ANalysis Of VAriance II

Dr Tom Ilvento
Department of Food and Resource Economics
University of Delaware

Overview

- Let's continue our journey through the ANOVA approach to data
  - Focus on Single Factor Models
  - Terms for the ANOVA Table
  - **R-square**
  - More single factor models
  - Strategies for **Multiple Comparisons**, including **Fisher's LSD**

What are the Sum of Squares called?

Terms Explained	Excel	JMP	SAS
SST - Sum of Squares Treatment	Between	Variable Name	Model
SSE - Sum of Squares Error	Within	Error	Error
SS_{Total} - Total Sum of Squares	Total	Total	Corrected Total
Factor Levels	Groups	Factors	Class

R-square

- R-square (R^2) is a measure of association
- Measures of Association reflect the relationship between two or more variables
- R^2 is a member of the class of measures of association called PRE measures – Proportion in the Reduction in Error
- It is based on fitting a model to the data, based on information from an independent variable (or set of variables) and comparing our model to a baseline model

R-square

- **The baseline model is the Grand Mean**

$$SS_{Total} = \sum_{i=1}^{n} (y_i - \overline{Y})^2$$

- **Our model is one that is based on knowledge of the Factors/treatments**

$$SST = \sum_{i=1}^{k} n_i(\overline{y}_i - \overline{Y})^2$$

- **R2 is a measure of the percent of the SS_{Total} that is due to the treatment**

$$R^2 = SST/SS_{Total}$$

R-square

- With R^2 we ask, "how much better do I understand the Response variable (dependent) by knowing something about the Factors/Treatments (independent variables)
- R^2 varies from 0 to 1
  - 0 means we explain nothing of the dependent variable
  - 1 means we explain it perfectly
- R^2 is a linear measure of association
- R^2 =
  - SST/SS_{Total}, or
  - $1 - SSE/SS_{Total}$

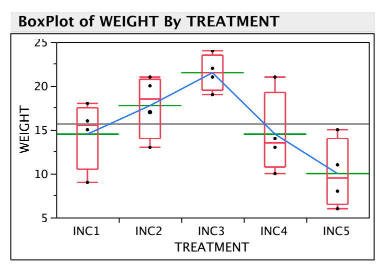
Another Problem

- An experiment is conducted to determine the differences in mean increases in plant growth from 5 different inoculums
- Inoculums are substances injected into a plant to fight disease.
- The experiment involved **20 cuttings** of a shrub (all of equal weight), with **4 cuttings** assigned to the five different inoculums
- The data represent the increase in weight in grams
- We will use α =.05

Incoculum Data

- Here is the way Excel would prefer the data
- We can add the means and variances
- And a box plot

I1	I2	I3	I4	I5
15	21	22	10	6
18	13	19	14	11
9	20	24	21	15
16	17	21	13	8



Excel results

- The results show that F* is 5.285 which has a p-value of .007
- Excel does not give us R-square, but it is easy to calculate:
  - **R^2 = 292.80/500.55 = .58496**
  - 58.5% of the variability in GROWTH is due to the type of inoculum
  - I can also solve R^2 as **1- 207.75/500.55 = .58496**

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
I1	4	58.00	14.50	15.00
I2	4	71.00	17.75	12.92
I3	4	86.00	21.50	4.33
I4	4	58.00	14.50	21.67
I5	4	40.00	10.00	15.33

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	292.80	4	73.200	5.285	0.007	3.056
Within Groups	207.75	15	13.850			
Total	500.55	19				

ANOVA Hypothesis Test for Incoculm Data

- **Ho:** **Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$**
- **Ha:** **Ha: At least two means are different**
- **Assumptions** **Equal variances, normal distribution**
- **Test Statistic** **$F^* = 5.285$ p =.007**
- **Rejection Region** **$F_{.05, 4, 15} = 3.056$**
- **Conclusion:** **$F^* > F_{.05, 4, 15}$**

or p = .007

Reject Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

Look at Output from JMP

- JMP shows
  - R-square
  - Adj R-square
  - Root Mean Square Error
  - Mean of Response (Grand Mean)
  - Number of observations
  - The ANOVA Table
  - Means and C.I.

