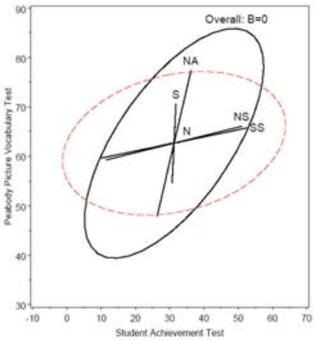
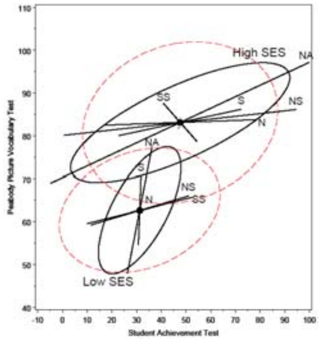
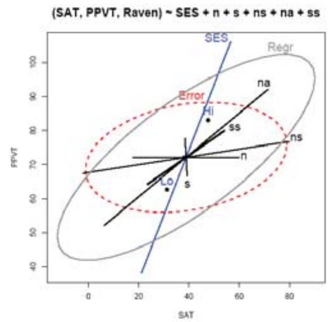
Multivariate multiple regression
& visualizing multivariate tests







Michael Friendly
Psychology 6140

Overview: Univariate & Multivariate Linear Models

	Dependent variables	
Independent variables	*1 Quantitative* **y** = **X** **β**	*2+ Quantitative* **Y** = **X** **B**
Quantitative	Regression	Multivariate regression
Categorical	ANOVA	MANOVA
Both	Reg. w/ dummy vars ANCOVA Homogeneity of regression	General MLM Homogeneity of regression MANCOVA

Today, just multivariate regression, with questions of homogeneity of regression.
Once we learn how to do multivariate tests, extensions to other contexts are easy

Multivariate regression

- When there are several (p>1) criterion variables, we could just fit *p* separate models

$$\left.\begin{aligned} \mathbf{y}_1 &= \mathbf{X}\boldsymbol{\beta}_1 \\ \mathbf{y}_2 &= \mathbf{X}\boldsymbol{\beta}_2 \\ &\cdots \\ \mathbf{y}_p &= \mathbf{X}\boldsymbol{\beta}_p \end{aligned}\right\}$$

```
proc reg;
   model y1-y4 = x1 – x10;
```

```
lm(cbind(y1,y2,y3,y4) ~ x[,1:10])
```

- But this:
  - Does not give simultaneous tests for all regressions
  - Does not take correlations among the y's into account

Why do multivariate tests?

- Avoid multiplying error rates, as in simple ANOVA
  - Overall test for multiple responses-- similar to overall test for many groups: g tests: $\alpha_{all} \approx g\ \alpha$
- Often, multivariate tests are more powerful, when the responses are correlated
  - Small, positively correlated effects can pool power.
  - If responses are uncorrelated, no need for multivariate tests
  - But this is rarely so
- Multivariate tests provide a way to understand the **structure** of relations across separate response measures. In particular:
  - how many "dimensions" of responses are important?
  - how do the predictors contribute to these?

Multivariate regression model

- The multivariate regression model is

$$\begin{bmatrix} \mathbf{y}_1 & \cdots & \mathbf{y}_p \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_q \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_p \end{pmatrix} + \mathbf{E}_{n\times p}$$

cols are coeffs for each criterion

rows, for each predictor

$$\mathbf{Y}_{n\times p} = \mathbf{X}_{n\times q} \ \mathbf{B}_{q\times p} + \mathcal{E}_{n\times p}$$

- The LS solution, $\mathbf{B} = (\mathbf{X'X})^{-1}\,\mathbf{X'Y}$ gives same coefficients as fitting *p* models separately.
- (Omitting here: consideration of model selection for each model)

Example: Rohwer data

- n=32 Lo SES kindergarten kids
- *p*=3 response measures of aptitude/ achievement: SAT, PPVT, Raven
- *q*=5 predictors: PA tests: n, s, ns, na, ss

SAS:

```
proc reg data=lo_ses;
     model sat ppvt raven = n s ns na ss;
 M1: mtest        /* all coeffs = 0 */
```

R:

```
mod1<-lm(cbind(SAT, PPVT, Raven) ~ n+s+ns+na+ss, data=lo_ses)
Manova(mod1)
```


Rohwer data: univariate regressions

- Separate univariate regressions

R^2

	SAT	PPVT	Raven
	0.208	0.511**	0.222

Coefficients:

	SAT	PPVT	Raven
(Intercept)	4.151	33.006	11.173
n	-0.609	-0.081	0.211
s	-0.050	-0.721	0.065
ns	-1.732	-0.298	0.216
na	0.495	1.470*	-0.037
ss	2.248*	0.324	-0.052

Overall tests for each response: H_0: $\boldsymbol{\beta}_i = \mathbf{0}$

Tests for predictors on each response

Publish or perish? Doesn't look like there is much predictive power here!

Rohwer data: multivariate regression

- Yet, the multivariate test is highly significant
  - Overall test for the multivariate model: H_0: $\mathbf{B} = \mathbf{0}$
  - Positive correlations among responses have made this test more powerful – pooling power!

