Displays for Statistics 5303

Lecture 10

September 25, 2002

Christopher Bingham, Instructor

612-625-7023 (St. Paul)
612-625-1024 (Minneapolis)

Class Web Page
http://www.stat.umn.edu/~kb/classes/5303

A question was asked in class as to how to do Exercise 5.2.

You are given means $\overline{y_{i\bullet}}$ = 3.2892, 10.256, 8.1157, 8.1825 and 7.5622 as results of a completely randomized design with g = 5 treatments and $n_1 = n_2 = \ldots = n_5 = 4$.

You are also told MSE = 4.012.

(a) Construct an ANOVA table for this experiment and test the null hypothesis that all treatments have the name mean.

Without the original data, there is no way to use **anova()** to do this. You have to fall back on formulas for SS_{trt} and SS_E.

By an equation on p. 46

$$SS_{trt} = \sum_{1\le i\le g} n_i(\overline{y}_{i\bullet} - \overline{y}_{\bullet\bullet})^2,$$

where

$$\overline{y}_{\bullet\bullet} = \sum_{1\le i\le g}\sum_{1\le j\le n_i} y_{ij}/N = \sum_{1\le i\le g} n_i\overline{y}_{i\bullet}/N$$

$$N = \sum_{1\le i\le g} n_i$$

Here's one way you could find the various quantities needed for an ANOVA table.

```
Cmd> ybars <- vector(3.2892,10.256,8.1157,8.1825,7.5622)
Cmd> n <- rep(4,5) # or vector(4,4,4,4,4), sample sizes
Cmd> N <- sum(n) # total number of cases
Cmd> g <- 5 # number of treatments
Cmd> df_trt <- g-1 # treatment DF
Cmd> df_error <- N - g # error DF
Cmd> grandmean <- sum(n*ybars)/N # from formula above
Cmd> ss_trt <- sum(n*(ybars - grandmean)^2) #from formula above
Cmd> ms_trt <- s_trt/df_trt
Cmd> ms_error <- 4.012 # given as MSE; = ss_error/df_error
Cmd> ss_error <- df_error * ms_error
Cmd> fstat <- ms_trt/ms_error # ratio of mean squares
Cmd> p_value <- 1 - cumF(fstat,df_trt,df_error)
```


You can now print out **ss_trt**, **ss_error**, **df_trt**, **df_error**, **ms_trt**, **ms_error**, **fstat** and **p_value** and arrange them in a table

(b) Test the null hypothesis that the average response in treatments 1 and 2 is the same as the average response in treatments 3, 4 and 5.

As always you need to express this symbolically. You are asked to test

$$H_0: (\mu_1 + \mu_2)/2 - (\mu_3 + \mu_4 + \mu_5)/3 = 0$$

This is a contrast in the group means with weights

$$\{w_i\} = \{1/2,1/2,-1/3,-1/3,-1/3)$$

To do the test you need a t-statistic of the form

t = estimate/(standard error of estimate)

The *estimate* is $\sum_{1\le i\le g} w_i\overline{y}_{i\bullet}$ with *estimated standard error* (see p. 68)

$$\hat{SE}[\sum_{1\le i\le g} w_i\overline{y}_{i\bullet}] = s_p\sqrt{\{\sum_{1\le i\le g} w_i^2/n_i\}},\ s_p^2 = MS_E$$

To do this in MacAnova you need to translate formulas into MacAnova commands:

```
Cmd> w <- vector(1/2,1/2,-1/3,-1/3,-1/3) #contrast weights

Cmd> sum(w) # sum is zero so it's a contrast
(1)   1.1102e-16

Cmd> estimate <- sum(w*ybars) # estimated contrast

Cmd> std_error <- sqrt(ms_error*sum(w^2/n))

Cmd> tstat <- estimate/std_error # t-statistic

Cmd> pval <- twotailt(tstat,df_error) # p-value
```

You can now use the P-value to decide whether you can reject H_0.

