     7.72    0.001
Error      42  0.008668  0.000206
Total      44  0.011853
```

12.40. **(a)** All four data sets appear to be reasonably close to Normal, although "3 promotions" seems to have a low outlier. Plots on the next page. **(b)** At right ($\bar{x}$, s, $s_{\bar{x}}$ in dollars). **(c)** The ratio of largest to smallest standard deviations is about 1.58, so the assumption of equal standard deviations is reasonable. **(d)** We test H_0: $\mu_1 = \cdots = \mu_4$ vs. H_a: not all μ_i are equal. Minitab output gives $F = 59.90$ with df 3 and 156 and $P < 0.0005$, so we reject the null hypothesis. With more promotions, expected price decreases.

Promotions	n	$\bar{x}$	s	$s_{\bar{x}}$
One	40	4.2240	0.2734	0.0432
Three	40	4.0627	0.1742	0.0275
Five	40	3.7590	0.2526	0.0399
Seven	40	3.5487	0.2750	0.0435

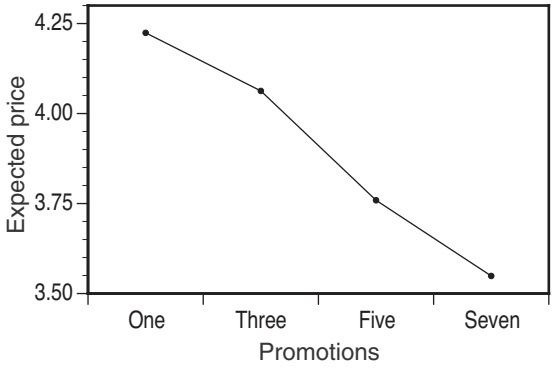


Minitab output

```
Analysis of Variance on ExpPrice
Source     DF        SS        MS        F        p
NumPromo    3   10.9885    3.6628    59.90    0.000
Error     156    9.5388    0.0611
Total     159   20.5273
```

12.41. We have six comparisons to make, and df $= 156$, so the Bonferroni critical value with $\alpha = 0.05$ is $t^{**} = 2.6723$. The pooled standard deviation is $s_p \doteq 0.2473$, so the standard error of each difference is $\text{SE}_D = s_p\sqrt{1/40 + 1/40} \doteq 0.05529$. Two means must differ by $t^{**}\text{SE}_D \doteq 0.1477$ in order to be significantly different; all six differences are significant. (Note that because the means decrease, we could consider only the differences in consecutive means, that is, $\bar{x}_1 - \bar{x}_3$, $\bar{x}_3 - \bar{x}_5$, and $\bar{x}_5 - \bar{x}_7$. Since these three differences are significant, it follows that the others must be, too.)

$\bar{x}_1 - \bar{x}_3 = 0.16125$	$t_{13} = 2.916$
$\bar{x}_1 - \bar{x}_5 = 0.46500$	$t_{15} = 8.410$
$\bar{x}_1 - \bar{x}_7 = 0.67525$	$t_{17} = 12.212$
$\bar{x}_3 - \bar{x}_5 = 0.30375$	$t_{35} = 5.493$
$\bar{x}_3 - \bar{x}_7 = 0.51400$	$t_{37} = 9.296$
$\bar{x}_5 - \bar{x}_7 = 0.21025$	$t_{57} = 3.802$



12.42. **(a)** At right. **(b)** We test H_0: $\mu_1 = \cdots = \mu_4$ vs. H_a: not all μ_i are equal. ANOVA gives $F = 9.24$ with df 3 and 74, for which $P < 0.0005$, so we reject the null hypothesis. The type of lesson does affect the mean score change; in particular, it appears that students who take piano lessons had significantly higher scores than the other students.

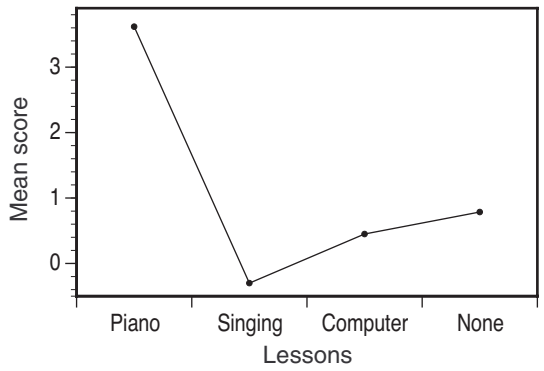
Lesson	n	$\bar{x}$	s	$s_{\bar{x}}$
Piano	34	3.6176	3.0552	0.5240
Singing	10	−0.3000	1.4944	0.4726
Computer	20	0.4500	2.2118	0.4946
None	14	0.7857	3.1908	0.8528

Minitab output

