Today’s topics

- Manova examples
  - Distinguishing among psychiatric groups
  - Robust MLMs
- Multivariate regression
  - PA tests & ability
  - Canonical correlation
  - MANCOVA & homogeneity of regression
- Homogeneity of variance

---

**Ex: Neuro- & Social-Cognitive measures in psychiatric groups**

- A study by Leah Hartman @York examined whether patients classified as ‘schizophrenic’ or ‘schizoaffective’ (on DSM-IV) could be distinguished from a normal, control sample on standardized tests in the following domains:
  - **Neuro-Cognitive**: processing speed, attention, verbal learning, visual learning, problem solving
  - **Social-cognitive**: managing emotions, theory of mind, externalizing bias, personalizing bias

- Research questions → MANOVA contrasts
  - Analyze neuro-cog (NC) and social-cog (SC) separately
  - Do the two psychiatric groups differ from the controls?
  - Do the psychiatric groups differ from each other?

  ```r
  library(heplots)
  library(candisc)
  data(NeuroCog, package="heplots")
  # fit the MANOVA model, test hypotheses
  NC.mlm <- lm(cbind(Speed, Attention, Memory, Verbal, Visual,ProbSolv) ~ Dx, data=NeuroCog)
  Anova(NC.mlm)
  Type II MANOVA Tests: Pillai test statistic
  Df test stat approx F num Df den Df Pr(>F)
  Dx 2 0.2992 6.8902 12 470 1.562e-11 ***
  ---
  Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
  
  So, the groups differ. But how?
  What about the research hypotheses?
  ```
### Neuro-cognitive measures

A simple result: Control (Schizophrenia | Schizoaffective)

```r
> print(linearHypothesis(NC.mlm, "Dx1"), SSP=FALSE)
```

Multivariate Tests:

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai</td>
<td>1</td>
<td>0.289</td>
<td>15.9</td>
<td>6</td>
<td>234</td>
<td>2.8e-15 ***</td>
</tr>
<tr>
<td>Wilks</td>
<td>1</td>
<td>0.711</td>
<td>15.9</td>
<td>6</td>
<td>234</td>
<td>2.8e-15 ***</td>
</tr>
<tr>
<td>Hotelling-Lawley</td>
<td>1</td>
<td>0.407</td>
<td>15.9</td>
<td>6</td>
<td>234</td>
<td>2.8e-15 ***</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```r
> print(linearHypothesis(NC.mlm, "Dx2"), SSP=FALSE)
```

Multivariate Tests:

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai</td>
<td>1</td>
<td>0.006</td>
<td>0.249</td>
<td>6</td>
<td>234</td>
<td>0.96</td>
</tr>
<tr>
<td>Wilks</td>
<td>1</td>
<td>0.994</td>
<td>0.249</td>
<td>6</td>
<td>234</td>
<td>0.96</td>
</tr>
<tr>
<td>Hotelling-Lawley</td>
<td>1</td>
<td>0.006</td>
<td>0.249</td>
<td>6</td>
<td>234</td>
<td>0.96</td>
</tr>
<tr>
<td>Roy</td>
<td>1</td>
<td>0.006</td>
<td>0.249</td>
<td>6</td>
<td>234</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Wow! All neuro-cog measures highly correlated in group means!
Only 1 dim. of $H$ variation

### Visualize me: in data space

- Bivariate view for any 2 responses:
  ```r
  heplot(NC.mlm, var=1:2, ...)
  ```
- HE plot matrix: for all responses
  ```r
  pairs(NC.mlm, ...)
  ```

### Visualize me: in canonical space

- Canonical HE plot is just the HE plot of canonical scores, $(z_1, z_2)$ in 2D,
or, $z_1, z_2, z_3$, in 3D.
- As in biplot, we add vectors to show relations of the $y_i$ response variables
to the canonical variates.
- Variable vectors here are structure coefficients = correlations of variables
  with canonical scores.

Very simple interpretation

Can1: normal vs. others
All vars highly correlated;
Can2: only 1.5%, NS; but perhaps suggestive

### Visualize me: canonical HE plots

- Canonical HE plot is just the HE plot of canonical scores, $(z_1, z_2)$ in 2D,
or, $z_1, z_2, z_3$, in 3D.
- As in biplot, we add vectors to show relations of the $y_i$ response variables
to the canonical variates.
- Variable vectors here are structure coefficients = correlations of variables
  with canonical scores.

The multivariate “juicer”
Shows just group means, $H$ ellipse & $E$ ellipse
Variable vectors offer interpretation of Can dimensions.
Social cognitive measures

- These measures deal with the person’s perception and cognitive processing of emotions of others
  - Scales: managing emotions, theory of mind, externalizing bias, personalizing bias
- Questions:
  - Do these differentiate normal from patient groups?
  - Can they distinguish between schizophrenic & schizoaffective
  - If so, this could be a major finding.

