Introduction to the Analysis of Variance

Sunthud Pornprasertmanit   Chulalongkorn University

Sometimes, researchers want to compare mean differences between three or more groups, either dependent groups or independent groups.

Null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$

If one pair of groups has significant mean differences, the null hypothesis is not tenable.

Alternative hypothesis $H_1: \mu_i \neq \mu_j$

The Analysis of Variance (ANOVA) proves whether the null hypothesis is tenable.

The Drawbacks of Multiple *t* Tests

Why the researcher do not used multiple t tests to prove this null hypothesis?

Null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$

Supposed that *k* =4, the group differences can be proved in C_2^k pairs.

$H_0: \mu_1 = \mu_2$ $H_0: \mu_1 = \mu_3$ $H_0: \mu_1 = \mu_4$

$H_0: \mu_2 = \mu_3$ $H_0: \mu_2 = \mu_4$ $H_0: \mu_3 = \mu_4$

If null hypothesis is true, the probability of accepting true null hypothesis is equal to 1-α.

If null hypothesis is true, the probability of all pairs accepting true null hypothesis is equal to $(1-\alpha)^C$ (C = number of pairs).

Therefore, the probability of rejecting at least one pair, if null hypothesis is true, is equal to

$$1 - (1 - \alpha)^C$$

If *k* = 3, this probability is equal to .14.

If *k* = 4, this probability is equal to .26.

Because of inflated type I error, the multiple *t* tests should not be used. The ANOVA method can control the probability of making a type I error equal to α (Familywise or Experimentwise Error Rate).

Author Note
This article was written in September 2007 for teaching in Introduction to Statistics in Psychology Class, Faculty of Psychology, Chulalongkorn University
Correspondence to Sunthud Pornprasertmanit. Email: psunthud@gmail.com

ANOVA as a Regression Analysis

No Predictor

You will see that when the predictor variable is equal to arithmetic mean, the sum of squared error is the least.

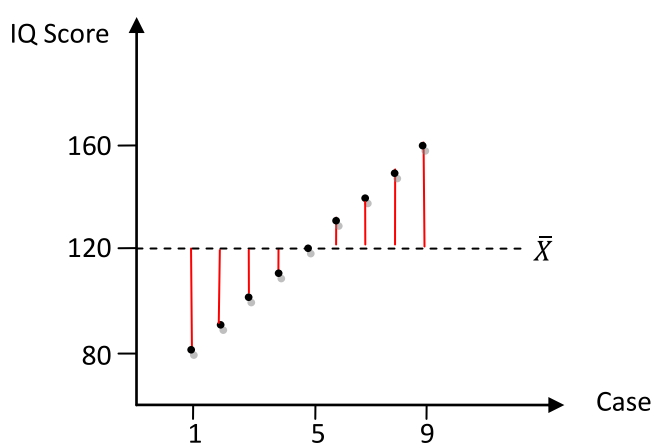
Arithmetic Mean

ID	IQ	Baseline Prediction	Error of Prediction	Squared error
1	80	120	-40	1600
2	90	120	-30	900
3	100	120	-20	400
4	110	120	-10	100
5	120	120	0	0
6	130	120	10	100
7	140	120	20	400
8	150	120	30	900
9	160	120	40	1600
Total	960		0	6000

Sum of Errors equal to zero.

$SS_{error} = SS_X = 6000$

$$SS_{error} = SS_X = \sum (X - \bar{X})^2$$



One Grouping Variable

If the predictor is interval, the linear transformation is used. This is called regression analysis.

However, if the predictor is categorical variable, the values that can predict all value leaving least error are group means.

$$X' = \bar{X}_{.j} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}$$

$$\bar{X}_{..} = \sum_{j=1}^{k} n_j \bar{X}_{.j} = \sum X/n$$

Criterion (*Y*)

Group Mean

ID	IQ	Class	Prediction Score	Error of Prediction	Squared error
1	80	1	90	-10	100
2	90	1	90	0	0
3	100	1	90	10	100
4	110	2	120	-10	100
5	120	2	120	0	0
6	130	2	120	10	100
7	140	3	150	-10	100
8	150	3	150	0	0
9	160	3	150	10	100
Total	960			0	600

Sum of Errors equal to zero.

