ONEWAY

For post hoc range tests and pairwise multiple comparisons, see Appendix 10.

Notation

The following notation is used throughout this chapter unless otherwise stated:

X_{lj}	Value of the *j*th observation in group *l*
w_{lj}	Weight for the *j*th observation in group *l*
$W_{l,j}$	Sum of weights of the first *j* cases in group *l*
W_l	Sum of weights of all cases in group *l*
k	Number of groups, determined as maximum group values minus minimum plus one
k'	Number of nonempty groups
n_l	Number of cases in group *l*
W	Sum of weights of cases in all groups

Group Statistics

Computation of Group Statistics

A weighted version of the Young-Cramer (1971) algorithm is used to compute recursively the corrected sum of squares for each group.

$$SSQ_{l,i} = SSQ_{l,i-1} + \frac{w_{li}\left(X_{li}W_{l,i-1} - \sum_{j=1}^{i-1} w_{lj}X_{lj}\right)^2}{W_{l,i-1}W_{li}}$$

The initial value is 0; the value for each group after the last observation has been processed is the corrected sum of squares.

$$SS_l = SSQ_{l,n_l}$$

The sum and mean for each group are

$$T_l = \sum_{i=1}^{n_l} X_{li} w_{li}$$

$$\overline{T}_l = T_l / W_l$$

The variance is

$$S_l^2 = SS_l \big/ \left(W_l - 1\right)$$

The grand sum is

$$G = \sum_{i=1}^{k} T_i$$

Group Statistics from Summary Statistics

With matrix data input, the user supplies sum of weights in each group $\left(W_l\right)$, means $\left(\overline{T}_l\right)$, and standard deviations $\left(S_l\right)$. From these,

$$T_l = W_l \overline{T}_l$$

$$SS_l = (W_l - 1)S_l^2$$

$$G = \sum_{i=1}^{k} T_i$$

If the user supplies the pooled variance S_p^2 and its degrees of freedom (D) instead of the individual S_l, and $D < 1$, the program will reset it to

$$D = \sum_{l=1}^{k} W_l - k'$$

The within-group sum of squares is

$$WSS = S_p^2 D$$

The ANOVA Table

Source of Variation	SS	df
Between (BSS)	$\sum_{l=1}^{k} T_l^2 / n_l - G^2 / W$	$k' - 1$
Within (WSS)	$\sum_{l=1}^{k} SS_l$ $\left(s_p^2 D \text{ for matrix input}\right)$	$W - k'$ (D)
Total (TSS)	$BSS + WSS$	$W - 1$

Mean squares are calculated by dividing each sum of squares by its degree of freedom. The *F* ratio for testing equality of group means is

$$F = \frac{\text{Mean Square Between}}{\text{Mean Square Within}} = \frac{BSSM}{WSSM}$$

The significance level is obtained from the F distribution with numerator and denominator degrees of freedom.

Basic Statistics

Descriptive Statistics

$$\text{Sample size} = W_q$$
$$\text{Mean} = \overline{T}_q$$
$$\text{Standard deviation} = S_q$$
$$\text{Standard error} = S_q \Big/ \sqrt{W_q}$$

95% Confidence Interval for the Mean

$$\overline{T}_q \pm t_{W_q - 1} S_q \Big/ \sqrt{W_q}$$

where t_{W_q-1} is the upper 2.5% critical value for the t distribution with $W_q - 1$ degrees of freedom.

Variance Estimates and Confidence Interval for Mean

Fixed-Effects Model

Pooled Standard Deviation

$$S_p = \sqrt{WSSM}$$

Standard Error

$$\text{Standard error} = \sqrt{WSSM/W}$$

95% Confidence Interval for the Mean

$$\overline{G} \pm t_{W-k'}\sqrt{WSSM/W}$$

where $t_{W-k'}$ is the upper 2.5% critical value for the t distribution with $W-k'$ degrees of freedom.

