Chapter 13: The Analysis of Variance

13.1 The summary statistics are: $\bar{y}_1$ = 1.875, s_1^2 = .6964286, $\bar{y}_2$ = 2.625, s_2^2 = .8392857, and $n_1 = n_2 = 8$. The desired test is: H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$, where μ_1, μ_2 represent the mean reaction times for Stimulus 1 and 2 respectively.

a. SST = $4(1.875 - 2.625)^2$ = 2.25, SSE = 7(.696428) + 7(.8392857) = 10.75. Thus, MST = 2.25/1 = 2.25 and MSE = 10.75/14 = .7679. The test statistic F = 2.25/.7679 = 2.93 with 1 numerator and 14 denominator degrees of freedom. Since $F_{.05}$ = 4.60, we fail to reject H_0: the stimuli are not significantly different.

b. Using the Applet, p–value = $P(F > 2.93)$ = .109.

c. Note that s_p^2 = MSE = .7679. So, the two–sample t–test statistic is $|t| = \frac{|1.875-2.625|}{\sqrt{.7679\left(\frac{2}{8}\right)}}$ = 1.712 with 14 degrees of freedom. Since $t_{.025}$ = 2.145, we fail to reject H_0. The two tests are equivalent, and since $F = T^2$, note that $2.93 \approx (1.712)^2$ (roundoff error).

d. We assumed that the two random samples were selected independently from normal populations with equal variances.

13.2 Refer to Ex. 10.77. The summary statistics are: $\bar{y}_1$ = 446, s_1^2 = 42, $\bar{y}_2$ = 534, s_2^2 = 45, and $n_1 = n_2 = 15$.

a. SST = $7.5(446 - 534)^2$ = 58,080, SSE = 14(42) + 14(45) = 1218. So, MST = 58,080 and MSE = 1218/28 = 1894.5. The test statistic F = 58,080/1894.5 = 30.64 with 1 numerator and 28 denominator degrees of freedom. Clearly, p–value < .005.

b. Using the Applet, p–value = $P(F > 30.64)$ = .00001.

c. In Ex. 10.77, t = –5.54. Observe that $(-5.54)^2 \approx 30.64$ (roundoff error).

d. We assumed that the two random samples were selected independently from normal populations with equal variances.

13.3 See Section 13.3 of the text.

13.4 For the four groups of students, the sample variances are: s_1^2 = 66.6667, s_2^2 = 50.6192, s_3^2 = 91.7667, s_4^2 = 33.5833 with $n_1 = 6$, $n_2 = 7$, $n_3 = 6$, $n_4 = 4$. Then, SSE = 5(66.6667) + 6(50.6192) + 5(91.7667) + 3(33.5833) = 1196.6321, which is identical to the prior result.

13.5 Since W has a chi–square distribution with r degrees of freedom, the mgf is given by

$$m_W(t) = E(e^{tW}) = (1-2t)^{-r/2}.$$

Now, $W = U + V$, where U and V are independent random variables and V is chi–square with s degrees of freedom. So,

$$m_W(t) = E(e^{tW}) = E(e^{t(U+V)}) = E(e^{tU})E(e^{tV}) = E(e^{tU})(1-2t)^{-s/2} = (1-2t)^{-r/2}.$$

Therefore, $m_U(t) = E(e^{tU}) = \frac{(1-2t)^{-r/2}}{(1-2t)^{-s/2}} = (1-2t)^{-(r-s)/2}$. Since this is the mgf for a chi–square random variable with $r - s$ degrees of freedom, where $r > s$, by the Uniqueness Property for mgfs U has this distribution.

13.6 **a.** Recall that by Theorem 7.3, $(n_i - 1)S_i^2/\sigma^2$ is chi–square with $n_i - 1$ degrees of freedom. Since the samples are independent, by Ex. 6.59, $\text{SSE}/\sigma^2 = \sum_{i=1}^{k}(n_i - 1)S_i^2/\sigma^2$ is chi–square with $n - k$ degrees of freedom.

b. If H_0 is true, all of the observations are identically distributed since it was already assumed that the samples were drawn independently from normal populations with common variance. Thus, under H_0, we can combine all of the samples to form an estimator for the common mean, $\overline{Y}$, and an estimator for the common variance, given by TSS/$(n - 1)$. By Theorem 7.3, TSS/σ^2 is chi–square with $n - 1$ degrees of freedom.

c. The result follows from Ex. 13.5: let $W = \text{TSS}/\sigma^2$ where $r = n - 1$ and let $V = \text{SSE}/\sigma^2$ where $s = n - k$. Now, SSE/σ^2 is distributed as chi–square with $n - k$ degrees of freedom and TSS/σ^2 is distributed as chi–square <u>under H_0</u>. Thus, $U = \text{SST}/\sigma^2$ is chi–square <u>under H_0</u> with $n - 1 - (n - k) = k - 1$ degrees of freedom.

d. Since SSE and TSS are independent, by Definition 7.3

$$F = \frac{\text{SST}\big/\left(\sigma^2(k-1)\right)}{\text{SSE}\big/\left(\sigma^2(n-k)\right)} = \frac{\text{MST}}{\text{MSE}}$$

has an F–distribution with $k - 1$ numerator and $n - k$ denominator degrees of freedom.

13.7 We will use R to solve this problem:

```
> waste <- c(1.65, 1.72, 1.5, 1.37, 1.6, 1.7, 1.85, 1.46, 2.05, 1.8,
1.4, 1.75, 1.38, 1.65, 1.55, 2.1, 1.95, 1.65, 1.88, 2)
> plant <- c(rep("A",5), rep("B",5), rep("C",5), rep("D",5))
> plant <- factor(plant)      # change plant to a factor variable
> summary(aov(waste~plant))
            Df  Sum Sq Mean Sq F value  Pr(>F)
plant        3 0.46489 0.15496  5.2002 0.01068 *
Residuals   16 0.47680 0.02980
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. The F statistic is given by F = MST/MSE = .15496/.0298 = 5.2002 (given in the ANOVA table above) with 3 numerator and 16 denominator degrees of freedom. Since $F_{.05} = 3.24$, we can reject H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ and conclude that at least one of the plant means are different.

b. The p–value is given in the ANOVA table: p–value = .01068.

13.8 Similar to Ex. 13.7, R will be used to solve the problem:

```
> salary <- c(49.3, 49.9, 48.5, 68.5, 54.0, 81.8, 71.2, 62.9, 69.0,
69.0, 66.9, 57.3, 57.7, 46.2, 52.2)
> type <- factor(c(rep("public",5), rep("private",5), rep("church",5)))
```

a. This is a completely randomized, one–way layout (this is sampled data, not a designed experiment).

b. To test H_0: $\mu_1 = \mu_2 = \mu_3$, the ANOVA table is given below (using R):

```
> summary(aov(salary~type))
            Df Sum Sq Mean Sq F value   Pr(>F)
type         2 834.98  417.49  7.1234 0.009133 **
Residuals   12 703.29   58.61
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the output, F = MST/MSE = 7.1234 with 3 numerator and 12 denominator degrees of freedom. From Table 7, .005 < p–value < .01.

c. From the output, p-value = .009133.

13.9 The test to be conducted is H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, where μ_i is the mean strength for the i^{th} mix of concrete, i = 1, 2, 3, 4. The alternative hypothesis at least one of the equalities does not hold.

a. The summary statistics are: TSS = .035, SST = .015, and so SSE = .035 – .015 = .020. The mean squares are MST = .015/3 = .005 and MSE = .020/8 = .0025, so the F statistic is given by F = .005/.0025 = 2.00, with 3 numerator and 8 denominator degrees of freedom. Since $F_{.05}$ = 4.07, we fail to reject H_0: there is not enough evidence to reject the claim that the concrete mixes have equal mean strengths.

b. Using the Applet, p–value = $P(F > 2)$ = .19266. The ANOVA table is below.

Source	d.f	SS	MS	F	p–value
Treatments	3	.015	.005	2.00	.19266
Error	8	.020	.0025		
Total	11	.035			

13.10 The test to be conducted is H_0: $\mu_1 = \mu_2 = \mu_3$, where μ_i is the mean score where the i^{th} method was applied, i = 1, 2, 3. The alternative hypothesis at least one of the equalities does not hold

a. The summary statistics are: TSS = 1140.5455, SST = 641.8788, and so SSE = 1140.5455 – 641.8788 = 498.6667. The mean squares are MST = 641.8788/2 = 320.939 and MSE = 498.6667/8 = 62.333, so the F statistic is given by F = 320.939/62.333 = 5.148, with 2 numerator and 8 denominator degrees of freedom. By Table 7, .025 < p–value < .05.

b. Using the Applet, p–value = $P(F > 5.148)$ = .03655. The ANOVA table is below.

Source	d.f	SS	MS	F	p–value
Treatments	2	641.8788	320.939	5.148	.03655
Error	8	498.6667	62.333		
Total	10	1140.5455			

c. With α = .05, we would reject H_0: at least one of the methods has a different mean score.

13.11 Since the three sample sizes are equal, $\bar{y} = \frac{1}{3}(\bar{y}_1 + \bar{y}_2 + \bar{y}_3) = \frac{1}{3}(.93 + 1.21 + .92) = 1.02$.

