Lect 15 STAT 102

- One Way Analysis of Variance (ANOVA)
  - **Read** Ch 9.1
- Comparison of means among I groups
- Individual t tests vs. multiple comparison:

  Two methods:

  Bonferroni and Tukey-Kramer

One-way **An**alysis **o**f **Va**riance

- One-way ANOVA: a technique designed to compare the means of two or more groups.
  - Extends the equal-variance two sample test discussed in Lecture 2
- Uses an F-test to determine whether there are – overall – any significant differences among the means
- Then (if there are any overall differences) uses special "multiple comparison" tests to determine which differences between pairs of means are significant.
- We'll discuss the theory in the context of an example.
- This example uses data printed in USA Today (~ 5 years ago) that reports the returns for prior years of a sample of Mutual Funds.

Stock Returns Example

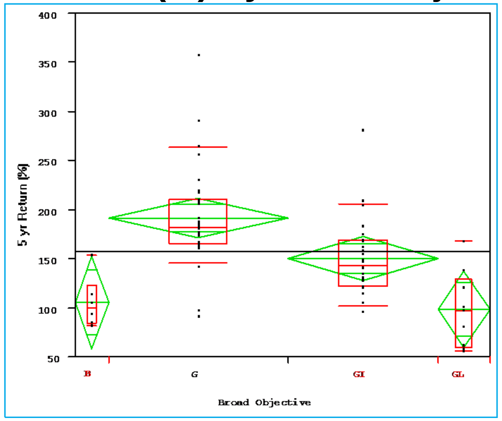
Look at the 5 yr. Returns in the USA Today stock fund data to see whether there are differences in 5 yr. Returns according to the Type of mutual fund.

In this data there are four main Types [*aka* "Broad Objectives"] and we will concentrate on these

B = Balanced, **GI** = Growth and Income,
G = Growth **GL** = Global

Here are side-by-side plots of the returns for the four major groups. This plot shows means diamonds and quantile box plots for each group. (The means diamonds are computed from the standard "assuming equal variance" analysis discussed below.)

5 yr Return (%) By Broad Objective



There are clearly noticeable differences among the returns. Overall, are they statistically significant? If so, which differences are significant?

Individual means

Here are the means and standard deviations for each group, and the SEs for the mean of each group as computed from the SD of that group.

Means and Std Deviations

Level	Number	Mean	Std Dev	S.E. Mean
B	6	106.2	26.23	10.71
G	31	192.6	51.07	9.17
GI	26	150.5	40.25	7.89
GL	9	98.44	38.94	12.98

One Way ANOVA (Theory)

Groups labeled $i = 1,\ldots,I$. Observations Y_{ij} in the i^{th} group, with $j = 1,\ldots,n_i$. $n = \sum n_i$ observations in all.

Model : $Y_{ij} = \mu_i + \varepsilon_{ij}$, where ε_{ij} indep. normal with mean=0 & var = σ^2

$$\mu_i = E(Y_{ij})$$

An alternate form of the model:

$$Y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

with $\mu = \frac{1}{I}\sum_i \mu_i$ and $\alpha_i = \mu_i - \mu$. (This implies that $\sum \alpha_i = 0$.)

Basic test is of

H_0: **All** the group means are the same - *ie*

$\mu_1 = \mu_2 = \ldots\ldots = \mu_I$ vs

H_a: They're not **all** the same.

Analysis of Variance

(Explanation of Calculations, Formulas, & Relation to Regression)

- The Model has $E\left(Y_{ij}\right) = \mu_i$.
- So, we'll estimate each μ_i by the mean of the corresponding $Y_{ij} : j = 1,..,n_i$; denoted by

$$\bar{Y}_{i\bullet} = n_i^{-1} \sum_j Y_{ij} \,.$$

- As with all our previous regression estimators, this is a Least Squares Estimator, *ie* it minimizes the total SSError:

$$SSE = \sum_{i,j} \left(Y_{ij} - \hat{Y}_{ij}\right)^2 = \sum (n_i - 1) s_i^2 = s_e^2 \text{where } \hat{Y}_{ij} = \bar{Y}_{i\bullet}$$

- As in other types of regression settings, this is compared to

$$SST = \sum (y_{ij} - \bar{Y}_{\bullet\bullet})^2$$

where $\bar{Y}_{\bullet\bullet}$ denotes the grand mean.

Test of $H_0 : \mu_1 = .. = \mu_I$

- We calculate the reduction in Sum of Squares due to the model:

SSR = SST – SSE.

