21

One-Way Analysis of Variance

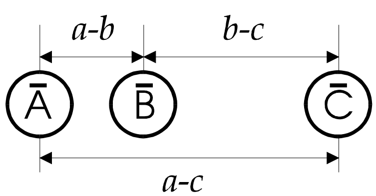
Why Not Multiple t-Tests?
F-Ratio Fundamentals
The F-Distribution Table
Computing the F-Ratio
Multiple Comparison Procedures

In Chapter 20 we presented two minor hypotheses from **Dr. John Babler's** dissertation on spiritual care provision through hospice agencies. His primary hypothesis was that "there would be significant differences in [provision of spiritual care scores] . . . between social workers, nurses, and spiritual care professionals."[1] The mean scores on provision of spiritual care for these three groups were 47. 23, 50.75, and 55.94 respectively. Applying the Analysis of Variance procedure, Babler found these three groups **differed significantly** in their provision of spiritual care (F=10.547, p=0.000). Application of the **(F)LSD multiple comparison test** revealed that the three groups differed significantly from each other.[2] Social workers scored lowest, professional spiritual care providers highest, and nurses in between.[3]

We established the fundamentals for parametric testing in Chapters 17 and 18. We learned how to apply one-sample z, t tests in Chapter 19. We extended these principles to two-sample tests in Chapter 20. The next logical step is testing the differences on a single dependent variable among ***three or more group means***. The procedure to use is one-way analysis of variance, more popularly known as "one-way ANOVA. "

Why Not Multiple t-tests?

But why learn another procedure? Why not just pair off the multiple group means and apply t-tests to each pair? The reason lies in the Type I error rate. If I have means A, B, and C, I could use three independent t-tests on A-B, A-C, and B-C, as show here:



[1]Babler, p. 32

[2]Ibid., p. 47. Note: use of the Least Significant Difference test is valid when the F-ratio is significant. This test was designed by Sir Ronald Fisher, developer of the ANOVA procedure (hence the name "F-Ratio"). Carmer and Swanson call this the Fisher-Protected LSD (FLSD).

[3]Ibid., p. 48

The multiple application of t-tests was used earlier in this century until Englishman R. A. Fisher showed that the Type I error rate expands from α to some larger value as the number of tests between paired means increases. The error rate expansion is constant and predictable, given by the following equation:

$$\mathbf{p = 1 - (1 - \alpha)^k}$$

where p is the actual Type I error rate of all tests together, α is the stated level of significance, and k is the number of tests performed. In the A-B-C example above, the true probability (p) of committing a Type I error using three t-tests (α=0.05) is given by

$$p = 1 - (1 - \alpha)^k = 1 - (1 - .05)^3 = 1 - .95^3 = 1 - 0.857 = \underline{\mathbf{0.143}}$$

In other words, we run a **<u>14.3%</u>** chance of wrongly declaring two means significantly different, even though we set the error rate (α) at 5%.

The problem grows with the number of means in the experiment. Suppose an experiment consists of ten groups. The researcher decides to apply the independent t-test (α = 0.05) to all pairs of means. The number of required t-tests equals (k)(k-1)/2 where k is the number of means in the experiment.With k=10 means, there are 10(9)/2 or ***45 t-tests*** to compute. The Type I error rate across these 45 tests (p) is

$$p = 1 - (1 - .05)45 = 1 - .9545 = 1 - .0994 = \underline{\mathbf{.9006}}$$

This means there is a *90% chance of committing a Type I error*, with α set at 5%! Since we want to lock the Type I error rate to α when testing multiple means, multiple t-tests should not be used. Sir Ronald Fisher proposed a solution, however, and he named his procedure the Analysis of Variance, or ANOVA. The "F" in "F-ratio" comes from his name.

