3.11.2 General Comments on the Rank Transformation

The procedure used in the previous section of replacing the observations by their ranks is called the **rank transformation**. It is a very powerful and widely useful technique. If we were to apply the ordinary F test to the ranks rather than to the original data, we would obtain

$$F_0 = \frac{H/(a-1)}{(N-1-H)/(N-a)} \tag{3.70}$$

as the test statistic [see Conover (1980), p. 337]. Note that as the Kruskal–Wallis statistic H increases or decreases, F_0 also increases or decreases, so the Kruskal–Wallis test is equivalent to applying the usual analysis of variance to the ranks.

The rank transformation has wide applicability in experimental design problems for which no nonparametric alternative to the analysis of variance exists. This includes many of the designs in subsequent chapters of this book. If the data are ranked and the ordinary F test is applied, an approximate procedure that has good statistical properties results [see Conover and Iman (1976, 1981)]. When we are concerned about the normality assumption or the effect of outliers or "wild" values, we recommend that the usual analysis of variance be performed on both the original data and the ranks. When both procedures give similar results, the analysis of variance assumptions are probably satisfied reasonably well, and the standard analysis is satisfactory. When the two procedures differ, the rank transformation should be preferred because it is less likely to be distorted by nonnormality and unusual observations. In such cases, the experimenter may want to investigate the use of transformations for nonnormality and examine the data and the experimental procedure to determine whether outliers are present and why they have occurred.

3.12 Problems

3.1. An experimenter has conducted a single-factor experiment with four levels of the factor, and each factor level has been replicated six times. The computed value of the F-statistic is $F_0 = 3.26$. Find bounds on the P-value.

3.2. An experimenter has conducted a single-factor experiment with six levels of the factor, and each factor level has been replicated three times. The computed value of the F- statistic is $F_0 = 5.81$. Find bounds on the P-value.

3.3. A computer ANOVA output is shown below. Fill in the blanks. You may give bounds on the P-value.

One-way ANOVA

Source	DF	SS	MS	F	P
Factor	3	36.15	?	?	?
Error	?	?	?		
Total	19	196.04			

3.4. A computer ANOVA output is shown below. Fill in the blanks. You may give bounds on the P-value.

One-way ANOVA

Source	DF	SS	MS	F	P
Factor	?	?	246.93	?	?
Error	25	186.53	?		
Total	29	1174.24			

3.5. An article appeared in *The Wall Street Journal* on Tuesday, April 27, 2010, with the title "Eating Chocolate Is Linked to Depression." The article reported on a study funded by the National Heart, Lung and Blood Institute (part of the National Institutes of Health) and conducted by faculty at the University of California, San Diego, and the University of California, Davis. The research was also published in the *Archives of Internal Medicine* (2010, pp. 699–703). The study examined 931 adults who were not taking antidepressants and did not have known cardiovascular disease or diabetes. The group was about 70% men and the average age of the group was reported to be about 58. The participants were asked about chocolate consumption and then screened for depression using a questionnaire. People who score less than 16 on the questionnaire are not considered depressed, while those with scores above 16 and less than or equal to 22 are considered possibly depressed, while those with scores above 22 are considered likely to be depressed. The survey found that people who were not depressed ate an average 5.4 servings of chocolate per month, possibly depressed individuals ate an average of 8.4 servings of chocolate per month, while those individuals who scored above 22 and were likely to be depressed ate the most chocolate, an average of 11.8 servings per month. No differentiation was made between dark and milk chocolate. Other foods were also examined, but no pattern emerged between other foods and depression. Is this study really a designed experiment? Does it establish a cause-and-effect link between chocolate consumption and depression? How would the study have to be conducted to establish such a cause-and effect link?

