Analysis of Variance
Two way ANOVA

Two-Factor Experiments

- **In one-factor experiments, the purpose is to determine whether varying the level of a single factor affects the response.**
- **Many experiments involve varying several factors, each of which may affect the response. We will consider two- factor experiments.**
- **If one factor is fixed and one is random, then we say that the experiment follows a mixed model.**
- **In the two factor case, the tests vary depending on whether the experiment follows a fixed effects model, a random effects model, or a mixed model.**
- **We discuss methods ONLY for experiments that follow a fixed effects model.**

Example

Yields of a Chemical Process with Various Combinations of Reagent and

Catalyst	Reagent 1	Reagent 2	Reagent 3
A	86.8, 82.4, 86.7, 83.5	93.4, 85.2, 94.8, 83.1	77.9, 89.6, 89.9, 83.7
B	71.9, 72.1, 80.0, 77.4	74.5, 87.1, 71.9, 84.1	87.5, 82.7, 78.3, 90.1
C	65.5, 72.4, 76.6, 66.7	66.7, 77.1, 76.7, 86.1	72.7, 77.8, 83.5, 78.8
D	63.9, 70.4, 77.2, 81.2	73.7, 81.6, 84.2, 84.9	79.8, 75.7, 80.5, 72.9

replicates

- Four runs of the process were made for each combination of three reagents and four catalysts.
- There are two factors, the catalyst and reagent.
- The catalyst is called the row factor since its values varies from row to row in the table. I = # levels of row factor = 4
- The reagent is called the column factor. J = # levels column factor = 3.
- We will refer to each combination of factors as a treatment (some call these treatment combinations).

Example

- **Experimental units assigned to a given treatment are called replicates.**
- **When the number of replicates is the same for each treatment , we will denote this number by *K*.**
- **When observations are taken on every possible treatment, the design is called a complete design or a full factorial design.**
- **Incomplete designs, in which there are no data for one or more treatments, can be difficult to interpret.**
- **When possible, complete designs should be used.**
- **When the number of replicates is the same for each treatment, the design is said to be balanced.**
- **With two-factor experiments, unbalanced designs are much more difficult to analyze than balanced designs.**
- **The factors may be fixed or random. We deal with fixed factors only.**

Set-Up

- **In a completely randomized design, each treatment represents a population, and the observation on that treatment are a simple random sample from that population.**

- **We will denote the sample values for the treatment corresponding to the *i*th level of the row factor and the *j*th level of the column factor by $X_{ij1},\ldots, X_{ijK}$.**

- **We will denote the population mean outcome for this treatment by μ_{ij}.**
- **The values μ_{ij} are called the treatment means.**

- **In general, the purpose of a two-factor experiment is to determine whether the treatment means are affected by varying either the row factor, the column factor, or both.**

- **The method of analysis appropriate for two-factor experiments is called the two-way analysis of variance.**

Parameterization for Two-Way Analysis of Variance

- For any level *i* of the row factor, the average of all the treatment means μ_{ij} in the *i*th row is denoted $\bar{\mu}_{i.}$. We express $\bar{\mu}_{i.}$ in terms of the treatment means as follows:

$$\bar{\mu}_{i.} = \frac{1}{J}\sum_{j=1}^{J}\mu_{ij}$$

- For level *j* of the column factor, the average of all the treatment means μ_{ij} in the *j*th row is denoted $\bar{\mu}_{.j}$. We express $\bar{\mu}_{.j}$ in terms of the treatment means as follows:

$$\bar{\mu}_{.j} = \frac{1}{I}\sum_{i=1}^{I}\mu_{ij}$$

Parameterization for Two-Way Analysis of Variance

- **We define the population grand mean, denoted by μ, which represents the average of all the treatment means μ_{ij}. The population grand mean can be expressed in terms of the previous means:**

$$\mu = \frac{1}{I}\sum_{i=1}^{I}\bar{\mu}_{i.} = \frac{1}{J}\sum_{j=1}^{J}\bar{\mu}_{.j} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\mu_{ij}$$