- And use $F = \dfrac{SSR/(I-1)}{SSE/(n-I)} \sim F_{I-1,n-I}$ to test H_0.
- **Degrees of Freedom**: The DF for the model is ***I* – 1**.
- This is because under H_0 the value of μ_1 is not restricted, but then **the remaining** $\mu_2,..,\mu_I$ are completely restricted to be this same value. There are ***I* – 1** completely restricted values under H_0; and hence ***I* – 1** DF.
- The alternate form of the model has $\mu_i = \mu + \alpha_i$ **with** $\sum \alpha_i = 0$. Hence $H_0 : \alpha_1 = .. = \alpha_I = 0$ is the same as $H_0 : \alpha_1 = .. = \alpha_{I-1} = 0$. As before there are only ***I* – 1** completely restricted values under H_0; and hence ***I* – 1** DF for the Model.

Fund Ex: Test of H_0

JMP tables for the One-way ANOVA.

Summary of Fit

RSquare	0.39	
Root Mean Square Error	**44.44**	= s_{pooled} (or s_e)
Observations	72	

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	86631	**28877**	**14.62** = **28877**/**1975**
Error	68	134305	**1975**	Prob>F
C Total	71	220936	3112	<.0001

The F-Ratio has 3 and 68 **DF**, and tests

H_0: ***All*** the group means are the same.

Versus the alternative

H_a: They're not ***all*** the same.

Reject H_0 at 0.05 since the P-value is <.0001

Multiple Comparison Tests (Intro)

- Since we have rejected the null hypothesis that all the means are the same we would like to go on to investigate the differences between each pair.
- A first step could be to examine the estimates of the means, and their SE^s.

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
B	6	106.17	18.14	69.96	142.37
G	31	192.61	7.98	176.69	208.54
GI	26	150.50	8.72	133.11	167.89
GL	9	98.44	14.81	68.88	128.01

Std Error uses a pooled estimate of error variance

- Note that the SE^s in this table are not the same as those on our p.4 – Why not?

- If we construct **standard t-test** $100(1-\alpha)\%$ confidence intervals for each pair (*assuming equal-variances*), we have

$$\text{For } \mu_i - \mu_j\text{: } \overline{Y}_{i\bullet} - \overline{Y}_{j\bullet} \pm t_{n-I,1-\alpha/2}\sqrt{MSE}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\,.$$

- The DF here for the t-statistic is $n - I = 68$, since this calculation uses $\sqrt{MSE} = s_e$, having 68 DF.
- For example,

$$\mu_G - \mu_B = 192.6 - 106.2 \pm 2 \times 44.4 \times \sqrt{\frac{1}{31}+\frac{1}{6}}$$

$$= 86.4 \pm 39.6 = 46.8 \text{ to } 126.0$$

- Note that the probability is 1-α for EACH such intervals that it contains the true value of the corresponding μ_i - μ_j.

JMP has several ways of displaying the results of this construction of confidence intervals:
(Use the Fit Y by X platform and then go there to the arrow command "Compare means -Each pair, Student's t".)

Here is one way to display all the results from these CI's.
(This shows which CI's for $\mu_i - \mu_j$ contain 0.)

Means Comparisons for each pair using Student's t

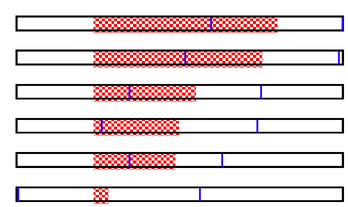
t	Alpha
1.995	0.05

Level				Mean
G	A			192.61
GI		B		150.50
B			C	106.17
GL			C	98.44

Levels not connected by same letter are significantly different

Here is another way. (This gives all the confidence intervals, both numerically and graphically – Note the blue lines (Lower CL and Upper CL) on the plot.)

Level	- Level	Difference	Lower CL	Upper CL	p-Value	Difference
G	GL	94.2	60.6	127.8	0.000	
G	B	86.5	46.9	126.0	0.000	
GI	GL	52.1	17.8	86.4	0.0035	
GI	B	44.3	4.2	84.5	0.0310	
G	GI	42.1	18.5	65.7	0.0007	
B	GL	7.7	-39.0	54.5	0.743	



P-Value is for Individual Tests of Difference = 0

- If we look at the above intervals **we might be tempted to claim** that we are 95% certain that

$60.59 \le \mu_G - \mu_{GL} \le 127.75$ **and** $46.89 \le \mu_G - \mu_B \le 126$ **and** $17.76 \le \mu_{GI} - \mu_{GL} \le 86.35$

and $4.17 \le \mu_{GI} - \mu_B \le 84.5$ **and**

$18.53 \le \mu_G - \mu_{GI} \le 65.70$ **and** $-39.02 \le \mu_B - \mu_{GL} \le 54.46$.