Thus, SST = $n_1\sum_{i=1}^{3}(\bar{y}_i - \bar{y})^2 = 14\sum_{i=1}^{3}(\bar{y}_i - 1.02)^2 = .7588$. Now, recall that the "standard error of the mean" is given by $s/\sqrt{n}$, so SSE can be found by

$$\text{SSE} = 13[14(.04)^2 + 14(.03)^2 + 14(.04)^2] = .7462.$$

Thus, the mean squares are MST = .7588/2 = .3794 and MSE = .7462/39 = .019133, so that the F statistic is F = .3794/.019133 = 19.83 with 2 numerator and 39 denominator degrees of freedom. From Table 7, it is seen that p–value < .005, so at the .05 significance level we reject the null hypothesis that the mean bone densities are equal.

13.12 The test to be conducted is H_0: $\mu_1 = \mu_2 = \mu_3$, where μ_i is the mean percentage of Carbon 14 where the i^{th} concentration of acetonitrile was applied, i = 1, 2, 3. The alternative hypothesis at least one of the equalities does not hold

a. The summary statistics are: TSS = 235.219, SST = 174.106, and so SSE = 235.219 – 174.106 = 61.113. The mean squares are MST = 174.106/2 = 87.053 and MSE = 235.219/33 = 1.852, so the F statistic is given by F = 87.053/1.852 = 47.007, with 2 numerator and 33 denominator degrees of freedom. Since $F_{.01} \approx 5.39$, we reject H0: at least one of the mean percentages is different and p–value < .005. The ANOVA table is below.

Source	d.f	SS	MS	F	p–value
Treatments	2	174.106	87.053	47.007	< .005
Error	33	61.113	1.852		
Total	35	235.219			

b. We must assume that the independent measurements from low, medium, and high concentrations of acetonitrile are normally distributed with common variance.

13.13 The grand mean is $\bar{y} = \frac{45(4.59)+102(4.88)+18(6.24)}{165} = 4.949$. So,

$$\text{SST} = 45(4.59 - 4.949)^2 + 102(4.88 - 4.949)^2 + 18(6.24 - 4.949)^2 = 36.286.$$

$$\text{SSE} = \sum_{i=1}^{3}(n-1)s_i^2 = 44(.70)^2 + 101(.64)^2 + 17(.90)^2 = 76.6996.$$

The F statistic is $F = \frac{\text{MST}}{\text{MSE}} = \frac{36.286/2}{76.6996/162} = 38.316$ with 2 numerator and 162 denominator degrees of freedom. From Table 7, p–value < .005 so we can reject the null hypothesis of equal mean maneuver times. The ANOVA table is below.

Source	d.f	SS	MS	F	p–value
Treatments	2	36.286	18.143	38.316	< .005
Error	162	76.6996	.4735		
Total	164	112.9856			

13.14 The grand mean is $\bar{y} = \frac{.032+.022+.041}{3} = 0.0317$. So,

$$\text{SST} = 10[(.032 - .0317)^2 + (.022 - .0317)^2 + (.041 - .0317)^2 = .001867.$$

$$\text{SSE} = \sum_{i=1}^{3}(n-1)s_i^2 = 9[(.014)^2 + (.008)^2 + (.017)^2] = .004941.$$

The F statistic is $F = 4.94$ with 2 numerator and 27 denominator degrees of freedom. Since $F_{.05} = 3.35$, we can reject H_0 and conclude that the mean chemical levels are different.

13.15 We will use R to solve this problem:

```
> oxygen <- c(5.9, 6.1, 6.3, 6.1, 6.0, 6.3, 6.6, 6.4, 6.4, 6.5, 4.8,
4.3, 5.0, 4.7, 5.1, 6.0, 6.2, 6.1, 5.8)
> location <- factor(c(1,1,1,1,1,2,2,2,2,2,3,3,3,3,3,4,4,4,4))
> summary(aov(oxygen~location))
            Df Sum Sq Mean Sq F value    Pr(>F)
location     3 7.8361  2.6120  63.656 9.195e-09 ***
Residuals   15 0.6155  0.0410
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

The null hypothesis is H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$, where μ_i is the mean dissolved O_2 in location i, $i = 1, 2, 3, 4$. Since the p–value is quite small, we can reject H_0 and conclude the mean dissolved O_2 levels differ.

13.16 The ANOVA table is below:

Source	d.f	SS	MS	F	p–value
Treatments	3	67.475	22.4917	.87	> .1
Error	36	935.5	25.9861		
Total	39	1002.975			

With 3 numerator and 36 denominator degrees of freedom, we fail to reject with α = .05: there is not enough evidence to conclude a difference in the four age groups.

13.17 $E(\overline{Y}_{i\bullet}) = \frac{1}{n_i}\sum_{j=1}^{n_i} E(Y_{ij}) = \frac{1}{n_i}\sum_{j=1}^{n_i}(\mu + \tau_i) = \frac{1}{n_i}\sum_{j=1}^{n_i}\mu_i = \mu_i$

$V(\overline{Y}_{i\bullet}) = \frac{1}{n_i^2}\sum_{j=1}^{n_i} V(Y_{ij}) = \frac{1}{n_i^2}\sum_{j=1}^{n_i} V(\varepsilon_{ij}) = \frac{1}{n_i}\sigma^2$.

13.18 Using the results from Ex. 13.17,

$$E(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) = \mu_i - \mu_{i'} = \mu + \tau_i - (\mu + \tau_{i'}) = \tau_i - \tau_{i'}$$
$$V(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) = V(\overline{Y}_{i\bullet}) + V(\overline{Y}_{i'\bullet}) = \left(\tfrac{1}{n_i} + \tfrac{1}{n_{i'}}\right)\sigma^2$$

13.19 **a.** Recall that $\mu_i = \mu + \tau_i$ for $i = 1, \ldots, k$. If all τ_i's = 0, then all μ_i's = μ. Conversely, if $\mu_1 = \mu_2 = \ldots = \mu_k$, we have that $\mu + \tau_1 = \mu + \tau_2 = \ldots = \mu + \tau_k$ and $\tau_1 = \tau_2 = \ldots = \tau_k$. Since it was assumed that $\sum_{i=1}^{k}\tau_i = 0$, all τ_i's = 0. Thus, the null hypotheses are equivalent.

b. Consider $\mu_i = \mu + \tau_i$ and $\mu_{i'} = \mu + \tau_{i'}$. If $\mu_i \neq \mu_{i'}$, then $\mu + \tau_i \neq \mu + \tau_{i'}$ and thus $\tau_i \neq \tau_{i'}$. Since $\sum_{i=1}^{k}\tau_i = 0$, at least one $\tau_i \neq 0$ (actually, there must be at least two). Conversely, let

$\tau_i \neq 0$. Since $\sum_{i=1}^{k} \tau_i = 0$, there must be at least one i' such that $\tau_i \neq \tau_{i'}$. With $\mu_i = \mu + \tau_i$ and $\mu_{i'} = \mu + \tau_{i'}$, it must be so that $\mu_i \neq \mu_{i'}$. Thus, the alternative hypotheses are equivalent.

13.20 **a.** First, note that $\bar{y}_1 = 75.67$ and $s_1^2 = 66.67$. Then, with $n_1 = 6$, a 95% CI is given by

$$75.67 \pm 2.571\sqrt{66.67/6} = 75.67 \pm 8.57 \text{ or } (67.10, 84.24).$$

b. The interval computed above is longer than the one in Example 13.3.
c. When only the first sample was used to estimate σ^2, there were only 5 degrees of freedom for error. However, when all four samples were used, there were 14 degrees of freedom for error. Since the critical value $t_{.025}$ is larger in the above, the CI is wider.

13.21 **a.** The 95% CI would be given by

$$\bar{y}_1 - \bar{y}_4 \pm t_{.025} s_{14} \sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_4}},$$

where $s_{14} = \sqrt{\frac{(n_1-1)s_1^2 + (n_4-1)s_4^2}{n_1+n_4-2}} = 7.366$. Since $t_{.025} = 2.306$ based on 8 degrees of freedom, the 95% CI is $-12.08 \pm 2.306(7.366)\sqrt{\tfrac{1}{6} + \tfrac{1}{4}} = -12.08 \pm 10.96$ or $(-23.04, -1.12)$.

b. The CI computed above is longer.

c. The interval computed in Example 13.4 was based on 19 degrees of freedom, and the critical value $t_{.025}$ was smaller.

13.22 **a.** Based on Ex. 13.20 and 13.21, we would expect the CIs to be shorter when all of the data in the one–way layout is used.

b. If the estimate of σ^2 using only one sample is much smaller than the pooled estimate (MSE) – so that the difference in degrees of freedom is offset – the CI width using just one sample could be shorter.

13.23 From Ex. 13.7, the four sample means are (again, using R):