```r
> data(SocialCog, package="heplots")
> SC.ml <- lm(cbind(MgeEmotions, ToM, ExtBias, PersBias) ~ Dx, data=SocialCog)
> Anova(SC.ml)

Type II MANOVA Tests: Pillai test statistic

<table>
<thead>
<tr>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dx 2</td>
<td>0.212</td>
<td>3.97</td>
<td>8</td>
<td>268</td>
<td>0.00018 ***</td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Test contrasts: Dx1 = Normal vs. Patient; Dx2 = Schizo vs. Schizoaffective

```r
> print(linearHypothesis(SC.ml, 'Dx1'), SSP=FALSE)

Multivariate Tests:

<table>
<thead>
<tr>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai</td>
<td>0.1355</td>
<td>5.212</td>
<td>4</td>
<td>133</td>
<td>0.000624 ***</td>
</tr>
</tbody>
</table>

> print(linearHypothesis(SC.ml, 'Dx2'), SSP=FALSE)

Multivariate Tests:

<table>
<thead>
<tr>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai</td>
<td>0.0697</td>
<td>2.493</td>
<td>4</td>
<td>133</td>
<td>0.0461 *</td>
</tr>
</tbody>
</table>

Visualize me: canonical space

Contrasts:
Dx1: Control vs. patients. Controls > patients on MgeEmotions, ExtBias, ToM
Dx2 : Schizo vs. schizoaffective.

Can1: group order
Can2: Schizoaffective vs. others
**Model checking & remedies**

- The MLM assumes residuals are multivariate normal
  - → Squared Mahalanobis distances
    \[ D_{M}^{2}(y) = (y_{i} - \bar{y})^T S^{-1} (y_{i} - \bar{y}) \sim \chi^2_p \text{ with p d.f.} \]
  - → a quantile – quantile plot of ordered \( D_{M}^{2}(y_{i}) \) vs. quantiles of \( \chi^2_p \) should plot as straight line
  - Outliers are readily apparent
  - plots: heplots::cqplot()

- Influence plots
  - mvinfluence::influence() calculates multivariate analogs of influence measures

---

**Social cog: cqplot**

heplots::cqplot() creates a chi-square QQ plot from a MLM

\[ cqplot(SC.mlm, id.n=2) \]

One observation appears as an extreme outlier.

This was a case w/ ExtBias = -33, but valid range = (-10, +10)

Refitting w/o case 15:
- Overall & DX1 tests still OK ☺
- Dx2 test: \( p=0.074 \), now NS ☹

---

**Social cog: Influence**

mvinfluence::influencePlot() creates a multivariate analog of an influence plot

\[ \text{library(mvinfluence)} \]
\[ \text{> influencePlot(SC.mlm, ...)} \]

Case 15 stands out as hugely influential

The 3 columns of circles correspond to the 3 groups

---

**Robust MLMs**

- Robust methods for univariate LMs are now well-developed and implemented
  - → proper SEs, CIs and hypothesis tests
- Analogous methods for multivariate LMs are a current hot research topic
- The heplots package now provides robmlm() for the fully general MLM (MANOVA, MMReg)
  - Uses simple M-estimator via IRLS
  - Weights: calculated from Mahalanobis \( D^2 \), a robust covariance estimator and weight function, \( \psi(D^2) \)
    \[ D^2 = (Y - \bar{Y})^T S_{\text{robust}}^{-1} (Y - \bar{Y}) \sim \chi^2_p \]
  - Downside: SEs, \( p \)-values only approximate
Robust MLMs

SC.rlm <- robmlm(cbind(MgeEmotions, ToM, ExtBias, PersBias) ~ Dx, data=SocialCog)

