$$\chi^2_s = \frac{(33-25)^2}{25} + \frac{(17-25)^2}{25} = 5.12$$

Using Table 9 with $df = 1$, the directional P-value is between 0.02 and 0.05. Thus, there is evidence that the two rats tend to agree with each other.

III.9 The 200 observations are not independent of one another. When considering the first pair of rats (as in Exercise III.7) we can think of the 50 observations as being 50 independent measurements of how well those two rats agree. But when we combine the data across four pairs of connected rats, we cannot think of the 200 observations as being 200 draws from a fixed population. Rather, we have four sets of 50 draws from (potentially) four populations.

• **III.10** (a) True. The standard error of the estimated proportion is largest when $p = 0.50$.

(b) False. A goodness-of-fit test can be conducted for any null hypothesis that specifies probabilities for each of the possible categories of a response variable, but there is no need for those probabilities to be equal.

(c) False. Either a goodness-of-fit test or a test of independence can be conducted for observational data or for experimental data.

(d) True. If the observed data perfectly agree with the expected values from the null hypothesis, then each term in the chi-square calculation will be zero, so the test statistic will be zero.

CHAPTER 11

Comparing the Means of Many Independent Samples

• **11.2.1** We have $n_. = 4 + 3 + 4 = 11$;

$$\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij} = 48 + 39 + 42 + 43 + 40 + 48 + 44 + 39 + 30 + 32 + 35 = 440;$$

$$\bar{\bar{y}} = \frac{440}{11} = 40.$$

(a) $SS(\text{between}) = (4)(43-40)^2 + (3)(44-40)^2 + (4)(34-40)^2 = 228;$

$$SS(\text{within}) = (48-43)^2 + (39-43)^2 + (42-43)^2 + (43-43)^2 + (40-44)^2 + (48-44)^2 + (44-44)^2 + (39-34)^2 + (30-34)^2 + (32-34)^2 + (35-34)^2 = 120.$$

(b) $$SS(\text{total}) = (48-40)^2 + (39-40)^2 + (42-40)^2 + (43-40)^2 + (40-40)^2 + (48-40)^2 + (44-40)^2 + (39-40)^2 + (30-40)^2 + (32-40)^2 + (35-40)^2 = 348.$$

Verification: SS(between) + SS(within) = SS(total);
228 + 120 = 348.

(c) $df(\text{between}) = I - 1 = 3 - 1 = 2$; $MS(\text{between}) = \frac{SS(\text{between})}{df(\text{between})} = \frac{228}{2} = 114;$

$df(\text{within}) = n_. - I = 11 - 3 = 8$; $MS(\text{within}) = \frac{SS(\text{within})}{df(\text{within})} = \frac{120}{8} = 15;$

$s_{pooled} = \sqrt{MS(\text{within})} = \sqrt{15} = 3.87.$

11.2.2 We have $n_{..} = 12$, $\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij} = 240$, and $\bar{\bar{y}} = 240/12 = 20$.

(a) $SS(\text{between}) = (4)(25-20)^2 + (3)(15-20)^2 + (5)(19-20)^2 = 180;$

$SS(\text{within}) = (23-25)^2 + (29-25)^2 + \ldots + (19-19)^2 = 72.$

(b) $SS(\text{total}) = (23-20)^2 + (29-20)^2 + \ldots + (19-20)^2 = 252;$

SS(between) + SS(within) = 180 + 72 = 252 = SS(total).

(c) df(between) = 2; MS(between) = 180/2 = 90;

df(within) = 9; MS(within) = 72/9 = 8;

$s_{pooled} = \sqrt{8} = 2.83$.

11.2.3 (a) SS(between) = SS(total) - SS(within) = 338.769 - 116 = 222.769.

(b) df(within) = 10; MS(within) = 116/10 = 11.6;

df(between) = 2; MS(between) = 222.769/2 = 111.3845.

$s_{pooled} = \sqrt{11.6} = 3.41$.

• **11.2.4** (a) We find SS(between) by subtraction: SS(between) = 472 - 337 = 135.

We find df(total) by adding df(between) and df(within): df(total) = 3 + 12 = 15.

We find MS(within) by division: MS(within) = SS(within)/df(within) = 337/12 = 28.08.

The completed table is

Source	df	SS	MS
Between groups	3	135	45
Within groups	12	337	28.08
Total	15	472	

(b) We have df(between) = 3 = I - 1, so I = 4.

(c) We have df(total) = 15 = n. - 1, so n. = 16.

11.2.5 (a)

Source	df	SS	MS
Between groups	4	159	39.75
Within groups	49	964	19.67
Total	53	1123	

(b) We have df(between) = 4 = I - 1, so I = 5.

(c) We have df(total) = 53 = n. - 1, so n. = 54.

11.2.6 (a)

Source	df	SS	MS
Between groups	3	258	86.00
Within groups	26	640	24.62
Total	29	898	

(b) We have df(between) = 3 = I - 1, so I = 4.

(c) We have df(total) = 29 = n. - 1, so n. = 30.

11.2.7 There are no single correct answers. Typical answers are:

(a)

	Sample		
	1	2	3
	1	2	3
	2	2	3
	3	3	3
	4	4	3
	5	4	3
$\bar{y}$	3	3	3

(b)

	Sample		
	1	2	3
	2	5	8
	2	5	8
	2	5	8
	2	5	8
	2	5	8
$\bar{y}$	2	5	8

11.4.1 (a) It appears that group I is shifted up from the others and that H_0 is false.

(b) There are 2 numerator and 39 denominator degrees of freedom.