- This would be associated with a **claim** that we are 95% certain that **the first five** of these differences are **ALL ≠ 0**.

Such claim would be unjustified!

- What is true is that each ***individual*** confidence interval has a 95% chance of being right. But this implies that

there is a much smaller chance that
<u>all</u> of these confidence intervals are right.

- An issue here is the difference between an

Individual Coverage Rate

and a

Family-wise Coverage rate

Simultaneous Confidence Intervals

- When several confidence intervals are considered *simultaneously* they constitute a ***family*** of CIs
- **Individual Coverage Rate**: The probability that any **individual** confidence interval in the family contains its true value
- **Family-wise Coverage Rate**: The probability that **every** confidence interval contains its true value

Simultaneous Test Procedures

Every set of simultaneous confidence intervals is associated with a family of simultaneous tests. Thus we have the **family** of tests

$$H_{0;i,j} : \mu_i - \mu_j = 0,\ \ i \neq j\ ,$$

and we reject the individual $H_{0;i,j}$ whenever the confidence interval for $\mu_i - \mu_j$ does not contain 0.

- **Individual Error Rate**: The probability for a single test in the family that the corresponding null hypothesis will be rejected if it is true
- **Family-wise Error Rate**: The probability for the entire family of tests that at least one true null hypothesis will be rejected.

When planning and carrying out a study such as a one-way ANOVA the recommended Best Practice is to use procedures guaranteeing the claimed Family-wise coverage and error rates.

The easiest way to attain this is to use

Bonferroni Confidence Intervals and Tests

This general method works for one-way ANOVA and for many other statistical settings

For one-way ANOVA the

Tukey-Kramer Method

gives slightly more powerful tests and slightly shorter confidence intervals.

Bonferroni Method (for Tests and CIs)

- A general method for doing multiple tests (or confidence intervals, resp.) for any family of ***k*** tests (confidence intervals):

In the context of one-way ANOVA there are ***I*** groups, and hence

$$k = \binom{I}{2} = \frac{I(I-1)}{2}$$

- Denote the desired family-wise error rate by α^* (desired family-wise coverage rate by $1-\alpha^*$)
- Compute individual tests (confidence intervals) at level

$$\alpha = \alpha^* / k$$

(confidence intervals at individual coverage $1-\alpha^*/k$)

This guarantees the family-wise error rate is **at most** α^* (and the family-wise coverage rate is **at least** $1-\alpha^*$)

Why Bonferroni Works

In the general case there are ***k*** null hypotheses. Label them $H_{0;j},\ j = 1,\ldots,k$.

The probability that an individual type I error is made on $H_{0;j}$ is $P_{H_{0;j}}\left(rej\left\{H_{0;j}\right\}\right) = P\left(E_j\right)$, say, where E_j denotes the event of rejecting $H_{0;j}$ given that $H_{0;j}$ is true.

The probability that any error is made in the entire family of tests is

$$(*) \quad P\left(E_1 \cup \ldots \cup E_k\right) \le \sum_{1}^{k} P\left(E_j\right) = k\alpha = k\left(\alpha^* \big/ k\right) = \alpha^*$$

Thus the family-wise error rate is $\le \alpha^*$, as desired.

NOTE that the inequality in (*) is generally a strict inequality. Hence one should expect the Bonferroni procedure to have family-wise error rate strictly less than the nominal α^* - but it is hard to know how much less.

The proof for a family of confidence intervals is similar.

To Use Bonferroni with JMP in a one-way ANOVA

- Determine ***k*** via $k = \binom{I}{2} = \frac{I(I-1)}{2}$, where I denotes the number of comparison groups.
- Choose α^*. (Usually α^*=0.05). Calculate $\alpha = {\alpha^*}/{k}$.
- Go to the arrow menu inside the Fit Y by X platform. Select "Set alpha level →other" and enter the value of $\alpha = {\alpha^*}/{k}$. Then perform the individual "Compare means-Each pair, Student's t" as before.
- This will give the desired confidence intervals, (C_{ij}, say) and the corresponding tests of $H_{0;i,j}$ can be performed by rejecting whenever $0 \notin C_{ij}$.

Fund Example (cont): Bonferroni

In the example there are 4 groups to compare. So $k = \frac{4 \times 3}{2} = 6$.