$SS_{error} = 600$

$$SS_{error} = \sum_{j-1}^{k}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_{.j}\right)^2$$

SS_{error} reduces from 6000 to 600. (SS_{error} may be called SS_{within})

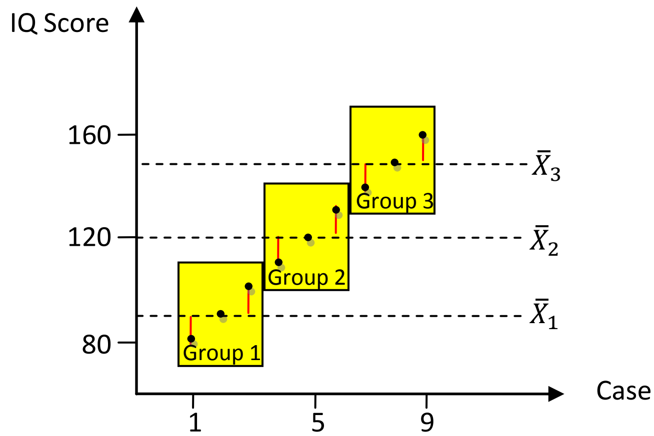
For ANOVA, SS_{error} from no predictor is SS_{total}. (As regression analysis)

The difference between SS_{error} from one predictor and SS_{total} is SS_{group}. ($SS_{regression}$ in regression analysis)

$$SS_{\text{total}} = SS_{\text{group}} + SS_{\text{error}}$$

The proportion of SS_{group} and SS_{total} is eta squared (the percentage of variance that group can be explained).

$$\eta^2 = \frac{SS_{\text{group}}}{SS_{\text{total}}} = 1 - \frac{SS_{\text{error}}}{SS_{\text{total}}}$$



From the table, you will see that the score can be divided into two components.

$$X_{ij} = \bar{X}_{.j} + e_{ij}$$

Score = Group mean + Error

$$X_{ij} = \bar{X}_{..} + (\bar{X}_{.j} - \bar{X}_{..}) + (X_{ij} - \bar{X}_{.j})$$

$$X_{ij} = \mu + \alpha_j + \varepsilon_{ij}$$

Score = Grand mean +Treatment effect + Error effect

This equation is called sample model equation.

For example

Case 1	$80 = 120 + (-30) + (-10)$
Case 6	$130 = 120 + (0) + (10)$

Basic Concepts of ANOVA for Testing Hypothesis

When researchers want to test hypothesis about equality of group means, the prefer statistic is one-way ANOVA.

Null hypothesis	$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$
Alternative hypothesis	$H_1: \mu_i \neq \mu_j$

For understanding logic of ANOVA, there are two key terms that researchers should know, degree of freedom and mean of squares.

The degrees of freedom in ANOVA are three components.

$$df_{\text{group}} = k - 1$$

$$df_{\text{error}} = df_{\text{error in each group}} = \sum_{j=1}^{k} (n_j - 1) = n - k$$

$$df_{\text{total}} = n - 1$$

$$df_{\text{total}} = df_{\text{group}} + df_{\text{error}}$$

The means of squared error (often called mean squares) are the sum of squared error divided by its degree of freedom. (As variance)

$$\hat{\sigma}^2 = MS_{\text{total}} = \frac{SS_X}{df_X} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}$$

$$MS_{\text{group}} = \frac{SS_{\text{group}}}{df_{\text{group}}} = \frac{\sum_{j=1}^{k} (\bar{X}_{.j} - \bar{X}_{..})^2}{k - 1}$$

$$MS_{\text{error}} = \frac{SS_{\text{error}}}{df_{\text{error}}} = \frac{\sum_{j-1}^{k} \sum_{i=1}^{n_j} \left(X_{ij} - \bar{X}_{.j}\right)^2}{n - k}$$

MS_{error} estimates variation within groups, such as the variation of participants who used the same diet.

MS_{group} estimates variation between groups, such as the variation of participants who use different diet.

If null hypothesis is true, the expected value of MS_{group} and MS_{error} is σ_ε^2.

$$E\left(MS_{\text{group}}\right) = E(MS_{\text{error}}) = \sigma_\varepsilon^2$$

Therefore, if null hypothesis is true,

$$F = \frac{MS_{\text{group}}}{MS_{\text{error}}} \approx 1$$

This ratio, when random sampled from population, is distributed in *F*-distribution with df_{group} and df_{error}. If null hypothesis is true, the *F* statistic is close to 1.