Approx test of Dx2 in robust model

> print(linearHypothesis(SC.rlm, "Dx2"), SSP=FALSE)

Multivariate Tests:

<table>
<thead>
<tr>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai</td>
<td>1</td>
<td>0.069</td>
<td>2.44</td>
<td>4</td>
<td>132</td>
</tr>
<tr>
<td>Wilks</td>
<td>1</td>
<td>0.931</td>
<td>2.44</td>
<td>4</td>
<td>132</td>
</tr>
</tbody>
</table>

Observation weights Approx test of Dx2 in robust model

Robust MLMs: Pottery data

> pottery.mod <- lm(cbind(Al,Fe,Mg,Ca,Na)~Site, data=Pottery)
> pottery.rmod <- robmlm(cbind(Al,Fe,Mg,Ca,Na)~Site, data=Pottery)

Observation weights overlaid HE plots

MMRA example: PA tasks & ability

- Rohwer data from Timm (1975)
- How well do paired associate (PA) tasks predict performance on measures of aptitude & achievement in kindergarten children?
  - Samples: 69 children in two groups (schools): ‘Lo’ | ‘Hi’ SES
  - Outcomes (Y):
    - Scholastic aptitude test (SAT)
    - Peabody picture vocabulary test (PPVT)
    - Raven progressive matrices (Raven)
  - Predictors (X): Scores (0—40) on PA tasks where the stimuli were:
    - named (n), still (s), named-still (ns), named-action (na), sentence-still (ss)

> data("Rohwer", package="heplots")
> car::some(Rohwer, n=5)

<table>
<thead>
<tr>
<th>group</th>
<th>SES</th>
<th>SAT</th>
<th>PPVT</th>
<th>Raven</th>
<th>n</th>
<th>s</th>
<th>ns</th>
<th>na</th>
<th>ss</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Lo</td>
<td>68</td>
<td>68</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>19</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>Lo</td>
<td>49</td>
<td>74</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>17</td>
<td>Lo</td>
<td>19</td>
<td>66</td>
<td>13</td>
<td>7</td>
<td>12</td>
<td>21</td>
<td>35</td>
<td>27</td>
</tr>
<tr>
<td>52</td>
<td>Hi</td>
<td>38</td>
<td>66</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>66</td>
<td>Hi</td>
<td>8</td>
<td>55</td>
<td>16</td>
<td>4</td>
<td>7</td>
<td>19</td>
<td>20</td>
<td>13</td>
</tr>
</tbody>
</table>

> Rohwer2 <- subset(Rohwer, subset=SES=="Hi")

Having a group factor makes the analysis more complicated (MANCOVA)

Start with analysis of the Hi SES group

Why not univariate models?

rohwer.mod1 <- lm(SAT ~ n + s + ns + na + ss, data = Rohwer2)
rohwer.mod2 <- lm(PPVT ~ n + s + ns + na + ss, data = Rohwer2)
rohwer.mod3 <- lm(Raven ~ n + s + ns + na + ss, data = Rohwer2)
library(stargazer)
stargazer(rohwer.mod1, rohwer.mod2, rohwer.mod3, type="text", ...)

Univariate regression models for Rohwer data

<table>
<thead>
<tr>
<th>Dependent variable:</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT</td>
</tr>
<tr>
<td>n</td>
</tr>
<tr>
<td>s</td>
</tr>
<tr>
<td>ns</td>
</tr>
<tr>
<td>na</td>
</tr>
<tr>
<td>ss</td>
</tr>
</tbody>
</table>

F Statistic (df = 5; 26) 0.56 0.35 0.31

Note: *p<0.05; **p<0.01; ***p<0.001

Results are disappointing

- Only model for SAT highly signif.
- Only a few coeffs. signif. ≠0
**MANOVA tests**

