Analysis of Variance

UCR GradQuant Workshop

11/4/2014

Roadmap

- Introduction to Analysis of Variance
- One-Way ANOVA
- Multiple Comparison Procedures
- Two-Way ANOVA

Introduction to Analysis of Variance

- Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means.

- The name is rather than “Analysis of Means” because Inferences about means are made by analyzing variance.

- We want to use the sample results to test the following hypotheses (k groups):

H_0: «$_1$ = «$_2$ = «$_3$ = · · · = «$_k$

H_a: Not all population means are equal

Introduction to Analysis of Variance

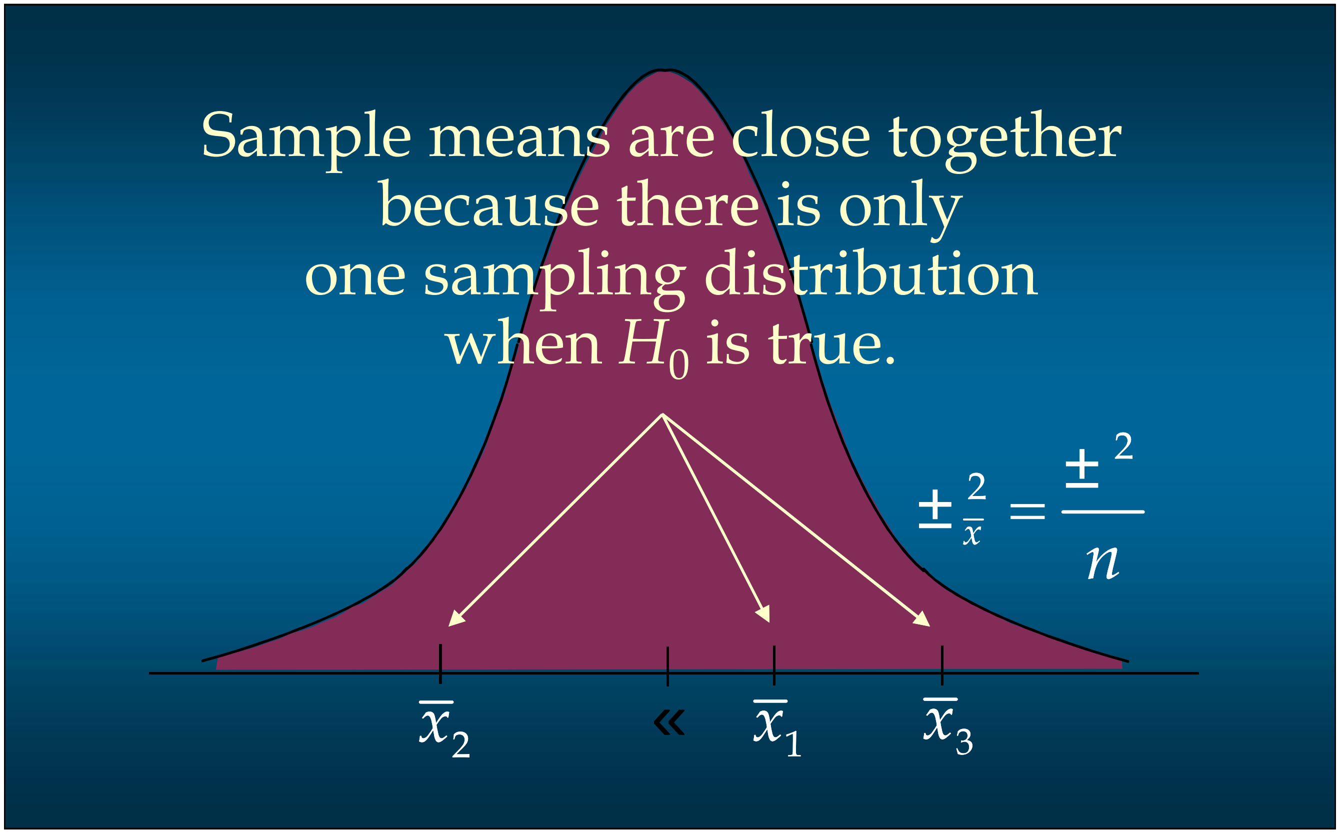
H_0: «$_1$ = «$_2$ = «$_3$ = · · · = «$_k$

H_a: Not all population means are equal

- If H_0 is rejected, we cannot conclude that *all* population means are different.
- Rejecting H_0 means that at least two population means have different values.

