CHAPTER 11

Section 11.1

1.

a. The null hypothesis H_0 says that the true means are the same for all five brands, and the alternative hypothesis says that they are not all the same. H_0 will be rejected if $f \ge F_{.05,4,30} = 2.69$ since $I - 1 = 4$, and $IJ - I = 35 - 5 = 30$. The computed value of F is $f = \frac{\text{MSTr}}{\text{MSE}} = \frac{123.5}{22.16} = 5.57$. Since $5.57 \ge 2.69$, H_0 is rejected. The data indicates a difference in the population mean driving distances of the different brands of golf balls.

b. Because $F_{.01,4,30} = 4.02$ and $F_{.001,4,30} = 6.12$, and our computed value of 5.57 is between those values, it can be said that $.001 < P\text{-value} < .01$.

3. With μ_i = true average elastic modulus of apple pieces when using the ith freezing method, we wish to test $H_0: \mu_1 = \mu_2 = \mu_3$ v. H_a: at least two μ_i's are different. The grand mean is $\bar{x}_{..} = (61 + 73 + 49)/3 = 61$, so SSTr = $8[(61-61)^2 + (73-61)^2 + (49-61)^2] = 2304$, MSTr = 2304/(3 – 1) = 1152, SSE = $(8-1)[10^2 + 20^2 + 10^2] = 4200$, MSE = 4200/(24 – 3) = 200, and $f = \frac{1152}{200} = 5.76$. The F critical value is $F_{.05,2,21} = 3.47$. Since $5.76 \ge 3.47$, we reject H_0 at the .05 level and conclude that there are indeed differences in the mean elastic modulus of apple pieces using the three freezing methods.

5.

a. The grand mean is $\bar{x}_{..} = (141 + 144 + \ldots + 129)/10 = 139.2$, from which the treatment sum of squares is SSTr = $4[(141-139.2)^2 + \cdots + (129-139.2)^2] = 4[2495.6] = 9982.4$. Then, MSTr = SSTr/($I$ – 1) = 9982.4/(10 – 1) = 1109.16.

b. The ANOVA table appears below. Since $8.53 \ge F_{.01,9,30} = 3.07$, we reject H_0 at the .01 level. We conclude that not all 10 bike helmet brands have the same mean PLA at these settings.

Source	**df**	**SS**	**MS**	***f***
Brand	9	9982.4	1109.16	8.53
Error	30	3900.0	130.00	
Total	39			

7. With μ_i = true average compression strength for the ith box type, we wish to test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ versus H_a: at least two μ_i's are different. Software provides the ANOVA table below. With a P-value of .000 to three decimal places, we clearly reject H_0 and conclude that not all four box types have the same mean compression strength. (This is consistent with the observation that Type 4 seems very different from the other three with respect to compression strength.)

Analysis of Variance

Source	DF	SS	MS	F-Value	P-Value
Type	3	127375	42458	25.09	0.000
Error	20	33839	1692		
Total	23	161214			

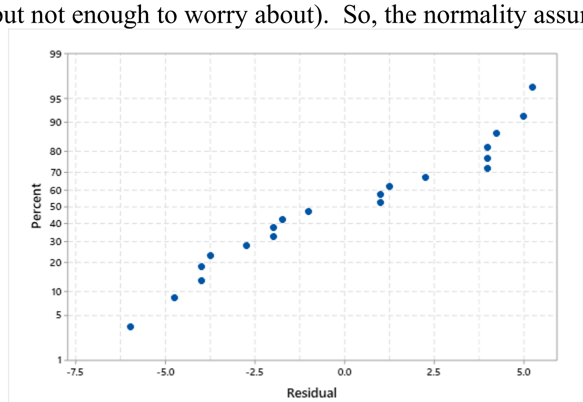
9.

a. The grand mean is $\bar{x}_{..} = (30 + 29 + 26 + 27)/4 = 28$, so SSTr = $27[(30-28)^2 + \cdots + (27-28)^2] = 270$. MSTr = SSTr/($I$ – 1) = 270/(4 – 1) = 90. SSE = $(27-1)[14^2 + 15^2 + 13^2 + 9^2]$ = 17,446, and finally MSE = SSE/(IJ – I) = 17,446/(108 – 4) = 167.75.

b. With μ_i = true mean back extension after 10 days using the ith treatment, we wish to test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ versus H_a: at least two μ_i's are different. The test statistic value is f = MSTR/MSE = 90/167.75 = 0.54 < $F_{.05,3,104} \approx 2.69$, so H_0 is not rejected at the .05 level. The data do not suggest that true mean back extension differs by treatment. (This is consistent with the observed sample means all being so close together.)

11.

a. The five group means are 202.75, 228.00, 142.00, 170.00, 209.75. Residuals for each observation are obtained by subtracting the group mean from the observed time to full charge. The accompanying normal probability plot of the 20 residuals is reasonably linear (there's a slight "pinch" toward the right-hand side, but not enough to worry about). So, the normality assumption is plausible.



Informally, equal variance is also plausible, since the ratio of the largest and smallest sample standard deviations is 4.90/3.37 < 2. Formally, with the aid of software we performed ANOVA on the absolute values of the residuals. The result was f = 0.42 and P-value = .792, suggesting there's no reason to doubt the equal variance assumption.

b. With μ_i = true mean time to full charge using the ith module, test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ versus H_a: at least two μ_i's are different. From software, f = [18797.5/4] / [251.5/15] = 4699.38/16.77 = 280.28. The enormous test statistic leaves no doubt: 280.28 ≥ $F_{.01,4,15}$ = 4.89, so H_0 is resoundingly rejected. The data provide clear evidence that these five modules do *not* require the same average amount of time to fully charge a 4.2-V battery.