Multivariate Statistics and F Approximations

S=3 M=0.5 N=13.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.343	2.54	15	80.46	0.0039
Pillai's Trace	0.825	2.35	15	93	0.0066
Hotelling-Lawley	1.449	2.72	15	49.77	0.0042
Roy's Max Root	1.055	6.54	5	31	0.0003

Multivariate General Linear Hypothesis (GLH)

- In addition to the overall test, $H_0 : \mathbf{B} = \mathbf{0}$, it is more often desired to test hypotheses about subsets of predictors or linear combinations of coefficients
- The GLH is a single, general method for all such tests

$$H_0 : \mathbf{L}_{r \times q} \quad \mathbf{B}_{q \times p} = \mathbf{0}_{r \times p}$$

where **L** specifies *r* linear combinations of the parameters

```
proc reg data=lo_ses;
   model sat ppvt raven = n s ns na ss;
   M1: mtest;           /* all coeffs = 0 */
   M2: mtest n,s,ns;    /* n,s,ns = 0     */
run;
```


Examples of GLH:

p=2 responses: y_1, y_2

q=3 predictors: $X_1 - X_3$

$$\mathbf{B} = \begin{array}{c} \\ \\ \end{array} \begin{array}{cc} y_1 & y_2 \\ \left(\begin{array}{cc} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{array}\right) & \end{array} = \begin{pmatrix} \boldsymbol{\beta}_0^T \\ \boldsymbol{\beta}_1^T \\ \boldsymbol{\beta}_2^T \\ \boldsymbol{\beta}_3^T \end{pmatrix} \begin{array}{l} \text{Intercept} \\ X_1 \\ X_2 \\ X_3 \end{array}$$

(a) No effect of X_2, X_3

$$H_0 : \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = \mathbf{0}_{2\times1} \Longrightarrow \mathbf{L} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow \mathbf{L\,B} = \begin{pmatrix} \boldsymbol{\beta}_2^T \\ \boldsymbol{\beta}_3^T \end{pmatrix} = \mathbf{0}$$

mtest x2, x3;

(b) Same coef. for X_2, X_3 $\Longrightarrow \mathbf{L} = \begin{pmatrix} 0 & 0 & 1 & -1 \end{pmatrix} \Rightarrow \mathbf{L\,B} = (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_3) = \mathbf{0}$

$$H_0 : \begin{cases} \beta_{21} = \beta_{31} \\ \beta_{22} = \beta_{32} \end{cases}$$

mtest x2-x3;

(Makes sense only if X_2, X_3 are commensurate)

In R, these tests are done with `car::linearHypothesis()`

Extended GLH

- The GLH can be extended to test subsets or linear combinations of coefficients *across the criteria*

$$H_0 : \quad \mathbf{L} \;\; \mathbf{B} \;\; \mathbf{M}_{p \times t} = \mathbf{0}$$

where the post-factor, **M**, specifies *t* linear combs. across criteria

- Previous slide: special case of $\mathbf{M}_{(p x p)} = \mathbf{I}$
- Overall test ($\mathbf{B} = \mathbf{0}$): $\mathbf{L}_{(q x q)} = \mathbf{I}$ and $\mathbf{M}_{(p x p)} = \mathbf{I}$

Example: Coeffs for Y1 = coeffs for Y2

$$\mathbf{L} = \mathbf{I}, \quad \mathbf{M} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \Rightarrow \quad \mathbf{L\,B\,M} = \begin{pmatrix} \beta_{01} - \beta_{02} \\ \beta_{11} - \beta_{12} \\ \beta_{21} - \beta_{22} \\ \beta_{31} - \beta_{32} \end{pmatrix} = \mathbf{0}$$

mtest y1-y2;

(Again, makes sense only if Y1 and Y2 are commensurable)

```
proc reg data=lo_ses;
  model sat ppvt raven = n s ns na ss;
 M3: mtest sat - ppvt;          /* SAT=PPVT */
run;
```

Note: In MANOVA designs:

- **L** specifies a set of contrasts or tests among 'between-group' effects
- **M** specifies contrasts among 'within-subject' effects (e.g., orthogonal polynomials or other within-S comparisons)

Tests of multivariate hypotheses

- In the general linear model, $\mathbf{Y} = \mathbf{X}\,\mathbf{B} + \boldsymbol{\varepsilon}$, all hypotheses are tested in the same way
- Calculate the q x q sum of squares and products matrices

$$\mathbf{H} \equiv SSP_H = (\mathbf{LB})^T [\mathbf{L}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{L}^T]^{-1}(\mathbf{LB})$$

$$\mathbf{E} \equiv SSP_E = \hat{\mathcal{E}}^T \hat{\mathcal{E}}$$

  - Diag elements of **H** & **E** are just the univariate SS
- The multivariate analog of the univariate F-test:

$$F = \frac{MS_H}{MS_E} \rightarrow (MS_H - F \times MS_E) = 0 \text{ is } \boxed{|\mathbf{H} - \lambda\mathbf{E}| = 0}$$

Pause: ANOVA → MANOVA tests

Recall that for a univariate response, the ANOVA table for the test of all predictors, $X_1 – X_p$, H_0: $\boldsymbol{\beta}$=0 looks like the following:

Source	SS	df	MS	*F*
Regression	SSR(X_1, .. X_p)	q	SSR/p	MSR/MSE
Error	SSE	n-q	SSE/n-p	

- The *F* statistic quantifies how big MSR is relative to MSE as evidence against the null hypothesis.
- It is referred to an F distribution with (q, n-q) df to give *p*-values
- The same MSE is used in all tests of sub-hypotheses, e.g., $\beta_1 = \beta_2 = 0$

In MANOVA tests, each SS becomes a *p* x *p* SSP matrix, **H** for the hypothesis, **E** for error. How big is **H** relative to **E**?

TABLE 4.7.2. MANOVA Table for Testing $\boldsymbol{\Gamma} = \mathbf{0}$ for Multivariate Regression Model

Source	df	SSP
α_0	1	$\begin{bmatrix} 36{,}179.7027 & & \\ 72{,}484.4865 & 145{,}219.5676 & \\ 15{,}322.4325 & 30{,}697.8378 & 6489.1892 \end{bmatrix}$
$\Gamma \mid \alpha_0$	5	$\begin{bmatrix} 3{,}653.7732 & & \\ 2{,}159.9966 & 2{,}883.6759 & \\ 63.3680 & 281.4842 & 76.5286 \end{bmatrix} = \mathbf{Q}_h$
Residual	31	Given by (4.7.3)
Total	37	$\begin{bmatrix} 57{,}363.0 & 0 & \\ 76{,}175.0 & 150{,}868.0 & \\ 15{,}843.0 & 31{,}193.0 & 6{,}834.0 \end{bmatrix} = \mathbf{Y}'\mathbf{Y}$

SSR(y_1)

SSR(y_3)

$$\mathbf{Q}_e = \mathbf{Y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}$$

$$= \begin{bmatrix} 13{,}929.5241 & & \\ 1530.5169 & 2764.7565 & \\ 457.1995 & 213.6780 & 268.2822 \end{bmatrix} \begin{array}{l} \text{SAT} \\ \text{PPVT} \\ \text{Raven} \end{array}$$

SSE(y_1) SSE(y_2) SSE(y_3)

Tests of multivariate hypotheses

- All multivariate test statistics are based on latent roots, λ_i of **H** in the metric of **E** (or of $\mathbf{HE}^{-1}$), or latent roots θ_i of $\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1}$
- These measure the "size" of **H** relative to **E** in up to *p* orthogonal dimensions
- Various test statistics differ in how the information is combined across dimensions
  - Wilks' Lambda: product
  - Trace criteria: sum
  - Roy's test: maximum

All ask "How big is **H** relative to **E**"?