More Notation

- **Using the quantities we just defined, we can decompose the treatment mean μ_{ij} as follows:**

$$\mu_{ij} = \mu + \left(\bar{\mu}_{i.} - \mu\right) + \left(\bar{\mu}_{.j} - \mu\right) + \left(\mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \mu\right)$$

- **This equation expresses the treatment mean μ_{ij} as a sum of four terms. In practice, simpler notation is used for the three rightmost terms in the above equation.**

$$\alpha_i = \bar{\mu}_{i.} - \mu$$

$$\beta_j = \bar{\mu}_{.j} - \mu$$

$$\gamma_{ij} = \mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \mu$$

Interpretations

$$\mu_{ij} = \mu + (\bar{\mu}_{i.} - \mu) + (\bar{\mu}_{.j} - \mu) + (\mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}_{.j} + \mu)$$

- **The quantity μ is the population grand mean, which is the average of all the treatment means.**

$$\alpha_i = \bar{\mu}_{i.} - \mu$$

- **The quantity α_i is called the *i*th row effect. It is the difference between the average treatment mean for the *i*th level of the row factor and the population grand mean. The value of α_i indicates the degree to which the *i*th level of the row factor tends to produce outcomes that are larger or smaller than the population mean.**

$$\beta_j = \bar{\mu}_{.j} - \mu$$

- **The quantity β_j is called the *j*th column effect. It is the difference between the average treatment mean for the *j*th level of the column factor and the population grand mean. The value of β_j indicates the degree to which the *j*th level of the column factor tends to produce outcomes that are larger or smaller than the population mean.**

Interpretations, contd.

$$\gamma_{ij} = \mu_{ij} - \overline{\mu}_{i.} - \overline{\mu}_{.j} + \mu$$

- **The quantity γ_{ij} is called the *ij* interaction. The effect of a level of a row (or column) factor may depend on which level of the column (or row) factor it is paired with. The interaction term measures the degree to which this occurs.**

For example, assume that level 1 of the row factor tends to produce a large outcome when paired with column level 1, but a small outcome when paired with column level 2. In this case γ_{11} would be positive, and γ_{12} would be negative.

More Set-Up

- Both row effects and column effects are called main effects to distinguish them from the interactions.
- Note that there are *I* row effects, one for each level of the row factor, *J* column effects, one for each level of the column factor, and *IJ* interactions, one for each treatment.

Furthermore, based on the re-parameterizations, the row effects, column effects, and interactions must satisfy the following constraints:

$$\sum_{i=1}^{I} \alpha_i = 0, \sum_{j=1}^{J} \beta_j = 0, \text{ and } \sum_{i=1}^{I} \gamma_{ij} = \sum_{j=1}^{J} \gamma_{ij} = 0$$

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

So, now we have:

- For each observation X_{ijk}, define $\varepsilon_{ijk} = X_{ijk} - \mu_{ij}$, the difference between the observation and its treatment mean. The quantities ε_{ijk} are called errors.
- It follows that $X_{ijk} = \mu_{ij} + \varepsilon_{ijk} \Rightarrow X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$

Continued….

- **When the interactions γ_{ij} are equal to zero, the additive model is said to apply. Under the additive model,**

$$X_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}.$$

- **When some or all of the interactions are not equal to zero, the additive model does not hold, and the combined effect of a row level and a column level cannot be determined from their individual main effects.**

Statistics

- The cell means are given by $$\bar{X}_{ij.} = \frac{1}{K}\sum_{k=1}^{K} X_{ijk}$$

- The row means are given by $$\bar{X}_{i..} = \frac{1}{J}\sum_{j=1}^{J} \bar{X}_{ij.} = \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K} X_{ijk}$$

- The column means are given by $$\bar{X}_{.j.} = \frac{1}{I}\sum_{i=1}^{I} \bar{X}_{ij.} = \frac{1}{IK}\sum_{j=1}^{J}\sum_{k=1}^{K} X_{ijk}$$

- The sample grand mean is given by $$\bar{X}_{...} = \frac{1}{I}\sum_{i=1}^{I} \bar{X}_{i..} = \frac{1}{J}\sum_{j=1}^{J} \bar{X}_{.j.}$$
$$= \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} \bar{X}_{ij.} = \frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} X_{ijk}$$