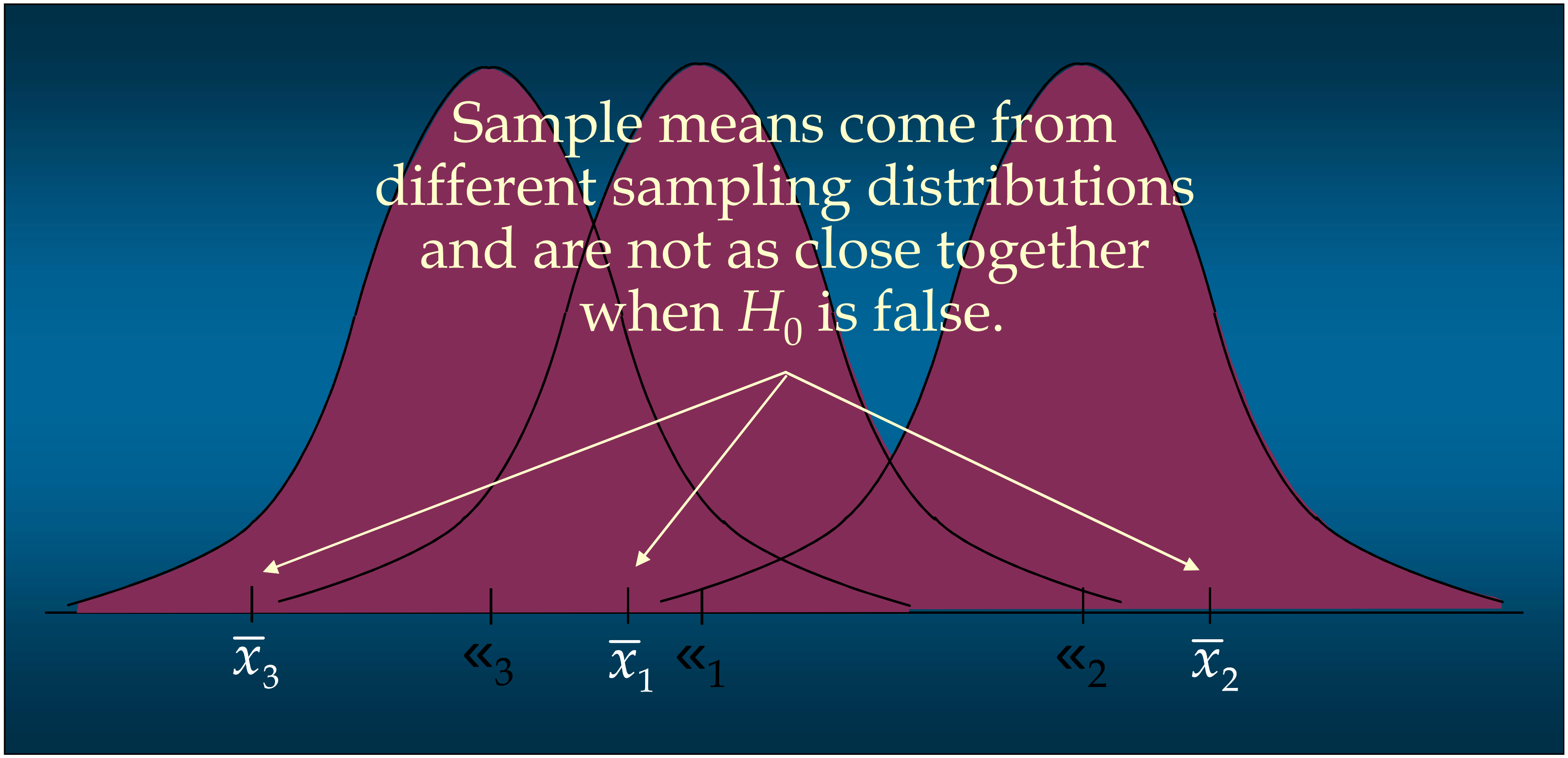
Introduction to Analysis of Variance

- Sampling Distribution of $\overline{x}$ Given H_0 is True



Introduction to Analysis of Variance

- Sampling Distribution of $\bar{x}$ Given H_0 is False



Assumptions for Analysis of Variance

- For each population, the response variable is normally distributed.
- The variance of the response variable, denoted $\pm\ ^2$, is the same for all of the populations.
- The observations must be independent.

One-Way ANOVA

- The test is based on two estimates of the population variance (σ^2)

  $\bar{x}_j = (\sum_{i=1}^{n_j} x_{ji})/n_j$ sample mean at the jth group

  $\bar{\bar{x}} = (\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ji})/n_T$ grand average

  $S_j^2 = [\sum_{i=1}^{n_j}(x_{ji} - \bar{x}_j)^2]/(n_j-1)$ sample variance at the jth group

Within-Samples Estimate of Population Variance

- The estimate of ± ² based on the variation of the sample observations within each sample is called the <u>mean square error</u> and is denoted by MSE.

$$\text{MSE} = \frac{\sum_{j=1}^{k}(n_j - 1)s_j^2}{n_T - k}$$

Denominator represents the <u>degrees of freedom</u> associated with SSE

Numerator is the <u>sum of squares due to error</u> and is denoted SSE

Between-Treatments Estimate of Population Variance

- A between-treatment estimate of $\pm^{2}$ is called the <u>mean square treatment</u> and is denoted MSTR.

$$\text{MSTR} = \frac{\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2}{k-1}$$

Denominator represents the <u>degrees of freedom</u> associated with SSTR

Numerator is the <u>sum of squares due to treatments</u> and is denoted SSTR

Comparing the Variance Estimates: The *F* Test

- MSTR only estimates σ^2 if the population means are equal. If population means are not equal, MSTR estimates a quantity larger than σ^2.
- If the null hypothesis is true and the ANOVA assumptions are valid, the sampling distribution of MSTR/MSE is an *F* distribution with MSTR d.f. equal to k - 1 and MSE d.f. equal to n_T - k.
- If the means of the k populations are not equal, the value of MSTR/MSE will be inflated because MSTR overestimates $\pm^{2}$.
- Hence, we will reject H_0 if the resulting value of MSTR/MSE appears to be too large to have been selected at random from the appropriate *F* distribution.

Test for the Equality of *k* Population Means

- Hypotheses

H_0: «$_1$ = «$_2$ = «$_3$ = · · · = «$_k$

H_a: Not all population means are equal

- Test Statistic

$$F = \text{MSTR}/\text{MSE}$$

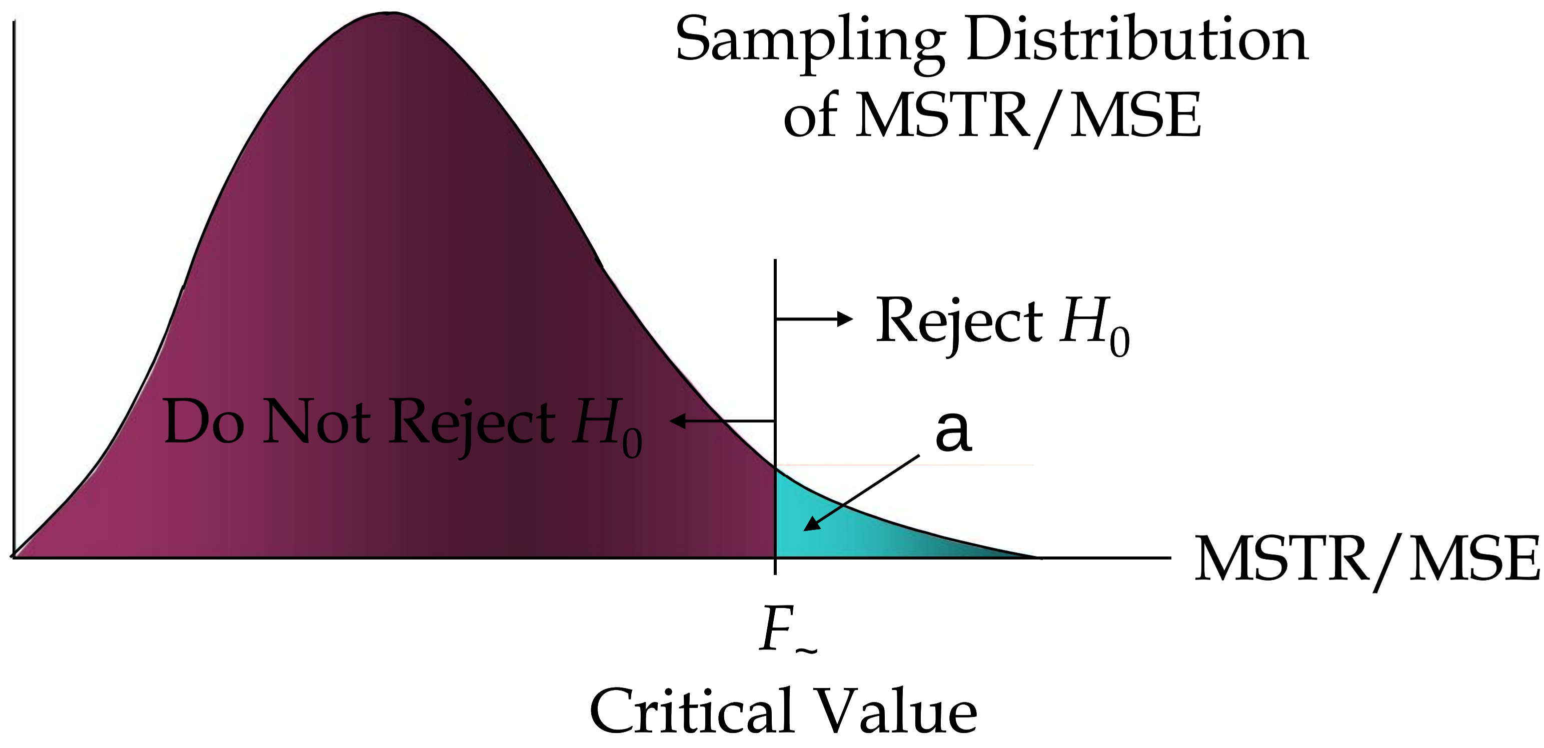
Test for the Equality of k Population Means

