Displays for Statistics 5401/8401

Lecture 20

October 21, 2005

Christopher Bingham, Instructor

612-625-1024, kb@umn.edu
372 Ford Hall

Class Web Page

http://www.stat.umn.edu/~kb/classes/5401

Recap: Testing linear hypothesis H_o using <u>hypothesis</u> and <u>error</u> <u>matrices</u> **H** and **E** with degrees of freedom f_h and f_e.

<u>Relative eigenvalues</u> $\hat{\lambda}_i$ are eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$.

Roy's maximum root test

Reject H_0 when $\hat{\lambda}_1 = \hat{\lambda}_{max}$ is "large"

I found estimates of $\hat{\lambda}_{max}(.10)$, $\hat{\lambda}_{max}(.05)$ and $\hat{\lambda}_{max}(.01)$ from 5000 simulated values in **lambda_max**.

```
Cmd> lambda_max[round(vector(.90,.95,.99)*M)]
(1)     0.076562     0.090821      0.12154
```

Actually **Roy** proposed the <u>canonical correlation form</u> of the statistic

$$\hat{\theta}_1 = \hat{\theta}_{max}$$

where $\hat{\theta}_j = \hat{\lambda}_j/(1 + \hat{\lambda}_j)$, j = 1, ..., p

```
Cmd> theta_max <- lambda_max/(1 + lambda_max)
Cmd> theta_max[round(vector(.90,.95,.99)*M)] # critical vals
(1)     0.071117     0.083259      0.10837   10%, 5%, 1%
```

This approach by simulation is always available with the right software

Alternatively you can use tables or charts where available.

I posted a handout with charts from D. L. Heck, Charts of Some Upper Percentage Points of the Distribution of the Largest Characteristic Root, *Ann.Math. Statist.* **31** (1960) 625-642.

These give upper 5% and 1% points of $\hat{\theta}_{max}$ for a range of situations.

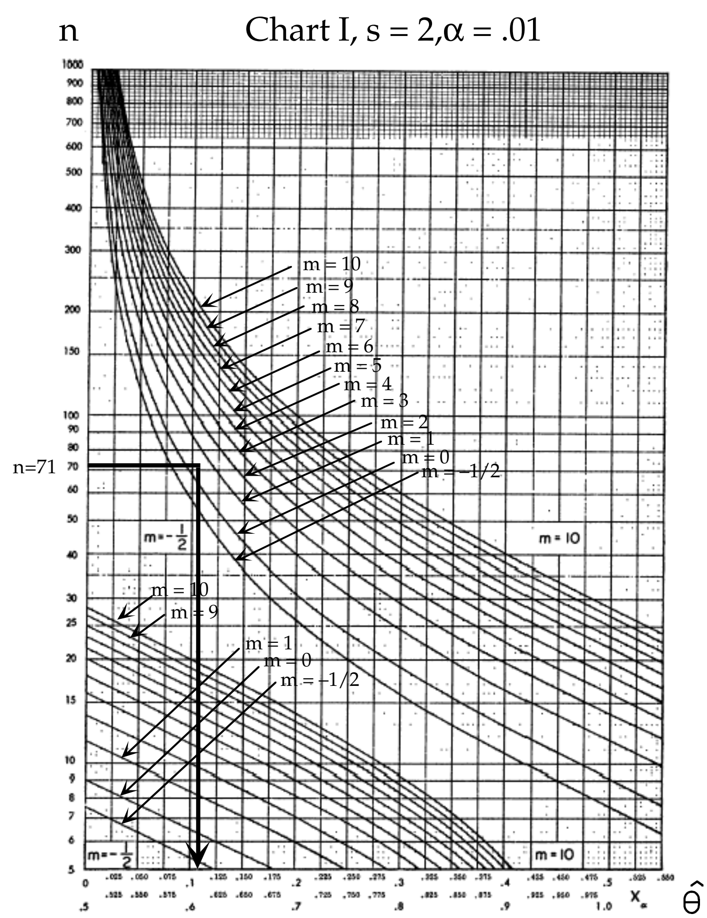
The null distributions of the $\hat{\lambda}_j$ and $\hat{\theta}_j$ (and of any statistic computed from them) depend on 3 quantities.