Oneway Anova

Summary of Fit

Rsquare	0.5850
Adj Rsquare	0.4743
Root Mean Square Error	3.7216
Mean of Response	15.6500
Observations (or Sum Wgts)	20.0000

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
TREATMENT	4	292.80000	73.2000	5.2852	0.0074*
Error	15	207.75000	13.8500		
C. Total	19	500.55000			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
INC1	4	14.5000	1.8608	10.534	18.466
INC2	4	17.7500	1.8608	13.784	21.716
INC3	4	21.5000	1.8608	17.534	25.466
INC4	4	14.5000	1.8608	10.534	18.466
INC5	4	10.0000	1.8608	6.034	13.966

Std Error uses a pooled estimate of error variance

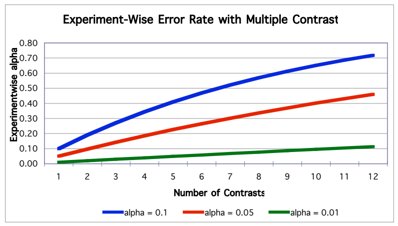
What's Next? Compare which means are different

- ANOVA just tests that at least two of the means are different
- ANOVA does not tell us which means are different
- **The next logical step is to ask which means are different from each other**
- We have five levels of the factor
- Resulting in 10 different comparisons of treatment means
  - 1 to 2; 1 to 3; 1 to 4; 1 to 5;
  - 2 to 3; 2 to 4; 2 to 5;
  - 3 to 4; 3 to 5;
  - 4 to 5

Difference of Means with Multiple Comparisons

- When we conduct a hypotheses test from a single experiment or sample, we set a level of Type I error for a comparison of two means.
- However, when we make many comparisons across treatments, the level of alpha increases in response to the number of comparisons.
- This is referred to as **Experiment-Wise Error Rate** (aka, family-wise error rate).
  - **$\alpha_e = 1-(1-\alpha)^c$**
  - where e is the experiment-wise error rate and c is the number of independent comparisons.

Number of Contrasts	Probability of a Type I Error on an Individual Test: 0.10	0.05	0.01
1	0.100	0.050	0.010
2	0.190	0.098	0.020
3	0.271	0.143	0.030
4	0.344	0.185	0.039
5	0.410	0.226	0.049
6	0.469	0.265	0.059
7	0.522	0.302	0.068
8	0.570	0.337	0.077
9	0.613	0.370	0.086
10	0.651	0.401	0.096
11	0.686	0.431	0.105
12	0.718	0.460	0.114



Fisher's Least Significant Difference (LSD)

- Fisher developed a strategy to deal with this issue using the concept of the Least Significant Difference (LSD).
- In this approach, an alpha rate is fixed and a least significant difference is calculated.
- Fisher's strategy was to develop a difference from which each comparison can be compared.
- **The difference between two means would need to be at least the size of the LSD_{ij} to be considered statistically significant.**
- The difference would take into account the experiment-wise error rate so that the researcher could be assured that for any comparison of two means, the overall level of alpha would be fixed at the desired level.

$$LSD_{ij} = t_{\alpha/2}\sqrt{s_w^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

$$LSD_{ij} = LSD = t_{\alpha/2}\sqrt{\frac{2s_w^2}{n}}$$

Fisher's LSD

$$LSD_{ij} = t_{\alpha/2}\sqrt{s_w^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

- α, the desired level of Type I error for each comparison. This level of is fixed by the LSD approach
- $t_{\alpha/2}$ a t-value associated with degrees of freedom error in the ANOVA table (set for a two tailed test in this example)
- s_w^2 the estimate of the pooled variance (MSE) from the ANOVA Table
- n_i the sample size for group i
- n_j the sample size for group j
- In the case where the sample size is the same for each group, we calculate a single LSD using

- **α = .05**
- **$t\alpha/2$, 15 d.f. = 2.131**
- **MSE = 13.85**
- **The n for all groups = 4**

$$LSD_{ij} = LSD = t_{\alpha/2}\sqrt{\frac{2s_w^2}{n}}$$

Fisher's LSD for Inoculum Data

- For any difference of means test of inoculums (INC1 to INC5), the difference must be at least 5.6078 to be significant at the .05 level.

$$LSD = 2.131\sqrt{\frac{2*13.85}{4}} = 5.6078$$

- Order means from lowest to highest:

INC5	INC1	INC4	INC2	INC3
10.0	14.5	14.5	17.75	21.50

- Compare the mean differences
- **INC5 to INC3 21.50 - 10.0 = 11.50 > LSD**
- **INC5 to INC2 17.75 - 10.0 = 7.75 > LSD**
- INC5 to INC4 14.5 - 10.0 = 4.5 < LSD
- **INC1 to INC3 21.50 - 14.50 = 7.00 > LSD**
- **INC4 to INC3 21.50 - 14.50 = 7.00 > LSD**
- INC1 to INC2 17.75 - 14.5 = 3.25 < LSD
- INC2 to INC3 21.50 - 17.75 = 3.75 < LSD

INC3 has the highest mean at 21.50.