- Rejection Rule
  - p-value Approach: Reject H_0 if p-value $\leq$ a
  - Critical Value Approach: Reject H_0 if $F \geq F_a$

    where the value of $F_{\sim:}$ is based on an
    F distribution with k - 1 numerator d.f.
    and n_T - k denominator d.f.

Sampling Distribution of MSTR/MSE

- Rejection Region



ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F
Treatment	SSTR	$k - 1$	MSTR	MSTR/MSE
Error	SSE	$n_T - k$	MSE	
Total	SST	$n_T - 1$		

SST is partitioned into SSTR and SSE.

SST's degrees of freedom (d.f.) are partitioned into SSTR's d.f. and SSE's d.f.

ANOVA Table

- SST divided by its degrees of freedom n_T – 1 is the overall sample variance that would be obtained if we treated the entire set of observations as one data set.

- With the entire data set as one sample, the formula for computing the total sum of squares, SST, is:

$$\mathrm{SST} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{\bar{x}})^2 = \mathrm{SSTR} + \mathrm{SSE}$$

ANOVA Table

- ANOVA can be viewed as the process of partitioning the total sum of squares and the degrees of freedom into their corresponding sources: treatments and error.

- Dividing the sum of squares by the appropriate degrees of freedom provides the variance estimates and the F value used to test the hypothesis of equal population means.

Test for the Equality of k Population Means

- Example: Reed Manufacturing

A simple random sample of five managers from each of the three plants was taken and the number of hours worked by each manager for the previous week is shown on the next slide.

Conduct an F test using $\alpha = .05$.

Test for the Equality of k Population Means

Observation	Plant 1 Buffalo	Plant 2 Pittsburgh	Plant 3 Detroit
1	48	73	51
2	54	63	63
3	57	66	61
4	54	64	54
5	62	74	56
Sample Mean	55	68	57
Sample Variance	26.0	26.5	24.5

Test for the Equality of *k* Population Means

- *p* -Value and Critical Value Approaches

1. Develop the hypotheses.

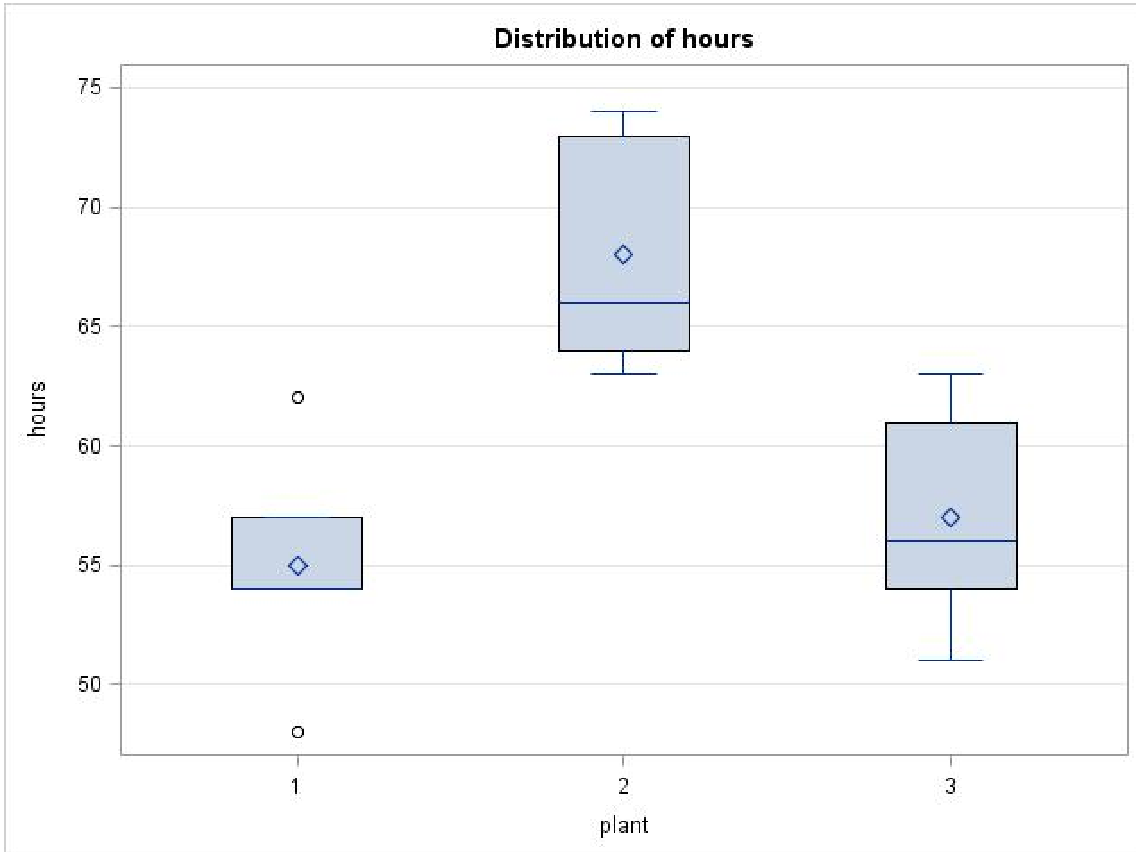
H_0: «$_1$ = «$_2$ = «$_3$
H_a: Not all the means are equal

where:

«$_1$ = mean number of hours worked per week by the managers at Plant 1
«$_2$ = mean number of hours worked per week by the managers at Plant 2
«$_3$ = mean number of hours worked per week by the managers at Plant 3

Test for the Equality of k Population Means

- *Box plot of sample means*



UCR GradSuccess
graduate.ucr.edu/success

Test for the Equality of k Population Means

- p -Value and Critical Value Approaches
  - 2. Specify the level of significance. $\alpha = .05$
  - 3. Compute the value of the test statistic.
    - <u>Mean Square Due to Treatments</u>

      (Sample sizes are all equal.)