If null hypothesis is not tenable, the expected value of MS_{group} is not equal to σ_ε^2.

$$E(MS_{\text{group}}) = \sigma_\varepsilon^2 + n\sigma_\alpha^2 = \sigma_\varepsilon^2 + \frac{n\sum(\mu_j - \mu)^2}{k-1}$$

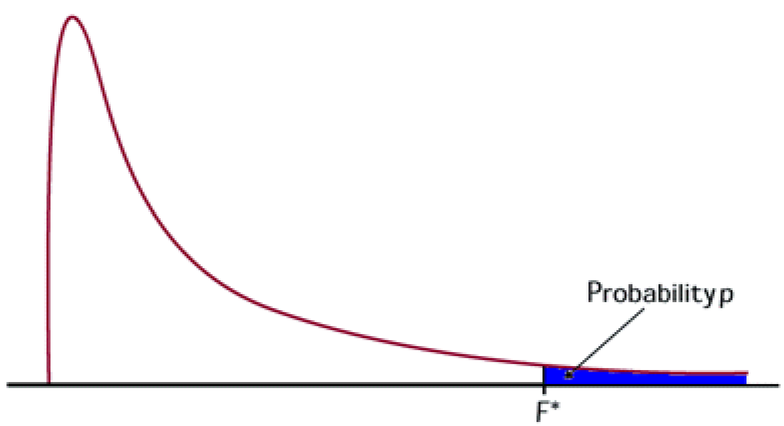
$$E(MS_{\text{error}}) = \sigma_\varepsilon^2$$

The expected value of MS_{group} includes a function of population treatment, $\mu_j - \mu$.

Therefore, if null hypothesis is not tenable, the *F* statistic should be larger than 1.

$$F = \frac{MS_{\text{group}}}{MS_{\text{error}}} \geq 1$$

How much larger than one should *F* be for a researcher to feel confident in rejecting the null hypothesis?



Then, this *F* statistic for comparing means has only one-tailed test.

The researcher should determine the probability that the null hypothesis is unlikely to be true or alpha level.

If the chance of type I error of the specified *F* (*p* value) is less than alpha level, the null hypothesis rejected.

If the chance of type I error of the specified *F* is more than alpha level, the null hypothesis is tenable and the researchers cannot draw the conclusion whether the research hypothesis is true.

The p value can be calculated from MS Excel.

FDIST(*X*, df_1, df_2)

Example

Assumption of one-way ANOVA

1) The model equation $X_{ij} = \mu_{..} + (\mu_{.j} - \mu_{..}) + (X_{ij} - \mu_{.j})$ reflects all the sources of variation affect X_{ij}.
2) Participants are random samples from the respective populations or the participants have been randomly assigned to the treatment levels.
3) The *j* = 1, 2, … , *k* populations are normally distributed.
4) The variances of the *j* = 1, 2, … , *k* populations are equal. (Homogeneity of variance)

The model equation is suitable for comparing means between independent groups. When the methods for dividing are more than one or when the groups are dependent, the other statistic should be used.

The one-way ANOVA is robust with respect to departures from normality. This is especially true when the populations are symmetrical and the sample sizes are equal and greater than 12.

The one-way ANOVA is robust with respect to violation of the homogeneity of variance assumption provided (1) there is an equal number of observations in each of the groups (2) the populations are normal, and (3) the ratio of the largest variance to the smallest variance does not exceed 3.

Otherwise, the Welch (or Brown-Forsyte) procedure in one-way ANOVA is preferred. (The general formula of independent *t* test when heterogeneity of variance)

The homogeneity of variance assumption can be checked by Levene test.

Fixed and Random Effect

Fixed Effect: Deliberately selected levels of independent variable

Random Effect: Randomly selected levels of independent variable

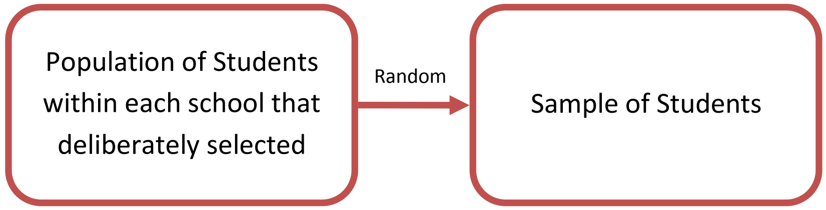
Fixed and random effects differ in level of generalization.

The significant result from fixed effect ANOVA tells that the dependent effect was different among exactly the same levels of independent variable.

The significant result from random effect ANOVA tells that the dependent effect was different among the population of the levels of independent variable.

