

A Primer on the Johnson-Neyman Technique: An Alternative Procedure to ANCOVA

Xuejun Ryan Ji

Texas A & M University

The homogeneity of regression coefficients is one of the most important assumptions for appropriate of ANCOVA (Huitema, 1980). However, this assumption is not always met, but usual to reveal the heterogeneity of slopes, and thus leading to biased statistical testing results. Hence, an alternative approach to ANCOVA is required. In the present paper, Johnson- Neyman technique is briefly introduced including a sequence of statistical tests and Statistical Package Syntax.

The homogeneity of regression assumption is one of the most important assumptions underlying the appropriate use of ANCOVA (Huitema, 1980). When this assumption is violated, the ANCOVA is inappropriate way to use. In the following numerical example, the ANOVA F is not statistically significant. If the homogeneity of regression assumption is not tested, the conclusion that there are no treatment effects among groups is erroneous. The biased F_{cal} is the result of violating the ANCOVA assumption. Johnson- Neyman is generally an appropriate alternative analysis. As Johnson and Fay (1950) pointed out,

Modern statistical procedures are primarily concerned with two basic problems –determining significance and estimating parameter values. The Johnson-Neyman technique may be used for both problems; however, its primary contribution is the determination of significance of differences in groups, and its unique and additional contribution is in defining the population, in terms of control variables, for which the conclusion of [statistically] significant difference of mean performance may be held. (p. 349)

In the present paper, a numerical example was presented to further explain the concept of J-N technique. The practical application of this procedure in the present paper was limited to two groups and one covariate. The system of tests described below was based on the procedure by Pedhazur (1997). The numerical example is based on the design matrix by Huitema (1980) shown in Table 1. The data in Table 1 are based on the experiment introduced by Huitema. The treatments are the two methods of therapy. The covariate is the scores on the sociability scale. The dependent variable is the aggressiveness score. Dummy code is used to categorize the group membership.

The overview procedure includes: 1) conduct a full regression model to test whether the main effects and interaction effects are sufficiently meaningful or not, and to obtain the sum of squares for calculating the limits of region of significance; 2) test the homogeneity assumption of slopes by testing the significance of interaction between group and covariate; 3) establish the region of significance.

Step 1. Is the Proportion of Variance Accounted for Meaningful?

This test serves a go/no-go criterion for continuing with the subsequent procedures. $R^2_{y,D,X,DX}$ indicates the proportion of variance accounted for by the main and interaction effects. Whether the $R^2_{y,D,X,DX}$ is statistically significant or not, if the value of $R^2_{y,D,X,DX}$ is too small to be substantively meaningful, the analysis should be terminated, otherwise it is worthy to proceed to the next step (Pedhazur, 1997). We obtained the R Square of .938 which means 93.8% of variance is explained by the main and interaction effects. This result is NOT too small to be meaningful. Hence, the analysis can proceed to the next step. The relevant SPSS command syntax is:

```
REGRESSION VAR Y X D DX/DES/STAT ALL/  
DEP Y/ENTER X/ENTER D/ ENTER DX .
```

The purpose of using STAT ALL subcommand is to obtain the Model Summary Table including the values of R Square Change, F Change and Sig. F Change, which equals the values obtained from ANOVA summary table. For instance, if we still want to know whether there are treatment effects or not, the correspondent F Change and p_{cal} can be obtained directly from the Model Summary Table. In the present case, F Change ratio on treatment effects is 2.282 ($p_{cal} = 0.142$) which is not significant at $\alpha=.05$. Or we can also perform this analysis by using SPSS command GLM. The relevant SPSS syntax is

GLM Y BY D WITH X
/PRINT=DESCRIPTIVE/CRITERIA=ALPHA (.05)/DESIGN.

The results of these three methods are identical, which are shown in Table 2, Table 3 and Table 4.