$$z = \frac{X-\mu}{\sigma}$$

$$z = \frac{X-\overline{X}}{s}$$

$$t = \frac{X-\overline{X}}{s}$$

$$z = \frac{\overline{X}-\mu}{\sigma_{\overline{X}}}$$

$$t = \frac{\overline{X}-\mu}{s_{\overline{X}}}$$

$$t = \frac{\overline{X}-\overline{Y}}{s_{\overline{X}-\overline{Y}}}$$

$$t = \frac{\overline{d}}{s_{\overline{d}}}$$

We have been walking down the "parametric road of differences" since chapter 16. From the simple z-score formula in chapter 16, through one- and two- sample parametric tests, there has been a common thread tying all these procedures together. That thread — perhaps you've already seen it — is that ***every procedure involves a ratio of "difference between" to "difference within."***

This z-equation for a score is a ratio of "difference between" an individual score and population mean in the numerator, and "difference within" the group (population standard deviation) in the denominator. If n > 30, s estimates σ, and $\overline{X}$ estimates μ, giving the second z-equation.

We can use the t-equation (3rd), especially when n<30. It uses the same ratio of "difference between" score and sample mean and "difference within" (estimated standard deviation).

The next z-equation (4th) is a ratio of "difference between" a sample mean ($\overline{X}$) and population or hypothesized mean (μ) in the numerator, and "difference within" the sampling distribution (standard error of the mean) in the denominator. The t-equation (5th) uses estimated values in the same ratio.

The next t-equation (6th) is a ratio of "difference between" two sample means in the numerator, and "difference within" both samples (standard error of difference) in the denominator.

While the form of the last t-equation (7th) changes, conceptually it maintains the ratio of "difference between" paired scores to "difference within" all scores together.

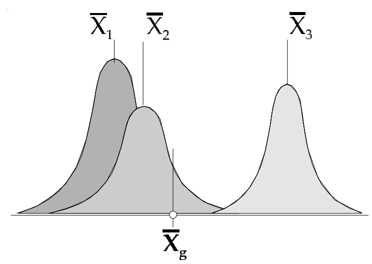
In all these cases, the ***between-to-within ratio*** remains constant. ANOVA continues this relationship. No matter how many groups are involved in an experiment, the ANOVA procedure breaks down the sum of squares of all experimental scores into two parts — a "difference between" part and a "difference within" part. The ratio of "between" to "within" differences forms the F-ratio, just as we have done all along.

Computing the F-Ratio

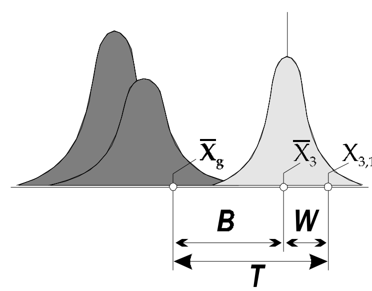
The process of calculating the F-ratio involves computation of sums of squares, degrees of freedom, and variance estimates which produce the F-ratio itself. Finally we will describe how to test the F-ratio for significance. Students ask, "Why do we need to study these details when computers will do this for us?" I answer, "You need more than numbers spit out of a computer -- you need to know what the computer is doing, at least on a basic level, in order to understand the printouts. Besides. . . *its neat!*"

Sums of Squares

We illustrate the process of dividing the **Total Sum of Squares (SS_t)** into two parts with the diagram at right. Here we find three groups of scores with grand mean $\overline{X}_g$ (the mean of all scores in the study) and sample means $\overline{X}_1$, $\overline{X}_2$, and $\overline{X}_3$.



Let's focus on one score in sample 3, indicated by the letter "$X_{3,1}$" at right. The distance between X and $\overline{X}_g$ ("*T*" in the diagram) equals this one score's part of the "total" deviation between the scores and the grand mean. When we subtract $\overline{X}_g$ from every score in the experiment and add these deviations together, we'll get 0 ($\Sigma x=0$). Square all the deviations and sum them to produce $\mathbf{SS_t}$ for the experiment. The full equation for SS_t is



$$SS_t = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_g)^2$$

where ***k*** is the number of groups, ***j*** is the group counter which increments from 1 to k, n_j is the number of scores in the jth group, and ***i*** is the score counter which increments from 1 to n_j. The equation says to subtract the grand mean from each score in each group, square the deviation, and sum them all up across groups to produce SS_t.