3.6. An article in *Bioelectromagnetics* ("Electromagnetic Effects on Forearm Disuse Osteopenia: A Randomized, Double-Blind, Sham-Controlled Study," Vol. 32, 2011, pp. 273–282) described a randomized, double-blind, sham-controlled, feasibility and dosing study to determine if a common pulsing electromagnetic field (PEMF) treatment could moderate the substantial osteopenia that occurs after forearm disuse. Subjects were randomized into four groups after a distal radius fracture, or carpal surgery requiring immobilization in a cast. Active or identical sham PEMF transducers were worn on the distal forearm for 1, 2, or 4h/day for 8 weeks starting after cast removal ("baseline") when bone density continues to decline. Bone mineral density (BMD) and bone geometry were measured in the distal forearm by dual energy X-ray absorptiometry (DXA) and peripheral quantitative computed tomography (pQCT). The data below are the percent losses in BMD measurements on the radius after 16 weeks for patients wearing the active or sham PEMF transducers for 1, 2, or 4h/day (data were constructed to match the means and standard deviations read from a graph in the paper).

(a) Is there evidence to support a claim that PEMF usage affects BMD loss? If so, analyze the data to determine which specific treatments produce the differences.

(b) Analyze the residuals from this experiment and comment on the underlying assumptions and model adequacy.

Sham	PEMF 1 h/day	PEMF 2 h/day	PEMF 4 h/day
4.51	5.32	4.73	7.03
7.95	6.00	5.81	4.65
4.97	5.12	5.69	6.65
3.00	7.08	3.86	5.49
7.97	5.48	4.06	6.98
2.23	6.52	6.56	4.85
3.95	4.09	8.34	7.26
5.64	6.28	3.01	5.92
9.35	7.77	6.71	5.58
6.52	5.68	6.51	7.91
4.96	8.47	1.70	4.90
6.10	4.58	5.89	4.54
7.19	4.11	6.55	8.18
4.03	5.72	5.34	5.42
2.72	5.91	5.88	6.03
9.19	6.89	7.50	7.04
5.17	6.99	3.28	5.17
5.70	4.98	5.38	7.60
5.85	9.94	7.30	7.90
6.45	6.38	5.46	7.91

3.7. The tensile strength of Portland cement is being studied. Four different mixing techniques can be used economically. A completely randomized experiment was conducted and the following data were collected:

Mixing Technique	Tensile Strength (lb/in^2)			
1	3129	3000	2865	2890
2	3200	3300	2975	3150
3	2800	2900	2985	3050
4	2600	2700	2600	2765

(a) Test the hypothesis that mixing techniques affect the strength of the cement. Use $\alpha = 0.05$.

(b) Construct a graphical display as described in Section 3.5.3 to compare the mean tensile strengths for the four mixing techniques. What are your conclusions?

(c) Use the Fisher LSD method with $\alpha = 0.05$ to make comparisons between pairs of means.

(d) Construct a normal probability plot of the residuals. What conclusion would you draw about the validity of the normality assumption?

(e) Plot the residuals versus the predicted tensile strength. Comment on the plot.

(f) Prepare a scatter plot of the results to aid the interpretation of the results of this experiment.

3.8(a) Rework part (c) of Problem 3.7 using Tukey's test with $\alpha = 0.05$. Do you get the same conclusions from Tukey's test that you did from the graphical procedure and/or the Fisher LSD method?

(b) Explain the difference between the Tukey and Fisher procedures.

3.9. Reconsider the experiment in Problem 3.7. Find a 95 percent confidence interval on the mean tensile strength of the Portland cement produced by each of the four mixing techniques. Also find a 95 percent confidence interval on the difference in means for techniques 1 and 3. Does this aid you in interpreting the results of the experiment?

3.10. A product developer is investigating the tensile strength of a new synthetic fiber that will be used to make cloth for men's shirts. Strength is usually affected by the percentage of cotton used in the blend of materials for the fiber. The engineer conducts a completely randomized experiment with five levels of cotton content and replicates the experiment five times. The data are shown in the following table.

Cotton Weight Percent	Observations				
15	7	7	15	11	9
20	12	17	12	18	18
25	14	19	19	18	18
30	19	25	22	19	23
35	7	10	11	15	11

(a) Is there evidence to support the claim that cotton content affects the mean tensile strength? Use $\alpha = 0.05$.

(b) Use the Fisher LSD method to make comparisons between the pairs of means. What conclusions can you draw?

(c) Analyze the residuals from this experiment and comment on model adequacy.