According to Pedhazur (1997), "...a test of significance can be conceived as an attempt to answer the question: Does additional information add significantly to the explanation of the dependent variable?" (p. 563). Applied to the present numerical example, this statistical test on treatment or the difference between groups can be viewed as the test of the increment in the regression sum of squares due to the dummy vectors, D. Because of the equation: $R^2 = \text{SOS}_{\text{reg}} / \text{SOS}_{\text{total}}$, the present test can also be viewed as the test of the increments in proportion of variance accounted for due to the dummy vectors, D. Hence, the same F ratio = 2.282 can be obtained through these three methods [only for heuristic purpose]. And the present F ratio is not statistically significant at $\alpha = 0.05$. If the homogeneity of regression assumption is met, the researchers can conclude that there are no treatment effects. However, if the assumption of the homogeneity of the regression slopes is rejected, then the conclusion that there are no treatment effects must be erroneous. Further details are discussed in Step 2.

Step 2. Is There a Statistically Significant Interaction? Or Is the Homogeneity of Regression Assumption Met?

In this step, we determine whether the ANCOVA is appropriate procedure for further analysis on difference between treatment groups based on the Yea/Nay of assumption of homogeneity of regression slopes. As Thompson (2006) explained,

What ANCOVA actually does is to create a single regression equation, ignoring groups, to predict dependent variable scores using the covariates... These computations are legitimate if ...homogeneity of regression assumption is met...If this assumption is not met, the use of an average, or "pooled" equation, for all the participants is unreasonable, creates e scores using a single regression equation that may not fit in any of the groups, and thus focuses the analysis on outcome scores that are inaccurate distortion for everybody. (p. 355)

Moreover, the test of ANCOVA assumption can also be conceived of as the test of the statistical significance of interaction effects. As Pedhazur (1997) noted,

A test of coefficient for the product vector [b_{DX}] is the test of the interaction and is equivalent to a test of increment in the proportion of variance accounted for by the product vector, over and above its constituent variables. Also, when one of the variables is categorical and the other is continuous, the test of an interaction addresses the question of whether the regression lines of the dependent variable on the continuous variable are parallel for all the categories of the categorical variable. In other words, a test of an interaction addresses the question of whether differences among regression coefficients from separate regression equations are statistically significant. (p. 583)

Hence, in Step 2, we can perform both two-sample t -test [on b_{DX}] and ANOVA via regression to test the statistical

Table 1. Demonstration Data with Two-Group One-Covariate Design

	Y	X	D	DX
1	10.0	1.0	1.0	1.0
2	10.0	2.0	1.0	2.0
3	11.0	2.0	1.0	2.0
4	10.0	3.0	1.0	3.0
5	11.0	4.0	1.0	4.0
6	11.0	5.0	1.0	5.0
7	10.0	5.0	1.0	5.0
8	11.0	6.0	1.0	6.0
9	11.5	6.0	1.0	6.0
10	12.0	7.0	1.0	7.0
11	12.0	8.0	1.0	8.0
12	11.0	8.0	1.0	8.0
13	11.0	9.0	1.0	9.0
14	12.5	10.0	1.0	10.0
15	12.0	11.0	1.0	11.0
16	5.0	1.0	0.0	0.0
17	6.0	1.5	0.0	0.0
18	6.0	2.5	0.0	0.0
19	7.0	3.5	0.0	0.0
20	8.0	4.5	0.0	0.0
21	9.0	4.5	0.0	0.0
22	9.0	5.0	0.0	0.0
23	9.0	6.0	0.0	0.0
24	10.5	6.0	0.0	0.0
25	11.0	7.0	0.0	0.0
26	12.5	7.0	0.0	0.0
27	12.5	7.5	0.0	0.0
28	14.0	8.0	0.0	0.0
29	14.5	9.0	0.0	0.0
30	16.0	10.0	0.0	0.0

significance of interaction effects, and expect to reach the same conclusion. As Thompson (2006) explained,

Multiple regression is the most general case of the univariate general linear model (GLM), and subsumes other univariate parametric analyses (i.e., two- sample t -test, ANOVA, ANCOVA, Pearson r) as special cases. This implies that *all* analyses are correlational, yield r^2 -type effect size, and apply weights to measured variables to estimate latent variable scores. (p. 390)