Next, notice that the "T" line in the diagram is equal to the sum of the other two parts. The first part is labelled "B" -- for **Between Sum of Squares (SS_b)** -- and is the deviation of $\overline{X}_3$ from $\overline{X}_g$. If we square these deviations, adjust for the number of subjects in each group, and sum them, we'll have $\mathbf{SS_b}$ for the experiment. The full equation for SS_b is

$$SS_b = \sum_{j=1}^{k} n_j(\overline{X}_j - \overline{X}_g)^2$$

The equation says to subtract the grand mean from each sample mean, square the difference, and multiply by the number of subjects in the sample. Add the *k* elements together to produce SS_b.

The second part is labelled "W" -- for **Within Sum of Squares (SS_w)** -- and is the deviation of $X_{3,1}$ from $\overline{X}_3$. If we square all of these deviations and sum them for Sample Three, and do the same for Samples One and Two, we'll have $\mathbf{SS_w}$ for the experiment. The equation for SS_w is

$$SS_w = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - \overline{X}_j)^2$$

The equation says to subtract each group's mean from each of the scores within that group, square the differences, and add them up. Add all these elements across all groups to produce SS_w. This gives us the combined "within" sum of squares for all the groups in the experiment. Notice that the "T"otal line in the diagram is equal to the "B"etween and "W"ithin segments, illustrating that the total sum of squares in any experiment of three or more groups can be divided into two parts, SS_b and SS_w, such that $\mathbf{SS_t = SS_b + SS_w}$.

Degrees of Freedom

Each sum of squares term (SS_b, SS_w, SS_t) in ANOVA has an associated df term (df_b, df_w, df_t).

The between df term is *k groups minus 1* (**df_b = k-1**).

The within df term is the *total number of scores in the experiment (N) minus the number of groups (k)* in the experiment (**df_w = N-k**).

In the **one**-sample test we lost **one** degree of freedom (df = n-**1**). In the **two**-sample test we lost **two** degrees of freedom (df = n + n **- 2**). It follows that when **k groups** are studied, we lose ***k*** degrees of freedom (**df_w = N-k**.)

The total df term equals *number of subjects minus 1* (**df_t = N-1**).

$SS_b + SS_w = SS_t$. In the same way, $df_b + df_w = df_t$.

k - 1 + N - k = N-1

Variance Estimates

Review: Recall from Chapter 16 that *variance (s^2) equals the sum of squares (Σx^2) divided by degrees of freedom (n-1).*

ANOVA computes a *between variance estimate*[4] (**MS_b**) and a *within variance estimate* (**MS_w**) from the SS and df terms defined above.

$$MS_w = \frac{SS_w}{df_w} = \frac{SS_w}{N-k} \qquad MS_b = \frac{SS_b}{df_b} = \frac{SS_b}{k-1}$$

The "MS" terms stand for "mean-square." Variance equals the average sum of squares. So, MS_b ("mean-square-between") stands for the mean of the squared deviations between groups. And MS_w ("mean-square-within") stands for the "mean of the squared deviations within all groups."

We do not take the square root of variance, as we did in the previous procedures. The F-ratio is built from these two variance estimates — hence the name, **Analysis of Variance.**[5]

The F-Ratio

The ***F-ratio*** of ANOVA is the ratio of MS_b and MS_w. This value is compared to a critical F value drawn from the F-distribution table to determine whether it is significant or not.

[4]Variance (**s^2**) equals sum of squares (**Σx^2**) divided by degrees of freedom (**n-1**). The two MS terms equals sum of squares divided by degrees of freedom.