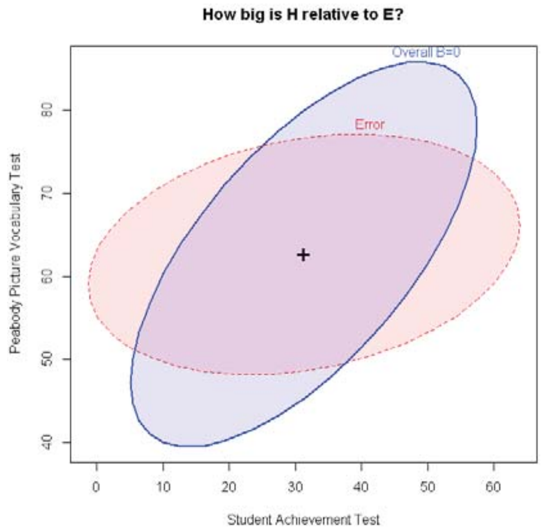
HE plots: Visualizing H & E

HE plots show the H & E matrices as data ellipsoids.

It is difficult to judge naively the size of **H** relative to **E**, but the eigenvalues of $\mathbf{HE}^{-1}$ capture the essential information.

Contributions of $s=\min(p, df_h)$ dimensions can be summarized in different kinds of "means."

As explained later, this plot provides a visual test of significance, based on Roy's test



Multivariate test statistics: overview

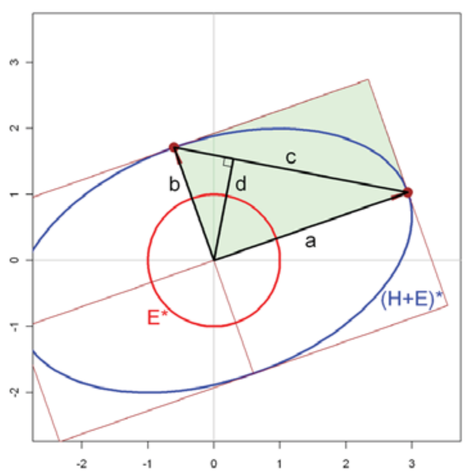
- How big is **H** relative to **E** across one or more dimensions?
- All test statistics can be seen as kinds of means of the $s=\min(p, df_h)$ non-zero eigenvalues of $\mathbf{HE}^{-1}$ or of $\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1}$

Table 1: Multivariate test statistics as functions of the eigenvalues λ_i solving $\det(\boldsymbol{H} - \lambda \boldsymbol{E}) = 0$ or eigenvalues ρ_i solving $\det[\boldsymbol{H} - \rho(\boldsymbol{H}+\boldsymbol{E})] = 0$.

Criterion	Formula	"mean" of ρ	Partial η^2
Wilks' Λ	$\Lambda = \prod_i^s \frac{1}{1+\lambda_i} = \prod_i^s (1-\rho_i)$	geometric	$\eta^2 = 1 - \Lambda^{1/s}$
Pillai trace	$V = \sum_i^s \frac{\lambda_i}{1+\lambda_i} = \sum_i^s \rho_i$	arithmetic	$\eta^2 = \frac{V}{s}$
Hotelling-Lawley trace	$H = \sum_i^s \lambda_i = \sum_i^s \frac{\rho_i}{1-\rho_i}$	harmonic	$\eta^2 = \frac{H}{H+s}$
Roy maximum root	$R = \lambda_1 = \frac{\rho_i}{1-\rho_i}$	supremum	$\eta^2 = \frac{\lambda_1}{1+\lambda_1} = \rho_1$

(This table uses ρ instead of θ for eigenvalues of $\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1}$)

Multivariate test statistics: geometry



Easiest to see if we transform **H** & **E** to "canonical space" where

- $\mathbf{E} \rightarrow \mathbf{E}^* = \mathbf{I}$ (stdized & uncorrelated)
- $(\mathbf{H}+\mathbf{E}) \rightarrow (\mathbf{H}+\mathbf{E})^* = \mathbf{E}^{-1/2} (\mathbf{H}+\mathbf{E}) \mathbf{E}^{-1/2}$
- Allows to focus on just size of $(\mathbf{H}+\mathbf{E})^*$

Then,

- Wilks' Λ ≡ test on area, ~ $(a \times b)^{-1}$
- HLT criterion ~ test on c
- Pillai trace criterion ~ test on d
- Roy's test ~ test on a alone

Multivariate test statistics: details

- n_h = df for hypothesis = # rows of **L**
- n_e = df for error
- $s = \min(n_h, p)$ = # non-zero roots = rank(**H**)
- $\lambda_1, \lambda_2, \ldots, \lambda_s$ = roots of $|\mathbf{H} - \lambda\, \mathbf{E}| = 0$
  $|\mathbf{HE}^{-1} - \lambda\, \mathbf{I}| = 0$
- $\theta_1, \theta_2, \ldots, \theta_s$ = roots of $|\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1} - \lambda\, \mathbf{I}| = 0$

$$\lambda_i = \frac{\theta_i}{1-\theta_i} \qquad \theta_i = \frac{\lambda_i}{1+\lambda_i}$$

