

The ANOVA table: Understand the table usually used to express the results of an Analysis of Variance. This same table will also be used for regression.

Traditional ANOVA table

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	838.5976	209.6494	15.38	0.0001
Error	20	272.6680	13.6334		
Corrected Total	24	1111.2656			

SEE SAS OUTPUT

Expected Mean Square

What do we estimate when we calculate a pooled variance estimate (MSE) or the sum of squared treatment (SSTreatments) effects divided by its d.f.?

The MSE estimates σ^2 , the random variation for individuals in the population.

If the null hypothesis is true, the MS for Treatments also estimate the same random variation, σ^2 . The F value should only reject the null hypothesis $\alpha*100\%$ of the time.

But what if the null hypothesis is NOT true? Then, the MSTreatments estimates σ^2 , PLUS some additional component due to a treatment effect.

For a random effect this additional component would be called σ_τ^2 . This is a variance.

For a FIXED effect the additional component is simple the sum of squared effects divided by the d.f., $\frac{\sum \tau_i^2}{t-1}$. This is not a variance component.

The ANOVA source table with its d.f. and Expected mean squares (for a balanced design).

Note: 1 tailed test, n influences power

Source	d.f.	EMS Random	EMS Fixed
Treatment	t-1	$\sigma_\epsilon^2 + n\sigma_\tau^2$	$\sigma_\epsilon^2 + n \frac{\sum \tau_i^2}{(t-1)}$
Error	t(n-1)	σ_ϵ^2	σ_ϵ^2
Total	tn-1		

We could also express our null hypothesis in terms of EMS [$H_0: \sigma_\tau^2 = 0$], particularly for the random effect since the variance component for treatments may be a value of interest.

Since for a fixed effect the individual means are usually of interest, the null hypothesis is usually expressed in terms of the means ($H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_t$).

Descriptions of post-hoc tests

Post-hoc or Post-ANOVA tests! Once you have found out some treatment(s) are “different”, how do you determine which one(s) are different?

If we had done a t-test on the individual pairs of treatments, the test would have been done as

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{MSE \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}. \text{ If the difference between } \bar{Y}_1 - \bar{Y}_2 \text{ was large}$$

enough, the t value would have been greater than the $t_{critical}$ and we would conclude that there was a significant difference between the means. Since we know the value of $t_{critical}$ we could figure out how large a difference is needed for significance for any particular values of MSE, n_1 and n_2 . We do this by replacing t with $t_{critical}$ and solving for $\bar{Y}_1 - \bar{Y}_2$.

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{MSE \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \text{ so}$$

$$t_{critical} \sqrt{MSE \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \bar{Y}_1 - \bar{Y}_2 \quad \text{or} \quad \bar{Y}_1 - \bar{Y}_2 = t_{critical} S_{\bar{Y}_1 - \bar{Y}_2}$$

This value is the exact width of an interval $\bar{Y}_1 - \bar{Y}_2$ which would give a t-test equal to $t_{critical}$. Any larger values would be “significant” and any smaller values would not. This is called the “Least Significant Difference”. $LSD = t_{critical} S_{\bar{Y}_1 - \bar{Y}_2}$

This least significant difference calculation can be used to either do pairwise tests on observed differences or to place a confidence interval on observed differences.

The LSD can be done in SAS in one of two ways. The MEANS statement produces a range test (LINES option) or confidence intervals (CLDIFF option), while the LSMEANS statement gives pairwise comparisons.

The LSD has an α probability of error on each and every test or comparison. The whole idea of ANOVA is to give a probability of error that is α for the whole experiment, so, much work in statistics has been dedicated to this problem. Some of the most common and popular alternatives are discussed below. Most of these are also discussed in your textbook. The LSD is said to have a “comparisonwise” error rate.

The LSD is the LEAST conservative of those discussed, meaning it is the one most likely to detect a difference and it is also the one most likely to make a Type I error when it finds a difference. However, since it is unlikely to miss a difference that is real, it is also the most powerful. The probability distribution used to produce the LSD is the t distribution.

Bonferroni's adjustment. Bonferroni pointed out that in doing k tests, each at a probability of Type I error equal to α , the overall experimentwise probability of Type I error will be NO MORE than $k \cdot \alpha$, where k is the number of tests. Therefore, if we do 7 tests, each at $\alpha = 0.05$, the overall rate of error will be NO MORE than $= 0.35$, or 35%. So, if we want to do 7 tests and keep an error rate of 5% overall, we can do each individual test at a rate of $\alpha/k = 0.05/7 = 0.007143$. For the 7 tests we have an overall rate of $7 \cdot 0.007143 = 0.05$. The probability distribution used to produce the LSD is the t distribution.

Duncan's multiple range test. This test is intended to give groupings of means that are not significantly different among themselves. The error rate is for each group, and has sometimes been called a familywise error rate. This is done in a manner similar to Bonferroni, except the calculation used to calculate the error rate is $[1-(1-\alpha)^{r-1}]$ instead of the sum of α . For comparing two means that are r steps apart, where for adjacent means $r=2$. Two means separated by 3 other means would have $r = 5$, and the error rate would be $[1-(1-\alpha)^{r-1}] = [1-(1-0.05)^4] = 0.1855$. The value of α needed to keep an error rate of α is the reverse of this calculation, $[1-(1-0.05)^{1/4}] = 0.0127$.