```
Analysis of Variance on Scores
Source     DF        SS        MS        F        p
Lesson      3    207.28     69.09     9.24    0.000
Error      74    553.44      7.48
Total      77    760.72
```

12.43. We have six comparisons to make, and df = 74, so the Bonferroni critical value with $\alpha = 0.05$ is $t^{**} = 2.7111$. The pooled standard deviation is $s_p \doteq 2.7348$. The table below shows the differences, their standard errors, and the t statistics.

The Piano mean is significantly higher than the other three, but the other three means are not significantly different.



$D_{PS} = 3.91765$	$D_{PC} = 3.16765$	$D_{PN} = 2.83193$
$\text{SE}_{PS} = 0.98380$	$\text{SE}_{PC} = 0.77066$	$\text{SE}_{PN} = 0.86843$
$t_{PS} = 3.982$	$t_{PC} = 4.110$	$t_{PN} = 3.261$
	$D_{SC} = -0.75000$	$D_{SN} = -1.08571$
	$\text{SE}_{SC} = 1.05917$	$\text{SE}_{SN} = 1.13230$
	$t_{SC} = -0.708$	$t_{SN} = -0.959$
		$D_{CN} = -0.33571$
		$\text{SE}_{CN} = 0.95297$
		$t_{CN} = -0.352$

12.44. We test the hypothesis H_0: $\psi = \mu_\text{P} - \frac{1}{3}(\mu_\text{S} + \mu_\text{C} + \mu_\text{N}) = 0$; the sample contrast is $c \doteq 3.618 - \frac{1}{3}(-0.300 + 0.450 + 0.786) = 3.306$. The pooled standard deviation estimate is $s_p = 2.735$, so $\text{SE}_c = 2.735\sqrt{1/34 + \frac{1}{9}/10 + \frac{1}{9}/20 + \frac{1}{9}/14} \doteq 0.6356$. Then $t = 3.306/0.6356 \doteq 5.20$, with df = 74. This is enough evidence ($P < 0.001$) to reject H_0 in favor of H_a: $\psi > 0$, so we conclude that mean score changes for piano students are greater than the average of the means for the other three groups.

12.45. **(a)** Pooling is reasonable, as the largest-to-smallest ratio is about 1.31. **(b)** ANOVA gives $F = 2.72$ (df 2 and 42), for which $P = 0.078$. We cannot reject H_0.

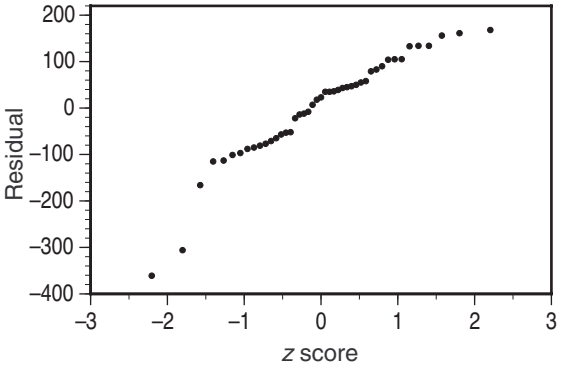
	n	$\bar{x}$	s
4 months	19	570.00	122.96
5 months	18	483.00	112.95
6 months	8	541.88	93.96

Minitab output

```
Analysis of Variance on energy
Source     DF        SS        MS        F        p
Treatment   2     71288     35644     2.72    0.078
Error      42    550811     13115
Total      44    622099
```

12.46. **(a)** Aside from two low outliers, the residuals are reasonably Normal. **(b)** We did not reject H_0, so we do not need comparisons to determine which means are different.

−3	6
−3	0
−2	
−2	
−1	6
−1	110
−0	9888776555
−0	2110
0	0123333444
0	555789
1	000333
1	566



12.47. **(a)** Pooling is reasonable, as the largest-to-smallest ratio is about 1.65. **(b)** ANOVA gives $F = 7.98$ (df 2 and 27), for which $P = 0.002$. We reject H_0 and conclude that not all means are equal.

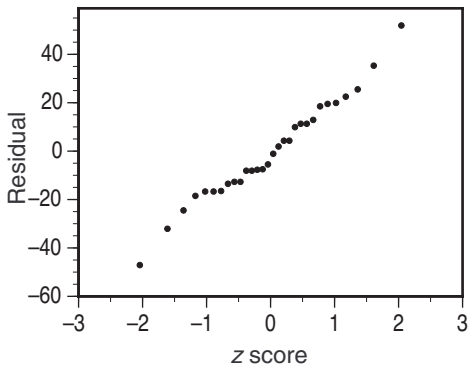
	n	$\bar{x}$	s
Control	10	601.1	27.364
Low jump	10	612.5	19.329
High jump	10	638.7	16.594

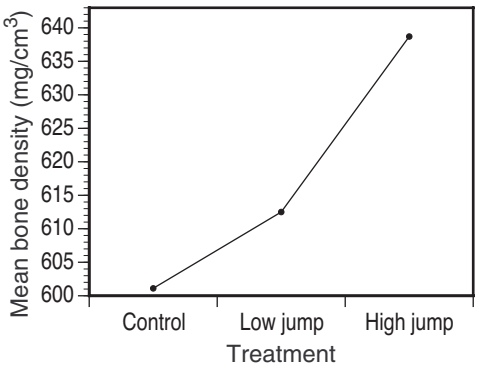
Minitab output

```
Analysis of Variance on density
Source     DF        SS        MS        F        p
Treatment   2      7434      3717     7.98    0.002
Error      27     12580       466
Total      29     20013
```

12.48. **(a)** The residuals appear to be reasonably Normal. **(b)** With df = 27, 3 comparisons, and $\alpha = 0.05$, the Bonferroni critical value is $t^{**} = 2.5525$. The pooled standard deviation is $s_p \doteq 21.5849$, and the standard error of each difference is $\text{SE}_D = s_p\sqrt{1/10 + 1/10} \doteq 9.6531$, so two means are significantly different if they differ by $t^{**}\text{SE}_D \doteq 24.6390$. The high-jump mean is significantly different from the other two.

−4	7
−3	2
−2	4
−1	8666322
−0	887751
0	1449
1	112899
2	25
3	5
4	
5	1





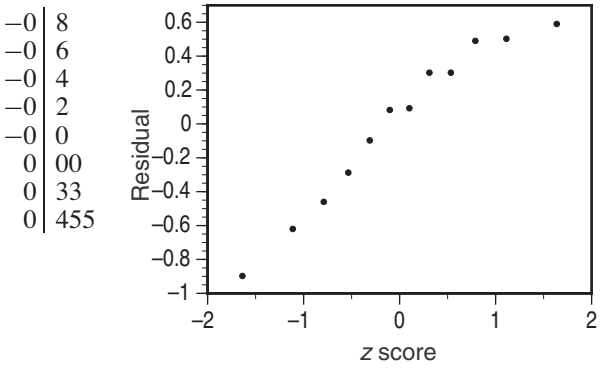
12.49. **(a)** Pooling is risky because $0.6283/0.2520 = 2.49 > 2$. **(b)** ANOVA gives $F = 31.16$ (df 2 and 9), for which $P < 0.0005$. We reject H_0 and conclude that not all means are equal.

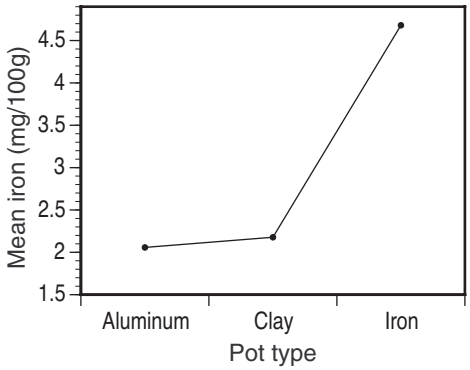
	n	$\bar{x}$	s
Aluminum	4	2.0575	0.2520
Clay	4	2.1775	0.6213
Iron	4	4.6800	0.6283

Minitab output

