One-Way Independent Samples ANOVA with SAS©

Download the file **ANOVA1.sas** from my [SAS programs page](#) and run it. The contrived data (I created them with SAS' normal random number generator) are within the program. We shall imagine that we are evaluating the effectiveness of a new drug (Athenopram HBr) for the treatment of persons with depressive and anxiety disorders. Our independent variable is the daily dose of the drug given to such persons, and our dependent variable is a measure of these persons' psychological illness after two months of pharmacotherapy. We have 20 scores in each of five treatment groups.

The Program

I could have written the INPUT statement this way: "INPUT DOSE ILLNESS;" and then entered one line of data for each subject, each line with two scores (for example, "0 101" for the first subject -- but that would make the file quite long -- 100 lines just for the data. You are certainly free to enter your data that way when you do ANOVAs, but I used a little trick here to make the data entry less tedious. Notice that I first used a statement that inputs treatment level (Dose) and sample size (N). SAS reads the first data line (0 20) and then executes the "**do loop**." That loop tells SAS to execute the following two statements (before the "END" statement) N times. So SAS skips to the next data line, reads the first illness score there, outputs it into the data set, and then, staying on the same line (@@), continues doing this for all N scores. Then it stops and returns to the first input statement, which skips to the next line, reads the Dose value and N there, then the do loop reads the N illness scores on the next line, etc. etc. Note that I could have made this a bit more simple: Since there were 20 scores in each group, I could have replaced the letter N in the do loop with the number 20, and then I would not have to read in N for each group -- but the way I have it set up here shows you that you could also use this trick to read data where there were not equal sample sizes. If you would like to see what the data set looks like after this data step is over, just add a PROC PRINT statement to your program

The first output I requested was a plot of the means. I requested this first because I did not want the plot to be split across pages, but I did want to suppress page ejects to save paper. PROC MEANS is used with the NOPRINT option and the OUTPUT command to create a new data set, named Sol, in which the mean of the illness variable, at each level of the dose variable, is stored in the variable Mean_Illness. PROC PLOT is then used to plot these means by dose, with the first character of the dose levels being used as the plotting symbol. Take a look at the plot. It appears that the drug is quite effective with 10 and 20 mg doses, but that increasing the dosage beyond that reduces its effectiveness (perhaps by creating problems opposite to those it was intended to alleviate).

I used PROC GLM (general linear model) to do the ANOVA. One can use the computationally more efficient PROC ANOVA for one-way ANOVAs, but PROC ANOVA does not support the use of the CONTRAST statement, and I wanted to make some contrasts. Note that I included in the PROC GLM statement specification of the name of the data set (Lotus) to be used. If you do not specify such a data set, the PROC will use the most

---

recently created data set, which, in this case, would be the inappropriate data set Sol. The CLASS variable tells SAS that the dose variable is a classification (categorical, grouping, independent) variable. The MODEL statement tells SAS that illness is the dependent variable and dose the independent variable. The option (after a slash), SS1, tells SAS to compute Type I sums of squares, which are sequential, computationally more efficient than other types of sums of squares, and appropriate for any one-way ANOVA. The EFFECTSIZE option produces a confidence interval for the proportion of variance explained by the model.

The first MEANS statement asks that the means, at levels of the dose variable, be compared with three different pairwise procedures, the Student-Newman-Keuls, the Tukey(a) HSD, and the REGWQ (Ryan/Einot/Gabriel/Welsch). I employed a second MEANS statement so I would get both means and standard deviations, and at the same time I asked for two tests of the null hypothesis that the 5 populations have identical variances (Levene and O'Brien) and also a Welch test (for an omnibus ANOVA without the homogeneity of variance assumption). The CONTRAST statements define a complete set of orthogonal contrasts. I can't really say that these contrasts are of exceptional interest for the current design, but I wanted to show you how to make such contrasts with SAS.

The second invocation of PROC GLM is used to conduct a **Trend Analysis**, using contrast coefficients appropriate for a 5 group design with the treatment levels equally spaced, as they are here. I think these orthogonal contrasts a lot more useful than those computed earlier, and also much more useful than the pairwise comparisons computed earlier.

The third and fourth invocations of PROC GLM are used to conduct the trend analysis as a **Polynomial Regression** analysis. The dose variable, after all, is really a continuous variable, not a categorical variable, so it makes perfect sense to remove it from the CLASS statement and treat it like the continuous variable it is. Doing trend analysis with contrast coefficients rather than directly as a polynomial regression really is just a holdover from past days when calculations were done by hand -- it was somewhat less tedious to do the analysis as an ANOVA with trend contrast coefficients than to do it directly as a polynomial regression.