[5]If you were to apply the independent-samples t-Test and ANOVA to the same two groups of scores, the resulting t, F values would have the relationship: $F = t^2$ or $t = \sqrt{F}$.

The F-Distribution Critical Value Table

In chapters 19 and 20 we used a "df" term to locate critical values with the t-distribution. The F-distribution table requires two "df" values to determine the critical value for an F-ratio. You will find part of an F-table in **A3-3** in the back of the text. The table is labelled "degrees of freedom in the numerator" on top and "degrees of freedom in the denominator" down the left side. These terms refer to the df_b (=k-1) and df_w (=N-k) from the ANOVA computations.

Suppose we test 4 groups with 28 subjects (k=4, N=28). The df_b is k-1, or 3; the df_w is N-k, or 24. Look down the column headed "3" until you cross the row labeled "24." The 0.05 critical value for F is ***3.01.***

The ANOVA Table

All of the F-Ratio elements we've discussed in this section are usually summarized in a concise chart called an ANOVA table. Tables such as the one below are commonly found in scientific literature. Study the relationships among the elements of the table and link the notations found here with the previous discussion of the computation of the F-Ratio. Here is the general format:

Source	*SS*	*df*	*MS*	*F*
Between	SS_b	df_b	$MS_b = SS_b/df_b$	$F = MS_b/MS_w$
Within	SS_w	df_w	$MS_w = SS_w/df_w$	
Total	SS_t	df_t		

Let's put all this together in an example problem.

An Example

At the beginning of the chapter we highlighted the findings of **Dr. John Babler's** dissertation on spiritual care provision. Here are two ANOVA tables from his dissertation.

TABLE 2[6]

ANALYSIS OF VARIANCE OF SCORE AND HOSPICE PROFESSION

Source	DF	Sum of Squares	Mean Squares	F Ratio	F Probability
Between Groups	2	1360.6566	680.3283	10.5465	.0000
Within Groups	193	12449.9302	64.5074		
Total	195	13810.5867			

The F-ratio of 10.5465 is significant because "p=.0000," that is, p(F) is very small, much less than α=0.05. This p(F) tells us that the spiritual care provision means of the three professions (47. 23, 50.75, and 55.94) were significantly different.

[6] Babler, p. 47

TABLE 8[7]

ANALYSIS OF VARIANCE OF SCORE AND AGE

Source	DF	Sum of Squares	Mean Squares	F Ratio	F Proba-bility
Between Groups	4	657.6667	219.2222	3.1952	.0247
Within Groups	191	13104.3128	68.6090		
Total	194	13761.9795			

In Table 8 we see an F of 3.1952 (p=.0247). This p(F) tells us that the spiritual care provision means of the four age groups of professionals were **significantly different.** These means were 52.08 (26-35), 48.39 (36-45), 52.46 (46-55) and 52.12 (over 55).

In these examples, you can see that the computer printout includes a ***p* value** for the computed F-ratio. This ***p*** is the *exact probability of obtaining the computed F-ratio.* It is easier to compare the computed ***p*** with ***α*** than it is to look up an F critical value in a table. ***If p < α**, then **reject the null hypothesis.***

Multiple Comparison Procedures

The purpose of ANOVA is to determine whether the sample means are indicative of experimental treatment effects or merely reflect chance variation. Two statistical conclusions are possible. Either the H_0: $\mu_1 = \mu_2 = . . . = \mu_k$ is tenable (means are equal), or it is rejected (means are different). A significant F-ratio leads to the rejection of the null hypothesis, but it *does not tells us which means differ significantly from the others.*

Procedures Defined

To determine *which means differ* sufficiently to produce the significant F-ratio, they study differences between pairs of means by using statistical techniques called ***multiple comparison procedures.*** These procedures vary in definition and procedure. We will briefly discuss four procedures here: the Least Significant Difference (LSD), the Tukey Honestly Significant Difference (HSD), the Student-Newman-Keuls (SNK) and the Fisher-Protected Least Significant Difference (FLSD).

