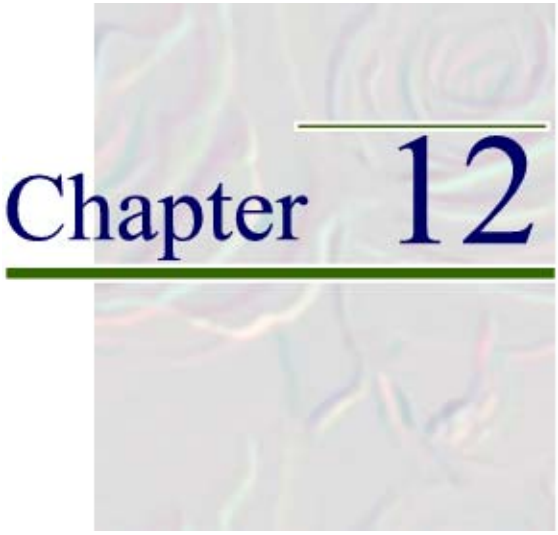


MULTISAMPLE INFERENCE

REVIEW OF KEY CONCEPTS

SECTION 12.1 General Concepts

In Chapter 8, we studied how to compare the means of two normally distributed samples using *t* test methods. In this chapter, we extend these methods to allow us to compare the means of more than two normally distributed samples.

Consider the data given in Table 8.4, Chapter 8, where we looked at the relationship between systolic blood pressure and alcohol consumption in a group of 30–39-year-old women not using oral contraceptives. The data are reproduced here:

	Systolic blood pressure (mm Hg)		
	Mean	sd	*n*
A. No alcohol use	110.5	13.3	357
B. ≤ 10 oz per week alcohol consumption	109.1	13.4	440
C. > 10 oz per week alcohol consumption	114.5	14.9	23

We could compare the means of each pair of groups, as we did in Problems 8.22–8.29. There are two important reasons for not doing so. If we have many ($k > 2$) groups and we compare the means of each pair of groups, then we will make a total of ${}_kC_2$ comparisons. Even for moderate k, ${}_kC_2$ is quite large and there is a high probability that some of the comparisons will be significant by chance alone. One approach to this problem is to use *one-way analysis-of-variance methods*. Using these methods, we will test the hypothesis $H_0: \mu_1 = \ldots = \mu_k$ versus H_1: at least two of the μ_i's are different, where μ_i is the underlying mean of the *i*th group. Only if we are able to reject H_0 and conclude that at least some of the means are different will we perform *t* tests comparing pairs of groups.

A second difference from the methods in Chapter 8 is that the *t* tests conducted in the context of the analysis of variance are different from ordinary two-sample *t* tests. First, the variance estimate used is based on a pooled-variance estimate over all groups, rather than simply the two groups being compared. Second, an explicit adjustment for the number of comparisons is sometimes made using what are known as *multiple-comparisons procedures*. We will study one popular method for adjusting for multiple comparisons known as the *Bonferroni method*.

Finally, there are two types of analysis-of-variance models, called *fixed-effects* and *random-effects models*. Under a fixed-effects model, we are primarily interested in comparing specific groups as in Table 8.4. Under a random-effects model, we are interested in determining what proportion of the total variation is attributable to between-group versus within-group variability. For example, in a dietary reproducibility study, where the same dietary instrument is administered on more than one occasion, we

might be interested in comparing the degree of between-person versus within-person variation for a specific nutrient (e.g., dietary cholesterol), but not in comparing dietary cholesterol between specific people.

SECTION 12.2 One-Way Analysis of Variance—Fixed-Effects Model

Let $y_{ij} = j$ th observation in the ith group; $i = 1, \ldots, k$, $j = 1, \ldots, n_i$.

The one-way analysis of variance fixed effects model is given by:

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

where

μ = overall mean

α_i = difference between mean of i th group and overall mean

$e_{ij} \sim N(0, \sigma^2)$

Differences among the α_i represent between-group variation, The e_{ij} represent within-group variation.

12.2.1 Overall Test of Significance

We with to test the hypothesis H_0: all $\alpha_i = 0$ versus H_1: some $\alpha_i \neq 0$. To test this hypothesis, we separate the total variation into a between-group and within-group component of variation. This can be written algebraically as

$$\sum_{ij} \left(y_{ij} - \bar{\bar{y}}\right)^2 = \sum_{ij} \left(y_{ij} - \bar{y}_i\right)^2 + \sum_{i} n_i \left(\bar{y}_i - \bar{\bar{y}}\right)^2$$

The terms in this equation are called *sums of squares* (*SS*), and the equation can be written in words as follows:

$$\text{Total SS} = \text{Within SS} + \text{Between SS}$$

The Within SS represents within-group variation; the Between SS represents between-group variation.

Another way to write the Within SS is

$$\text{Within SS} = \sum_{i=1}^{k} (n_i - 1) s_i^2$$

where s_i^2 is the sample variance for the ith sample.

Another way to write the Between SS is

$$\text{Between SS} = \sum_{i=1}^{k} n_i \bar{y}_i^2 - \frac{\left(\sum_{i=1}^{k} n_i \bar{y}_i\right)^2}{\sum_{i=1}^{k} n_i}$$

To test the overall hypothesis that at least two of the group means are different, we use the test statistic

$$F = \frac{\text{Between MS}}{\text{Within MS}} \sim F_{k-1,\, n-k} \text{ under } H_0$$

where MS stands for mean square and

$$\text{Between MS} = \frac{\text{Between SS}}{k-1}$$
$$\text{Within MS} = \frac{\text{Within SS}}{n-k}$$

The values $k-1$ and $n-k$ are referred to as the *df* pertaining to the Between-group and Within-group components of variation. We will reject H_0 if F is large and accept H_0 if F is small. The p-value $= Pr\left(F_{k-1,\, n-k} > F\right)$. The results are typically displayed in an analysis-of-variance (or ANOVA) table as follows:

ANOVA Table

	SS	df	MS	F-stat	p-value
Between					
Within					
Total					

Example: Determine if there are significant differences for mean SBP among the 3 alcohol-use groups. The overall mean SBP over all groups is

$$\frac{110.5(357) + 109.1(440) + 114.5(23)}{820} = 109.86\,.$$

Therefore,

$$\text{Between SS} = 357(110.5 - 109.86)^2 + \ldots + 23(114.5 - 109.86)^2 = 895.55$$
$$\text{Between MS} = \frac{895.55}{2} = 447.78$$
$$\text{Within SS} = 356(13.3)^2 + \ldots + 22(14.9)^2 = 146{,}683.9$$
$$\text{Within MS} = \frac{146{,}683.9}{820 - 3} = 179.54$$
$$F = \frac{447.78}{179.54} = 2.49 \sim F_{2,\,817} \text{ under } H_0$$
$$p\text{-value} = Pr\left(F_{2,\,817} > 2.49\right) = .083$$

Thus, there is no significant difference among the means of the 3 groups. This is the same conclusion as was reached in Problems 8.25–8.29, where groups A vs. B and A vs. C were compared. Under the analysis of variance, if the overall F test is not significant as was the case here, then one is not justified in looking at any 2 group comparisons, i.e., all 2 group comparisons are considered not significantly different. If the overall F test is significant, then one is justified in looking at which pairs of groups are significantly different as is explained in detail in the next three sections.

SECTION 12.3 Comparisons of Specific Groups—Least Significant Difference (LSD) Approach

Suppose we consider the data in Table 8.9 (Chapter 8). We wish to compare the mean blood-lead levels at 12 months among infants in the low, medium, and high cord blood-lead-level groups.