13. Following the instructions provided,

$$\text{SST} = \sum_i \sum_j (x_{ij} - \bar{x}_{..})^2 = \sum_i \sum_j [(x_{ij} - \bar{x}_{i.}) + (\bar{x}_{i.} - \bar{x}_{..})]^2$$
$$= \sum_i \sum_j [(x_{ij} - \bar{x}_{i.})^2 + 2(x_{ij} - \bar{x}_{i.})(\bar{x}_{i.} - \bar{x}_{..}) + (\bar{x}_{i.} - \bar{x}_{..})^2] = \text{SSE} + 2\sum_i \sum_j (x_{ij} - \bar{x}_{i.})(\bar{x}_{i.} - \bar{x}_{..}) + \text{SSTr}$$

It remains to show the middle term is 0. Continuing with the middle sum,

$$\sum_i \sum_j (x_{ij} - \bar{x}_{i.})(\bar{x}_{i.} - \bar{x}_{..}) = \sum_i \left[(\bar{x}_{i.} - \bar{x}_{..}) \sum_j (x_{ij} - \bar{x}_{i.})\right] = \sum_i [(\bar{x}_{i.} - \bar{x}_{..}) \cdot 0] \qquad \text{from the hint}$$
$$= \sum_i 0 = 0, \text{ as desired.}$$

Section 11.2

15. $Q_{.05,5,15} = 4.37$, and $4.37\sqrt{272.8/4} = 36.09$. The brands seem to divide into two groups: 1, 3, and 4; and 2 and 5; with no significant differences within each group but all between group differences are significant.

3	1	4	2	5
437.5	462.0	469.3	512.8	532.1

17. Brand 1 does not differ significantly from 3 or 4, 2 does not differ significantly from 4 or 5, 3 does not differ significantly from1, 4 does not differ significantly from 1 or 2, 5 does not differ significantly from 2, but all other differences (e.g., 1 with 2 and 5, 2 with 3, etc.) do appear to be significant.

3	1	4	2	5
427.5	462.0	469.3	502.8	532.1

19. $Q_{.05,4,20}$ = 3.958 and MSE = 1692, from which $3.958\sqrt{1692/6} = 66.47$. The means are displayed below: mean compression strength for type 4 is significantly different from all others, but no other difference is significant (e.g., 756.93 – 698.07 = 58.86 < 66.47). This is consistent with the parallel boxplots of the data, which show that type 4 boxes have much lower strength than the other three types, but types 1-3 appear fairly comparable.

4	3	1	2
562.02	698.07	713.00	756.93

21. With $c_1 = c_2 = \frac{1}{2}$ and $c_3 = c_4 = c_5 = -\frac{1}{3}$, $\sum c_i^2 = \frac{5}{6} \approx .8333$. The point estimate for θ is $\hat{\theta} = \frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \frac{1}{3}(\bar{x}_3 + \bar{x}_4 + \bar{x}_5) = \frac{1}{2}(95.08 + 87.53) - \frac{1}{3}(76.85 + 82.50 + 89.08) = 8.495$. The t critical value is $t_{.025,15}$ = 2.131. Thus, a 95% CI for θ is $8.495 \pm 2.131\sqrt{6.949 \cdot (.8333)/4}$ = 8.495 ± 2.094 = (6.401, 10.589).

23.

a. MSE = SSE/(40 – 10) = 130 and $Q_{.05,10,30} = 4.824$, from which $4.824\sqrt{130/4} = 27.50$. The results of Tukey's method are displayed below (similar information appears in the original article).

SOO	BSP	CW	SWE	BMIPS	ST	BSF	N	GMIPS	GS
117	122	127	129	141	142	144	147	148	175

b. For the given contrast, $\sum c_i^2 = 7\cdot(\frac{1}{7})^2 + 3\cdot(-\frac{1}{3})^2 = \frac{10}{21} = .4762$. Also, $t_{.025,30} = 2.042$. The estimated contrast is $\frac{1}{7}(144+122+\cdots+129) - \frac{1}{3}(141+148+147) = -8.76$. Thus, a 95% CI for the contrast is $-8.76 \pm 2.042\sqrt{130\cdot(.4762)/4} = -8.76 \pm 8.03 = (-16.79, -0.73)$.

25. MSTr = 140, error df = 12, so $f = \dfrac{140}{\text{SSE}/12} = \dfrac{1680}{\text{SSE}}$ and $F_{.05,2,12} = 3.89$.

$Q_{.05,3,12}\sqrt{\text{MSE}/J} = 3.77\sqrt{\text{SSE}/60} = .4867\sqrt{\text{SSE}}$. Thus we wish $\dfrac{1680}{\text{SSE}} > 3.89$ (significant f) and $.4867\sqrt{\text{SSE}} > 10$ (= 20 – 10, the difference between the extreme $\bar{x}_{i.}$'s, so no significant differences are identified). These become 431.88 > SSE and SSE > 422.16, so SSE = 425 will work.

Section 11.3

27. Let μ_i = true mean tomato yield (kg/plot) at the ith salinity level. We test $H_0: \mu_1 = \cdots = \mu_4$ versus H_a: not all μ's are equal. Summary quantities are $\bar{x}_{..} = 52.44$, $\bar{x}_{i.} = 58.28, 55.40, 50.85, 45.50$ for i = 1, 2, 3, 4, and s_i = 3.60, 2.66, 2.43, 2.90 for i = 1, 2, 3, 4. From these,
SSTr = $5(58.28-52.44)^2 + 4(55.40-52.44)^2 + 4(50.85-52.44)^2 + 5(45.50-52.44)^2 = 465.5$ and
SSE = $(5-1)3.60^2 + (4-1)2.66^2 + (4-1)2.43^2 + (5-1)2.90^2 = 124.5$. dfTr = I – 1 = 3 and dfError = $n - I$ = 18 – 4 = 14, so MSTr = 465.5/3 = 152.17, MSE = 124.5/14 = 8.89, and f = 17.12.
Because $17.12 \ge F_{.05,3,14} = 3.34$, H_0 is rejected at level .05. There is a difference in true average yield of tomatoes for the four different levels of salinity.

