3. *Use of a pooled SD* We have seen that pooling all of the within-sample variability into a single pooled SD leads to a better estimate of the common population SD and thus to a more precise analysis. This is particularly advantageous if the individual sample sizes (n's) are small, in which case the individual SD estimates are quite imprecise. Of course, using a pooled SD is proper only if the population SDs are equal. It sometimes happens that one cannot take advantage of pooling the SDs because the assumption of equal population SDs is not tenable. One approach that can be helpful in this case is to analyze a transformed variable, such as $\log(Y)$; the SDs may be more nearly equal in the transformed scale.

OTHER EXPERIMENTAL DESIGNS

The techniques of this chapter are valid only for independent samples. But the basic idea—partitioning variability within and between treatments into interpretable components—can be applied in many experimental designs. For instance, all the techniques discussed in this chapter can be adapted (by suitable modification of the SE calculation) to analysis of data from an experiment with more than two experimental factors or situations for which all or some experimental factors are numeric rather than categorical. These and related techniques belong to the large subject called *analysis of variance*, of which we have discussed only a small part.

NONPARAMETRIC APPROACHES

There are k-sample analogs of the Wilcoxon-Mann-Whitney test and other nonparametric tests (e.g., the Kruskal–Wallis test). These tests have the advantage of not assuming underlying normal distributions. However, many of the advantages of the parametric techniques—such as the use of linear combinations—do not easily carry over to the nonparametric setting.

Supplementary Exercises 11.S.1–11.S.20

(*Note*: Exercises preceded by an asterisk refer to optional sections.)

11.S.1 Consider the research described in Exercise 11.4.6 in which 10 women in an aerobic exercise class, 10 women in a modern dance class, and a control group of 9 women were studied. One measurement made on each woman was change in fat-free mass over the course of the 16-week training period. Summary statistics are given in the following table.[8] The ANOVA SS(between) is 2.465 and the SS(within) is 50.133.

	Aerobics	Modern dance	Control
Mean	0.00	0.44	0.71
SD	1.31	1.17	1.68
n	10	10	9

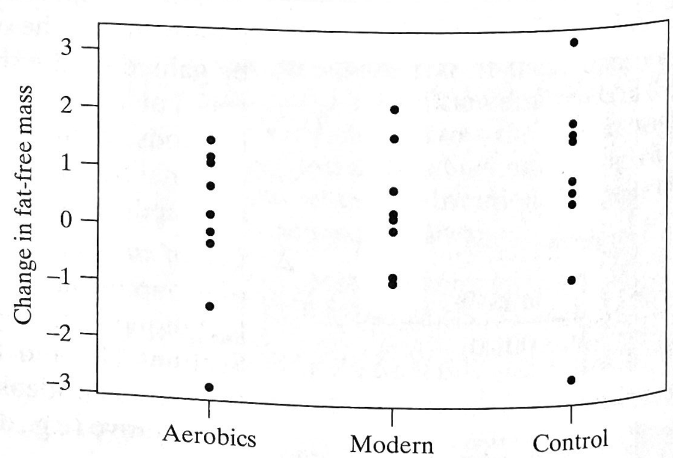
(a) State in words, in the context of this problem, the null hypothesis that is tested by the ANOVA.

(b) Construct the ANOVA table and test the null hypothesis. Let $\alpha = 0.05$.

11.S.2 Refer to Exercise 11.S.1. The F test is based on certain conditions concerning the population distributions.

(a) State the conditions.

(b) The following dotplots show the raw data. Based on these plots and on the information given in Exercise 11.S.1, does it appear that the F test conditions are met? Why or why not?



11.S.3 In a study of the eye disease retinitis pigmentosa (RP), 211 patients were classified into four groups according to the pattern of inheritance of their disease. Visual acuity (spherical refractive error, in diopters) was measured for each eye, and the two values were then averaged to give one observation per person. The accompanying table shows the number of persons in each group and the group mean refractive error.[38] The ANOVA of the 211 observations yields SS(between) = 129.49 and SS(within) = 2,506.8. Construct the ANOVA table and carry out the F test at $\alpha = 0.05$.

Group	Number of persons	Mean refractive error
Autosomal dominant RP	27	+0.07
Autosomal recessive RP	20	−0.83
Sex-linked RP	18	−3.30
Isolate RP	146	−0.84
Total	211	

11.S.4 (*Continuation of Exercise 11.S.3*) Another approach to the data analysis is to use the eye, rather than the person, as the observational unit. For the 211 persons there were 422 measurements of refractive error; the accompanying table summarizes these measurements. The ANOVA of the 422 observations yields SS(between) = 258.97 and SS(within) = 5,143.9.

