Paper SP04

Simultaneous Nonparametric Inference in a One-Way Layout Using the SAS® System

Paul Juneau, Pfizer Global Research & Development, Ann Arbor, MI

ABSTRACT

The process of discovering a novel medicine is one fraught with many unknowns. Even if a fundamental understanding of the biological systems involved in the disease process exists, the effect of a novel agent or agents on an organism may be difficult to predict or characterize. Moreover, knowledge about the corresponding measurement properties (e.g., distribution) is generally very limited. For the statistician, this setting is a perfect one to apply nonparametric statistical methods for data analysis.

A one-way layout with three or more treatments is a common study design employed in basic drug discovery research for the analysis of continuous responses. The goal of such an experimental design is often to make inferences amongst several treatments (e.g., all pair-wise comparisons, all comparisons to a control, etc.). In an attempt to address the needs of drug discovery researchers, the author has developed a set of SAS macros to perform simultaneous nonparametric inference in the one-way layout. During his presentation, he will summarize the flow and processing of his macros and their application to a few examples from drug discovery projects.

INTRODUCTION

The process of discovering a novel medicine is one fraught with many unknowns. Even if a fundamental understanding of the biological systems involved in the disease process exists, the effect of a novel agent or agents on an organism may be difficult to predict or characterize. Moreover, knowledge about the properties (e.g., distribution) of measurements selected to describe the agent's response is generally very limited. Compounding these difficulties is the fact that the investigations are generally carried out with relatively small sample sizes due to limited availability of the test substance at this phase of research. For the statistician, this setting is a perfect one to apply nonparametric methods for data analysis.

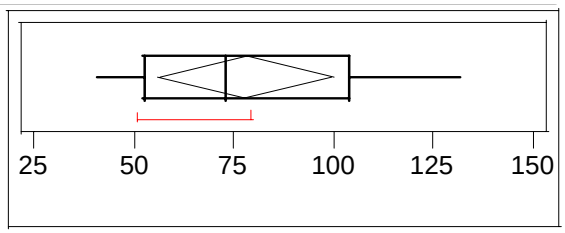
A second common feature of many drug discovery investigations is the choice of experimental design. A one-way layout with three or more treatments is a study design frequently employed in basic drug discovery research for the analysis of continuous responses (Juneau, 2003). The goal of such an experimental design is often to make inferences amongst several treatments (e.g., all pair-wise comparisons, all comparisons to a control, etc.).

Consider the following example. Samples of PC12 cells were randomized to one of three groups. The PC12 cells were cultured in one of three media: the first medium was infected with a particular strain of bacteria postulated to be associated with the development of Parkinson's disease. The second group used a medium cultured with a second strain of bacteria. The third group of cells was cultured in a normal uninfected medium. All cells were incubated for 24 hours and harvested to determine the dopamine concentration. Due to some unanticipated circumstances, some samples were lost during processing and the resultant sample sizes were unequal. The results of the experiment are displayed below via horizontal box plots with summary statistics*.

*The raw data for all examples presented in this work are found in the *Appendix*.

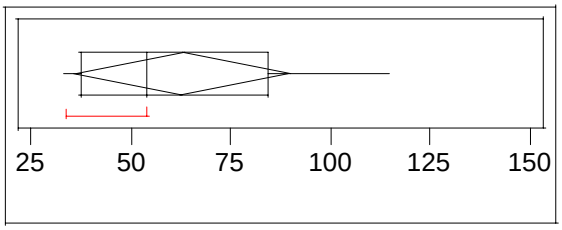
Figure 1: Dopamine Response Box Plot Summary

Strain II



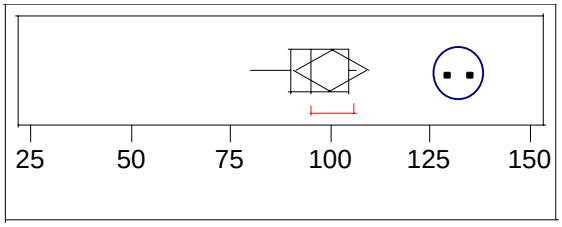
Mean = 78.04 Median = 72.92
Standard Deviation = 29.81 N = 10

Strain I



Mean = 63.12 Median = 54.29
Standard Deviation = 28.66 N = 7

Control



Mean = 100.00 Median = 95.24
Standard Deviation = 15.14 N = 14

Note the skewness of the Strain I group and its imbalance with respect to the sample sizes of the other two groups. Moreover, note the two extreme values measured in the controls (circled in blue). Suppose that the investigator is interested in performing all pair-wise comparisons of the location parameters of the three groups.

A simple linear model may be employed to relate the dopamine concentration to the type of cell media:

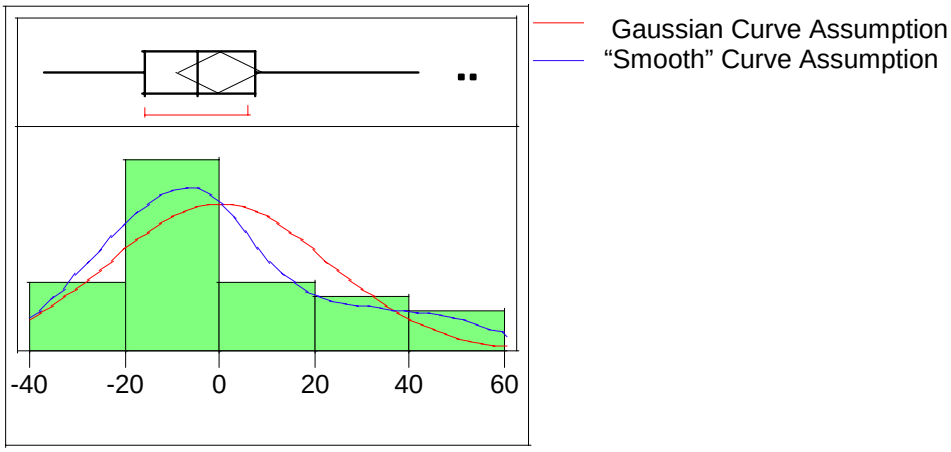
$$X_{ij} = \mu + \tau_i + \varepsilon,$$

