Published in final edited form as:

West J Nurs Res. 2004 November ; 26(7): 804–812. doi:10.1177/0193945904266733.

The Johnson-Neyman Procedure as an Alternative to ANCOVA¹

Karen T. D'Alonzo, Ph.D., R.N., A.P.N.C.[Assistant professor]

College of Nursing, Rutgers, The State University of New Jersey

Abstract

ANCOVA is a widely used statistical procedure that is particularly useful in analyzing data from experimental designs. There are, however, a number of assumptions that must be tested before proceeding with the ANCOVA. Of particular concern is the assumption of homogeneity of regression slopes (HOS). When the HOS assumption has been violated, the researcher needs to look for an alternative approach to the ANCOVA. The Johnson-Neyman procedure (J-N) is presented as such an alternative. Although the calculations for the procedure are somewhat tedious and are not currently a standard feature of statistical software packages, an alternative approach using SAS syntax codes is presented.

Keywords

Johnson-Neyman procedure; ANCOVA; experimental designs

ANCOVA is a useful statistical procedure that is commonly utilized in experimental designs. ANCOVA is essentially a hybrid form of multiple regression and ANOVA and is used to make comparisons between two or more group means after statistically removing the effect of one or more extraneous variables (covariates) on the dependent variable (DV) (Tabachnick & Fidell, 2001). As Huitema (1980) and others have discussed, the purpose of the ANCOVA in experimental designs is to increase the sensitivity of the test of main effects and interactions by reducing the error term. In ANCOVA, the error term is adjusted for the relationship between the DV and the covariate (CV). Tabachnick & Fidell (2001) noted that with ANCOVA, CVs are used to assess the “noise” or undesirable variance in the DV that is estimated by scores on the CV.

ASSUMPTIONS

There are a number of assumptions surrounding the use of ANCOVA (Huitema, 1980). In many instances, violations of the assumptions are not considered serious if there is a relatively large sample size and equal groups. It is of paramount importance that the researcher assesses these assumptions before proceeding with data analysis. Information regarding the testing of assumptions is often omitted in journal articles. Dorsey & Soeken (1996) found that of 30 articles that used ANCOVA and appeared in *Nursing Research* from 1986 to 1996, only one mentioned testing for statistical assumptions.

The first assumption, randomization to groups, is of course the hallmark of a true experiment. This is the assumption most often violated with ANCOVA, which has been used by many researchers to statistically equate two groups when random assignment to groups has not been

possible. The use of ANCOVA in quasi-experimental designs remains a controversial issue (Huitema, 1980; Pedhazur, 1997; Tabachnick & Fidell, 2001). This article focuses mainly on violation of the assumption of homogeneity of regression slopes (HOS). It is this assumption that forms the basis of the Johnson-Neyman procedure.

Most commercially available software programs do not routinely test for the assumption of homogeneity of regression slopes. The researcher can rely on two methods to assess this assumption before proceeding to data analysis.

Scatterplots

A scatterplot of the data is the best way to assess HOS while examining descriptive statistics. Figure 1 shows that when the DV is plotted against the CV, Groups 1 and 2 have similar slopes. These slopes are considered homogenous. Heterogeneity of regression slopes is said to occur when there are differences in the slopes of the two lines. When the difference is small and group sizes are equal, Keppel (1991) and others argued this type of heterogeneity is usually not a significant problem and the ANCOVA remains robust.

In contrast, the type of heterogeneity of regression slopes depicted in Figure 2 presents a potential problem. Significant differences between the group regression coefficients indicate the presence of an interaction between the treatment (grouping variable) and the CV. This implies that the magnitude of the treatment effect is not the same for different levels of the CV. In Figure 2 there are three separate regions of interest to the researcher; a region where the estimated mean for Group 1 is greater than Group 2, a region where the estimated mean for Group 2 is greater than Group 1, and a middle region where the estimated means do not differ significantly. In Figure 2, the shaded area represents this area of insignificance. Misinterpretations of findings can occur when such an interaction is present; for example, it may be concluded that there are no treatment effects because the adjusted means are equal.

p Value of ANOVA Interaction Term

If the scatterplot reveals evidence of a treatment-covariate interaction, the researcher can perform an ANOVA and examine the *p* value of the interaction term. If the *p* value is less than .10, an interaction is present. If the *p* value is slightly below .10, the interaction is usually not problematic. In such cases, the interaction may occur at either end of the scatterplot or it may represent a “theoretical” interaction, in an area of the scatterplot where there are no data points. When this occurs, it is important that the researcher restrict the generalization of results to the range of CV values included in the sample (Huitema, 1980).

If a true treatment-covariate interaction exists, the next step is to determine:

1. What is the region of insignificance for group differences when the treatment effect differs depending on the level of the covariate?
2. For what values of the covariate do the groups differ significantly with regard to the dependent variable?