The Output

We have already discussed the plot. If you remember from high school algebra what a quadratic function is, you should recognize it in this plot -- that is what I had in mind when I generated the population means that I used in my simulator.

The first ANOVA clearly shows that dose is significantly related to illness. The R^2 reported there, .467, is what we have called η^2. Take a look at the critical ranges (least significant differences) for the three pairwise procedures. The Tukey procedure is the least powerful (most conservative) procedure in this group, with the LSD set at 7.95 for all comparisons. That value is the LSD for the outside layer in the other two procedures, with inner layer comparisons having lesser LSDs. For the $r = 4$ comparison, the SNK has the same LSD as does the REGWQ, but for the innermost two layers the SNK has smaller LSDs. Because it does not hold familywise error at the nominal value for a partially true null hypothesis (the REGWQ does and is more powerful than Tukey), the SNK has fallen out of favor with some statisticians. These paranoids seem to think that they will burn in hell if they ever make a Type I error, but they worry too little about the error they are aggravating, the

Type II error. Do note that the very conservative Tukey procedure differs from the other two with respect to which comparisons it declares to be significant.

Note that neither of the homogeneity of variance tests indicates that dose of drug has a significant effect on the variance of the illness scores. Also note that the Welch test, which does not assume homogeneity of variance, does find a significant effect of dose on mean illness. The error *df* for this test are not integer, 47.0276.

If you add up the Contrast sums of squares, you will find that their sum is identical to the treatment (model, dose) sum of squares, as expected, since these contrast coefficients provide a complete and orthogonal partitioning of the treatment sums of squares.

Below is an example of how to write up these results. While the underlining means method of presenting pairwise comparisons is dandy when you are writing by hand, it is cumbersome when you are using a word processor, and you never see it in published manuscripts. Instead, I present such results in a table, using super- or sub-scripts to indicate which means differ significantly from which other means. I elected to report the Tukey test, not because I prefer that test, but because it gave me the opportunity to show you how to display "overlapping lines" in a table like this.

Table 1

Psychological Illness of Patients

As a Function of Dose of Athenopram

Dose (mg)	M	SD
40	101.8^A	10.66
0	100.8^A	8.82
30	92.5^B	7.24
10	85.0^{BC}	11.01
20	81.1^C	6.60

Note. Means with the same letter in their superscripts do not differ significantly from one another according to a Tukey test with a .05 limit on familywise error rate.

An analysis of variance indicated that dose of Athenopram significantly affected psychological illness of our patients, $F(4, 95) = 20.78$, $MSE = 81.71$, $p < .001$, $\omega^2 = .44$, 90% CI [.32, .54]. As shown in Table 1, Tukey's procedure indicated that low doses of the drug were associated with significantly better mental health than were high doses or placebo treatment.

If you elected to conduct the orthogonal contrasts rather than the pairwise comparisons, you could write a report like this:

Table 1.

Orthogonal Contrasts

Contrast	*F*	*p*	95% Confidence Interval for *d*
Placebo vs Drug	22.42	< .001	.67, 1.70
Low vs. High	48.32	< .001	1.06, 2.04
10 mg vs. 20 mg	1.19	.17	-0.19, 1.06
30 mg vs. 40 mg	10.47	.002	0.38, 1.66

An analysis of variance indicated that dose of Athenopram significantly affected psychological illness of our patients, $F(4, 95) = 20.78$, $MSE = 1698$, $p < .001$. As shown in Table 1, Athenopram was effective in reducing the psychological illness of our patients. Low doses of the drug were much more effective than were higher doses. Several patients treated with the higher doses, especially with the 40 mg dose, showed classic symptoms of hypomania during treatment, and a few manifested full blown manic episodes.

I could have computed for each contrast an η^2 by dividing the contrast *SS* by the total SS. Had I done this, I also could have reported a confidence interval for each of those. I decided to report standardized contrasts instead, that is, estimates of the standardized difference between the means of the two sets of groups being compared. For each contrast, one can compute the standardized difference by finding the difference between the two means and dividing by the pooled standard deviation (square root of the *MSE*). For example, for the Low versus High contrast, the two means are 100.8 and 90.1. The pooled standard deviation is 9.039. That yields $g = 1.18$. Instead of reporting the point values of the standardized contrasts, I reported 95% confidence intervals for the standardized contrasts, using my program Conf-Interval-Contrast.sas. For each contrast, if you average the endpoints of the confidence interval you obtain the point estimate of the standardized contrast.

