

Multivariate multiple regression

This example concerns predicting measures of aptitude and achievement for a sample of low SES kindergarten students from measures of performance on paired associate memory tests. The response variables are:

- PPVT – Peabody picture vocabulary test
- SAT – A scholastic aptitude test
- Raven – the Raven progressive matrices test

The predictors are 5 paired associate tasks called N, S, NS, NA, SS.

1. Run the following to create the LO_SES data set.

```
%include data(lo_ses);
```

2. The following proc reg step carries out a series of analyses. Run each analysis line **individually** and examine the output before going on to the next.

```
proc reg data=lo_ses;
  model sat ppvt raven = n s ns na ss; run; *-- step 1;
  M1: mtest / print canprint; run; *-- step 2;
  M2: mtest n,s,ns / print canprint; run; *-- step 3;
  M3: mtest sat-ppvt, na,ns; run; *-- step 4;
```

3. For each step, indicate what statistical hypothesis is/are being tested. What do you conclude from each step? In particular,
 - Which PA tests are significant predictors for **each** response by univariate multiple regression tests?
 - Which PA tests are significant predictors for **all** responses by multivariate tests?
4. Now open the file `mreg3.sas` from `N:\psyc6140\examples\regress`. Run the same multivariate tests with PROC GLM:

```
title 'Using PROC GLM';
proc glm data=lo_ses outstat=stats;
  model sat ppvt raven = n s ns na ss / ss3 nouni;
  manova h=_all_/ printh printe short;
  *-- obtain predicted values and residuals;
  output out=results p=p_sat p_ppvt p_raven
  r=r_sat r_ppvt r_raven;
run;
```

5. Examine the **E** matrix. What are the diagonal entries? What do the “Partial correlations from the Error SSCP Matrix” mean? Examine the **H** matrix for each predictor. What are the diagonal entries? [Hint: find the (1,1) diagonal entries among the regression output for SAT.]
6. Run the remaining steps plotting the predicted values and residuals.

```
*-- Visualizing the H and E variation for SAT and PPVT;
%gdispla(OFF);
title 'Predicted values';
%ellipses(data=results, x=p_sat, y=p_ppvt, colors=black);
```

MMRA

```
title 'Residuals';
%ellipses(data=results, x=r_sat, y=r_ppvt);
%gdispla(ON);
%panels(rows=1, cols=2);

*--Visualize H & E matrices;
%scatmat(data=results, var=p_sat p_ppvt p_raven, anno=ellipse, interp=rl);
%scatmat(data=results, var=r_sat r_ppvt r_raven, anno=ellipse, interp=rl);
```

7. Finally, create an HE plot that shows how all predictors contribute to the multivariate test (for two predictors). An H effect that projects outside the E ellipse is significant by one (Roy's) multivariate test.

```
*-- HE plot for individual predictors;
title 'HE plot for SAT and PPVT';
%heplots(data=lo_ses, stat=stats, var=sat ppvt, class=,
          effects=n s ns na ss);
```

Multivariate tests in R

The data used in this exercise is a subset of the data frame **Rohwer** in the **heplots** package.

```
library(heplots)
data(Rohwer)
head(Rohwer)
```

Fit the multivariate regression model using **lm()**, test with Anova & linearHypothesis:

```
# just use Lo SES group here (via subset=)
rohwer.ses1 <- lm(cbind(SAT, PPVT, Raven) ~ n + s + ns + na + ss,
data=Rohwer, subset=SES=="Lo")
```

```
# MANOVA results for each predictor
Anova(rohwer.ses1)
Anova(rohwer.ses1, test="Roy")
# more detailed output:
summary(Anova(rohwer.ses1))
# overall MANOVA tests for all predictors jointly (B=0)
predictors <- c("n", "s", "ns", "na", "ss")
linearHypothesis(rohwer.ses1, hypothesis=predictors)
```

HE plots:

```
heplot(rohwer.ses1)
# Add ellipse to test all 5 regressors
heplot(rohwer.ses1, hypotheses=list("Regr" = predictors))
# View all pairs
pairs(rohwer.ses1, hypotheses=list("Regr" = predictors))
# or 3D plot
col <- c("pink", "black", "blue", "cyan", "magenta", "brown", "gray")
heplot3d(rohwer.ses1, hypotheses=list("Regr" = predictors), col=col)
```