```
Analysis of Variance on iron
Source     DF        SS        MS        F        p
Pot         2    17.539     8.770    31.16    0.000
Error       9     2.533     0.281
Total      11    20.072
```

12.50. **(a)** There are no clear violations of Normality, but the number of residuals is so small that is difficult to draw any conclusions. **(b)** With df = 9, 3 comparisons, and $\alpha = 0.05$, the Bonferroni critical value is $t^{**} = 2.9333$. The pooled standard deviation is $s_p \doteq 0.5305$, and the standard error of each difference is $\text{SE}_D = s_p\sqrt{1/4 + 1/4} \doteq 0.3751$, so two means are significantly different if they differ by $t^{**}\text{SE}_D \doteq 1.1003$. The iron mean is significantly higher than the other two.





12.51. **(a)** Pooling is risky because $8.66/2.89 = 3 > 2$. **(b)** ANOVA gives $F = 137.94$ (df 5 and 12), for which $P < 0.0005$. We reject H_0 and conclude that not all means are equal.

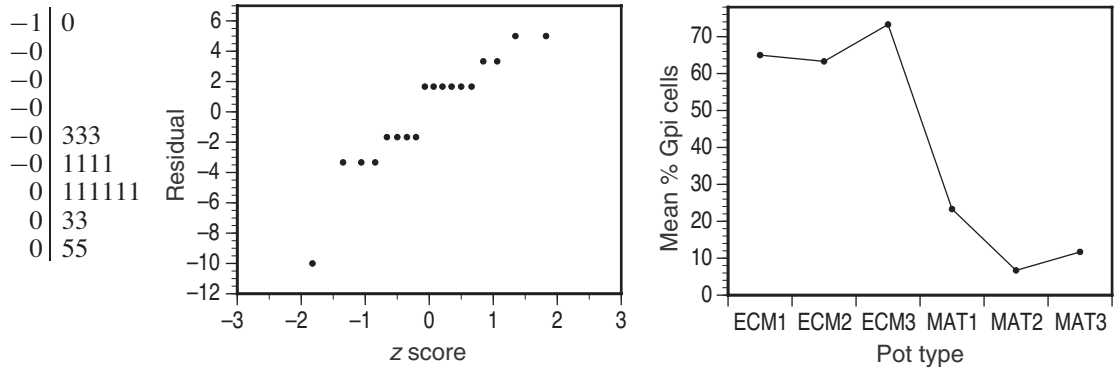
	n	$\bar{x}$	s
ECM1	3	65.0%	8.6603%
ECM2	3	$63.\overline{3}\%$	2.8868%
ECM3	3	$73.\overline{3}\%$	2.8868%
MAT1	3	$23.\overline{3}\%$	2.8868%
MAT2	3	$6.\overline{6}\%$	2.8868%
MAT3	3	$11.\overline{6}\%$	2.8868%

Minitab output