Wilks' Lambda: details

- A likelihood-ratio test of H_0: **L B = 0**

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{H}+\mathbf{E}|} = \prod_{i=1}^{s} \frac{1}{1+\lambda_i} = \prod_{i=1}^{s} (1-\theta_i)$$

- Rao's *F* approximation (exact if $s \leq 2$)

$$F = \frac{1-\Lambda^{1/t}}{\Lambda^{1/t}} \times \frac{mt-2k}{pn_h} \sim F(pn_h, mt-2k)$$

NB: df not always integers

- Association: $\eta^2 = 1 - \Lambda^{1/s}$ = geometric mean

Pillai & Hotelling-Lawley trace criteria

- Based on sum (or average) of λ_i or θ_i
- Pillai:

$$V = tr[\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1}] = \sum_{i=1}^{s} \theta_i = \sum \frac{\lambda_i}{1+\lambda_i}$$

$\eta^2 = V/s$

$$F = \frac{2n+s+1}{2m+s+1} \times \frac{V}{s-V} \sim F[s(2m+s+1), s(2n+s+1)]$$

- Hotelling-Lawley:

$$H = tr[\mathbf{HE}^{-1}] = \sum_{i=1}^{s} \lambda_i = \sum \frac{\theta_i}{1-\theta_i}$$

$\eta^2 = H/(H+s)$

$$F = \frac{2(ns+1)}{s^2(2m+s+1)} \times H \sim F[s(2m+s+1), 2(ns+1)]$$

Roy's maximum root test

- Most powerful when there is one large dimension of **H** relative to **E**
- $R = \lambda_1$

$\eta^2 = R/(R+1)$

$$F = \frac{n_e + n_h - s}{s} \sim F(s, n_e + n_h - s)$$

(Exact if s=1)

- Simplicity makes this useful for visual tests of significance in HE plots

Multivariate tests for individual predictors

- H_0 : $\boldsymbol{\beta}_i = \mathbf{0}$ – simultaneous test, for all *p* responses, of predictor $\mathbf{x}_i$
  - $\mathbf{L} = (0, 0, \ldots, 1, 0, \ldots 0)_{(1 \times q)}$ in GLH
  - $\mathbf{H} = \boldsymbol{\beta}_i^{\mathsf{T}} (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1} \boldsymbol{\beta}_i$ – a rank 1 matrix (s=1)
  - All multivariate tests are exact & have the same *p*-values
  - More parsimonious than separate univariate tests for each response.

Example: Rohwer data (SAS)

```
proc reg data=lo_ses;
   model sat ppvt raven = n s ns na ss ;
   Mn:  mtest n;   /* n=0 for all responses */
   Mna: mtest na;  /* na=0 for all responses */
run;
```

Output for NA:

```
 Multivariate Statistics and Exact F Statistics
                       S=1     M=0.5     N=13.5

Statistic           Value   F Value   Num DF   Den DF   Pr > F

Wilks' Lambda       0.733   3.53        3        29     0.0271
Pillai's Trace      0.267   3.53        3        29     0.0271
Hotelling-Lawley    0.365   3.53        3        29     0.0271
Roy's Max Root      0.365   3.53        3        29     0.0271
```


Example: Rohwer data (R)

```
> Manova(mod1)

Type II MANOVA Tests: Pillai test statistic
   Df test stat approx F num Df den Df  Pr(>F)
n   1    0.0384   0.3856      3     29 0.76418
s   1    0.1118   1.2167      3     29 0.32130
ns  1    0.2252   2.8100      3     29 0.05696 .
na  1    0.2675   3.5294      3     29 0.02705 *
ss  1    0.1390   1.5601      3     29 0.22030
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note: `Manova()` and `Anova()` in the **car** package are identical

They give a compact summary for *all* predictors, for a *given* test statistic

Gory details are available from the **`summary()`** method

Canonical analysis: How many dimensions?

- Sequential tests for the latent roots λ_i of **HE** $^{-1}$ indicate the number of dimensions of the **y**s predicted by the **x**s.
- Canonical correlations: correlation of best linear combination of **y**s with best of **x**s

$$\lambda_i = \frac{\rho^2}{1-\rho^2} \qquad \rho^2 = \frac{\lambda}{1-\lambda}$$

```
proc reg data=lo_ses;
   model sat ppvt raven = n s ns na ss;
   M1: mtest / canprint;       /* all coeffs = 0 */
   run;
```


Canonical analysis: How many dimensions?

```
                         Adjusted     Approximate         Squared
          Canonical     Canonical        Standard       Canonical
        Correlation   Correlation           Error     Correlation

  1        0.716526      0.655198        0.081098        0.513409
  2        0.490621      0.414578        0.126549        0.240709
  3        0.266778      0.211906        0.154805        0.071170

              Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)

        Eigenvalue    Difference    Proportion    Cumulative

  1         1.0551        0.7381        0.7283        0.7283
  2         0.3170        0.2404        0.2188        0.9471
  3         0.0766                      0.0529        1.0000

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

        Likelihood   Approximate
             Ratio       F Value    Num DF    Den DF    Pr > F