```
> tapply(waste,plant,mean)
    A     B     C     D 
1.568 1.772 1.546 1.916 
>
```

a. In the above, the sample mean for plant A is 1.568 and from Ex. 13.7, MSE = .0298 with 16 degrees of freedom. Thus, a 95% CI for the mean amount of polluting effluent per gallon for plant A is

$$1.568 \pm 2.12\sqrt{.0298/5} = 1.568 \pm .164 \text{ or } (1.404, 1.732).$$

There is evidence that the plant is exceeding the limit since values larger than 1.5 lb/gal are contained in the CI.

b. A 95% CI for the difference in mean polluting effluent for plants A and D is

$$1.568 - 1.916 \pm 2.12\sqrt{.0298\left(\tfrac{2}{5}\right)} = -.348 \pm .231 \text{ or } (-.579, -.117).$$

Since 0 is not contained in the CI, there is evidence that the means differ for the two plants.

13.24 From Ex. 13.8, the three sample means are (again, using R):

```
> tapply(salary,type,mean)
 church private  public
  56.06   70.78   54.04
```

Also, MSE = 58.61 based on 12 degrees of freedom. A 98% CI for the difference in mean starting salaries for assistant professors at public and private/independent universities is

$$54.04 - 70.78 \pm 2.681\sqrt{58.61\left(\tfrac{2}{5}\right)} = -16.74 \pm 12.98 \text{ or } (-29.72, -3.76).$$

13.25 The 95% CI is given by $.93 - 1.21 \pm 1.96(.1383)\sqrt{2/14} = -.28 \pm .102$ or $(-.382, -.178)$ (note that the degrees of freedom for error is large, so 1.96 is used). There is evidence that the mean densities for the two groups are different since the CI does not contain 0.

13.26 Refer to Ex. 13.9. MSE = .0025 with 8 degrees of freedom.

a. 90% CI for μ_A: $2.25 \pm 1.86\sqrt{.0025/3} = 2.25 \pm .05$ or $(2.20, 2.30)$.

b. 95% CI for $\mu_A - \mu_B$: $2.25 - 2.166 \pm 2.306\sqrt{.0025\left(\tfrac{2}{3}\right)} = .084 \pm .091$ or $(-.007, .175)$.

13.27 Refer to Ex. 13.10. MSE = 62.233 with 8 degrees of freedom.

a. 95% CI for μ_A: $76 \pm 2.306\sqrt{62.333/5} = 76 \pm 8.142$ or $(67.868, 84.142)$.

b. 95% CI for μ_B: $66.33 \pm 2.306\sqrt{62.333/3} = 66.33 \pm 10.51$ or $(55.82, 76.84)$.

c. 95% CI for $\mu_A - \mu_B$: $76 - 66.33 \pm 2.306\sqrt{62.333\left(\tfrac{1}{5}+\tfrac{1}{3}\right)} = 9.667 \pm 13.295$.

13.28 Refer to Ex. 13.12. MSE = 1.852 with 33 degrees of freedom

a. $23.965 \pm 1.96\sqrt{1.852/12} = 23.962 \pm .77$.

b. $23.965 - 20.463 \pm 1.645\sqrt{1.852\left(\tfrac{2}{12}\right)} = 3.502 \pm .914$.

13.29 Refer to Ex. 13.13. MSE = .4735 with 162 degrees of freedom.

a. $6.24 \pm 1.96\sqrt{.4735/18} = 6.24 \pm .318$.

b. $4.59 - 4.58 \pm 1.96\sqrt{.4735\left(\tfrac{1}{45}+\tfrac{1}{102}\right)} = -.29 \pm .241$.

c. Probably not, since the sample was only selected from one town and driving habits can vary from town to town.

13.30 The ANOVA table for these data is below.

Source	d.f	SS	MS	F	p–value
Treatments	3	36.7497	12.2499	4.88	< .05
Error	24	60.2822	2.5118		
Total	27	97.0319			

a. Since $F_{.05} = 3.01$ with 3 numerator and 24 denominator degrees of freedom, we reject the hypothesis that the mean wear levels are equal for the four treatments.

b. With $\bar{y}_2$ = 14.093 and $\bar{y}_3$ = 12.429, a 99% CI for the difference in the means is

$$14.093 - 12.429 \pm 2.797\sqrt{2.5118\left(\tfrac{2}{7}\right)} = 1.664 \pm 2.3695.$$

c. A 90% CI for the mean wear with treatment A is
$11.986 \pm 1.711\sqrt{2.5118\left(\tfrac{1}{7}\right)}$ = 11.986 ± 1.025 or (10.961, 13.011).

13.31 The ANOVA table for these data is below.

Source	d.f	SS	MS	F	p–value
Treatments	3	18.1875	2.7292	1.32	> .1
Error	12	24.75	2.0625		
Total	15	32.9375			

a. Since $F_{.05}$ = 3.49 with 3 numerator and 12 denominator degrees of freedom, we fail to reject the hypothesis that the mean amounts are equal.

b. The methods of interest are 1 and 4. So, with $\bar{y}_1$ = 2 and $\bar{y}_4$ = 4, a 95% CI for the difference in the mean levels is
$2 - 4 \pm 2.052\sqrt{2.0625\left(\tfrac{2}{4}\right)}$ = –2 ± 2.21 or (–.21, 4.21).

13.32 Refer to Ex. 13.14. MSE = .000183 with 27 degrees of freedom. A 95% CI for the mean residue from DDT is $.041 \pm 2.052\sqrt{.000183/10}$ = .041 ± .009 or (.032, .050).

13.33 Refer to Ex. 13.15. MSE = .041 with 15 degrees of freedom. A 95% CI for the difference in mean O2 content for midstream and adjacent locations is

$$6.44 - 4.78 \pm 2.131\sqrt{.041\left(\tfrac{2}{5}\right)} = 1.66 \pm .273 \text{ or } (1.39, 1.93).$$

13.34 The estimator for $\theta = \frac{1}{2}(\mu_1 + \mu_2) - \mu_4$ is $\hat{\theta} = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) - \bar{y}_4$. So, $V(\hat{\theta}) = \frac{1}{4}\left(\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}\right) + \frac{\sigma^2}{n_4}$.
A 95% CI for θ is given by $\frac{1}{2}(\bar{y}_1 + \bar{y}_2) - \bar{y}_4 \pm t_{.025}\sqrt{\text{MSE}\left(\frac{1}{4n_1} + \frac{1}{4n_2} + \frac{1}{n_4}\right)}$. Using the supplied data, this is found to be .235 ± .255.

13.35 Refer to Ex. 13.16. MSE = 25.986 with 36 degrees of freedom.

a. A 90% CI for the difference in mean heart rate increase for the 1st and 4th groups is

$$30.9 - 28.2 \pm 1.645\sqrt{25.986\left(\tfrac{2}{10}\right)} = 2.7 \pm 3.75.$$

b. A 90% CI for the 2nd group is

$$27.5 \pm 1.645\sqrt{25.986/10} = 27.5 \pm 2.652 \text{ or } (24.85, 30.15).$$

13.36 See Sections 12.3 and 13.7.

13.37 **a.** $\frac{1}{bk}\sum_{i=1}^{k}\sum_{j=1}^{b}E(Y_{ij}) = \frac{1}{bk}\sum_{i=1}^{k}\sum_{j=1}^{b}(\mu + \tau_i + \beta_j) = \frac{1}{bk}(bk\mu + b\sum_{i=1}^{k}\tau_i + k\sum_{j=1}^{b}\beta_j = \mu$.

b. The parameter μ represents the overall mean.

13.38 We have that:

$$\overline{Y}_{i\bullet} = \tfrac{1}{b}\sum_{j=1}^{b} Y_{ij} = \tfrac{1}{b}\sum_{j=1}^{b}(\mu + \tau_i + \beta_j + \varepsilon_{ij})$$
$$= \mu + \tau_i + \tfrac{1}{b}\sum_{j=1}^{b}\beta_j + \tfrac{1}{b}\sum_{j=1}^{b}\varepsilon_{ij} = \mu + \tau_i + \tfrac{1}{b}\sum_{j=1}^{b}\varepsilon_{ij}\ .$$

Thus: $E(\overline{Y}_{i\bullet}) = \mu + \tau_i + \tfrac{1}{b}\sum_{j=1}^{b}E(\varepsilon_{ij}) = \mu + \tau_i = \mu_i$, so $\overline{Y}_{i\bullet}$ is an unbiased estimator.

$$V(\overline{Y}_{i\bullet}) = \tfrac{1}{b^2}\sum_{j=1}^{b}V(\varepsilon_{ij}) = \tfrac{1}{b}\sigma^2\ .$$

13.39 Refer to Ex. 13.38.

a. $E(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) = \mu + \tau_i - (\mu + \tau_{i'}) = \tau_i - \tau_{i'}$.

b. $V(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) = V(\overline{Y}_{i\bullet}) + V(\overline{Y}_{i'\bullet}) = \tfrac{2}{b}\sigma^2$, since $\overline{Y}_{i\bullet}$ and $\overline{Y}_{i'\bullet}$ are independent.

13.40 Similar to Ex. 13.38, we have that

$$\overline{Y}_{\bullet j} = \tfrac{1}{k}\sum_{i=1}^{k} Y_{ij} = \tfrac{1}{k}\sum_{i=1}^{k}(\mu + \tau_i + \beta_j + \varepsilon_{ij})$$
$$= \mu + \tfrac{1}{k}\sum_{i=1}^{k}\tau_i + \beta_j + \tfrac{1}{k}\sum_{i=1}^{k}\varepsilon_{ij} = \mu + \beta_j + \tfrac{1}{k}\sum_{i=1}^{k}\varepsilon_{ij}\ .$$

a. $E(\overline{Y}_{\bullet j}) = \mu + \beta_j = \mu_j$, $V(\overline{Y}_{\bullet j}) = \tfrac{1}{k^2}\sum_{i=1}^{k}V(\varepsilon_{ij}) = \tfrac{1}{k}\sigma^2$.

b. $E(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet j'}) = \mu + \beta_j - (\mu + \beta_{j'}) = \beta_j - \beta_{j'}$.

c. $V(\overline{Y}_{\bullet j} - \overline{Y}_{\bullet j'}) = V(\overline{Y}_{\bullet j}) + V(\overline{Y}_{\bullet j'}) = \tfrac{2}{k}\sigma^2$, since $\overline{Y}_{\bullet j}$ and $\overline{Y}_{\bullet j'}$ are independent.

13.41 The sums of squares are Total SS = 1.7419, SST = .0014, SSB = 1.7382, and SSE = .0023. The ANOVA table is given below:

Source	d.f	SS	MS	F
Program	5	1.7382	.3476	772.4
Treatments	1	.0014	.0014	3.11
Error	5	.0023	.00045	
Total	11	1.7419		

a. To test H_0: $\mu_1 = \mu_2$, the F–statistic is $F = 3.11$ with 1 numerator and 5 denominator degrees of freedom. Since $F_{.05} = 6.61$, we fail to reject the hypothesis that the mean CPU times are equal. This is the same result as Ex. 12.10(b).

b. From Table 7, p–value > .10.

c. Using the Applet, p–value = $P(F > 3.11) = .1381$.

d. Ignoring the round–off error, s_D^2 = 2MSE.

13.42 Using the formulas from this section, TSS = 674 – 588 = 86, SSB = $\frac{20^2+36^2+28^2}{4}$ – CM = 32, SST = $\frac{21^2+\ldots+18^2}{3}$ – CM = 42. Thus, SSE = 86 – 32 – 42 = 12. The remaining calculations are given in the ANOVA table below.

Source	d.f	SS	MS	F
Treatments	3	42	14	7
Blocks	2	32	16	
Error	6	12	2	
Total	11	86		

The F–statistic is $F = 7$ with 3 and 6 degrees of freedom. With $\alpha = .05$, $F_{.05} = 4.76$ so we can reject the hypothesis that the mean resistances are equal. Also, $.01 < p$–value $< .025$ from Table 7.

13.43 Since the four chemicals (the treatment) were applied to three different materials, the material type could add unwanted variation to the analysis. So, material type was treated as a blocking variable.

13.44 Here, R will be used to analyze the data. We will use the letters A, B, C, and D to denote the location and the numbers 1, 2, 3, 4, and 5 to denote the company.