Approaches to Heterogeneity of Regression Slopes

There are two basic approaches to address the problem of heterogeneity of regression slopes. The first and simplest procedure is to perform a two-factor ANOVA and divide the CV scores into either a high or low category. For example, a researcher wishing to study the effects of an exercise intervention on lowering systolic blood pressure (BP) in hypertensive women might divide the women into high systolic BP and low systolic BP based on the CV values obtained during preintervention testing. Although this approach is simple and straightforward, it does

not provide specific information about the region of insignificance for group differences associated with different levels of treatment.

Johnson & Neyman (1936) developed an alternative procedure for establishing regions of insignificance associated with a test of the difference between two treatments at any specific point on the X continuum. Further modifications to the Johnson-Neyman procedure were made by Pothoff (1964) and Rogosa (1977). Assumptions for the J-N procedure include all those previously mentioned for the ANCOVA, with the exception of the homogeneity of regression slopes (Huitema, 1980). The three steps in the computation of the Johnson-Neyman procedure are as follows (Aikens & West, 1991):

Step I: Regression Analyses—The continuous predictor variable is regressed on the outcome variable using the data from each group. In an experimental study:

$$\begin{aligned}\text{Group 1 (Experimental): } Y_E &= b_{1(E)} (\text{Predictor variable}) + b_{0(E)} \\ \text{Group 2 (Control): } Y_C &= b_{1(C)} (\text{Predictor variable}) + b_{0(C)}\end{aligned}$$

Step II: Calculation of Intermediate Quantities

$$\begin{aligned}A &= \frac{-2F_{2,N-4}}{N-4} (SS_{RES}) \left[\frac{1}{SS_{X(1)}} + \frac{1}{SS_{X(2)}} + [b_{1(1)} - b_{1(2)}]^2 \right] \\ B &= \frac{2F_{2,N-4}}{N-4} (SS_{RES}) \left[\frac{M_1}{SS_{X(1)}} + \frac{M_2}{SS_{X(2)}} + [b_{0(1)} - b_{0(2)}][b_{1(1)} - b_{1(2)}] \right] \\ C &= \frac{-2F_{2,N-4}}{N-4} (SS_{RES}) \left[\frac{N}{n_1 n_2} + \frac{M_1^2}{SS_{X(1)}} + \frac{M_2^2}{SS_{X(2)}} + [b_{0(1)} - b_{0(2)}]^2 \right]\end{aligned}$$

Step III: Calculation of Region Cutoffs

$$x = \frac{-B + (B^2 - AC)^{1/2}}{A} \text{ and } x = \frac{-B - (B^2 - AC)^{1/2}}{A}$$

X represents the region of insignificance for group differences when the treatment effect differs. In the previous example, a researcher may find that the exercise intervention produced no significant changes in systolic BP in those participants with preintervention systolic BPs less than 145 mm Hg, suggesting that the intervention was most effective for those women with BPs above 145 mm Hg. (Figure 3). In this case, if the sample contained many participants with mild to moderate hypertension, the results of the ANCOVA alone would have revealed that the intervention produced no significant differences between the control and experimental groups.

Calculation of the formulas for the Johnson-Neyman technique is tedious and is currently not a standard feature of any statistical software package. Pedhazur (1997) wrote syntax commands to run the procedure using the 1993 version of the Statistical Package for the Social Sciences (SPSS), however these commands cannot be used with more recent editions of the SPSS software. Jenn-Yun Tein (in Aiken & West, 1991) developed syntax codes for Steps 2 and 3 of the J-N procedure using SAS. These commands are presented here, using hypothetical variables and data from the previous example. (The reader is referred to the Aiken and West text, Appendix C: SAS Program for Test of Critical Regions, for a complete description of the procedure).

Variables are entered in the order below separated by a space (free format). The values of each of the variables used in the hypothetical exercise intervention study appear in lines 0023 and 0024 of the program.

```

DEPVAR (short name of dependent variable)
ALLN = N (total N combining two groups)
N

1

= n

1

(number of subjects in group 1)
N

2

= n

2

(number of subjects in group 2)
SXSQR1 = SSX (1) (sum of squares predicted in Group 1)
SXSQR2 = SSX (2) (sum of squares predicted in Group 2)
MEAN1 = X (1) (mean of predictor in Group 1)
MEAN2 = X (2) (mean of predictor in Group 2)
F = F

2, N - 4

(value of F from table)
SSRES = SS

res

(sum of squares residual - add-up values of SS

res

from
Groups 1 and 2
)
B1 = B

1(1)

(slope for Group 1)
B01 = B

```

```

0(1)

(intercept for Group 1)
B2 = B

```

```

1(2)

(slope for Group 2)
B02 = B

```

```

0(2)

(intercept for Group 2)