It would be a lot easier with all the data, since then you could do something like:

```
Cmd> anova("y = treat", fstat:T)

Cmd> result <- contrast(treat,w)

Cmd> tstat <- result$estimate/result$se

Cmd> pval <- twotailt(tstat,DF[3])
```

Here **DF[3]** is the third element of variable **DF** created by **anova()** and containing the DF column from the ANOVA table.

More on multiple comparisons

Several multiple comparison methods are based on the distribution of the **Studentized Range**.

Mathematically, the Studentized range distribution is defined as follows:

- Let $X_1, X_2, \ldots, X_K$ be a random sample from $N(\mu, \sigma^2)$
- Let S^2 be an estimate of σ^2 distributed as $\sigma^2\chi_{df}^2/df$ *independent* of $\{X_i\}$.

$Q = \text{Range}(\{X_i\})/S = (\max(\{X_i\}) - \max(\{X_i\}))/S$ has the **Studentized range distribution**.

Comment: S^2 is an unbiased estimate of σ^2, that is $\mu_{S2} = \sigma^2$.

Note that all the X_i's must have the same variance.

The distribution of Q is characterized by

- K = number of observations in range
- df = degrees of freedom associated with S^2

The distribution does not depend on σ or on μ.

Table D.8 on Oehlert p. 633-634 has upper 5% and 1% critical values for Q for K = 1, 2, ..., 10, 15, 20, 30, and 50 and degrees of freedom df = ν = 1, 2, ..., 30, 35, 40, 50, 100 and ∞.

$\nu = \infty$ corresponds to the case when σ^2 is known and the ratio is

$$Q = (\max(\{X_i\})-\max(\{X_i\}))/\sigma,$$

that is the actual value of σ is used instead of an estimate.

Table D.8: Percent points for the Studentized range

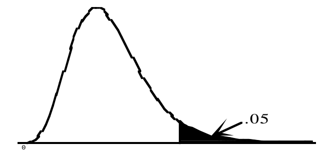

Table entries are $q_{.05}(K,\nu)$

	K												
ν	2	3	4	5	6	7	8	9	10	15	20	30	50
1	18.0	27.0	32.8	37.1	40.4	43.1	45.4	47.4	49.1	55.4	59.6	65.1	71.7
2	6.09	8.33	9.80	10.9	11.7	12.4	13.0	13.5	14.0	15.7	16.8	18.3	20.1
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	10.5	11.2	12.2	13.4
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.66	9.23	10.0	10.9
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.72	8.21	8.87	9.67
6	3.46	4.34	4.90	5.31	5.63	5.90	6.12	6.32	6.49	7.14	7.59	8.19	8.91
7	3.34	4.17	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.76	7.17	7.73	8.40
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.48	6.87	7.40	8.03
9	3.20	3.95	4.42	4.76	5.02	5.24	5.43	5.59	5.74	6.28	6.64	7.14	7.75
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	6.11	6.47	6.95	7.53
11	3.11	3.82	4.26	**4.57**	4.82	5.03	5.20	5.35	5.49	5.98	6.33	6.79	7.35
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.88	6.21	6.66	7.21
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.79	6.11	6.55	7.08
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.71	6.03	6.46	6.98
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.65	5.96	6.38	6.89
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.59	5.90	6.31	6.81
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.54	5.84	6.25	6.74
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.50	5.79	6.20	6.68
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.46	5.75	6.15	6.63
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.43	5.71	6.10	6.58
21	2.94	3.56	3.94	4.21	4.42	4.60	4.74	4.87	4.98	5.40	5.68	6.07	6.53
22	2.93	3.55	3.93	4.20	4.41	4.58	4.72	4.85	4.96	5.37	5.65	6.03	6.49
23	2.93	3.54	3.91	4.18	4.39	4.56	4.70	4.83	4.94	5.34	5.62	6.00	6.46
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.32	5.59	5.97	6.42
25	2.91	3.52	3.89	4.15	4.36	4.53	4.67	4.79	4.90	5.30	5.57	5.94	6.39
26	2.91	3.51	3.88	4.14	4.35	4.51	4.65	4.77	4.88	5.28	5.55	5.92	6.36
27	2.90	3.51	3.87	4.13	4.33	4.50	4.64	4.76	4.86	5.26	5.53	5.89	6.34
28	2.90	3.50	3.86	4.12	4.32	4.49	4.62	4.74	4.85	5.24	5.51	5.87	6.31
29	2.89	3.49	3.85	4.11	4.31	4.47	4.61	4.73	4.84	5.23	5.49	5.85	6.29
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	5.21	5.48	5.83	6.27
35	2.87	3.46	3.81	4.07	4.26	4.42	4.56	4.67	4.77	5.15	5.41	5.76	6.18
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	5.11	5.36	5.70	6.11
45	2.85	3.43	3.77	4.02	4.21	4.36	4.49	4.61	4.71	5.07	5.32	5.66	6.06
50	2.84	3.42	3.76	4.00	4.19	4.34	4.47	4.58	4.68	5.04	5.29	5.62	6.02
100	2.81	3.36	3.70	3.93	4.11	4.26	4.38	4.48	4.58	4.92	5.15	5.46	5.83
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.80	5.01	5.30	5.65

