April 2003: Think beyond simple ANOVA when a factor is time or dose—think ANCOVA. Case A: One-way ANOVA (New Rule, 6.13).
A few corrections have been inserted in blue.

[*At times I encounter information that suggests a useful new rule—evidence that not all the rules have been covered in the book. I will number such new rules according to the chapter in which the rule fits best. So far I have not found rules for which I would create a new chapter, but that possibility is not excluded either, of course*.]

Introduction

If factors such as dose and time are used in a study, the simple analysis of variance does not take into account this structure in the covariate. Hence the statistical rule of thumb below. I will discuss two cases: the one-way ANOVA in this month's discussion and a factorial ANOVA in next month's. The underlying principle is the same.

Rule of Thumb

Think beyond simple ANOVA when a factor is time or dose—think ANCOVA.

Illustration

This is a made-up example for two reasons. First, to make the illustration more graphic. Second, to protect the guilty. The example deals with shelf-life of a standard dose of aspirin which is supposed to contain 325 mg of active ingredient. In this example tablets are stored and then randomly withdrawn from the container for analysis at four month intervals up to two years. The data in Table 1 have been "obtained." (I have created additional structure in the data but will not reveal it until next month's discussion.)

Table 1. Active ingredient (in mgs) in aspirin tablets stored for four (T4), eight (T8), twelve (T12), sixteen (T16), twenty (T20), or twenty-four (T24) months. T0 is the value at baseline.

	T0	T4	T8	T12	T16	T20	T24
	334	332	325	344	321	321	316
	337	345	322	323	327	324	322
	345	325	342	334	325	317	319
	325	332	341	338	337	337	323
	332	336	332	324	329	328	335
	328	334	335	325	331	330	328
n	6	6	6	6	6	6	6
Mean	333.5	334.0	332.8	331.3	328.3	326.2	323.8
S.D.	7.06	6.54	8.18	8.66	5.47	7.08	6.79

A one-way analysis of variance (editing output from STATA) is run on these data with the following results.

Table 2. One-way analysis of variance of amount of aspirin by time. Data from Table 1.

Source of variation	Degrees of freedom	Sum of Squares	Mean Square	F	Prob>F
Time	6	561.33	93.556	1.81	0.12
Error	35	1804.67	51.562		
Total	41	2366.00			

This is not significant and we "conclude" that there is no deterioration over time. This, of course, ignores the time structure in the data. We have also violated Rule 7.3: Always graph the data. If we graph the data we see the following:

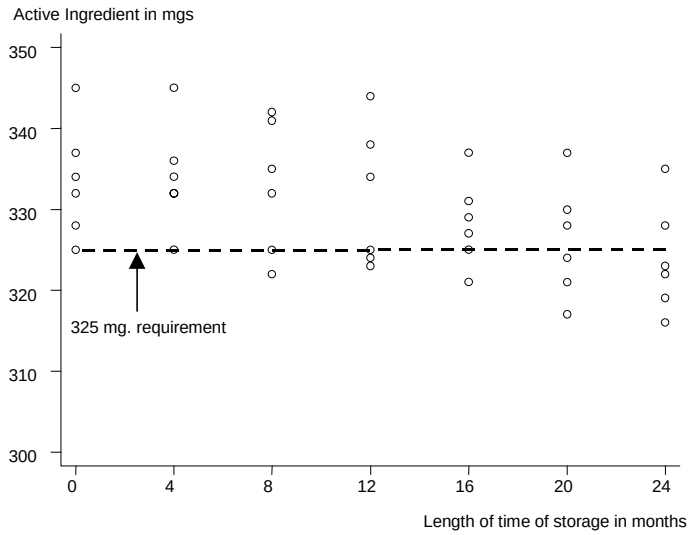

Figure 1. Level of active ingredient in aspirin tablets stored for up to 24 months.

The line at Y=325 defines the required concentration of each tablet. The graphs suggests deterioration in quality. For example, at 24 months all but two of the tablets are below the 325 line.

This suggests trying an analysis of covariance (ACOVA) on these data. The following table was constructed from a STATA regression run on the variable Time and then combining the results with the ANOVA table. (There probably is an easier way to do this.)

Table 3. Analysis of covariance of amount of aspirin, with time as the predictor variable. Data from Table 1.

Source of variation	Degrees of freedom	Sum of Squares	Mean Square	F	Prob>F
Time	6	561.33	93.556		
Linear	1	518.01	518.01	10.05	0.003
Remainder	5	43.32	8.66	0.17	0.98
Error	35	1804.67	51.562		
Total	41	2366.00			

Thus the further partitioning of the treatment sum of squares into a linear component and a remainder confirms the suggestion in the graph of a significant decline in potency.

