Statistical Techniques II

EXST7015

Post-ANOVA or Post-Hoc Tests

Overview of ANOVA

- **Recall that we are testing for differences among indicator variables.**
  - **The treatments may be fixed or random.**
  - **H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ for fixed effects.**
  - **H_0: $\sigma^2_\tau = 0$ for random effects.**
- **Assume $e_i \sim NIDrv(0,\sigma^2)$. Remember that this covers 3 separate assumptions.**
- **Also, assume no block "interactions" for the RBD.**

Overview *(continued)*

- **Every analysis can be expressed as a model with appropriate notation and subscripting.**
- **CRD : $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$**
- **For the moment we will be concerned only with examining for differences among the treatment levels.**
- **We will assume that we have already detected a significant difference among treatments levels with ANOVA.**

Overview *(continued)*

- **Treatments levels may be fixed or random. Determining appropriate tests depends on recognizing correctly.**
- **With random effects we are probably not interested in individual treatment levels. We are likely to be interested in the variability among the treatment levels and the distribution of the levels.**
- **With fixed effects we will probably want to compare individual levels.**

Post ANOVA tests

- **Having rejected the Null hypothesis we wish to determine how the treatment levels interrelate. This is the "post-ANOVA" part of the analysis.**
- **These tests fall into two general categories.**
  - **Post hoc tests (LSD, Tukey, Scheffé, Duncan's, Dunnett's, etc.)**
  - ***A priori* tests or pre-planned comparisons (contrasts)**

Post ANOVA *(continued)*

- ***A priori*** **tests are better. These are tests that the researcher plans on doing before they gather data, and if we dedicate 1 d.f. to each one we generally feel comfortable doing each at some specified level of alpha.**

Post ANOVA *(continued)*

- **However, since multiple tests do entail risks of higher experiment wide error rates, it would not be unreasonable to apply some technique, like Bonferroni's adjustment, to insure an experimentwise error rate of the desired level of alpha (α).**
- **So how might we do these "post hoc" tests?**

Post ANOVA *(continued)*

- **The simplest approach would be to do pairwise test of the treatments using something like the two-sample t-test.**
- **This tests examines the null hypothesis**
  - **H_0: $\mu_1 = \mu_2$ or H_0: $\mu_1 - \mu_2 = 0$,**
  - **against the alternative**
  - **H_a: $\mu_1 - \mu_2 \neq 0$, or H_a: $\mu_1 - \mu_2 \geq 0$ or H_a: $\mu_1 - \mu_2 \leq 0$.**

Post ANOVA *(continued)*

- **Recall two things about the two-sample t-test.**
  - **First, in a t-test we had to determine if the variance was equal for the two populations tested.**
  - **Second, the variance of the test (variance of the difference between μ_1 and μ_2) was equal to $\sigma^2_1/n_1 + \sigma^2_2/n_2$. If the variance are equal (as they MUST be for ANOVA) then the variance is $\sigma^2(1/n_1+1/n_2)$. We estimate σ^2 with MSE.**

Post ANOVA *(continued)*

- **So, we would test each pair of means using the two sample t-test as**
- **$t = (\bar{Y}_1 - \bar{Y}_2) / \sqrt{(MSE((1/n_1 + 1/n_2)))}$.**
- **If the design is balanced we can simplify this to $t = (\bar{Y}_1 - \bar{Y}_2)/\sqrt{(2MSE/n)}$.**

Post ANOVA *(continued)*

- **Notice that if the value of t is greater than the tabular value of t, we would reject the null hypothesis.**
- **If the value of t is less than the tabular value we would fail to reject.**
- **Lets call the tabular value t*, and write the case for rejection of the Null Hypothesis (H_0) as;**
- **$t^* \leq (\bar{Y}_1 - \bar{Y}_2) / \sqrt{(MSE((1/n_1+1/n_2)))}$.**

Post ANOVA *(continued)*

- **So we would reject H_0 if**
  - **$t^* \leq (\bar{Y}_1 - \bar{Y}_2) / \sqrt{(MSE((1/n_1+1/n_2)))}$**
  - **$t^*[\sqrt{(MSE((1/n_1+1/n_2)))}] \leq (\bar{Y}_1 - \bar{Y}_2)$**
  - **$(\bar{Y}_1 - \bar{Y}_2) \geq t^*[\sqrt{(MSE((1/n_1+1/n_2)))}]$**
- **So, for any difference $(\bar{Y}_1 - \bar{Y}_2)$ that is greater than $t^*[\sqrt{(MSE((1/n_1+1/n_2)))}]$ we find the difference statistically different (reject Ho), and for any value less we find the difference consistent with the null hypothesis. Right?**