Random-Effects Model

Between-Groups Component of Variance (Snedecor and Cochran 1967)

$$\omega^2 = \frac{(BSSM - WSSM)(W(k'-1))}{\left(W^2 - \sum_{i=1}^{k} W_i^2\right)}$$

Standard Error of the Mean = $\sqrt{V(\overline{G})}$ (Brownlee 1965)

where

$$V(\overline{G}) = \frac{\left(\sum_{i=1}^{k} W_i^2\right)(k'-1)(BSSM - WSSM)}{W\left(W^2 - \sum_{i=1}^{k} W_i^2\right)} + \frac{WSSM}{W}$$

If $BSSM < WSSM$, $V\left(\overline{G}\right) = \frac{WSSM}{W}$ and a warning is printed that the variance component estimate is negative.

95% Confidence Interval for the Mean

$$\overline{G} \pm t_{k'-1}\sqrt{V\left(\overline{G}\right)}$$

where $t_{k'-1}$ is the upper 2.5% critical value for the t distribution with $k'-1$ degrees of freedom.

Tests for Homogeneity of Variances

Levene Test

$$L = \frac{(W-k')\sum_{i=1}^{k'} W_i\left(\overline{Z}_i - \overline{Z}\right)^2}{(k'-1)\sum_{i=1}^{k'}\sum_{l=1}^{n_i} w_{il}\left(Z_{il} - \overline{Z}_i\right)^2}$$

where

$$Z_{il} = \left|X_{il} - \overline{T}_i\right|$$

$$\overline{Z}_i = \frac{\sum_{l=1}^{n_i} w_{il} Z_{il}}{W_i}$$

$$\overline{Z} = \frac{\sum_{i=1}^{k'} W_i \overline{Z}_i}{W}$$

User-Supplied Contrasts

Let C_1 through C_k be the coefficients for a particular contrast. If the sum of the coefficients is not 0, a warning is printed and the contrast number is starred. For each contrast the following are printed.

Value of the Contrast

$$V = \sum_{i=1}^{k} \overline{T}_i C_i$$

Pooled Variance Statistics

Standard Error

$$SE = \sqrt{S_p^2 \sum_{i=1}^{k} C_i^2 / W_i}$$

t Value

$$t = V / SE$$

Degrees of Freedom

$$W - k'$$

Two-tailed significance level based on the *t* distribution with $W - k'$ degrees of freedom

Separate Variance Statistics

Standard Error

$$SE = \sqrt{\sum_{i=1}^{k} C_i^2 \left(S_i^2 / W_i \right)}$$

t Value

$$t = V / SE$$

Degrees of Freedom (Brownlee 1965)

$$df = \frac{\left(\sum_{i=1}^{k} C_i^2 S_i^2 / W_i \right)^2}{\sum_{i=1}^{k} \left(C_i^2 S_i^2 / W_i \right)^2 / (W_i - 1)}$$

Two-tailed significance level based on the *t* distribution with *df* degrees of freedom

Polynomial Contrasts (Speed 1976)

If the specified degree of the polynomial (*NP*) is less than or equal to 0, or greater than 5, a message is printed and the procedure is terminated. If the degree of the polynomial specified is greater than the number of nonempty groups, it is set to $k' - 1$. If the sums of the weights in each group are equal, only the WEIGHTED contrasts will be generated. For unequal sample sizes with equal spacing between groups, both WEIGHTED and UNWEIGHTED contrasts are computed. For unequal sample sizes and unequal spacing, only WEIGHTED contrasts are computed. The metric for the polynomial is the group code.

