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 $\sigma^2$, the random variation for individuals in the population.

If the null hypothesis is true, the MS for Treatments also estimate the same random variation, $\sigma^2$. 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 $\sigma^2$, PLUS some additional component due to a treatment effect.

For a random effect this additional component would be called $\sigma_\tau^2$. This is a variance.

For a FIXED effect the additional component is simple the sum of squared effects divided by the d.f., $\Sigma\tau_i^2 \Big/ t-1$. 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 | $\sigma_\varepsilon^2 + n\sigma_\tau^2$ | $\sigma_\varepsilon^2 + n\Sigma\tau_i^2 \Big/ (t-1)$ |
| Error | t(n–1) | $\sigma_\varepsilon^2$ | $\sigma_\varepsilon^2$ |
| 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 ( $H_0$: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_t$ ).

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

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

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

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 $\alpha$ probability of error on each and every test or comparison. The whole idea of ANOVA is to give a probability of error that is $\alpha$ 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 $\alpha$, the overall experimentwise probability of Type I error will be NO MORE than $k*\alpha$, 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 $\alpha/k = 0.055/7 = 0.007143$. For the 7 tests we have an overall rate of $7*0.007143 = 0.05$. 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 $\alpha$. 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 a needed to keep an error rate of $\alpha$ 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 $\alpha$. "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.

## Homogeniety 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

which gets the negative is not usually important. OURS gets a zero and is excluded from the calculations.

Contrast 2: Contrast the mean of HIS and HERS to the mean of OURS;

$H_0$: $\frac{\mu_{His} + \mu_{Hers}}{2} = \mu_{Ours}$. The multipliers are 1/2 and 1/2 for his and hers, and –1 for OURS (or negative on the 1/2s and positive on the 1). But we could also test $H_0$: $\mu_{His} + \mu_{Hers} = 2\mu_{Hers}$ and get the same results. The multipliers are now 1, 1 and –2 (or –1, –1 and 2).

| Contrast | HIS | HERS | OURS | SUM |
|---|---|---|---|---|
| **contrast 1** | **–1** | **1** | **0** | **0** |
| **contrast 2** | **–0.5** | **–0.5** | **1** | **0** |
| **alternative to 2** | **–1** | **–1** | **2** | **0** |

## Contrast calculations

A calculation similar to the LSD, but extended to more than just 2 means, is called a contrast. Suppose we wish to test the mean of the first two means against the mean of the last 3 means.

1) $H_0$: $\frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4 + \mu_5}{3}$ or $\frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4 + \mu_5}{3} = 0$ or

$$\left(\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2\right) - \left(\frac{1}{3}\mu_3 + \frac{1}{3}\mu_4 + \frac{1}{3}\mu_5\right) = 0 \text{ or}$$

$$\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 + \left(-\frac{1}{3}\right)\mu_3 + \left(-\frac{1}{3}\right)\mu_4 + \left(-\frac{1}{3}\right)\mu_5 = 0 \text{ or}$$

$$3\mu_1 + 3\mu_2 + (-2)\mu_3 + (-2)\mu_4 + (-2)\mu_5 = 0$$

This expression is what is a "linear model", and the last expression of this linear model is the easiest form for us to work with. We can evaluate the linear model, and if we can find the variance we can test the linear model. Generically, the variance of a linear model is "the sum of the variances", however there are a few other details. As with the transformations discussed earlier in the semester, when we multiply a value by "a" the mean changes by "a", but the variance changes by "$a^2$". Also, if there are covariances between the observations these must also be included in the variance. For our purposes, since we have assumed independence, there are no covariances.

The linear expression to evaluate is then: $a_1T_1+a_2T_2+a_3T_3+a_4T_4+...+a_kT_k$ where the "a" are the coefficients and the "T" are the treatment means (sums can also be used).