Cord blood-lead group	12-month blood-lead level mean ± sd	*n*
Low	5.8 ± 5.1	69
Medium	8.5 ± 7.6	70
High	8.8 ± 6.4	60

We first perform the overall *F* test for one-way ANOVA and obtain an *F* statistic

$$\frac{183.04}{41.69} = 4.39 \sim F_{2,\,196} \text{ under } H_0 .$$

The *p*-value $= Pr\left(F_{2,\,196} > 4.39\right) = .014$. Thus, we reject H_0 and conclude that some of the group means are significantly different. To compare specific groups-for example, the 1st and 2nd groups—we test the hypothesis $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$ and compute the test statistic

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\text{Within MS}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n-k} \text{ under } H_0$$

The *p*-value $= 2 \times Pr(t_{n-k} > t)$ if $t \geq 0$, or $= 2 \times Pr(t_{n-k} < t)$ if $t < 0$.

In the blood-lead example, we compare each pair of groups as follows:

Groups compared	Test statistic	*p*-value
Low, medium	$t = \frac{5.8 - 8.5}{\sqrt{41.69(\frac{1}{69} + \frac{1}{70})}} = \frac{-2.7}{1.10} = -2.47 \sim t_{196}$	.015
Low, high	$t = \frac{5.8 - 8.8}{\sqrt{41.69(\frac{1}{69} + \frac{1}{60})}} = \frac{-3.0}{1.14} = -2.63 \sim t_{196}$	.009
Medium, high	$t = \frac{8.5 - 8.8}{\sqrt{41.69(\frac{1}{70} + \frac{1}{60})}} = \frac{-0.3}{1.14} = -0.26 \sim t_{196}$	.79

Thus, there are significant differences between the low group and each of the medium and high groups, while there is no significant difference between the medium and high groups.

SECTION 12.4 Comparisons of Specific Groups—Bonferroni Approach

Some authors feel that first performing the overall *F* test for one-way ANOVA is not sufficient protection against the multiple-comparisons problem. They recommend that an additional adjustment for multiple comparisons be performed when comparing specific groups. One of the most commonly used such methods is the *Bonferroni approach*. The goal is that the *experiment-wise type I error*, which is defined as the probability that at least one pair of groups will be significantly different under H_0, be set at α.

Under this approach, if we wish to achieve an experiment-wise type I error $= \alpha$, then we should use a critical value $= t_{n-k,\,1-\alpha^*/2}$, where

$$\alpha^* = \frac{\alpha}{\binom{k}{2}}$$

for comparisons between specific pairs of groups where $k =$ number of groups. The rationale for the definition of α^* is that there are $\binom{k}{2}$ possible 2 group comparisons among the *k* groups. To conduct the significance test, we reject H_0 if $|t| > t_{n-k,\, 1-\alpha^*/2}$, and accept H_0 otherwise.

Suppose we apply this approach to the blood-lead example. If we set $\alpha = .05$, then since $k = 3$, $\binom{k}{2} = \binom{3}{2} = 3$ and $\alpha^* = .05/3 = .0167$ and $1 - \alpha^*/2 = 1 - .0167/2 = .99167$. Thus, the critical value $= t_{196,\, .99167} = 2.41$ by computer. Since $|t| = 2.47$ for low vs. medium and $|t| = 2.63$ for low vs. high, and both 2.47 and 2.63 are greater than 2.41, these contrasts would remain statistically significant even under the Bonferroni approach.

In general, it is harder to achieve statistical significance with the Bonferroni approach than with the LSD approach, particularly if the number of groups *k* is large. At the present time, it is controversial as to which approach is preferable.

SECTION 12.5 Linear Contrasts

More complex comparisons are possible in one-way ANOVA. Suppose we wish to compare the mean blood-lead level at 12 months for low cord blood-lead children with the mean blood-lead level at 12 months for the combined groups of medium and high blood-lead children. Assume that there are an equal number of medium and high blood-lead children, and that we denote the true mean blood-lead level at 12 months for the low group by μ_1, the medium group by μ_2, and the high group by μ_3. We can form a *linear contrast* $L = \bar{y}_1 - 0.5(\bar{y}_2 + \bar{y}_3)$ to represent this comparison and test the hypothesis $H_0: \mu(L) = 0$ versus $H_1: \mu(L) \neq 0$, where $\mu(L) = \mu_1 - 0.5(\mu_2 + \mu_3)$.

In general, a *linear contrast* can be written in the form $L = \sum_{i=1}^{k} \lambda_i \bar{y}_i$, where $\sum_{i=1}^{k} \lambda_i = 0$. To test the preceding hypothesis, we compute the test statistic

$$t = \frac{L}{se(L)} = \frac{L}{\sqrt{\text{Within MS}\left(\sum_{i=1}^{k} \frac{\lambda_i^2}{n_i}\right)}} \sim t_{n-k} \text{ under } H_0$$

The *p*-value $= 2 \times Pr(t_{n-k} > t)$ if $t > 0$ or $= 2 \times Pr(t_{n-k} < t)$ if $t \leq 0$.

In this case,

$$L = 5.8 - 0.5(8.5 + 8.8) = -2.85$$

$$se(L) = \sqrt{41.69\left(\frac{1}{69} + \frac{0.5^2}{70} + \frac{0.5^2}{60}\right)} = 0.963$$

$$t = \frac{-2.85}{0.963} = -2.96 \sim t_{196} \text{ under } H_0$$

$$p\text{-value} = 2 \times Pr(t_{196} < -2.96) = .003$$

Thus, the children with low cord blood-lead levels have significantly lower mean blood-lead levels at 12 months than the children with medium or high cord blood-lead levels.

A multiple comparisons procedure also exists for testing the significance of linear contrasts. Under *Scheffe's procedure*, the experiment-wise type I error is preserved at α, where the experiment-wise type I error is defined as the probability under H_0, that at least one linear contrast, among all possible linear contrasts, will be declared statistically significant (see section 12.4.4, text, Chapter 12).

SECTION 12.6 One-Way Analysis of Variance—Random-Effects Model

In the blood-lead example, we were interested in comparing specific groups, e.g., low versus high cord blood-lead groups. In some instances, we are interested in determining what proportion of the total variation is attributable to between-group versus within-group variation, but are not specifically interested in comparing individual groups.

A dietary reproducibility study was performed among 173 U.S. nurses. Each woman filled out a *diet record* on four separate weeks, approximately 3 months apart, over a 1-year period. The diet record consists of an enumeration of the type and quantity of each food eaten over a 1-week period. The question is, How reproducible are the reported intakes of specific nutrients such as dietary cholesterol?

To assess between- and within-person variability, we use the model

$$y_{ij} = \mu + \alpha_i + e_{ij},\ i = 1,\ \ldots,\ k = \text{person};\ j = 1,\ \ldots,\ n = \text{week}$$

and assume $\alpha_i \sim N(0,\ \sigma_A^2)$, $e_{ij} \sim N(0,\ \sigma^2)$. σ_A^2 and σ^2 are referred to as the between-person and within-person variance components. σ_A^2 is a measure of the true variation between different people on dietary cholesterol. σ^2 is a measure of the week-to-week variation in dietary cholesterol for an individual. To estimate the variance components, we run a one-way ANOVA and estimate the variance components by

$$\hat{\sigma}^2 = \text{Within MS}$$
$$\hat{\sigma}_A^2 = \max\left(\frac{\text{Between MS} - \text{Within MS}}{n},\ 0\right)$$

The following data were obtained for the 1st day of intake of each of 4 weeks for dietary cholesterol:

	SS	*df*	MS	*F*-stat	*p*-value
Between	936.7	172	5.446	1.59	< .001
Within	1780.2	519	3.430		

Thus, as expected, there is a significant difference in mean dietary cholesterol between different individuals. The estimated within-person variance $= \hat{\sigma}^2 = 3.430$. The estimated between-person variance $= \hat{\sigma}_A^2 = (5.446 - 3.430)/4 = 0.504$. Thus, the within-person variation is six times as large as the between-person variation.