(c) $F_s = 68.06/10.72 = 6.35$

(d) The hypotheses are

H_0: Mean MAO is the same for all three diagnoses ($\mu_1 = \mu_2 = \mu_3$)
H_A: Mean MAO is not the same for all three diagnoses (the μ's are not all equal)

We reject H_0 because the P-value is smaller than 0.05. We have strong evidence (P=0.004) that mean MAO in not the same for all three diagnoses.

(e) $s_{pooled} = \sqrt{418.25/39} = \sqrt{10.72} = 3.27$.

• **11.4.2** **(a)** The hypotheses are

H_0: The stress conditions all produce the same mean lymphocyte concentration ($\mu_1 = \mu_2 = \mu_3 = \mu_4$)
H_A: Some of the stress conditions produce different mean lymphocyte concentrations (the μ's are not all equal)

The number of groups is I = 4 and the total number of observations is n.= 48. Thus, df(between) = I - 1 = 3 and df(within) = n.- I = 44.

Source	df	SS	MS
Between groups	3	89.036	29.68
Within groups	44	340.24	7.733
Total	47	429.28	

The test statistic is $F_s = \frac{\text{MS(between)}}{\text{MS(within)}} = \frac{29.68}{7.733} = 3.84$. With df = 3 and 40 (the closest value to 44), Table 10 gives $F_{0.02} = 3.67$ and $F_{0.01} = 4.31$. Thus, we have $0.01 < P < 0.02$. Since $P < \alpha$, we reject H_0. There is sufficient evidence ($0.01 < P < 0.02$) to conclude that some of the stress conditions produce different mean lymphocyte concentrations.

(b) $s_{pooled} = \sqrt{\text{MS(within)}} = \sqrt{7.733} = 2.78$ cells/ml x 10^{-6}.

• **11.4.3** **(a)** The null hypothesis is

H_0: Mean HBE is the same in all three populations

(b) In symbols, the null hypothesis is

H_0: $\mu_1 = \mu_2 = \mu_3$

(c) The number of groups is I = 3 and the total number of observations is n. = 36. Thus, df(between) = I - 1 = 2 and df(within) = n. - I = 33.

Source	df	SS	MS
Between groups	2	240.69	120.35
Within groups	33	6887.6	208.72
Total	35	7128.3	

The test statistic is $F_s = \frac{\text{MS(between)}}{\text{MS(within)}} = \frac{120.35}{208.72} = 0.58$. With df = 2 and 30 (the closest value to 33), Table 10 gives $F_{0.20} = 1.70$. Thus, we have $P > 0.20$. Since $P > \alpha$, we do not reject H_0. There is insufficient evidence ($P > 0.20$) to conclude that mean HBE is not the same in all three populations.

(d) $s_{pooled} = \sqrt{\text{MS(within)}} = \sqrt{208.7} = 14.4$ pg/ml.

11.4.4 **(a)** The null hypothesis is

H_0: Mean time until alleviation of symptoms is the same in all three populations

(b) In symbols, the null hypothesis is

H_0: $\mu_1 = \mu_2 = \mu_3$

(c) There are 3 numerator and 259 denominator degrees of freedom.

(d) We reject H_0. There is sufficient evidence (P = 0.034) to conclude that mean time until alleviation of symptoms is not the same in all three populations.

(e) $s_{pooled} = \sqrt{\text{MS(within)}} = \sqrt{7.885} = 2.80$.

11.4.5 **(a)** The dotplots give the impression of differences between the five groups. In particular, the Open group appears to have a smaller mean than the others.

(b) In symbols, the null hypothesis is

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

(c) I = 5, n. = 65.

Source	df	SS	MS
Between groups	4	871.408	217.852
Within groups	60	3588.54	59.809
Total	64	4459.948	

The test statistic is $F_s = 217.852/59.809 = 3.64$. With df = 4 and 60, Table 10 gives $F_{0.02} = 3.16$ and $F_{0.01} = 3.65$. Thus, we have $0.01 < P < 0.02$, so we reject H_0. There is sufficient evidence ($0.01 < P < 0.02$) to conclude that mean stem lengths of daffodils is not the same in all five locations.

11.4.6 **(a)** The dotplots show differences between the sample means, with the modern dance mean being higher than the other two.

(b) In symbols, the null hypothesis is

H_0: $\mu_1 = \mu_2 = \mu_3$

(c) I = 3, n. = 29.

Source	df	SS	MS
Between groups	2	7.04	3.52
Within groups	26	15.08	0.58
Total	28	22.12	

The test statistic is $F_s = 3.52/0.58 = 6.07$. With df = 2 and 26, Table 10 gives $F_{0.01} = 5.53$ and $F_{0.001} = 9.12$. Thus, we have $0.001 < P < 0.01$, so we reject H_0. There is strong evidence ($0.001 < P < 0.01$) to conclude that flexibility differs depending on the group (aerobic class, modern dance class, or control) that one is in.

11.4.7 **(a)** I = 3.

(b) H_0 is not rejected. There is insufficient evidence (P = 0.68) to conclude that the (population) mean yields of the varieties differ from one another.

(c) $s_{pooled} = \sqrt{95.5432} = 9.77$.

11.4.8 **(a)** The data in plot II will have a larger value of F_S. While the within group variability is similar for the data in both plots, the between group variability is much larger in plot II.

(b) The data in plot I will have a larger value of F_S. While the between group variability is similar across the groups in the two plots, the within group variability is smaller in plot I.

11.4.9 **(a)** The null hypothesis is that average height of a radish seedling after 9 days is the same for any of the four lighting conditions used in this experiment..

(b) There is very strong evidence ($P = 2.9 \times 10^{-12}$) that the four lighting conditions result in different average growth.

(c) Yes. This was an experiment so we can speak in terms of causation.