The variance is then: $a^2_1Var(T_1)+a^2_2Var(T_2)+a^2_3Var(T_3)+a^2_4Var(T_4)+...+a^2_kVar(T_k)$

In an ANOVA, the best estimate of the variance is the MSE, and the variance of a treatment mean is MSE/n, where n is the number of observations in that treatment. We can therefore factor out MSE, and in the balanced case (1/n) can also be factored out. The result is $\left(MSE/n\right)\left(a_1^2+a_2^2+a_3^2+a_4^2+...+a_k^2\right)$.

If we were to use a t-test to test the linear combination against zero, the t-test would be:

$$\frac{a_1 T_1 + a_2 T_2 + a_3 T_3 + a_4 T_4 + \ldots + a_k T_k}{\sqrt{\frac{MSE}{n}\left(a_1^2 + a_2^2 + a_3^2 + a_4^2 + \ldots + a_k^2\right)}} = \frac{\sum_{i=1}^{k} a_i T_i}{\sqrt{\frac{MSE}{n}\sum_{i=1}^{k} a_i^2}}$$

This is the test done with treatment means. If treatment totals are used the equation is modified slightly to $\sum_{i=1}^{k} a_i T_i \Big/ \sqrt{nMSE\sum_{i=1}^{k} a_i^2}$ and will give the same result.

One final modification. If we calculate our "contrasts" as above without the "MSE" in the denominator, then we calculate $Q = \sum_{i=1}^{k} a_i T_i \Big/ \sqrt{n\sum_{i=1}^{k} a_i^2}$ , without the MSE, then all that would remain to complete the t-test is to divide by $\sqrt{\text{MSE}}$ .

The value called "Q", when divided by $\sqrt{\text{MSE}}$ gives a t statistic. If we calculate $Q^2$ and divide by MSE we get an F statistic. SAS uses F tests. All we need provide SAS is the values of "a", the coefficients, in the correct order, and it will calculate and test the "Contrast" with an F statistic.

## Another example

Suppose we are comparing hemoglobin concentrations for various animals with diverse lifestyles. The animals included in our study are: Wrens, Dogs, Whales, People, Cod, Turkeys and Turtles.

We want to contrast 1) People to Others, 2) Aquatic species to others, and 3) Bird species to others.

1) People to Others – 1 category versus 6

2) Aquatic species to others – 3 categories versus 4

3) Bird species to others – 2 categories versus 5

| Contrast | Wrens | Dogs | Whale | People | Cod | Turkey | Turtle |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | –6 | 1 | 1 | 1 |
| 2 | 3 | 3 | –4 | 3 | –4 | 3 | –4 |
| 3 | –5 | 2 | 2 | 2 | 2 | –5 | 2 |

Note that all contrasts sum to zero.

In SAS, the contrast statements follow the PROC MIXED or PROC GLM statement.

SAS checks that they sum to zero (to 8 decimal places)

```
proc mixed data=clover order=data; class treatmnt;
     TITLE2 'ANOVA with PROC MIXED - separate variances';
   model percent = treatmnt / htype=3 DDFM=Satterthwaite outp=resids;
   repeated / group = treatmnt;
   lsmeans treatmnt / adjust=tukey pdiff;
** treatments in order=data ==========> 3DOk1 3DOk4 3DOk5 3DOk7 3DOk13;
   contrast '3 low vrs 2 high' treatmnt  -2    -2    -2     3     3;
   contrast 'odd vrs even'     treatmnt  -1     4    -1    -1    -1;
   contrast '1st vrs 2nd'      treatmnt  -1     1     0     0     0;
run;
```

## More on Contrasts and Orthogonality

Under some conditions, contrast sum of squares (SS) may add up to less than the treatment SS or they may add up to MORE than the treatment SS. The most satisfying condition is when they sum to equal the treatment SS. This is not necessarily a problem, as long as the contrasts are testing the hypotheses that you are interested in testing.

If we do only a few contrasts, fewer than the d.f. for the treatments, the contrast SS will probably add up to less than the treatment SS. No problem.

If we do MANY contrasts, more than the number of d.f. for treatments, the contrast SS will probably add up to more than the treatment SS. You are data-dredging? Consider a Scheffé adjustment.