A useful measure of reproducibility is the *intraclass correlation coefficient*, which is defined by

$$\rho_I = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2}$$

This measure can be interpreted as the proportion of the total variance that is attributable to true between-person variation. For a very reproducible measure, ρ_I will be near 1, while for a very noisy measure, ρ_I will be near 0. For dietary cholesterol, ρ_I is estimated by $0.504/(0.504 + 3.430) = 0.504/3.934 = .13$. This is very low and indicates a lot of within-person variation relative to between-person variation for dietary cholesterol. Only about 13% of the total variation is attributable to true between-person variation.

The one-way ANOVA random-effects model can also be extended to the case of an unequal number of replicates per subject, such as might occur if some individuals did not complete all 4 weeks of diet records. See Section 12.8 (Chapter 12 in the main text) for a discussion of this topic.

SECTION 12.7 The Kruskal-Wallis Test

In the Data Set BETACAR.DAT (in the text) data are provided on the bioavailability of four different preparations of beta carotene (Solatene, Roche, BASF-30, and BASF-60). Serum beta-carotene levels were measured at baseline (two replicates) and at 6, 8, 10, and 12 weeks of follow-up. There were 5–6 subjects for each of the four preparations. Different subjects were used for the testing of each preparation. To measure bioavailability, we compute

$$x = \text{mean (6, 8-, 10-, and 12-week levels)/mean (2 baseline levels)}$$

The results are displayed here:

Subject	SOL	ROCHE	BASF_30	BASF_60
1	0.91787	1.98214	2.45614	1.27528
2	1.99630	1.12273	1.39323	2.33745
3	1.60106	1.38500	2.09024	2.14773
4	1.58108	1.44565	1.68466	1.15605
5	1.91837	1.79009	2.22024	3.43939
6	2.62264	1.82796		2.68493

The distributions are somewhat skewed. If we wish to compare the four preparations, then we could transform x to a different scale (e.g., the ln scale) and perform the fixed-effects one-way ANOVA on the transformed scale. Another approach is to leave the data in the original scale, but perform a nonparametric test to compare the four groups. The nonparametric analogue to the one-way ANOVA is called the *Kruskal-Wallis test*. We wish to test the hypothesis H_0: average rank (x, group 1) = ... = average rank (x, group 4) versus H_1: at least two of the average ranks are different. The test statistic (in the absence of ties) is

$$H = \frac{12}{N(N+1)} \times \left(\sum_{i=1}^{k} \frac{R_i^2}{n_i} \right) - 3(N+1) \sim \chi^2_{k-1} \text{ under } H_0$$

where R_i is the rank sum in the ith group, and the ranks are based on the pooled sample over all groups

$$n_i = \text{sample size in the } i\text{th group}$$
$$N = \sum n_i$$
$$k = \text{number of groups}$$

If there are ties, then the test statistic is divided by

$$1 - \frac{\sum_{j=1}^{R} \left(t_j^3 - t_j \right)}{N^3 - N}$$

where t_j = the number of observations in the jth group of tied values. The p-value $= Pr\left(\chi^2_{k-1} > H \right)$. This test should only be used if the minimum group size is ≥ 5.

Using MINITAB we have applied this procedure to the beta-carotene data, with results given as follows:

MTB > Kruskal-Wallis C9 C1.

Group	Nobs	Median	Avg. Rank
1	6	1.760	11.2
2	6	1.618	8.5
3	5	2.090	14.0
4	6	2.243	14.7
Overall	23		12.0
H = 3.05	d.f . = 3	p = 0.384	

We see that the beta-carotene level has increased in all groups (median increase = 76%, solatene group (group 1); 62%, Roche group (group 2); 109%, BASF-30 group (group 3); 124%, BASF-60 group (group 4)). The test statistic $= H = 3.05 \sim \chi^2_3$. The p-value $= Pr(\chi^2_3 > 3.05) = .384$. Thus, there was no significant difference in bioavailability among the four preparations.

If the Kruskal-Wallis test had been significant, then we could compare specific groups using the Dunn procedure [see Equation 12.28, Chapter 12, text]. Finally, if at least one of the group sizes is < 5, then special small-sample tables must be used to evaluate significance. In Table 15 (Appendix, text) we provide such a table for selected group sizes in the special case of 3 groups.

PROBLEMS

Nutrition, Arthritis
A comparison was made of mean protein intake among three groups of premenopausal women: (1) women eating a standard American diet (STD), (2) women eating a lacto-ovo-vegetarian diet (LAC), and (3) women eating a strict vegetarian diet (VEG). The mean and sd for protein intake (mg) are presented in Table 12.1.

Table 12.1 Protein intake (mg) among three dietary groups of premenopausal women

Group	Mean	sd	*n*
STD	74	16	10
LAC	56	16	10
VEG	55	9	10

12.1 What parametric procedure can be used to compare the underlying means of the three groups?

12.2 Implement the procedure in Problem 12.1 using the critical-value method.

12.3 Compare the underlying means of each specific pair of groups using the *t* test (LSD) methodology.

12.4 Suppose that in the general population 70% of vegetarians are lacto-ovo-vegetarians, whereas 30% are strict vegetarians. Perform a statistical procedure to test if the contrast $L = 0.7\bar{y}_2 + 0.3\bar{y}_3 - \bar{y}_1$ is significantly different from 0. What does this contrast mean?

12.5 Using the data in Table 12.1, perform a multiple-comparisons procedure to identify which specific underlying means are different.

12.6 Perform a multiple-comparisons procedure to test if the linear contrast in Problem 12.4 is significantly different from 0.

Cardiovascular Disease
Physical activity has been shown to have beneficial effects on cardiovascular-disease outcomes. As part of a study of physical-activity assessment methodology, the number of hours of sleep, light activity, moderate activity, hard activity, and very hard activity were computed for each person in a study [1]. These categories were defined in terms of levels of energy, or MET, where MET = ratio of working metabolic rate to resting metabolic rate. Using this classification, sleep = 1 MET; light activity = 1.1–2.9 MET; moderate activity = 3.0–5.0 MET; hard activity = 5.1–6.9 MET; very hard activity ≥ 7.0 MET. In Table 12.2, data relating the number of hours of moderate activity per week to age for males are presented.

12.7 What is the appropriate method of analysis to test for the effect of age on the mean number of hours of moderate activity per week?

12.8 Perform the test mentioned in Problem 12.7 and report a *p*-value.

12.9 Comment on which specific age groups are different based on the data in Table 12.2.

12.10 Perform a test for whether or not there is a general trend for increasing or decreasing amounts of moderate activity by age based on the data in Table 12.2.

Table 12.2 Hours per week of moderate activity for men by age

Age group											
20–34			35–49			50–64			65–74		
Mean	sd	*n*	Mean	sd	*n*	Mean	sd	*n*	Mean	sd	*n*
8.1	10.4	487	9.7	10.2	233	7.9	10.2	191	5.8	10.3	82

Source: Reprinted with permission from the *American Journal of Epidemiology*, **121**(1), 91–106,1985.