You can compute critical values (upper percent points) in MacAnova using **`invstudrng()`**:

```
Cmd> invstudrng(1 - .05,5, 11)
(1)      4.5737        Upper 5% point of Q
```

You can get an upper tail probability (P-value) using **`cumstudrng()`**:

```
Cmd> q_obs <- 5.123; 1 - cumstudrng(q_obs, 5, 11)
(1)    0.026489        P(Q ≥ 5.123)
```

In the multiple comparison situation, when H_0: $\mu_1 = \mu_2 = \ldots = \mu_g = \mu$ is true, and $n_1 = n_2 = \ldots = n_g = n$ (equal sample sizes)

- $\overline{y}_{1\bullet}, \overline{y}_{2\bullet}, \ldots, \overline{y}_{g\bullet}$ are independent $N(\mu,\sigma^2/n)$
- $MS_E/n = s_p^2/n = \hat{\sigma}^2/n$ is independent of $\{\overline{y}_{i\bullet}\}$ with distribution $(\sigma^2/n)\chi_{N-g}^2/(N-g)$

Identifying $\overline{y}_{i\bullet}$ with X_i and s_p^2/n with S^2

$$Q = \{\max(\overline{y}_{i\bullet}) - \min(\overline{y}_{i\bullet})\}/(s/\sqrt{n})$$

has the Studentized range distribution with $K = g$ and df = $N-g$. The "range" is the range of sample means.

This provides an alternative way (to an F statistic) to test H_0: $\mu_1 = \mu_2 = \ldots = \mu_g$ when the sample sizes are equal:

Reject H_0: when $Q \geq q_\alpha(g,N-g)$

```
Cmd> data33 <- read("","pr3.3",quiet:T) # Problem 3.3 data
Read from file "TP1:Stat5303:Data:OeCh03.dat"

Cmd> treat <- factor(data33[,1]) # create treatment factor

Cmd> longevity <- vector(data33[,2]) # create response vector

Cmd> anova("longevity=treat",fstat:T)
Model used is longevity=treat
                DF          SS          MS          F      P-value
CONSTANT         1      2782.4      2782.4 1349.49826      < 1e-08
treat            4      243.16       60.79   29.48371   5.9878e-07
ERROR1          15      30.928      2.0618

Cmd> dfe <- DF[3]; mse <- SS[3]/dfe # mse = 30.928/15

Cmd> vector(mse, dfe) # same as in ERROR1 line of table
      ERROR1      ERROR1
      2.0618          15

Cmd> ybars <- tabs(longevity,treat, means:T);ybars#sample means
(1)         18          12      11.975           9           8

Cmd> tabs(longevity,treat,count:T) # sample sizes
(1)          4           4           4           4           4

Cmd> n <- 4 # common value of sample sizes

Cmd> g <- 5 # number of groups

Cmd> q <- (max(ybars) - min(ybars))/sqrt(mse/n); q
(1)     13.928   Studentized range Q

Cmd> invstudrng(1 - .01,5, dfe) # Critical value
(1)     5.5563   Q = 13.928 >> 5.5563; reject at 1% level

Cmd> 1 - cumstudrng(q,5,dfe) # P-value
(1)   1.3828e-05  Very small P-value => Reject H0
```