The Trend Analysis

If one's independent variable is fixed but quantitative rather than qualitative, one may be interested in determining not only whether the independent variable accounts for a significant proportion of the variance in the dependent variable, but also the shape of the function relating the dependent variable to the independent variable. For our example, if the relationship between dose (X in the equations that follow) and illness (Y) can be well described by a function in the form of $\hat{Y} = a + b_1 X$, then we would describe the relationship as **linear**. The relationship might, however, be better described by a higher order polynomial, such as $\hat{Y} = a + b_1 X + b_2 X^2$, a **quadratic**. For the X-values -1.5, -1.0, -0.5, 0, 0.5, and 1.0, compute $Y = X + X^2$. Plot the Y-values on the ordinate and the X-values on the abscissa. Notice that there is one reversal in the slope of the quadratic function. A **cubic** function, of the form $\hat{Y} = a + b_1 X + b_2 X^2 + b_3 X^3$, would have two reversals in slope, a quartic three reversals, and so on.

One can determine whether the function is significantly linear, quadratic, cubic, etc. by employing orthogonal polynomial contrast coefficients. A general method for constructing these is presented in Keppel (*Design and Analysis*, 1973, Prentice-Hall, pages 581-588). When the levels of X are equally spaced (as 0, 20, 40, and 60 seconds are), one can use coefficients found in tables such as that on page 599 of Keppel. I provide here coefficients for designs with up to six levels of X:

Levels of X	Trend	Coefficients
3	Linear	−1 0 1
3	Quadratic	1 −2 1
4	Linear	−3 −1 1 3
4	Quadratic	1 −1 −1 1
4	Cubic	−1 3 −3 1
5	Linear	−2 −1 0 1 2
5	Quadratic	2 −1 −2 −1 2
5	Cubic	−1 2 0 −2 1
5	Quartic	1 −4 6 −4 1
6	Linear	−5 −3 −1 1 3 5
6	Quadratic	5 −1 −4 −4 −1 5
6	Cubic	−5 7 4 −4 −7 5
6	Quartic	1 −3 2 2 −3 1
6	Quintic	−1 5 −10 10 −5 1

Try plotting these coefficients and note the shape of the resulting curve—no bend = linear, one bend = quadratic, two = cubic, etc.

We have five levels of X, so we can obtain contrasts for linear, quadratric, cubic, and quartic components of the relationship between X and Y. One way to do this is first to obtain the four treatment means and then find the Pearson r^2 between the linear coefficients and the treatment means. For our example, the linear coefficient - treatment mean pairs to be correlated are (−2, 100.8), (−1, 85.05), (0, 81.1), (1, 92.5), and (2, 101.75). The r^2 is .0257. Next, obtain the among groups sum of squares for an ANOVA on Y with X as the classification variable. Multiply the linear r^2 just obtained by that among sum of squares and you have the sum of squares due to the linear component of the relationship between X and Y, (.0257)(6791.54) = 175. An *F* to test the significance of this component is obtained by dividing this sum of squares by the overall ANOVA *MSE*. It has one *df* in the numerator.

The quadratic and cubic components are tested in the same way, using the appropriate coefficients.

If you are using SAS to do the ANOVA, it is simple to get it to do the trend analysis as well - just provide it with the appropriate orthogonal polynomial contrast coefficients. Take a look at the output from ANOVA1.sas. Look at the output from the second set of contrast statements used with PROC GLM. You see that although both the quadratic and cubic components are significant, the quadratic component accounts for a much larger proportion of the variance in Y. Of course, we don't know what would have happened had we had additional, higher levels of dose. The curve probably would continue upwards (until the drug killed our patients, at which point their problems would be over), but there is a hint that its slope in declining beyond 40 mg.

Please note that several of the computational steps in the analysis as we have done it so far require that the levels of the independent variable be equally spaced and that there be an equal number of subjects in each group. If that is not true, which is especially likely in nonexperimental research, the most straight-forward approach is a **polynomial regression analysis** (see Pedhazur, *Multiple Regression in Behavioral Research*, 2nd Edition, 1982, especially pages 420-433). The analysis will be a <u>sequential</u> analysis. We shall first obtain the sum of squares due to a strictly linear effect of dose. Then we shall add $dose^2$ to the model and obtain the increase in SS due to $dose^2$. Then we shall add $dose^3$ to the model (which already has dose and $dose^2$) and obtain the increase in SS due to adding $dose^3$. We continue up to the highest power (number of levels of dose minus one), or we can at each step test the contribution of all powers not yet in (simultaneously) and stop as soon as that yields a nonsignificant result. This latter method can save you a lot of work if you are doing the analysis by hand, but with a computer doing the work, time is not an issue.