  1     0.34316907          2.54        15    80.458    0.0039
  2     0.70525204          1.43         8        60    0.2025
  3     0.92882959          0.79         3        31    0.5078
```

Wilks' Lambda

only 1 signif. dim

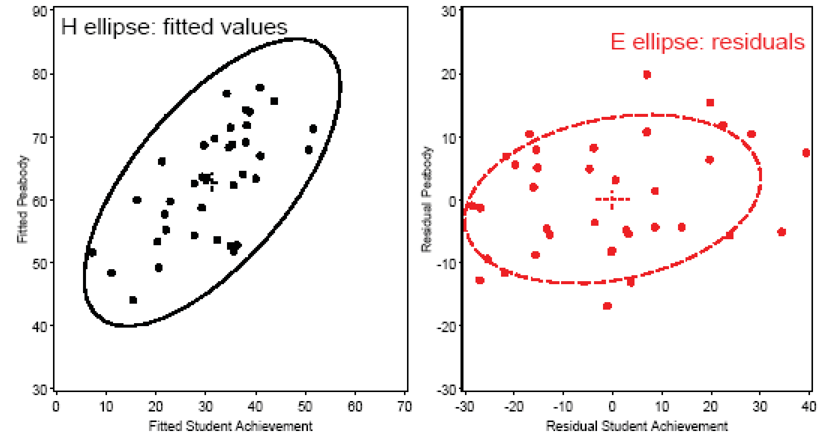
Visualizing multivariate tests: HE plots

- The **H** and **E** matrices in the GLH summarize the (co)variation of the fitted values and residuals for a given effect

$$\mathbf{H} = \hat{\mathbf{Y}}_{eff}^{T} \hat{\mathbf{Y}}_{eff} \qquad \mathbf{E} = \mathcal{E}^{T} \mathcal{E}$$

  - For two variables, we can visualize their size & shape with data ellipses
  - For *p=3* these display as ellipsoids
  - For *p*>2 can use an HE-plot matrix

Data ellipses for H & E



How big is H relative to E?

How to make them comparable?

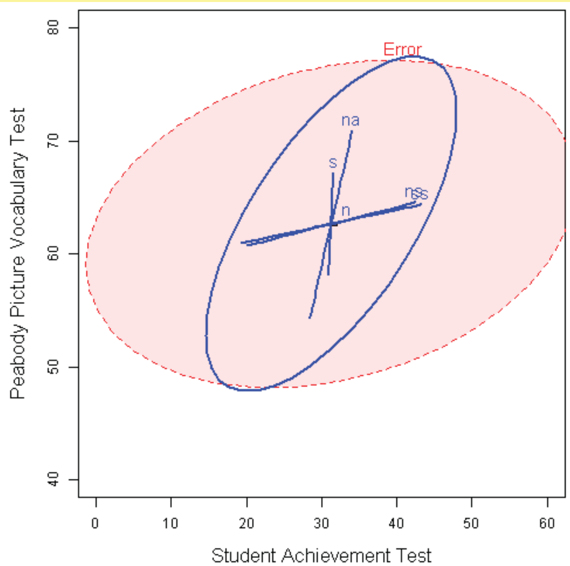
Animation:
http://www.datavis.ca/gallery/animation/manova/

HE plot: effect scaling

- Scale: **E**/df_e, **H**/df_e
- Center: shift to centroid
- Plot: 68% data ellipses

For each predictor, the data ellipse degenerates to a line (rank **H**: s=1)

- **Orientation**: how x_i contributes to prediction of y_1, y_2
- **Length**: relative strength of relation



HE plot: significance scaling

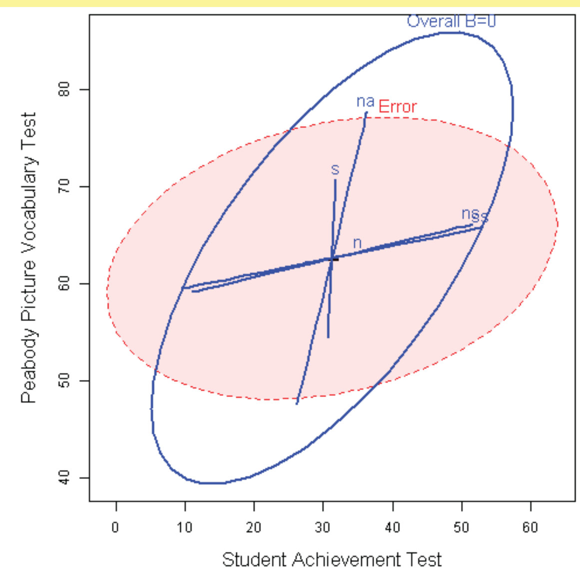
Scale:

- E: E/df_e
- H: H/df_e λ_α

λ_α = critical value of Roy statistic at level α

→ any **H** ellipse will protrude beyond **E** ellipse *iff* effect is significant at level α