If you do a number of contrasts equal to the number of treatment d.f., then the contrast SS can add up to more or less than the treatment SS. However, if the contrasts are orthogonal they will sum to exactly the treatment SS.

Contrasts are orthogonal if all their pairwise cross products sum to zero.

The cross products of a set of paired numbers is simply the product of the pairs. For example, take the following contrasts. Where the treatment levels are A1, A2, A3 and A4, write contrasts for A1 versus A2, A1 and A2 versus A3 and A4, and A3 versus A4.

These contrasts are given below.

| **Contrast** | **a1** | **a2** | **a3** | **a4** | **Sum** |
|---|---|---|---|---|---|
| **a1 v a2** | **–1** | **1** | **0** | **0** | **0** |
| **a1&a2 v a3&a4** | **–1** | **–1** | **1** | **1** | **0** |
| **a3 v a4** | **0** | **0** | **–1** | **1** | **0** |
| **Cross products** | | | | | |
| **c1 & c2** | **1** | **–1** | **0** | **0** | **0** |
| **c1 & c3** | **0** | **0** | **0** | **0** | **0** |
| **c2 & c3** | **0** | **0** | **–1** | **1** | **0** |

These contrasts are orthogonal. How about the set below?

Where the treatment levels are A1, A2, A3 and A4, write contrasts for A1 versus A2 and A3, A1 and A2 versus A3 and A4, and A3 versus A4.

If any one set of cross products do not sum to zero, the contrasts are not orthogonal. Orthogonality is a nice property, but not necessary. Write the contrasts that you want to test, orthogonal if possible.

Remember the ANOVA source table with its d.f. and Expected mean squares?

Well, a more "modern" approach involves estimating the variance components directly (PROC MIXED).

| Source | d.f. | EMS Random | EMS Fixed |
|---|---|---|---|
| Treatment | t–1 | $\sigma^2_\varepsilon + n\sigma^2_\tau$ | $\sigma^2_\varepsilon + n\Sigma\tau_i^2 \big/ t-1$ |
| Error | t(n–1) | $\sigma^2_\varepsilon$ | $\sigma^2_\varepsilon$ |
| Total | tn–1 | | |

Since the components are estimated directly there is no "sum of squares" for each line in the table. The model is fitted interatively (maximum likelihood).

### 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 | | | |

Te results of the tests and contrasts are usually the same. However, the mixed model analysis is capable of addressing issues tht PROC GLM cannot, so when differences exist in the analysis PROC MIXED is likely to give the better result.

PROC MIXED ANOVA table

Type 3 Tests of Fixed Effects

| Effect | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| treatmnt | 4 | 7.08 | 25.64 | 0.0003 |

Contrasts

| Label | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|
| 3 low vrs 2 high | 1 | 12.7 | 21.59 | 0.0005 |
| odd vrs even | 1 | 5.55 | 11.66 | 0.0161 |
| 1st vrs 2nd | 1 | 7.21 | 19.87 | 0.0027 |

## Summary

Understand the post-hoc tests. The range tests and contrasts. Be able to interpret these from SAS output.

Understand the differences between the post-hoc tests (error rates). Only one is correct for a particular objective.

Understand that contrasts are best done as *a priori* tests, and there is less concern with inflated Type I error rates if these are *a priori* tests. What is the error rate for contrasts by the way? The ANOVA was summarized. Note those aspects that I consider most important.

Understand Orthogonality.

Understand Expected mean squares. These will become extremely important in discussing larger designs. Fortunately SAS will give us the EMS (later), we need only understand them.

## The Factorial Treatment Arrangement

Also known as "two-way" ANOVA, this analysis has two (or more) treatments. For example, treatment A with two levels (a1 and a2) and treatment B with two levels (b1 and b2). The treatments are cross-classified such that each level of one treatment occurs in combination with each level of the other treatment (e.g. $a_1b_1$, $a_1b_2$, $a_2b_1$, $a_2b_2$).