- s = min(f_h, p) = rank(**H**) $\geq$ 1, integer
- m = ($|f_h - p|$ - 1)/2 $\geq$ -1/2, integer or half integer
- n = (f_e - p - 1)/2 $\geq$ -1, integer or half integer (n is *not* the sample size)

You can check that

- m + s = (f_h + p - 1)/2

Chart I, s = 2, α = .01



Using the charts of probability points

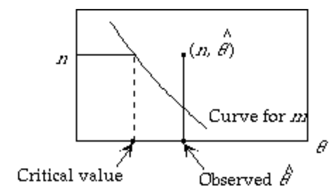
- Each chart is a set of double-width graphs . The curves in the lower left continue (with a little overlap) the curves that start in the upper left.
- Each chart goes with one value of s = 2, 3, 4, or 5 and one value of α = .05 or .01.
- Each curve on a chart goes with one value of $m = -1/2, 0, 1, 2, 3, \ldots, 10$

  The two bottom curves in each group are for $m = -1/2$ and $m = 0$; the others are for $m = 1, 2, \ldots$, stepping by 1. For in between values you need to interpolate. So for $m = 5/2$ interpolate between $m = 2$ and $m = 3$ curves.
- x-axis (two scales) is $\hat\theta_{max}$
- y axis represents $n = (f_e - p - 1)/2$ from 5 to 1000. Large f_e leads to large n. Large p reduces n.

The *two* scales below the X-axis *both* represent values of $\hat\theta_1$. The upper scale, from 0 to .550 goes with the upper set of curves. The lower scale, from .500 to 1, goes with the lower curves.

- Find a **critical value**:

Find where the curve for m crosses the horizontal line corresponding to n. The horizontal position of the intersection is the critical value.



- **Significance test** of an observed $\hat\lambda_1$: Compute $\hat\theta_1 = \hat\lambda_1/(1 + \hat\lambda_1)$, s, m and n. Find point $(\hat\theta_1, n)$ on chart for s and α. Reject H_0 when it is to the right of the curve for m; otherwise don't reject H_0.

Application to Fisher Data.

N = 150, g = 3, p = 4

- $f_e = N - g = 147$
- $f_h = g - 1 = 2$
- $s = \min(4, 2) = 2$
- $m = (|4 - 2| - 1)/2 = 1/2 = 1/2$
- $n = (147 - 4 - 1)/2 = 142/2 = 71$

On Chart for s = 2 and α = .01, the heavy line traces from n = 71 to the critical value. Since the curves for m = 0 and 1 are 2nd and 3rd curves, we need to interpolate between them.

The intersection with the m = 0 line is approximately at **.095** and the intersection with the m = 1 line is approximately at **.118**. So the critical value is about (0.095+.118)/2 = **.1065**

```
Cmd> thetahat <- lambdahat/(1 + lambdahat)
Cmd> thetahat[1]
(1)     0.96987    >> .1065, v. signif.
```

Simulated 0.1084 is close to .1065.

Pillai's V-statistic is

$$V = (f_h + f_e)\mathrm{tr}(\mathbf{H} + \mathbf{E})^{-1}\mathbf{H}$$
$$= (f_h + f_e)\sum_{1\le i\le s}\hat\lambda_i/(1 + \hat\lambda_i) \cong \chi^2_{f_h p}$$

Continuing with the artificial data:

```
Cmd> v <- (fh + fe)*trace(solve(h+e,h)); v
(1)      23.576
Cmd> (fh+fe)*sum(eigvals/(1 + eigvals))
(1)      23.576      Computed from relative eigenvalues
Cmd> cumchi(v,fh*p,upper:T)
(1)    0.023213
```

This is a large sample P-value computed from χ^2.

You can use **cumtrace()** with keyword phrase **pillai:T** to get a more exact P-value:

```
Cmd> cumtrace(trace(solve(h+e,h)),fh,fe,p,pillai:T,upper:T)
(1)    0.013864
```

Note: The degrees of freedom for the large sample χ^2 approximation to the null distribution of Hotelling's T_0^2, Pillai's V and the log LR test are all the same, $f = f_h \times p$ = the number of scalar coefficients or linear combinations of coefficients that are being tested.

Complex MANOVA situations

Suppose you have a complex MANOVA situation that is more complex than one-way MANOVA.

For example, you might need to analyze multivariate data from a completely randomized 3- way factorial (3 way ANOVA) or a split plot design.

If you know how to do a univariate ANOVA for the situation, you know how to do a MANOVA.

Suppose you are analyzing a split plot experiment with whole block factor A arranged in a randomized block design and sub plot factor B.

The correct analysis has <u>two error terms</u>, whole plot and subplot.