Where X_{ij} = the dopamine concentration response of the *jth* experiment unit ($1 \le j \le n_j$) to the *ith* treatment ($1 \le i \le 3$), μ is the overall mean effect, τ_i is the effect of the *ith* media and ε is the error associated with the measurement. Typically, ε is assumed to be an independent and identically distributed Gaussian (normal) random variable. This model may be re-parameterized by substituting μ_i for $\mu + \tau_i$:

$$X_{ij} = \mu_i + \varepsilon.$$

One approach to this problem would be to apply the Tukey-Kramer multiple comparison procedure (Kramer, 1956) in the context of the analysis of variance to these data to simultaneously compare the three mean responses of the groups and preserve the family-wise error rate for the three statistical inferences. Examination of the summary statistics indicates that the measurements are not identically distributed (note the unequal standard deviations); however, the difference in standard deviations is less than a factor of 2, and thus, may not cause significant consequences on the desired family of statistical inferences (Hochberg and Tamhane, 1987). Another chief assumption is that the errors in the measurement follow a Gaussian (normal) distribution. It has been shown that this assumption may be relaxed to the point that only symmetry of the distribution is required (Scheffé, 1959).

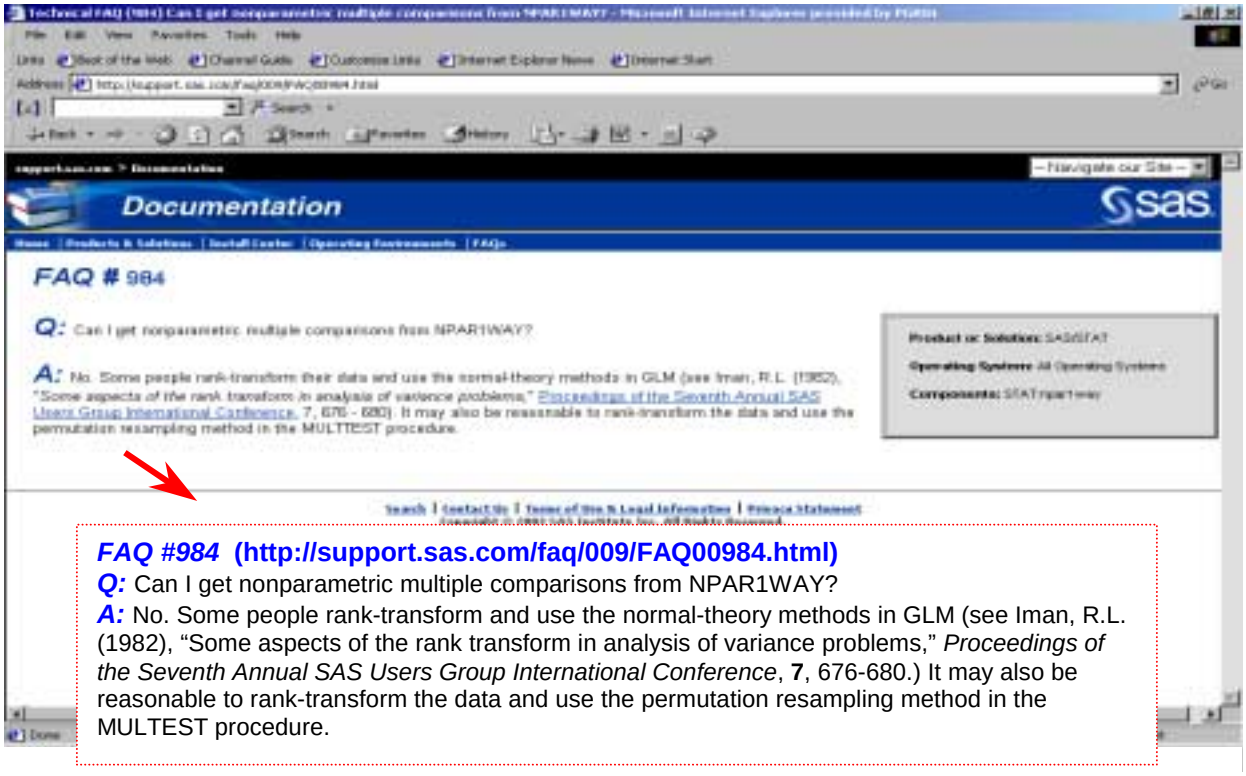
Suppose that a simple one-way analysis of variance (ANOVA) model is fit to these data. Examine the residuals (in composite) from the model fit:



The two plots of the residuals above (box plot and histogram) suggest that the distribution is probably not Gaussian and may not even be symmetrical about its mean. Thus, the assumption of symmetry necessary for the Tukey-Kramer procedure is not met for our example setting.

The problem of a lack of symmetry in this setting may be addressed by the application of nonparametric statistical methods. Nonparametric methods do not assume that the distribution of the errors in a one-way layout follows a distribution that is symmetric about its mean.

Currently, the SAS Institute website makes the following suggestion for nonparametric simultaneous inference:



FAQ #984 **(http://support.sas.com/faq/009/FAQ00984.html)**
Q: Can I get nonparametric multiple comparisons from NPAR1WAY?
A: No. Some people rank-transform and use the normal-theory methods in GLM (see Iman, R.L. (1982), "Some aspects of the rank transform in analysis of variance problems," *Proceedings of the Seventh Annual SAS Users Group International Conference*, **7**, 676-680.) It may also be reasonable to rank-transform the data and use the permutation resampling method in the MULTEST procedure.