The Least Significant Difference

The Least Significant Difference (LSD) is a form of the "multiple t-test" we discussed at the beginning of the chapter. Since the LSD focused on each pair of means individually, it uses a *comparison-wise* Type I error rate.[8] Fisher developed the LSD and recommended it as a multiple comparison test, but stipulated that ***LSD should be used only if the F-ratio were significant***. The stipulation of "significant F" locks Type I error rate to α and ***prevents the inflated error rates*** often reported in textbooks discussing multiple comparison tests. *Why*, reasoned Fisher, *would anyone use the LSD to find significant differences in an experiment that failed to produce a significant F?* The

[7]Babler, p. 54

[8]The concepts of comparison- and experiment-wise error rates is beyond the scope of this book. Suffice it so say that c-wise procedures (LSD) applies α to each pair, which can lead to an inflated Type I error rate; e-wise procedures (HSD) applies α to all pairs in the experiment, leading to low power per pair.

answer, of course, lies in the demands for educator-researchers to conduct and publish research. Since refereed journals demand articles reporting significant findings, many professors sought ways to produce significant findings more often. One way this was done was to use the LSD ***without*** a prior significant F-ratio, violating Fisher's guidelines. The result was an explosion of Type I errors and false positives in the literature. The problem was not the LSD test *per sec*, but its misuse. Other theorists sought ways to reduce the problem of excessive Type I error rate. The LSD generates the lowest critical value (and the highest level of power) of those discussed here, *when used as Fisher directed.*

The Honestly Significant Difference

J. W. Tukey developed the Honestly Significant Difference (HSD) in response to the inflated Type I error rates produced by misuse of the LSD. This procedure is based on an *experimentwise* Type I error rate (an error rate for all pairs taken together) which holds Type I error at α no matter how many groups are in the experiment. As the number of groups increases, the critical value used to test differences increases. The problem with the HSD is that, while it protects against Type I errors, it also *yields less power* in detecting "significant differences." The more groups tested, the less power achieved. The HSD generates the highest critical value (and the lowest level of power) of those discussed here.

Multiple Range Tests

A compromise approach between the low critical value ("liberal Type I error rate") of the LSD and the high critical value ("conservative Type I error rate") of the HSD was to create a *range* procedure. This multiple comparison procedure generates a ***range of critical values*** from conservative (equal to HSD) to liberal (equal to LSD). The Student-Newman-Keuls (SNK) is the most popular range statistic. Most statisticians discourage the use of range comparisons because it confuses the error values. Nonetheless, I found through a Dissertation Abstracts search in 1984 that the ***SNK was by far the most common multiple comparison used*** in graduate school research at that time.

Fisher-Protected Least Significance Difference

Displeased with the loss of power with the HSD, seeking ways to avoid the explosion of Type I errors with the (often misused) LSD, and avoiding the confusion of range statistics, some statisticians returned to Fisher's original prescription: *use the LSD only if the F-ratio is significant*.

Applying the LSD only to **significant F-ratios** ***limits experiment-wise error rate to*** α. At the same time, it ***maximizes the power*** of the test to detect pairwise differences. This modified procedure was called the Fisher-Protected Least Significant Difference (FLSD).[9]

Procedures Computed

In each of these procedures, a value is computed and compared to the differences between paired means. If the difference between two means is greater than the mul-

[9]Yount, "A Monte Carlo Analysis...," 26-37

tiple comparison value, the two means are declared "significantly different."

We will forego the specific formulas for each of these procedures since this computational work will most likely be done by computer. However, given specific elements of an ANOVA example, multiple comparison critical values can be compared with each other. The following table displays the values:

r	(F)LSD	SNK	HSD
5	12.522*	17.531	17.531**
4	12.522	16.502	17.531
3	12.522	15.026	17.531
2	12.522	12.522	17.531

Notice that the (F)LSD procedure produces the smallest critical value (12.522), producing the greatest power. The HSD procedure produces the largest critical value (17.531), producing the least power. The SNK procedure produces a range between the two.