Table 12.3 Relationship between protein concentration (mg/mL) of duodenal secretions to pancreatic function as measured by trypsin secretion (u/(kg/hr))

Trypsin secretion (u/(kg/hr))					
< 50		51–1000		>1000	
Subject number	Protein concentration	Subject number	Protein concentration	Subject number	Protein concentration
1	1.7	1	1.4	1	2.9
2	2.0	2	2.4	2	3.8
3	2.0	3	2.4	3	4.4
4	2.2	4	3.3	4	4.7
5	4.0	5	4.4	5	5.0
6	4.0	6	4.7	6	5.6
7	5.0	7	6.7	7	7.4
8	6.7	8	7.6	8	9.4
9	7.8	9	9.5	9	10.3
		10	11.7		

Source: Reprinted with permission of the *New England Journal of Medicine*, **312**(6), 329–334, 1985.

Gastroenterology

A 1985 study was performed focusing on the protein concentration of duodenal secretions from patients with cystic fibrosis [2]. Table 12.3 provides data relating protein concentration to pancreatic function as measured by trypsin secretion.

12.11 What statistical procedure can be used to compare the protein concentration among the three groups?

12.12 Perform the test mentioned in Problem 12.11 and report a *p*-value.

12.13 Compare pairs of specific groups as to mean protein concentration using the LSD procedure.

12.14 Provide a 95% CI for the mean difference in protein concentration between each pair of specific groups.

Pulmonary Function

In the same study referred to in Example 12.1 (in Chapter 12, text), the authors also obtained other measures of pulmonary function on the 1050 men. In particular, the FEV data for these men are presented in Table 12.4.

12.15 Are the mean FEVs different overall among the six groups?

12.16 Analyze the data for between-group differences using the conventional *t* test procedure.

12.17 Analyze the data for between-group differences using a multiple-comparisons procedure.

Table 12.4 FEV data for smoking and nonsmoking males in the White and Froeb study [3]

Group number (l)	Group name	Mean FEV (L)	sd FEV (L)	n_i
1	NS	3.72	0.65	200
2	PS	3.54	0.61	200
3	NI	3.56	0.76	50
4	LS	3.49	0.62	200
5	MS	3.08	0.61	200
6	HS	2.77	0.60	200
Overall		3.331		1050

Source: Reprinted with permission of the *New England Journal of Medicine*, **302**(13), 720–723, 1980.

Diabetes

Collagen-linked fluorescence was measured in skin biopsy specimens from 41 subjects with longstanding type I diabetes [4]. Diabetics were subdivided by the severity of diabetic complications. In particular, diabetics were graded according to (a) level of retinopathy (ocular abnormalities), where grade 0 = normal, grade 1 = background retinopathy, grade 2 = extensive (proliferative) retinopathy, and (b) level of nephropathy (kidney abnormalities), where grade 0 = urinary protein < 0.5 g/24 hrs, grade 1 = urinary protein ≥ 0.5 g/24 hrs, < 1 g/24 hrs, grade 2 = urinary protein ≥ 1 g/24 hrs. The results are presented in Table 12.5.

Table 12.5 Collagen-linked fluorescence in relation to the type and severity of diabetic complications

Retinopathy grade	*n*	Mean ± sd	Nephropathy grade	*n*	Mean ± sd
0	11	447 ± 17	0	28	487 ± 24
1	16	493 ± 30	1	6	481 ± 16
2	14	551 ± 35	2	7	567 ± 24

Source: Reprinted with permission of the *New England Journal of Medicine*, **314**(7), 403–408, 1986.

12.18 Assess whether there is any overall difference in mean fluorescence level by retinopathy grade.

12.19 Identify which specific retinopathy-grade levels are different using both ordinary *t* tests and the method of multiple comparisons.

12.20 Answer Problem 12.18 when patients are grouped by nephropathy grade.

12.21 Answer Problem 12.19 when patients are grouped by nephropathy grade.

Pulmonary Disease
Twenty-two young asthmatic volunteers were studied to assess the short-term effects of sulfur dioxide (SO_2) exposure under various conditions [5]. The baseline data in Table 12.6 were presented regarding bronchial reactivity to SO_2 stratified by lung function (as defined by FEV_1/FVC) at screening.

12.22 Suppose we do not wish to assume normality. What nonparametric test can be used to compare the three groups?

12.23 Implement the test in Problem 12.22 and report a *p*-value.

Table 12.6 Relationship of bronchial reactivity to SO_2 (cm H_2O/s) grouped by lung function at screening among 22 asthmatic volunteers

Lung function group		
Group A	Group B	Group C
$FEV_1/FVC \le 74\%$	FEV_1/FVC 75–84%	$FEV_1/FVC \ge 85\%$
20.8	7.5	9.2
4.1	7.5	2.0
30.0	11.9	2.5
24.7	4.5	6.1
13.8	3.1	7.5
	8.0	
	4.7	
	28.1	
	10.3	
	10.0	
	5.1	
	2.2	

Source: Reprinted with permission of the *American Review of Respiratory Disease*, **131**(2), 221–225, 1985.

Pharmacology
Refer to Table 12.28 (in Chapter 12, text) where data are presented on the reduction in fever for patients getting different doses of aspirin.

12.24 Use nonparametric methods to test for significant differences among groups. Identify which groups are significantly different.

12.25 Compare your results for this problem using parametric (see Problems 12.21–12.23, text) and nonparametric methods. Which do you think is the more appropriate method here?

Pulmonary Disease
In Table 6.3 (Chapter 6 of this Study Guide) we have provided replicate measures for each of two different methods of measuring flow rates at 50% of forced vital capacity.

12.26 Estimate the intraclass correlation coefficient between replicate measures, using the manual method, and provide a 95% Cl about this estimate.

12.27 Answer the same questions posed in Problem 12.26 based on the digitizer method.

SOLUTIONS

12.1 A fixed-effects one-way ANOVA can be used.

12.2 The test statistic is given by $F =$ Between MS/Within MS $\sim F_{k-1,\, n-k}$ under H_0.

$$\text{Between SS} = 10(74)^2 + 10(56)^2 + 10(55)^2 - \frac{[10(74) + 10(56) + 10(55)]^2}{30}$$
$$= 54{,}760 + 31{,}360 + 30{,}250 - 114{,}083.33$$
$$= 116{,}370 - 114{,}083.33 = 2286.7$$

$$\text{Between MS} = \frac{2286.7}{2} = 1143.3$$

Within MS = s^2, obtained from the pooled-variance estimate,

$$s^2 = \frac{9(16)^2 + 9(16)^2 + 9(9)^2}{27} = \frac{5337}{27} = 197.7$$

Therefore,

$$F = \frac{1143.3}{197.7} = 5.78 \sim F_{2,\,27} \text{ under } H_0$$

Since $F_{2,\,27,\,.95} < F_{2,\,20,\,.95} = 3.49 < F$, it follows that $p < .05$, and H_0 is rejected at the 5% level.

12.3 Use the test statistic

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n-k}$$

under H_0. The results are given in the table below.