29.

a. With such large sample sizes, population normality is not important — the Central Limit Theorem ensures that averages will be approximately normally distributed. Equal variance is still required, and the sample sd's affirm the plausibility of this assumption: the max-to-min ratio is 11.32/9.13 < 2.

b. Let μ_i = true mean accounting anxiety level for students at the ith grade level (1 = freshman, 5 = graduate). We test H_0: $\mu_1 = \cdots = \mu_5$ vs H_a: not all μ's are equal. We need the following:

$$\bar{x}_{..} = \frac{86(48.95)+224(51.45)+225(52.89)+198(52.92)+287(45.55)}{86+224+225+198+287} = 50.182$$

SSTr = $86(48.95-50.182)^2 + \cdots + 287(45.55-50.182)^2 = 9782.7$, MSTr = SSTR/(5 – 1) = 2445.7

SSE = $(86-1)9.13^2 + \cdots + (287-1)10.1^2$ = 118,632.6, MSE = SSE/(1020 – 5) = 116.9

Hence f = 2445.7/116.9 = 20.92. Since 20.92 ≥ $F_{.05,4,1015}$ = 2.38, H_0 is rejected. The data provide convincing statistical evidence that true mean accounting anxiety level differs by year in school.

c. $Q_{.05,5,1015} \approx 3.86$, so $d_{ij} = 3.86\sqrt{\dfrac{116.9}{2}\left(\dfrac{1}{J_i}+\dfrac{1}{J_j}\right)}$. The honestly significant differences are summarized below; the ith and jth means are significantly different iff $\left|\bar{x}_{i\cdot} - \bar{x}_{j\cdot}\right| > d_{ij}$.

Graduate	Freshman	Sophomore	Junior	Senior
45.55	48.95	51.45	52.89	52.92

31. Let μ_i = true mean impact of social media, as a percentage of sales, for the ith category (i = 1 for two or fewer platforms, i = 2 for three to five platforms, i = 3 for six or more). We test H_0: $\mu_1 = \mu_2 = \mu_3$ versus H_a: not all μ's are equal. We need the following:

$$\bar{x}_{\cdot\cdot} = \frac{107(12.76)+164(17.23)+90(21.56)}{107+164+90} = 16.985$$

SSTr = $107(12.76-16.985)^2+\cdots+90(21.56-16.985)^2$ = 3804, MSTr = SSTR/(3 – 1) = 1902

SSE = $(107-1)13.00^2+\cdots+(90-1)17.35^2$ = 76,973, MSE = SSE/(361 – 3) = 215

Hence f = 1902/215 = 8.85. Since 8.85 ≥ $F_{.01,2,358} \approx 4.66$, H_0 is rejected at the .01 significance level. The data provide convincing statistical evidence that an association exists between social media presence and mean sales impact.

33.

a. The distributions of the polyunsaturated fat percentages for each of the four regimens must be normal with equal variances. There's insufficient to check for normality, but the similarity of the sample sd's suggests equal population variances is plausible.

b. We have all the $\bar{x}_{i\cdot}$s , and we need the grand mean:

$$\bar{x}_{\cdot\cdot} = \frac{8(43.0)+13(42.4)+17(43.1)+14(43.5)}{52} = \frac{2236.9}{52} = 43.017;$$

$$\text{SSTr} = \sum J_i\left(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}\right)^2 = 8(43.0-43.017)^2+13(42.4-43.017)^2$$

$$+17(43.1-43.017)^2+13(43.5-43.017)^2 = 8.334 \text{ and MSTr} = \frac{8.334}{3} = 2.778$$

$$\text{SSE} = \sum(J_i-1)s_i^2 = 7(1.5)^2+12(1.3)^2+16(1.2)^2+13(1.2)^2 = 77.79 \text{ and MSE} = \frac{77.79}{48} = 1.621. \text{ Then}$$

$$f = \frac{\text{MSTr}}{\text{MSE}} = \frac{2.778}{1.621} = 1.714.$$

Since $1.714 < F_{.10,3,50} = 2.20$, we can say that the P-value is > .10. We do not reject the null hypothesis at significance level .10 (or any smaller level), so we conclude that the data suggests no difference in the percentages for the different regimens.

35. Let μ_i = true mean change in CMS under the ith treatment. We test H_0: $\mu_1 = \mu_2 = \mu_3$ versus H_a: not all μ's are equal. We need the following:

$$\bar{x}_{..} = \frac{32(5.1)+33(6.4)+34(6.5)}{99} = 6.01$$

SSTr = $32(5.1-6.01)^2 + \cdots + 34(6.5-6.01)^2$ = 39.68, MSTr = SSTR/(3 – 1) = 19.84

SSE = $(32-1)10.3^2 + \cdots + (34-1)18.1^2$ = 16,867.6, MSE = SSE/(99 – 3) = 175.70

Hence f = 19.84/175.70 = 0.1129, an extremely small F-value. In particular, since 0.1129 < $F_{.05,2,96} \approx 3.09$, H_0 is not rejected at the .05 level. The data do not suggest that true mean change in CMS differs by treatment.

37. From the previous exercise,

$$E(\text{SSTr}) = E\left(\sum_i J_i \bar{X}_{i\cdot}^2 - n\bar{X}_{..}^2\right) = \sum_i J_i E\left(\bar{X}_{i\cdot}^2\right) - nE\left(\bar{X}_{..}^2\right)$$
$$= \sum_i J_i \left(V(\bar{X}_{i\cdot}) + [E(\bar{X}_{i\cdot})]^2\right) - n\left(V(\bar{X}_{..}) + [E(\bar{X}_{..})]^2\right)$$
$$= \sum_i J_i \left(\frac{\sigma^2}{J_i} + \mu_i^2\right) - n\left(\frac{\sigma^2}{n} + \left[\frac{1}{n}\sum_i J_i E(\bar{X}_{i\cdot})\right]^2\right) = \sum_i J_i \left(\frac{\sigma^2}{J_i} + \mu_i^2\right) - n\left(\frac{\sigma^2}{n} + \left[\frac{\sum_i J_i \mu_i}{n}\right]^2\right)$$
$$= I\sigma^2 + \sum_i J_i(\mu + \alpha_i)^2 - \sigma^2 - \frac{1}{n}\left[\sum_i J_i(\mu + \alpha_i)\right]^2$$
$$= (I-1)\sigma^2 + \mu^2 \sum_i J_i + 2\sum_i J_i \alpha_i + \sum_i J_i \alpha_i^2 - \frac{1}{n}\left[\mu \sum_i J_i + \sum_i J_i \alpha_i\right]^2$$
$$= (I-1)\sigma^2 + \mu^2 n + 2(0) + \sum_i J_i \alpha_i^2 - \frac{1}{n}\left[\mu n + 0\right]^2 = (I-1)\sigma^2 + \sum_i J_i \alpha_i^2$$