There is a great deal of confusion in the literature concerning multiple comparisons. My Ph.D. dissertation* focused on six multiple comparison tests and analyzed their error rates by a computerized Monte Carlo technique. I generated 28,000 sets of random data, 1000 tests for each of 28 n- and k-combinations related to educational research. My findings indicated **the best multiple comparison procedure, under all conditions, was the (F)LSD.** It consistently provided the **greatest power** and an **error rate closest to α.** HSD was too conservative (consistently produced low power). The Scheffe method (not discussed in the chapter, but included in the study) consistently decreased the level of significance below α, reducing the power of the test more than any other. Scheffe consistently reduces the likelihood of finding "true differences" and should be avoided except under very narrow conditions.

If you need a multiple comparison test, **my unqualified recommendation is the (F)LSD.** If using SPSS, check the box marked "LSD" under multiple comparison tests, but **only use the results if the F-ratio is significant.**

Summary

In this chapter we established the general procedure for use of the one-way Analysis of Variance (ANOVA) test. We explained the problem of using multiple t-tests. We illustrated the breakdown of total sum of squares into between and within parts. We showed how each element in an ANOVA table is computed. We discussed the ANOVA table and the relationships among the various table elements. Finally, we introduced the concept of multiple comparison procedures and illustrated their use.

Examples

Dr. John Babler's Table 2, p. 21-5, displayed a significant F between the means of the three professions. But which professional group differed significantly from the others? Babler used a computerized LSD test (with a significant F) to determine that

*Yount. "A Monte Carlo Analysis..." Ph.D. diss., University of North Texas, 1985, 45-46 each of the three means differed significantly from the others.[10]

We can put them in a difference table to see the paired-differences more clearly. 1 = Spiritual Care professionals, 2 = nurses, 3 = social workers.

		2	1	ranks
ranks	means	50.75	55.94	
3	47.23	**5.19**	**8.71**	<--- largest difference
2	55.75		**3.52**	<--- smallest difference

The largest difference (8.71) is between the highest mean (55.94) and the lowest (47.23). You can see the differences between all three paired means above. All of these differences exceeded the critical value, and were declared significant.

Dr. Gail Linam's dissertation (see Chapter 24-1 for full reference and Chapter 25 for the two-way ANOVA findings) compared reading comprehension in children grades 4-6 across three translations of Scripture, the King James, New International and New Century versions. She found that the KJV produced significantly lower comprehension scores than either NIV or NCV. She applied the FLSD procedure to determine exact differences between versions.

For the Old Testament Retelling (OTR) scores, mission children scored significantly **lower** with KJV than NCV (7.00 < 23.11) and main campus children significantly **lower** with KJV than either NCV or NIV (18.81 < 34.41, 18.81 < 32.95).

For the New Testament Retelling (NTR) scores, mission children scored significantly **lower** with KJV than either NCV or NIV (7.25<29.44, 7.25<21.56) and main campus children the same (25.55<37.55, 25.55<34.60).[11]

For Old Testament Cloze (OTC) scores, mission children scored significantly **lower** with KJV than either NCV or NIV (4.22<17.56, 4.22<13.33), and main campus children the same (6.55<23.50, 6.55<23.27).[12]

For New Testament Cloze (NTC) scores, mission children scored significantly **lower** with the KJV than with either NCV or NIV (0.38<16.11, 0.38<10.78) and main campus children scored the same (11.05<23.50, 11.05<22.82).[13]

In every case and under every condition, the **KJV produced significantly lower reading comprehension scores** using two different types of testing procedures and stories from both Old and New Testaments. Older children (4th-6th grades) simply do not understand the King James translation as well as the NIV or NCV versions.