For a *univariate* split plot ANOVA with whole plots arranged in a RBD, in MacAnova you would use a command like the following.

```
anova("y=reps+a+E(reps.a)+b+a.b")
```

where **y** is N by 1 and **reps**, **a** and **b** are factors coding <u>replications</u>, the <u>whole plot factor</u> and <u>subplot factor</u>.

Variable `SS` will contain hypotheses sums of squares `SS[1]` = SS_h for `CONSTANT`, `SS[2]` = **reps** = SS_h for reps, `SS[3]` = SS_h for **a**, `SS[5]` = SS_h for **b** (`SS[5]`), and `SS[3]` = SS_h for **a.b** (interaction). The two error SS are `SS[4]` = SSE_{WP} and `SS[7]` = SSE_{SP}. Their degrees of freedom are in `DF`.

You use `SS[4]/DF[4]` as the error MS to test A main effects.

You use `SS[7]/DF[7]` as the error MS to test B main effects and AB interaction.

When **y** is N by p is (*multivariate*), you would do a split plot *MANOVA* by

```
manova("y=reps+a+E(reps.a)+b+a.b")
```

Variable `SS` will contain hypotheses Matrices **H** for `CONSTANT` (`SS[1,,]`) **reps** (`SS[2,,]`), **a** (`SS[3,,]`), **b** (`SS[5,,]`), and **a.b** (interaction, (`SS[6,,]`)) and two error matrices $\mathbf{E}_{WP}$ = `SS[4]` and $\mathbf{E}_{SP}$ = `SS[7]`.

You use `SS[3,,] for` **H** and `SS[4,,]` for **E** with f_h = `DF[3]` and f_e = `DF[4]` to test A mean effects.

You use `SS[7,,]` for **E** with f_e = `DF[7]` to test B main effects and AB interaction.

For every hypothesis, you have the full range of tests, all based on some comparison of an **H** and an **E** --

- Bonferronized F comparing diagonals h_{ii} and e_{ii}
- Roy's maximum root
- Wilks' likelihood ratio
- Hotelling's T_0^2 (Hotelling's trace test)
- Pillai's V (Pillai's trace test).

But there are other possible tests as well.

Bonferronize T^2-tests

Example: In the g group situation,

$$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_g$$

is equivalent to the g-1 hypotheses

$$\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = 0,\ \boldsymbol{\mu}_1 - \boldsymbol{\mu}_3 = 0,\ \ldots,\ \boldsymbol{\mu}_1 - \boldsymbol{\mu}_g = 0$$

each of which you can test by

$$T_{1j}^2 = (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_j)'(\hat{V}[\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_j])^{-1}(\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_j)$$

where $\hat{V}[\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_j] = (1/n_1+1/n_j)\mathbf{S}$, $\mathbf{S} = \mathbf{E}/f_e$.

Note that this uses $\mathbf{S}$, an estimate of $\boldsymbol{\Sigma}$ pooling over *all* g groups, not just groups 1 and j.

You Bonferronize using factor g-1, that is find critical values using $\alpha' = \alpha/(g-1)$.

Or, in the spirit of multiple comparisons, you could Bonferronize *all* g(g-1)/2 statistics T_{ij}^2, $\leq i < j \leq g$ testing $\boldsymbol{\mu}_i = \boldsymbol{\mu}_j$. Bonferronized critical values use $\alpha' = \alpha/\{g(g-1)/2\}$.

Bonferronize univariate t-tests

Since an alternative to a T^2 is to Bonferronize p t-tests, you could Bonferronize p×(g-1) t-tests based on

$$t_{1jk}=(\overline{x}_{k1} - \overline{x}_{kj})/\hat{S}E[\overline{x}_{k1} - \overline{x}_{kj}],$$
$$j = 2,\ldots,g,\ k = 1,\ldots,p$$

where $\hat{S}E[\overline{x}_{k1} - \overline{x}_{kj}] = \sqrt{\{s_{kk}(1/n_1+1/n_j)\}}$

Or you could Bonferronize all p×g×(g-1)/2 t statistics t_{ijk}, $1 \leq i < j \leq g$, k = 1,...,p.

There are a *lot* of options.

Bonferronizing t or T^2 is more interpretable, but can lose power, especially if there is high correlation.

When there is one variable which strongly violate H_0 and for the other variables, H_0 is (nearly) true, Bonferronizing F-tests or even t-tests may have good power.

Choosing a MANOVA test

The Wilks (LR), Hotelling trace, and Pillai trace tests are all general, having a completely unspecified alternative H_1. Moreover, they tend to have similar behavior and give similar conclusions so it's hard to come up with good reasons for preferring one over another.