from which $E(\text{MSTr}) = \dfrac{E(\text{SSTr})}{I-1} = \sigma^2 + \dfrac{1}{I-1}\sum_i J_i \alpha_i^2$.

When H_0 is true, all the α_i's are 0, and E(MSTr) = σ^2. Otherwise, E(MSTr) > σ^2.

39. $\alpha_1 = -\sigma$, $\alpha_2 = \alpha_3 = 0$, $\alpha_4 = \sigma$, so $\lambda = \dfrac{\sum J_i \alpha_i^2}{\sigma^2} = \dfrac{5(-\sigma)^2 + 4(0)^2 + 4(0)^2 + 5(\sigma)^2}{\sigma^2} = 10$. Also $\nu_1 = 3$, $\nu_2 = 14$, and $F_{.05,3,14}$ = 3.344. From R, β = `pf(3.344,df1=3,df2=14,ncp=10)` = .372, and so power = 1 – β = 1 – .372 = .628.

41.

a. The sample standard deviations are very different. In particular, 10.4/3.0 = 3.5 > 2, so our max-to-min convention suggests that the equal variance ANOVA assumption is not valid here.

b. Notice that the standard deviations are now closer together. While .41/.18 = 2.3 > 2, that's far less of a violation than in part **a**. From the given log-transformed information, $\bar{y}_{..}$ = 2.46, SSTr = 26.104, MSTr = SSTr/(5 – 1) = 6.526, SSE = 7.748, MSE = SSE/(89 – 5) = 0.0922, and f = MSTr/MSE = 70.752. With such a massive F-statistic, we clearly reject H_0; in particular, $f \geq F_{.05,4,84} \approx 2.48$. Therefore, at the .05 significance level, we conclude that the mean of ln(Hg Concentration) differs across these five Canadian reservoirs. Equivalently, the *median* mercury concentration is not the same at all five reservoirs.

c. Still using the log-transformed data for each pair of (transformed) means we compare $\left|\bar{x}_{i.} - \bar{x}_{j.}\right|$ to $d_{ij} = Q_{.05,5,84}\sqrt{\frac{\text{MSE}}{2}\left(\frac{1}{J_i}+\frac{1}{J_j}\right)} = 3.94\sqrt{\frac{0.0922}{2}\left(\frac{1}{J_i}+\frac{1}{J_j}\right)}$. The results of the 10 pairwise comparisons are summarized below.

PW	MK	NW	50K	FR
1.30	2.28	2.62	2.75	2.82

43. $g(x) = x\left(1-\frac{x}{n}\right) = nu\left(1-u\right)$ where $u = \frac{x}{n}$, so $h(x) = \int\left[u\left(1-u\right)\right]^{-1/2} du$. From a table of integrals, this gives $h(x) = \arcsin\left(\sqrt{u}\right) = \arcsin\left(\sqrt{\frac{x}{n}}\right)$ as the appropriate transformation.

Section 11.4

45.

a. $\text{MSA} = \frac{30.6}{4} = 7.65$, $\text{MSE} = \frac{59.2}{12} = 4.93$, $f_A = \frac{7.65}{4.93} = 1.55$. Since $1.55 < F_{.05,4,12} = 3.26$, don't reject H_{0A}. There is no significant difference in true average tire lifetime due to different makes of cars.

b. $\text{MSB} = \frac{44.1}{3} = 14.70$, $f_B = \frac{14.70}{4.93} = 2.98$. Since $2.98 < F_{.05,2,12} = 3.49$, don't reject H_{0B}. There is no significant difference in true average tire lifetime due to different brands of tires.

47.

a. With 17 participants (blocks), df(Blocks) = 17 – 1 = 16. Also, df(Method) = I – 1 = 6 – 1 = 5 and df(Total) = (17)(6) – 1 = 101. Using the fact that df and SS columns add, we can complete the ANOVA table. For each row, MS = SS/df.

Source	df	SS	MS	f
Method	5	596,748	119,349.6	9.67
Block	16	529,100	3306.9	0.27
Error	80	987,380	12342.3	
Total	101	2,113,228		

b. To test H_0: $\alpha_1 = \cdots = \alpha_5 = 0$ versus H_a: not all α's are 0, use the first f-value in the table. Since $9.67 \geq F_{.01,5,80} = 3.255$, we reject H_0 at the .01 level and conclude there is a statistically significant "method effect."

c. With $I = 6$, $J = 17$, MSE = 12342.3, and $Q_{.01,6,80} \approx 4.93$, the honestly significant difference for the methods is $4.93\sqrt{12342.3/17} = 132.8$. The resulting underscoring scheme appears below. Although the means appear far apart the large MSE implies that only the wrist and hip accelerators without LFE (the bottom two categories) *are* honestly significantly different from the hand tally.

Wrist acc.	Hip acc.	Pedometer	Wrist + LFE	Hip + LFE	Hand tally
449	466	557	579	606	668

49.

a. Using software or the by-hand formulas presented in this section yields the following ANOVA table.