UNWEIGHTED Contrasts and Statistics

The coefficients for the orthogonal polynomial are calculated recursively from the following relations:

$$c_{i,q} = \left(i - A_q\right)c_{i,q-1} - C_q c_{i,q-2}$$

for

$$q = 1, 2, \ldots, NP$$

$$i = 1, 2, \ldots, k$$

with the initial values

$$c_{i,-1} = 0, \quad c_{i,0} = 1$$

and

$$A_q = \frac{\sum_{i=1}^{k} i c_{i,q-1}^2}{\sum_{i=1}^{k} c_{i,q-1}^2}$$

$$C_q = \frac{\sum_{i=1}^{k} c_{i,q-1}^2}{\sum_{i=1}^{k} c_{i,q-2}^2} \quad \text{for q} \geq 2$$

$$C_q = 0 \quad \text{for q} = 1$$

The F statistic for the qth degree contrast is computed as

$$F = \frac{\left[\sum_{i=1}^{k} \left(\overline{T}_i - \overline{G} \right) c_{i,q} \right]^2}{\sum_{i=1}^{k} c_{i,q}^2 \Big/ W_i} \Bigg/ WSSM$$

where $WSSM$ is the mean square within. The significance level is obtained from the F distribution with 1 and $W - k'$ degrees of freedom.

WEIGHTED Contrasts and Statistics (Emerson 1968; Robson 1959)

The contrast for the qth degree polynomial component is computed from the following recursive relations:

$$d_{i,q} = \left(i - A'_q\right) d_{i,q-1} - C'_q d_{i,q-2}$$

for

$q = 1, 2, \ldots, NP.$
$i = 1, 2, \ldots, k.$

with initial values

$$d_{i,0} = 1,\ d_{i,-1} = 0$$

$$A'_q = \frac{\sum_{i=1}^{k} i W_i d_{i,q-1}^2}{\sum_{i=1}^{k} W_i d_{i,q-1}^2}$$

$$C'_q = \frac{\sum_{i=1}^{k} i W_i d_{i,q-1} d_{i,q-2}}{\sum_{i=1}^{k} W_i d_{i,q-2}^2} \quad \text{for } q \geq 2$$

$$C'_q = 0 \quad \text{for } q = 1$$

The test for the contribution of the qth degree orthogonal polynomial component is based on

$$F = D_q / WSSM$$

where

$$D_q = \frac{\left(\sum_{i=1}^{k} W_i \overline{T}_i d_{i,q}\right)^2}{\sum_{i=1}^{k} W_i d_{i,q}^2}$$

The significance level is computed from the F distribution with degrees of freedom 1 and $W - k'$.

The test for deviation from the qth degree polynomial is based on

$$F = DD_q \big/ WSSM$$

where

$$DD_q = \left(BSS - \sum_{j=1}^{q} D_j\right) \Bigg/ (k' - q - 1)$$

The significance level is computed from the F distribution with degrees of freedom $k' - q - 1$ and $W - k'$. The highest degree printed will be the minimum of $(k' - 2)$ and 5.

Multiple Comparisons (Winer 1971)

Generation of Ranges (R_r)

The Student-Newman-Keuls (SNK), TUKEY, and TUKEYB procedures are all based on the studentized range, $S_{r,f}$, where r is the number of steps between means and f is the degrees of freedom for the within-groups mean square. For the above tests, only $\alpha = 0.05$ can be used.

The appropriate range of values for the tests are

SNK	$R_r = S_{r,f}$	$r = 2, \ldots, k'$
TUKEY	$R_r = S_{k',f}$	
TUKEYB	$R_r = \dfrac{\left(S_{r,f} + S_{k',f}\right)}{2}$	

For the DUNCAN procedure, alphas of 0.01, 0.05, and 0.10 can be used. The ranges $\left(D_{r,f}\right)$ are generated using the algorithm of Gebhardt (1966).

DUNCAN	$R_r = D_{r,f}$	$r = 2, \ldots, k'$

The Scheffé, LSD, and modified LSD procedures all use critical points from the *F* distribution. Any $\alpha \le 0.5$ can be used.