3.11. Reconsider the experiment described in Problem 3.10. Suppose that 30 percent cotton content is a control. Use Dunnett's test with $\alpha = 0.05$ to compare all of the other means with the control.

3.12. A pharmaceutical manufacturer wants to investigate the bioactivity of a new drug. A completely randomized single-factor experiment was conducted with three dosage levels, and the following results were obtained.

Dosage	Observations			
20 g	24	28	37	30
30 g	37	44	31	35
40 g	42	47	52	38

(a) Is there evidence to indicate that dosage level affects bioactivity? Use $\alpha = 0.05$.

(b) If it is appropriate to do so, make comparisons between the pairs of means. What conclusions can you draw?

(c) Analyze the residuals from this experiment and comment on model adequacy.

3.13. A rental car company wants to investigate whether the type of car rented affects the length of the rental period. An experiment is run for one week at a particular location, and 10 rental contracts are selected at random for each car type. The results are shown in the following table.

Type of Car	Observations									
Subcompact	3	5	3	7	6	5	3	2	1	6
Compact	1	3	4	7	5	6	3	2	1	7
Midsize	4	1	3	5	7	1	2	4	2	7
Full size	3	5	7	5	10	3	4	7	2	7

(a) Is there evidence to support a claim that the type of car rented affects the length of the rental contract? Use $\alpha = 0.05$. If so, which types of cars are responsible for the difference?

(b) Analyze the residuals from this experiment and comment on model adequacy.

(c) Notice that the response variable in this experiment is a count. Should this cause any potential concerns about the validity of the analysis of variance?

3.14. I belong to a golf club in my neighborhood. I divide the year into three golf seasons: summer (June–September), winter (November–March), and shoulder (October, April, and May). I believe that I play my best golf during the summer (because I have more time and the course isn't crowded) and shoulder (because the course isn't crowded) seasons, and my worst golf is during the winter (because when all of the part-year residents show up, the course is crowded, play is slow, and I get frustrated). Data from the last year are shown in the following table.

Season	Observations									
Summer	83	85	85	87	90	88	88	84	91	90
Shoulder	91	87	84	87	85	86	83			
Winter	94	91	87	85	87	91	92	86		

(a) Do the data indicate that my opinion is correct? Use $\alpha = 0.05$.

(b) Analyze the residuals from this experiment and comment on model adequacy.

3.15. A regional opera company has tried three approaches to solicit donations from 24 potential sponsors. The 24 potential sponsors were randomly divided into three groups of eight, and one approach was used for each group. The dollar amounts of the resulting contributions are shown in the following table.

Approach	Contributions (in $)							
1	1000	1500	1200	1800	1600	1100	1000	1250
2	1500	1800	2000	1200	2000	1700	1800	1900
3	900	1000	1200	1500	1200	1550	1000	1100

(a) Do the data indicate that there is a difference in results obtained from the three different approaches? Use $\alpha = 0.05$.

(b) Analyze the residuals from this experiment and comment on model adequacy.

3.16. An experiment was run to determine whether four specific firing temperatures affect the density of a certain type of brick. A completely randomized experiment led to the following data:

Temperature	Density				
100	21.8	21.9	21.7	21.6	21.7
125	21.7	21.4	21.5	21.4	
150	21.9	21.8	21.8	21.6	21.5
175	21.9	21.7	21.8	21.4	

(a) Does the firing temperature affect the density of the bricks? Use $\alpha = 0.05$.

(b) Is it appropriate to compare the means using the Fisher LSD method (for example) in this experiment?

(c) Analyze the residuals from this experiment. Are the analysis of variance assumptions satisfied?

(d) Construct a graphical display of the treatment as described in Section 3.5.3. Does this graph adequately summarize the results of the analysis of variance in part (a)?

3.17. Rework part (d) of Problem 3.16 using the Tukey method. What conclusions can you draw? Explain carefully how you modified the technique to account for unequal sample sizes.

3.18. A manufacturer of television sets is interested in the effect on tube conductivity of four different types of coating for color picture tubes. A completely randomized experiment is conducted and the following conductivity data are obtained:

Coating Type	Conductivity			
1	143	141	150	146
2	152	149	137	143
3	134	136	132	127
4	129	127	132	129

(a) Is there a difference in conductivity due to coating type? Use $\alpha = 0.05$.