```
Analysis of Variance on gpi
Source      DF        SS        MS        F        p
Treatment    5   13411.1    2682.2   137.94    0.000
Error       12     233.3      19.4
Total       17   13644.4
```

12.52. **(a)** The residuals have one low outlier, and a lot of granularity, so Normality is difficult to assess. **(b)** With df = 12, 15 comparisons, and $\alpha = 0.05$, the Bonferroni critical value is $t^{**} = 3.6489$. The pooled standard deviation is $s_p \doteq 4.4096\%$, and the standard error of each difference is $\text{SE}_D = s_p\sqrt{1/3 + 1/3} \doteq 3.6004\%$, so two means are significantly different if they differ by $t^{**}\text{SE}_D \doteq 13.1375\%$. The three ECM means are significantly higher than the three MAT means. **(c)** The contrast is $\psi = \frac{1}{3}(\mu_{\text{ECM1}} + \mu_{\text{ECM2}} + \mu_{\text{ECM3}}) - \frac{1}{3}(\mu_{\text{MAT1}} + \mu_{\text{MAT2}} + \mu_{\text{MAT3}})$, and the hypotheses are H_0: $\psi = 0$ vs. H_a: $\psi \neq 0$. For the test, we compute $c \doteq 53.33\%$, $\text{SE}_c \doteq 2.0787\%$, and

$t \doteq 25.66$ with df $= 12$. This has a tiny P-value; the difference between ECM and MAT is highly significant. This is consistent with the Bonferroni results from part (b).



12.53. Let μ_1 be the placebo mean, μ_2 and μ_3 be the means for low and high doses of Drug A, and μ_4 and μ_5 be the means for low and high doses of Drug B. Recall that $s_p \doteq 3.1544$. **(a)** The first contrast is $\psi_1 = \mu_1 - 0.5\mu_2 - 0.5\mu_4$; the second is $\psi_2 = \mu_3 - \mu_2 - (\mu_5 - \mu_4)$. **(b)** The estimated contrasts are $c_1 = 14.00 - 0.5(15.25) - 0.5(16.75) = -2$ and $c_2 = (18.25 - 15.25) - (22.50 - 16.75) = -2.75$. The respective standard errors are:

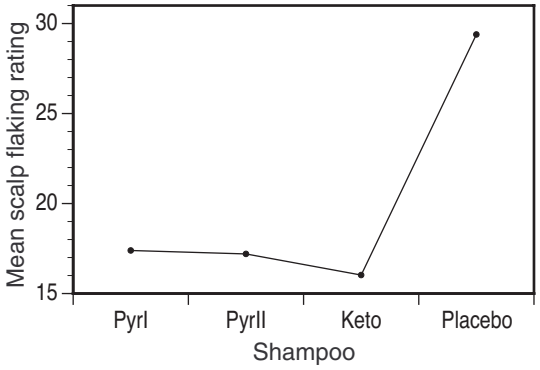
$$\text{SE}_{c_1} = s_p\sqrt{\frac{1}{4} + \frac{0.25}{4} + \frac{0}{4} + \frac{0.25}{4} + \frac{0}{4}} \doteq 1.9316 \text{ and}$$

$$\text{SE}_{c_2} = s_p\sqrt{\frac{0}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}} \doteq 3.1544$$

(c) Neither contrast is significant ($t_1 \doteq -1.035$ and $t_2 \doteq -0.872$, for which the P-values are 0.1584 and 0.1985). We do not have enough evidence to conclude that low doses increase activity level over a placebo, nor can we conclude that activity level changes due to increased dosage are different between the two drugs.

12.54. **(a)** Below. **(b)** To test H_0: $\mu_1 = \cdots = \mu_4$ vs. H_a: not all μ_i are equal, ANOVA (Minitab output below) gives $F = 967.82$ (df 3 and 351), which has $P < 0.0005$. We conclude that not all means are equal; specifically, the "Placebo" mean is much higher than the other three means.

Shampoo	n	$\bar{x}$	s	$s_{\bar{x}}$
PyrI	112	17.393	1.142	0.108
PyrII	109	17.202	1.352	0.130
Keto	106	16.028	0.931	0.090
Placebo	28	29.393	1.595	0.301

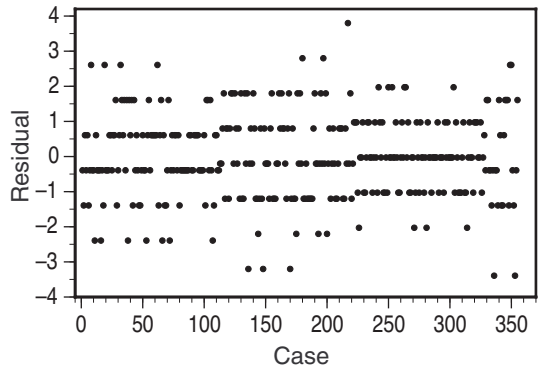


Minitab output

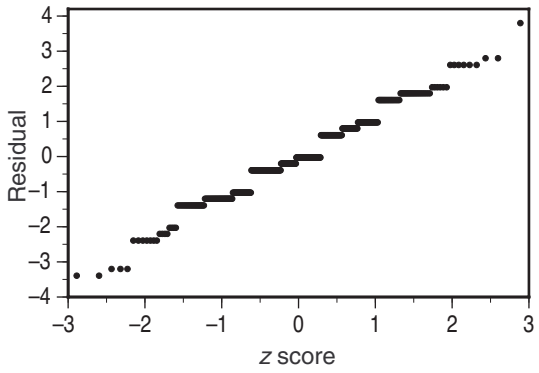
```
Analysis of Variance on Flaking
Source     DF        SS         MS        F       p
Code        3   4151.43    1383.81   967.82   0.000
Error     351    501.87       1.43
Total     354   4653.30
```

12.55. **(a)** The plot (below) shows granularity (which varies between groups), but that should not make us question independence; it is due to the fact that the scores are all integers. **(b)** The ratio of the largest to the smallest standard deviations is $1.595/0.931 \doteq 1.714$—less than 2. **(c)** Apart from the granularity, the quantile plots (below) are reasonably straight. **(d)** Again, apart from the granularity, the quantile plot looks pretty good.

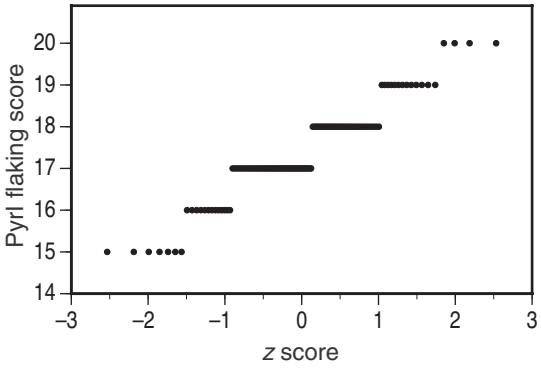
12.55(a)



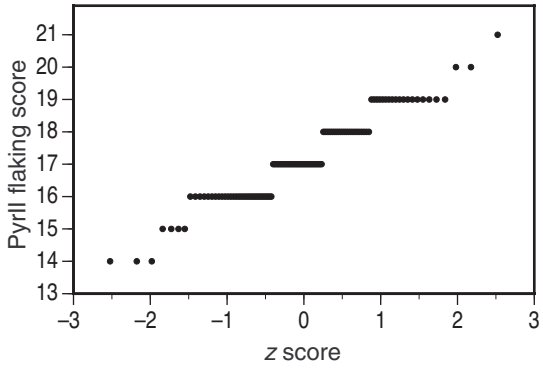
12.55(d)



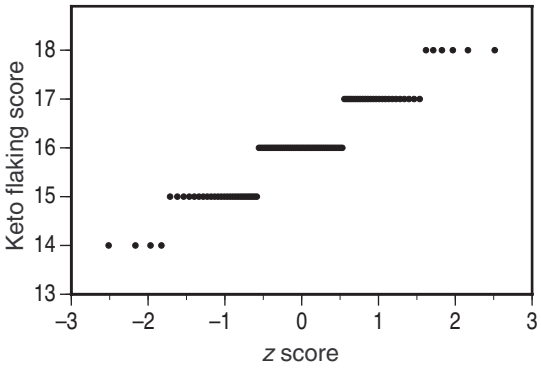
12.55(c)–PyrI



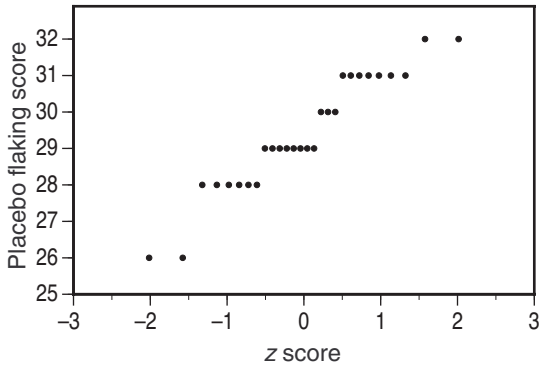
12.55(c)–PyrII



12.55(c)–Keto



12.55(c)–Placebo



12.56. We have six comparisons to make, and df = 351, so the Bonferroni critical value with $\alpha = 0.05$ is $t^{**} = 2.6533$. The pooled standard deviation is $s_p = \sqrt{\text{MSE}} \doteq 1.1958$; the differences, standard errors, and t statistics are below. The only *non*significant difference is between the two Pyr treatments (meaning the second application of the shampoo is of little benefit). The Keto shampoo mean is the lowest; the placebo mean is by far the highest.

$D_{\text{Py1}-\text{Py2}} = 0.19102$	$D_{\text{Py1}-\text{K}} = 1.36456$	$D_{\text{Py1}-\text{P}} = -12.0000$
$\text{SE}_{\text{Py1}-\text{Py2}} = 0.16088$	$\text{SE}_{\text{Py1}-\text{K}} = 0.16203$	$\text{SE}_{\text{Py1}-\text{P}} = 0.25265$
$t_{\text{Py1}-\text{Py2}} = 1.187$	$t_{\text{Py1}-\text{K}} = 8.421$	$t_{\text{Py1}-\text{P}} = -47.497$
	$D_{\text{Py2}-\text{K}} = 1.17353$	$D_{\text{Py2}-\text{P}} = -12.1910$
	$\text{SE}_{\text{Py2}-\text{K}} = 0.16312$	$\text{SE}_{\text{Py2}-\text{P}} = 0.25334$
	$t_{\text{Py2}-\text{K}} = 7.195$	$t_{\text{Py2}-\text{P}} = -48.121$
		$D_{\text{K}-\text{P}} = -13.3646$
		$\text{SE}_{\text{K}-\text{P}} = 0.25407$
		$t_{\text{K}-\text{P}} = -52.601$

12.57. **(a)** The three contrasts are:

$$\psi_1 = \tfrac{1}{3}\mu_{\text{Py1}} + \tfrac{1}{3}\mu_{\text{Py2}} + \tfrac{1}{3}\mu_{\text{K}} - \mu_{\text{P}}$$
$$\psi_2 = \tfrac{1}{2}\mu_{\text{Py1}} + \tfrac{1}{2}\mu_{\text{Py2}} - \mu_{\text{K}}$$
$$\psi_3 = \mu_{\text{Py1}} - \mu_{\text{Py2}}$$

$c_1 = -12.51$	$c_2 = 1.269$	$c_3 = 0.191$
$\text{SE}_{c_1} \doteq 0.2355$	$\text{SE}_{c_2} \doteq 0.1413$	$\text{SE}_{c_3} \doteq 0.1609$
$t_1 = -53.17$	$t_2 = 8.98$	$t_3 = 1.19$
$P_1 < 0.0005$	$P_2 < 0.0005$	$P_3 \doteq 0.2359$

(b) The pooled standard deviation is $s_p = \sqrt{\text{MSE}} \doteq 1.1958$. The estimated contrasts and their standard errors are in the table. For example:

$$\text{SE}_{c_1} = s_p\sqrt{\tfrac{1}{9}/112 + \tfrac{1}{9}/109 + \tfrac{1}{9}/106 + 1/28} \doteq 0.2355$$

(c) We test H_0: $\psi_i = 0$ vs. H_a: $\psi_i \neq 0$ for each contrast. The t- and P-values are given in the table.

The Placebo mean is significantly higher than the average of the other three, while the Keto mean is significantly lower than the average of the two Pyr means. The difference between the Pyr means is not significant (meaning the second application of the shampoo is of little benefit)—this agrees with our conclusion from Exercise 12.56.

12.58. **(a)** At right. **(b)** Each new value (except for n) is simply:

$$(\text{old value})/64 \times 100\%$$

Condition	n	$\bar{x}$	s	$s_{\bar{x}}$
Immediate	2	76.102%	2.3975%	1.6953%
One day	2	65.555%	3.3256%	2.3516%
Three days	2	34.055%	1.2043%	0.8516%
Five days	2	19.398%	1.6904%	1.1953%
Seven days	2	13.000%	0.4198%	0.2969%

(c) The SS and MS entries differ from those of Exercise 12.29—by a factor of $(100/64)^2$. However, everything else is the same: $F = 367.74$ with df 4 and 5; $P < 0.0005$, so we (again) reject H_0 and conclude that vitamin C content decreases over time.

Minitab output