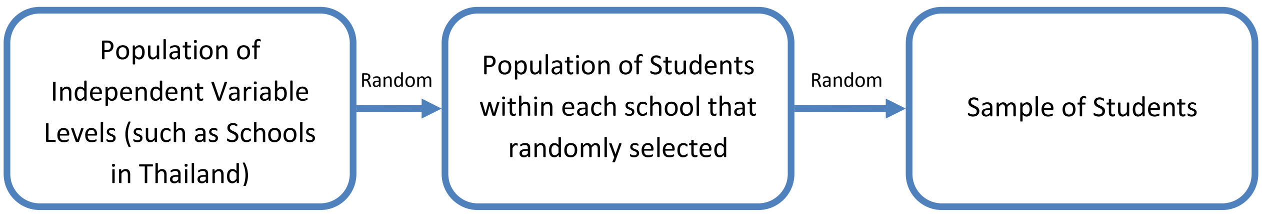
The fixed effect generally has more power than the random effect. If the selected levels of independent variables are similar to population of independent variables levels (the population is clear), use the fixed effect ANOVA. However, if not (or the population is not clear), use the random effect ANOVA.

Fixed Effect



Generalization: Within levels of independent variable that deliberately selected (for example, are school A, B, and C different in socioeconomic status?)

Random Effect



Generalization: Population of independent variable levels (for example, are schools in Thailand different in socioeconomic status?)

Multiple Comparison Procedures

If null hypothesis $H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$ is rejected, which population means are not equal?

The group of procedure for comparing group means is multiple comparisons.

If the multiple comparisons are done before *F* test, it is called priori contrast. Otherwise, it is called post hoc test.

The null hypothesis in multiple comparisons has two types (such as in 3 groups)

Pairwise contrast	Nonpairwise contrast
$H_0: \mu_1 - \mu_2 = 0;\ \mu_1 = \mu_2$	$H_0: \frac{1}{2}(\mu_1 + \mu_2) - \mu_3 = 0;\ \frac{1}{2}(\mu_1 + \mu_2) = \mu_3$
$H_0: \mu_1 - \mu_3 = 0; \mu_1 = \mu_3$	$H_0: \frac{1}{2}(\mu_1 + \mu_3) - \mu_2 = 0; \frac{1}{2}(\mu_1 + \mu_3) = \mu_2$
$H_0: \mu_2 - \mu_3 = 0;\ \mu_2 - \mu_3$	$H_0: \frac{1}{2}(\mu_2 + \mu_3) - \mu_1 = 0; \frac{1}{2}(\mu_2 + \mu_3) = \mu_1$

A contrast or comparison among means is a difference among means.

The general formula of contrast is

$$\psi = \sum_{j=1}^{k} c_j \mu_j$$

For example

$$\psi_1 = (1)\mu_1 + (-1)\mu_2 + (0)\mu_3 = \mu_1 - \mu_2$$

$$\psi_2 = (1)\mu_1 + (0)\mu_2 + (-1)\mu_3 = \mu_1 - \mu_3$$

$$\psi_3 = (0)\mu_1 + (1)\mu_2 + (-1)\mu_3 = \mu_2 - \mu_3$$

$$\psi_4 = \left(\frac{1}{2}\right)\mu_1 + \left(\frac{1}{2}\right)\mu_2 + (-1)\mu_3 = \frac{1}{2}(\mu_1 + \mu_2) - \mu_3$$

$$\psi_5 = \left(\frac{1}{2}\right)\mu_1 + (-1)\mu_2 + \left(\frac{1}{2}\right)\mu_3 = \frac{1}{2}(\mu_1 + \mu_3) - \mu_2$$

$$\psi_6 = \left(\frac{1}{2}\right)\mu_1 + \left(\frac{1}{2}\right)\mu_2 + (-1)\mu_3 = \frac{1}{2}(\mu_2 + \mu_3) - \mu_1$$

In general, the coefficient of a contrast should be

1) $\sum_{j=1}^{k} c_j = 0$
2) $\sum_{j=1}^{k} |c_j| = 2$

The general formula of null hypothesis of null hypothesis is

Null hypothesis	$H_0: \psi_i = 0$	
Alternative hypothesis	$H_1: \psi_i \neq 0$	(Two-tailed)
	$H_1: \psi_i > 0; \psi_i < 0$	(One-tailed)

Orthogonal Contrast

That is

SPSS Contrasts Grouping

Deviation (first)/ Deviation (last) → Effect coding with first or last as a reference group

Simple (first)/Simple (last) → Dummy coding with first or last as a reference group

Repeated → Each category is compared to the previous category

Helmert/ Reverse Helmert → Each category (except the last or first) is compared to the mean effect to all subsequent or previous categories.