(b) Estimate the overall mean and the treatment effects.

(c) Compute a 95 percent confidence interval estimate of the mean of coating type 4. Compute a 99 percent confidence interval estimate of the mean difference between coating types 1 and 4.

(d) Test all pairs of means using the Fisher LSD method with $\alpha = 0.05$.

(e) Use the graphical method discussed in Section 3.5.3 to compare the means. Which coating type produces the highest conductivity?

(f) Assuming that coating type 4 is currently in use, what are your recommendations to the manufacturer? We wish to minimize conductivity.

3.19. Reconsider the experiment from Problem 3.18. Analyze the residuals and draw conclusions about model adequacy.

3.20. An article in the *ACI Materials Journal* (Vol. 84, 1987, pp. 213–216) describes several experiments investigating the rodding of concrete to remove entrapped air. A 3-inch × 6-inch cylinder was used, and the number of times this rod was used is the design variable. The resulting compressive strength of the concrete specimen is the response. The data are shown in the following table:

Rodding Level	Compressive Strength		
10	1530	1530	1440
15	1610	1650	1500
20	1560	1730	1530
25	1500	1490	1510

(a) Is there any difference in compressive strength due to the rodding level? Use $\alpha = 0.05$.

(b) Find the P-value for the F statistic in part (a).

(c) Analyze the residuals from this experiment. What conclusions can you draw about the underlying model assumptions?

(d) Construct a graphical display to compare the treatment means as described in Section 3.5.3.

3.21. An article in *Environment International* (Vol. 18, No. 4, 1992) describes an experiment in which the amount of radon released in showers was investigated. Radon-enriched water was used in the experiment, and six different orifice diameters were tested in shower heads. The data from the experiment are shown in the following table:

Orifice Diameter	Radon Released (%)			
0.37	80	83	83	85
0.51	75	75	79	79
0.71	74	73	76	77
1.02	67	72	74	74
1.40	62	62	67	69
1.99	60	61	64	66

(a) Does the size of the orifice affect the mean percentage of radon released? Use $\alpha = 0.05$.

(b) Find the *P*-value for the *F* statistic in part (a).

(c) Analyze the residuals from this experiment.

(d) Find a 95 percent confidence interval on the mean percent of radon released when the orifice diameter is 1.40.

(e) Construct a graphical display to compare the treatment means as described in Section 3.5.3 What conclusions can you draw?

3.22. The response time in milliseconds was determined for three different types of circuits that could be used in an automatic valve shutoff mechanism. The results from a completely randomized experiment are shown in the following table:

Circuit Type	Response Time				
1	9	12	10	8	15
2	20	21	23	17	30
3	6	5	8	16	7

(a) Test the hypothesis that the three circuit types have the same response time. Use $\alpha = 0.01$.

(b) Use Tukey's test to compare pairs of treatment means. Use $\alpha = 0.01$.

(c) Use the graphical procedure in Section 3.5.3 to compare the treatment means. What conclusions can you draw? How do they compare with the conclusions from part (b)?

(d) Construct a set of orthogonal contrasts, assuming that at the outset of the experiment you suspected the response time of circuit type 2 to be different from the other two.

(e) If you were the design engineer and you wished to minimize the response time, which circuit type would you select?

(f) Analyze the residuals from this experiment. Are the basic analysis of variance assumptions satisfied?

3.23. The effective life of insulating fluids at an accelerated load of 35 kV is being studied. Test data have been obtained for four types of fluids. The results from a completely randomized experiment were as follows:

Fluid Type	Life (in h) at 35 kV Load					
1	17.6	18.9	16.3	17.4	20.1	21.6
2	16.9	15.3	18.6	17.1	19.5	20.3
3	21.4	23.6	19.4	18.5	20.5	22.3
4	19.3	21.1	16.9	17.5	18.3	19.8

(a) Is there any indication that the fluids differ? Use $\alpha = 0.05$.