```
Analysis of Variance on VitCPct
Source     DF        SS         MS        F        p
Days        4   6263.97    1565.99   367.74    0.000
Error       5     21.29       4.26
Total       9   6285.26
```

12.59. Transformed values for vitamin A are at right; each value is:

$$(\text{old value})/5 \times 100\%$$

(or equivalently, multiply by 20%). The transformation has no effect on vitamin E since the number of milligrams remaining is also the percentage of the original 100 mg.

Condition	n	$\bar{x}$	s	$s_{\bar{x}}$
Immediate	2	67.0%	0.2828%	0.2%
One day	2	64.8%	1.1314%	0.8%
Three days	2	64.2%	1.4142%	1.0%
Five days	2	66.1%	1.5556%	1.1%
Seven days	2	59.3%	1.2728%	0.9%

For vitamin A, the SS and MS entries differ from those of Exercise 12.31—by a factor of $(100/5)^2 = 400$. Everything else is the same: $F = 12.09$ with df 4 and 5; $P = 0.009$, so we (again) reject H_0 and conclude that vitamin A content decreases over time.

Since the vitamin E numbers are unchanged, the ANOVA table is unchanged, and we again fail to reject H_0 ($F = 0.69$ with df 4 and 5; $P = 0.630$).

In summary, transforming to percents (or doing any linear transformation) has no effect on the results of the ANOVA.

Minitab output