The focus of this work is to supplement the SAS Institute's suggested approach with some SAS macro solutions designed to perform approximate (large sample) simultaneous nonparametric inference. The focus of this manuscript will be all pair-wise comparisons and all pair-wise comparisons with a designated control group.

DUNN'S PROCEDURE: ONE FORM OF SIMULTANEOUS NONPARAMETRIC INFERENCE IN THE ONE-WAY LAYOUT FOR ALL PAIR-WISE COMPARISONS

If one is interested in comparing the location parameters of the three previously described experimental groups (μ_1, μ_2, and μ_3, respectively) simultaneously and preserving the family-wise error rate, he/she could use an approach suggested by Dunn (Dunn, 1964) for the linear model stated above:

Conclude $\mu_i \neq \mu_1$ if $\left|\bar{R}_i - \bar{R}_l\right| > z_{\frac{\alpha}{(K(K-1))}} \sqrt{\frac{N(N+1)}{12}} \sqrt{\frac{1}{n_i} + \frac{1}{n_l}},$

Where $\bar{R}_i$ is the mean of the joint ranks* for group *i* , $\bar{R}_l$ is the mean of the joint ranks for group *l* ($1 \leq i < l \leq k$), n_i and n_l are sample sizes for groups *i* and *l*, respectively, *N* = the total sample size, K = the total number of comparisons desired (in our case, K=3) and $z_{\frac{\alpha}{(K(K-1))}}$ is the $\alpha/(K(K-1))$ *th* quantile from a standard Gaussian distribution.

Dunn's Procedure offers the following advantages: (a) the symmetry assumption, which is often difficult to assess in drug discovery settings with small sample sizes may be relaxed or ignored; (b) equal sample sizes are not required; (c) relatively small total sample sizes may be analyzed with this technique (3 groups with 5 experimental units/group, or >3 groups with 4 units/group - see Lehman, 1975). A simple set of macros may be constructed in SAS to perform Dunn's Procedure for all pair-wise comparisons. A macro to perform this analysis is the %DUNN macro described below.

A SUMMARY OF THE *%DUNN* MACRO FLOW**

The %DUNN macro consists of a body of code containing one embedded macro (%GROUPS). The embedded macro determines the number of groups present and assigns that value to a macro variable (&NGRPS). If a group in the SAS data set does not contain at least one response value (i.e., all values "missing"), it will not be included in the analysis. The embedded macro also creates one global macro variable that contains the group labels (&GRPVEC) for the levels of the class variable. The main body of the SAS code determines summary statistics (e.g., average ranks, sample sizes, etc.). This information is employed to calculate the pair-wise test statistics. The corresponding cutoff for the test statistic is calculated with the PROBIT function. Results are then printed out using **PROC PRINT**.

*Recall that the "joint ranking" is determined by ranking all of the *N* observations together from smallest to largest.

**Due to space limitations imposed by the conference *Proceedings*, all SAS code written to conduct the analyses described in this paper was excluded. The reader is encouraged to contact the author for a printed or electronic copy of the macros. Contact information is supplied towards the end of this paper.

ANALYSIS OF THE PC12 DATA SET FROM THE INTRODUCTION

The PC12 data presented in the introduction may be analyzed using the %DUNN macro designed to perform the desired simultaneous pair-wise nonparametric comparisons of all treatments. The results of the macro's execution are illustrated below:

Large Sample Approximation Multiple Comparison Procedure
Designed for Unbalanced Data
3 Groups: Control StrainI StrainII (Respective Sample Sizes: 14 7 10)
Alpha = 0.05

Method Suggested by Dunn (1964)

Group Comparisons	Comparison Number	Difference in Average Ranks	Cutoff at Alpha=0.05	Significant Difference = **
Control - StrainI	1	11.8571	10.0759	**
Control - StrainII	2	7.6429	9.0121	
StrainI - StrainII	3	4.2143	10.7266	

The %DUNN macro produces title statements that state the number of class levels, a list of each of the class levels with non-missing values, and the corresponding group sample sizes. Moreover, the macro generates output that contains the relevant test statistic for each comparison (a function of the average ranks/class level), the corresponding cutoff for the chosen level of family-wise error, and a symbol indicating whether the results of the statistical inference are statistically significant.

From this analysis, one would conclude that a statistically significant difference existed between the median dopamine levels in the controls and samples treated with the first strain of bacteria.

DWASS, STEELE, CRITCHLOW-FLIGNER PROCEDURE: A SECOND FORM OF SIMULTANEOUS NONPARAMETRIC INFERENCE IN THE ONE-WAY LAYOUT FOR ALL PAIR-WISE COMPARISONS (HOLLANDER & WOLFE, 1999)

If one is interested in comparing the location parameters of the *k* experimental groups ($\mu_1, \mu_2, \mu_3, \ldots, \mu_i, \ldots, \mu_k$, respectively) simultaneously and preserving the family-wise error rate, he/she could use an approach suggested by Dwass (Dwass, 1960) and Steele (Steele, 1960) for the linear model stated above. The procedure begins by calculating the k(k-1)/2 pairs of Wilcoxon Rank Sum statistics, $W_{i,l}$ (Wilcoxon,1945) for each pair, *i* and *l* ($1 \le i < l \le k$). The Wilcoxon statistics *should* include an adjustment for tied values.

Conclude $\mu_i \neq \mu_1$ if $\left| DSCF_{i,l} \right| > q_\alpha$,

Where: $DSCF_{i,l} = \sqrt{2}\left(\dfrac{W_{i,l} - \left(\dfrac{(\min(n_i, n_l)(n_i + n_l + 1)}{2} \right)}{\sqrt{\text{var}(W_{i,l})}} \right)$, $\text{var}(W_{i,l})$ is the tie-adjusted variance for the Wilcoxon statistic, and q_α is the α *th* quantile from the *Studentized Range Distribution*.