```
> rate <- c(736, 745, 668, 1065, 1202, 836, 725, 618, 869, 1172, 1492,
1384,1214, 1502, 1682, 996, 884, 802, 1571, 1272)
> location <- factor(c(rep("A",5),rep("B",5),rep("C",5),rep("D",5)))
> company <- factor(c(1:5,1:5,1:5,1:5))
> summary(aov(rate ~ company + location))
            Df  Sum Sq Mean Sq F value    Pr(>F)
company      4  731309  182827  12.204 0.0003432 ***
location     3 1176270  392090  26.173 1.499e-05 ***
Residuals   12  179769   14981
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. This is a randomized block design (applied to sampled data).
b. The F–statistic is $F = 26.173$ with a p–value of .00001499. Thus, we can safely conclude that there is a difference in mean premiums.
c. The F–statistic is $F = 12.204$ with a p–value of .0003432. Thus, we can safely conclude that there is a difference in the locations.
d. See parts b and c above.

13.45 The treatment of interest is the soil preparation and the location is a blocking variable. The summary statistics are:
$CM = (162)^2/12 = 2187$, $TSS = 2298 - CM = 111$, $SST = 8900/4 - CM = 38$,
$SSB = 6746/3 - CM = 61.67$. The ANOVA table is below.

Source	d.f	SS	MS	F
Treatments	2	38	19	10.05
Blocks	3	61.67	20.56	10.88
Error	6	11.33	1.89	
Total	11	111		

a. The F–statistic for soil preparations is F = 10.05 with 2 numerator and 6 denominator degrees of freedom. From Table 7, p–value < .025 so we can reject the null hypothesis that the mean growth is equal for all soil preparations.

b. The F–statistic for the locations is F = 10.88 with 3 numerator and 6 denominator degrees of freedom. Here, p–value < .01 so we can reject the null hypothesis that the mean growth is equal for all locations.

13.46 The ANOVA table is below.

Source	d.f	SS	MS	F
Treatments	4	.452	.113	8.37
Blocks	3	1.052	.3507	25.97
Error	12	.162	.0135	
Total	19	1.666		

a. To test for a difference in the varieties, the F–statistic is F = 8.37 with 4 numerator and 12 denominator degrees of freedom. From Table 7, p–value < .005 so we would reject the null hypothesis at $\alpha = .05$.

b. The F–statistic for blocks is 25.97 with 3 numerator and 12 denominator degrees of freedom. Since $F_{.05} = 3.49$, we reject the hypothesis of no difference between blocks.

13.47 Using a randomized block design with locations as blocks, the ANOVA table is below.

Source	d.f	SS	MS	F
Treatments	3	8.1875	2.729	1.40
Blocks	3	7.1875	2.396	1.23
Error	9	17.5625	1.95139	
Total	15	32.9375		

With 3 numerator and 9 denominator degrees of freedom, $F_{.05} = 3.86$. Thus, neither the treatment effect nor the blocking effect is significant.

13.48 Note that there are $2bk$ observations. So, let y_{ijl} denote the l^{th} observation in the j^{th} block receiving the i^{th} treatment. Therefore, with CM = $\dfrac{\left(\sum_{i,j,l} y_{ijl}\right)^2}{2bk}$,

TSS = $\sum_{i,j,l} y_{ijl}^2$ – CM with $2bk - 1$ degrees of freedom,

SST = $\dfrac{\sum_i y_{i\bullet\bullet}^2}{2b}$ – CM with $k - 1$ degrees of freedom,

SSB = $\dfrac{\sum_j y_{\bullet j\bullet}^2}{2k}$, with $b - 1$ degrees of freedom, and

SSE = TSS – SST – SSB with $2bk - b - k - 1$ degrees of freedom.

13.49 Using a randomized block design with ingots as blocks, the ANOVA table is below.

Source	d.f	SS	MS	F
Treatments	2	131.901	65.9505	6.36
Blocks	6	268.90	44.8167	
Error	12	124.459	10.3716	
Total	20	524.65		

To test for a difference in the mean pressures for the three bonding agents, the F–statistic is $F = 6.36$ with 2 numerator and 12 denominator degrees of freedom. Since $F_{.05} = 3.89$, we can reject H_0.