Roy's is also general, but has best power when the alternative is one dimensional - all means or effects tested are close to a straight line in p-dimensional space.

When you have prior information about which alternative hypotheses are likely, you can sometimes get tests with higher power than the general tests.

For example, suppose you believe that H_0 might be substantially false for specific linear combinations $y_{\mathbf{u}_j} \equiv \mathbf{u}_j'\mathbf{y}$ for one or more $\mathbf{u}_j$'s. Then you might include these $\mathbf{y}_{\mathbf{u}_j}$'s among a larger set $\mathbf{u}_1'\mathbf{y}$, $\mathbf{u}_2'\mathbf{y}$, ... of linear combinations to be analyzed by Bonferronized F-tests.

Example

In a repeated measures one-way MANOVA case when p = 5, suppose you believe that the means of linear combination

$$y_{\mathbf{u}} = \mathbf{u}'\mathbf{y} = -2y_1 - y_2 + y_4 + 2y_5$$

differed greatly among groups. That is

$\mu_{1u} = -2\mu_{11} - \mu_{21} + \mu_{41} + 2\mu_{51}$ (group 1 value)

$\mu_{2u} = -2\mu_{12} - \mu_{22} + \mu_{42} + 2\mu_{52}$ (group 2 value)

. . . .,

$\mu_{gu} = -2\mu_{1g} - \mu_{2g} + \mu_{4g} + 2\mu_{5g}$ (group g value)

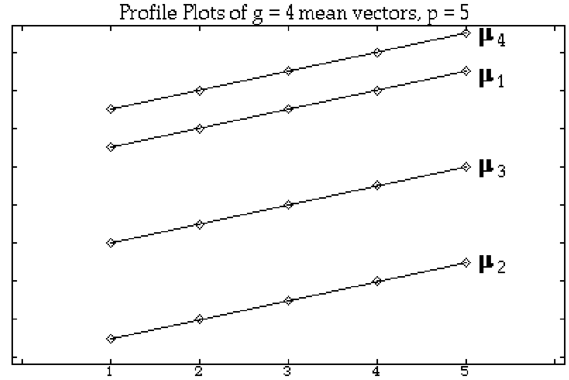
are very different ($\mathbf{u}$ = [-2,-1, 0, 1, 2]).

Then an F-test computed from the values of $y_{\mathbf{u}}$ might have high power, even when Bonferronized because you test other linear combinations such as

$$\mathbf{1}_5'\mathbf{y} = y_1 + y_2 + y_3 + y_4 + y_5.$$

You might be tempted to restrict testing to just this one linear combination, so you didn't need to Bonferronize. You would run the risk of having no power if you were wrong and $\mathbf{u}'\boldsymbol{\mu}_1 \cong \mathbf{u}'\boldsymbol{\mu}_2 \cong \ldots \cong \mathbf{u}'\boldsymbol{\mu}_g$, even though the $\boldsymbol{\mu}_j$'s were very different. For example, suppose the profile plots of the mean vectors were like this:

Profile Plots of g = 4 mean vectors, p = 5



For these $\mathbf{u}'\boldsymbol{\mu}_j$ = 0, j = 1,2,3,4 but $\mathbf{1}_5'\boldsymbol{\mu}_1$, $\mathbf{1}_5'\boldsymbol{\mu}_2$, $\mathbf{1}_5'\boldsymbol{\mu}_3$ and $\mathbf{1}_5'\boldsymbol{\mu}_4$ differ.

MANOVA CANONICAL VARIABLES

Eigenvalues $\hat{\lambda}_i$ of $\mathbf{H}$ *relative* to $\mathbf{E}$ tell us about the relative sizes of $\mathbf{H}$ and $\mathbf{E}$.

As computed by MacAnova, $\hat{\mathbf{u}}_i$ & $\hat{\lambda}_i$ satisfy

$\mathbf{E}^{-1}\mathbf{H}\hat{\mathbf{u}}_i = \hat{\lambda}_i\hat{\mathbf{u}}_i$	$\hat{\mathbf{u}}_i'\mathbf{H}\hat{\mathbf{u}}_i = \hat{\lambda}_i$	$\hat{\mathbf{u}}_i'\mathbf{E}\hat{\mathbf{u}}_i = 1$
$\mathbf{H}\hat{\mathbf{u}}_i = \hat{\lambda}_i\mathbf{E}\hat{\mathbf{u}}_i$	$\hat{\mathbf{u}}_i'\mathbf{H}\hat{\mathbf{u}}_j = 0,\ i \neq j$	$\hat{\mathbf{u}}_i'\mathbf{E}\hat{\mathbf{u}}_j = 0,\ i \neq j$