Suppose that for all $n_i + n_l$ observations in the comparison, ξ tied groups of size t_τ $(1,2,\ldots,\tau,\ldots,\xi)$ exist. The variance in the denominator of the statistic is as follows:

$$\frac{n_i n_l}{24}\left(n_i + n_l + 1 - \frac{\sum_{\tau=1}^{\xi}(t_\tau - 1)t_\tau(t_\tau + 1)}{(n_i + n_l)(n_i + n_l - 1)} \right)$$ (Hollander and Wolfe, 1999).

An analysis using the Dwass, Steele, Critchlow-Fligner method may be conducted using a SAS macro called %DSCF.

A SUMMARY OF THE *%DSCF* MACRO FLOW

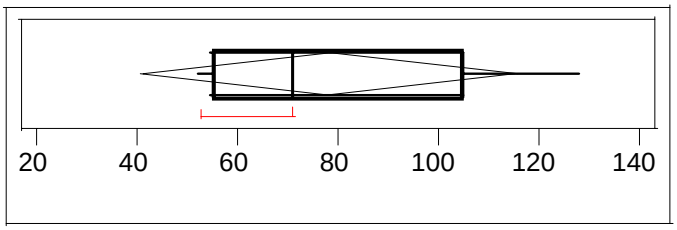
The %DSCF macro consists of a body of code containing one embedded macro (%GROUPS). The embedded macro determines the number of groups present (&NGRPS) as in the %DUNN macro. If a group in the input data set does not contain at least one response value, it will be excluded from the analysis. The embedded macro also creates two global macro variables that contain the group labels (&GRPVEC) for the levels of the class variable and information about the sample size for each group (&NVEC). The main body of the code calculates the necessary summary statistics (e.g., Wilcoxon Rank Sum test statistics) and the number of ties present in each pair-wise comparison. This information is then utilized to calculate the pair-wise test statistics. The cutoff for the test statistic is calculated with the PROBMC function (using the *Studentized Range* argument). As the macro iterates between all pair-wise comparisons it concatenates successive results in a data set called STAT. The final results are then printed out with **PROC PRINT**.

ANALYSIS OF THE TRIGLYCERIDE DATA SET

Consider the following experiment. Subjects were randomized to one of four treatment groups: three active agents and one untreated vehicle control group. The goal of the experiment was to determine whether the agents could affect triglyceride level (in mg/dl) relative to the vehicle controls and to determine whether evidence existed to declare one agent different from another with respect to triglyceride response. Each subject was treated and after a fixed post-treatment period the blood triglyceride level was measured for each subject. The results of the experiment are displayed below via box plots with summary statistics. Note the balanced sample sizes. The Dwass, Steele, Critchlow-Fligner multiple comparison procedure works optimally under settings with equal sample sizes/group.

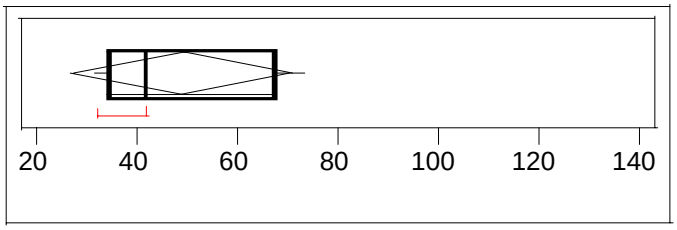
Figure 2: Triglyceride Response Box Plot Summary

Treatment A



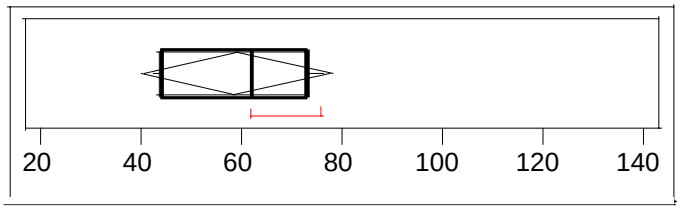
Mean = 78.2 Median = 71
Standard Deviation = 29.75 N = 5

Treatment B



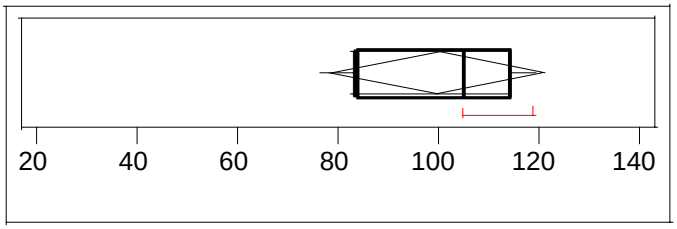
Mean = 49.0 Median = 42
Standard Deviation = 17.33 N = 5

Treatment C



Mean = 59.0 Median = 62
Standard Deviation = 15.00 N = 5

Vehicle



Mean = 99.8 Median = 105
Standard Deviation = 16.71 N = 5

The triglyceride data presented previously may be analyzed using the %DSCF macro designed to perform the desired simultaneous pair-wise nonparametric comparisons of all treatments. The results of the macro's execution are illustrated below:

Large Sample Approximation Multiple Comparison Procedure
For All Treatment Pairs
Based Upon Pairwise Rankings
4 Groups: TreatmentA TreatmentB TreatmentC Vehicle (Respective Sample Sizes: 5 5 5 5)
Alpha = 0.05

Method Suggested by Dwass, Steel, Critchlow & Fligner

Group Comparisons	Comparison Number	Test Statistic	Test Statistic Absolute Value	Cutoff at Alpha=0.05	Significant Difference = **
TreatmentA – TreatmentB	1	-2.21565	2.21565	3.63316	
TreatmentA – TreatmentC	2	-1.62481	1.62481	3.63316	
TreatmentA – Vehicle	3	1.92023	1.92023	3.63316	
TreatmentB – TreatmentC	4	1.92023	1.92023	3.63316	
TreatmentB – Vehicle	5	3.69274	3.69274	3.63316	**
TreatmentC – Vehicle	6	3.69274	3.69274	3.63316	**

The %DSCF macro produces title statements that state the number of class levels, a list of each of the class levels with non-missing values, and the corresponding group sample sizes. Moreover, the macro generates output that contains the relevant test statistic for each comparison (a function of the Wilcoxon statistic), the corresponding cutoff for the chosen level of family-wise error, and a symbol indicating whether the results of the statistical inference are statistically significant.