Hence, the conclusive results via the t -test and ANOVA should be identical. We will conduct the regression analysis by using SPSS syntax. Then we will compare the results obtained from the output table. The relevant syntax is

```
REGRESSION VAR Y X D DX/DES/STAT ALL
/DEP Y/ENTER X/ENTER D/ ENTER DX .
```

or

```
GLM Y BY D WITH X /PRINT=DESCRIPTIVE/CRITERIA=ALPHA (.05)
/DESIGN= X D D*X .
```

We can also build an ANOVA summary table by subtracting the smaller sum of square from the larger one, then producing an F ratio. Table 3, Table 5 and Table 6 present that the obtained F ratio of test of ANOVA assumption via two methods is identical, with value of 141.896, which is statistically significant, when compared with $F_{crit} = 4.23$ ($\alpha = .05$, $df_{reg}=1$, $df_{error}= 26$).

Moreover, the obtained t_{cal} is -11.912, which is statistical significant, when compared with the $p_{cal}=0.000$ [INPUT “= T.DIST.2T (11.912, 26)” in EXCEL (As for some previous versions, INPUT “=TDIST(11.912, 26,2) where the 11.912 is the t_{cal} , 26 is the degree of freedom, and 2 indicates that the test is two tailed)]. Because of the equation: $t_{cal}^2 = F_{cal}$, $(-11.912)^2 = 141.896$, the test results of homogeneity of regression slope via three different methods are identical, all of which are statically significant at $\alpha = .05$. We can conclude that the regression slopes are heterogeneous and can determine that ANCOVA is quite inappropriate which are “an inaccurate distortion to everybody” (Thompson, 2006, p. 583). The alternative technique is required.

Step 3. Establish Region of Significance via Johnson-Neyman Technnique

When the homogeneity of regression assumption is rejected, the ANCOVA is not appropriate procedure for further analysis. In the present paper, the Johnson- Neyman Technique, as an alternative approach, is briefly introduced, with emphasis on two groups, one covariate. The procedure below was similar to the one presented by Huitema (1980).

To establish the limits of the region of non-significance (or significance), two X values need to be computed by using:

$$\begin{aligned} X1 &= \frac{-B - \sqrt{B^2 - AC}}{A} \\ X2 &= \frac{-B + \sqrt{B^2 - AC}}{A} \end{aligned} \quad (1)$$

$$\begin{aligned} A &= \frac{-F_{\alpha(1, N-4)}}{N-4} (SOS_{res}) \left(\frac{1}{\sum x_1^2} + \frac{1}{\sum x_2^2} \right) + (b_1 - b_2)^2 \\ B &= \frac{F_{\alpha(1, N-4)}}{N-4} (SOS_{res}) \left(\frac{\bar{X}_1}{\sum x_1^2} + \frac{\bar{X}_2}{\sum x_2^2} \right) + (a_1 - a_2)(b_1 - b_2) \\ C &= \frac{-F_{\alpha(1, N-4)}}{N-4} (SOS_{res}) \left(\frac{N}{n_1 n_2} + \frac{\bar{X}_1^2}{\sum x_1^2} + \frac{\bar{X}_2^2}{\sum x_2^2} \right) + (a_1 - a_2)^2 \end{aligned}$$

Where $X1$ and $X2$ are limits of non-significance region $F_{\alpha(2, N-4)}$ = critical value of F statistic for desired level of α and 1 and $N-4$ degrees of freedom (N = total number of participants)

SOS_{res} = individual residual sum of squares

\bar{X}_1, \bar{X}_2 = covariate means for sample 1 and sample 2 respectively

$\sum x_1^2, \sum x_2^2$ = covariate sums of squares for sample and sample 2 respectively

a_1 = regression intercept for sample 1

a_2 = regression intercept for sample 2

b_1 = regression slope for sample 1

b_2 = regression slope for sample 2

The required values for the application of the formula to Table 1 are:

$$F_{\alpha(2, N-4)} = 4.225 \quad \bar{X}_1 = 5.8 \quad \bar{X}_2 = 5.533 \quad \sum x_1^2 = 130.4 \quad \sum x_2^2 = 99.23$$

$$SOS_{res} = 10.920 \quad a_1 = 9.857 \quad a_2 = 3.155 \quad b_1 = 0.209 \quad b_2 = 1.237$$

The values for A, B, and C are: A

$$A = 1.0253304 \quad B = -6.7155607 \quad C = 43.71518698$$

$$X1 = 6.04 \quad X2 = 7.06$$

The detailed computations were omitted by Huitema (1980). The author of the present paper conducted the calculation by using the SPSS syntax. Figure 1 presents the SPSS syntax relevant to the Johnson- Neyman technique.