```
Analysis of Variance on VitAPct
Source     DF        SS        MS        F        p
Days        4     71.58     17.89    12.09    0.009
Error       5      7.40      1.48
Total       9     78.98
```

12.60. There is no effect on the test statistic, df, P-value, and conclusion. The degrees of freedom are not affected, since the number of groups and sample sizes are unchanged; meanwhile, the SS and MS values change (by a factor of b^2), but this change does not affect F since the factors of b^2 cancel out in the ratio $F = \text{MSG}/\text{MSE}$. With the same F- and df values, the P-value and conclusion are necessarily unchanged.

Proof of these statements is not too difficult, but it requires careful use of the SS formulas. For most students, a demonstration with several choices of a and b would probably be more convincing than a proof. However, here is the basic idea: Using results of Chapter 1, we know that the means undergo the same transformation as the data ($\bar{x}_i^* = a + b\bar{x}_i$), while the standard deviations are changed by a factor of $|b|$. Let $\bar{x}$ be the average of *all* the data; note that $\bar{x}^* = a + b\bar{x}$. Now $\text{SSG} = \sum_{i=1}^{I} n_i(\bar{x}_i - \bar{x})^2$, so:

$$\text{SSG}^* = \textstyle\sum_i n_i(\bar{x}_i^* - \bar{x}^*)^2 = \sum_i n_i(b\,\bar{x}_i - b\,\bar{x})^2 = \sum_i n_i\, b^2(\bar{x}_i - \bar{x})^2 = b^2\text{SSG}$$

Similarly, we can establish that $\text{SSE}^* = b^2\text{SSE}$ and $\text{SST}^* = b^2\text{SST}$. Since the MS values are merely SS values divided by the (unchanged) degrees of freedom, these also change by a factor of b^2.

12.61. A table of means and standard deviations is given on the next page. Quantile plots are not shown, but apart from the granularity of the scores and a few possible outliers, there are no marked deviations from Normality. Pooling is reasonable for both PRE1 and PRE2; the ratios are 1.24 and 1.48.

For both PRE1 and PRE2, we test H_0: $\mu_B = \mu_D = \mu_S$ vs. H_a: at least one mean is different. Both tests have df 2 and 63. For PRE1, $F = 1.13$ and $P = 0.329$; for PRE2, $F = 0.11$ and $P = 0.895$. There is no reason to believe that the mean pretest scores differ between methods.

		PRE1		PRE2	
Method	n	$\bar{x}$	s	$\bar{x}$	s
Basal	22	10.5	2.9721	$5.\overline{27}$	2.7634
DRTA	22	$9.\overline{72}$	2.6936	$5.\overline{09}$	1.9978
Strat	22	$9.1\overline{36}$	3.3423	$4.9\overline{54}$	1.8639

Minitab output

```
Analysis of Variance on pre1
Source     DF        SS        MS        F        p
Group       2     20.58     10.29     1.13    0.329
Error      63    572.45      9.09
Total      65    593.03