Group	Number of eyes	Mean refractive error
Autosomal dominant RP	54	+0.07
Autosomal recessive RP	40	−0.83
Sex-linked RP	36	−3.30
Isolate RP	292	−0.84
Total	422	

(a) Construct the ANOVA table and bracket the P-value for the F test. Compare with the P-value obtained in Exercise 11.S.3. Which of the two P-values is of doubtful validity, and why?

(b) The mean refractive error for the sex-linked RP patients was −3.30. Calculate the standard error of this mean two ways: (i) regarding the person as the observational unit and using s_{pooled} from the ANOVA of Exercise 11.S.3; (ii) regarding the eye as the observational unit and using s_{pooled} from the ANOVA of this exercise. Which of these standard errors is of doubtful validity, and why?

***11.S.5** In a study of the mutual effects of the air pollutants ozone and sulfur dioxide, Blue Lake snap beans were grown in open-top field chambers. Some chambers were fumigated repeatedly with sulfur dioxide. The air in some chambers was carbon filtered to remove ambient ozone. There were three chambers per treatment combination, allocated at random. After one month of treatment, total yield (kg) of bean pods was recorded for each chamber, with results shown in the accompanying table.[39] For these data, SS(between) = 1.3538 and SS(within) = 0.27513.

	Ozone absent		Ozone present	
	Sulfur dioxide		Sulfur dioxide	
	Absent	Present	Absent	Present
	1.52	1.49	1.15	0.65
	1.85	1.55	1.30	0.76
	1.39	1.21	1.57	0.69
Mean	1.587	1.417	1.340	0.700
SD	0.237	0.181	0.213	0.056

(a) Construct the ANOVA table.

(b) The P-value for a one-way ANOVA F test is 0.0002. If $\alpha = 0.05$, what is your conclusion regarding the null hypothesis?

***11.S.6** Consider the data from Exercise 11.S.5. For these data, SS(ozone) = 0.696, SS(sulfur) = 0.492, SS(interaction) = 0.166, and SS(within) = 0.275.

(a) Prepare an interaction graph (like Figure 11.7.3).

(b) Construct the ANOVA table.

(c) What is the value of the F test statistic for interactions?

(d) The P-value for the interactions F test is 0.058. If $\alpha = 0.05$, what is your conclusion regarding the null hypothesis?

(e) The P-value for testing that ozone has no effect is 0.0019. If $\alpha = 0.05$, what is your conclusion regarding the null hypothesis?

***11.S.7** Refer to Exercise 11.S.5. Define contrasts to measure each effect specified, and calculate the value of each contrast.

(a) The effect of sulfur dioxide in the absence of ozone

(b) The effect of sulfur dioxide in the presence of ozone

(c) The interaction between sulfur dioxide and ozone

***11.S.8** (*Continuation of Exercises 11.S.6 and 11.S.7*) For the snap-bean data, use a t test to test the null hypothesis of no interaction against the alternative that sulfur dioxide is more harmful in the presence of ozone than in its absence. Let $\alpha = 0.05$. How does this compare with the F test of Exercise 11.S.6(b) (which has a nondirectional alternative)?

***11.S.9** (*Computer exercise*) Refer to the snap-bean data of Exercise 11.S.5. Apply a reciprocal transformation to the data. That is, for each yield value Y, calculate $Y' = 1/Y$.

(a) Calculate the ANOVA table for Y' and carry out the F test.

(b) It often happens that the SDs are more nearly equal for transformed data than for the original data. Is this true for the snap-bean data when a reciprocal transformation is used?

(c) Make a normal quantile plot of the residuals, $(y'_{ij} - \bar{y}'_i)$. Does this plot support the condition that the populations are normal?

***11.S.10** (*Computer exercise—continuation of Exercises 11.S.8 and 11.S.9*) Repeat the test in Exercise 11.S.7 using Y' instead of Y, and compare with the results of Exercise 11.S.7.

11.S.11 Suppose a drug for treating high blood pressure is to be compared to a standard blood pressure drug in a study of humans.

(a) Describe an experimental design for a study that makes use of blocking. Be careful to note which parts of the design involve randomness and which parts do not.