Source	DF	SS	MS	F-Value	P-Value
Spindle speed	2	16106	8052.8	10.47	0.026
Feed rate	2	2156	1077.8	1.40	0.346
Error	4	3078	769.4		
Total	8	21339			

b. The test statistic and P-value for H_{0A}: $\alpha_1 = \alpha_2 = \alpha_3 = 0$ vs H_{aA}: not all α's = 0 are $f = 10.47$ and $P = .026$. Since $.026 \le .05$, we reject H_{0A} at the .05 level and conclude that mean temperature varies with spindle speed.

c. The test statistic and P-value for H_{0B}: $\beta_1 = \beta_2 = \beta_3 = 0$ vs H_{aB}: not all β's = 0 are $f = 1.40$ and $P = .346$. Since $.346 > .05$, we do not reject H_{0B} at the .05 level and conclude that feed rate has no statistically significant effect on mean temperature.

51.

a. Students' mouths may get warmer as the activity proceeds, because their closed mouths act like ovens. So, it's possible a student's third chip will melt faster simply because the student's mouth is hotter. Randomization balances this effect across the three flavors. (If everyone melted the chips in alphabetical order and white melted fastest, we wouldn't know if that was due to white chips' chemistry or simply due to increasing mouth temperature over time.)

b. Every person's mouth is different. In particular, a student who drank coffee/tea right before class might have an especially hot mouth, while someone who had soda might be cold. The effect that this person-to-person temperature variation has on melt times can be captured using blocking. Otherwise, it would be a source of unaccounted-for, extraneous variation, increasing the error and making it more difficult to detect differences due to chip flavor.

c. From the sample means, $\bar{x}_{..} = (88.15 + 60.49 + 72.35)/3 = 73.66$ and SSA = $54[(88.15-73.66)^2 + (60.49-73.66)^2 + (72.35-73.66)^2] = 20{,}797$. Also, dfA = $I - 1 = 2$, dfB = $J - 1 = 53$, and dfError = (2)(53) = 106.

Source	df	SS	MS	f
Flavor	2	20,797	10,398.5	35.0
Block (Subject)	53	135,833	2,562.9	8.6
Error	106	31,506	297.2	
Total	161	188,136		

With $f = 35.0 \geq F_{.01,2,106} \approx 4.81$, H_{0A}: $\alpha_1 = \alpha_2 = \alpha_3 = 0$ is rejected at the .01 level. There is a statistically significant "flavor effect"; i.e., the mean melt times for these three chip flavors are not the same.

d. Yes: $f = 8.6$ is a large F-value, suggesting that blocking by subject allowed us to account for a significant amount of variation in melt times. (Otherwise, the block sum of squares would be absorbed into the error sum of squares, and the F-statistic for the flavor effect would be much smaller.)

53. Software provides the following ANOVA table.

Source	DF	SS	MS	F-Value	P-Value
Current	2	106.78	53.39	0.19	0.833
Voltage	2	56.05	28.03	0.10	0.907
Error	4	1115.75	278.94		
Total	8	1278.58			

According to the ANOVA table, neither factor has a statistically significant effect at the .10 level: both P-values are > .10.

55. Software provides the information in the following ANOVA table.

Source	df	SS	MS	f
Method	2	23.23	11.61	8.69
Batch	9	86.79	9.64	7.22
Error	18	24.04	1.34	
Total	29	134.07		

$F_{.01,2,18} = 6.01 < 8.69 < F_{.001,2,18} = 10.39$, so .001 < P-value < .01, which is statistically significant. At least two of the curing methods produce differing average compressive strengths. (With P-value < .001, there are differences between batches as well, but we are not as interested in that.)

Next, let's examine the three pairwise differences:

$Q_{.05,3,18} = 3.61$, so $d = 3.61\sqrt{1.34/10} = 1.32$.

Method A	Method B	Method C
29.49	31.31	31.40

Methods B and C produce strengths that are not significantly different, but Method A produces strengths that are different (less) than those of both B and C.

57. $\text{MSB} = \frac{113.5}{4} = 28.38$, $\text{MSE} = \frac{25.6}{8} = 3.20$, $f_B = 8.87$, $F_{.01,4,8} = 7.01$, and since $8.87 \geq 7.01$, we reject H_0 at the .01 level and conclude that $\sigma_B^2 > 0$.

59. Use the definition of $\overline{X}_{i.}$, the fact that $\overline{X}_{..} = \frac{1}{I}\sum_i \overline{X}_{i.}$, and linearity of expectation.

$$E\left(\overline{X}_{i.}\right) = \frac{1}{J}\sum_j E(X_{ij}) = \frac{1}{J}\sum_j\left(\mu+\alpha_i+\beta_j\right) = \frac{1}{J}\sum_j \mu + \frac{1}{J}\sum_j \alpha_i + \frac{1}{J}\sum_j \beta_j$$

$$= \frac{1}{J}J\mu + \frac{1}{J}J\alpha_i + \frac{1}{J}0 = \mu + \alpha_i$$

$$E\left(\overline{X}_{..}\right) = \frac{1}{I}\sum_i E\left(\overline{X}_{i.}\right) = \frac{1}{I}\sum_i\left(\mu+\alpha_i\right) = \frac{1}{I}\sum_i \mu + \frac{1}{I}\sum_i \alpha_i$$

$$= \frac{1}{I}I\mu + \frac{1}{I}0 = \mu$$

Therefore, $E\left(\overline{X}_{i.} - \overline{X}_{..}\right) = \mu + \alpha_i - \mu = \alpha_i$, as desired.

Section 11.5

61.

a.

Source	**df**	**SS**	**MS**	***f***
A	2	30,763.0	15,381.50	3.79
B	3	34,185.6	11,395.20	2.81
AB	6	43,581.2	7263.53	1.79
Error	24	97,436.8	4059.87	
Total	35	205,966.6		

b. $f_{AB} = 1.79 < F_{.05,6,24} = 2.51$, so H_{0AB} cannot be rejected, and we conclude that no statistically significant interaction is present. We many proceed to examining the main effects.

c. $f_A = 3.79 \geq F_{.05,2,24} = 3.40$, so H_{0A} is rejected at level .05.

d. $f_B = 2.81 < F_{.05,3,24} = 3.01$, so H_{0B} is not rejected.

e. $Q_{.05,3,24} = 3.53$, so $d = 3.53\sqrt{4059.87/12} = 64.93$.