From this analysis, one would conclude that a statistically significant difference existed between the median triglyceride levels in the controls and samples treated with Treatments B and C. Statistically significant differences did not exist between the three active agents (Treatments A, B and C).

A COMPARISON OF THE TWO MULTIPLE COMPARISON PROCEDURES FOR ALL PAIR-WISE COMPARISONS

The triglyceride data used previously as an example for the application of the Dwass, Steele, Critchlow-Fligner method was then analyzed with Dunn's procedure for all pair-wise comparisons. The results of this analysis are shown below:

Large Sample Approximation Multiple Comparison Procedure
Designed for Unbalanced Data
4 Groups: TreatmentA TreatmentB TreatmentC Vehicle (Respective Sample Sizes: 5 5 5 5)
Alpha = 0.05

Method Suggested by Dunn (1964)

Comparison Number	Group Comparisons	Difference in Average Ranks	Cutoff at Alpha=0.05	Significance Difference = **
1	TreatmentA – TreatmentB	6.6	9.87145	
2	TreatmentA – TreatmentC	3.6	9.87145	
3	TreatmentA – Vehicle	5.0	9.87145	
4	TreatmentB – TreatmentC	3.0	9.87145	
5	TreatmentB – Vehicle	11.6	9.87145	**
6	TreatmentC – Vehicle	8.6	9.87145	

Recall that in the previous analysis using the Dwass, Steele, Critchlow-Fligner method, Treatment C was also declared to be statistically significantly different from the vehicle control group for the median triglyceride response. A comparison of the properties of these two methods may shed some light on the reason for the differing inferential conclusions. First, the Dunn method uses a Bonferroni-like correction to the family-wise error rate (Miller, 1981) and might be a bit too conservative. Second, the Dunn procedure employs *joint ranking*, and thus, the comparison of two groups is highly influenced by the behavior of other groups in the experiment as the data are initially ranked over the entire experiment (Hollander and Wolfe, 1999). The balanced sample sizes for all groups also suggest that the Dwass, Steele, Critchlow-Fligner method might be the most appropriate technique to employ for all pair-wise nonparametric comparisons.

SIMULTANEOUS NONPARAMETRIC INFERENCE IN THE ONE-WAY LAYOUT *FOR ALL GROUP COMPARISONS WITH A DESIGNATED CONTROL GROUP*

If one is interested in comparing the location parameters of several experimental groups ($\mu_1, \mu_2, \mu_3, \ldots, \mu_i, \ldots, \mu_{k-1,}$ respectively) to a designated control group (μ_c) simultaneously and preserving the family-wise error rate, he/she could use one of two approaches originally suggested by Dunn (Dunn, 1964) or Miller (Miller, 1966) for the linear model stated above.

DUNN'S PROCEDURE FOR ALL GROUP COMPARISONS TO A DESIGNATED CONTROL GROUP (FOR UNEQUAL SAMPLE SIZES)

Conclude $\mu_i \neq \mu_c$ if $\left|\bar{R}_i - \bar{R}_c\right| > z_{\frac{\alpha}{(2(K-1))}} \sqrt{\frac{N(N+1)}{12}} \sqrt{\frac{1}{n_c} + \frac{1}{n_i}},$

Where $\bar{R}_i$ is the mean of the joint ranks for group *i*, $\bar{R}_c$ is the mean of the joint ranks for the control group *c*, and n_i and n_c are sample sizes for group *i* and the control group, *c*, respectively, *N* = the total sample size, K = the total number of comparisons desired and $z_{\frac{\alpha}{(2(K-1))}}$ is the $\alpha/(2(K-1))$ *th* quantile from a standard Gaussian distribution.

MILLER'S PROCEDURE FOR ALL GROUP COMPARISONS TO A DESIGNATED CONTROL GROUP (FOR EQUAL SAMPLE SIZES)

Conclude $\mu_i \neq \mu_c$ if $\left|\bar{R}_i - \bar{R}_c\right| > \left|M\right|_{\alpha,k-1,\infty} \sqrt{\frac{N(N+1)}{12}} \sqrt{\frac{1}{n_c} + \frac{1}{n_i}},$

Where $\bar{R}_i$ is the mean of the joint ranks for group *i*, $\bar{R}_c$ is the mean of the joint ranks for the control group *c*, and n_i and n_c are sample sizes for group *i* and the control group, *c*, respectively, *N* = the total sample size, K = the total number of comparisons desired and $\left|M\right|_{\alpha,k-1,\infty}$ is α *th* quantile from the *Studentized Maximum Modulus* distribution.

All nonparametric pair-wise comparisons to a designated control group in a one-way layout may be performed with the %NPARMCC macro.