Estimating Effects

- We estimate the row effects by $\hat{\alpha}_i = \bar{X}_{i..} - \bar{X}_{...}$

- We estimate the column effect by $\hat{\beta}_j = \bar{X}_{.j.} - \bar{X}_{...}$

- We estimate the interactions with

$$\hat{\gamma}_{ij} = \bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...}$$

Example 2

Using the yield data, compute the estimated row effects, column effects, and interactions. ANOVA: Yield versus Catalyst, Reagent

Factor	Type	Levels	Values
Catalyst	fixed	4	A, B, C, D
Reagent	fixed	3	1, 2, 3

Analysis of Variance for Yield

Source	DF	SS	MS	F	P
Catalyst	3	877.56	292.52	9.36	0.000
Reagent	2	327.14	163.57	5.23	0.010
Catalyst*Reagent	6	156.98	26.16	0.84	0.550
Error	36	1125.33	31.26		
Total	47	2487.02			

S = 5.59099 R-Sq = 54.75% R-Sq(adj) = 40.93%

Means

Catalyst N Yield

Catalyst	N	Yield
A	12	86.417
B	12	79.800
C	12	75.050
D	12	77.167

Reagent	N	Yield
1	16	75.919
2	16	81.569
3	16	81.338

Example 2 contd

Catalyst	N	Yield
A	12	86.417
B	12	79.800
C	12	75.050
D	12	77.167

Row means

Reagent	N	Yield
1	16	75.919
2	16	81.569
3	16	81.338

Column means

Computed separately: Mean of Yield_1 = 79.6085 ≈ 79.61

Effects: Row effects: $\alpha 1$ = 86.42 – 79.61 = 6.81, etc.

Column effects: $\beta 1$ = 75.92 – 79.61 = -3.69, etc.

Interaction effects: $\gamma 11$ = 84.85 – 86.42 – 75.92 + 79.61 = 2.12, etc.

Using the Two-Way ANOVA to Test Hypotheses

A two-way ANOVA is designed to address three main questions:

1. **Does the additive model hold?**
   We test the null hypothesis that all the interactions are equal to zero:
   H_0: $\gamma_{11} = \gamma_{12} = \ldots = \gamma_{IJ} = 0$. If this null hypothesis is true, the additive model holds.

2. **If so, is the mean outcome the same for all levels of the row factor?**
   We test the null hypothesis that all the row effects are equal to zero:
   H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_I = 0$. If this null hypothesis is true, the mean outcome is the same for all levels of the row factor.

3. **If so, is the mean outcome the same for levels of the column factor?**
   We test the null hypothesis that all the column effects are equal to zero:
   H_0: $\beta_1 = \beta_2 = \ldots = \beta_J = 0$. If this null hypothesis is true, the mean outcome is the same for all levels of the column factor.

Assumptions

The standard two-way ANOVA hypothesis test are valid under the following conditions:

1. **The design must be complete.**
2. **The design must be balanced.**
3. **The number of replicates per treatment, *K*, must be at least 2.**
4. **Within each treatment, the observations are a simple random sample from a normal population.**
5. **The population variance is the same for all treatments. We denote this variation σ^2.**

Notation: SSA is the sum of squares for the rows. SSB is the sum of squares for the column. The interaction sum of squares is SSAB, and the error sum of squares is SSE. The sum of all of these is the total sum of squares (SST).

ANOVA Table

Source	Degrees of Freedom	Sum of Squares
Rows (SSA)	$I-1$	$JK\sum_{i=1}^{I}\hat{\alpha}_i^2 = JK\sum_{i=1}^{I}\bar{X}_{i..}{}^2 - IJK\bar{X}_{...}{}^2$
Columns (SSB)	$J-1$	$IK\sum_{j=1}^{J}\hat{\beta}_j^2 = IK\sum_{j=1}^{J}\bar{X}_{.j.}{}^2 - IJK\bar{X}_{...}{}^2$
Interactions ($SSAB$)	$(I-1)(J-1)$	$K\sum_{i=1}^{I}\sum_{j=1}^{J}\hat{\gamma}_{ij}^2$ $= K\sum_{i=1}^{I}\sum_{j=1}^{J}\bar{X}_{ij.}{}^2 - JK\sum_{i=1}^{I}\bar{X}_{i..}{}^2 - IK\sum_{j=1}^{J}\bar{X}_{.j.}{}^2 + IJK\bar{X}_{...}{}^2$
Error (SSE)	$IJ(K-1)$	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(X_{ijk}-\bar{X}_{ij.})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - K\sum_{i=1}^{I}\sum_{j=1}^{J}\bar{X}_{ij.}{}^2$
Total (SST)	$IJK-1$	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(X_{ijk}-\bar{X}_{...})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}X_{ijk}^2 - IJK\bar{X}_{...}{}^2$