      $\bar{\bar{x}} = (55 + 68 + 57)/3 = 60$

      $SSTR = 5(55 - 60)^2 + 5(68 - 60)^2 + 5(57 - 60)^2 = 490$

      $MSTR = 490/(3 - 1) = 245$

Test for the Equality of k Population Means

- p -Value and Critical Value Approaches

  - 3. Compute the value of the test statistic. (continued)

    - Mean Square Due to Error

$$SSE = 4(26.0) + 4(26.5) + 4(24.5) = 308$$

$$MSE = 308/(15 - 3) = 25.667$$

$$F = MSTR/MSE = 245/25.667 = 9.55$$

Test for the Equality of k Population Means

- ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F
Treatment	490	2	245	9.55
Error	308	12	25.667	
Total	798	14		

Test for the Equality of k Population Means

- p –Value Approach

4. Compute the p –value.

   With 2 numerator d.f. and 12 denominator d.f., the p-value is .01 for F = 6.93. Therefore, the p-value is less than .01 for F = 9.55.

5. Determine whether to reject H_0.

   The p-value $\leq$ .05, so we reject H_0.

   We have sufficient evidence to conclude that the mean number of hours worked per week by department managers is not the same at all 3 plant.

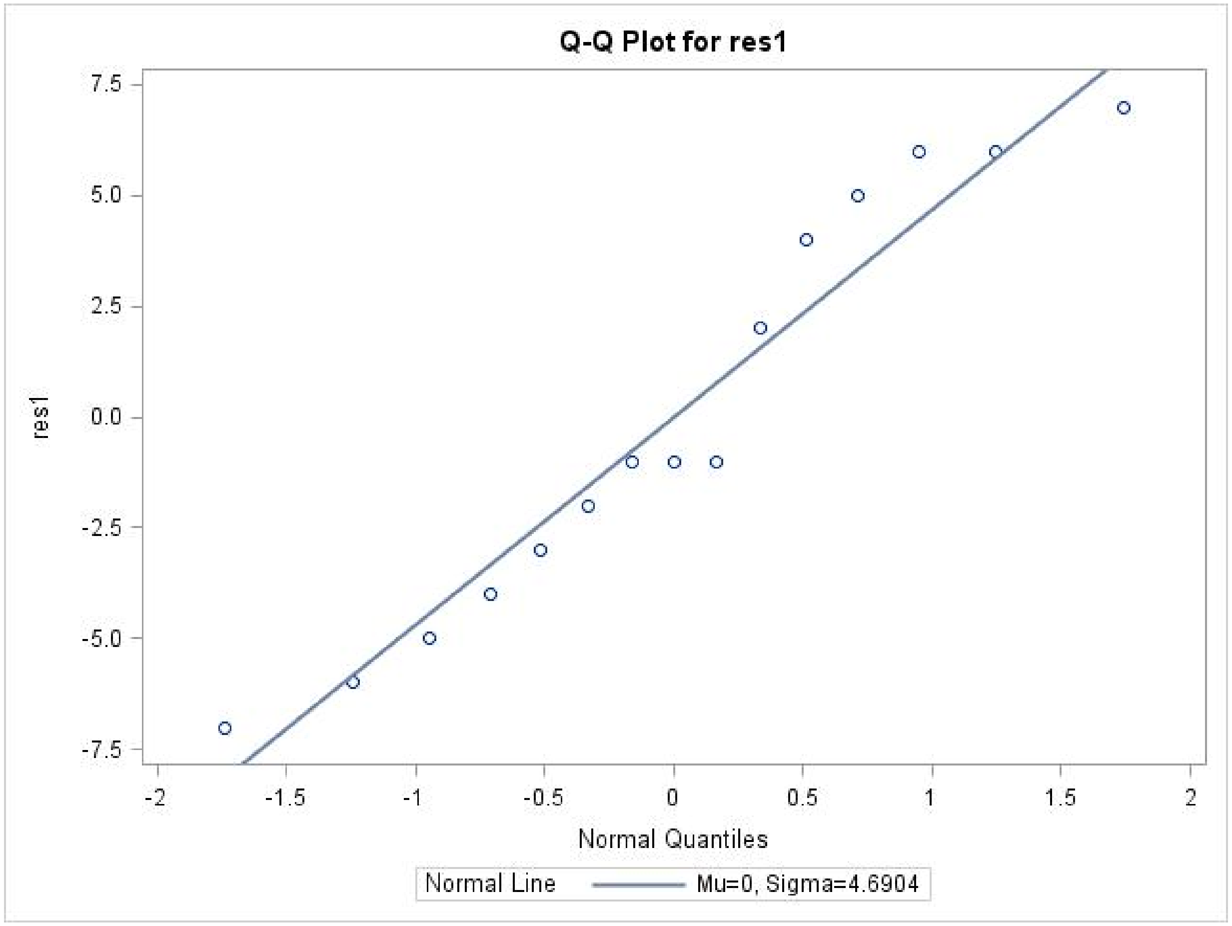
Test for the Model Assumptions

- *Normality assumption: Q-Q plot*

Observation	Plant 1 Buffalo	Plant 2 Pittsburgh	Plant 3 Detroit
1	48 (-7)	73 (5)	51 (-6)
2	54 (-1)	63 (-5)	63 (6)
3	57 (2)	66 (-2)	61 (4)
4	54 (-1)	64 (-4)	54 (-3)
5	62 (7)	74 (6)	56 (-1)
Sample Mean	55	68	57
Sample Variance	26.0	26.5	24.5