A SUMMARY OF THE *%NPARMCC* MACRO FLOW

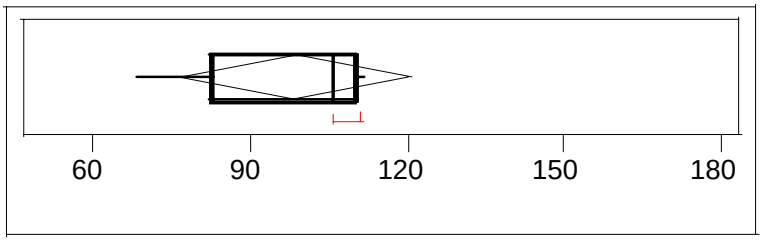
The macro consists of a body of code containing one embedded macro (%GROUPS). The embedded macro determines the number of groups present (&NGRPS) as in the two previous macros. If a group in the SAS data set does not contain at least one response value, it will be excluded from the analysis. The embedded macro also creates one global macro variable that contains the group labels (&GRPVEC) for the levels of the class variable. The main body of the macro determines the necessary summary statistics (e.g., average ranks, sample sizes, etc.). This information is then employed to calculate the pair-wise test statistics. The cutoff for the test statistic is calculated with the PROBIT function for Dunn's procedure and the PROBMC function (with the *Studentized Maximum Modulus* argument) for Miller's procedure. The results are then printed out with a **PROC PRINT** procedure.

ANALYSIS OF THE ACTIVATED CLOTTING TIME DATA SET

An experiment was designed to study the effect of increasing the dose of a novel agent on activated clotting time (ACT). Subjects were randomized to one of four groups: a vehicle control group, a low dose group, a medium dose group, and a high dose group. About 200 minutes after receiving treatment, the ACT for each subject was measured (in seconds). The results of the experiment are illustrated below (as before, with box plots and summary statistics):

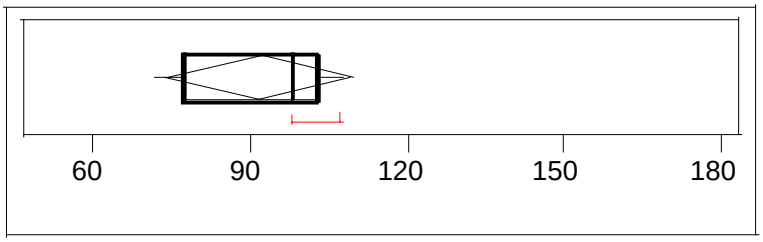
Figure 3: Activated Clotting Time Response Box Plot Summary

Low Dose



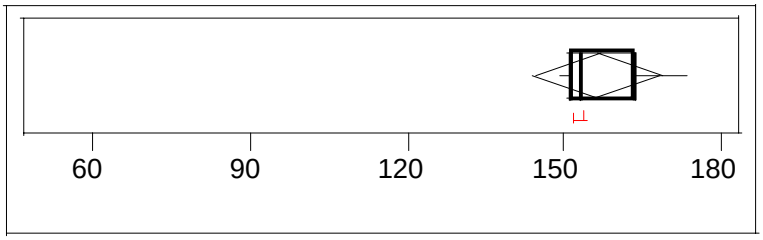
Mean = 98.4 Median = 106
Standard Deviation = 17.47
N = 5

Medium Dose



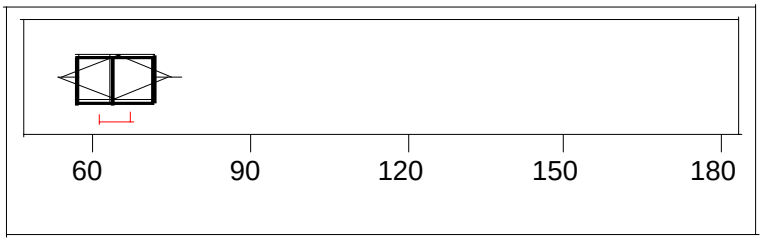
Mean = 91.8 Median = 98
Standard Deviation = 14.06
N = 5

High Dose



Mean = 156.4 Median = 153
Standard Deviation = 9.40
N = 5

Vehicle Controls



Mean = 64.2 Median = 63
Standard Deviation = 8.11
N = 5

The ACT data presented in the introduction may be analyzed using the %NPARMCC macro designed to perform the desired simultaneous pair-wise nonparametric comparisons of all groups against a designated control group. The results of the macro's execution are illustrated below:

Large Sample Approximation Multiple Comparison Procedure
For All Treatments Compared Against a Control (Control Group = Control)
4 Groups: Control DrugHigh DrugLow DrugMed (Respective Sample Sizes: 5 5 5 5)
Alpha = 0.05

Method Suggested by Dunn (1964) for UNBALANCED Sample Sizes
Method Suggested by Miller (1966) for Balanced Sample Sizes

Group Comparisons	Comparison Number	Difference in Average Ranks	Dunn Cutoff at Alpha=0.05	Miller Cutoff at Alpha=0.05	Significant Difference Dunn's = *D*	Significant Difference Miller's = *M*
DrugHigh – Control	1	14.6	7.96242	8.93410	*D*	*M*
DrugLow – Control	2	7.6	7.96242	8.93410		
DrugMed – Control	3	6.2	7.96242	8.93410		

The %NPARMCC macro produces title statements that state the number of class levels, a list of each of the class levels with non-missing values, and the corresponding group sample sizes. Moreover, the macro generates output that contains the relevant test statistic for each comparison (the difference in the mean ranks), the corresponding cutoff for the chosen level of family-wise error, and a symbol indicating whether the results of the statistical inference are statistically significant by Dunn's procedure (*D*) and/or Miller's procedure (*M*).

From this analysis, one would conclude that a statistically significant difference existed between the median activated clotting time in the controls and samples treated with high dose of the new agent. As the experiment consisted of balanced sample sizes, the conclusion inferred by the Miller approach would be considered the appropriate one to report.