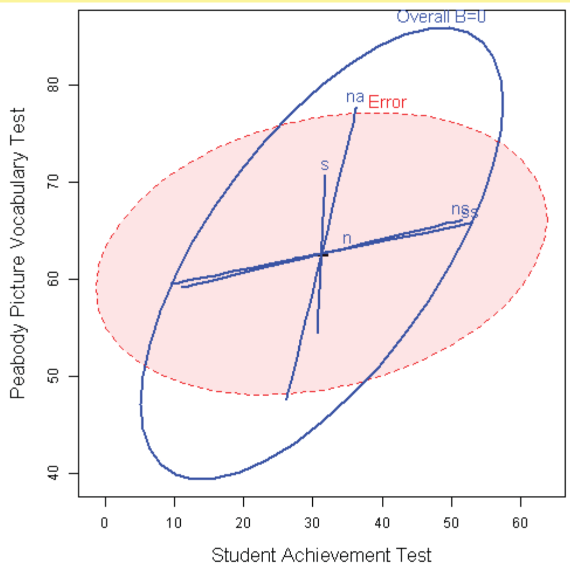
Directions of **H**s show *how* predictors contribute to responses



HE plot: significance scaling

Rohwer data, low SES gp

- Overall test highly significant
- Only NA individually significant (in this view)
- NA & S contribute to predicting PPVT
- NS & SS contribute to SAT
- N doesn't contribute at all

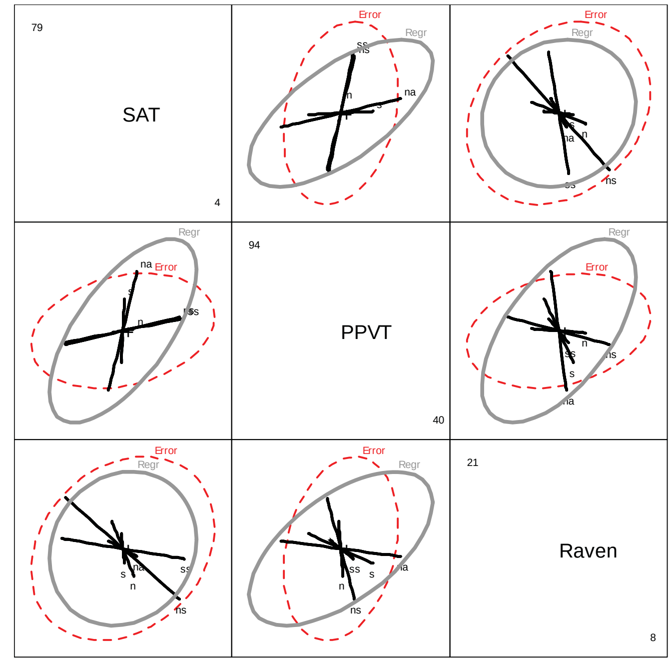


HE plot matrix

All pairwise views

An effect is significant if **H** projects outside **E** in *any* view

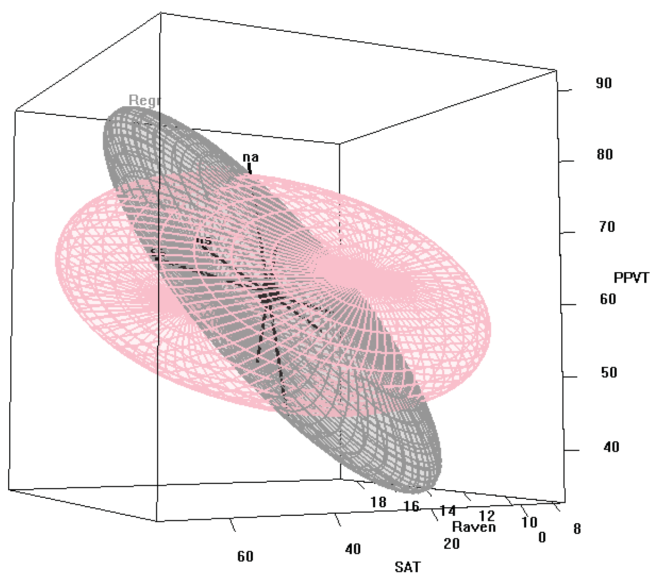
That applies to any rotation, not just the bivariate views shown here.



3D HE plot

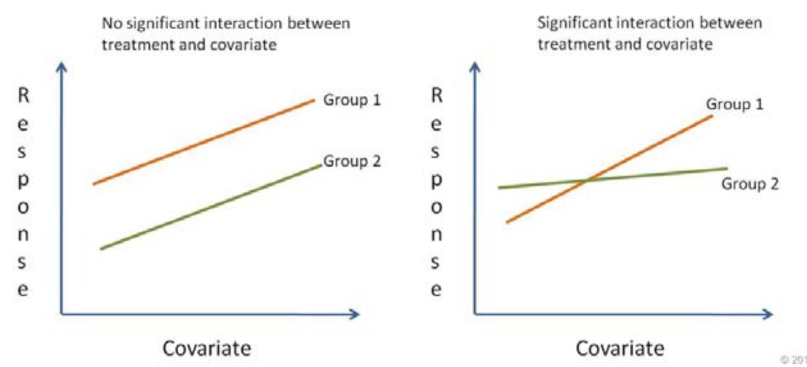
In the R version (heplots package), 3D plots can be rotated dynamically

In this view, we see NA poking out beyond the **E** ellipsoid



Homogeneity of regression: Univariate

- With 2+ groups there are several hypotheses of interest
  - equal slopes: no group * X interaction
  - equal means: no group "main effect"
  - equal slopes and means (same regression lines)
- ANCOVA: Test equal means, assuming equal slopes



Homogeneity of Regression: Multivariate

- When there are several groups, we often want to test hypotheses of “homogeneity”:
  - equal slopes for the predictors (interactions)?
  - equal intercepts for groups (same means)?
  - equal slopes & intercepts (coincident regressions)?