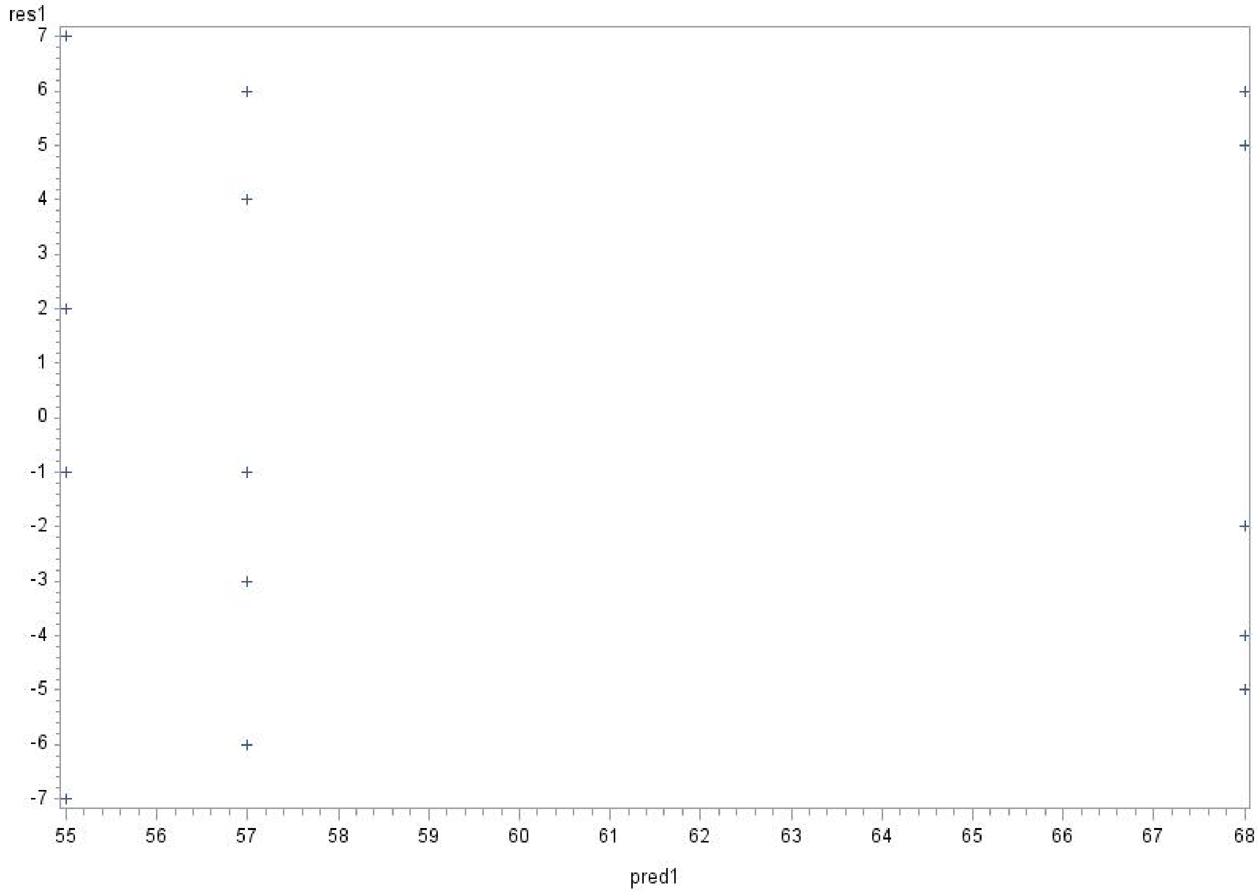
Test for the Model Assumptions

- *Normality assumption: Q-Q plot*



Test for the Model Assumptions

- *Equal variance assumption*



Nonnormal Responses and Transformations

- Square root transformation: $y_{ij}^{*} = \sqrt{y_{ij}}$
  Observations are poisson distributed
- Logarithmic transformation: $y_{ij}^{*} = log y_{ij}$
  Observations are lognormal distributed
- Box-Cox transformation

$$y^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\gamma & \lambda\, not = 0 \\ \log(y) & \lambda = 0 \end{cases}$$

Find value of λ that minimizes SSE(λ).

Nonnormal Responses and Transformations

- An example:

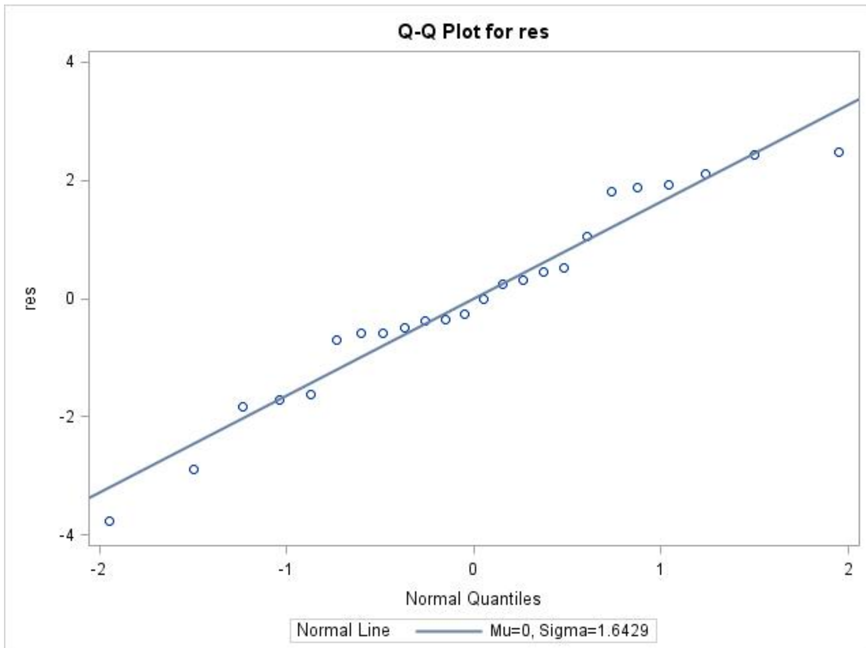
  A civil engineer is interested in determining whether four different methods of estimating flood flow frequency produce equivalent estimates of peak discharge when applied to the same watershed.