(b) Which fluid would you select, given that the objective is long life?

(c) Analyze the residuals from this experiment. Are the basic analysis of variance assumptions satisfied?

3.24. Four different designs for a digital computer circuit are being studied to compare the amount of noise present. The following data have been obtained:

Circuit Design	Noise Observed				
1	19	20	19	30	8
2	80	61	73	56	80
3	47	26	25	35	50
4	95	46	83	78	97

(a) Is the same amount of noise present for all four designs? Use $\alpha = 0.05$.

(b) Analyze the residuals from this experiment. Are the analysis of variance assumptions satisfied?

(c) Which circuit design would you select for use? Low noise is best.

3.25. Four chemists are asked to determine the percentage of methyl alcohol in a certain chemical compound. Each chemist makes three determinations, and the results are the following:

Chemist	Percentage of Methyl Alcohol		
1	84.99	84.04	84.38
2	85.15	85.13	84.88
3	84.72	84.48	85.16
4	84.20	84.10	84.55

(a) Do chemists differ significantly? Use $\alpha = 0.05$.

(b) Analyze the residuals from this experiment.

(c) If chemist 2 is a new employee, construct a meaningful set of orthogonal contrasts that might have been useful at the start of the experiment.

3.26. Three brands of batteries are under study. It is suspected that the lives (in weeks) of the three brands are different. Five randomly selected batteries of each brand are tested with the following results:

Weeks of Life		
Brand 1	Brand 2	Brand 3
100	76	108
96	80	100
92	75	96
96	84	98
92	82	100

(a) Are the lives of these brands of batteries different?

(b) Analyze the residuals from this experiment.

(c) Construct a 95 percent confidence interval estimate on the mean life of battery brand 2. Construct a 99 percent confidence interval estimate on the mean difference between the lives of battery brands 2 and 3.

(d) Which brand would you select for use? If the manufacturer will replace without charge any battery that fails in less than 85 weeks, what percentage would the company expect to replace?

3.27. Four catalysts that may affect the concentration of one component in a three-component liquid mixture are being investigated. The following concentrations are obtained from a completely randomized experiment:

Catalyst			
1	2	3	4
58.2	56.3	50.1	52.9
57.2	54.5	54.2	49.9
58.4	57.0	55.4	50.0
55.8	55.3		51.7
54.9			

(a) Do the four catalysts have the same effect on the concentration?

(b) Analyze the residuals from this experiment.

(c) Construct a 99 percent confidence interval estimate of the mean response for catalyst 1.

3.28. An experiment was performed to investigate the effectiveness of five insulating materials. Four samples of each material were tested at an elevated voltage level to accelerate the time to failure. The failure times (in minutes) are shown below:

Material	Failure Time (minutes)			
1	110	157	194	178
2	1	2	4	18
3	880	1256	5276	4355
4	495	7040	5307	10,050
5	7	5	29	2

(a) Do all five materials have the same effect on mean failure time?

(b) Plot the residuals versus the predicted response. Construct a normal probability plot of the residuals. What information is conveyed by these plots?

(c) Based on your answer to part (b) conduct another analysis of the failure time data and draw appropriate conclusions.

3.29. A semiconductor manufacturer has developed three different methods for reducing particle counts on wafers. All three methods are tested on five different wafers and the after treatment particle count obtained. The data are shown below:

Method	Count				
1	31	10	21	4	1
2	62	40	24	30	35
3	53	27	120	97	68

(a) Do all methods have the same effect on mean particle count?

(b) Plot the residuals versus the predicted response. Construct a normal probability plot of the residuals. Are there potential concerns about the validity of the assumptions?

(c) Based on your answer to part (b) conduct another analysis of the particle count data and draw appropriate conclusions.

3.30. A manufacturer suspects that the batches of raw material furnished by his supplier differ significantly in calcium content. There are a large number of batches currently in the warehouse. Five of these are randomly selected for study. A chemist makes five determinations on each batch and obtains the following data:

Batch 1	Batch 2	Batch 3	Batch 4	Batch 5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.40	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.40	23.50	23.49	23.39	23.38

(a) Is there significant variation in calcium content from batch to batch? Use $\alpha = 0.05$.