The $\hat{\mathbf{u}}_i$'s are part of the definition of ***MANOVA canonical variables***:

$$\hat{z}_1 = \hat{\mathbf{u}}_1'\mathbf{y},\ \hat{z}_2 = \hat{\mathbf{u}}_2'\mathbf{y},\ \ldots,\ \hat{z}_p = \hat{\mathbf{u}}_p'\mathbf{y}$$

Each $\hat{z}_j = \sum_{1 \leq \ell \leq p} \hat{u}_{\ell j} y_\ell$ is a *linear combination* of the original variables

- $SS_h(\hat{z}_i) = \hat{\mathbf{u}}_i'\mathbf{H}\hat{\mathbf{u}}_i = \hat{\lambda}_i$
- $SS_e(\hat{z}_i) = \hat{\mathbf{u}}_i'\mathbf{E}\hat{\mathbf{u}}_i = 1$
- $f_e\hat{\lambda}_i/f_h = (SS_h(\hat{z}_i)/f_h)/(SS_e(\hat{z}_i)/f_e)$ = ANOVA F-statistic computed from $\hat{z}_i$
- $\hat{\mathbf{u}}_i'\mathbf{E}\hat{\mathbf{u}}_j = 0,\ i \neq j \Rightarrow \hat{z}_i$ and $\hat{z}_j$ have estimated within-group correlation 0.

Reminder:

$\mathbf{H}$ is a hypothesis matrix for a *specific* null hypothesis H_0. This means the relative eigenvalues and eigenvectors $\hat{\lambda}_i$ and $\hat{\mathbf{u}}_i$ are specific to H_0.

When you test a different H_0 on the basis of the same data the relative eigenvalues and vectors *and MANOVA canonical variables* are different.

For example in a two factor experiment with main effect terms for A and B and interaction effect AB, you would have three sets of canonical variables, $\hat{z}$'s computed from $\mathbf{H}_A$ and $\mathbf{E}$, $\hat{z}$'s computed from $\mathbf{H}_B$ and $\mathbf{E}$, and $\hat{z}$'s computed from $\mathbf{H}_{AB}$ and $\mathbf{E}$.

Properties:

- $F_{\hat{z}_1} = f_e\hat{\lambda}_1/f_h$ is the largest possible F-statistic F_u of *any* linear combination **u'y**. It is *not* distributed as F.
- $F_{\hat{z}_2}$ is largest F_u based on **u** for which **u'y** is *uncorrelated* with $\hat{z}_1$.
- $F_{\hat{z}_3}$ is largest F_u based on **u** for which **u'y** is *uncorrelated* with $\hat{z}_1$ and $\hat{z}_2$.
- And so on.

Thus $\hat{z}_1$ is the linear combination for which the null hypothesis appears to be ***most violated***.

$\hat{z}_2$ is the linear combination uncorrelated with $\hat{z}_1$ that most violates H_0, and so on.

Examination of the canonical variables may help find ways in which H_0 is false, just as finding the variable y_j with the largest F does.

Statistics 5401 Lecture 20 October 21, 2005

Example using Fisher iris data.

```
Cmd> manova("y=varieties", silent:T)
Cmd> h <- SS[2,,]; e <- SS[3,,] # H and E
Cmd> fh <- DF[2]; fe <- DF[3] # hypothesis and error D.F.
Cmd> eigs <- releigen(h,e) # eigs is structure
Cmd> eigs$values # last 2 are essentially 0
(1)      32.192     0.28539  7.5171e-15  1.7978e-16
Cmd> u1hat <- eigs$vectors[,1]; z1 <- y %*% u1hat # Can Var 1
Cmd> u2hat <- eigs$vectors[,2]; z2 <- y %*% u2hat # Can Var 2
Cmd> u3hat <- eigs$vectors[,3]; z3 <- y %*% u3hat # Can Var 3
Cmd> # y is 150 by 4, each zi is 150 by 1
```

There are $s = \min(2,4) = 2$ non-zero $\hat{\lambda}_i$.

```
Cmd> plot(z1,z2,symbols:varieties,\
     title:"Fisher iris data canonical variable plot",\
     xlab:"z1_hat = canonical variable 1",\
     ylab:"CanVar 2",yaxis:F)
```