Groups compared	Test statistic	p-value
STD, LAC	$t = \frac{74-56}{\sqrt{197.7(\frac{1}{10}+\frac{1}{10})}} = \frac{18}{6.288} = 2.86 \sim t_{27}$	$.001 < p < .01$
STD, VEG	$t = \frac{74-55}{6.288} = \frac{19}{6.288} = 3.02 \sim t_{27}$	$.001 < p < .01$
LAC, VEG	$t = \frac{56-55}{6.288} = \frac{1}{6.288} = 0.16 \sim t_{27}$	NS

12.4 This contrast is a comparison of the general vegetarian population with the general nonvegetarian population. Compute the test statistic

$$t = \frac{L}{se(L)} \sim t_{n-k} \text{ under } H_0$$

$$= \frac{0.7(56) + 0.3(55) - 74}{\sqrt{197.7\left(\frac{0.7^2}{10} + \frac{0.3^2}{10} + \frac{(-1)^2}{10}\right)}} = \frac{-18.3}{\sqrt{31.231}} = \frac{-18.3}{5.589}$$

$$= -3.27 \sim t_{27} \text{ under } H_0$$

Since $t_{27,\,.995} = 2.771$, and $t_{27,\,.9995} = 3.690$, it follows that $.001 < p < .01$, and we reject H_0 and conclude that among premenopausal women, the general vegetarian population has a significantly lower mean protein intake than the general nonvegetarian population.

12.5 We use the Bonferroni approach. For a 5% level of significance, the critical values are given by $t_{27,\,\alpha^*/2}$, $t_{27,\,1-\alpha^*/2}$, where

$$\alpha^* = \frac{.05}{\binom{3}{2}} = \frac{.05}{3} = .0167.$$

Thus, $t_{27,\,\alpha^*/2} = t_{27,\,.0083}$ and $t_{27,\,1-\alpha^*/2} = t_{27,\,.9917}$. Using a computer program (e.g., the inverse CDF program of MINITAB), we find that $t_{27,\,.9917} = 2.55$. From the solution to Problem 12.3,

$$|t(\text{STD, LAC})| = 2.86 > 2.55,$$
$$|t(\text{STD, VEG})| = 3.02 > 2.55$$

and $|t(\text{LAC, VEG})| = 0.16 < 2.55$. Thus, it follows that both the lactovegetarians and the strict vegetarians have significantly lower mean protein intake than the non-vegetarians, among premenopausal women.

12.6 From Problem 12.4, we have a t statistic $= -3.27$. Using the Scheffé multiple-comparisons procedure (text, Equation 12.16, Chapter 12), the critical values are $c_2 = \sqrt{2F_{2,\,27,\,.95}}$, $c_1 = -c_2$. Using the computer, we obtain $F_{2,\,27,\,.95} = 3.35$. Therefore, $c_2 = \sqrt{6.70} = 2.59$, $c_1 = -2.59$. Since $t < c_1$, it follows that $p < .05$ and the linear contrast is significantly different from 0.

12.7 The fixed-effects one-way analysis of variance.

12.8 We have the test statistic

$$F = \frac{\text{Between MS}}{\text{Within MS}} \sim F_{k-1,\,n-k} \text{ under } H_0,$$

where

$$\text{Between SS} = 487(8.1)^2 + \ldots + 82(5.8)^2 - \frac{[487(8.1) + \ldots + 82(5.8)]^2}{993}$$

$$= 68{,}553.83 - \frac{(8189.3)^2}{993}$$

$$= 68{,}553.83 - 67{,}537.40 = 1016.43$$

$$\text{Between MS} = \frac{1016.43}{3} = 338.81$$

$$\text{Within MS} = s^2 = \frac{(10.4)^2 486 + \ldots + (10.3)^2 81}{989} = 106.23$$

Thus, $F = 338.81/106.23 = 3.19 \sim F_{3,\ 989}$ under H_0. We note from the F table that

$$F > F_{3,\ 120,\ .95} = 2.68 > F_{3,\ 989,\ .95}$$

Thus, the means are significantly different (i.e., $p < .05$).

12.9 We now perform t tests for each pair of groups to determine which specific groups are different using the LSD approach to obtain critical values. The results are given in the table.

Groups	Test statistic	p-value
20–34, 35–49	$t = \frac{8.1 - 9.7}{\sqrt{106.23(\frac{1}{487} + \frac{1}{233})}} = \frac{-1.6}{0.82} = -1.95 \sim t_{989}$	$.05 < p < .10$
20–34, 50–64	$t = \frac{8.1 - 7.9}{\sqrt{106.23(\frac{1}{487} + \frac{1}{191})}} = \frac{0.2}{0.88} = 0.23 \sim t_{989}$	NS
20–34, 65–74	$t = \frac{8.1 - 5.8}{\sqrt{106.23(\frac{1}{487} + \frac{1}{82})}} = \frac{2.3}{1.23} = 1.87 \sim t_{989}$	$.05 < p < .10$
35–49, 50–64	$t = \frac{9.7 - 7.9}{\sqrt{106.23(\frac{1}{233} + \frac{1}{191})}} = \frac{1.8}{1.01} = 1.79 \sim t_{989}$	$.05 < p < .10$
35–49, 65–74	$t = \frac{9.7 - 5.8}{\sqrt{106.23(\frac{1}{233} + \frac{1}{82})}} = \frac{3.9}{1.32} = 2.95 \sim t_{989}$	$.001 < p < .01$
50–64, 65–74	$t = \frac{7.9 - 5.8}{\sqrt{106.23(\frac{1}{191} + \frac{1}{82})}} = \frac{2.1}{1.36} = 1.54 \sim t_{989}$	NS

12.10 We note from Problem 12.9 that only two of the groups are significantly different (35–49, 65–74). However, most of the comparisons show a trend in the direction of less activity with increasing age. To test for this general trend, we form the linear contrast

$$L = 27.5\bar{y}_1 + 42.5\bar{y}_2 + 57.5\bar{y}_3 + 70\bar{y}_4,$$

where the weights are chosen to be the midpoints of the four respective age groups. Since the coefficients of the contrast must add up to 0, we subtract $(27.5 + \ldots + 70)/4 = 49.375$ from each coefficient to obtain the contrast

$$\begin{aligned} L &= (27.5 - 49.375)\bar{y}_1 + (42.5 - 49.375)\bar{y}_2 \\ &\quad + (57.5 - 49.375)\bar{y}_3 + (70 - 49.375)\bar{y}_4 \\ &= -21.875\bar{y}_1 - 6.875\bar{y}_2 + 8.125\bar{y}_3 + 20.625\bar{y}_4 \end{aligned}$$

We wish to test the hypothesis H_0: $E(L) = 0$ vs. H_1: $E(L) \neq 0$. We have the test statistic $L/se(L) \sim t_{n-k}$ under H_0, where

$$\begin{aligned} L &= -21.875(8.1) + \ldots + 20.625(5.8) = -60.06 \\ se(L) &= \sqrt{106.23\left(\frac{21.875^2}{487} + \ldots + \frac{20.625^2}{82}\right)} \\ &= \sqrt{106.23(6.719)} = \sqrt{713.75} = 26.72 \end{aligned}$$

Thus, $t = -60.06/26.72 = -2.25 \sim t_{989}$ under H_0. Since $|t| > t_{120,\ .975} = 1.980 > t_{989,\ .975}$, it follows that

$$p < .05.$$

Furthermore, since $|t| < t_{\infty,\ .99} = 2.326 < t_{989,\ .99}$, it follows that $p > .02$. Therefore, we have

$$.02 < p < .05$$

and the linear trend is significant. Thus, there is a significant trend with the average amount of overall activity declining with age. In this case, the linear contrast provides a less confusing message than did the comparison of individual groups in Problem 12.9.