(b) Estimate the components of variance.

(c) Find a 95 percent confidence interval for $\sigma_\tau^2 / (\sigma_\tau^2 + \sigma^2)$.

(d) Analyze the residuals from this experiment. Are the analysis of variance assumptions satisfied?

3.31. Several ovens in a metal working shop are used to heat metal specimens. All the ovens are supposed to operate at the same temperature, although it is suspected that this may not be true. Three ovens are selected at random, and their temperatures on successive heats are noted. The data collected are as follows:

Oven	Temperature					
1	491.50	498.30	498.10	493.50	493.60	
2	488.50	484.65	479.90	477.35		
3	490.10	484.80	488.25	473.00	471.85	478.65

(a) Is there significant variation in temperature between ovens? Use $\alpha = 0.05$.

(b) Estimate the components of variance for this model.

(c) Analyze the residuals from this experiment and draw conclusions about model adequacy.

3.32. An article in the *Journal of the Electrochemical Society* (Vol. 139, No. 2, 1992, pp. 524–532) describes an experiment to investigate the low-pressure vapor deposition of polysilicon. The experiment was carried out in a large-capacity reactor at Sematech in Austin, Texas. The reactor has several wafer positions, and four of these positions are selected at random. The response variable is film thickness uniformity. Three replicates of the experiment were run, and the data are as follows:

Wafer Position	Uniformity		
1	2.76	5.67	4.49
2	1.43	1.70	2.19
3	2.34	1.97	1.47
4	0.94	1.36	1.65

(a) Is there a difference in the wafer positions? Use $\alpha = 0.05$.

(b) Estimate the variability due to wafer positions.

(c) Estimate the rendom error component.

(d) Analyze the residuals from this experiment and comment on model adequacy.

3.33. Consider the vapor-deposition experiment described in Problem 3.32.

(a) Estimate the total variability in the uniformity response.

(b) How much of the total variability in the uniformity response is due to the difference between positions in the reactor?

(c) To what level could the variability in the uniformity response be reduced if the position-to-position variability in the reactor could be eliminated? Do you believe this is a significant reduction?

3.34. An article in the *Journal of Quality Technology* (Vol. 13, No. 2, 1981, pp. 111–114) describes an experiment that investigates the effects of four bleaching chemicals on pulp brightness. These four chemicals were selected at random from a large population of potential bleaching agents. The data are as follows:

Oven	Temperature				
1	77.199	74.466	92.746	76.208	82.876
2	80.522	79.306	81.914	80.346	73.385
3	79.417	78.017	91.596	80.802	80.626
4	78.001	78.358	77.544	77.364	77.386

(a) Is there a difference in the chemical types? Use $\alpha = 0.05$.

(b) Estimate the variability due to chemical types.

(c) Estimate the variability due to random error.

(d) Analyze the residuals from this experimental and comment on model adequacy.

3.35. Consider the single-factor random effects model discussed in this chapter. Develop a procedure for finding a $100(1 - \alpha)\%$ confidence interval on the ratio $\sigma^2/(\sigma_\tau^2 + \sigma^2)$. Assume that the experiment is balanced.

3.36. Consider testing the equality of the means of two normal populations, where the variances are unknown but are assumed to be equal. The appropriate test procedure is the pooled t-test. Show that the pooled t-test is equivalent to the single-factor analysis of variance.

3.37. Show that the variance of the linear combination $\sum_{i=1}^{a} c_i y_{i.}$ is $\sigma^2 \sum_{i=1}^{a} n_i c_i^2$.

3.38. In a fixed effects experiment, suppose that there are n observations for each of the four treatments. Let Q_1^2, Q_2^2, Q_3^2 be single-degree-of-freedom components for the orthogonal contrasts. Prove that $SS_{\text{Treatments}} = Q_1^2 + Q_2^2 + Q_3^2$.

3.39. Use Bartlett's test to determine if the assumption of equal variances is satisfied in Problem 3.24. Use $\alpha = 0.05$. Did you reach the same conclusion regarding equality of variances by examining residual plots?