- **Note that SST = SSA + SSB + SSAB + SSE**

Mean Square Errors

- The mean square error for rows is MSA = SSA / (I – 1).
- The mean square error for columns is MSB = SSB / (J – 1).
- The mean square error for interaction is MSAB = SSAB / ((I – 1)(J – 1)).
- The mean square error is MSE = SSE / (IJ(K – 1)).
- The test statistics for the three null hypotheses are quotients of MSA, MSB, and MSAB with MSE.

Test Statistics

- **Under H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_I = 0$, the statistic MSA / MSE has an $F_{I-1,\, IJ(K-1)}$ distribution.**

- **Under H_0: $\beta_1 = \beta_2 = \ldots = \beta_I = 0$, the statistic MSB / MSE has an $F_{J-1,\, IJ(K-1)}$ distribution.**

- **Under H_0: $\gamma_{11} = \gamma_{12} = \ldots = \gamma_{IJ} = 0$, the statistic MSAB / MSE has an $F_{(I-1)(J-1),\, IJ(K-1)}$ distribution.**

Output

Two-way ANOVA: Yield versus Catalyst, Reagent

Source	DF	SS	MS	F	P
Catalyst	3	877.56	292.521	9.36	0.000
Reagent	2	327.14	163.570	5.23	0.010
Interaction	6	156.98	26.164	0.84	0.550
Error	36	1125.33	31.259		
Total	47	2487.02			

S = 5.591 R-sq = 54.75% R-Sq(adj) = 40.93%

- **The labels DF, SS, MS, F, and P refer to degrees of freedom, sum of squares, mean square, *F* statistic, and *P*-value, respectively.**

- **The MSE is an estimate of the error variance.**

Example 2 cont.

Determine whether the additive model is plausible for the yield data. If the additive model is plausible, can we conclude that either the catalyst or the reagent affects the yield?

Answer: Additive model works if the interactions are not significant. In our case, the interaction p-value = 0 832> 0.05, so we do not reject H_0: $\gamma_{11} = \gamma_{12} = \ldots = \gamma_{IJ} = 0$, and conclude that we can assume that the additive model works.

Now we can deal with catalyst and reagent effects. The p-values for both catalyst and reagent are smaller than 0.05, so we reject the null hypotheses:

Ho: means for all rows are the same (no catalyst effect)

Ho: means for all columns are the same (no reagent effect)

Conclusion: Both change in catalyst and change in reagent affect the yield.

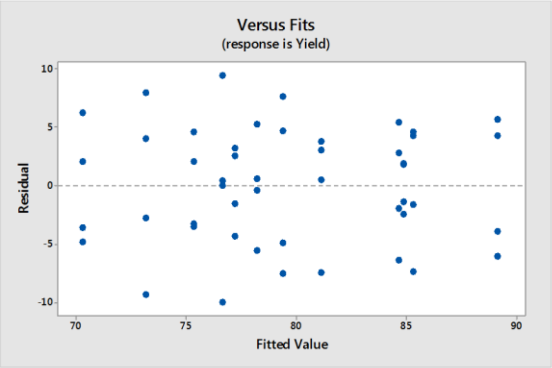
Checking the Assumptions

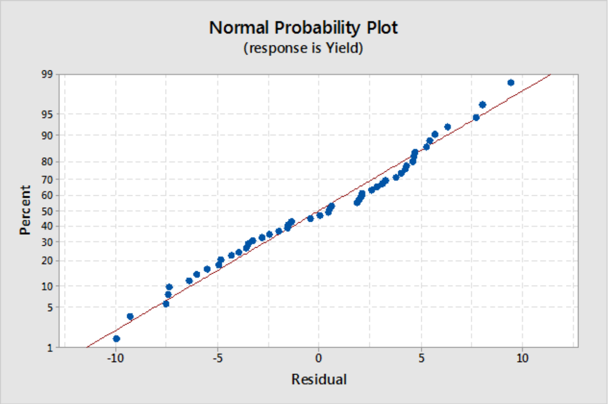
- **A residual plot is used to check the assumptions of equal variance.**

