

## Week 5: Lecture

To determine if the assumption of homoscedasticity is tenable, we can use the Fmax test. To illustrate this consider the following output from SPSS:

Group	Mean	Std Deviation
1	25.6667	12.6659
2	40.6833	6.1712
3	38.0000	8.4423
4	42.7500	3.1659
5	52.4833	1.8809
6	43.3333	5.9135
Total	40.4861	10.6982

Source	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4624.569	5	925.914	17.434	.000
Within Groups	3501.417	66	53.052		
Total	8125.986	71			

$$F_{\max} = (12.6659)^2 / (1.8809)^2 = (160.425) / 3.5377 = 45.346$$

If  $F_{\max}$  is greater than  $F_{\max}$ -critical, we have evidence that the variances are not homogeneous. To Find  $F_{\max}$ -critical, we look into the F-max Table. The degrees of freedom  $v_1$  and  $v_2$  are

$v_1 = k$  and  $v_2 = n - 1$ . For our example,  $v_1 = 6$  and  $v_2 = 12 - 1 = 11$ . The  $F_{\max}$  critical value is found on page 812 of the Kirk text. The value is approximately 6.32. Since  $45.346 > 6.32$ , there is evidence the group variances are not homogeneous. So the assumption of homoscedasticity is not tenable for this problem. Later we will see what to do about this.

For the example above, we would reject the null hypothesis:  $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$ . In the anova summary table under "SIG." the value reported is ".000". We would reject the null hypothesis at the  $\alpha = .05$  level if that value was less than 0.05. We would reject the null hypothesis at the  $\alpha = .01$  level if the value reported there was less than 0.01.

In the traditional way, we would reject the null hypothesis at the  $\alpha = .05$  level if the F-value computed and reported in the ANOVA summary table exceeded the value found in a table for the F distribution. In consulting such a table we would need two degrees of freedom terms:  $v_1$  and  $v_2$ . The  $v_1$  for the completely randomized design would be  $k - 1$  and  $v_2$  would be  $N - k$ . For the example above:  $v_1$  is  $6 - 1 = 5$  and  $v_2$  is  $72 - 6 = 66$ . These would correspond to the degrees of freedom found in the ANOVA summary table under "degrees of freedom" for "Between groups" and "Within groups."

In a table for the F-distribution, the critical value would be  $F = 2.36$  for  $\alpha = .05$  and  $F = 3.31$   $\alpha = .01$ . In each case, the computed F-value = 17.434 exceeds both of them.

The equation for the computed F can be expressed as:

$$F = \frac{\sigma_e^2 + n\sigma_y^2}{\sigma_e^2} \text{ where } \sigma_e^2 \text{ can be thought of as the within groups variance and } \sigma_y^2 \text{ is the variance for}$$

the treatment. Note that if there is no treatment variance, the resulting F value is 1.00 Also note that with a

fixed  $\sigma_e^2$  and  $\sigma_y^2$  F can be increased by increasing the sample size. In fact, in some instances, with a small  $\sigma_y^2$  a significant F can be obtained if n is large enough. In this case, treatment really didn't have much of an effect and that the significant test was just due to sample size. In different words, there was no real effect even though there is a statistically significant effect. In order to determine whether or not we have a real or a sample effect, we can compute one of several strength of association indices. We already mentioned them in last weeks lecture notes, so here we will apply them to our sample problem.

Using the ANOVA Summary table,

$$\eta^2 = \frac{4624.569}{8125.986} = .569 \quad \hat{\rho}I = \frac{924.914 - 53.052}{924.914 + 11(53.052)} = \frac{871.862}{1508.486} = .578$$

$$\omega^2 = \frac{4624.569 - 5(53.052)}{8125.986 + 53.052} = \frac{4359.309}{8179.038} = .533$$

In each case here, we can think of these as the coefficient of determination (correlation coefficient-squared) between the independent and dependent variables. All three indicate that the independent variable accounts for over 50% of the variation in the dependent variable. An interpretation of this index depends on the situation and the phenomena under study. In the study above, one independent variable accounted for over 50% of the variation. There is evidence that the effect is real and not an artifact of large sample sizes. The easiest one to use is eta-squared.

In the 1960's and early 1970's a number of researcher examined the effect size of studies reported in psychology research journals. Their discoveries were shocking in that many studies were highly significant statistically, but showed a very small effect. It was concluded that many of these studies were worthless when it came to making decisions based on the findings. Charles W. Simon, for example, examined 15 years of research studies that appeared in the journal: *Human Factors*. In addition to a number of other methodological problems with Human Factors research, effect size was one of them. Human factors research had such small effect sizes that they were deemed unusable in any practical sense. Philip Gallo at San Diego State University found similar patterns with research done in other areas of psychology.