Analysis of Variance on pre2
Source     DF        SS        MS        F        p
Group       2      1.12      0.56     0.11    0.895
Error      63    317.14      5.03
Total      65    318.26
```

12.62. Some of the distributions have mild outliers or skewness, but there are no serious violations of Normality evident. Pooling is appropriate for all three response variables.

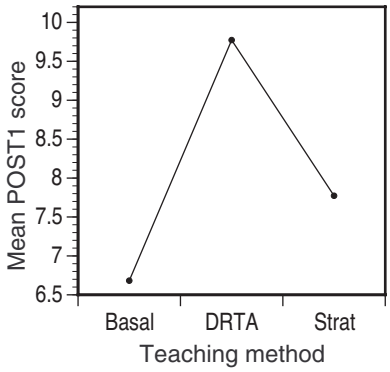
Variable	s_p	SE_D	$t^{**}\text{SE}_D$
POST1	3.1885	0.9614	2.3646
POST2	2.3785	0.7171	1.7639
POST3	6.3141	1.9038	4.6825

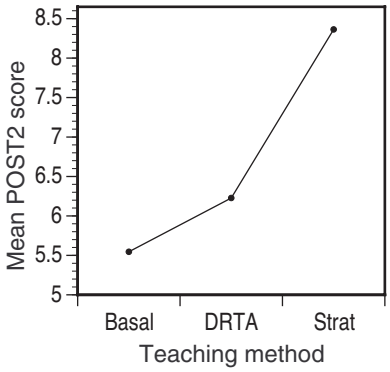
The three F statistics, all with df 2 and 63, are 5.32 ($P = 0.007$), 8.41 ($P = 0.001$), and 4.48 ($P = 0.015$). We conclude that at least one mean is different for each posttest.

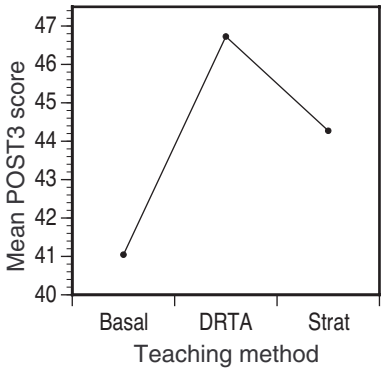
For multiple comparisons, we have three comparisons, df 63, and $\alpha = 0.05$, so the Bonferroni critical value is $t^{**} = 2.4596$. In the table on the right are the pooled standard deviations, standard error of each difference, and values of $t^{**}\text{SE}_D$ (the "minimum significant difference," or MSD) for each response variable. For both POST1 and POST3, DRTA is significantly greater than Basal, but no other comparisons are significant. For POST2, Strat is significantly greater than both Basal and DRTA.

We also may examine the contrasts $\psi_1 = -\mu_\text{B} + \frac{1}{2}(\mu_\text{D} + \mu_\text{S})$, which is positive if the average of the new methods is greater than the basal mean, and $\psi_2 = \mu_\text{D} - \mu_\text{S}$, which compares the two new methods. Estimated contrasts, standard errors, t, and P are given in the table below. (The P-values are one-sided for ψ_1 and two-sided for ψ_2.) We see that c_1 is significantly positive for all three variables. The second contrast was included in our multiple comparisons, where we found the difference significant only for POST2, but this time, when we are testing *only* this difference, rather than all three possible differences, we conclude that $\mu_\text{D} > \mu_\text{S}$ for POST1, in addition to the difference for POST2.

Variable	c_1	SE_{c_1}	t_1	P_1	c_2	SE_{c_2}	t_2	P_2
POST1	2.0909	0.8326	2.51	0.0073	2.0000	0.9614	2.08	0.0415
POST2	1.7500	0.6211	2.82	0.0032	−2.1364	0.7171	−2.98	0.0041
POST3	4.4545	1.6487	2.70	0.0044	2.4545	1.9038	1.29	0.2019







Test	Method	$\bar{x}$	s
POST1	Basal	6.6818	2.7669
	DRTA	9.7727	2.7243
	Strat	7.7727	3.9271
POST2	Basal	5.5455	2.0407
	DRTA	6.2273	2.0915
	Strat	8.3636	2.9040
POST3	Basal	41.0455	5.6356
	DRTA	46.7273	7.3884
	Strat	44.2727	5.7668

Basal/POST1	
1	
2	0
3	0
4	000
5	00000
6	0
7	00
8	000
9	000
10	0
11	
12	00

DRTA/POST1	
1	
2	
3	
4	
5	00
6	
7	000
8	0000
9	
10	0000
11	00
12	000
13	00
14	00

Strat/POST1	
1	0
2	
3	0
4	0000
5	00
6	0
7	000
8	0
9	00
10	
11	00
12	00
13	0
14	0
15	0

Basal/POST2	
0	
1	
2	
3	000
4	00000
5	000000
6	0
7	00
8	000
9	0
10	0

DRTA/POST2	
0	0
1	
2	
3	0
4	
5	00
6	0000000000
7	0000
8	00
9	0
10	
11	0

Strat/POST2	
0	
1	0
2	
3	
4	0
5	00
6	0
7	00
8	000
9	0000
10	000
11	00
12	00
13	0

Basal/POST3	
3	
3	223
3	5
3	66
3	99
4	0011
4	23
4	4555
4	66
4	9
5	
5	
5	4

DRTA/POST3	
3	01
3	
3	7
3	
4	01
4	23
4	
4	7
4	8889999
5	0
5	33
5	455
5	7

Strat/POST3	
3	
3	33
3	4
3	
3	8
4	1
4	2223
4	455
4	7
4	888999
5	01
5	3

Minitab output

```
Analysis of Variance on POST1
Source     DF        SS        MS        F        p
Method      2     108.1      54.1     5.32    0.007
Error      63     640.5      10.2
Total      65     748.6

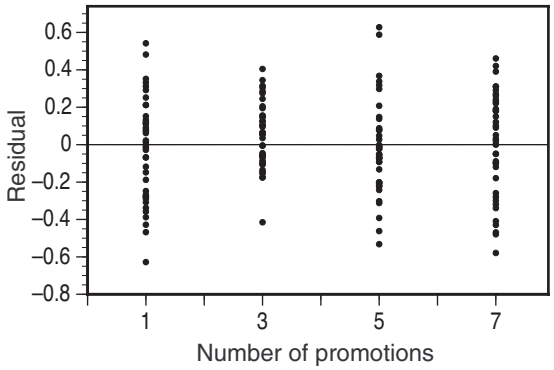
Analysis of Variance on POST2
Source     DF        SS        MS        F        p
Method      2     95.12     47.56     8.41    0.001
Error      63    356.41      5.66
Total      65    451.53