  We plot the residuals versus fitted values.

  residuals = individual obs of the response – average of the col and row factor obs

- **A normal plot of the residuals is used to check normality.**





Both plots support the corresponding assumptions.

Comments

- **In a two-way analysis of variance, if the additive model *is not* rejected, then the hypothesis tests for the main effects can be used to determine whether the row and column factors affect the outcome.**

- **In a two-way analysis of variance, if the additive model *is* rejected, then the hypothesis tests for the main effects should not be used. Instead, the cell means must be examined to determine how various combinations of row and column levels affect the outcome.**

- **When there are two factors, a two factor design must be used. Examining one factor at a time cannot reveal interactions between the factors.**

Example 3

The thickness of the silicon dioxide layer on a semiconductor wafer is crucial to its performance. Oxide layer thicknesses were measured for three types of wafers. In addition, several furnace locations were used to grow the oxide layer. A two-way ANOVA for three runs at one wafer site for the three types of wafers at three furnace locations were performed.

Two-way ANOVA for Thickness versus Wafer, Location

Source	DF	SS	MS	F	P
Wafer	2	5.8756	2.9378	2.07	0.155
Location	2	4.1089	2.0544	1.45	0.262
Interaction	4	21.349	5.3372	3.76	0.022
Error	18	25.573	1.4207		
Total	26	56.907			

Example 3 cont.

Since recycled wafers are cheaper, the company hopes that there is no difference in the oxide layer thickness among the three types of wafers. If possible, determine whether the data is consistent with the hypothesis of no difference. If not possible, explain why not.

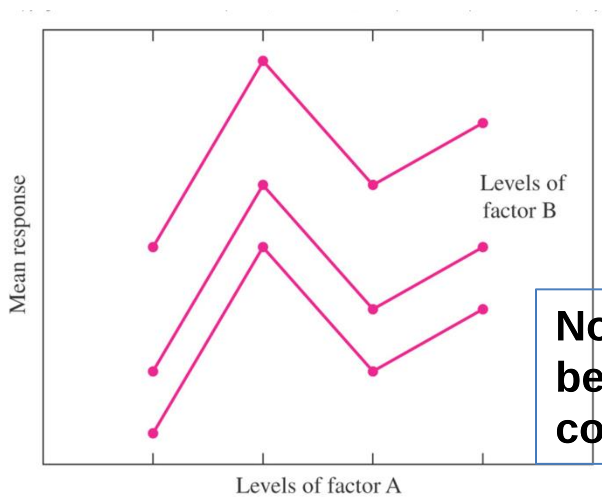
Solution: **Note that the p-value for the interactions is 0.02 < 0.05, so we reject**

H: no interactions.

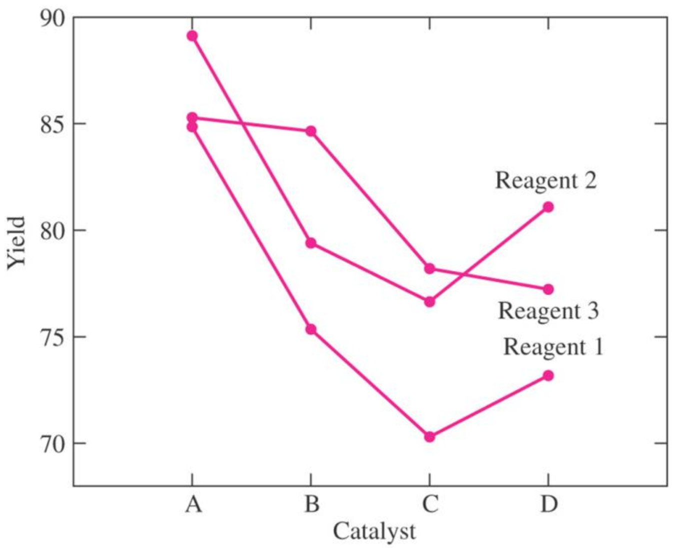
Thus, we have interaction between the wafers and location. We can not interpret/use main effects. The next step is to compare means for individual cells. We see that thicknesses vary among wafers, but no particular kind of wafer produced consistently thicker oxide layers.