```

The program prints the name of the dependent variable, the limit of Region 1 (XL1) and the limit of Region 2 (XL2).

```

SAS Program Commands:
00001 (local system Job Control Language [JCL])
00002 (local system JCL)
00003 (local system JCL)
00004 DATA HYPERT:
00005 INPUT DEPVBL POSTSYSBP LLN N1 N2 SXSQR1 SXSQR2
MEANX1 00006 MEANX2 F;
00006 SSRES B1 B01 B2 B02;
00007 MXSQR1 = MEANX1**2;
00008 MXSQR2 = MEANX2**2;
00009 SUM1 = (1/SXSQR1) + (1/SXSQR2);
00010 SUM2 = (MEANX1/SXSQR1) + (MEANX2/SXSQR2);
00011 SUM3 = (ALLN/(N1*N2)) + (MXSQR1/SXSQR1) + (MXSQR2/
SXSQR2
);
00012 SUMB1 = B1-B2;
00013 SUMB0 = B01-B02;
00014 SUMB1SQ = SUMB1**2;
00015 SUMB0SQ = SUMB0**2;
00016 A = (((-2*F)/ALLN-4)) * SSRES * SUM1) + SUMB1SQ;
00017 B = (((2*F)/ALLN-4)) * SSRES * SUM2) + (SUMB0 * SUMB1);
00018 C = (((2*F)/ALLN-4)) * SSRES * SUM3) + SUMB0SQ;
00019 SQRTB2AC = ((B**2) - (A*C)**.5;
00020 XL1 = (-B-SQRTB2AC)/A;
00021 XL2 = (-B+SQRTB2AC)/A;
00022 CARDS
00023 POSTSYSBP 30 15 15 2312.5 3001.4 154.40 150.53 3.37
00024 779.4 51.67 .37 .63 1.02
00025 PROC PRINT; VAR DEPVBL XL1 XL2;
00026 RUN;
00027 (local system JCL)

```

Calculation of the J-N region of insignificance is subject to a number of factors, including the variability of the CV and DV measurements within and between groups, and the number of

participants per group. The sensitivity of the cut points for the region of insignificance can be examined by calculating the standard error of the critical point. This can be carried out using a resampling technique called bootstrapping, a process where statistics are generated over a large number of replications, with samples drawn with replacement from a data set (Tabachnick & Fidell, 2001). Bootstrapping is most commonly performed using statistical software packages. Using the data from the previous example, 2000 bootstrap samples (each with the same sample size as the observed data) were generated with replacement from the observed data. One thousand samples were examined to determine the critical point of the distribution. The calculated critical point was determined to be 145 mm Hg, with a 95% confidence interval of 138 to 154 mm Hg.

The J-N technique does not always yield one solution within the effective range of the CV. There may be 0, 1, or 2 regions within the possible range of the CV in which the predicted values of the two regression lines differ, depending on the nature of the interaction. For example, in Figure 2, there are two regions where the estimated means for Group 1 and Group 2 are significantly different. In Figure 3, there is only one region ($XL\ 2 = 144.91$) where the group means are significantly different. The reader is referred to Huitema (1980) for a comprehensive discussion of this problem, as well as the use of the J-N technique with several groups, several covariates, two-factor designs, multiple independent variables and correlated samples.

As noted, the J-N technique is most useful in situations where there is a suspected treatment-covariate interaction. Examples of potential uses for the J-N technique in nursing research would be in the evaluation of treatments or medications that may have different effects depending upon an individual's physiological characteristics, psychological attributes or functional status.

SUMMARY

The Johnson-Neyman technique is the strongest alternative to ANCOVA in experimental designs when the assumption of homogeneity of regression slopes has been violated. The J-N technique provides the researcher with additional information regarding the region of insignificance with different treatment effects. Although commercially available computer software does not routinely perform the procedure, the J-N technique can be easily performed using customized SAS commands.

References

- Aikens, LS.; West, SG. Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage; 1991.
- Dorsey SG, Soeken KL. Use of the Johnson-Neyman procedure as an alternative to analysis of covariance. *Nursing Research*. 1996; 45(6):363–366. [PubMed: 8941311]
- Huitema, BE. The analysis of covariance and alternatives. New York: John Wiley; 1980.
- Johnson PO, Neyman J. Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs*. 1936; 1:57–93.
- Keppel, G. Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice Hall; 1991.
- Pedhazur, EJ. Multiple regression in behavioral research. 3. Fort Worth, TX: Holt, Rinehart & Winston; 1997.
- Pothoff RF. On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*. 1964; 29:241–256.
- Rogosa, DR. Unpublished doctoral dissertation. Stanford University; Stanford, CA: 1977. Some results for the Johnson-Neyman procedure.
- Tabachnick, BG.; Fidell, LS. Using multivariate statistics. 4. Boston: Allyn & Bacon; 2001.

Figure 1.
Scatterplot Illustrating Homogeneity of Regression Slopes

Figure 2. Scatterplot Illustrating Heterogeneity of Regression Slopes: Treatment-Covariate Interaction (2 Regions)

Figure 3. Scatterplot Illustrating Heterogeneity of Regression Slopes: Treatment-Covariate Interaction (1 Region) in an Exercise Intervention to Lower Systolic Blood Pressure