```r
> rohwer.mlm <- lm(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss, data=Rohwer2)
> Anova(rohwer.mlm)
```

Type II MANOVA Tests: Pillai test statistic

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt; F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>1</td>
<td>0.202</td>
<td>2.02</td>
<td>3</td>
<td>24</td>
<td>0.1376</td>
</tr>
<tr>
<td>s</td>
<td>1</td>
<td>0.310</td>
<td>3.59</td>
<td>3</td>
<td>24</td>
<td>0.0284 *</td>
</tr>
<tr>
<td>ns</td>
<td>1</td>
<td>0.358</td>
<td>4.46</td>
<td>3</td>
<td>24</td>
<td>0.0126 *</td>
</tr>
<tr>
<td>na</td>
<td>1</td>
<td>0.465</td>
<td>6.96</td>
<td>3</td>
<td>24</td>
<td>0.0016 **</td>
</tr>
<tr>
<td>ss</td>
<td>1</td>
<td>0.089</td>
<td>0.78</td>
<td>3</td>
<td>24</td>
<td>0.5173</td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Can also test overall hypothesis, H0: B = 0 (all coefs = 0)

```r
> print(linearHypothesis(rohwer.mlm, +                        c("n", "s", "ns", "na", "ss")), SSP=FALSE)
```

Multivariate Tests:

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt; F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai</td>
<td>5</td>
<td>1.0386</td>
<td>2.753</td>
<td>15</td>
<td>68.655</td>
<td>0.001192 **</td>
</tr>
<tr>
<td>Wilks</td>
<td>5</td>
<td>0.2431</td>
<td>2.974</td>
<td>15</td>
<td>66.650</td>
<td>0.001154 **</td>
</tr>
<tr>
<td>Hotelling-Lawley</td>
<td>5</td>
<td>2.0615</td>
<td>3.115</td>
<td>15</td>
<td>68.000</td>
<td>0.000697 ***</td>
</tr>
<tr>
<td>Roy</td>
<td>5</td>
<td>1.4654</td>
<td>7.620</td>
<td>5</td>
<td>26.000</td>
<td>0.000180 ***</td>
</tr>
</tbody>
</table>

Strongly reject H0 by all criteria

**Visualize me!**

```r
cols <- c("red", "blue", "black", "darkgreen", "darkcyan", "magenta", "gray20")
hyp <- list("Regr" = c("n", "s", "ns", "na", "ss"))  # Test of B = 0
heplot(rohwer.mlm, hypotheses = hyp, fill=TRUE, fill.alpha=0.1, col=cols,lwd=c(1,3))
```

Each predictor gives a 1 df test -> H ellipse is a line E here is a 3D ellipsoid (rank(E) = min(p,q))

Interpretation:

- Any H ellipse that protrudes outside E ellipse is significant by Roy’s test
- Length of each H line = strength of association
- Orientation of each H line shows relation of Xi to the two Yj responses shown.

**pairs.mlm() plot**

View all pairwise HE plots

```r
pairs(rohwer.mlm, hypotheses=hyp, col=cols, fill=TRUE, fill.alpha=0.1, ...)
```

Can now understand more subtle aspects

SAT is best predicted overall, but relation with PA tests varies
The na & ns tasks are strongest for SAT
Raven is weakly predicted

**Canonical correlations**

For quantitative (X, Y) data, canonical correlation analysis is an alternative to MMRA
It finds the weighted sums of the Y variables most highly correlated with the Xs