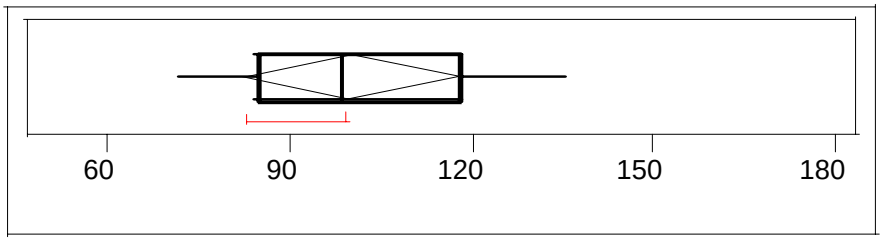
A COMPARISON OF THE TWO MULTIPLE COMPARISON PROCEDURES FOR ALL PAIR-WISE COMPARISONS VS A DESIGNATED CONTROL GROUP

It is interesting to compare the behavior of Dunn and Miller's respective methods using a data set that is not balanced with respect to sample size. Suppose that we conducted a similar ACT experiment as described previously a second time, yet this time with unequal sample sizes.

Consider the following data:

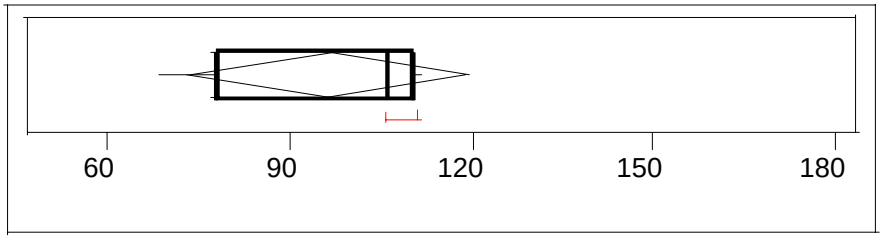
Figure 4: Activated Clotting Time (Unbalanced Sample Sizes) Response Box Plot Summary

Low Dose



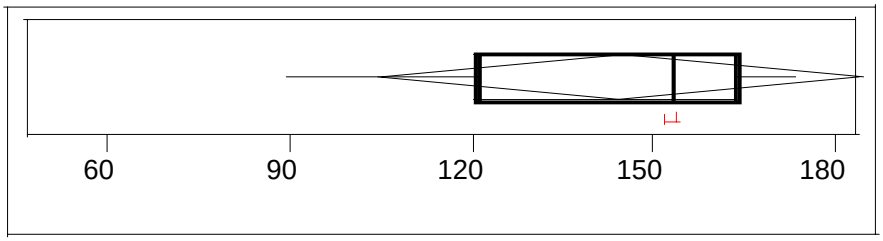
Mean = 100 Median = 98.5
Standard Deviation = 20.87 N = 8

Medium Dose



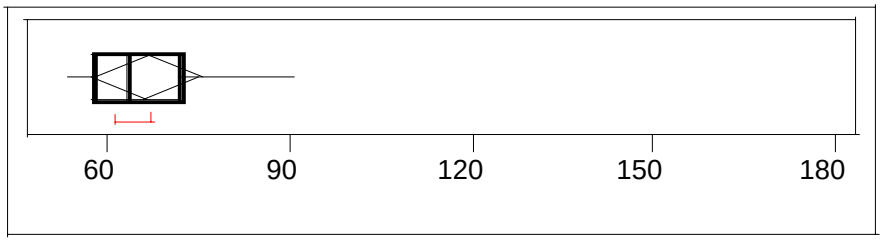
Mean = 96.4 Median = 104
Standard Deviation = 18.37 N = 5

High Dose



Mean = 144.4 Median = 153
Standard Deviation = 31.63 N = 5

Control



Mean = 62.2 Median = 63
Standard Deviation = 11.10 N = 9

The data were once again analyzed using the %NPARMCC macro. For this data set, the control group and the low dose group have different sample sizes than in the first analysis of triglyceride data presented previously (note sample sizes circled in red).

The results of the analysis are illustrated below:

Large Sample Approximation Multiple Comparison Procedure
For All Treatments Compared Against a Control (Control Group = Control)
4 Groups: Control DrugHigh DrugLow DrugMed (Respective Sample Sizes: 9 5 8 5)
Alpha = 0.05

Method Suggested by Dunn (1964) for UNBALANCED Sample Sizes
Method Suggested by Miller (1966) for Balanced Sample Sizes

Group Comparisons	Comparison Number	Difference in Average Ranks	Dunn Cutoff at Alpha=0.05	Miller Cutoff at Alpha=0.05	Significant Difference Dunn's = *D*	Significant Difference Miller's = *M*
DrugHigh – Control	1	17.4667	9.42126	10.5710	*D*	*M*
DrugLow – Control	2	10.2917	8.20747	9.2091	*D*	*M*
DrugMed – Control	3	10.1667	9.42126	10.5710	*D*	

Note that Miller's procedure does not declare the median response for the medium dose group (104) to be statistically significantly different than the control group (63) while Dunn's procedure *does* declare this result to be statistically significant. The difference in the inferential conclusions can be attributed to the fact that Dunn's procedure was derived to handle the unequal sample size setting, whereas the Miller procedure, using the Studentized Maximum Modulus cutoff works optimally under conditions of equal sample size (Hochberg and Tamhane, 1987).

SUMMARIZATION OF THE NONPARAMETRIC MULTIPLE COMPARISON PROCEDURES PRESENTED IN THIS PAPER

The following table provides advice on the procedure to use given the study design of a particular investigation.