Method for Multiple Comparisons

Although, for each comparison, the type I error rate is α, for multiple comparisons, the familywise error rate inflates to $1 - (1 - \alpha)^C$ (if comparisons are independent or orthogonal).

If the comparisons are dependent, the $1 - (1 - \alpha)^C$ still approximate familywise error rate.

In general, the researchers specify the familywise error rate not over significance level (such as .05 or .01). Then, there are a lot of methods to test multiple contrasts that restrict type I error not over significance level. The differences of these procedures are

1) Priori or post hoc contrasts

   The priori contrast is the method that specified the tested comparison before doing research.

   The post hoc test (posteriori contrast) is the method that tested all pairwise or all possible comparisons after found that the group means were significantly different.

   According to Howell (2007), the difference between two methods is vague. In priori contrast, the researchers test a small number of comparisons; however, in post hoc test, the researchers test all pairwise or all possible comparisons. They do not care whether specifying the comparison before doing research or not.

2) Overall *F* required or not

   A few methods require the overall *F* test to be significant before testing. However, most methods (esp. the priori contrast) do not require for testing overall *F* test because the familywise error rate is not over significance level.

3) Specifed or pairwise or nonpairwise contrasts
   a. Orthogonal Contrast
   b. Specifying the best levels of independent variable
   c. k -1 contrasts with a control group means (Special form of orthogonal contrast)
   d. C contrasts
   e. All pairwise contrasts
   f. All possible (pairwise and nonpairwise) contrasts
4) Equal or not equal sample size in each group
5) Homogeneity or Heterogeneity of variance
6) One-tailed available
7) Confidence intervals available
8) Power

The most multiple comparison tests control the probability of making one or more Type I errors at or less than α for a collection of tests.

Priori Contrast

Multiple *t*-test or *F*-test (with one degree of freedom).

In comparing a pairwise contrast, the researcher can use t-test for comparing this difference. If squared the t-test with m degrees of freedom, the distribution is F-distribution with 1 and m degrees of freedom.

The general formula for testing individual contrasts is

$$F = \frac{MS_{\text{Contrast}}}{MS_{\text{error}}} = \frac{n\psi^2/\sum c_j^2}{MS_{\text{error}}}$$

Use MSerror in overall *F* test

This method is useful if combining with Bonferreni or Dunn-Sidak method.

Bonferreni *t*.

Bonferreni inequality showed that

Error rate per comparison (α') ≤ Familywise error rate $[1 – (1 - \alpha')^C]$ ≤ $c\alpha'$

Then, error rate per comparison = /c

Dunn-Sidak Test

Holm and Larzelere and Mulaik Test

Post Hoc Contrast

LSD

Tukey Test

Dunnett's C

Games-Howell

Newman-Keuls Test

Ryan Procedure (REGWQ)

Hsu

Dunnett's Test

Scheffe

Brown-Forsythe

Hochberg's GT2 → Sample sizes are unequal, not robust to heterogeneity of variance

Gabriel → Sample sizes are unequal, more powerful, became too liberal when the sample size are very different.