3.40. Use the modified Levene test to determine if the assumption of equal variances is satisfied in Problem 3.26. Use $\alpha = 0.05$. Did you reach the same conclusion regarding the equality of variances by examining residual plots?

3.41. Refer to Problem 3.22. If we wish to detect a maximum difference in mean response times of 10 milliseconds with a probability of at least 0.90, what sample size should be used? How would you obtain a preliminary estimate of σ^2?

3.42. Refer to Problem 3.26.

(a) If we wish to detect a maximum difference in battery life of 10 hours with a probability of at least 0.90, what sample size should be used? Discuss how you would obtain a preliminary estimate of σ^2 for answering this question.

(b) If the difference between brands is great enough so that the standard deviation of an observation is increased by 25 percent, what sample size should be used if we wish to detect this with a probability of at least 0.90?

3.43. Consider the experiment in Problem 3.26. If we wish to construct a 95 percent confidence interval on the difference in two mean battery lives that has an accuracy of ± 2 weeks, how many batteries of each brand must be tested?

3.44. Suppose that four normal populations have means of $\mu_1 = 50$, $\mu_2 = 60$, $\mu_3 = 50$, and $\mu_4 = 60$. How many observations should be taken from each population so that the probability of rejecting the null hypothesis of equal population means is at least 0.90? Assume that $\alpha = 0.05$ and that a reasonable estimate of the error variance is $\sigma^2 = 25$.

3.45. Refer to Problem 3.44.

(a) How would your answer change if a reasonable estimate of the experimental error variance were $\sigma^2 = 36$?

(b) How would your answer change if a reasonable estimate of the experimental error variance were $\sigma^2 = 49$?

(c) Can you draw any conclusions about the sensitivity of your answer in this particular situation about how your estimate of σ affects the decision about sample size?

(d) Can you make any recommendations about how we should use this general approach to choosing n in practice?

3.46. Refer to the aluminum smelting experiment described in Section 3.8.3. Verify that ratio control methods do not affect average cell voltage. Construct a normal probability plot of the residuals. Plot the residuals versus the predicted values. Is there an indication that any underlying assumptions are violated?

3.47. Refer to the aluminum smelting experiment in Section 3.8.3. Verify the ANOVA for pot noise summarized in Table 3.16. Examine the usual residual plots and comment on the experimental validity.

3.48. Four different feed rates were investigated in an experiment on a CNC machine producing a component part used in an aircraft auxiliary power unit. The manufacturing engineer in charge of the experiment knows that a critical part dimension of interest may be affected by the feed rate. However, prior experience has indicated that only dispersion effects are likely to be present. That is, changing the feed rate does not affect the *average* dimension, but it could affect dimensional variability. The engineer makes five production runs at each feed rate and obtains the standard deviation of the critical dimension (in 10^{-3} mm). The data are shown below. Assume that all runs were made in random order.

Feed Rate (in/min)	Production Run 1	2	3	4	5
10	0.09	0.10	0.13	0.08	0.07
12	0.06	0.09	0.12	0.07	0.12
14	0.11	0.08	0.08	0.05	0.06
16	0.19	0.13	0.15	0.20	0.11

(a) Does feed rate have any effect on the standard deviation of this critical dimension?

(b) Use the residuals from this experiment to investigate model adequacy. Are there any problems with experimental validity?

3.49. Consider the data shown in Problem 3.22.

(a) Write out the least squares normal equations for this problem and solve them for $\hat{\mu}$ and $\hat{\tau}_i$, using the usual constraint $\left(\sum_{i=1}^{3} \hat{\tau}_i = 0\right)$. Estimate $\tau_1 - \tau_2$.

(b) Solve the equations in (a) using the constraint $\hat{\tau}_3 = 0$. Are the estimators $\hat{\tau}_i$ and $\hat{\mu}$ the same as you found in (a)? Why? Now estimate$\tau_1 - \tau_2$ and compare your answer with that for (a). What statement can you make about estimating contrasts in the τ_i?

(c) Estimate $\mu + \tau_1$, $2\tau_1 - \tau_2 - \tau_3$, and $\mu + \tau_1 + \tau_2$ using the two solutions to the normal equations. Compare the results obtained in each case.

