Assumptions of the ANOVA F-test:

● Again, most assumptions involve the ε_{ij}'s (the error terms).

(1) The model is correctly specified.
(2) The ε_{ij}'s are normally distributed.
(3) The ε_{ij}'s have mean zero and a common variance, σ^2.
(4) The ε_{ij}'s are independent across observations.

● With multiple populations, detection of violations of these assumptions requires examining the residuals rather than the *Y*-values themselves.

● An estimate of ε_{ij} is:

● Hence the residual for data value Y_{ij} is:

● We can check for non-normality or outliers using residual plots (and normal Q-Q plots) from the computer.

● Checking the equal-variance assumption may be done with a formal test:

H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2$
H_a: at least two variances are not equal

● The Levene test is a formal test for unequal variances that is robust to the normality assumption.

● It performs the ANOVA F-test on the absolute residuals from the sample data.

Example pictures:

Remedies to Stabilize Variances

● If the variances appear unequal across populations, using transformed values of the response may remedy this. (Such transformations can also help with violations of the normality assumption.)

● The drawback is that interpretations of results may be less convenient.

Suggested transformations:

● **If the standard deviations of the groups increase proportionally with the group means, try:** $Y_{ij}^* = \log(Y_{ij})$

● **If the variances of the groups increase proportionally with the group means, try:** $Y_{ij}^* = \sqrt{Y_{ij}}$

● **If the responses are proportions (or percentages), try:** $Y_{ij}^* = \arcsin(\sqrt{Y_{ij}})$

● **If none of these work, may need to use a nonparametric procedure (e.g., Kruskal-Wallis test).**

<u>Making Specific Comparisons Among Means</u>

● **If our F-test rejects H_0 and finds there are significant differences among the population means, we typically want more specific answers:**

(1) Is the mean response at a specified level superior to (or different from) the mean response at other levels?

(2) Is there some natural grouping or separation among the factor level mean responses?

● **Question (1) involves a "pre-planned" comparison and is tested using a contrast.**

● **Question (2) is a "post-hoc" comparison and is tested via a "Post-Hoc Multiple Comparisons" procedure.**

Contrasts

● A contrast is a linear combination of the population means whose coefficients add up to zero.

Example (t = 4):

● Often a contrast is used to test some meaningful question about the mean responses.

Example (Rice data): Is the mean of variety 4 different from the mean of the other three varieties?

We are testing:

What is the appropriate contrast?

Now we test:

We can estimate L by:

Under H_0, and with balanced data, the variance of a contrast

is:

• Also, when the data come from normal populations, $\hat{L}$ is normally distributed.

• Replacing σ^2 by its estimate MSW:

For balanced data:

• To test H_0: $L = 0$, we compare t^* to the appropriate critical value in the t-distribution with $t(n - 1)$ d.f.

• Our software will perform these tests even if the data are unbalanced.

Example:

• Note: When testing multiple contrasts, the specified α (= P{Type I error}) applies to each test individually, not to the <u>series</u> of tests collectively.

<u>Post Hoc Multiple Comparisons</u>

● When we specify a significance level α, we want to limit P{Type I error}.

● What if we are doing many simultaneous tests?

● Example: We have $\mu_1, \mu_2, \ldots, \mu_t$. We want to compare <u>all pairs</u> of population means.

● <u>Comparisonwise error rate</u>: The probability of a Type I error on <u>each comparison</u>.

● <u>Experimentwise error rate</u>: The probability that the simultaneous testing results in <u>at least one</u> Type I error.

● We only do post hoc multiple comparisons if the overall F-test indicates a difference among population means.

● If so, our question is: Exactly <u>which</u> means are different?

● We test:

● The <u>Fisher LSD procedure</u> performs a t-test for each pair of means (using a common estimate of σ^2, MSW).

● The Fisher LSD procedure declares μ_i and μ_j significantly different if:

● **Problem: Fisher LSD only controls the <u>comparisonwise</u> error rate.**
● **The <u>experimentwise</u> error rate may be <u>much larger</u> than our specified α.**

● **<u>Tukey's Procedure</u> controls the <u>experimentwise</u> error rate to be only equal to α.**

● **Tukey procedure declares μ_i and μ_j significantly different if:**

● **$q_\alpha(t, \text{df})$ is a critical value based on the studentized range of sample means:**

● **Tukey critical values are listed in Table A.7.**

● **Note: $q_\alpha(t, \text{df})$ is larger than**

→ Tukey procedure will declare a significant difference between two means _______ often than Fisher LSD.

→ Tukey procedure will have _______ experimentwise error rate, but Tukey will have _______ power than Fisher LSD.

→ Tukey procedure is a _______ conservative test than Fisher LSD.

Some Specialized Multiple Comparison Procedures

● **Duncan multiple-range test: An adjustment to Tukey's procedure that reduces its conservatism.**
● **Dunnett's test: For comparing several treatments to a "control".**
● **Scheffe's procedure: For testing "all possible contrasts" rather than just all possible pairs of means.**

Notes: ● When appropriate, preplanned comparisons are considered superior to post hoc comparisons (more power).

● Tukey's procedure can produce simultaneous CIs for all pairwise differences in means.

Example:

<u>Random Effects Model</u>

● Recall our ANOVA model:

● If the *t* levels of our factor are the only levels of interest to us, then $\tau_1, \tau_2, \ldots, \tau_t$ are called <u>fixed effects</u>.

● If the *t* levels represent a random selection from a <u>large population</u> of levels, then $\tau_1, \tau_2, \ldots, \tau_t$ are called <u>random effects</u>.

<u>Example</u>: From a population of teachers, we randomly select 6 teachers and observe the standardized test scores for their students. Is there <u>significant variation</u> in student test score <u>among the population</u> of teachers?

● If $\tau_1, \tau_2, \ldots, \tau_t$ are random variables, the F-test no longer tests:

Instead, we test:

<u>Question of interest</u>: Is there significant variation among the effects for the different levels in the population?

● For the one-way ANOVA, the test statistic is exactly the same, F* = MSB / MSW, for the random effects model as for the fixed effects model.