Furnace	Wafer Type			
Location	Virgin	In-House	External	Row Mean
1	90.067	89.933	91.967	90.656
2	90.067	91.233	88.633	89.978
3	88.300	90.533	90.367	89.733
Column Mean	89.478	90.566	90.322	

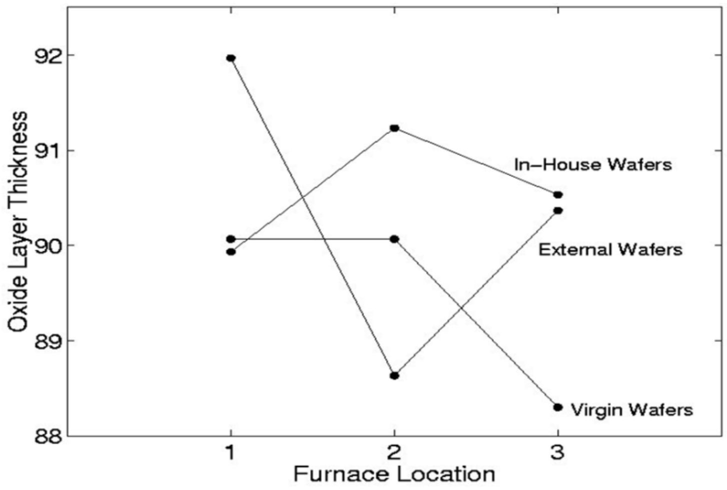
Interaction plots – help visualize the interactions.



No interaction between row and column factors



Yield example: not significant interaction



Oxide layer example: significant interaction

Tukey's Method for Simultaneous Confidence Intervals

- **Let I be the number of levels of the row factor, J be the number of levels of the column factor, and K be the sample size for each treatment. Then, if the additive model is plausible, the Tukey 100(1 − α)% simultaneous confidence intervals for all differences $\alpha_i - \alpha_j$ (or all differences $\beta_i - \beta_j$) are**

$$\hat{\alpha}_i - \hat{\alpha}_j \pm q_{I,IJ(K-1),\alpha}\sqrt{\frac{MSE}{JK}}$$

$$\hat{\beta}_i - \hat{\beta}_j \pm q_{J,IJ(K-1),\alpha}\sqrt{\frac{MSE}{IK}}$$

- **We are 100(1 − α)% confident that the Tukey confidence intervals contain the true value of the difference $\alpha_i - \alpha_j$ (or $\beta_i - \beta_j$) for every i and j.**

Example 2 cont.

If appropriate, use Tukey's method to determine which pairs of catalysts and which pairs of reagents can be concluded to differ, at the 5% level, in their effect on yield.

Two-way ANOVA: Yield versus Catalyst, Reagent

Source	DF	SS	MS	F	P
Catalyst	3	877.56	292.521	9.36	0.000
Reagent	2	327.14	163.570	5.23	0.010
Interaction	6	156.98	26.164	0.84	0.550
Error	36	1125.33	31.259		
Total	47	2487.02			

S = 5.591 R-sq = 54.75% R-Sq(adj) = 40.93%

Interaction not significant, so we can use Tukey's method for 95% CIs for differences between row effects and differences between column effects. (discussion)

Comments

- **The *F* tests that we have discussed require the assumption that the sample size *K* for each treatment be at least 2. When $K = 1$, a two-way ANOVA cannot be performed unless it is certain that the additive model holds.**

- **We focused on experiments where both factors were fixed. The methods of analysis and the null hypothesis differ among the fixed effect models, random effects model, and mixed effect models.**

- **The methods we discussed do not apply to unbalanced designs.**