September 20, 2010

Documentation for ANOVA test

Minn M. Soe, MD, MPH, MCTM : msoe@sph.emory.edu
Kevin M. Sullivan, PhD, MPH, MHA: cdckms@sph.emory.edu

This analysis is conducted to observe whether there is a significant difference in means among two or more independent samples, given values for standard deviation or standard error.

The input values requested are, for each group, sample size, sample means, and either standard deviation or the standard error of the sample mean.

Calculate

Clear

Analysis of Variance (ANOVA)					
	N (counts)	Mean	Std. Dev.	(or)	Std. Error
Group 1	63	55.1	10.93		
Group 2	17	47.59	7.08		
Group 3	15	49.4	10.2		
Group 4					
Group 5					
Group 6					
Group 7					
Group 8					
Group 9					
Group 10					
Enter your summary data in any respective group, no need to be consecutive. (see the example above)					

The results of the calculation are shown below:

ANOVA Table

Source of variation	Sum of squares	d.f	Mean square	*F* statistics	p-
Between Groups	967.821	2	483.91	4.60609	0.0
Within Groups	9665.41	92	105.059		
Total	10633.2	94			

	Chi square	d.f	p-value[1]
Test for equality of variance	3.91791	2	0.141005

		95% CI of individual sample mean		95% CI assumin variance	
Group	Mean	Lower Limit	Upper Limit	Lower Limit	Upp
1	55.1	52.3473	57.8527	52.5186	57
2	47.59	43.9498	51.2302	42.3201	52
3	49.4	43.7515	55.0485	43.7239	55
4					
5					
6					
7					
-					

The interpretation is that there is an overall significant difference among the mean MAXFWT scores (maximum finger-wrist tapping score) in the three groups. It is noteworthy that before interpreting as above, the Chi square test for the equality of variances from these independent, normally distributed samples should be first checked. If the sample variances are significantly different, i.e. p-value from test for equality of variance is <0.05, the use of the ANOVA test may not be justified and the alternative non-parametric test should be done (e.g. Kruskal-Wallis test). In the example above, the sample variances are not significantly different, i.e. p-value from test of equality of variance is 0.138, and therefore, the ANOVA test can be used to compare the MAXFWT scores in the three sample groups, assuming the data are normally distributed.

The ANOVA module also displays the confidence intervals of individual sample means in two formats, based on the standard error of the observed samples and on Mean Square Within (Mean Square Error).

The formulae and definitions for one-way ANOVA test are from Bernard Rosner. Fundamentals of Biostatistics (5th edition).

Equation 12.5

Short Computational Form for the between SS and Within SS

$$Between\ \ SS = \sum_{i=1}^{k} n_i \overline{Y}_i^{\,2} - \frac{\left(\sum_{i=1}^{k} n_i \overline{Y}_i\right)^2}{n} = \sum_{i=1}^{k} n_i \overline{Y}_i^{\,2} - \frac{Y^2..}{n}$$

$$Within\ \ SS = \sum_{i=1}^{k} (n_i - 1) S_i^2$$

Y.. = sum of the observations across all groups-ie, the grand total of all observations over all groups; n= total number of observations over all groups.

Definition 12.5: Between Mean Square= Between MS= Between SS/(k-1).
(k= number of comparison groups)

Definition 12.6: Within Mean Square= Within MS= Within SS/(n-k).

Display of one-way ANOVA results

Source of variation	Sum of Square	*df*	Mean Square	F statistic	p-value
Between	$\sum_{i=1}^{k} n_i \overline{Y}_i^{\,2} - \frac{Y^2..}{n} = A$	$k-1$	$\frac{A}{k-1}$	$\frac{A/(k-1)}{B/(n-k)} = F$	$\Pr(F_{k-1,n-k} > F)$
Within	$\sum_{i=1}^{k} (n_i - 1) S_i^2 = B$	$n-k$	$\frac{B}{n-k}$		
Total	Between SS + Within SS				

<u>Statistical formula for equality of variance (Bartlett's test)</u>

$$B = {\{\sum[(n_i - 1) LnS^2] - \sum[(n_i - 1) LnS_i^2]\}}\big/ C$$

In the above, S_i^2 is the variance of the ith group, n_i is the sample size of the i^{th} group, k is the number of groups, and S^2 is the pooled variance. The pooled variance is a weighted average of the group variances and is defined as:

$$S^2 = \sum[(n_i - 1)S_i^2] \Big/ \sum[(n_i - 1)] \quad \text{over all } i = 1, 2, \ldots\ldots\ldots, k$$

$$C = 1 + \{\sum[1/(n_i - 1)] - 1/\sum[1/(n_i - 1)]\} \Big/ 3(k+1)$$

<u>Calculation of 95% confidence intervals are based on *t* distribution.</u>

- X ± $t_{(df, alpha=0.5)}$ * standard error of each sample

- X ± $t_{(df, alpha=0.5)}$ * (Mean square within/n_i $)^{1/2}$, assuming equal variance.

Reference:
Bernard Rosner. Fundamentals of Biostatistics (5th edition).

Acknowledgement:
Default values were obtained from table 12.5 (pg. 533) described in 'Fundamentals of Biostatistics' (5th edition) by Bernard Rosner.

February 27 2008 - fixed an error in the code where the *t*-value used in confidence intervals used *ni* rather than *ni*-1 for the degrees of freedom.