When is Q significant? When the range of $\overline{y}_{i\bullet}$'s is large enough. Specifically, when

$$\max(\overline{y}_{i\bullet}) - \min(\overline{y}_{i\bullet}) \geq HSD$$

where the **H**onestly **S**ignificant **D**ifference HSD is defined to be

$$HSD = q_\alpha(g,N-g)s_p/\sqrt{n}$$

Now $\hat{SE}[\overline{y}_{i\bullet} - \overline{y}_{j\bullet}] = \sqrt{\{2\times s_p^2/n\}} = \sqrt{2}\times s_p/\sqrt{n}$ so another expression for the HSD is

$$HSD = q_\alpha(g,N-g)\times\hat{SE}[\overline{y}_{i\bullet}-\overline{y}_{j\bullet}]/\sqrt{2}.$$

Note the quantity $\sqrt{2}$ in the denominator.

Obviously, if any $|\overline{y}_{i\bullet} - \overline{y}_{j\bullet}| > HSD$, then

$$\max(\overline{y}_{i\bullet}) - \min(\overline{y}_{i\bullet}) > HSD$$

so another way to test H_0 is reject H_0 if $|\overline{y}_{i\bullet} - \overline{y}_{j\bullet}| > HSD$ for any $i \neq j$.

The nice thing about this is that when you reject H_0 you have information about which means are different. Specifically, for any $i \neq j$, when $|\overline{y}_{i\bullet} - \overline{y}_{j\bullet}| > HSD$ you reject H_{0ij}: $\mu_i = \mu_j$.

This procedure is the basis of the **HSD multiple comparisons method**, also known as the **Tukey** method and the **Studentized range** method.

Tukey named it the *Honestly significance difference* because he believed the most widely used method, the LSD or **L**east **S**ignificant **D**ifference method, often found more significant differences than was really supported by the data and thus was not quite "honest".

The **(protected) LSD method** is the oldest multiple comparisons method. It was first formalized by R. A. Fisher.

Suppose your sample sizes are equal $n_1 = n_2 = \ldots = n_g = n$.

Then the <u>naive</u> method rejects H_{0ij}: $\mu_i = \mu_j$ when $|t_{ij}| \geq t_{\alpha/2,N-g}$, where t_{ij} is a t-statistic defined as:

$$t_{ij} = (\overline{y}_{i\bullet} - \overline{y}_{j\bullet})/\widehat{SE}[\overline{y}_{i\bullet} - \overline{y}_{j\bullet}]$$
$$= (\overline{y}_{i\bullet} - \overline{y}_{j\bullet})/\sqrt{(2s_p^2/n)}$$
$$s_p^2 = MS_E \text{ from ANOVA}$$

This is the same as rejecting H_{0ij} when

$$|\overline{y}_{i\bullet} - \overline{y}_{j\bullet}| \geq LSD = t_{\alpha/2,N-g} \times \sqrt{(2s_p^2/n)}$$

the **L**east **S**ignificant **D**ifference.

Thus you might call this the **naive LSD method**. Its per comparison error is α but its experimentwise error rate can be very high.

The *protected* LSD method has 2 steps.

1. Do an ANOVA. If F is not significant at level α then you are done; there is no evidence that any means differ.
2. Only if F is significant, compute the LSD and reject H_{0ij} if $|\overline{y}_{i\bullet} - \overline{y}_{j\bullet}| > LSD$

With this procedure, the only way you can make a type I error is if you get to step 2 and then find $|\overline{y}_{i\bullet} - \overline{y}_{j\bullet}| > LSD$, and even then it may not be a type I error.

When all the means are equal,

$$P(\text{get to step 2}) = P(F > F_\alpha) = \alpha,$$

so

$$P(\text{any type I error}) \leq \alpha,$$

This means the experimentwise error rate cannot be greater than α. However, the strong experimentwise error rate can be much bigger.