12.11 The fixed-effects one-way ANOVA model is appropriate here. We use the overall F test for one-way ANOVA.

12.12 We have the following mean and sd within each group:

Group	Mean	sd	n
≤ 50	3.93	2.21	9
51–1000	5.41	3.38	10
> 1000	5.94	2.55	9
Overall	5.11		28

The test statistic is

$$F = \frac{\text{Between MS}}{\text{Within MS}} \sim F_{2,\ 25} \text{ under } H_0,$$

where

$$\begin{aligned} \text{Between SS} &= 9(3.93 - 5.11)^2 + 10(5.41 - 5.11)^2 \\ &\quad + 9(5.94 - 5.11)^2 \\ &= 19.627 \end{aligned}$$

$$\text{Between MS} = \frac{19.627}{2} = 9.814$$

$$\text{Within SS} = 8(2.21)^2 + 9(3.38)^2 + 8(2.55)^2 = 193.991$$

$$\text{Within MS} = \frac{193.991}{25} = 7.760$$

$$F = \frac{9.814}{7.760} = 1.265 \sim F_{2,\ 25} \text{ under } H_0$$

$$p\text{-value} = Pr\left(F_{2,\ 25} > 1.265\right) = .30$$

Thus, there is no significant difference among the mean protein concentrations for the 3 groups.

12.13 Since the F test in Problem 12.12 was not significant, we declare that each pair of means is not significanitly different.

12.14 A 95% Cl for the mean difference between each pair of groups is given as follows:

95% CI for mean difference in protein concentration between trypsin-secretion groups

Groups compared	95% CI
≤ 50, 51–1000	$3.93 - 5.41 \pm t_{25,\ .975}\sqrt{7.76\left(\frac{1}{9}+\frac{1}{10}\right)} = -1.48 \pm 2.060(1.28) = (-4.1,\ 1.2)$
≤ 50, > 1000	$3.93 - 5.94 \pm 2.060\sqrt{7.76\left(\frac{1}{9}+\frac{1}{9}\right)} = -2.01 \pm 2.060(1.31) = (-4.7,\ 0.7)$
51–1000, > 1000	$5.41 - 5.94 \pm 2.060\sqrt{7.76\left(\frac{1}{10}+\frac{1}{9}\right)} = -0.53 \pm 2.64 = (-3.2,\ 2.1)$

12.15 We wish to test the hypothesis $H_0: \mu_1 = \ldots = \mu_6$ versus H_1: at least 2 of the μ_i's are unequal. We use the F test as follows:

$$\text{Between SS} = 200(3.72)^2 + \ldots + 200(2.77) - \frac{[200(3.72) + \ldots + 200(2.77)]^2}{1050}$$

$$= 122.22$$

$$\text{Between MS} = \frac{122.22}{5} = 24.44$$

$$\text{Within MS} = s^2, \text{ where}$$

$$s^2 = \frac{199(0.65)^2 + \ldots + 199(0.60)^2}{1050 - 6} = \frac{408.61}{1044} = 0.391$$

Thus, we have the test statistic

$$F = \frac{\text{Between MS}}{\text{Within MS}} = \frac{24.44}{0.391} = 62.46 \sim F_{5,\ 1044} \text{ under } H_0.$$

Since $F_{5,\ 1044,\ .999} < F_{5,\ 120,\ .999} = 4.42 < F$, it follows that $p < .001$ and the 6 means are significantly different.

12.16 We use the LSD t test criterion in Equation 12.12 (text, Chapter 12) for each pair of groups. The results are summarized on the next page.

The p-values were all evaluated using the $N(0, 1)$ distribution because it is sufficiently close to a t_{1044} distribution. The results show that (1) nonsmokers are significantly different from passive smokers and light, moderate, and heavy inhaling smokers and are not significantly different from noninhalers; (2) passive smokers are significantly different from moderate and heavy inhaling smokers, but not from light smokers or noninhalers; (3) noninhalers are significantly different from moderate and heavy smokers but not from light smokers; and (4) light, moderate, and heavy inhaling smokers are all significantly different from each other.

12.17 Using the Bonferroni approach, the critical values are $c_1 = t_{1044,\ .025/15} = t_{1044,\ .0017}$ and $c_2 = t_{1044,\ .9983}$. If we approximate t_{1044} by an $N(0, 1)$ distribution and refer to Table 3 (Appendix, text), we see that $c_1 = -2.94$, $c_2 = 2.94$. Thus, the results are the same as with the LSD procedure in Problem 12.16, with the exception of the (NS, PS) comparison, which is no longer significant. However, the (NS, PS) comparison is the most important result in the paper, so that the LSD and Bonferroni approaches yield very different conclusions.

12.18 We use the fixed-effects one-way ANOVA as follows:

$$\text{Between SS} = 447^2(11) + 493^2(16) + 551^2(14) - \frac{[447(11) + 493(16) + 551(14)]^2}{41}$$

$$= 10{,}337{,}097 - 10{,}269{,}008.80 = 68{,}088.20$$

$$\text{Between MS} = \frac{\text{Between SS}}{2} = 34{,}044.10$$

$$\text{Within MS} = s^2, \text{ where}$$

$$s^2 = \frac{10(17)^2 + \ldots + 13(35)^2}{38} = \frac{32{,}315}{38} = 850.39$$

Thus, the test statistic

$$F = \frac{34{,}044.10}{850.39} = 40.03 \sim F_{2,\ 38} \text{ under } H_0.$$

Since $F > F_{2,\ 30,\ .999} = 8.77 > F_{2,\ 38,\ .999}$, it implies that $p < .001$ and there is a significant overall difference in mean fluorescence level by retinopathy grade.

Groups	Test statistic	p-value
NS, PS	$t = \frac{3.72 - 3.54}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.18}{0.063} = 2.88 \sim t_{1044}$	.004
NS, NI	$t = \frac{3.72 - 3.56}{\sqrt{0.391(\frac{1}{200} + \frac{1}{50})}} = \frac{0.16}{0.099} = 1.62 \sim t_{1044}$	NS
NS, LS	$t = \frac{3.72 - 3.49}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.23}{0.063} = 3.68 \sim t_{1044}$	$< .001$
NS, MS	$t = \frac{3.72 - 3.08}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.64}{0.063} = 10.23 \sim t_{1044}$	$< .001$
NS, HS	$t = \frac{3.72 - 2.77}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.95}{0.063} = 15.19 \sim t_{1044}$	$< .001$
PS, NI	$t = \frac{3.54 - 3.56}{\sqrt{0.391(\frac{1}{200} + \frac{1}{50})}} = \frac{-0.02}{0.099} = -0.20 \sim t_{1044}$	NS
PS, LS	$t = \frac{3.54 - 3.49}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.05}{0.063} = 0.80 \sim t_{1044}$	NS
PS, MS	$t = \frac{3.54 - 3.08}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.46}{0.063} = 7.35 \sim t_{1044}$	$< .001$
PS, HS	$t = \frac{3.54 - 2.77}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.77}{0.063} = 12.31 \sim t_{1044}$	$< .001$
NI, LS	$t = \frac{3.56 - 3.49}{\sqrt{0.391(\frac{1}{50} + \frac{1}{200})}} = \frac{0.07}{0.099} = 0.71 \sim t_{1044}$	NS
NI, MS	$t = \frac{3.56 - 3.08}{\sqrt{0.391(\frac{1}{50} + \frac{1}{200})}} = \frac{0.48}{0.099} = 4.85 \sim t_{1044}$	$< .001$
NI, HS	$t = \frac{3.56 - 2.77}{\sqrt{0.391(\frac{1}{50} + \frac{1}{200})}} = \frac{0.79}{0.099} = 7.99 \sim t_{1044}$	$< .001$
LS, MS	$t = \frac{3.49 - 3.08}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.41}{0.063} = 6.55 \sim t_{1044}$	$< .001$
LS, HS	$t = \frac{3.49 - 2.77}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.72}{0.063} = 11.51 \sim t_{1044}$	$< .001$
MS, HS	$t = \frac{3.08 - 2.77}{\sqrt{0.391(\frac{1}{200} + \frac{1}{200})}} = \frac{0.31}{0.063} = 4.96 \sim t_{1044}$	$< .001$