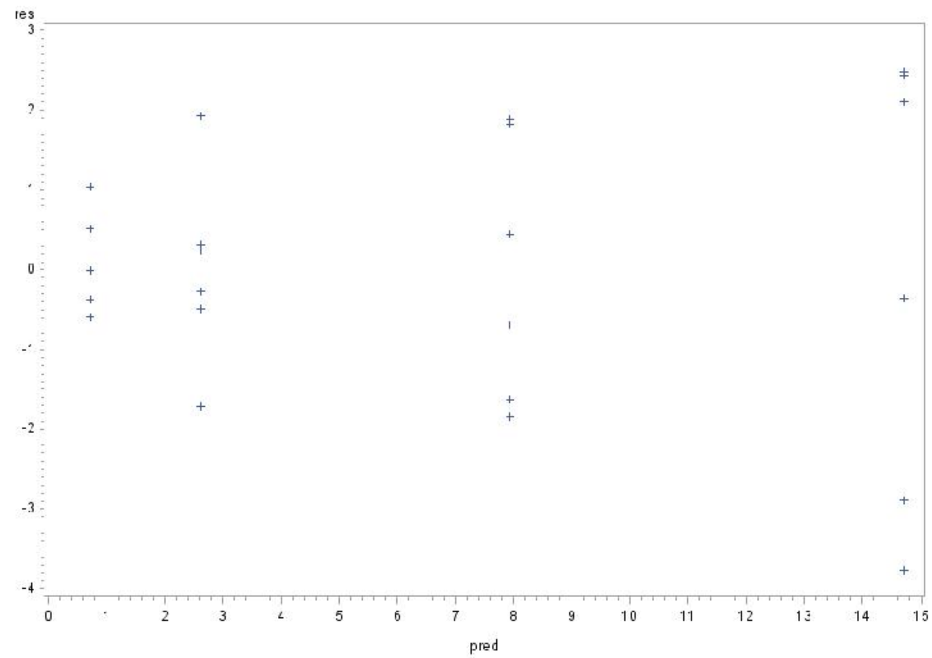
Nonnormal Responses and Transformations

Method	Observations					
1	0.34	0.12	1.23	0.70	1.75	0.12
2	0.91	2.94	2.14	2.36	2.86	4.55
3	6.31	8.37	9.75	6.09	9.82	7.24
4	17.15	11.82	10.95	17.20	14.35	16.82