```
*-- test equal slopes, by allowing interactions (separate slopes for each group);
proc glm data=rohwer;
   class SES;
   model sat ppvt raven = SES|n SES|s SES|ns SES|na SES|ss /ss3 nouni;
   manova h=SES*n SES*s SES*ns SES*na SES*ss;       test all interactions
run;
```

```
*-- MANCOVA model: test intercepts (means), assuming equal slopes;
proc glm data=rohwer;
   class SES;
   model sat ppvt raven = SES n s ns na ss /ss3 nouni;
   manova h=_all_;
run;
```

NB: better than reporting separate results and making “eyeball” comparisons

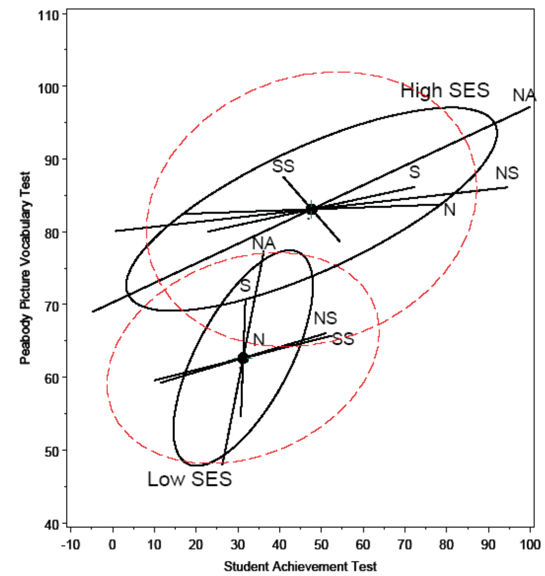
HE plots: Homogeneity of regression

Rohwer data: Lo (n=32) & Hi (n=37) SES groups:

- Fit separate regressions for each group
- Are slopes the same?
- Are intercepts the same?
- Are regressions coincident? (equal slopes and intercepts)

Here, slopes for NS are similar; most others seem to differ, but only NA is signif.

Intercepts (means) clearly differ.

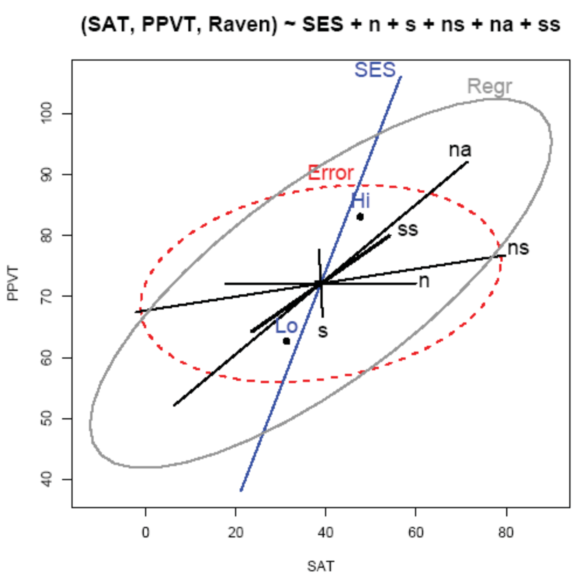


HE plots: MANCOVA model

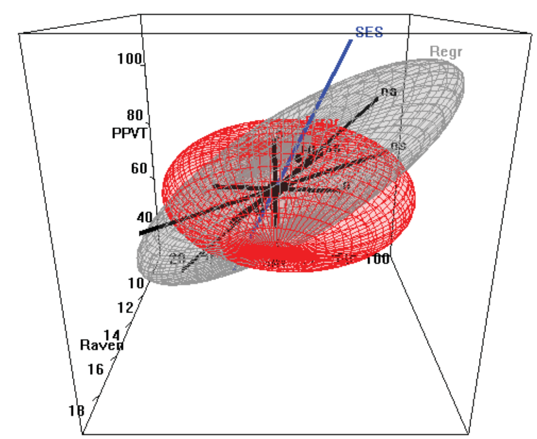
Alternatively, we can fit a model that assumes **equal** slopes for both SES groups, but allows unequal intercepts

From the ANOVA view, this is equivalent to an analysis of covariance model, with group effect and quantitative predictors

If the main interest is in the SES effect, the MANCOVA test relies on the assumption of equal slopes.



HE plots: MANCOVA model



Nature vs Nurture: IQ of adopted children MMReg + Repeated measures

- Data from an observational, longitudinal, study on adopted children (n=62).
- Is child's intelligence related to intelligence of the biological mother and the intelligence of the adoptive mother?
- The child's intelligence was measured at age 2, 4, 8, and 13
- How does intelligence change over time?
- How are these changes related to intelligence of the birth and adoptive mother?


```
> some(Adopted)
   AMED BMIQ Age2IQ Age4IQ Age8IQ Age13IQ
3    14   89    126    115    113      90
6     8   64    125    109     96      87
23    6   92    116    121    119     109
30   13   78    108     90     86      80
31   16   87    113     97    101     109
32   15   63    127    121    119     101
40    8   95    140    130    126     118
42   15   65    110    111    114      95
52   13   74    121    132    132     113
58   11   88    112    107    110     103
```

AMED: Adoptive mother educ. (proxy for IQ)

BMIQ: Birth mother IQ

Treat as multivariate regression problem:

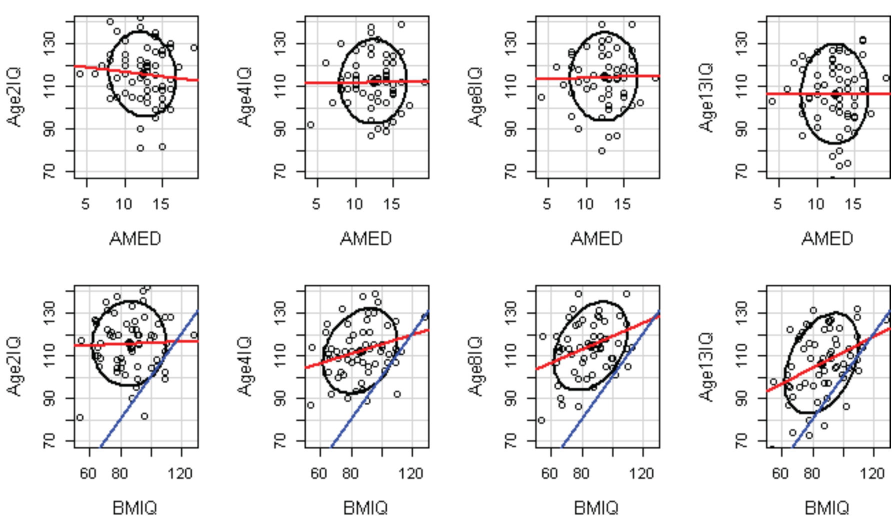
```
> Adopted.mod <- lm(cbind(Age2IQ, Age4IQ, Age8IQ, Age13IQ) ~ AMED + BMIQ, data=Adopted)
> Adopted.mod

Call:
lm(formula = cbind(Age2IQ, Age4IQ, Age8IQ, Age13IQ) ~ AMED +     BMIQ, data = Adopted)

Coefficients:
             Age2IQ     Age4IQ     Age8IQ     Age13IQ
(Intercept)  117.63046   93.33771   88.03739   76.84827
AMED          -0.44136   -0.02073   -0.01216   -0.16063
BMIQ           0.04001    0.22172    0.30961    0.36747
```

$= \mathbf{B}_{(3 \times 4)}$

What can we tell from this?

Scatterplots of child IQ vs. AMED and BMIQ

- Regression lines (red) show the fitted (univariate) relations
- Data ellipses: visualize strength of relations
- Blue lines: equality of child IQ and BMIQ



Multivariate tests of each predictor: $\boldsymbol{\beta}_{AMED} = \mathbf{0}$; $\boldsymbol{\beta}_{BMIQ} = \mathbf{0}$

```
> Manova(Adopted.mod)

Type II MANOVA Tests: Pillai test statistic
     Df test stat approx F num Df den Df  Pr(>F)
AMED  1   0.01722  0.24535      4     56 0.91129
BMIQ  1   0.17759  3.02320      4     56 0.02504 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Conclusions from this:

- Birth mother IQ significantly predicts child IQ at these ages: $\boldsymbol{\beta}_{BMIQ} \neq \mathbf{0}$
- Adoptive mother ED does not: $\boldsymbol{\beta}_{AMED} = \mathbf{0}$

How to understand the nature of these relations?