3	1	2
3960.02	4010.88	4029.10

Only times 2 and 3 yield significantly different strengths.

63.

Source	**df**	**SS**	**MS**	***f***	***F*.05**	***F*.01**
Formulation	1	2,253.44	2,253.44	376.2	4.75	9.33
Speed	2	230.81	115.41	19.27	3.89	6.93
Form. x Speed	2	18.58	9.29	1.55	3.89	6.93
Error	12	71.87	5.99			
Total	17	2,574.7				

a. There appears to be no interaction between the two factors: $f = 1.55 < 3.89$.

b. Both formulation and speed appear to have a highly statistically significant effect on yield: both *f*-values far exceed the .01 critical value.

c. Let Factor A = formulation and Factor B = speed. Begin by estimating the μ's:

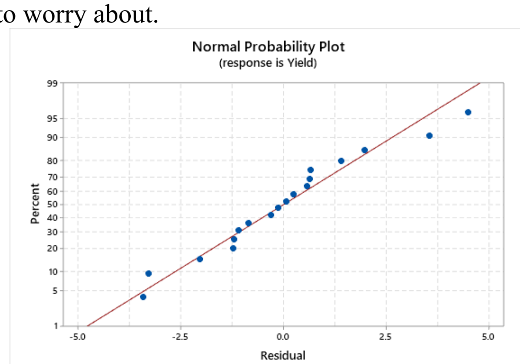
$\hat{\mu} = \bar{x}... = 175.84$; $\hat{\mu}_1. = \frac{1}{3}\sum_j \bar{x}_{1j}. = 187.03$ and $\hat{\mu}_2. = 164.66$; $\hat{\mu}._1 = \frac{1}{2}\sum_i \bar{x}_{i1}. = 177.83$, $\hat{\mu}._2 = 170.82$, and $\hat{\mu}._3 = 178.88$.

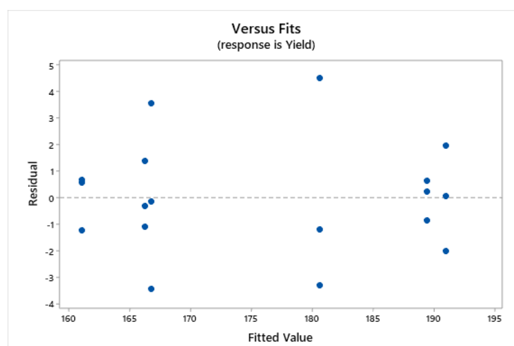
Since $\alpha_i = \mu_i. - \mu$, $\hat{\alpha}_1 = 187.03 - 175.84 = 11.19$ and $\hat{\alpha}_2 = 164.66 - 175.84 = -11.18$; these sum to 0 except for rounding error. Similarly, $\hat{\beta}_1 = \hat{\mu}._1 - \hat{\mu} = 177.83 - 175.84 = 1.99$, $\hat{\beta}_2 = -5.02$, and $\hat{\beta}_3 = 3.04$; these sum to 0 except for rounding error.

d. Using $\gamma_{ij} = \mu_{ij} - (\mu + \alpha_i + \beta_j)$ and techniques similar to above, we find the following estimates of the interaction effects: $\hat{\gamma}_{11} = .45$, $\hat{\gamma}_{12} = -1.41$, $\hat{\gamma}_{13} = .96$, $\hat{\gamma}_{21} = -.45$, $\hat{\gamma}_{22} = 1.39$, and $\hat{\gamma}_{23} = -.97$. Again, there are some minor rounding errors.

Observed	Fitted	Residual	Observed	Fitted	Residual
189.7	189.47	0.23	161.7	161.03	0.67
188.6	189.47	–0.87	159.8	161.03	–1.23
190.1	189.47	0.63	161.6	161.03	0.57
165.1	166.20	–1.1	189.0	191.03	–2.03
165.9	166.20	–0.3	193.0	191.03	1.97
167.6	166.20	1.4	191.1	191.03	0.07
185.1	180.60	4.5	163.3	166.73	–3.43
179.4	180.60	–1.2	166.6	166.73	–0.13
177.3	180.60	–3.3	170.3	166.73	3.57

e. and **f.** The requested plots appear below. The normal probability plot is fairly straight, suggesting that it's plausible the ε_{ijk}'s are normally distributed. The plot of residuals versus predicted values on the right shows some variation in spread at different treatment combinations, but probably not enough to worry about.

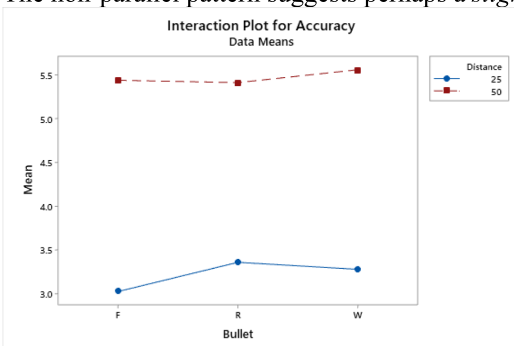




65.

a. One factor is firing distance, with levels 25 yards and 50 yards. The other factor is bullet brand, with levels Federal, Remington, and Winchester. Together, they make 6 treatment combinations: (25,Fed), (25,Rem), (25,Win), (50,Fed), (50,Rem), and (50,Win).

b. The interaction plot suggests a huge distance effect — not surprisingly, accuracy is better (lower values) when shooting at a closer distance. There appears to be very little bullet manufacturer effect. The non-parallel pattern suggests perhaps a *slight* interaction effect.