The scores lie within the range of 6.04 and 7.06 are not statistically significantly different across treatments. We can conclude that the two methods of therapies are not different for participants whose sociability scores fall at the specific point within this range. However, there are two regions of significance, one for X above 7.06 and one for X scores below 6.04. If a score falls at the point below 6.04, we can conclude that method 1 is superior to method 2 for participants whose sociability scores fall at the specific point below 6.04. If a score is above 7.06, we can conclude that method 2 is superior (Huitema, 1980).

```

SET PRINTBACK=LISTING.
DATA LIST FILE= 'C:\SPSSWIN\huitema.txt' records=1/1
Y 1-4 X 6-9 D 11-13 .
compute DX= D*X.
LIST VARIABLE= ALL/ FORMAT=NUMBERED.
COMMENT ESTABLISH THE REGIONN OF SIGNIFICANCE.
REGRESSION DEPENDENT= Y/ ENTER D X DX.
TEMPORARY.
SPLIT FILE BY D .
REGRESSION VARIABLES = X Y/DEP Y/ENTER X .
TEMPORARY.
SPLIT FILE BY D .
REGRESSION VARIABLES = X Y/DEP X/ENTER Y.
COMMENT COMPUTATIONS J-N METHOD .
DATA LIST FREE
/Fcrit SSres n1 avg1 SSx1 a1 b1 n2 avg2 SSx2 a2 b2 .
BEGIN DATA
4.225 10.920 15 5.8 130.4 9.857 0.209 15 5.53 99.233 3.155 1.237
END DATA .
list variables =Fcrit SSres n1 avg1 SSx1 a1 b1 n2 avg2 SSx2 a2
b2/FORMAT=NUMBERED.
COMPUTE T1 = (-1)* (Fcrit/(n1+n2-4))*SSres.
COMPUTE T2= (-1) * T1 .
COMPUTE A = T1* ((1/SSx1)+(1/SSx2))+ (b1-b2)**2 .
COMPUTE B = T2* ((avg1/SSx1)+(avg2/SSx2)) + ((a1-a2) * (b1-b2)).
COMPUTE C = T1* (((n1+n2)/(n1*n2))+ ((avg1**2)/SSx1)+((avg2**2)/SSx2)) + (a1-
a2)**2 .
LIST VARIABLES = A B C / FORMAT=NUMBERED.
COMPUTE X1= ((-1)*B +sqrt((B**2)-A*C))/A .
COMPUTE X2 = ((-1)*B - sqrt((B**2)-A*C)) /A .
LIST VARIABLE= X1 X2/FORMAT=NUMBERED.

```

Figure 1. Johnson- Neyman technique SPSS syntax.

Group Differences (Main Effects)

The interaction between group and covariate is statistically significant, indicating that group differences vary upon the level of covariate. Within the region of significance, the group differences are statistically significant across the levels of covariate. While outside the region of significance, the group differences are small or not statistically significant or even zero across the levels of covariate. The group comparison, in fact, is the difference of least squares means (or estimated marginal means) at the mean of covariate. Researcher can also evaluate the group differences (or effects) at other specific values of covariate or even of multiple covariates.

Least squares means (LSmeans in SAS PROC GLM) or estimated marginal means (EMMEANS in SPSS) are the estimated marginal means adjusted for the covariate. Note that these statistics are predicated rather than observed means. One can easily obtain these statistics from SAS or SPSS.

```
/*SPSS SYNTAX*/
GET FILE "TABLE1.SAV"
UNIANOVA Y BY D WITH X
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/EMMEANS=TABLES(D)WITH(X=
8) COMPARE(D)
/CRITERIA=ALPHA(.05)
/DESIGN=X D X*D.
```

```
/*SAS SYNTAX*/
PROC GLM DATA=JNT.TABLE1;
CLASS D;
MODEL Y = D X D*X/SOLUTION;
LSMEANS D/ PDIFF=ALL TDIFF
CL AT X=8;
RUN;
QUIT;
```