13.50 Here, R will be used to analyze the data. The carriers are the treatment levels and the blocking variable is the shipment.

```
> time <- c(15.2,14.3, 14.7, 15.1, 14.0, 16.9, 16.4, 15.9, 16.7, 15.6,
17.1, 16.1, 15.7, 17.0, 15.5)     # data is entered going down columns
> carrier <- factor(c(rep("I",5),rep("II",5),rep("III",5)))
> shipment <- factor(c(1:5,1:5,1:5))
> summary(aov(time ~ carrier + shipment))
            Df Sum Sq Mean Sq F value    Pr(>F)
carrier      2 8.8573  4.4287  83.823 4.303e-06 ***
shipment     4 3.9773  0.9943  18.820  0.000393 ***
Residuals    8 0.4227  0.0528
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

To test for a difference in mean delivery times for the carriers, from the output we have the F–statistic $F = 83.823$ with 2 numerator and 8 denominator degrees of freedom. Since the p–value is quite small, we can conclude there is a difference in mean delivery times between carriers.

A randomized block design was used because different size/weight shipments can also affect the delivery time. In the experiment, shipment type was blocked.

13.51 Some preliminary results are necessary in order to obtain the solution (see Ex. 13.37–40):

(1) $E(Y_{ij}^2) = V(Y_{ij}) + [E(Y_{ij})]^2 = \sigma^2 + (\mu + \tau_i + \beta_j)^2$

(2) With $\overline{Y}_{\bullet\bullet} = \frac{1}{bk}\sum_{i,j} Y_{ij}$, $E(\overline{Y}_{\bullet\bullet}) = \mu$, $V(\overline{Y}_{\bullet\bullet}) = \frac{1}{bk}\sigma^2$, $E(\overline{Y}_{\bullet\bullet}^2) = \frac{1}{bk}\sigma^2 + \mu^2$

(3) With $\overline{Y}_{\bullet j} = \frac{1}{k}\sum_i Y_{ij}$, $E(\overline{Y}_{\bullet j}) = \mu + \beta_j$, $V(\overline{Y}_{\bullet j}) = \frac{1}{k}\sigma^2$, $E(\overline{Y}_{\bullet j}^2) = \frac{1}{k}\sigma^2 + (\mu + \beta_j)^2$

(4) With $\overline{Y}_{i\bullet} = \frac{1}{b}\sum_j Y_{ij}$, $E(\overline{Y}_{i\bullet}) = \mu + \tau_i$, $V(\overline{Y}_{i\bullet}) = \frac{1}{b}\sigma^2$, $E(\overline{Y}_{i\bullet}^2) = \frac{1}{b}\sigma^2 + (\mu + \tau_i)^2$

a. $$E(\text{MST}) = \frac{b}{k-1}E\left[\sum_i (\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2\right] = \frac{b}{k-1}\left[\sum_i E\left(\overline{Y}_{i\bullet}^2\right) - kE\left(\overline{Y}_{\bullet\bullet}^2\right)\right]$$

$$= \frac{b}{k-1}\left[\sum_i \left(\frac{\sigma^2}{b} + \mu^2 + 2\mu\tau_i + \tau_i^2\right) - k\left(\frac{\sigma^2}{bk} + \mu^2\right)\right] = \sigma^2 + \frac{b}{k-1}\sum_i \tau_i^2 .$$

b. $E(\text{MSB}) = \frac{k}{b-1} E\left[\sum_j (\overline{Y}_{\bullet j} - Y_{\bullet\bullet})^2\right] = \frac{k}{b-1}\left[\sum_j E\left(\overline{Y}_{\bullet j}^2\right) - bE\left(\overline{Y}_{\bullet\bullet}^2\right)\right]$

$$= \frac{k}{b-1}\left[\sum_j \left(\frac{\sigma^2}{k} + \mu^2 + 2\mu\beta_j + \beta_j^2\right) - b\left(\frac{\sigma^2}{bk} + \mu^2\right)\right] = \sigma^2 + \frac{k}{b-1}\sum_j \beta_j^2 .$$

c. Recall that TSS = $\sum_{i,j} Y_{ij}^2 - bk\overline{Y}_{\bullet\bullet}^2$. Thus,

$$E(\text{TSS}) = \sum_{i,j}\left(\sigma^2 + \mu^2 + \tau_i^2 + \beta_j^2\right) - bk\left(\frac{\sigma^2}{bk} + \mu^2\right) = (bk-1)\sigma^2 + b\sum_i \tau_i^2 + k\sum_j \beta_j^2 .$$

Therefore, since E(SSE) = E(TSS) – E(SST) – E(SSB), we have that

$$E(\text{SSE}) = E(\text{TSS}) - (k-1)E(\text{MST}) - (b-1)E(\text{MSB}) = (bk - k - b + 1)\sigma^2.$$

Finally, since $\text{MST} = \frac{\text{SSE}}{bk - k - b + 1}$, $E(\text{MST}) = \sigma^2$.

13.52 From Ex. 13.41, recall that MSE = .00045 with 5 degrees of freedom and $b = 6$. Thus, a 95% CI for the difference in mean CPU times for the two computers is

$$1.553 - 1.575 \pm 2.571\sqrt{.00045\left(\tfrac{2}{6}\right)} = -.022 \pm .031 \text{ or } (-.053, .009).$$

This is the same interval computed in Ex. 12.10(c).

13.53 From Ex. 13.42, MSE = 2 with 6 degrees of freedom and $b = 3$. Thus, the 95% CI is

$$7 - 5 \pm 2.447\sqrt{2\left(\tfrac{2}{3}\right)} = 2 \pm 2.83.$$

13.54 From Ex. 13.45, MSE = 1.89 with 6 degrees of freedom and $b = 4$. Thus, the 90% CI is

$$16 - 12.5 \pm 1.943\sqrt{1.89\left(\tfrac{2}{4}\right)} = 3.5 \pm 1.89 \text{ or } (1.61, 5.39).$$

13.55 From Ex. 13.46, MSE = .0135 with 12 degrees of freedom and $b = 4$. The 95% CI is

$$2.689 - 2.544 \pm 2.179\sqrt{.0135\left(\tfrac{2}{4}\right)} = .145 \pm .179.$$

13.56 From Ex. 13.47, MSE = 1.95139 with 9 degrees of freedom and $b = 4$. The 95% CI is

$$2 \pm 2.262\sqrt{1.95139\left(\tfrac{2}{4}\right)} = 2 \pm 2.23.$$

This differs very little from the CI computed in Ex. 13.31(b) (without blocking).

13.57 From Ex. 13.49, MSE = 10.3716 with 12 degrees of freedom and $b = 7$. The 99% CI is

$$71.1 - 75.9 \pm 3.055\sqrt{10.3716\left(\tfrac{2}{7}\right)} = -4.8 \pm 5.259.$$

13.58 Refer to Ex. 13.9. We require an error bound of no more than .02, so we need n such that

$$2\sqrt{\sigma^2\left(\tfrac{2}{n}\right)} \le .02,$$

The best estimate of σ^2 is MSE = .0025, so using this in the above we find that $n \ge 50$. So the entire number of observations needed for the experiment is $4n \ge 4(50) = 200$.

13.59 Following Ex. 13.27(a), we require $2\sqrt{\frac{\sigma^2}{n_A}} \le 10$, where $2 \approx t_{.025}$. Estimating σ^2 with MSE = 62.333, the solution is $n_A \ge 2.49$, so at least 3 observations are necessary.

13.60 Following Ex. 13.27(c), we require $2\sqrt{\sigma^2\left(\frac{2}{n}\right)} \le 20$ where $2 \approx t_{.025}$. Estimating σ^2 with MSE = 62.333, the solution is $n \ge 1.24$, so at least 2 observations are necessary. The total number of observations that are necessary is $3n \ge 6$.

13.61 Following Ex. 13.45, we must find b, the number of locations (blocks), such that

$$2\sqrt{\sigma^2\left(\tfrac{2}{b}\right)} \le 1,$$

where $2 \approx t_{.025}$. Estimating σ^2 with MSE = 1.89, the solution is $b \ge 15.12$, so at least 16 locations must be used. The total number of locations needed in the experiment is at least 3(16) = 48.

13.62 Following Ex. 13.55, we must find b, the number of locations (blocks), such that

$$2\sqrt{\sigma^2\left(\tfrac{2}{b}\right)} \le .5,$$

where $2 \approx t_{.025}$. Estimating σ^2 with MSE = 1.95139, the solution is $b \ge 62.44$, so at least 63 locations are needed.

13.63 The CI lengths also depend on the sample sizes n_i and $n_{i'}$, and since these are not equal, the intervals differ in length.

13.64 **a.** From Example 13.9, $t_{.00417} = 2.9439$. A 99.166% CI for $\mu_1 - \mu_2$ is

$$75.67 - 78.43 \pm 2.9439(7.937)\sqrt{\tfrac{1}{6} + \tfrac{1}{7}} = -2.76 \pm 13.00.$$

b. The ratio is $\dfrac{2(12.63)}{2(13.00)} = .97154$.

c. The ratios are equivalent (save roundoff error).

d. If we divide the CI length for $\mu_1 - \mu_3$ (or equivalently the margin of error) found in Ex. 13.9 by the ratio given in part b above, a 99.166% CI for $\mu_1 - \mu_3$ can be found to be

$$4.84 \pm 13.11/.97154 = 4.84 \pm 13.49.$$

13.65 Refer to Ex. 13.13. Since there are three intervals, each should have confidence coefficient 1 – .05/3 = .9833. Since MSE = .4735 with 162 degrees of freedom, a critical value from the standard normal distribution can be used. So, since $\alpha = 1 - .9833 = .0167$, we require $z_{\alpha/2} = z_{.00833} = 2.39$. Thus, for pairs (i, j) of (1, 2), (1, 3) and (2, 3), the CIs are

$$\begin{aligned}
(1,2):&\quad -0.29 \pm 2.39\sqrt{.4735\left(\tfrac{1}{45} + \tfrac{1}{102}\right)} \quad \text{or } -0.29 \pm .294 \\
(1,3):&\quad -1.65 \pm 2.39\sqrt{.4735\left(\tfrac{1}{45} + \tfrac{1}{18}\right)} \quad \text{or } -1.65 \pm .459\,. \\
(2,3):&\quad -1.36 \pm 2.39\sqrt{.4735\left(\tfrac{1}{102} + \tfrac{1}{18}\right)} \quad \text{or } -1.36 \pm .420
\end{aligned}$$

The simultaneous coverage rate is at least 95%. Note that only the interval for (1, 2) contains 0, suggesting that μ_1 and μ_2 could be equal.

13.66 In this case there are three pairwise comparisons to be made. Thus, the Bonferroni technique should be used with $m = 3$.