Post ANOVA *(continued)*

- **This value of $t^*[\sqrt{(MSE((1/n_1+1/n_2)))}]$ is what R. A. Fisher called the "Least Significant Difference", commonly called the LSD (not to be confused with the Latin Square Design = LSD).**
- **We calculate this value for each pair of differences and if the observed difference is less, the treatments are "not significantly different". If greater they are "significantly different".**

Post ANOVA *(continued)*

- **One last detail. If the design is balanced then the value of $t^*[\sqrt{(MSE((1/n_1+1/n_2)))}]$ simplifies to $t^*[\sqrt{(2MSE/n)}]$. This is nice because all pairwise comparisons would use the same test value.**
- **It is nice, but not necessary.**
- **This is the first of our post ANOVA tests, it is called the "LSD".**

Post ANOVA *(continued)*

- **But hey, wait a minute! Didn't Fisher invent ANOVA in the first place to avoid doing a bunch of separate t-tests? So, now we are doing a bunch of separate t-tests.**
- **What is wrong with this picture?**

Post ANOVA *(continued)*

- **So, Fisher comes up with this.**
- **OK. When we do a bunch of separate t-tests, we don't know if there are any real differences at the α level. When we do the LSD as a post ANOVA test we SHOULD know that there are some differences. So we only do the LSD if the ANOVA says that there are differences, otherwise, don't do the LSD.**

Post ANOVA *(continued)*

- **This is called "Fisher's Protected LSD". We can use the LSD ONLY if the ANOVA shows differences, otherwise we are NOT justified in using the LSD.**
- **Makes sense. But there were still a lot of nervous statisticians looking for something a little better. As a result there are MANY alternative calculations. We will discuss the "classic" solutions.**

Post ANOVA *(continued)*

- **Basically, we calculate the LSD with our chosen value of α. We then do our mean comparisons. Each test has a pairwise error rate of α.**
- **We have already seen one alternative, the Bonferroni adjustment. If we do 5 tests, or 10 tests, our error rate is no more than 5(α/2) or 10(α/2).**
- **Generally, for g tests our error rate is no more than gα/2.**

Post ANOVA *(continued)*

- **To keep an experiment wide error rate of α, we simply do each comparison using a t value for an α equal to $\alpha/2g$.**
- **For two tailed tests (which the LSD almost always is) we do each test at $\alpha/2$ and the Bonferroni test would use a t for an error rate of $\alpha/2g$.**
- **One tailed tests are possible.**

Post ANOVA *(continued)*

- **The Bonferroni adjustment is fine if we are only doing a few tests. However, it is an upper boundary of the error, the highest that the error can be. The real probability of error is actually less.**
- **So if we are doing very many tests, Bonferroni gets very conservative, giving us an actual error rate much lower than the α we really want.**

Post ANOVA *(continued)*

- **So we seek alternatives.**
- **The big ones are Tukey's and Scheffé's. We will also consider Dunnett's and Duncan's since they are commonly used.**

Post ANOVA *(continued)*

- **Tukey is my favorite. This test basically allows for <u>all pairwise tests.</u>**
- **Tukey developed his own tables. The table values are similar to t values, but gives the correct value to use for given values of α, the number of tests and d.f. error.**
- **Note SAS puts "HSD" by Tukey's. This stands for "Honest Significant Difference".**

Post ANOVA *(continued)*

- **Scheffé came up with a very conservative test. This test allows for all possible tests (all possible contrasts). Not only can we test all pairwise tests, but all combinations of tests (including contrasts to be discussed later).**
  - **e.g. H_0: $(\mu_1+\mu_2)/2 = (\mu_3+\mu_4+\mu_5)/3$**
- **This test is appropriate for "data dredging".**

Post ANOVA *(continued)*

- **Note that if you want to do a couple of pairwise tests you can calculate Bonferroni and compare to Tukey's. Tukey's is conservative for fewer than all possible pairwise tests and Bonferroni is conservative because it is a bound.**
- **For other sets of tests including some that are not pairwise, compare Bonferroni to Scheffé.**

Post ANOVA *(continued)*

- **Comparison wise error rate: LSD**
- **Experiment wise error rate: Tukey (all pairwise), Bonferroni (selected tests), Scheffé (all possible contrasts).**
- **When doing pairwise tests, the LSD is the test most likely to find differences, and the one most likely to be wrong when it finds a difference.**
- **Scheffé is the test least likely to find a difference, and least likely to be wrong.**