12.19 We first perform ordinary t tests using critical values based on the LSD approach in the table on the next page. Clearly, all pairs of groups are significantly different using ordinary t tests. We now use the Bonferroni method for multiple comparisons to determine the critical values. Since there are 3 groups, the critical values are $c_1 = t_{38,\ \alpha^*/2}$, $c_2 = t_{38,\ 1-\alpha^*/2}$ where

$$\alpha^* = \frac{0.05}{\binom{3}{2}} = .0167$$

Thus, $c_1 = t_{38,\ .0083}$, $c_2 = t_{38,\ .9917}$. Using the computer, we determine that $c_1 = -2.50$, $c_2 = 2.50$. Since the test statistics for all 3 comparisons are < -2.50, it follows that using the Bonferroni approach, there are also significant differences between the mean fluorescence level for each pair of retinopathy groups.

12.20 We use the fixed-effects one-way ANOVA. We have

$$\text{Between SS} = 487^2(28) + 481^2(6) + 567^2(7) - \frac{[487(28) + 481(6) + 567(7)]^2}{41}$$
$$= 10{,}279{,}321 - 10{,}241{,}001.98 = 38{,}319.02$$

$$\text{Between MS} = \frac{38{,}319.02}{2} = 19{,}159.51$$

$$\text{Within MS} = s^2, \text{ where}$$

$$s^2 = \frac{27(24)^2 + \ldots + 6(24)^2}{38} = \frac{20{,}288}{38} = 533.89$$

Thus, the F statistic

$$F = \frac{19{,}159.51}{533.89} = 35.89 \sim F_{2,\,38} \text{ under } H_0 .$$

Since $F > F_{2,\,30,\,.999} = 8.77 > F_{2,\,38,\,.999}$, it implies that $p < .001$. Thus, there is a significant overall difference in mean fluorescence level by nephropathy grade.

12.21 We first perform ordinary t tests using critical values based on the LSD approach as shown below. Thus, there are significant differences between group 0 and group 2 and between group 1 and group 2 ($p < .001$), but no significant difference between groups 0 and 1. We now use the method of multiple comparisons based on the Bonferroni approach. Since there are 3 groups and 38 *df*, the critical values are the same as in Problem 12.19, namely, $c_1 = -2.50$, $c_2 = 2.50$. Since

$$|0.58| < 2.50, \; |-8.19| > 2.50 \text{ and } |-6.69| > 2.50,$$

it follows that there are significant differences in mean fluorescence levels between nephropathy grade 2 vs. each of nephropathy grades 0 and 1, respectively, while there is no significant difference between patients in nephropathy grades 0 and 1.

Retinopathy groups compared	Test statistic	p-Value
0, 1	$t = \frac{447 - 493}{\sqrt{850.39(\frac{1}{11} + \frac{1}{16})}} = \frac{-46}{11.422} = -4.03 \sim t_{38}$	$< .001$
0, 2	$t = \frac{447 - 551}{\sqrt{850.39(\frac{1}{11} + \frac{1}{14})}} = \frac{-104}{11.750} = -8.85 \sim t_{38}$	$< .001$
1, 2	$t = \frac{493 - 551}{\sqrt{850.39(\frac{1}{16} + \frac{1}{14})}} = \frac{-58}{10.672} = -5.43 \sim t_{38}$	$< .001$

Nephropathy groups compared	Test statistic	p-Value
0, 1	$t = \frac{487 - 481}{\sqrt{533.89(\frac{1}{28} + \frac{1}{6})}} = \frac{6}{10.395} = 0.58 \sim t_{38}$	NS
0, 2	$t = \frac{487 - 567}{\sqrt{533.89(\frac{1}{28} + \frac{1}{7})}} = \frac{-80}{9.764} = -8.19 \sim t_{38}$	$< .001$
1, 2	$t = \frac{481 - 567}{\sqrt{533.89(\frac{1}{6} + \frac{1}{7})}} = \frac{-86}{12.855} = -6.69 \sim t_{38}$	$< .001$

12.22 The Kruskal-Wallis test.

12.23 We have the test statistic

$$H = \frac{H^*}{1 - \left(\sum t_j^3 - t_j\right)/(N^3 - N)} \sim \chi^2_{k-1} \text{ under } H_0$$

where

$$H^* = \frac{12}{N(N+1)}\left(\sum \frac{R_i^2}{n_i}\right) - 3(N+1)$$

and

$$R_i = \text{rank sum in the } i\text{th group.}$$

We pool the three samples and compute ranks in the combined sample as follows:

Value	Frequency	Rank range	Average rank
2.0	1	1	1.0
2.2	1	2	2.0
2.5	1	3	3.0
3.1	1	4	4.0
4.1	1	5	5.0
4.5	1	6	6.0
4.7	1	7	7.0
5.1	1	8	8.0
6.1	1	9	9.0
7.5	3	10–12	11.0
8.0	1	13	13.0
9.2	1	14	14.0
10.0	1	15	15.0
10.3	1	16	16.0
11.9	1	17	17.0
13.8	1	18	18.0
20.8	1	19	19.0
24.7	1	20	20.0
28.1	1	21	21.0
30.0	1	22	22.0

We now compute the rank sum in each of the three samples:

$$R_1 = 19.0 + \ldots + 18.0 = 84.0$$
$$R_2 = 11.0 + \ldots + 2.0 = 131.0$$
$$R_3 = 14.0 + \ldots + 11.0 = 38.0$$

Thus, the test statistic is given by

$$H^* = \frac{12}{22(23)} \times \left(\frac{84^2}{5} + \frac{131^2}{12} + \frac{38^2}{5} \right) - 3(23)$$
$$= 0.0237(1411.2 + 1430.1 + 288.8) - 69$$
$$= 0.0237(3130.1) - 69 = 5.231$$

Finally, we apply the correction for ties and obtain

$$H = \frac{5.231}{1 - \frac{3^3 - 3}{22^3 - 22}} = \frac{5.231}{1 - \frac{24}{10{,}626}}$$
$$= \frac{5.231}{.9977} = 5.24 \sim \chi^2_2 \text{ under } H_0$$

Since $\chi^2_{2,\ .95} = 5.99 > 5.24$, it follows that $p > .05$ and there is no significant difference among the three means.