13.67 Refer to Ex. 13.45. There are three intervals to construct, so with $\alpha = .10$, each CI should have confidence coefficient $1 - .10/3 = .9667$. Since MSE = 1.89 with 6 degrees of freedom, we require $t_{.0167}$ from this t–distribution. As a conservative approach, we will use $t_{.01} = 3.143$ since $t_{.0167}$ is not available in Table 5 (thus, the simultaneous coverage rate is at least 94%). The intervals all have half width $3.143\sqrt{1.89\left(\frac{2}{4}\right)} = 3.06$ so that the intervals are:

$(1, 2)$: -3.5 ± 3.06 or $(-6.56, -.44)$
$(1, 3)$: $.5 \pm 3.06$ or $(-2.56, 3.56)$
$(2, 3)$: 4.0 ± 3.06 or $(.94, 7.06)$

13.68 Following Ex. 13.47, MSE = 1.95139 with 9 degrees of freedom. For an overall confidence level of 95% with 3 intervals, we require $t_{.025/3} = t_{.0083}$. By approximating this with $t_{.01}$, the half width of each interval is $2.821\sqrt{1.95139\left(\frac{2}{4}\right)} = 2.79$. The intervals are:

$(1, 4)$: -2 ± 2.79 or $(-4.79, .79)$
$(2, 4)$: -1 ± 2.79 or $(-3.79, 1.79)$
$(3, 4)$: $-.75 \pm 2.79$ or $(-3.54, 2.04)$

13.69 **a.** $\beta_0 + \beta_3$ is the mean response to treatment A in block III.
b. β_3 is the difference in mean responses to chemicals A and D in block III.

13.70 **a.** The complete model is $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, where

$$x_1 = \begin{cases} 1 & \text{if method } A \\ 0 & \text{otherwise} \end{cases}, \qquad x_2 = \begin{cases} 1 & \text{if method } B \\ 0 & \text{otherwise} \end{cases}.$$

Then, we have

$$\boldsymbol{Y} = \begin{bmatrix} 73 \\ 83 \\ 76 \\ 68 \\ 80 \\ 54 \\ 74 \\ 71 \\ 79 \\ 95 \\ 87 \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad \boldsymbol{X'X} = \begin{bmatrix} 11 & 5 & 3 \\ 5 & 5 & 0 \\ 3 & 0 & 3 \end{bmatrix} \qquad \hat{\boldsymbol{\beta}} = \begin{bmatrix} 87 \\ -11 \\ 20.67 \end{bmatrix}$$

Thus, $\text{SSE}_c = \boldsymbol{Y'Y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X'Y}$ = 65,286 – 54,787.33 = 498.67 with 11 – 3 = 8 degrees of freedom. The reduced model is $Y = \beta_0 + \varepsilon$, so that $\boldsymbol{X}$ is simply a column vector of eleven 1's and $(\boldsymbol{X'X})^{-1} = \frac{1}{11}$. Thus, $\hat{\beta} = \bar{y}$ = 76.3636. Thus, SSE_R = 65,286 – 64,145.455 = 1140.5455. Thus, to test H_0: $\beta_1 = \beta_2 = 0$, the reduced model F–test statistic is

$$F = \frac{(1140.5455 - 498.67)/2}{498.67/8} = 5.15$$

with 2 numerator and 8 denominator degrees of freedom. Since $F_{.05} = 4.46$, we reject H_0.

b. The hypotheses of interest are H_0: $\mu_A - \mu_B = 0$ versus a two–tailed alternative. Since MSE = $\text{SSE}_c/8$ = 62.333, the test statistic is

$$|t| = \frac{|76-66.33|}{\sqrt{62.333\left(\frac{1}{5}+\frac{1}{3}\right)}} = 1.68.$$

Since $t_{.025} = 2.306$, the null hypothesis is not rejected: there is not a significant difference between the two mean levels.

c. For part **a**, from Table 7 we have .025 < p–value < .05. For part **b**, from Table 5 we have 2(.05) < p–value < 2(.10) or .10 < p–value < .20.

13.71 The complete model is $Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \varepsilon$, where x_1 and x_2 are dummy variables for blocks and x_3, x_4, x_5 are dummy variables for treatments. Then,

$$\boldsymbol{Y} = \begin{bmatrix} 5\\3\\8\\4\\9\\8\\13\\6\\7\\4\\9\\8 \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1&1&0&1&0&0\\ 1&1&0&0&1&0\\ 1&1&0&0&0&1\\ 1&1&0&0&0&0\\ 1&0&1&1&0&0\\ 1&0&1&0&1&0\\ 1&0&1&0&0&1\\ 1&0&1&0&0&0\\ 1&0&0&1&0&0\\ 1&0&0&0&1&0\\ 1&0&0&0&0&1\\ 1&0&0&0&0&0 \end{bmatrix} \quad \boldsymbol{X'X} = \begin{bmatrix} 12&4&4&3&3&3\\ 4&4&0&1&1&1\\ 4&0&4&1&1&1\\ 3&1&1&3&0&0\\ 3&1&1&0&3&0\\ 3&1&1&0&0&3 \end{bmatrix} \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} 6\\-2\\2\\1\\-1\\4 \end{bmatrix}$$

Thus, SSE_c = 674 – 662 = 12 with 12 – 6 = 6 degrees of freedom. The reduced model is $Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \varepsilon$, where x_1 and x_2 are as defined in the complete model. Then,

$$(\boldsymbol{X'X})^{-1} = \begin{bmatrix} .25 & -.25 & -.25\\ -.25 & .5 & .25\\ -.25 & .25 & .5 \end{bmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{bmatrix} 7\\2\\-2 \end{bmatrix}$$

so that SSE_R = 674 – 620 = 54 with 12 – 3 = 9 degrees of freedom. The reduced model F–test statistic is $F = \frac{(54-12)/3}{12/6} = 7$ with 3 numerator and 6 denominator degrees of freedom. Since $F_{.05} = 4.76$, H_0 is rejected: the treatment means are different.

13.72 (Similar to Ex. 13.71). The full model is $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$, where x_1, x_2, and x_3 are dummy variables for blocks and x_4 and x_5 are dummy variables for treatments. It can be shown that $SSE_c = 2298 - 2286.6667 = 11.3333$ with $12 - 6 = 6$ degrees of freedom. The reduced model is $Y = \beta_0 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon$, and $SSE_R = 2298 - 2225 = 73$ with $12 - 3 = 9$ degrees of freedom. Then, the reduced model F–test statistic is $F = \frac{(73-11.3333)/3}{11.3333/6} = 10.88$ with 3 numerator and 6 denominator degrees of freedom. Since Since $F_{.05} = 4.76$, H_0 is rejected: there is a difference due to location.

13.73 See Section 13.8. The experimental units within each block should be as homogenous as possible.

13.74 **a.** For the CRD, experimental units are randomly assigned to treatments.
b. For the RBD, experimental units are randomly assigned the k treatments *within each block*.

13.75 **a.** Experimental units are the patches of skin, while the three people act as blocks.
b. Here, MST = 1.18/2 = .59 and MSE = 2.24/4 = .56. Thus, to test for a difference in treatment means, calculate $F = .59/.56 = 1.05$ with 2 numerator and 4 denominator degrees of freedom. Since $F_{.05} = 6.94$, we cannot conclude there is a difference.

13.76 Refer to Ex. 13.9. We have that CM = 58.08, TSS = .035, and SST = .015. Then, SSB = $\frac{(8.9)^2+(8.6)^2+(8.9)^2}{4} - CM = .015$ with 2 degrees of freedom. The ANOVA table is below:

Source	d.f	SS	MS	F
Treatments	3	.015	.00500	6.00
Blocks	2	.015	.00750	9.00
Error	6	.005	.000833	
Total	11	.035		

a. To test for a "sand" effect, this is determined by an F–test for blocks. From the ANOVA table $F = 9.00$ with 2 numerator and 6 denominator degrees of freedom. Since $F_{.05} = 5.14$, we can conclude that the type of sand is important.
b. To test for a "concrete type" effect, from the ANOVA table $F = 6.00$ with 3 numerator and 6 denominator degrees of freedom. Since $F_{.05} = 4.76$, we can conclude that the type of concrete mix used is important.
c. Compare the sizes of SSE from Ex. 13.9 and what was calculated here. Since the experimental error was estimated to be much larger in Ex. 13.9 (by ignoring a block effect), the test for treatment effect was not significant.

13.77 Refer to Ex. 13.76
a. A 95% CI is given by $2.25 - 2.166 \pm 2.447\sqrt{.000833\left(\frac{2}{3}\right)} = .084 \pm .06$ or (.024, .144).
b. Since the SSE has been reduced by accounting for a block effect, the precision has been improved.

13.78 **a.** This is not a randomized block design. There are 9 treatments (one level of drug 1 and one level of drug 2). Since both drugs are factors, there could be interaction present.

b. The second design is similar to the first, except that there are two patients assigned to each treatment in a completely randomized design.

13.79 **a.** We require $2\sigma\frac{1}{\sqrt{n}} \leq 10$, so that $n \geq 16$.

b. With 16 patients assigned to each of the 9 treatments, there are 16(9) – 9 = 135 degrees of freedom left for error.

c. The half width, using $t_{.025} \approx 2$, is given by $2(20)\sqrt{\frac{1}{16}+\frac{1}{16}}$ = 14.14.

13.80 In this experiment, the car model is the treatment and the gasoline brand is the block. Here, we will use R to analyze the data:

```
> distance <- c(22.4, 20.8, 21.5, 17.0, 19.4, 18.7, 19.2, 20.2, 21.2)
> model <- factor(c("A","A","A","B","B","B","C","C","C"))
> gasoline <- factor(c("X","Y","Z","X","Y","Z","X","Y","Z"))
> summary(aov(distance ~ model + gasoline))
            Df  Sum Sq Mean Sq F value  Pr(>F)
model        2 15.4689  7.7344  6.1986 0.05951 .
gasoline     2  1.3422  0.6711  0.5378 0.62105
Residuals    4  4.9911  1.2478
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. To test for a car model effect, the F–test statistic is F = 6.1986 and by the p–value this is not significant at the α = .05 level.

b. To test for a gasoline brand effect, the F–test statistic is F = .5378. With a p–value of .62105, this is not significant and so gasoline brand does not affect gas mileage.

13.81 Following Ex. 13.81, the R output is

```
> summary(aov(distance~model))
            Df  Sum Sq Mean Sq F value  Pr(>F)
model        2 15.4689  7.7344  7.3274 0.02451 *
Residuals    6  6.3333  1.0556
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a. To test for a car model effect, the F–test statistic is F = 6.1986 with p–value = .02451. Thus, with α = .05, we can conclude that the car model has an effect on gas mileage.

b. In the RBD, SSE was reduced (somewhat) but 2 degrees of freedom were lost. Thus MSE is larger in the RBD than in the CRD.

c. The CRD randomly assigns treatments to experimental units. In the RBD, treatments are randomly assigned to experimental units within each block, and this is not the same randomization procedure as a CRD.

13.82 **a.** This is a completely randomized design.

b. The sums of squares are: TSS = 183.059, SST = 117.642, and SSE = 183.059 – 117.642 = 65.417. The ANOVA table is given below

Source	d.f	SS	MS	F
Treatments	3	117.642	39.214	7.79
Error	13	65.417	5.032	
Total	16	183.059		

To test for equality in mean travel times, the F–test statistic is $F = 7.79$ with 3 numerator and 13 denominator degrees of freedom. With $F_{.01} = 5.74$, we can reject the hypothesis that the mean travel times are equal.

c. With $\bar{y}_1 = 26.75$ and $\bar{y}_3 = 32.4$, a 95% CI for the difference in means is

$$26.75 - 32.4 \pm 2.160\sqrt{5.032\left(\tfrac{1}{4}+\tfrac{1}{5}\right)} = -5.65 \pm 3.25 \text{ or } (-8.90, -2.40).$$

13.83 This is a RBD with digitalis as the treatment and dogs are blocks.

a. TSS = 703,681.667, SST = 524,177.167, SSB = 173,415, and SSE = 6089.5. The ANOVA table is below.

Source	d.f	SS	MS	F
Treatments	2	524,177.167	262,088.58	258.237
Blocks	3	173,415	57,805.00	56.95
Error	6	6,089.5	1,014.9167	
Total	11	703,681.667		

b. There are 6 degrees of freedom for SSE.

c. To test for a digitalis effect, the F–test has $F = 258.237$ with 2 numerator and 6 denominator degrees of freedom. From Table 7, p–value < .005 so this is significant.

d. To test for a dog effect, the F–test has $F = 56.95$ with 3 numerator and 6 denominator degrees of freedom. From Table 7, p–value < .005 so this is significant.

e. The standard deviation of the difference between the mean calcium uptake for two levels of digitalis is $s\sqrt{\frac{1}{n_i}+\frac{1}{n_j}} = \sqrt{1014.9167\left(\tfrac{1}{4}+\tfrac{1}{4}\right)} = 22.527$.

f. The CI is given by $1165.25 - 1402.5 \pm 2.447(22.53) = -237.25 \pm 55.13$.

13.84 We require $2\sqrt{\sigma^2\left(\frac{2}{b}\right)} \le 20$. From Ex. 13.83, we can estimate σ^2 with MSE = 1014.9167 so that the solution is $b \ge 20.3$. Thus, at least 21 replications are required.

13.85 The design is completely randomized with five treatments, containing 4, 7, 6, 5, and 5 measurements respectively.

a. The analysis is as follows:

CM = $(20.6)^2/27 = 15.717$

TSS = 17,500 – CM = 1.783

SST = $\frac{(2.5)^2}{4} + \ldots + \frac{(2.4)^2}{5}$ – CM = 1.212, d.f. = 4

SSE = 1.783 – 1.212 = .571, d.f. = 22.

To test for difference in mean reaction times, $F = \frac{1.212/4}{.571/22} = 11.68$ with 4 numerator and 22 denominator degrees of freedom. From Table 7, p–value < .005.

b. The hypothesis is H_0: $\mu_A - \mu_D = 0$ versus a two–tailed alternative. The test statistic is

$$|t| = \frac{|.625-.920|}{\sqrt{.02596\left(\frac{1}{4}+\frac{1}{5}\right)}} = 2.73.$$

The critical value (based on 22 degrees of freedom) is $t_{.025} = 2.074$. Thus, H_0 is rejected. From Table 5, $2(.005) < p$–value $< 2(.01)$.

13.86 This is a RBD with people as blocks and stimuli as treatments. The ANOVA table is below.

Source	d.f	SS	MS	F
Treatments	4	.787	.197	27.7
Blocks	3	.140	.047	
Error	12	.085	.0071	
Total	19	1.012		

To test for a difference in the mean reaction times, the test statistic is $F = 27.7$ with 4 numerator and 12 denominator degrees of freedom. With $F_{.05} = 3.25$, we can reject the null hypothesis that the mean reaction times are equal.

13.87 Each interval should have confidence coefficient $1 - .05/4 = .9875 \approx .99$. Thus, with 12 degrees of freedom, we will use the critical value $t_{.005} = 3.055$ so that the intervals have a half width given by $3.055\sqrt{.0135\left(\frac{2}{4}\right)} = .251$. Thus, the intervals for the differences in means for the varieties are

$\mu_A - \mu_D$: $.320 \pm .251$ $\quad\quad$ $\mu_B - \mu_D$: $.145 \pm .251$
$\mu_C - \mu_D$: $.023 \pm .251$ $\quad\quad$ $\mu_E - \mu_D$: $-.124 \pm .251$

13.88

$$\begin{aligned}
\text{TSS} &= \sum_{j=1}^{b}\sum_{i=1}^{k}(Y_{ij}-\overline{Y})^2 = \sum_{j=1}^{b}\sum_{i=1}^{k}(Y_{ij}-Y_{i\bullet}+Y_{i\bullet}-Y_{\bullet j}+Y_{\bullet j}-\overline{Y}+\overline{Y}-\overline{Y})^2 \\
&= \sum_{j=1}^{b}\sum_{i=1}^{k}(\underline{Y_{i\bullet}-\overline{Y}}+\underline{Y_{\bullet j}-\overline{Y}}+\underline{Y_{ij}-Y_{i\bullet}-Y_{\bullet j}+\overline{Y}})^2 \qquad \leftarrow \text{expand as shown} \\
&= \sum_{j=1}^{b}\sum_{i=1}^{k}(Y_{i\bullet}-\overline{Y})^2+\sum_{j=1}^{b}\sum_{i=1}^{k}(Y_{\bullet j}-\overline{Y})^2+\sum_{j=1}^{b}\sum_{i=1}^{k}(Y_{ij}-Y_{i\bullet}-Y_{\bullet j}+\overline{Y})^2 \\
&\quad + \text{cross terms } (=C) \\
&= b\sum_{i=1}^{k}(Y_{i\bullet}-\overline{Y})^2+k\sum_{j=1}^{b}(Y_{\bullet j}-\overline{Y})^2+\sum_{j=1}^{b}\sum_{i=1}^{k}(Y_{ij}-Y_{i\bullet}-Y_{\bullet j}+\overline{Y})^2+C \\
&= \text{SST}+\text{SSB}+\text{SSE}+C.
\end{aligned}$$

So, it is only left to show that the cross terms are 0. They are expressed as

$$\begin{aligned}
C &= 2\sum_{j=1}^{b}(\overline{Y}_{\bullet j}-\overline{Y})\sum_{i=1}^{k}(\overline{Y}_{i\bullet}-\overline{Y}) && (1)\\
&\quad + 2\sum_{j=1}^{b}(\overline{Y}_{\bullet j}-\overline{Y})\sum_{i=1}^{k}(Y_{ij}-\overline{Y}_{i\bullet}-\overline{Y}_{\bullet j}+\overline{Y}) && (2)\\
&\quad + 2\sum_{i=1}^{k}(\overline{Y}_{i\bullet}-\overline{Y})\sum_{j=1}^{b}(Y_{ij}-\overline{Y}_{i\bullet}-\overline{Y}_{\bullet j}+\overline{Y}). && (3)
\end{aligned}$$

Part (1) is equal to zero since