Basis of the rule

The basis of the rule is that a statistical analysis should incorporate explicit structure in data. As indicated in the introduction, the usual ANOVA ignores the ordering in a factor. Thus, if there is trend in the data the noise may obscure the trend. It thus becomes a question of power. The single degree of freedom test for linear trend has more power than the omnibus test for unequal means.

Discussion and Extensions

1. An omnibus alternative hypothesis, such as the test for Time as a categorical factor in Table 2, has less power than a specific alternative hypothesis, such as the test for trend.

2. Why not use simple linear regression instead of the seemingly more abstruse analysis of covariance? Good question. Here is the result of a simple linear regression of amount of aspirin on time.

Table 4. Regression analysis of amount of aspirin on time. Data from Table 1.

Source of variation	Degrees of freedom	Sum of Squares	Mean Square	F	Prob>F
Time	1	518.01	518.01	11.21	0.002
Error	40	1847.99	46.20		
Total	41	2366.00			

The regression line is,

Amount = 335.3 - 0.439*Time.

The mean square for time in Table 4, MS=518.01, is identical to the mean square for time in Table 3. But the error mean squares differ in their

degrees of freedom and their values. The reason is that there are 6 repeated observations at each time point that the regression analysis does not take into account. The result is that while the slope and intercept estimates are validly estimated by the regression analysis, the precision of those estimates does not take that part of the structure of the data into account. Hence an analysis of covariance is more appropriate—and more informative in automatically producing a test for non-linearity.

3. There should not be a slavish adherence to the rule of not investigating effects further if the overall test of significance—as in Table 2—produces a *p*-value larger than the nominal level, in this case say 0.05. This requires statistical judgment because it may be possible to construct a data-based partitioning that has no scientific basis yet produces a significant result.

4. The interpretation of the linear trend depends on the mechanism that is envisioned. It is NOT useful to start doing pairwise *t*-tests to determine where two times are beginning to differ. The implication of the trend is that they differ at different times. It may be useful from a regulatory perspective to ask about the estimated shelf-life of the product. This could involve a statement that a maximum allowed percentage of the population can be below the 325mg level. If this is suitably formulated a statistical calculation can be made determine the time when this maximum is exceeded. Other criteria are also possible.

5. Partitioning the sum of squares for Time in the analysis of variance can be continued by using orthogonal polynomials or some other rationale for partitioning. In this example, up to a sixth power polynomial is possible. Beyond the linear polynomial none are significant.

6. Since pairwise tests are sometimes done it may be useful to list their limitations. First, they ignore the pattern or trend. Second, they do not make full use of the estimate of error available from variability within times. For the example above, if Time0 is compared with Time4 by means of an unpaired *t*-test the number of degrees of freedom for the error term is only 10 when, in fact, 35 degrees of freedom are available from the analysis of variance and covariance. Finally, doing a series of pairwise tests ignores the multiple comparison problem.

7. Note that in Table 3, the *F*-test for the remainder has a *p*-value of 0.99. This value is too good to be true. In a future note I will discuss such values. In this example it suggests that I was not careful enough in constructing this data set—it is not “random enough.”

8. The simulated data above were inspired by real data in Shao and Chow [1994]. They carry out appropriate, and correct, analyses for shelf-life

data. They also discuss additional issues such as predicting future shelf-life.

9. There are non-parametric analogues for the analysis of covariance. See Hollander and Wolfe [1999] for a discussion. They also discuss a general non-parametric test for trend, the Jonckheere ordered alternative test.

10. These analyses are useful when there is some kind of ordering in one of the factors. It need not be time or dose. Some other examples of ordered categories are education, socio-economic status, or grade of tumors. In these cases a linear trend may not be the appropriate model for this ordering, although it may not be a bad first-order approximation.

Acknowledgements
A version of the above rule was suggested by Dr. Harvey Motulsky, author of the excellent introductory text, *Intuitive Biostatistics*, Oxford University Press, New York, and president of GraphPad Software, a company with several outstanding statistical software packages (see his website at http://www.graphpad.com/). Rules for April and May 2003 have been substantially improved because of very helpful comments byDr. Motulsky on earlier drafts.

I am also indebted to Dr. William Griffith, University of Washington, for discussion of this point and his explanation for the reasons this rule is not followed (see below).

References

Hollander, M. and Wolfe, D.A. [1999]. *Nonparametric Statistics*, second edition. John Wiley and Sons, New York.

Motulsky, H. [1995]. *Intuitive Biostatistics*. Oxford University Press, New York.

Shao, J. and Chow, S-C. [1994]. Statistical inference in stability analysis. *Biometrics*, **50:** 753-763.