Each treatment may be fixed or random (independently).

The combinations of treatments are still assigned at random to experimental units, so the design is still a CRD. For example, the 4 combinations in the example given ($a_1b_1$, $a_1b_2$, $a_2b_1$, $a_2b_2$)

would be assigned at random to the available experimental units, preferably in equal numbers to achieve a balanced design.

This treatment arrangement is called a "factorial", and the dimensions are usually given as 2 by 2 (above), 2 by 3, 3 by 3, etc. A schematic of a 3 by 3 factorial is given below.

| Treatments | A1 | A2 | A3 |
|---|---|---|---|
| B1 | a1b1 | a2b1 | a3b1 |
| B2 | a1b2 | a2b2 | a3b2 |
| B3 | a1b3 | a2b3 | a3b3 |

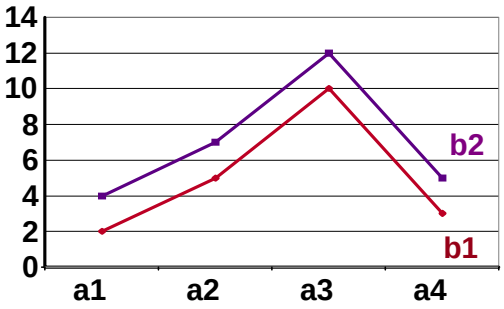
## Interactions

The principle treatments (A and B in the previous examples) are called main effects. The main effect for treatment A will be calculated from the marginal means or sums of the A treatment, averaged across the B treatment. Likewise, the main effect of treatment B will be calculated from the marginal means for treatment B average across the levels of A.

Marginal sums or means are used to evaluate the main effects.

| Treatments | A1 | A2 | A3 | B Means |
|---|---|---|---|---|
| B1 | a1b1 | a2b1 | a3b1 | b1 mean |
| B2 | a1b2 | a2b2 | a3b2 | b2 mean |
| B3 | a1b3 | a2b3 | a3b3 | b3 mean |
| A Means | a1 mean | a2 mean | a3 mean | |

Calculations for the main effects (Uncorrected treatment SS) are the same as for the CRD. There is however one new issue. It is possible for the same main effects to arise from various different cell patterns.

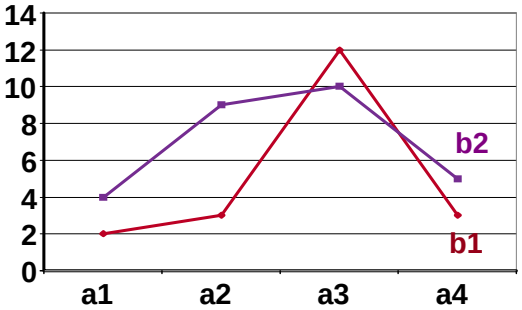



Plotting the means for the first case.

| Treatment | a1 | a2 | a3 | a4 | Means |
|---|---|---|---|---|---|
| b1 | 2 | 5 | 10 | 3 | 5 |
| b2 | 4 | 7 | 12 | 5 | 7 |
| Means | 3 | 6 | 11 | 4 | |

Plotting the means for the second case.

| Treatment | a1 | a2 | a3 | a4 | Means |
|---|---|---|---|---|---|
| b1 | 2 | 3 | 12 | 3 | 5 |
| b2 | 4 | 9 | 10 | 5 | 7 |
| Means | 3 | 6 | 11 | 4 | |




This lack of consistency in the cells is caused when the marginal means are not strictly additive. When additivity exists if some treatment marginal mean (#1) is larger by 2 units than some other mean (#2), each cell will in treatment #1 be 2 units higher than the corresponding mean of the treatment #2. This would represent additivity, or no interaction between the treatments.

If, however, the increases and decreases are not consistent, with the marginal means, then there is an interaction, or a lack of additivity. The marginal means (or sums) are used to calculate the main effects of the treatments. The cell to cell variation is used to measure the interaction