In this example, the EMMENAS statement in SPSS and LSMEANS statement in SAS report the marginal means of Y at each level of the grouping variable D at a specific value of X, say 8. Moreover, these two statements also report the group difference along with statistical significance testing. Both statements output the same results: LSmean = 13.051 (D = 0), 95% CIs [12.575,13.528], LSmean = 11.526 (D = 1), 95% CIs [11.096, 11.955]. The group difference at the value of 8 of covariate, X , is - 1.526, 95% CIs [-2.167, -.884]. The outputs are consistent with the results reported in the example illustrated by Huitema (1980). Moreover, the score of 8 is above 7.06, and it is located in the region of significance. Therefore, the group difference at this specific value of X is significant. Of course, one can also calculate the least squares means of each group, and group differences at the overall mean of covariate by simply inputting "X=MEAN" in SPSS or "At means" in SAS.

Multiple Groups and Multiple Covariates

If more than two groups involved, researchers should employ adjustment technique when conducting pairwise comparisons. SPSS users can specify the option "ADJ" with Bonferroni or Sidak correction tests. SAS users can specify the option "ADJUST=" with Bonferroni, Sidak, Tukey and Scheffe correction tests.

If more than one covariate involved, researchers can conduct the group differences at the specific value of each covariate. To illustrate this case, I replicated the example in the textbook by Huitema (1980, p287). In this example, the researcher attempted to determine the treatment effect at the three covariate scores: X1 = 4, X2 = 9, and X3 = 3. SAS PROC GLM allows users to specify the values on multiple covariates. The SAS syntax is displayed as follows.

```
/*SAS SYNTAX*/
PROC GLM DATA=JNT.TABLE3COV;
CLASS D;
MODEL Y = D X1 X2 X3 X1*D X2*D X3*D/ SOLUTION;
LSMEANS D / PDIFF TDIFF CL AT (X1 X2 X3)=(4 9 3);
RUN;QUIT;
```

The SAS outputs are matched with the results reported by Huitema (1980). The treatment effect is - 3.485 with 95% CIs [-4.363,-2.606].

Region of Significance

Originally, Johnson and Neyman (1936) designed a procedure to test the difference between treatment effects at a specific point on X. The established region of significance by the conventional procedure in the present case is non-simultaneous. The modified version was proposed by Potthoff (1964), who extend the procedure for establishing simultaneous region, which enable one to “state with 95 percent confidence that there is a nonzero difference between the two groups simultaneously for all points in the region” (Potthoff, p. 241). To calculate simultaneous regions of significance, replace $F_{\alpha(1,N-4)}$ with $2F_{\alpha(2,N-4)}$.

Plotting the Region of Significance. Data visualization can be a conducive way for understanding the story told by the data. One way to plot the region of significance is using STATA plotting function.

```
use"table1.dta", clear
sum X Y
regress Y D##c.X
margins, dydx(D) at(X=(1(1)12)) noatlegend // get points for
graphing
marginsplot, recast(line) recastci(rline) ciopts(color(gs10)) ///
yline(0) name(vv4, replace)
```

Figure 2 presented the group difference line, its confidence interval area, and region of significance based on the Table 1 data. As shown in Figure 2, when group differences confidence interval contains the value of zero, the corresponding upper limit and lower limit values are around 6.04 and 7.06 respectively. At X = 8, the corresponding Y value (group difference) is around -1.52. Hereinafter is the STATA Syntax.

If multiple group and multiple covariates involved, SPSS, SAS and STATA cannot visualize the multi-dimension data. Instead, one can plot the regression hyperspace by using Mathematica. Readers can consult the paper authored by Hunka and Leighton (1997).