Post ANOVA *(continued)*

- **There are two other tests that are used in particular circumstances.**
  - **Dunnett's is used to compare one treatment to all other treatments.**
  - **Duncan's intended to give a groupwise or family wise error rate. When means are grouped according to which are different and which are not, this test should have only a α% chance of error for each group.**

Post ANOVA *(continued)*

- **All of these tests can be expressed in one of two ways.**
- **If the analysis is BALANCED, then there is a popular expression of pairwise tests that starts with ranked means.**
- **Suppose we calculate a value of the LSD equal to 8, and we have sorted the means of treatment levels and have 5, 14, 17, 23, 29, and 38.**

Post ANOVA *(continued)*

LSD = 8

Treatment Level	Mean	Groups
3	38	
1	29	
6	23	
5	17	
2	14	
4	5	

Post ANOVA *(continued)*

LSD = 8

Treatment Level	Mean	Groups
3	38	A
1	29	B
6	23	B C
5	17	D C
2	14	D
4	5	E

Post ANOVA *(continued)*

Tukey adjusted = 10

Treatment Level	Mean	Groups
3	38	A
1	29	A B
6	23	B C
5	17	C
2	14	C
4	5	D

Post ANOVA *(continued)*

Scheffé adjusted = 15

Treatment Level	Mean	Groups
3	38	A
1	29	A B
6	23	A B
5	17	B C
2	14	B C
4	5	C

Examples

- **Test the effects of fumigants on wire worms. Treatments are two fumigants (C and S) and a control (0).**
- **In SAS these are done from the GLM using the means statement. They can be done from MIXED using the LSMeans statement.**

Examples *(continued)*

- **SAS statements**

```
PROC GLM DATA=FUMIGANT;
    CLASSES FUMIGANT BLOCK REP;
    TITLE3 'Analysis of fumigant with RBD';
  MODEL WORMS = FUMIGANT BLOCK FUMIGANT*BLOCK;
  MEANS FUMIGANT / DUNNETT('0')
        E=FUMIGANT*BLOCK;
  MEANS FUMIGANT / LSD DUNCAN TUKEY SCHEFFÉ
        E=FUMIGANT*BLOCK; RUN; QUIT;
```

Examples *(continued)*

- **Results with the LSD**

```
Alpha                                    0.05
Error Degrees of Freedom                    8
Error Mean Square                    24.52917
Critical Value of t                   2.30600
Least Significant Difference           3.6116

Grouping    Mean        N     FUMIGANT
A          9.700       20     0
B          5.250       20     C
B          4.800       20     S
```

Examples *(continued)*

- **Results with Duncan's MRT.**

```
Alpha                                0.05
Error Degrees of Freedom                8
Error Mean Square               24.52917
Number of Means                        2          3
Critical Range                     3.612      3.764

Grouping    Mean        N    FUMIGANT
A          9.700       20    0
B          5.250       20    C
B          4.800       20    S
```

Examples *(continued)*

- **Results with Tukey's.**

Alpha	0.05
Error Degrees of Freedom	8
Error Mean Square	24.52917
Critical Value of Studentized Range	4.04101
Minimum Significant Difference	4.4752

Grouping	Mean	N	FUMIGANT
A	9.700	20	0
B A	5.250	20	C
B	4.800	20	S

Examples *(continued)*

- **Results with Scheffé's.**

Alpha	0.05
Error Degrees of Freedom	8
Error Mean Square	24.52917
Critical Value of F	4.45897
Minimum Significant Difference	4.6771

Grouping	Mean	N	FUMIGANT
A	9.700	20	0
B A	5.250	20	C
B	4.800	20	S

Examples *(continued)*

- **Results with Dunnett's.**

```
Alpha                                    0.05
Error Degrees of Freedom                    8
Error Mean Square                       24.52917
Critical Value of Dunnett's t            2.67281
Minimum Significant Difference           4.1861

                Difference          Simultaneous
 FUMIGANT          Between         95% Confidence
Comparison          Means             Limits
C    - 0           -4.450         -8.636  -0.264  ***
S    - 0           -4.900         -9.086  -0.714  ***
```

Post ANOVA *(continued)*

- **Comparison of ranked means works very well if the analysis is balanced. If the analysis is not balanced there can be a problem.**
- **It is possible that means that are close together are significantly different, while means that have a greater difference are not significantly different.**