Homogeneous Subset

D

Trend Analysis

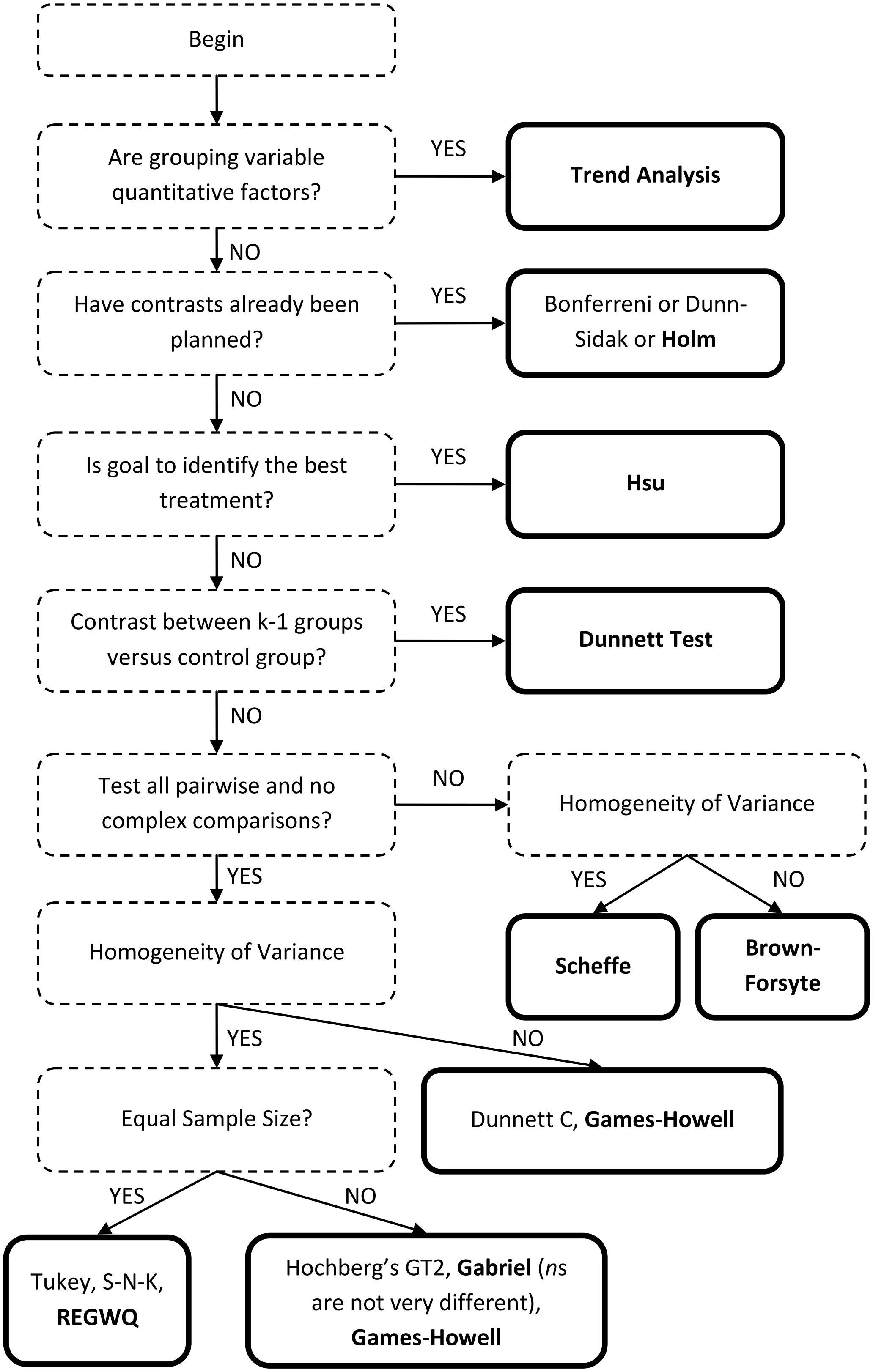
Polynomial

Linear

Quadratic

Cubic

Guideline for Using Multiple Comparison Procedures



Practical Significance

A measure of strength of association that is used in ANOVA F test is eta squared and omega squared.

The eta squared is the proportion of the population variance in the dependent variable that is accounted for by the *k* treatment levels.

$$\eta^2 = \frac{SS_{\text{group}}}{SS_{\text{total}}} = 1 - \frac{SS_{\text{error}}}{SS_{\text{total}}}$$

The eta squared is similar to the coefficient of determination, r^2, in regression analysis.

However, the eta squared cannot estimate the proportion of the population variance in the dependent variable that is accounted for by grouping variables. The number of groups (*k*) and the numbers of sample size (*n*) affect the estimation.

The omega squared is the corrected form of eta squared that can estimate this proportion.

$$\omega^2 = \frac{\sigma_\alpha^2}{\sigma_{TOTAL}^2} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$$

	FIXED EFFECT	RANDOM EFFECT
$\boldsymbol{\sigma_\alpha^2}$	$(k-1)(MS_{group} - MS_{error})/nk$	$(MS_{group} - MS_{error})/n$
$\boldsymbol{\sigma_\varepsilon^2}$	MS_{error}	MS_{error}
$\boldsymbol{\omega^2}$	$\frac{(k-1)(F-1)}{(k-1)(F-1)+nk}$	$\frac{(F-1)}{(F-1)+n}$
		(Squared Intraclass Correlation)

Cohen (1988) has suggested the following guidelines for interpreting strength of association.

$\omega^2 = .010$ is a small association.

$\omega^2 = .059$ is a medium association.

$\omega^2 = .138$ is a large association.

Hedges' g statistic can be used to determine the effect size of contrasts among the diets.

$$g = \frac{|\hat{\psi}_i|}{\sqrt{MS_{\text{error}}}}$$