12.24 We use the Kruskal-Wallis test to assess overall differences between groups. We combine the three samples and compute ranks in the pooled sample as follows:

Value	Rank	Value	Rank	Value	Rank
–1.0	1.0	0.3	6.5	1.2	11.0
–0.4	2.0	0.3	6.5	1.3	12.0
–0.2	3.0	0.5	8.0	1.6	13.0
0.2	4.5	0.6	9.0	2.0	14.0
0.2	4.5	1.1	10.0	2.1	15.0

We now compute the rank sum in each of the three groups:

Drug A		Drug B		Drug C	
Value	Rank	Value	Rank	Value	Rank
2.0	14.0	0.5	8.0	1.1	10.0
1.6	13.0	1.2	11.0	–1.0	1.0
2.1	15.0	0.3	6.5	–0.2	3.0
0.6	9.0	0.2	4.5	0.2	4.5
1.3	12.0	–0.4	2.0	0.3	6.5
Rank sum	63.0		32.0		25.0

The test statistic is given by

$$H = \frac{1}{C}\left[\frac{12}{15(16)} \left(\frac{63.0^2}{5} + \frac{32.0^2}{5} + \frac{25.0^2}{5} \right) - 3(16) \right]$$
$$= \frac{1}{C}[0.05(793.8 + 204.8 + 125.0) - 48]$$
$$= \frac{1}{C}[0.05(1{,}123.6) - 48] = \frac{1}{C}(56.18 - 48) = \frac{8.18}{C}$$

The correction for ties (C) is

$$C = 1 - \frac{(2^3 - 2) + (2^3 - 2)}{15^3 - 15} = 1 - \frac{12}{3360} = .9964$$

Thus, $H = 8.18/.9964 = 8.21 \sim \chi^2_2$ under H_0. Since $\chi^2_{2,\ .975} = 7.38$, $\chi^2_{2,\ .99} = 9.21$ and $7.38 < 8.21 < 9.21$, it follows that $.01 < p < .025$ and some of the groups are significantly different.

To identify which groups are different, we use the Dunn procedure (text, Chapter 12). The test statistics and p-values for the comparison of each pair of groups are given as follows:

Groups	Test statistic	Critical value	p-value
A, B	$z = \frac{\frac{63.0-32.0}{5}}{\sqrt{\frac{15(16)}{12} \times \left(\frac{1}{5} + \frac{1}{5}\right)}} = \frac{6.2}{2.83} = 2.19$	2.394[a]	NS
A, C	$z = \frac{\frac{63.0-25.0}{5}}{2.83} = \frac{7.6}{2.83} = 2.69$	2.394	$p < .05$
B, C	$z = \frac{\frac{32.0-25.0}{5}}{2.83} = \frac{1.4}{2.83} = 0.49$	2.394	NS

[a] $z_{1-0.05/6} = z_{.9917} = 2.394$

Thus, under this procedure only groups A and C are significantly different.

12.25 Using parametric methods, group A was significantly different from both groups B and C (text, Answers Section). Using nonparametric methods, group A was only significantly different from group C. A stem-and-leaf plot of the temperature differences over all groups is given as follows:

–1	0
–0	42
0	653223
1	6321
2	01

From the plot, we see that the normality assumption seems to be reasonably satisfied with these data and thus, the parametric ANOVA procedure is probably more appropriate here.

12.26 We first run a one-way ANOVA based on the manual-method replicates. The results, using MINITAB, are given as follows:

Analysis of variance of manual method replicates					
Source	DF	SS	MS	F	P
Person	9	11.71510	1.30168	452.76	0.000
Error	10	0.02875	0.00288		
Total	19	11.74385			

From Equation 12.33 (text, Chapter 12), we have $\hat{\sigma}^2 = 0.00288$,
$\hat{\sigma}_A^2 = (1.30168 - 0.00288)/2 = 0.64940$. The estimated intraclass correlation

$$\hat{\rho}_{I,\text{ manual}} = \frac{0.64940}{0.64940 + 0.00288} = \frac{0.64940}{0.65228} = .996.$$

From Equation 12.36 (text, Chapter 12), a 95% CI for $\hat{\rho}_{I,\text{ manual}}$ is given by (c_1, c_2), where

$$c_1 = \max\left(\frac{\frac{F}{F_{9,\,10,\,.975}} - 1}{2 + \frac{F}{F_{9,\,10,\,.975}} - 1},\ 0\right)$$
$$c_2 = \max\left(\frac{\frac{F}{F_{9,\,10,\,.025}} - 1}{2 + \frac{F}{F_{9,\,10,\,.025}} - 1},\ 0\right)$$

Using MINITAB, we estimate that $F_{9,\,10,\,.975} = 3.78$ and $F_{9,\,10,\,.025} = 0.253$. Thus,

$$c_1 = \max\left(\frac{\frac{452.76}{3.78} - 1}{\frac{452.76}{3.78} + 1},\ 0\right) = \frac{118.78}{120.78} = .983$$
$$c_2 = \max\left(\frac{\frac{452.76}{0.253} - 1}{\frac{452.76}{0.253} + 1},\ 0\right) = \frac{1788.57}{1790.57} = .999$$

Therefore, the 95% CI for $\hat{\rho}_{I,\text{ manual}} = (0.983,\ 0.999)$.

12.27 We run a one-way ANOVA based on the digitizer-method replicates as follows:

Analysis of variance on digitizer					
Source	DF	SS	MS	F	P
Person	9	12.67584	1.40843	3658.25	0.000
Error	10	0.00385	0.00039		
Total	19	12.67969			

We have $\hat{\sigma}^2 = 0.00039$,

$$\hat{\sigma}_A^2 = \frac{1.40843 - 0.00039}{2} = 0.704021.$$

Thus,

$$\hat{\rho}_{I,\text{ digitizer}} = \frac{0.704021}{0.704021 + 0.00039} = \frac{0.704021}{0.704406} = .9995.$$

From Equation 12.36 (text, Chapter 12) a 95% CI for $\hat{\rho}_{I,\text{ digitizer}} = (c_1, c_2)$, where

$$c_1 = \max\left(\frac{\frac{F}{F_{9,\,10,\,.975}} - 1}{\frac{F}{F_{9,\,10,\,.975}} + 1},\ 0\right)$$
$$c_2 = \max\left(\frac{\frac{F}{F_{9,\,10,\,.025}} - 1}{\frac{F}{F_{9,\,10,\,.025}} + 1},\ 0\right)$$

Thus,

$$c_1 = \max\left(\frac{\frac{3658.25}{3.78} - 1}{\frac{3658.25}{3.78} + 1},\ 0\right) = \frac{966.79}{968.79} = .9979$$
$$c_2 = \max\left(\frac{\frac{3658.25}{0.253} - 1}{\frac{3658.25}{0.253} + 1},\ 0\right) = \frac{14,458.50}{14,460.50} = .9999$$

Therefore, the 95 % CI for

$$\hat{\rho}_{I,\text{ digitizer}} = (.9979,\ .9999).$$

REFERENCES

[1] Sallis, J. F., Haskell, W. L., Wood, P. D., Fortmann, S. P., Rogers, T., Blair, S. N., & Paffenbarger, R. S., Jr. (1985). Physical activity assessment methodology in the five-city project. *American Journal of Epidemiology*, **121**(1), 91–106.

[2] Kopelman, H., Durie, P., Gaskin, K., Weizman, Z., & Forstner, G. (1985). Pancreatic fluid secretion and protein hyperconcentration in cystic fibrosis. *New England Journal of Medicine*, **312**(6), 329–334.

[3] White, J. R., & Froeb, H. F. (1980). Small-airways dysfunction in nonsmokers chronically exposed to tobacco smoke. *New England Journal of Medicine*, **302**(13), 720–723.

[4] Monnier, V. M., Vishwanath, V., Frank, K. E., Elmets, C. A., Dauchot, P., & Kohn, R. R. (1986). Relation between complications of type I diabetes mellitus and collagen-linked fluorescence. *New England Journal of Medicine*, **314**(7), 403–408.

[5] Linn, W. S., Shamoo, D. A., Anderson, K. R., Whynot, J. D., Avol, E. L., & Hackney, J. D. (1985). Effects of heat and humidity on the responses of exercising asthmatics to sulfur dioxide exposure. *American Review of Respiratory Disease*, **131**(2), 221–225.