Nonnormal Responses and Transformations

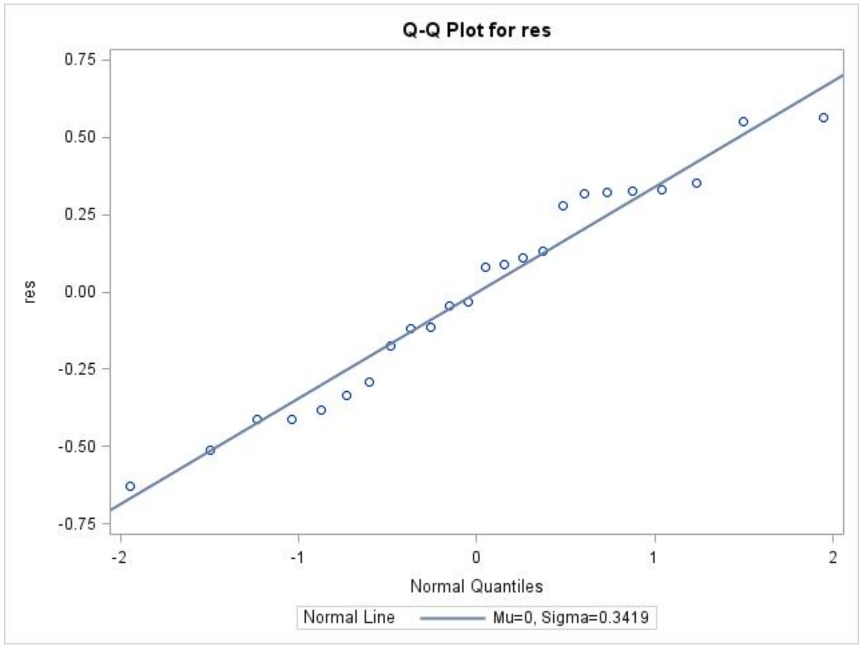
- Before transformation

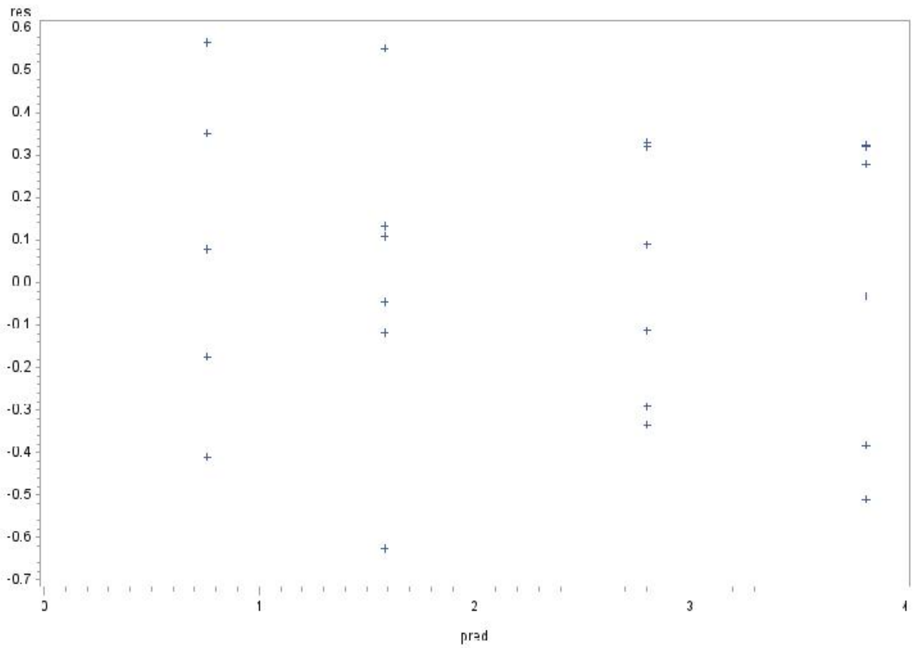




Nonnormal Responses and Transformations

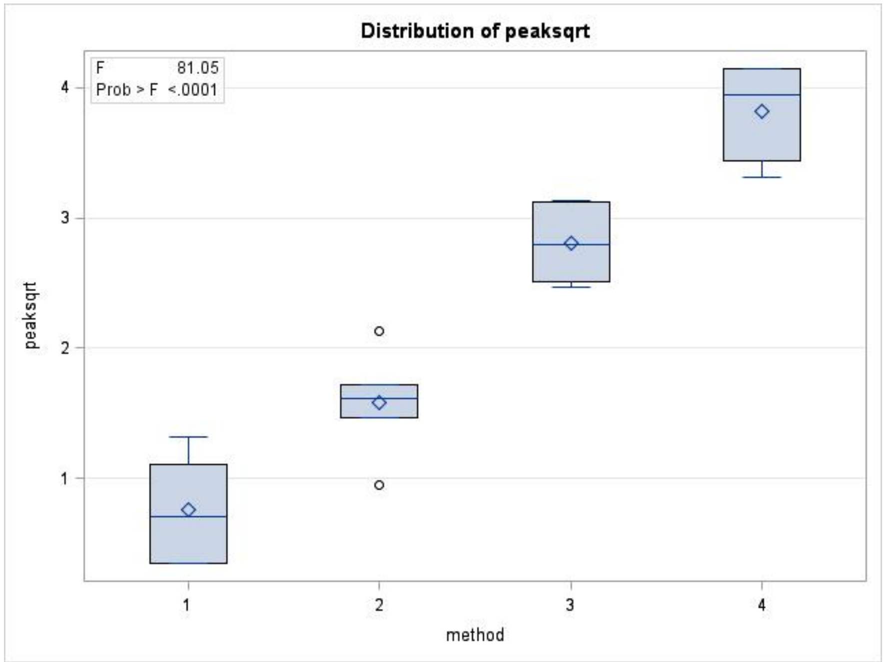
- After transformation: square root





UCR GradSuccess graduate.ucr.edu/success

Nonnormal Responses and Transformations



Source	DF	Type III SS	Mean Square	F Value	Pr > F
method	3	32.68421267	10.89473756	81.05	<.0001

post hoc Tests

- ANOVA compares all individual mean differences simultaneously, in one test
- A significant *F*-ratio indicates that at least one difference in means is statistically significant
  - *Does not* indicate which means differ significantly from each other!
- *post hoc* tests are follow up tests done to determine exactly which mean differences are significant, and which are not

Type I error rate

- The <u>comparisonwise Type I error rate</u> α indicates the level of significance associated with a single pairwise comparison.
- The <u>experimentwise Type I error rate</u> α_{EW} is the probability of making a Type I error on at least one of the $k(k-1)/2$ pairwise comparisons.

$$\alpha_{EW} = 1 - (1 - \alpha)^{k(k-1)/2}$$

- The experimentwise Type I error rate gets larger for problems with more populations (larger *k*).

Multiple Comparisons

- Tukey's Test: test all pairwise mean comparisons

$$HSD = q_{\alpha}(k, f)\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

- Back to Reed manufacturing example

$$HSD = 3.77\sqrt{\frac{25.6667}{5}} = 8.542$$

$$\text{plant 1 vs. 2}: 55 - 68 = -13^*$$

$$\text{plant 1 vs. 3}: 55 - 57 = 2$$

$$\text{plant 2 vs. 3}: 68 - 57 = 11^*$$

The starred values indicate the pairs of means are significantly different.

Multiple Comparisons

- Dunnett's Test: compare treatment means with a control

$$d = d_{\sim}(k-1, f)\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_a}\right)}$$

Two-Way ANOVA

- **It is more efficient to conduct one study that includes both independent variables.**
- **It allows a test of the *interaction* between the variables.**

Two-Way ANOVA

- **An example**:

  The researchers were interested in whether the weight of a companion of a job applicant would affect judgments of a male applicant's qualifications for a job. Two *independent variables* were investigated: (1) whether the companion was obese or of typical weight and (2) whether the companion was a girlfriend or just an acquaintance.

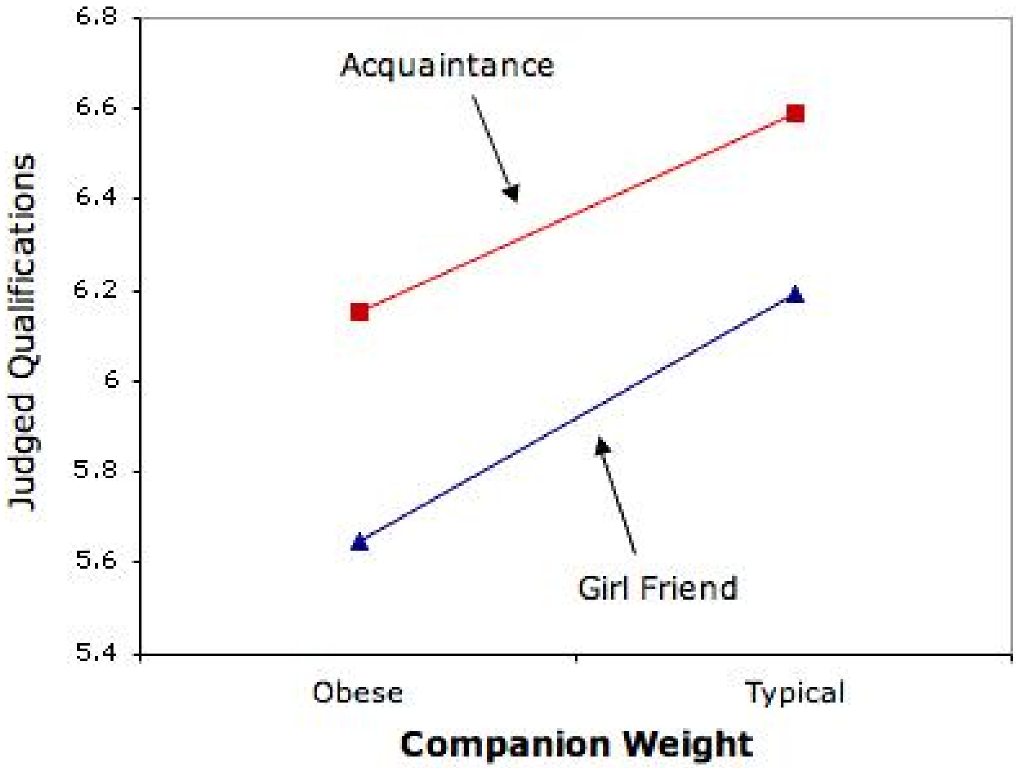
Two-Way ANOVA

- There are three effects of interest in this experiment:
  - Weight: Are applicants judged differently depending on the weight of their companion?
  - Relationship: Are applicants judged differently depending on their relationship with their companion?
  - Weight x Relationship Interaction: Does the effect of weight differ depending on the relationship with the companion?