The practical application of the LSD method, HSD method as well as other methods starts with ordering the means from smallest to largest, say

$$\overline{y}_{(1)\bullet} \le \overline{y}_{(2)\bullet} \le \ldots \le \overline{y}_{(g)\bullet}$$

corresponding to means $\mu_{(1)}, \mu_{(2)}, \ldots, \mu_{(g)}$.

You need to keep track of which treatment $\overline{y}_{(i)\bullet}$ and $\mu_{(i)}$ go with.

You first find all means $\overline{y}_{(i)\bullet}$, if any, that are not significantly from $\overline{y}_{(1)\bullet}$. These are all the means such that treatments such that $\overline{y}_{(i)\bullet} < \overline{y}_{(1)\bullet}$ + LSD. Often a line is drawn under these. Then all means $\overline{y}_{(i)\bullet}$ with i > 2 such that $\overline{y}_{(i)\bullet} < \overline{y}_{(2)\bullet}$ + LSD are considered not significantly different from $\mu_{(2)}$, and a line drawn under them, and so on. If a line is completely under another line it is not drawn.


```
Cmd> lsd <- sqrt(2*mse/n)*invstu(1 - .05/2,dfe); lsd
(1)       2.1641   5% Least Significant Difference

Cmd> sort(ybars) # ordered means
(1)          8          9       11.975         12          18

Cmd> sort(ybars)[-5] + lsd
(1)       10.164      11.164      14.139      14.164
```

A line connects the first two means because 8 + 2.164 = 10.164 > 9 and 10.164 < 11.975. Since 9 + 2.164 = 11.164 < 11.975 no line is drawn connecting the 2nd and 3rd mean. And so on.

You can use **grade(ybars)** to recover the treatment numbers of each mean.

```
Cmd> grade(ybars)
(1)          5          4          3          2          1
```

Macro **pairwise()** provides a black box way to do the comparison, orienting things vertically rather than horizontally.

```
Cmd> pairwise("treat",.05,lsd:T)
    |      5    -3.79     lsd:T directs that the LSD
    |      4    -2.79     method be used
      |    3     0.18
      |    2    0.205
           1     6.21
```

The first column of numbers are treatment numbers and the last column are effects $\hat{\alpha}_i$, not sample means.

```
Cmd> ybars - sum(ybars)/5 # alpha_hats
(1)       6.205       0.205       0.18      -2.795      -3.795
```

The HSD method is done the same using HSD instead of LSD

```
Cmd> hsd <- invstudrng(1 - .05, 5, dfe)*sqrt(mse/n);hsd
(1)      3.1354   5% Honestly significant difference

Cmd> sort(ybars)[-5] + hsd # ordered means + HSD
(1)      11.135      12.135       15.11      15.135

Cmd> sort(ybars) # ordered means
(1)            8           9      11.975          12          18
               ____________
                                  _________________________
```

Now a line is drawn under the 2^{nd}, 3^{rd} and 4^{th} means because 9 + HSD = 9 + 3.135 = 12.135 > 12.

```
Cmd> pairwise("treat",.05,hsd:T)
  |       5    -3.79
  | |     4    -2.79
    |     3     0.18
    |     2    0.205
          1     6.21
```


The Bonferroni method as applied to multiple comparisons can also be expressed in terms a significant difference, BSD = **Bonferroni significant difference**.

BSD is like the LSD except a *Bonferronized* Student's t critical value $t_{\alpha/K,N-g}$ is used, taking account of there being K = g(g-1)/2 different comparisons.

```
Cmd> bsd <- invstu(1 - .025/10, dfe)*sqrt(2*mse/n); bsd
(1)      3.3364   5% Bonferroni significant difference

Cmd> sort(ybars)[-5] + bsd
(1)      11.336      12.336      15.311      15.336

Cmd> sort(ybars) # ordered means
(1)            8           9      11.975          12          18
               ____________
                                  _________________________

Cmd> pairwise("treat",.05,bsd:T)
  |       5    -3.79
  | |     4    -2.79
    |     3     0.18
    |     2    0.205
          1     6.21
```