```r
> X <- Rohwer2[, 6:10]  # X variables for High SES students
> Y <- Rohwer2[, 3:5]   # Y variables for High SES students
> (cc <- cancor(X, Y, set.names=c("PA", "Ability")))
```

<table>
<thead>
<tr>
<th></th>
<th>CanR</th>
<th>CanRSQ</th>
<th>Eigen percent</th>
<th>cum percent</th>
<th>scree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7710</td>
<td>0.5944</td>
<td>1.4654</td>
<td>71.080</td>
<td>71.08 ******************************</td>
</tr>
<tr>
<td>2</td>
<td>0.5465</td>
<td>0.2987</td>
<td>0.4259</td>
<td>20.659</td>
<td>91.74 *********</td>
</tr>
<tr>
<td>3</td>
<td>0.3815</td>
<td>0.1455</td>
<td>0.1703</td>
<td>8.261</td>
<td>100.00 ***</td>
</tr>
</tbody>
</table>

Test of H0: The canonical correlations in the current row and all that follow are zero

<table>
<thead>
<tr>
<th></th>
<th>CanR LR test stat approx F numDF denDF Pr(&gt; F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.77096         0.24307       2.9738           15 66.655</td>
</tr>
<tr>
<td>2</td>
<td>0.54652         0.59926       1.8237           8 50.000</td>
</tr>
<tr>
<td>3</td>
<td>0.18147         0.85444       1.4759           3 26.000</td>
</tr>
</tbody>
</table>

Two dimensions acct for 91.7% of (X,Y) association

Only Can1 is significant
**Visualize CCA in HE plot**

```r
cols <- c("red", "blue", "black", "darkgreen", "darkcyan", "magenta", "gray20")
heplot(cc, hypotheses=list("na+ns"=c("na", "ns")), fill = TRUE, fill.alpha=0.1, col=cols, label.pos = c(3, rep(1,5), .1), cex=1.4, var.cex=1.25, var.lwd=3, var.col="black")
```

Residuals are uncorrelated in canonical space

- **H** ellipses for X terms same as in ordinary HE plots – outside E ellipse iff signif. by Roy’s test
- Variable vectors for Ys: correlations with canonical variables Ycan1, Ycan2
  - SAT & PPVT: mainly Ycan1
  - Raven: more aligned with Ycan2

**MANCOVA & homogeneity of regression**

- With a group variable (SES) can test differences in means (intercepts)
  - `rohwer.mod <- lm(cbind(SAT, PPVT, Raven) ~ SES + n + s + ns + na + ss, data=Rohwer)`
  - This assumes that slopes (B) are the same for both groups (homogeneity of regression)
- Can test for equal slopes by adding interactions of SES with Xs
  - `rohwer.mod1 <- lm(cbind(SAT, PPVT, Raven) ~ SES * (n + s + ns + na + ss))`
- Or, fit separate models for each group
  - `rohwer.ses1 <- lm(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss, data = Rohwer, subset = SES == "Hi")`
  - `rohwer.ses2 <- lm(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss, data = Rohwer, subset = SES == "Lo")`

**MANCOVA**

Fit the MANCOVA model & test hypotheses

```r
> rohwer.mod <- lm(cbind(SAT, PPVT, Raven) ~ SES + n + s + ns + na + ss, + data=Rohwer)
> Anova(rohwer.mod)
```

**Type II MANOVA Tests: Pillai test statistic**

<table>
<thead>
<tr>
<th></th>
<th>DF test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SES</td>
<td>0.379</td>
<td>12.18</td>
<td>3</td>
<td>60</td>
<td>3.7e-06 ***</td>
</tr>
<tr>
<td>n</td>
<td>0.040</td>
<td>0.84</td>
<td>3</td>
<td>60</td>
<td>0.4773</td>
</tr>
<tr>
<td>s</td>
<td>0.093</td>
<td>2.04</td>
<td>3</td>
<td>60</td>
<td>0.1173</td>
</tr>
<tr>
<td>ns</td>
<td>0.193</td>
<td>4.78</td>
<td>3</td>
<td>60</td>
<td>0.0047 **</td>
</tr>
<tr>
<td>na</td>
<td>0.231</td>
<td>6.02</td>
<td>3</td>
<td>60</td>
<td>0.0012 **</td>
</tr>
<tr>
<td>ss</td>
<td>0.050</td>
<td>1.05</td>
<td>3</td>
<td>60</td>
<td>0.3770</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Visualize effects**