3.50. Apply the general regression significance test to the experiment in Example 3.5. Show that the procedure yields the same results as the usual analysis of variance.

3.51. Use the Kruskal–Wallis test for the experiment in Problem 3.23. Compare the conclusions obtained with those from the usual analysis of variance.

3.52. Use the Kruskal–Wallis test for the experiment in Problem 3.23. Are the results comparable to those found by the usual analysis of variance?

3.53. Consider the experiment in Example 3.5. Suppose that the largest observation on etch rate is incorrectly recorded as 250 Å/min. What effect does this have on the usual analysis of variance? What effect does it have on the Kruskal–Wallis test?

3.54. A textile mill has a large number of looms. Each loom is supposed to provide the same output of cloth per minute. To investigate this assumption, five looms are chosen at random, and their output is noted at different times. The following data are obtained:

Loom	Output (lb/min)				
1	14.0	14.1	14.2	14.0	14.1
2	13.9	13.8	13.9	14.0	14.0
3	14.1	14.2	14.1	14.0	13.9
4	13.6	13.8	14.0	13.9	13.7
5	13.8	13.6	13.9	13.8	14.0

(a) Explain why this is a random effects experiment. Are the looms equal in output? Use $\alpha = 0.05$.

(b) Estimate the variability between looms.

(c) Estimate the experimental error variance.

(d) Find a 95 percent confidence interval for $\sigma_\tau^2 / (\sigma_\tau^2 + \sigma^2)$.

(e) Analyze the residuals from this experiment. Do you think that the analysis of variance assumptions are satisfied?

(f) Use the REML method to analyze this data. Compare the 95 percent confidence interval on the error variance from REML with the exact chi-square confidence interval.

3.55. A manufacturer suspects that the batches of raw material furnished by his supplier differ significantly in calcium content. There are a large number of batches currently in the warehouse. Five of these are randomly selected for study.

A chemist makes five determinations on each batch and obtains the following data:

Batch 1	Batch 2	Batch 3	Batch 4	Batch 5
23.46	23.59	23.51	23.28	23.29
23.48	23.46	23.64	23.40	23.46
23.56	23.42	23.46	23.37	23.37
23.39	23.49	23.52	23.46	23.32
23.40	23.50	23.49	23.39	23.38

(a) Is there significant variation in calcium content from batch to batch? Use $\alpha = 0.05$.

(b) Estimate the components of variance.

(c) Find a 95 percent confidence interval for $\sigma_\tau^2 / (\sigma_\tau^2 + \sigma^2)$.

(d) Analyze the residuals from this experiment. Are the analysis of variance assumptions satisfied?

(e) Use the REML method to analyze this data. Compare the 95 percent confidence interval on the error variance from REML with the exact chi-square confidence interval.

CHAPTER 4

Randomized Blocks, Latin Squares, and Related Designs

CHAPTER OUTLINE

The supplemental material is on the textbook website www.wiley.com/college/montgomery.

4.1 The Randomized Complete Block Design

In any experiment, variability arising from a nuisance factor can affect the results. Generally, we define a **nuisance factor** as a design factor that probably has an effect on the response, but we are not interested in that effect. Sometimes a nuisance factor is **unknown and uncontrolled**; that is, we don't know that the factor exists, and it may even be changing levels while we are conducting the experiment. **Randomization** is the design technique used to guard against such a "lurking" nuisance factor. In other cases, the nuisance factor is **known but uncontrollable**. If we can at least observe the value that the nuisance factor takes on at each run of the experiment, we can compensate for it in the statistical analysis by using the **analysis of covariance**, a technique we will discuss in Chapter 14. When the nuisance source of variability is **known and controllable**, a design technique called **blocking** can be used to systematically eliminate its effect on the statistical comparisons among treatments. Blocking is an extremely important design technique used extensively in industrial experimentation and is the subject of this chapter.

To illustrate the general idea, reconsider the hardness testing experiment first described in Section 2.5.1. Suppose now that we wish to determine whether or not four different tips produce different readings on a hardness testing machine. An experiment such as this might be part of a