Desired Comparison	Study Design Feature	Nonparametric Multiple Comparison Technique	Designed SAS Macro to Perform Analysis
All Pair-wise	Equal Group Sample Sizes	*Dwass, Steele Critchlow-Fligner* Technique	%DSCF
	Unequal Group Sample Sizes	*Dunn's* Technique	%DUNN
Comparisons with a Designated Control Group	Equal Group Sample Sizes	*Miller's* Technique	%NPARMCC
	Unequal Group Sample Sizes	*Dunn's* Technique	

CONCLUSION

The nonparametric multiple comparison procedures (simultaneous inference procedures) presented offer the following advantages to a data analyst working in the arena of drug discovery: (1) the symmetry assumption, which is often difficult to assess in drug discovery settings with small sample sizes, may be relaxed or ignored; (2) equal sample sizes are not required as procedures exist for both the balanced and unbalanced sample size cases; (c) relatively small total sample sizes may be analyzed with these techniques. The macro code to perform these analyses is relatively simple in its structure and easy to use; moreover, the macros presented supplement the SAS Institute's currently suggested methods for nonparametric multiple comparisons. Coupled with PROC NPAR1WAY, the macros presented in this paper offer the statistician new opportunities for data analysis in settings where the properties of the response measures are not well characterized because of the early stage in the scientific investigation.

REFERENCES

Dunn, O.J. “Multiple comparisons using rank sums”. *Technometrics* **6** (1964) pp. 241-252.

Dwass, M. “Some k-sample rank-order statistics”. *Contributions to Probability and Statistics* (I. Olkin, S.G. Ghurye, H Hoeffding, W.G. Madow, and H.B. Mann, editors). Stanford: Stanford University Press, 1960, pp. 198-202.

Hochberg, Y. and Tamhane, A.C. *Multiple Comparison Procedures*. New York: John Wiley & Sons, 1987.

Hollander, M., and Wolfe, D.A. *Nonparametric Statistical Methods*, 2/e. New York: John Wiley & Sons, 1999, pp. 240-249.

Juneau, P. “Using SAS® to Perform a Single-Stage Multiple Comparison Procedure for All Pair-wise Comparisons in a One-Way Layout with Unequal Variances” *Proceedings of the PharmaSUG 2003 Annual Conference, Miami, Florida*. Paper 15, p. 485.

Kramer, C.Y. “Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications”. *Biometrics* **12** (1956), pp. 307-310.

Lehman, E.L. *Nonparametrics: Statistical Methods Based Upon Ranks*. New York: Holden-Day, 1975, pp. 206-207.

Miller, R.G., Jr. *Simultaneous Statistical Inference*, 1/e. New York: Springer-Verlag, 1966.

Miller, R.G., Jr. *Simultaneous Statistical Inference*, 2/e. New York: Springer-Verlag, 1981.

Scheffé, H. The Analysis of Variance. New York: John Wiley & Sons, 1959, pp. 334-353.

Steele, R.G.D. “A rank sum test for comparing all pairs of treatments”. *Technometrics* **2** (1960) pp. 197-207.

Wilcoxon, F. “Individual comparisons by ranking methods”. *Biometrics* **1** (1945), pp. 80-83.

ACKNOWLEDGMENTS

The author would like to thank Drs. Thomas J. Vidmar, Cyrus Hoseyni, Fang Dong and Arthur J. Roth, of the division of Biostatistics and Reporting at Pfizer Global Research & Development for their constructive comments that improved the quality of this manuscript. He would also like to thank Dr. Mohan Beltangady, Chief Statistician, Pfizer Global Research & Development, for his encouragement, Mr. Thomas McClanahan for data examples from cardiovascular research and Dr. Dianne Camp of the William Beaumont Research Institute for the original data example from her Parkinson’s research program and for her ubiquitous editorial assistance.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Paul Juneau
Associate Director
Midwest Nonclinical Statistics Department
Pfizer Global Research & Development – Michigan Laboratories
2800 Plymouth Road
Ann Arbor, MI 48105
Work Phone: 734-622-1791
Fax: 734-622-3153
Email: paul.juneau@pfizer.com

APPENDIX: DATA SET EXAMPLES DISCUSSED IN THIS WORK

THE PC12 CELL DATA SET

Group	Dopamine Concentration
Control	80.32
Control	103.39
Control	96.72
Control	85.05
Control	134.52
Control	105.99
Control	104.19
Control	89.82
Control	128.57
Control	90.48
Control	95.24
Control	95.24
Control	95.24
Control	95.24
StrainI	47.55
StrainI	33.56
StrainI	37.76
StrainI	114
StrainI	70.57
StrainI	54.29
StrainI	84.14
StrainII	51.05
StrainII	60.39
StrainII	68.82
StrainII	77.13
StrainII	53.25
StrainII	41.72
StrainII	119.3
StrainII	79.24
StrainII	131.16
StrainII	98.37

THE TRIGLYCERIDE DATA SET

Group	Triglyceride Level
Vehicle	89
Vehicle	77
Vehicle	105
Vehicle	119
Vehicle	109
TreatmentA	127
TreatmentA	71
TreatmentA	57
TreatmentA	83
TreatmentA	53
TreatmentB	37
TreatmentB	32
TreatmentB	61
TreatmentB	73
TreatmentB	42
TreatmentC	62
TreatmentC	76
TreatmentC	70
TreatmentC	43
TreatmentC	44

THE ACT DATA SET – BALANCED SIZES

Group	Activated Clotting Time
Control	63
Control	67
Control	54
Control	76
Control	61
DrugHigh	154
DrugHigh	173
DrugHigh	153
DrugHigh	150
DrugHigh	152
DrugLow	110
DrugLow	111
DrugLow	96
DrugLow	69
DrugLow	106
DrugMed	98
DrugMed	99
DrugMed	83
DrugMed	107
DrugMed	72

THE ACT DATA SET – *UNBALANCED SAMPLE SIZES*

Group	Activated Clotting Time
Control	63
Control	67
Control	54
Control	76
Control	61
Control	90
Control	63
Control	55
Control	67
DrugHigh	154
DrugHigh	173
DrugHigh	153
DrugHigh	90
DrugHigh	152
DrugMed	110
DrugMed	111
DrugMed	86
DrugMed	69
DrugMed	106
DrugLow	125
DrugLow	89
DrugLow	98
DrugLow	99
DrugLow	83
DrugLow	135
DrugLow	72
DrugLow	99