```r
pairs(rohwer.mod, hypotheses=list("Regr" = c("n", "s", "ns", "na", "ss")), fill=TRUE, fill.alpha=0.1)
```

The SES effect is positive for all Y variables

Hi SES group > Lo SES group
Fit model with interactions

Fit heterogeneous regression model with SES interactions

```r
> rohwer.mod1 <- lm(cbind(SAT, PPVT, Raven) ~ SES * (n + s + ns + na + ss),
>                    data=Rohwer)
> Anova(rohwer.mod1)
```

Type II MANOVA Tests: Pillai test statistic

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SES</td>
<td>1</td>
<td>0.391</td>
<td>11.78</td>
<td>3</td>
<td>55</td>
<td>4.5e-06 ***</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>0.079</td>
<td>1.57</td>
<td>3</td>
<td>55</td>
<td>0.20638</td>
</tr>
<tr>
<td>s</td>
<td>1</td>
<td>0.125</td>
<td>2.62</td>
<td>3</td>
<td>55</td>
<td>0.05952 ***</td>
</tr>
<tr>
<td>ns</td>
<td>1</td>
<td>0.254</td>
<td>6.25</td>
<td>3</td>
<td>55</td>
<td>0.00100 ***</td>
</tr>
<tr>
<td>na</td>
<td>1</td>
<td>0.307</td>
<td>8.11</td>
<td>3</td>
<td>55</td>
<td>0.00015 ***</td>
</tr>
<tr>
<td>ss</td>
<td>1</td>
<td>0.060</td>
<td>1.17</td>
<td>3</td>
<td>55</td>
<td>0.32813</td>
</tr>
<tr>
<td>SES:n</td>
<td>1</td>
<td>0.142</td>
<td>3.43</td>
<td>3</td>
<td>55</td>
<td>0.00010 **</td>
</tr>
<tr>
<td>SES:s</td>
<td>1</td>
<td>0.099</td>
<td>2.02</td>
<td>3</td>
<td>55</td>
<td>0.12117</td>
</tr>
<tr>
<td>SES:ns</td>
<td>1</td>
<td>0.118</td>
<td>2.44</td>
<td>3</td>
<td>55</td>
<td>0.07383 .</td>
</tr>
<tr>
<td>SES:na</td>
<td>1</td>
<td>0.148</td>
<td>3.18</td>
<td>3</td>
<td>55</td>
<td>0.03081 *</td>
</tr>
<tr>
<td>SES:ss</td>
<td>1</td>
<td>0.057</td>
<td>1.12</td>
<td>3</td>
<td>55</td>
<td>0.35094</td>
</tr>
<tr>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signif. codes:</td>
<td>0 ‘<em><strong>’ 0.001 ‘</strong>’ 0.01 ‘</em>’ 0.05 ‘.’ 0.1 ‘ ’ 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OK, as expected

Hmm ???

Test interactions

Can test all interactions simultaneously with linearHypothesis()

Do I need any interaction terms?

I use ‘grep’ trick here to find the names of coefficients like ‘SES:’ containing a ‘:

```r
> coefs <- rownames(coef(rohwer.mod1))  # store coefficient names in a vector
> linearHypothesis(rohwer.mod1,    # only test for interaction effects
>                          c
> coefs[.grep(":", coefs)], SSP=FALSE)
```

Multivariate Tests:

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai</td>
<td>5</td>
<td>0.4179</td>
<td>1.845</td>
<td>15</td>
<td>171.0</td>
<td>0.03209 *</td>
</tr>
<tr>
<td>Wilks</td>
<td>5</td>
<td>0.6236</td>
<td>1.894</td>
<td>15</td>
<td>152.2</td>
<td>0.02769 *</td>
</tr>
<tr>
<td>Hotelling-Lawley</td>
<td>5</td>
<td>0.5387</td>
<td>1.927</td>
<td>15</td>
<td>161.0</td>
<td>0.02396 *</td>
</tr>
<tr>
<td>Roy</td>
<td>5</td>
<td>0.3846</td>
<td>4.385</td>
<td>5</td>
<td>57.0</td>
<td>0.00019 **</td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Evidence shows that some slopes differ for Hi/Lo SES

Fit separate models

Fitting a model for each group allows all slopes to differ
Also allows within-group covariances to differ