Analysis of Variance on POST3
Source     DF        SS        MS        F        p
Method      2     357.3     178.7     4.48    0.015
Error      63    2511.7      39.9
Total      65    2869.0
```

12.63. The regression equation is:

$$\hat{y} = 4.3645 - 0.1165x$$

($\hat{y}$ is the expected price and x is the number of promotions). The regression is significant (i.e., the slope is significantly different from 0): $t = -13.31$ with df $= 158$, giving $P < 0.0005$. The regression on number of promotions explains $r^2 = 52.9\%$ of the variation in expected price. (This is similar to the ANOVA value: $R^2 = 53.5\%$.)



The granularity of the "Number of promotions" observations makes interpreting the residual plot a bit tricky. For five promotions, the residuals seem to be more likely to be negative (in fact, 26 of the 40 residuals are negative), while for 3 promotions, the residuals are weighted toward the positive side. (We also observe that in the plot of mean expected price vs. number of promotions—see the solution to Exercise 12.40—the mean for 3 promotions is not as small as one would predict from a line near the other three points.) This suggests that a linear model may not be appropriate.

Minitab output

```
The regression equation is ExpPrice = 4.36 - 0.116 NumPromo

Predictor       Coef       Stdev    t-ratio        p
Constant     4.36452     0.04009     108.87    0.000
NumPromo   -0.116475    0.008748     -13.31    0.000

s = 0.2474      R-sq = 52.9%     R-sq(adj) = 52.6%

Analysis of Variance

SOURCE       DF          SS          MS         F        p
Regression    1      10.853      10.853    177.26    0.000
Error       158       9.674       0.061
Total       159      20.527
```

12.64. The pooled standard deviation s_p is found by looking at the spread of each observation about its *group* mean $\bar{x}_i$. The "total" standard deviation s given in Exercise 12.26 is the spread about the grand mean (the mean of all the data values, ignoring distinctions between groups). When we ignore group differences, we have more variation (uncertainty) in our data, so s is *almost always* larger than s_p.

This can be made clearer (to sufficiently mathematical students) by noting that the total variance s^2 can be found in the ANOVA table: Just as $s_p^2 = \dfrac{\text{SSE}}{\text{DFE}} = \text{MSE}$, $s^2 = \dfrac{\text{SST}}{\text{DFT}} = \text{MST}$. (The total mean square is not included in the ANOVA table but is easily computed from the values on the bottom line.) Because SSM + SSE = SST, we always have SSE $\leq$ SST, with equality only when the model is completely worthless (that is, when all group means equal the grand mean). Because DFE $<$ DFT, it might be that MSE $\geq$ MST but that does not happen very often.

12.66. With $\sigma = 7$ and means $\mu_1 = 40$, $\mu_2 = 47$, and $\mu_3 = 43$, we have $\bar{\mu} = \frac{40+47+43}{3} = 43.\overline{3}$ and noncentrality parameter:

$$\lambda = \frac{n\sum(\mu_i - \bar{\mu})^2}{\sigma^2} = \frac{(10)\left[(40 - 43.\overline{3})^2 + (47 - 43.\overline{3})^2 + (43 - 43.\overline{3})^2\right]}{49} = \frac{(10)(24.\overline{6})}{49} \doteq 5.0340$$

(The value of λ in the G•Power output below is slightly different due to rounding.) The degrees of freedom and critical value are the same as in Example 12.27: df 2 and 27, $F^* = 3.35$. Software reports the power as about 46%. Samples of size 10 are not adequate for this alternative; we should increase the sample size so that we have a better chance of detecting it. (For example, samples of size 20 give nearly 80% power for this alternative.)

G•Power output

```
Post-hoc analysis for "F-Test (ANOVA)", Global, Groups: 3:
Alpha: 0.0500
Power (1-beta): 0.4606
Effect size "f": 0.4096
Total sample size: 30
Critical value: F(2,27) = 3.3541
Lambda: 5.0332
```

12.67. (a) Sampling plans will vary but should attempt to address how cultural groups will be determined: Can we obtain such demographic information from the school administration? Do we simply select a large sample then poll each student to determine if he or she belongs to one of these groups? **(b)** Answers will vary with choice of H_a and desired power. For example, with the alternative $\mu_1 = \mu_2 = 4.4$, $\mu_3 = 5$, and standard deviation $\sigma = 1.2$, three samples of size 75 will produce power 0.89. (See G•Power output below.) **(c)** The report should make an attempt to explain the statistical issues involved; specifically, it should convey that sample sizes are sufficient to detect anticipated differences among the groups.

G•Power output

```
Post-hoc analysis for "F-Test (ANOVA)", Global, Groups: 3:
Alpha: 0.0500
Power (1-beta): 0.8920
Effect size "f": 0.2357
Total sample size: 225
Critical value: F(2,222) = 3.0365
Lambda: 12.4998
```

12.68. Recommended sample sizes will vary with choice of H_a and desired power. For example, with the alternative $\mu_1 = \mu_2 = 0.22$, $\mu_3 = 0.24$, and standard deviation $\sigma = 0.015$, three samples of size 10 will produce power 0.84, and samples of size 15 increase the power to 0.96. (See G•Power output below.) The report should make an attempt to explain the statistical issues involved; specifically, it should convey that sample sizes are sufficient to detect anticipated differences among the groups.

G•Power output

```
Post-hoc analysis for "F-Test (ANOVA)", Global, Groups: 3:
Alpha: 0.0500
Power (1-beta): 0.8379
Effect size "f": 0.6285
Total sample size: 30
Critical value: F(2,27) = 3.3541
Lambda: 11.8504
Note: Accuracy mode calculation.

Post-hoc analysis for "F-Test (ANOVA)", Global, Groups: 3:
Alpha: 0.0500
Power (1-beta): 0.9622
Effect size "f": 0.6285
Total sample size: 45
Critical value: F(2,42) = 3.2199
Lambda: 17.7756
```

12.69. The design can be similar, although the types of music might be different. Bear in mind that spending at a casual restaurant will likely be less than at the restaurants examined in Exercise 12.24; this might also mean that the standard deviations could be smaller. A pilot study might be necessary to get an idea of the size of the standard deviations. Decide how big a difference in mean spending you would want to detect, then do some power computations.