Tukey's adjustment The Tukey adjustment allows for **all possible pairwise tests**, which is often what an investigator wants to do. Tukey developed his own tables (see Appendix table A.7 in your book for “percentage points of the studentized range”). For “ t ” treatments and a given error degrees of freedom the table will provide 5% and 1% error rates that give an experimentwise rate of Type I error.

Scheffé's adjustment This test is the most conservative. It allows the investigator to do not only all pairwise tests, but **all possible tests**, and still maintain an experimentwise error rate of α . “All possible” tests includes not only all pairwise tests, but comparisons of all possible combinations of treatments with other combinations of treatments (see CONTRASTS below). The calculation is based on a square root of the F distribution, and can be used for range type tests or confidence intervals. The test is more general than the others mentioned, for the special case of pairwise comparisons, the statistic is $\sqrt{(t-1)*F_{t-1, n(t-1)}}$ for a balanced design with t treatments and n observations per treatment.

Place the post-hoc tests above in order from the one most likely to detect a difference (and the one most likely to be wrong) to the one least likely to detect a difference (and the one least likely to be wrong). **LSD is first, followed by Duncan's test, Tukey's and finally Scheffé's.** Dunnett's is a special test that is similar to Tukey's, but for a specific purpose, so it does not fit well in the ranking. The Bonferroni approach produces an upper bound on the error rate, so it is conservative for a given number of tests. It is a useful approach if you want to do a few tests, fewer than allowed by one of the others (e.g. you may want to do just a few and not all possible pairwise). In this case, the Bonferroni may be better.

Evaluating the assumptions for ANOVA.

We have already discussed some techniques for the evaluation of data for homogeneous variance. The assumption of independence is somewhat more difficult to evaluate. Random sampling is the best guarantee of independence and should be used as much as possible.

The third assumption is normality. The observations are assumed to be normally distributed within each treatment, but how the treatments come together to form the dependent variable Y_{ij} may cause them to look non-normal. The best way to test for normality is to examine the residuals, pooling the normal distribution across the treatments to a common mean of zero. SAS will output the residuals with an output statement, and PROC UNIVARIATE has a number of tools to evaluate normality.

Homogeneity of Variance

Your textbook discusses one test by Hartley. It is one of the simplest tests, but not usually the best. To do this test we calculate the largest observed variance divided by the

smallest observed variance. This statistics is tested with a special table by Hartley (Appendix Table 5.A in your Freund & Wilson textbook).

A number of other tests are available in SAS, but only for a simple CRD (i. e. a One-way ANOVA). These test are briefly discussed below.

To get all of the tests available in SAS, use the following statement following PROC GLM.

```
MEANS your_treatment_name / HOVTEST=BARTLETT
HOVTEST=BF HOVTEST=LEVENE(TYPE=ABS)
HOVTEST=LEVENE(TYPE=SQUARE) HOVTEST=OBRIEN WELCH;
```

Levene's Test: This test is basically an ANOVA of the squared deviations (TYPE=SQUARE). It can also be done with absolute values (TYPE=ABS). This is one of the most popular HOV tests.

O'Brien's Test: This test is a modification of Levene's with an additional adjustment for kurtosis.

Brown and Forsythe's Test: This test is similar to Levene's, but uses absolute deviations from the median instead of more ANOVA like means. There is a “nonparametric” ANOVA that employs deviations from the median instead of the usual deviations from the mean used for the normal ANOVA.

Bartlett's Test for Equality: This test is similar to Hartley's, but uses a likelihood ratio test instead of an F test. This test can be inaccurate if the data is not normally distributed.

Welch's ANOVA: It is **not** a test of homogeneity of variance; this test is a weighted ANOVA. This ANOVA weights the observations by an inverse function of the variances and is intended to address the problem of non-homogeneous variance and to be use when the variance is not homogeneous.

The Homogeniety of Variance (HOV) tests discussed above can be done in SAS (PROC GLM). Note that the last one is NOT an HOV test, it is another type of ANOVA called a weighted ANOVA.

Contrasts and Orthogonality

A priori contrasts are one of the most useful and powerful techniques in ANOVA. There are a few additional considerations that should be made.

So what is a contrast? As described in the handout, it is a comparison of some means against some other means. The comparison is a linear combination.

When we set these up in SAS, we only need to give the multipliers in the CORRECT ORDER, and SAS will complete the calculations.

The multipliers must sum to zero, and they can be given as fractions or as integers.

For example, compare pounds of laundry where the treatments are HIS, HERS and OURS, we want to contrast HIS to HERS to each other and we want to contrast HIS and HERS combined to OURS.

Contrast 1: Contrast the mean of HIS to the mean of HERS, excluding the mean for OURS.

$H_0: \mu_{His} = \mu_{Hers}$. The multipliers are -1 and 1 for his and hers, which gets the positive and