INC3 is significantly different from INC5, INC1, and INC4.

INC2 is significantly different from INC5.

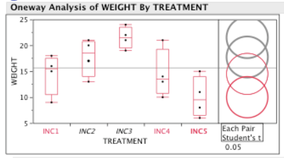
No other means were significantly different from each other.

All comparisons were significant at α=.05 controlling for multiple comparisons using Fisher's LSD.

Look at how JMP does this test

- The first matrix shows the difference minus the LSD
- Values that are positive show a difference that is significant
- The second table is also a popular way to show the same results
- Move down the columns to find significant differences

Oneway Analysis of WEIGHT By TREATMENT



Oneway Anova

Summary of Fit

Rsquare	0.584957
Adj Rsquare	0.474278
Root Mean Square Error	3.721559
Mean of Response	15.65
Observations (or Sum Wgts)	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
TREATMENT	4	292.80000	73.2000	5.2852	0.0074*
Error	15	207.75000	13.8500		
C. Total	19	500.55000			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
INC1	4	14.5000	1.8608	10.534	18.466
INC2	4	17.7500	1.8608	13.784	21.716
INC3	4	21.5000	1.8608	17.534	25.466
INC4	4	14.5000	1.8608	10.534	18.466
INC5	4	10.0000	1.8608	6.034	13.966

Std Error uses a pooled estimate of error variance

Means Comparisons

Comparisons for each pair using Student's t

Abs(Dif)-LSD

	INC3	INC2	INC1	INC4	INC5
INC3	-5.6090	-1.8590	1.3910	1.3910	5.8910
INC2	-1.8590	-5.6090	-2.3590	-2.3590	2.1410
INC1	1.3910	-2.3590	-5.6090	-5.6090	-1.1090
INC4	1.3910	-2.3590	-5.6090	-5.6090	-1.1090
INC5	5.8910	2.1410	-1.1090	-1.1090	-5.6090

Positive values show pairs of means that are significantly different.

Level				Mean
INC3	A			21.500000
INC2	A	B		17.750000
INC1		B	C	14.500000
INC4		B	C	14.500000
INC5			C	10.000000

Levels not connected by same letter are significantly different.

Experiment-Wise Error Rate

- **There are many other methods of comparison**
  - **Scheffe**
  - **Tukey**
  - **Bonferoni**
  - **Tukey-Kramer**
- **Most of the multiple comparison strategies use the following approach**

1. Fix alpha at some level
2. Adjust the comparisons to reflect an overall alpha
3. Compare the selected means (or all of them) using a difference of means test using a pooled variance
4. Many show the result in terms of a confidence interval
5. If the Confidence Interval overlaps with zero – there isn't a difference

Example for you to run and work on - ANOVA Golf.xls or ANOVA Golf.jmp

- The USGA wants to compare the mean distances of several brands of golf balls struck by a driver.
- They set up an experiment where a 10 balls are randomly picked from an allotment of four different brands of golf balls.
- To hold constant the effect of the golfer, they use a mechanical robotic golfer using the same driver.
- The distance the ball traveled is recorded as the response variable.
- Use an alpha level of .01.

