

Testing mean differences: Hotelling's T^2 and MANOVA

This example deals with the simplest multivariate situation for testing mean differences – a single-sample design, which could be treated either as a univariate ANOVA, a repeated measures ANOVA, or a MANOVA. It also introduces you to the ideas of testing contrasts, and the necessity to reshape data from “wide” to “long” format for different analyses.

The data concerns a ‘probe’ experiment, where subjects listened to tape-recorded sentences, then a probe word, and had to respond with the next word in the sentence. The response is a measure of speed, and the goal is to see if speed varies with grammatical function in the sentence. More specifically, we want to see **how** speed varies with grammatical function.

1. Open the file `probe0.sas` from `N:\psyc6140\examples\glm`, and submit the data step to create the PROBE1 data set.
2. The data has been read in “wide” format, with 5 variables, P1-P5, corresponding to the positions in a sentence at which a probe word has been tested. This is the data format used for both repeated measures and multivariate analysis.

```

P1  P2  P3  P4  P5
datalines;
1  51  36  50  35  42
2  27  20  26  17  27
....

```

3. PROC TRANSPOSE is used to convert from wide format to “long” format, with all the scores in a single variable, SPEED, and the positions in a second variable, POSITION. This is the format required for plotting the data and also for univariate ANOVA.

```

proc transpose data=probel out=long prefix=speed;
  by subjno;
  var p1-p5;
data long;
  set long;
  rename speed1 = Speed
         _name_ = Position;

```

4. As always, a first step is to plot the data. Here, a boxplot of speed by position is useful. This uses the long format.

```
%boxplot(data=long, var=Speed, class=Position, connect=3);
```

5. A standard univariate, repeated measures analysis is easily carried out here with the data in the long format. But this relies on the fact that for a

MeanDiff

1-factor, repeated measures design, the correct error term for testing the POSITION effect is essentially the POSITION * SUBJNO interaction. In more complex designs, you would have to specify different error terms for each effect. This analysis also requires additional assumption (compound symmetry) that is not tested in this analysis.

```
proc glm data=long;
  class Position subjno;
  model Speed = Position subjno;
  contrast 'SubjPred' Position 1 1 -1 -1 0;
  contrast 'AdjNoun' Position 1 -1 1 -1 0;
  contrast 'SPxAN' Position 1 -1 -1 1 0;
  contrast 'RelPn' Position 1 1 1 1 -4;
run;
```

In the output, pay attention to the tests of the contrasts. Try to interpret these in relation to what you saw in the boxplot.

6. An alternative is a MANOVA analysis, which requires no additional assumptions, and uses the data in the wide format. The same contrasts can be tested here, but the syntax is different, and the output format is more complex.

```
proc glm data=probel;
  model p1-p5 = / nouni;
  manova H= intercept /* No between subject effect*/
         M= P1 + P2 - P3 - P4, /* Subject vs. Predicate */
         P1 - P2 + P3 - P4, /* Adjs vs Nouns */
         P1 - P2 - P3 + P4, /* SubPred x AdjNoun */
         P1 + P2 + P3 + P4 - 4*P5 /* Relative clause */
         MNames = SubPred AdjNoun SPxAN RelPn
         / summary short ;
run;
```

In the output, find the overall test of the position effect, and the tests of the contrasts in position. Are the conclusions the same or different as from the univariate tests?