Look back at the program. Note that I have computed $dose^2$ (variable named Quadratic) , $dose^3$ (Cubic) and $dose^4$ (Quartic) in the data step DATA POLYNOMIAL. Dose is no longer identified as being a CLASSification variable and the model includes Dose (linear effect), Quadratic, Cubic, and Quartic. Look at the output. The Type I (sequential) SS are the same SS previously obtained using contrast coefficients. If we sum the SS for all four trends, we get the total treatment SS, 6791.54, so the nonlinear trends, not yet in, jointly account for a SS of $6791.54 - 174.845 = 6616.695$. This is a large proportion of the total SS due to the effect of dose of the drug ($6616.695/6791.54 = .97$), so it appears that we need to add higher powers of dose. We can use an F-test to see if there is a significant nonlinear (powers of dose above one) effect. Each power has one df, so the F for the joint contribution of Quadratic, Cubic, and Quartic, is $(6616.695/3)/(81.71263) = 26.99$ on 3, 95 df. This is significant, so we do indeed need to consider higher powers. Were the nonlinear effect of dose not significant, we could stop right here. In that case we might well want to put the SS and df that belong to the higher trends into the error term. For our example, that would raise the error df from 95 to 98 and the error SS from 7762.7 to 14,379.395, and the F for the linear only model would be computed use those error SS and df.

Now, look at the SS contributed by $dose^2$ when added to the model containing dose. That SS, 6101, is the same obtained earlier with the quadratic contrast coefficients. Do we need add yet higher powers? The SS from all higher powers (now just $dose^3$ and $dose^4$) is 515.81, a relatively small proportion of the total effect of the dose of the drug ($515.81/6791.54 = .08$) but a statistically "significant" one, $F(1, 95) = (515.81/2)/(81.71263) = 3.156$, $p = .047$, so we keep the cubic trend in the model.

Now we are ready to evaluate the significance of the trends. One method for doing this is to evaluate each power using an error term that includes only variance not accounted for by any of the $K–1$ (levels of Dose minus one) powers that could have been included (do note that including all $K–1$ powers accounts for all of the variance due to the independent variable, linear and nonlinear, thus giving you the same model and error SS obtained with a standard ANOVA using Dose as a CLASSification variable). This is what is most often done, and is what is shown in our SAS output with all four trends retained.

An alternative method, which I think has considerable merit, is to use an error term that includes only variance not accounted for by the powers actually retained in the model, after dropping from the model effects which are trivial and/or not statistically significant. This has the advantage of gaining error df (and thus power) without gaining much increase in error SS (since powers not retained should have low SS). Please note that since this is a polynomial regression, one must retain all powers lower than the highest power retained, even if they are nonsignificant. Our dose^2 effect is significant but dose^1 not, but dose^1 is retained in the model, since the quadratic polynomial is $\hat{Y} = a + b_1 X + b_2 X^2$, not just

$\hat{Y} = a + bX^2$. For our example, this entails putting the dose^4 SS and df back into the error term. The output of the last invocation of PROC GLM in ANOVA1.sas shows the results of that analysis.

Here is an example of how to write up the results APA Style, using the model where all trends were retained. I assume that you have already prepared Figure 1, a plot of the group means. Note that for each effect I have presented eta-squared, the effect's sum of squares divided by the total sum of squares.

Table 1. *Trend Analysis Source Table*

Source	SS	df	MS	F	p	η^2
Dose	6,792	4	1,698	20.786	<.001	.47
Linear	175	1	175	2.14	.147	.01
Quadratic	6,101	1	6,101	74.66	<.001	.42
Cubic	389	1	389	4.76	.032	.03
Quartic	127	1	127	1.55	.216	.01
Error	7,763	95	81.71			
Total	14,554	99				

As shown in Table 1, the trend analysis indicated that dose of Athenopram significantly affected the number of correct responses, with significant quadratic and cubic trends. Dose of the drug explained 47% of the variance in the patient's illness, with the quadratic trend being much stronger than the cubic trend. As shown in Figure 1, mean illness declined sharply with a dose of 10 mg, declined moderately with an additional 10 mg, but then rose sharply with further increases in dose.

You could also present confidence intervals about the reported values of eta-squared, but I have not done so here. Some would prefer to report omega-squared.

Annotated Output – produced by SAS 9.3, HTML with ODS graphics