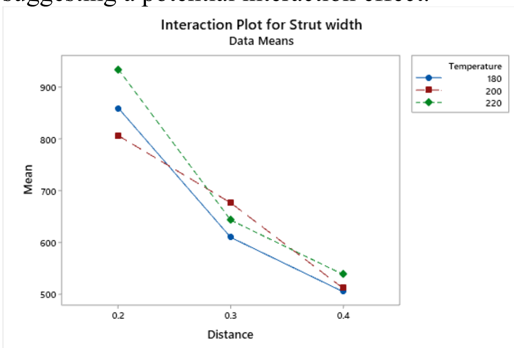
c. Software provided the accompanying ANOVA table (from which we extracted the sums of squares in the text). Consistent with the interaction plot, there is no significant interaction ($f = 0.53$, $P = .589$) and also no significant bullet manufacturer effect ($f = 0.63$, $P = .531$). There is, however, an extremely statistically significant distance effect ($f = 242.56$, $P = .000$).

Source	DF	SS	MS	F-Value	P-Value
Distance	1	568.97	568.969	242.56	0.000
Bullet	2	2.97	1.487	0.63	0.531
Distance*Bullet	2	2.48	1.242	0.53	0.589
Error	444	1041.49	2.346		
Total	449	1615.92			

67. This is a mixed effects model. In particular, the relevant null hypotheses are $H_{0A}: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$; $H_{0B}: \sigma_B^2 = 0$; $H_{0AB}: \sigma_G^2 = 0$. Software gives the ANOVA table below. Interaction between brand and writing surface has no significant effect on the lifetime of the pen. Since neither f_A nor f_B is greater than its respective critical value, we can conclude that neither the surface nor the brand of pen has a significant effect on the writing lifetime.

Source	df	SS	MS	f	$F_{.05}$
A	3	1,387.5	462.5	$\frac{MSA}{MSAB} = .34$	4.76
B	2	2,888.08	1,444.04	$\frac{MSB}{MSAB} = 1.07$	5.14
AB	6	8,100.25	1,350.04	$\frac{MSAB}{MSE} = 1.97$	3.00
Error	12	8,216.0	684.67		
Total	23	20,591.83			

69. Let's first examine an interaction plot. The plot shows a very strong distance effect, but perhaps not much of a temperature effect — the three lines overlap substantially. The "bends" of the three lines are different, suggesting a potential interaction effect.



With the aid of software, a two-way ANOVA with interaction was performed, resulting in the following ANOVA table.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Distance	2	562424	281212	360.70	0.000
Temperature	2	11757	5879	7.54	0.004
Distance*Temperature	4	21715	5429	6.96	0.001
Error	18	14033	780		
Total	26	609930			

The ANOVA table indicates a highly statistically significant interaction effect ($f = 6.96$, P-value = .001). In the presence of the interaction, the two main effects can't be directly interpreted. But the interaction by itself indicates that both nozzle-bed distance and temperature play a significant role in determining strut width. We may apply Tukey's method here to the nine (distance, temperature) pairs to identify honestly significant differences due to treatment combinations. The results from software appear below.

Grouping Information Using the Tukey Method and 95% Confidence

Distance*Temperature	N	Mean	Grouping				
0.2 220	3	935.000	A				
0.2 180	3	860.000	A	B			
0.2 200	3	806.667		B			
0.3 200	3	676.667			C		
0.3 220	3	643.333			C		
0.3 180	3	610.000			C	D	
0.4 220	3	538.333				D	E
0.4 200	3	511.667					E
0.4 180	3	505.000					E

Means that do not share a letter are significantly different.

71.

a. Using $\Sigma_i\alpha_i = 0, \Sigma_j\beta_j = 0, \Sigma_j\gamma_{ij} = 0$,

$$E\left(\bar{X}_{i..} - \bar{X}_{...}\right) = \frac{1}{JK}\sum_j\sum_k E\left(X_{ijk}\right) - \frac{1}{IJK}\sum_i\sum_j\sum_k E\left(X_{ijk}\right) = \frac{1}{JK}\sum_j\sum_k\left(\mu+\alpha_i+\beta_j+\gamma_{ij}\right) - \frac{1}{IJK}\sum_i\sum_j\sum_k\left(\mu+\alpha_i+\beta_j+\gamma_{ij}\right)$$

$$= \frac{1}{JK}\sum_j K\left(\mu+\alpha_i+\beta_j+\gamma_{ij}\right) - \frac{1}{IJK}\sum_i\sum_j K\left(\mu+\alpha_i+\beta_j+\gamma_{ij}\right)$$

$$= \frac{1}{J}\left(J\mu+J\alpha_i+0+0\right) - \frac{1}{IJ}\sum_i\left(J\mu+J\alpha_i+0+0\right) = \mu+\alpha_i - \frac{1}{I}\sum_i\left(\mu+\alpha_i\right)$$

$$= \mu+\alpha_i - \frac{1}{I}(I\mu+0) = \alpha_i$$

b. Similarly,

$$E\left(\hat{\gamma}_{ij}\right) = \frac{1}{K}\sum_k E\left(X_{ijk}\right) - \frac{1}{JK}\sum_j\sum_k E\left(X_{ijk}\right) - \frac{1}{IK}\sum_i\sum_k E\left(X_{ijk}\right) + \frac{1}{IJK}\sum_i\sum_j\sum_k E\left(X_{ijk}\right)$$

$$= \frac{1}{K}\sum_k\left(\mu+\alpha_i+\beta_j+\gamma_{ij}\right) - \left(\mu+\alpha_i\right) - (\mu+\beta_j) + \mu$$

$$= \mu+\alpha_i+\beta_j+\gamma_{ij} - \mu - \alpha_i - \mu - \beta_j + \mu = \gamma_{ij}$$

73.

a. $\dfrac{E(\text{MSAB})}{E(\text{MSE})} = \dfrac{\sigma^2 + K\sigma_G^2}{\sigma^2} = 1$ if $\sigma_G^2 = 0$ and > 1 if $\sigma_G^2 > 0$, so $\dfrac{\text{MSAB}}{\text{MSE}}$ is the appropriate F ratio.

b. $\dfrac{E(\text{MSA})}{E(\text{MSAB})} = \dfrac{\sigma^2 + K\sigma_G^2 + JK\sigma_A^2}{\sigma^2 + K\sigma_G^2} = 1$ if $\sigma_A^2 = 0$ and > 1 if $\sigma_A^2 > 0$, so $\dfrac{\text{MSA}}{\text{MSAB}}$ is the appropriate F ratio for H_{0A} versus H_{aA}. Similarly, $\dfrac{\text{MSB}}{\text{MSAB}}$ is the appropriate F ratio for H_{0B} versus H_{aB}.