Experimental Design

1 Factor: Golfball Brand

4 Treatments

10 replications per treatment

Experimental Unit: Golfball

Measurement Unit: Golfball

Total Sample Size: 40

Results from Excel

- This is the way Excel prefers the data
- Looking at the means, I see Ball C went the furthest on average, and Ball D the shortest
- The Variances are similar - no ratio greater than 2.2
- I used TOOLS, DATA ANALYSIS, ANOVA Single Factor to run the ANOVA
- **F^* = 43.989, $p < .001$**
- **R^2 = 2794.39/3556.69 = .7857**
- **78.6% of the variability in driving distance is due to the ball type**

	Ball A	Ball B	Ball C	Ball D
	251.2	263.2	269.7	251.6
	245.1	262.9	263.2	248.6
	248.0	265.0	277.5	249.4
	251.1	254.5	267.4	242.0
	260.5	264.3	270.5	246.5
	250.0	257.0	265.5	251.3
	253.9	262.8	270.7	261.8
	244.6	264.4	272.9	249.0
	254.6	260.6	275.6	247.1
	248.8	255.9	266.5	245.9
Mean	**251.00**	**261.63**	**270.33**	**249.70**
Variance	**24.68**	**13.12**	**21.14**	**28.83**

ANOVA Hypothesis Test for Golfball Data

Ho:	**Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$**
Ha:	**Ha: At least two means are different**
Assumptions	**Equal variances, normal distribution**
Test Statistic	**$F^* = 43.989 \quad p < .001$**
Rejection Region	**$F_{.01,\,3,\,36} = 4.377$**
Conclusion:	**$F^* > F_{.01,\,3,\,36}$**
	or $p < .001$
	Reject Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

Next: Which Golfballs are different from each other?

- We will use Fisher's LSD
  - There are 6 contrasts
  - $\alpha = .01$
  - MSE = 21.18
  - n = 10
  - $t_{.01/2,\ 36\ d.f.} = 2.719$
- Comparisons
- D to C 270.33 - 249.70 = 20.63 > LSD
- D to B 261.63 - 249.70 = 11.93 > LSD
- D to A 251.00 - 249.70 = 1.30 < LSD
- A to C 270.33 - 251.00 = 19.33 > LSD
- A to B 261.63 - 251,00 = 10.63 > LSD
- B to A 270.33 - 261.63 = 8.70 > LSD

$$LSD = 2.719\sqrt{\frac{2*21.18}{10}} = 5.59612$$

Ball D	Ball A	Ball B	Ball C
249.70	251.00	261.63	270.33

Ball C has the highest mean distance at 270.33.

Ball C is significantly different from Ball D, Ball A, and Ball B.

Ball B is significantly different from Ball A and Ball D

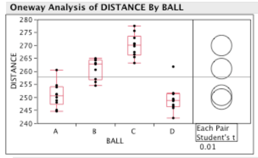
No other means were significantly different from each other.

All comparisons were significant at α=.01 controlling for multiple comparisons using Fisher's LSD.

Results from JMP

- JMP (any advanced software) gives a complete analysis
  - R-square
  - ANOVA and F*
  - Mean Comparisons
- Software like JMP would also
  - Test the assumption about equal variances
  - Different Mean comparisons

Oneway Analysis of DISTANCE By BALL



Oneway Anova

Summary of Fit

Rsquare	0.785671
Adj Rsquare	0.76781
Root Mean Square Error	4.601633
Mean of Response	257.7775
Observations (or Sum Wgts)	40

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
BALL	3	2794.3887	931.463	43.9887	<.0001*
Error	36	762.3010	21.175		
C. Total	39	3556.6897			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 99%	Upper 99%
A	10	250.780	1.4552	246.82	254.74
B	10	261.060	1.4552	257.10	265.02
C	10	269.950	1.4552	265.99	273.91
D	10	249.320	1.4552	245.36	253.28

Std Error uses a pooled estimate of error variance

Means Comparisons

Comparisons for each pair using Student's t

t	Alpha
2.71948	0.01

Abs(Dif)-LSD

	C	B	A	D
C	-5.596	3.294	13.574	15.034
B	3.294	-5.596	4.684	6.144
A	13.574	4.684	-5.596	-4.136
D	15.034	6.144	-4.136	-5.596

Positive values show pairs of means that are significantly different.

Level				Mean
C	A			269.95000
B		B		261.06000
A			C	250.78000
D			C	249.32000

Levels not connected by same letter are significantly different.

Summary

- We looked at some more single-factor problems and the way to look at the results
- We introduced R^2 as a measure of association, which shows us how much of the variability in the response variable is explained by the factor levels.
- After we establish some of the treatment means differ from each other, we want to know which means are different.
- To do this we use a of "Experiment or Family-wise error rate" to make multiple comparisons of differences of means.
- We introduced Fisher's LSD as a simple way to make multiple comparisons and control the overall "Experiment-Wise Error Rate."