Post ANOVA *(continued)*

Variance = MSE($1/n_1 + 1/n_2$)

mse	= 25					
tmt	mean	n	test	diff	se	2*se
1	18	5	1 v 2	5	2.29	4.58
2	13	100	2 v 3	1	2.29	4.58
3	12	5	1 v 3	6	3.16	6.32

Post ANOVA *(continued)*

- **For unbalanced tests the best way to check for difference is to calculate a confidence interval for each mean and see if the confidence intervals overlap.**
- **By default, SAS will use this approach for unbalanced means.**

Example

- **Typhoid strain example. Number of days to mouse mortality was the dependent variable.**
- **SAS statements**

```
PROC MIXED DATA=OnebyOne cl; CLASSES STRAIN;
TITLE3 'ANALYSIS OF VARIANCE with PROC MIXED';
MODEL DAYS = STRAIN / htype=3 DDFM=Satterthwaite;
repeated / group=strain;
LSMEANS STRAIN / ADJUST=TUKEY pdiff;
LSMEANS STRAIN / ADJUST=SCHEFFE pdiff;
LSMEANS STRAIN / ADJUST=BON pdiff;
```

- ***NOTE that normally only one post-ANOVA examination would be done. We have done several here for comparison.;**

Example *(continued)*

- **First the LSMeans statement prints the means**

Least Squares Means

Effect	STRAIN	Estimate	Standard Error	DF	t Value	Pr > |t|
STRAIN	11C	7.3667	0.3126	59	23.57	<.0001
STRAIN	9D	4.0323	0.2475	30	16.29	<.0001
STRAIN	DSC1	7.7970	0.2233	132	34.91	<.0001
STRAIN	11C	7.3667	0.3126	59	23.57	<.0001
STRAIN	9D	4.0323	0.2475	30	16.29	<.0001
STRAIN	DSC1	7.7970	0.2233	132	34.91	<.0001
STRAIN	11C	7.3667	0.3126	59	23.57	<.0001
STRAIN	9D	4.0323	0.2475	30	16.29	<.0001
STRAIN	DSC1	7.7970	0.2233	132	34.91	<.0001

  - **These are repeated 3 times because there are 3 LSMeans statements in the SAS program.**

Example *(continued)*

- **Results for the LSD and the Tukey adjustment.**

Differences of Least Squares Means

Effect	STRAIN	_STRAIN	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
STRAIN	11C	9D	3.3344	0.3987	88.1	8.36	<.0001	Tukey-Kramer	<.0001
STRAIN	11C	DSC1	-0.4303	0.3842	121	-1.12	0.2649	Tukey-Kramer	0.5037
STRAIN	9D	DSC1	-3.7647	0.3334	85.8	-11.29	<.0001	Tukey-Kramer	<.0001

Example *(continued)*

- **Results for the Bonferroni and the Scheffé adjustment**

Differences of Least Squares Means

Effect	STRAIN	_STRAIN	Estimate	Standard Error	DF	t Value	Pr > |t|	Adjustment	Adj P
STRAIN	11C	9D	3.3344	0.3987	88.1	8.36	<.0001	Scheffe	<.0001
STRAIN	11C	DSC1	-0.4303	0.3842	121	-1.12	0.2649	Scheffe	0.5358
STRAIN	9D	DSC1	-3.7647	0.3334	85.8	-11.29	<.0001	Scheffe	<.0001
STRAIN	11C	9D	3.3344	0.3987	88.1	8.36	<.0001	Bonferroni	<.0001
STRAIN	11C	DSC1	-0.4303	0.3842	121	-1.12	0.2649	Bonferroni	0.7950
STRAIN	9D	DSC1	-3.7647	0.3334	85.8	-11.29	<.0001	Bonferroni	<.0001

Post-ANOVA tests

- **The test we have seen so far are often (usually?) done with no *a priori* hypotheses in mind. We do not have certain comparisons in mind before doing the experiment, we want to examine many, or all, levels of the treatments for differences from one another.**
- **The experimentwise error rate is intended to allow this (except for the LSD).**

Post-ANOVA tests *(continued)*

- **However, sometimes we do have some particular comparisons in mind when we do an experiment.**
- **When we want some lesser number of comparisons, and they are determined *a priori* (without looking at the data), then we can use a less stringent criteria.**

Post-ANOVA tests *(continued)*

- **We generally feel comfortable with one test per degree of freedom at some specified level of alpha (α), just as we did in regression (looking at each regression coefficient with an α level of error).**
- **This is the case with *a priori* contrasts.**