```r
rohwer.ses1 <- lm(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss,
                    data = Rohwer, subset = SES == "Hi")
rohwer.ses2 <- lm(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss,
                    data = Rohwer, subset = SES == "Lo")
```

For SAT & PPVT:
- means higher for Hi SES
- within-group covariance larger for Hi SES
- slopes of predictors smaller for Hi SES → SAT more important for this group.

Homogeneity of (co)variances

- ANOVA assumes equality of residual variances
  \[
  \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_g^2
  \]
  Levine's test: ANOVA of \( z_i = \gamma_i - \overline{y}_i \)

- MANOVA: assumes equality of covariance matrices
  \[
  \Sigma_1 = \Sigma_2 = \ldots = \Sigma_g
  \]
  Box's M test:
  \[
  M = (N-g)\ln |S_0| - \sum_i^{g} (n_i-1) \ln |S_i| \\
  -2 \ln(M) \sim \chi^2_{df}
  \]

See: [http://www.datavis.ca/papers/EqCov-TAS.pdf](http://www.datavis.ca/papers/EqCov-TAS.pdf)
Visualizing covariance matrices

Visualize covariance ellipses in data space
Center to see pure differences in size & shape

```r
covEllipses(iris[,1:2], iris$Species, ...)
covEllipses(iris[,3:4], iris$Species, center=TRUE, ...)
```

In all cases, *setosa* stands out as different from the others
- sometimes correlation differs
- sometimes smaller variance(s)

Visualize in PCA space

PCA projects the data into an orthogonal space accounting for maximum variance
Covariance ellipses show the differences among groups in this space

Surprisingly, the small dimensions contribute largely to Box’s M test.

```r
iris.pca <- prcomp(iris[,1:4])
covEllipses(iris.pca$x, iris$Species, ...)
covEllipses(iris.pca$x, iris$Species, center=TRUE, ...)
covEllipses(iris.pca$x, iris$Species, center=TRUE, variables=3:4, ...)
```

Visualizing Box’s M test

Box’s test is based on a comparison of the log $|S|$ relative to log $|S_p|$: `plot them!`

```r
iris.box <- boxM(iris[,1:4], iris[, "Species"])
plot(iris.box, gplabel="Species")
```

Box’s test is based on a comparison of the log $|S|$ relative to log $|S_p|$: `plot them!`

```r
iris.box <- boxM(iris[,1:4], iris[, "Species"])
plot(iris.box, gplabel="Species")
```

CI: based on an asymptotic CLT = distribution of ln $|S|$ (Cai, Liang, and Zhou 2016)
(Thx: Augustine Wong)

Unsolved: Bootstrap CI
Diabetes data

covEllipses(Diabetes[,2:5], Diabetes$group, fill=TRUE, pooled=FALSE, col=c("blue", "red", "darkgreen"), variables=1:3)

Two groups of diabetic patients and a normal group were measured on blood glucose and insulin.

The differences in correlation and variances are dramatic here.

Summary

- MANOVA tests of MLMs are easily visualized in HE plots
  - Contrasts among groups can be easily shown
  - Canonical plots show data in 2D/3D space of max. group differences
  - Robust methods can help guard against outliers

- MMRA models
  - Visualize effects of quant. predictors as lines in data space
  - Test & visualize any linear hypothesis
  - Canonical correlations: visualize in 2D/3D of max. (X, Y) correlations

- Homogeneity of covariances
  - Visualize within-group $S_i$ and pooled $S_p$ by data ellipses
  - Visualize Box’s M test by simple dot plot of $|S_p|$ and $|S_i|$