(b) Can the experiment you described in part (a) involve blinding? If so, explain how blinding could be used.

11.S.12 In a study of balloon angioplasty, patients with coronary artery disease were randomly assigned to one of four treatment groups: placebo, probucol (an experimental drug), multivitamins (a combination of beta carotene, vitamin E, and vitamin C), or probucol combined with multivitamins. Balloon angioplasty was performed on each of the patients. Later, "minimal luminal diameter" (a measurement of how well the angioplasty did in dilating the artery) was recorded for each of the patients. Summary statistics are given in the following table.[40]

	Placebo	Probucol	Multi-vitamins	Probucol and multi-vitamins
n	62	58	54	56
Mean	1.43	1.79	1.40	1.54
SD	0.58	0.45	0.55	0.61

(a) Complete the ANOVA table and bracket the P-value for the F test.

Source	df	SS	MS	F
Between treatments	____	5.4336	____	____
Within treatments	____	____	____	____
Total	229	73.9945	____	____

(b) What is the value of the F test statistic for testing that the effects of the four treatments are equal?

(c) How many degrees of freedom are there for the test from part (b)?

(d) The P-value for a one-way ANOVA F test is 0.0006. If $\alpha = 0.01$, what is your conclusion regarding the null hypothesis?

***11.S.13** Refer to Exercise 11.S.12. Define contrasts to measure each effect specified, and calculate the value of each contrast.

(a) The effect of probucol in the absence of multivitamins

(b) The effect of probucol in the presence of multivitamins

(c) The interaction between probucol and multivitamins

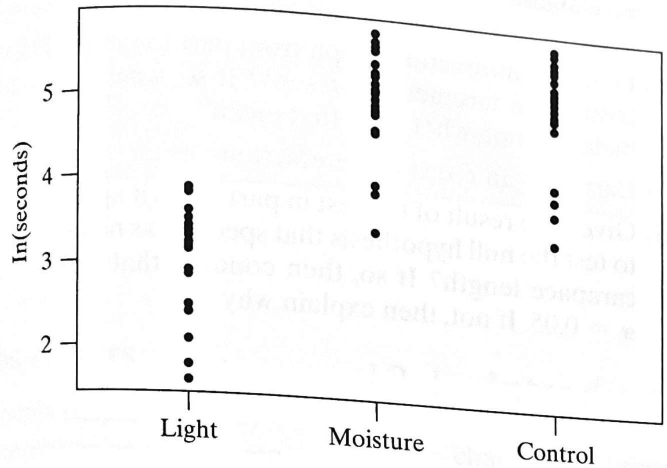
***11.S.14** Refer to Exercise 11.S.12. Construct a 95% confidence interval ($\alpha_{cw} = 0.05$) for the effect of probucol in the absence of multivitamins. That is, construct a 95% confidence interval for $\mu_{\text{probucol}} - \mu_{\text{placebo}}$.

***11.S.15** Refer to Exercise 11.S.12. Assuming all possible comparisons of group means will be computed, use the Bonferroni method to construct a 95% confidence interval for the effect of probucol in the absence of multivitamins. That is, construct a Bonferroni-adjusted 95% ($\alpha_{ew} = 0.05$) confidence interval for $\mu_{\text{probucol}} - \mu_{\text{placebo}}$.

***11.S.16** Three college students collected several pill-bugs from a woodpile and used them in an experiment in which they measured the time, in seconds, that it took for a bug to move 6 inches within an apparatus they had created. There were three groups of bugs: one group was exposed to strong light, for one group the stimulus was moisture, and a third group served as a control. The data are shown in the following table.[41]

	Light	Moisture	Control
	23	170	229
	12	182	126
	29	286	140
	12	103	260
	5	330	330
	47	55	310
	18	49	45
	30	31	248
	8	132	280
	45	150	140
	36	165	160
	27	206	192
	29	200	159
	33	270	62
	24	298	180
	17	100	32
	11	162	54
	25	126	149
	6	229	201
	34	140	173
Mean	23.6	169.2	173.5
SD	12.3	83.5	86.0
n	20	20	20

Clearly the SDs show that the variability is not constant between groups, so a transformation is needed. Taking the natural logarithm of each observation results in the following dotplots and summary statistics.



	Light	Moisture	Control
Mean	2.99	4.98	4.99
SD	0.65	0.62	0.66

For the transformed data, the ANOVA SS(between) is 53.1103 and the SS(within) is 23.5669.