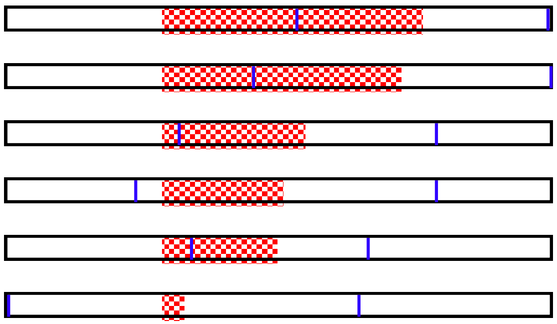
For $\alpha^* = 0.05$ we have $\alpha = \alpha^* / k = 0.05/6 = .00833$. We get the output:

Comparisons for each pair using Student's t

t	Alpha
2.718	0.00833

(Note that **critical t-value** here is 2.718, compared to the earlier t= 1.995 for α=0.05.)

Level	- Level	Diff'ce	Lower CL	Upper CL	Difference
G	GL	94.17	48.44	139.90	
G	B	86.45	32.58	140.32	
GI	GL	52.06	5.34	98.77	
GI	B	44.33	-10.37	99.04	
G	GI	42.11	10.00	74.23	
B	GL	7.72	-55.94	71.38	



HERE we can only reject **4** of the null hypotheses instead of **5** as with the individual t-tests procedure.

We also conclude Fund G is the best; better than all others.

Tukey-Kramer Method

- Note that Bonferroni uses simultaneous CIs of the form:

For $\mu_i - \mu_j$: $\overline{Y}_{i\bullet} - \overline{Y}_{j\bullet} \pm \mathbf{t}_{Bonf} \sqrt{MSE} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$

where $\mathbf{t}_{Bonf} = t_{n-I;1-(\alpha^*/k)/2}$.

- Tukey-Kramer uses CIs of the same form, but with a different (slightly smaller) value of **t.** Thus it has the form:

For $\mu_i - \mu_j$: $\overline{Y}_{i\bullet} - \overline{Y}_{j\bullet} \pm q^*_{T-K} \sqrt{MSE} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$

where q^*_{T-K} is specially chosen to give an error rate at most α^*.

- More precisely, the T-K procedure has family-wise error rate EXACTLY α^* when all n_i are the same. Otherwise it has error rate AT MOST α^*. (This was conjectured by Tukey & Kramer in the 50s and proven in the 70s by Hayter.)

- JMP performs the T-K procedure automatically. Use the command "Compare means-All Pairs, Tukey HSD". Be sure the Alpha Level is set at α^*, **and not at** $\alpha = \alpha^*/k$.

Example (cont) The T-K Method

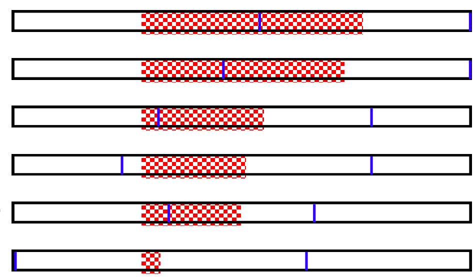
For $\alpha^* = 0.05$ we make sure the Alpha Level command is at 0.05, and then request the T-K output. We get:

Comparisons for all pairs using Tukey-Kramer HSD

q*	Alpha
2.63 (t_{BON}=2.718)	0.05

NOTE that this q* is slightly less than that for the Bonferroni procedure; hence the confidence intervals are slightly shorter.

Level	- Level	Difference	Lower CL	Upper CL	Difference
G	GL	94.17	49.85	138.49	
G	B	86.45	34.24	138.65	
GI	GL	52.06	6.79	97.32	
GI	B	44.33	-8.68	97.35	
G	GI	42.11	10.99	73.24	
B	GL	7.72	-53.97	69.41	



These CIs are *slightly* shorter (and hence more precise) than the Bonferroni ones. It turns out that we can still reject only the same **4** hypotheses as with Bonferroni. (Note: The difference between Bonferroni and T-K becomes more pronounced as the number of groups grows larger.)

Other Issues to be Addressed in Lecture (Optional additional material)

1. How (and where) to find CI^s for the different factor means?
2. How (and where) to find prediction CI^s for future observations on a given factor?
3. Where to find estimates for the parameters μ and α_i? *[Hint: Use "Fit Model" and the Drop-down "Expanded Estimates" option.]*
4. How to validate the model for homoscedasticity and normality?
5. Would it have been preferable to use Log(Return) here, rather than return?
6. Why isn't "linearity" a validation issue here, as it was in ordinary regression or multiple regression?
7. How does JMP (and other standard statistical software) use "indicator variables" to produce the Least Squares analysis? *[See Chapter 7 for an introduction to indicator variables. We won't need to master this material because JMP performs these operations automatically.]*