SCHEFFE	$R_r = \sqrt{2(k'-1)F_{1-\alpha}(k'-1, f)}$
LSD	$R_r = \sqrt{2F_{1-\alpha}(1, f)}$
MODLSD	$R_r = \sqrt{2F_{1-\alpha'}(1, f)}$

where

$$\alpha' = \frac{2\alpha}{k'(k'-1)}$$

Compute the multiplier of the ranges for the difference of means *i* and *j*.

$$M_{i,j} = S_p \sqrt{\frac{1}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \quad \text{(default)}$$

$$M_{i,j} = S_p \sqrt{\frac{\sum_{l=1}^{k} 1/n_l}{k'}} \quad \text{(harmonic mean for all groups)}$$

Establishment of Homogeneous Subsets

If the sample sizes in all groups are equal, or the harmonic mean for all groups has been selected, or the multiple comparison procedure is SNK or DUNCAN, homogeneous subsets are established as follows:

The means are sorted into ascending order from $\overline{T}_{(1)}$ to $\overline{T}_{(k')}$. Values of i and q such that

$$\left|\overline{T}_{(q)} - \overline{T}_{(i)}\right| \le R_{q-i+1} M_{q,i} \qquad (*)$$

are systematically searched for and

$$\left\{\overline{T}_{(i)}, \ldots, \overline{T}_{(q)}\right\}$$

is considered a homogeneous subset. The search procedure is as follows:

At each step t, the value of i is incremented by 1 (the starting value is 1), and $q = k'$. The value of q is then decremented by one until $(*)$ is true. Call this value q_t. If $q_t > q_{t-1}$ and $(*)$ is true,

$$\left\{\overline{T}_{(i)}, \ldots, \overline{T}_{q_t}\right\}$$

is considered homogeneous. Otherwise i is incremented and the next is step done. The procedure terminates when $i = k$ or $q_t = k$.

In all other situations, all nonredundant pairs of groups are compared using the criteria of $(*)$. A table containing all pairs of groups is printed with symbols indicating group means that are significantly different.

Welch Test

In Welch (1947,1951), he derived the an approximate test for equality of means without the homogeneous variance assumption. The statistic is given by

$$F_{Welch} = \frac{\sum_{l=1}^{k} \omega_l \left[(\overline{T}_l - \tilde{X})^2 / (k-1) \right]}{1 + \frac{2(k-2)}{(k^2-1)} \sum_{l=1}^{k} \left[(1 - \frac{\omega_l}{\mathrm{u}})^2 / (W_l - 1) \right]}$$

where $\omega_l = W_l / S_l^2$, $\mathrm{u} = \sum_{l=1}^{k} \omega_l$, and $\tilde{X} = \sum_{l=1}^{k} \omega_l \overline{T}_l / \mathrm{u}$.

The Welch statistic has an approximate F distribution with k-1 and f degrees of freedom, where

$$\mathrm{f} = \left[\frac{3}{k^2 - 1} \sum_{l=1}^{k} (1 - \frac{\omega_l}{\mathrm{u}})^2 / (W_l - 1) \right]^{-1}.$$

This statistics and the chisquare approximation proposed by James (1951) have close connection with the asymptotic chi-square test. The asymptotic chi-square statistic has a very simple formula as below,

$$\mathrm{C} = \sum_{l=1}^{k} \omega_l (\overline{T}_l - \tilde{X})^2.$$

As $W_1, \ldots, W_k \to \infty$, the statistic C has an asymptotic chi-square distribution with $k-1$ degrees of freedom. The Welch statistic is derived to improve the small sample of behavior of C by using a F approximation. One can still see the similarity between the numerators of chisquare statistic and Welch statistic.

Since the weight used in Welch statistic is $\omega_l = W_l / S_l^2$, one cannot compute the statistic if any one group has zero standard deviation. Moreover, sample sizes of all groups have to be greater than or equal to zero.