(a) State the null hypothesis in symbols.

(b) Construct the ANOVA table and test the null hypothesis. Let $\alpha = 0.05$.

(c) Calculate the pooled standard deviation, s_{pooled}.

***11.S.17** Mountain climbers often experience several symptoms when they reach high altitudes during their climbs. Researchers studied the effects of exposure to high altitude on human skeletal muscle tissue. They set up a 2×2 factorial experiment in which subjects trained for 6 weeks on a bicycle. The first factor was whether subjects trained under hypoxic conditions (corresponding to an altitude of 3,850 m) or normal conditions. The second factor was whether subjects trained at a high level of energy expenditure or at a low level (25% less than the high level). There were either 7 or 8 subjects at each combination of factor levels. The accompanying table shows the results for the response variable "percentage change in vascular endothelial growth factor mRNA."[42]

	Hypoxic		Normal	
Energy	Low level	High level	Low level	High level
Mean	117.7	173.2	95.1	114.6
No. of patients	7	7	8	8

Prepare an interaction graph (like Figure 11.7.3).

***11.S.18** Consider the data from Exercise 11.S.17.

(a) Complete the following ANOVA table.

Source	df	SS	MS	F ratio
Between hypoxic and normal	1	12126.5	___	___
Between energy level	1	10035.7	___	___
Interaction	1	___	___	___
within groups	26	56076.0	___	___
Total	29	80738.7	___	___

(b) What is the value of the F test statistic used for testing for interactions?

(c) The P-value for the interactions F test is 0.29. If $\alpha = 0.05$, what is your conclusion regarding the null hypothesis?

(d) Based on your conclusion from part (c), is it sensible to examine the main effects of condition and of energy level?

(e) The P-value for testing the null hypothesis that energy level has no effect on the response is 0.04. If $\alpha = 0.05$, what is your conclusion regarding the null hypothesis?

(f) What is the value of the F test statistic used to test whether the effect on the response of hypoxic training is the same as the effect on the response of normal training?

(g) The P-value for the hypoxic training versus normal training F test is 0.025. If $\alpha = 0.05$, what is your conclusion regarding the null hypothesis?

***11.S.19** In a study to examine the utility of using ammonia gas to sanitize animal feeds, researchers inoculated corn silage with a strain of *Salmonella*. Next, two petri dishes of 5 g of contaminated feed were exposed to concentrated anhydrous ammonia gas and two control petri dishes of 5 g of contaminated feed were not treated with the gas. This experiment was repeated twice, for a total of three trials, as only two petri dishes could be placed in the pressurized gas chamber at any given time. Twenty-four hours after inoculation and gassing, the number of bacterial colonies (colony forming units or cfu) on each dish were counted. Because the data were highly skewed, the log(cfu) was analyzed.[43]

(a) Identify the blocking, treatment, and response variables in this problem.

(b) Complete the following ANOVA table for this blocked analysis.

	df	SS	MS	F ratio
Between treatments	1	1.141	1.141	7.107
Between trials	2	3.611	___	___
Within groups	8	___	___	
Total	11	6.036		

(c) Using the complete table from part (b), is there evidence that the ammonia gas treatment affects the contamination level (i.e., mean log cfu)? Use $\alpha = 0.05$.

(d) Do the preceding analysis and information allow you to infer that ammonia reduces contamination? If not, what other information would be necessary to make such a claim?

11.S.20 A biologist measured the post-orbital carapace length (mm) for specimens of two species of crayfish: Rusty crawfish and Sanborn crawfish. Summary statistics are given below.[44]

	Rusty, female	Rusty, male	Sanborn, female	Sanborn, male
Mean	16.78	21.62	16.64	15.69
SD	2.81	4.26	1.76	2.45
N	9	13	14	28

For these data, SS(Sex) = 11.5 SS(Species) = 198.8, SS(interaction) = 113.5, and SS(within) = 482.7.

(a) Prepare an interaction graph (like Figure 11.7.3) to investigate whether the male versus female difference is the same for Rusty as for Sanborn crayfish.

(b) Does the interaction graph from part (a) suggest that there is an interaction present? If so, what does that mean? If not, what does that mean?

(c) Carry out an F test for interactions; use $\alpha = 0.10$.

(d) Given the result of the test in part (c), is it appropriate to test the null hypothesis that species has no effect on carapace length? If so, then conduct that test using $\alpha = 0.05$. If not, then explain why not.