Vocabulary

ANOVA	Analysis of Variance: tests difference among 3 or more ind't samples means
df_b	degrees of freedom between: df between the means: =k-1
df_t	degrees of freedom total: df for whole experiment: =N-1
df_w	degrees of freedom within: (n-1 per group)(k groups) = N-k
FLSD	Fisher-protected LSD -- LSD test protected by a prior significant F-ratio
HSD	Tukey Honestly Significant Difference mcp: very conservative
LSD	Least Significant Difference—high Type I error without significant F-ratio

[10]Babler, 49 [11]Linam, 109 [12]Ibid., 111 [13]Ibid.

MS_w	mean square within: variance within-all-groups-combined
MS_b	mean square between: variance between means and grand mean
multiple t-tests	applying t-tests to multiple pairs of means in an experiment with three+ groups
SNK	Student-Newman-Keuls mcp which uses a range of critical values
SS_b	between sum of squares: SS term between grand mean and group means
SS_t	total sum of squares: SS term between grand mean and all scores
SS_w	within sum of squares: SS between scores and their respective group means

Study Questions

Dr. Martha Bergen studied attitudes toward computer-enhanced learning for seminary education among full-time professors at Southwestern Baptist Theological Seminary in 1989.[14] One of her hypotheses was that there would be a "significant difference [in attitude toward computer-enhanced learning] between the professors in the religious education, theology, and church music schools." Scores were generated from an attitude scale Dr. Bergen developed for the study. The mean attitude scores for the three schools were 118 (highest) in the Religious Education faculty, 117 in the church music faculty, and 114 (lowest) in the theology faculty. But were these differences in attitude significant? Here is the ANOVA table she generated:[15]

SOURCE OF VARIATION	SUM OF SQUARES	df	MS	F	p
Between	323.387	2	161.694	.472	.626
Within	25018.652	73	342.721		
Total	25342.039	75			

I. Using the problem and printout above, answer these questions:

1. Is the F-ratio significant? Explain why you say this.
2. Explain this F-ratio in terms of the three group means: 114, 117, 118.
3. How do you explain the differences in the school mean scores?
4. Dr. Bergen did not apply multiple comparisons tests to see if any single mean was significantly different from the others. Why ? Was she correct in doing so?

II. General Chapter Questions:

1. Explain the problem of using several t-tests to determine significant differences among pairs of means.

2. Since the FLSD is a modified multiple t-test, explain how it overcomes the problem explained in #1.

3. Explain in your own words how ANOVA divides the total sum-of-squares into "be tween" and "within" parts. (Use the "deviation" explanation).

4. **Fill in the ANOVA table below.** You are testing the means of 4 groups of 10 subjects each. The SS_b=480.0 and SS_t=1440.0. Compute the F-ratio. Determine the critical value (α=0.05). Is the F-ratio significant?

14 Bergen, Cover Page

15 Ibid., 87

Sample Test Questions

1. If you have **five** group means and apply (**5**x4/2=) 10 t-tests to the paired means, the level of Type I error will be
   A. equal to that of a one-way ANOVA
   B. greater than that of a one-way ANOVA
   C. less than that of a one-way ANOVA
   D. unknown

2. An experimentwise Type I error rate is the probability of committing a Type I error
   A. in each paired t-test individually
   B. in all paired t-tests together
   C. in the experimental groups only
   D. in the overall F-ratio test

3. The proper order of experimentwise error rate, high to low, of the following is
   A. LSD, HSD, SNK
   B. LSD, SNK, HSD
   C. HSD, SNK, FLSD
   D. HSD, LSD, SNK

4. The "mean-square" terms in the ANOVA are most closely associated with which of the following?
   A. Σx^2
   B. s^2
   C. ΣX
   D. σ

5. The "SS_b" term refers to Σx^2...
   A. between the sample means and the grand mean, summed
   B. among all scores within a given sample
   C. among all scores within the entire experiment
   D. between the scores and their own sample means, summed