Supplementary Exercises

75.

Source	**df**	**SS**	**MS**	***f***
Treatment	3	81.1944	27.0648	22.36
Block	8	66.5000	8.3125	6.87
Error	24	29.0556	1.2106	
Total	35	176.7500		

Since 22.36 > $F_{.05,3,24} = 3.01$, reject H_{0A}. There is an effect due to treatments. Next, $Q_{.05,4,24} = 3.90$, so Tukey's HSD is $3.90\sqrt{1.2106/9} = 1.43$.

1	4	3	2
8.56	9.22	10.78	12.44

77.

a. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ v. H_a: at least two of the μ_i's are different; $f = 3.68 < F_{.01,3,20} = 4.94$, thus fail to reject H_0. The means do not appear to differ.

b. We reject H_0 when the P-value $\le \alpha$. Since .029 > .01, we still fail to reject H_0.

79. Let μ_i = true mean exam score for all students using the ith lesson delivery mode (VR, AR, or 3D tablet). We test H_0: $\mu_1 = \mu_2 = \mu_3$ versus H_a: not all μ's are equal. From the available information,

$$\bar{x}_{..} = \frac{20(12.9)+17(12.5)+22(13.3)}{20+17+22} = \frac{763.1}{59} = 12.934$$

SSTr = $20(12.9-12.934)^2 + 17(12.5-12.934)^2 + 22(13.3-12.934)^2 = 6.172 \rightarrow$ MSTr = 6.172/2 = 3.086

SSE = $(20-1)4.3^2 + (17-1)4.5^2 + (22-1)4.2^2 = 1045.75 \rightarrow$ MSE = 1045.75/(59 – 3) = 18.674

Thus the observed test statistic value is $f = 3.086/18.674 = 0.165 < F_{.05,2,56} = 3.16$. With such a small f-value, the data provide absolutely no evidence in favor of H_a; the data do *not* suggest there's a difference in true mean exam score based on lesson delivery method.

81.

a. $I = 5$, $J = 6$, $\bar{x}_{..} = 2.448$, from which SSTr = 0.93 and SSE = SST – SSTr = 3.62 – 0.93 = 2.69.

Source	df	SS	MS	f
Treatments	4	0.93	.233	2.16
Error	25	2.69	.108	
Total	29	3.62		

Since $2.16 < F_{.05,4,25} = 2.76$, do not reject H_0 at level .05.

b. $\hat{\theta} = 2.58 - \frac{2.63+2.13+2.41+2.49}{4} = .165$, $t_{.025,25} = 2.060$, MSE = .108, and $\Sigma c_i^2 = (1)^2 + (-.25)^2 + (-.25)^2 + (-.25)^2 + (-.25)^2 = 1.25$, so a 95% confidence interval for θ is $.165 \pm 2.060\sqrt{(.108)(1.25)/6} = .165 \pm .309 = (-.144, .474)$. This interval does include zero, so 0 is a plausible value for θ.

c. $\mu_1 = \mu_2 = \mu_3, \mu_4 = \mu_5 = \mu_1 - \sigma$, so $\mu = \mu_1 - \frac{2}{5}\sigma$, $\alpha_1 = \alpha_2 = \alpha_3 = \frac{2}{5}\sigma$, $\alpha_4 = \alpha_5 = -\frac{3}{5}\sigma$. Then $\lambda = 6 \cdot \frac{3\left(\frac{2}{5}\sigma\right)^2 + 2\left(-\frac{3}{5}\sigma\right)^2}{\sigma^2} = 7.2$, $\nu_1 = 4$, $\nu_2 = 25$, and $F_{.05,4,25} = 2.79$. With the aid of R software, $\beta =$ `pf(2.79,df1=4,df2=25,ncp=7.2)` $\approx .54$.

83. Let μ_i = true mean increase in 6MWT distance (meters) for all COPD sufferers in the ith weight category.

a. We test H_0: $\mu_1 = \mu_2 = \mu_3$ versus H_a: not all μ's are equal. From the available information,

$$\bar{x}_{\cdot\cdot} = \frac{53(61)+39(67)+63(41)}{155} = 54.38$$

SSTr = $53(61-54.38)^2+39(67-54.38)^2+63(41-54.38)^2$ = 19812.6 → MSTr = 19812.6/2 = 9906.3

SSE = $(53-1)80^2+(39-1)86^2+(63-1)87^2$ = 1,083,126 → MSE = 1,083,126 /(155 – 3) = 7125.8

Thus f = 9906.3/7125.8 = 1.30 < $F_{.05,2,152}$ = 3.06, and H_0 is not rejected. The data does not provide convincing statistical evidence that mean increase in 6MWT distance varies by weight category. (Notice that although the mean increase for obese patients was much less than for the other two groups, there is an enormous amount of "noise" as seen in the sample standard deviations.)

b. $Q_{.05,3,152} \approx 3.347$ and $d_{ij} = 3.347\sqrt{\frac{7125.8}{2}\left(\frac{1}{J_i}+\frac{1}{J_j}\right)}$ for each pair. This gives d_{12} = 42.1, d_{13} = 37.2, and d_{23} = 40.7. None of the sample means are nearly this far apart, so Tukey's method provides no statistically significant differences. This is consistent with the results in part **a**.