$$\sum_{j=1}^{b}(\overline{Y}_{\bullet j}-\overline{Y}) = \sum_{j=1}^{b}\left(\tfrac{1}{k}\Sigma_i Y_{ij} - \tfrac{1}{bk}\Sigma_{ij}Y_{ij}\right) = \tfrac{1}{k}\Sigma_{ij}Y_{ij} - \tfrac{b}{bk}\Sigma_{ij}Y_{ij} = 0.$$

Part (2) is equal to zero since

$$\sum_{i=1}^{k}(Y_{ij} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j} + \overline{Y}) = \sum_{i=1}^{k}\left(Y_{ij} - \tfrac{1}{b}\Sigma_j Y_{ij} - \tfrac{1}{k}\Sigma_i Y_{ij} + \tfrac{1}{bk}\Sigma_{ij} Y_{ij}\right)$$
$$= \Sigma_i Y_{ij} - \tfrac{1}{b}\Sigma_{ij} Y_{ij} - \Sigma_i Y_{ij} + \tfrac{1}{b}\Sigma_{ij} Y_{ij} = 0.$$

A similar expansion will shown that part (3) is also equal to 0, proving the result.

13.89 **a.** We have that Y_{ij} and $Y_{ij'}$ are normally distributed. Thus, they are independent if their covariance is equal to 0 (recall that this only holds for the normal distribution). Thus,

$$\text{Cov}(Y_{ij}, Y_{ij'}) = \text{Cov}(\mu + \tau_i + \beta_j + \varepsilon_{ij}, \mu + \tau_i + \beta_{j'} + \varepsilon_{ij'}) = \text{Cov}(\beta_j + \varepsilon_{ij}, \beta_{j'} + \varepsilon_{ij'})$$
$$= \text{Cov}(\beta_j, \beta_{j'}) + \text{Cov}(\beta_j, \varepsilon_{ij'}) + \text{Cov}(\varepsilon_{ij}, \beta_{j'}) + \text{Cov}(\varepsilon_{ij}, \varepsilon_{ij'}) = 0,$$

by independence specified in the model. The result is similar for Y_{ij} and $Y_{i'j'}$.

b. $\text{Cov}(Y_{ij}, Y_{i'j}) = \text{Cov}(\mu + \tau_i + \beta_j + \varepsilon_{ij}, \mu + \tau_{i'} + \beta_j + \varepsilon_{i'j}) = \text{Cov}(\beta_j + \varepsilon_{ij}, \beta_j + \varepsilon_{i'j})$

$= V(\beta_j) = \sigma_B^2$, by independence of the other terms.

c. When $\sigma_B^2 = 0$, $\text{Cov}(Y_{ij}, Y_{i'j}) = 0$.

13.90 **a.** From the model description, it is clear that $E(Y_{ij}) = \mu + \tau_i$ and $V(Y_{ij}) = \sigma_B^2 + \sigma_\varepsilon^2$.

b. Note that $\overline{Y}_{i\bullet}$ is the mean of b independent observations in a block. Thus,

$$E(\overline{Y}_{i\bullet}) = E(Y_{ij}) = \mu + \tau_i \text{ (unbiased) and } V(\overline{Y}_{i\bullet}) = \tfrac{1}{b}V(Y_{ij}) = \tfrac{1}{b}(\sigma_B^2 + \sigma_\varepsilon^2).$$

c. From part b above, $E(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) = \mu + \tau_i - (\mu + \tau_{i'}) = \tau_i - \tau_{i'}$.

d. $V(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}) = V\left[\mu + \tau_i + \frac{1}{b}\sum_{j=1}^{b}\beta_j + \frac{1}{b}\sum_{i=1}^{b}\varepsilon_{ij} - \left(\mu + \tau_{i'} + \frac{1}{b}\sum_{j=1}^{b}\beta_j + \frac{1}{b}\sum_{i=1}^{b}\varepsilon_{i'j}\right)\right]$

$$= V\left[\frac{1}{b}\sum_{i=1}^{b}\varepsilon_{ij} - \frac{1}{b}\sum_{i=1}^{b}\varepsilon_{i'j}\right] = \frac{1}{b^2}V\left[\sum_{i=1}^{b}\varepsilon_{ij}\right] + \frac{1}{b^2}V\left[\sum_{i=1}^{b}\varepsilon_{i'j}\right] = \frac{2\sigma_\varepsilon^2}{b}.$$

13.91 First, $\overline{Y}_{\bullet j} = \frac{1}{k}\sum_{i=1}^{k}(\mu + \tau_i + \beta_j + \varepsilon_{ij}) = \mu + \frac{1}{k}\sum_{i=1}^{k}\tau_i + \beta_j + \frac{1}{k}\sum_{i=1}^{k}\varepsilon_{ij} = \mu + \beta_j + \frac{1}{k}\sum_{i=1}^{k}\varepsilon_{ij}$.

a. Using the above, $E(\overline{Y}_{\bullet j}) = \mu$ and $V(\overline{Y}_{\bullet j}) = V(\beta_j) + \frac{1}{k^2}\sum_{i=1}^{k}V(\varepsilon_{ij}) = \sigma_B^2 + \frac{1}{k}\sigma_\varepsilon^2$.

b. $E(\text{MST}) = \sigma_\varepsilon^2 + \left(\frac{b}{k-1}\right)\sum_{i=1}^{k}\tau_i^2$ as calculated in Ex. 13.51, since the block effects cancel here as well.

c. $E(\text{MSB}) = kE\left[\dfrac{\sum_{j=1}^{b}(\overline{Y}_{\bullet j} - \overline{Y})^2}{b-1}\right] = \sigma_\varepsilon^2 + k\sigma_B^2$

d. $E(\text{MSE}) = \sigma_\varepsilon^2$, using a similar derivation in Ex. 13.51(c).

13.92 **a.** $\hat{\sigma}^2_\varepsilon = \text{MSE}$.

b. $\hat{\sigma}^2_B = \dfrac{\text{MSB - MSE}}{k}$. By Ex. 13.91, this estimator is unbiased.

13.93 **a.** The vector $\boldsymbol{AY}$ can be displayed as

$$\boldsymbol{AY} = \begin{bmatrix} \dfrac{\sum_i Y_i}{\sqrt{n}} \\ \dfrac{Y_1 - Y_2}{\sqrt{2}} \\ \dfrac{Y_1 + Y_2 - 2Y_3}{\sqrt{2\cdot 3}} \\ \vdots \\ \dfrac{(Y_1 + Y_2 + \ldots + Y_{n-1} - (n-1)Y_n}{\sqrt{n(n-1)}} \end{bmatrix} = \begin{bmatrix} \sqrt{n}\overline{Y} \\ U_1 \\ U_2 \\ \vdots \\ U_{n-1} \end{bmatrix}$$

Then, $\sum_{i=1}^{n} Y_i^2 = \boldsymbol{Y'Y} = \boldsymbol{Y'A'AY} = n\overline{Y}^2 + \sum_{i=1}^{n-1} U_i$.

b. Write $L_i = \sum_{j=1}^{n} a_{ij} Y_j$, a linear function of $Y_1, \ldots, Y_n$. Two such linear functions, say L_i and L_k are pairwise orthogonal if and only if $\sum_{j=1}^{n} a_{ij} a_{kj} = 0$ and so L_i and L_k are independent (see Chapter 5). Let $L_1, L_2, \ldots, L_n$ be the n linear functions in $\boldsymbol{AY}$. The constants a_{ij}, $j = 1, 2, \ldots, n$ are the elements of the i^{th} row of the matrix $\boldsymbol{A}$. Moreover, if any two rows of the matrix $\boldsymbol{A}$ are multiplied together, the result is zero (try it!). Thus, $L_1, L_2, \ldots, L_n$ are independent linear functions of $Y_1, \ldots, Y_n$.

c. $\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2 = n\overline{Y}^2 + \sum_{i=1}^{n-1} U_i - n\overline{Y}^2 = \sum_{i=1}^{n-1} U_i$. Since U_i is independent of $\sqrt{n}\overline{Y}$ for $i = 1, 2, \ldots, n - 1$, $\sum_{i=1}^{n} (Y_i - \overline{Y})^2$ and $\overline{Y}$ are independent.

d. Define

$$W = \frac{\sum_{i=1}^{n} (Y_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y} + \overline{Y} - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{\sigma^2} + \frac{n(\overline{Y} - \mu)^2}{\sigma^2} = X_1 + X_2.$$

Now, W is chi–square with n degrees of freedom, and X_2 is chi–square with 1 degree of freedom since $X_2 = \left(\dfrac{\overline{Y} - \mu}{\sigma/\sqrt{n}}\right)^2 = Z^2$. Since X_1 and X_2 are independent (from part **c**), we can use moment generating functions to show that

$$(1 - 2t)^{-n/2} = m_W(t) = m_{X_1}(t) m_{X_2}(t) = m_{X_1}(t)(1 - 2t)^{-1/2}.$$

Thus, $m_{X_1}(t) = (1 - 2t)^{-(n-1)/2}$ and this is seen to be the mgf for the chi–square distribution with $n - 1$ degrees of freedom, proving the result.

13.94 **a.** From Section 13.3, SSE can be written as $\text{SSE} = \sum_{i=1}^{k}(n_i - 1)S_i^2$. From Ex. 13.93, each $\overline{Y}_i$ is independent of $S_i^2 = \sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2$. Therefore, since the *k* samples are independent, $\overline{Y}_1, \ldots, \overline{Y}_k$ are independent of SSE.

b. Note that SST = $\sum_{i=1}^{k} n_i(\overline{Y}_i - \overline{Y})^2$, and $\overline{Y}$ can be written as

$$\overline{Y} = \frac{\sum_{i=1}^{k} n_i \overline{Y}_i}{n}.$$

Since SST can be expressed as a function of only $\overline{Y}_1, \ldots, \overline{Y}_k$, by part (a) above we have that SST and SSE are independent. The distribution of $F = \frac{\text{MST}}{\text{MSE}}$ was derived in Ex. 13.6.