Brown-Forsythe Test

In Brown and Forsythe (1974a,1974b), a test statistic for equal means was proposed. The statistic as the following form,

$$\mathrm{F}_{BF} = \frac{\sum_{l=1}^{k} W_l (\overline{T}_l - \overline{G})^2}{\sum_{l=1}^{k} (1 - W_l / W) S_l^2}.$$

The statistic has an approximate F distribution with (k-1) and f degrees of freedom, where

$$\frac{1}{\mathrm{f}} = \sum_{l=1}^{k} \mathrm{c}_l^2 / (W_l - 1)$$

and

$$\mathrm{c}_l = \frac{(1 - W_l / W) S_l^2}{\sum_{l=1}^{k} (1 - W_l / W) S_l^2}.$$

When we look at the denominator of F_{BF}, we can see that it tries to estimate the ‘pooled variance’ by

$$\mathrm{S}_{\mathrm{pool}}^2 = \sum_{l=1}^{k} \omega_l^* S_l^2 ,$$

where

$$\omega_l^* = \frac{(W - W_l)}{W(k-1)}.$$

We can see that the weight of i-th group decreases as sample size increases. Moreover, the weight tends to ‘average out’ the differences in sample sizes. This can be visualized in the following example.

n_i	10	20	30	40	50
w_i^*	.233	.217	.200	.183	.167
n_i/N	.067	.133	.200	.267	.333

With such a weighting scheme, the pooled variance will not be dominate by a group with very large sample size and thus preventing over/under estimate of the pooled variance. Under the assumption of unequal variances, S^2_{pool} does not have chi-square distribution anymore. However, by using Satterthwaite’s method (1946), one can show that S^2_{pool} (with proper scaling) has an approximate chi-square distribution with f degree of freedom.

The Brown & Forsythe statistic cannot be computed if all groups have zero standard deviation or any group has sample size less than or equal to 1. In the situation that some groups have zero standard deviations, the statistic can be computed but the approximation may not work.

References

Brown, M. B. and A. B. Forsythe (1974a). "The small sample behavior of some statistics which test the equality of several means", *Technometrics*, v16, 129-132.

Brown, M. B. and A. B. Forsythe (1974b). "Robust tests for the equality of variances", *J. American Statistical Association*, v69, 364-367.

Brownlee, K. A. 1965. *Statistical theory and methodology in science and engineering*. New York: John Wiley & Sons, Inc.

Duncan, D. B. 1955. Multiple range and multiple *F* tests. *Biometrics*, 11: 1–42.

Eisenhart, C. 1947. Significance of the largest of a set of sample estimates of variance. In: *Techniques of Statistical Analysis*, C. Eisenhart, M. W. Hastay, and N. A. Wallis, eds. New York: McGraw-Hill.

Emerson, P. L. 1968. Numerical construction of orthogonal polynomials from a general recurrence formula, *Biometrics*, 24: 695–701.

Gebhardt, F. 1966. Approximation to the critical values for Duncan's multiple range test. *Biometrics*, 22: 179–182.

James, G. S. (1951). "The comparison of several groups of observations when the ratios of the population variances are unknown", *Biometrika*, v38,324-329.

Kramer, C. Y. 1956. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 12: 307–310.

Miller, R. G., Jr. 1966. *Simultaneous statistical inference*. New York: McGraw-Hill.

Robson, D. S. 1959. A simple method for construction of orthogonal polynomials when the independent variable is unequally spaced. *Biometrics*, 15: 187–191.

Satterthwaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin*, 2, 110 -114.

Snedecor, G. W., and Cochran, W. G. 1967. *Statistical methods*. Ames, Iowa: Iowa State University Press.

Speed, F. M. 1976. Response curves in the one-way classification with unequal numbers of observations per cell. *Proceedings of the Statistical Computing Section, American Statistical Association*. 270–272.

Welch, B. L. (1947). "The generalization of Student's problem when serveral different population variances are involved", *Biometrika*, v34, 28-35.

Welch, B. L. (1951). "On the comparison of several mean values: an alternative approach", *Biometrika*, v38, 330-36.

Winer, B. J. 1971. *Statistical principles in experimental design*, 2nd ed. New York: McGraw-Hill.

Young, E. A., and Cramer, E. W. 1971. Some results relevant to choice of sum and sum-of-product algorithms. *Technometrics*, 13: 657–665.