Related to the topic of sample size is statistical power. Remember for a fixed sample size, the probability of type 1 and type 2 errors vary inversely with each other. That is if the risk of a type 1 error is reduced, the risk of the type 2 error is increased. Many research studies are only concerned with the type 1 error, but doing these studies have been criticized in recent years. In particular, Jacob Cohen's 1994 article in *The American Psychologist* attacks the concern over the type 1 error only. A number to top businesses such as Proctor and Gamble take great efforts in doing their internal research to protect themselves from both types of errors. A type 2 error to P&G means a lost opportunity. P&G has remained amongst one of the top companies in the world because of their attention to the type 2 error. Remember we use  $\alpha$  to represent the probability of a type 1 error and  $\beta$  to represent the probability of a type 2 error.  $1 - \beta$  is referred to as the "power of a test." The idea here is to obtain a test that will have a high level of power and also a low amount of  $\alpha$ . In order to do this one of the recommended methods is to increase the sample size. We can also through some formulas and tables from the non-central F-distribution estimate the power of a study we have just completed. Power is related to effect size.

The general formulas used by Kirk and others are

$$\phi' = \frac{\sqrt{\frac{\sum \beta_i^2}{k}}}{\hat{\sigma}_e / \sqrt{n}} \quad \text{where } \sum \beta_i^2 = \frac{k-1}{n} (MS_{Between} - MS_{Within})$$

and

$$\hat{\sigma}_e = \sqrt{MS_{within}}$$

For our example above,  $\sum \beta_i^2 = \frac{6-1}{12} (924.914 - 53.052) = 363.276$

$$\hat{\sigma}_e = \sqrt{53.052} = 7.284 \quad \sqrt{n} = \sqrt{12} = 3.46 \quad \sqrt{\frac{\sum \beta_i^2}{k}} = \sqrt{60.546} = 7.781$$

$$\hat{\sigma}_e / \sqrt{n} = \frac{7.284}{3.46} = 2.105 \quad \phi' = \frac{7.781}{2.105} = 3.696$$

Now with  $v_1 = 5$ ,  $v_2 = 66$  and  $\alpha = .05$  and approximate power of .99 we get a  $\phi = 2.20$  when using the table for the non-central F-distribution. Our  $\phi'$  is 3.696 and it is greater than 2.20. This tells us that we have way too many subjects in our sample. With our  $\phi'$  we would have a power ( $1 - \beta$ ) that is beyond 0.99.

So if we only want a power equal to .97, we are going to have to drastically reduce our sample size. We can go back to the original formula and substitute different values of  $n$  to see what happens to the  $\phi'$  value. We can also take a different approach. Consider from the table,  $v_1 = 5$ ,  $v_2 = 30$  for  $1 - \beta = .97$ . The  $\phi$  there is about 2.15. This would correspond to an  $n = 6$  for our study. If we go back and substitute  $n = 6$  into our equations and get a new  $\phi'$ , we would get

$$\hat{\sigma}_e / \sqrt{n} = \frac{7.284}{2.45} = 2.973 \quad \phi' = \frac{7.781}{2.973} = 2.62$$

this is still larger than 2.15 but certainly closer than  $n = 12$ . We can then try  $n = 5$  and this would give us  $v_1 = 5$  and  $v_2 = 24$  and  $1 - \beta = .97$  and  $\phi = 2.10$  to compare with. The

$$\phi' = \frac{7.781}{(7.284/2.236)} = 2.15$$

This tells us that we would do quite well with  $n = 5$  in obtaining a power of .97. This is made possible because the effect size is quite large. So the effect size,  $\eta^2$  we computed above can be evaluated as being large since a smaller sample size yields a good power value.

### Week 5: LAB

In our ANOVA, we found that the homogeneity of variance assumption is not tenable with our data. The  $F_{max}$  and the Levene test (in SPSS output) states that the group-by-group variances are significantly different. So what can we do about this. We can cite the research that was done years ago by Norton (in Linquist, 1953) and others that showed that ANOVA is robust to violations of its assumptions so long as we have equal sample sizes or something close to equal sample sizes. Without equal sample sizes, the robustness of the ANOVA weakens sufficiently.

Another recommendation is to perform a transformation on the dependent variable in an effort to equalize the variances between groups. The five most common transformations are

1. square root
2. logarithm (natural or base 10)
3. inverse.
4. square
5. exponential

We will try each of these. The results will show that only the square and the exponential will yield a slight improvement over the raw scores. Through experimentation and the use of other approaches will be find a transformation that is suitable.