Two-Way ANOVA

		Companion Weight		Marginal Mean
		Obese	Typical	
Relationship	Girlfriend	5.65	6.19	5.92
	Acquaintance	6.15	6.59	6.37
Marginal Mean		5.90	6.39	

Two-Way ANOVA



Two-Way ANOVA

- ANOVA Table

Source	df	SS	MS	F	p
Weight	1	10.4673	10.4673	6.214	0.0136
Relation	1	8.8144	8.8144	5.233	0.0234
W x R	1	0.1038	0.1038	0.062	0.8043
Error	172	289.7132	1.6844		
Total	175	310.1818			

Two-Way ANOVA

- **Another example**: Twelve subjects were selected from a population of high-self-esteem subjects and an additional 12 subjects were selected from a population of low-self-esteem subjects. Subjects then performed on a task and (independent of how well they really did) half in each esteem category were told they succeeded and the other half were told they failed. Therefore, there were six subjects in each of the four esteem/outcome combinations and 24 subjects in all. After the task, subjects were asked to rate (on a 10-point scale) how much of their outcome (success or failure) they attributed to themselves as opposed to being due to the nature of the task.

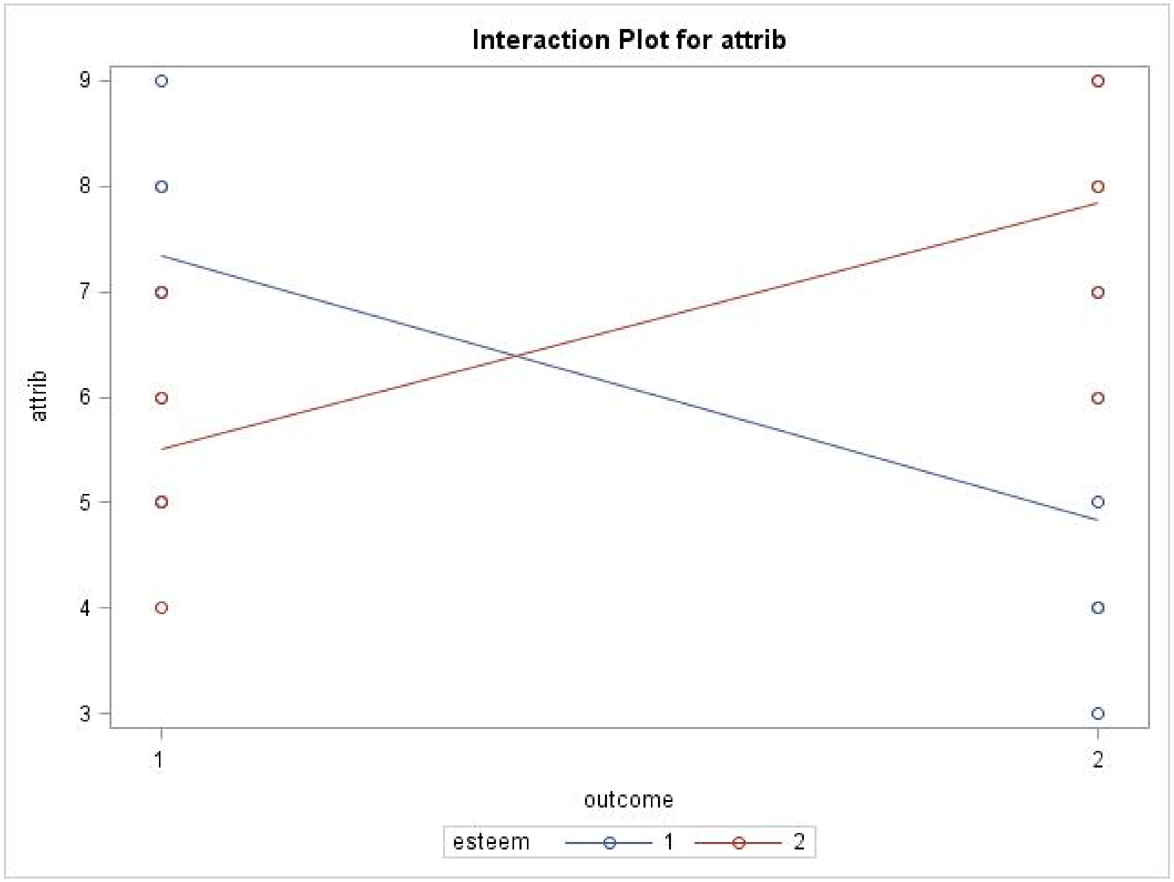
Two-Way ANOVA

		Esteem	
		High	Low
Outcome	Success	7	6
		8	5
		7	7
		8	4
		9	5
		5	6
	Failure	4	9
		6	8
		5	9
		4	8
		7	7
		3	6

Two-Way ANOVA

		Esteem		
		High	Low	Marginal Mean
Outcome	Success	7.333	5.500	6.467
	Failure	4.833	7.833	6.333
	Marginal Mean	6.083	6.667	

Two-Way ANOVA



High self-esteem:1
Low self-esteem: 2
Success: 1
Failure: 2

Nonparallel lines indicate interaction. The significance test for the interaction determines whether it is justified to conclude that the lines in the population are not parallel.

Two-Way ANOVA

- ANOVA Table

Source	df	SS	MS	F	p
Outcome	1	0.0417	0.0417	0.0256	0.8744
Esteem	1	2.0417	2.0417	1.2564	0.2756
O x E	1	35.0417	35.0417	21.5641	0.0002
Error	20	32.5000	1.6250		
Total	23	69.6250			

Unequal Sample Size

- An example:

Rep	Treatment A			Treatment B			Treatment C		
	Block 1	Block 2	Block 3	Block 1	Block 2	Block 3	Block 1	Block 2	Block 3
1	17	43	16	21	39	19	22	46	26
2	28	30		21	45	22	30		31
3	19	39		24	42	16	33		26
4	21	44		25	47		31		33
5	19	44							29
6									25

Unequal Sample Size

- Weighted and unweighted means

Rep	Treatment A			Treatment B			Treatment C		
	Block 1	Block 2	Block 3	Block 1	Block 2	Block 3	Block 1	Block 2	Block 3
1	17	43	16	21	39	19	22	46	26
2	28	30		21	45	22	30		31
3	19	39		24	42	16	33		26
4	21	44		25	47		31		33
5	19	44							29
6									25
Mean at each block	20.8	40	16	22.8	43.3	19	29	46	28.3
unweighted	25.6			28.3			34.4		
$(\sum_j \sum_k Y_{ijk})/n_i$	29.09091			29.18182			30.18182		

Unequal Sample Size

- Weighted and unweighted means
  - In the case to summarizing and comparing groups for one-way and balanced designs, weighted and unweighted means are equivalent.
  - In unbalanced designs with more than one effect, the weighted mean for a group might not accurately reflect the "typical" response for that group, since it does not take other effects into account. Unweighted means are more appropriate.
  - Statistical analysis programs use different terms for means that are computed controlling for other effects. SPSS calls them *estimated marginal means*, whereas SAS and SAS JMP call them *least squares means*.