Where these tests come from:

```
> linearHypothesis(Adopted.mod, c("BMIQ"))
```

$$\mathbf{L} = \begin{array}{c} \qquad\qquad \text{BMIQ} \\ \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \end{array}$$

H

```
Sum of squares and products for the hypothesis:
         Age2IQ    Age4IQ    Age8IQ   Age13IQ
Age2IQ   24.78808  137.3590  191.8035  227.6471
Age4IQ  137.35902  761.1521 1062.8471 1261.4684
Age8IQ  191.80350 1062.8471 1484.1238 1761.4719
Age13IQ 227.64710 1261.4684 1761.4719 2090.6499
```

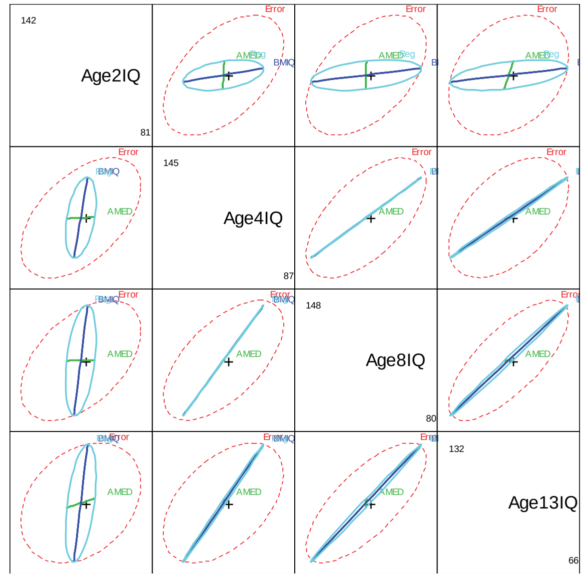
E

```
Sum of squares and products for error:
         Age2IQ   Age4IQ   Age8IQ   Age13IQ
Age2IQ  10242.157 5137.843 5000.888  3430.234
Age4IQ   5137.843 9561.649 5929.696  5316.677
Age8IQ   5000.888 5929.696 9875.424  8141.506
Age13IQ  3430.234 5316.677 8141.506 12312.409

Multivariate Tests:
                 Df test stat  approx F num Df den Df   Pr(>F)
Pillai            1 0.1775928 3.0231979      4     56 0.025038 *
Wilks             1 0.8224072 3.0231979      4     56 0.025038 *
Hotelling-Lawley  1 0.2159427 3.0231979      4     56 0.025038 *
Roy               1 0.2159427 3.0231979      4     56 0.025038 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$\mathbf{B} = \begin{pmatrix} \beta_{01} & \beta_{02} & \beta_{03} & \beta_{04} \\ \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14} \\ \beta_{21} & \beta_{22} & \beta_{23} & \beta_{24} \end{pmatrix}$$

Age: 2 4 8 13


```
> pairs(Adopted.mod, hypotheses=list("Reg"=c("AMED", "BMIQ")))
```



Pairwise HE plots showing tests of AMED & BMIQ + overall regression

- Signif. of BMIQ largely from older ages

Repeated measures analysis

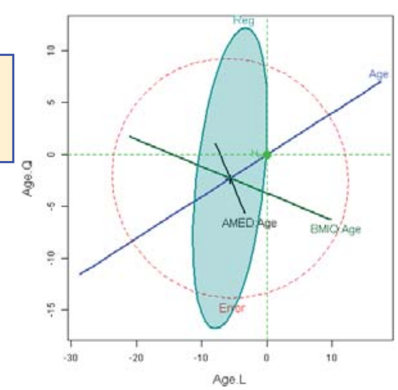
- Because Age is a quantitative factor, we can use it in a multivariate trend analysis.
- This amounts to analysis of **Y M**, where **M** comes from

$$\mathbf{M} = \begin{bmatrix} 2 & 2^2 & 2^3 \\ 4 & 4^2 & 4^3 \\ 8 & 8^2 & 8^3 \\ 13 & 13^2 & 13^3 \end{bmatrix}$$

- This gives tests of linear, quadratic & cubic trends of IQ in relation to AMED and BMIQ
- Interactions– AMED*Age & BMIQ*Age test for equal slopes over Age


```
> # Treat IQ at different ages as a repeated measure factor
> Age <- data.frame(Age=ordered(c(2,4,8,13)))
> Anova(Adopted.mod, idata=Age, idesign=~Age, test="Roy")

Type II Repeated Measures MANOVA Tests: Roy test statistic
         Df test stat approx F num Df den Df   Pr(>F)
AMED      1    0.0019   0.1131      1     59 0.737878
BMIQ      1    0.1265   7.4612      1     59 0.008302 **
Age       1    0.7120  13.5287      3     57 8.91e-07 ***
AMED:Age  1    0.0143   0.2718      3     57 0.845454
BMIQ:Age  1    0.1217   2.3114      3     57 0.085792 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> heplot(Adopted.mod, idata=Age,
idesign=~Age, iterm="Age",
+  hypoth=list("Reg"=c("AMED", "BMIQ")))
> mark.H0()
```



HE plots: software

- SAS macros
  - See: Friendly (2006): *Data ellipses, HE plots …*, http://www.jstatsoft.org/v17/i06/
  - `heplots, hemreg, hemat`: http://www.math.yorku.ca/SCS/sasmac/
- R packages
  - See: Fox et al (2007): *Visual hypothesis tests in MLMs…* http://www.math.yorku.ca/SCS/Papers/dsc-paper.pdf
  - `heplots` & `car` packages: http://www.r-project.org/

Summary

- MMRA → multivariate tests for a collection of p responses, each in up to s dimensions
  - Different test statistics combine these in different ways, to say how big is **H** vs **E**
  - Canonical analysis: How many dimensions of Ys are predicted by the Xs?
  - HE plots → visualize the relations of responses to the predictors
- These methods generalize to all linear models: MANOVA, MANCOVA, etc.