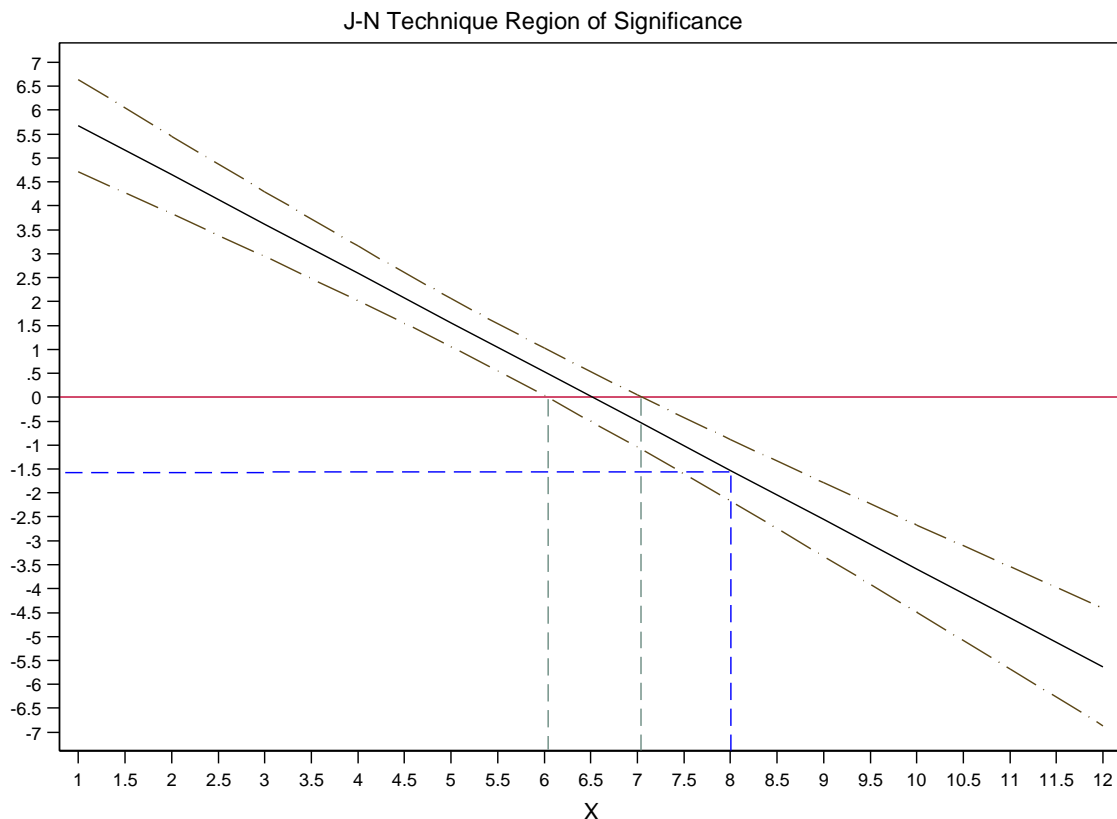


Figure 2. Johnson-Neyman technique region of significance.

Summary

As Huitema (1980) pointed out, “heterogeneous regression slopes associated with ANCOVA present interpretation problems because the magnitude of treat effect is not the same at different levels X” (p. 270). Hence, prior to performing ANCOVA analysis, the assumption of homogeneity of regression must be met. For instance, Even if ANCOVA F is statistically not significant, the conclusion that there are no treatment effects might be incorrect, without testing the homogeneity of regression slopes. If the ANCOVA assumption is rejected, the alternative method, such as Johnson-Neyman technique can be adopted to do the further analysis by establishing region of significance. The present procedure can allow a researcher to make statements about the regions of non-significance and significance by identifying the values of X that are associated with significant group differences on Y.

References

- Huitema, B. E. (1980). *The analysis of covariance and alternatives*. New York: Wiley.
- Hunka, S., & Leighton, J. (1997). Defining Johnson-Neyman regions of significance in the three-covariate ANCOVA using Mathematica. *Journal of Educational and Behavioral Statistics*, 22(4), 361-387.
- Johnson, P. O., & Fay, L. C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika*, 15(4), 349-367.
- Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, 1, 57-93.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: MHarcourt Brace College Publishers.
- Potthoff, R. F. (1964). On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*, 29(3), 241-256.
- Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: Guilford Press.

Send correspondence to:

Xuejun Ryan Ji
 Texas A&M University
 Email: ryanji329@tamu.edu
