STAT 802 Applied Multivariate Analysis Nov 14, 2005

Student Presentations Today:

Linnea Duke	12.6	Correspondence Analysis	14-Nov
Elizabeth Juarez	10.4	Interpreting Canonical Correlation	14-Nov

If time permits:

Intro to Ch 7: Multivariate Linear Regression
Familiar Multiple Regression model – box p 356 and Result 7.1 p 358
Note assumed independence of rows of the n x p data matrix. Also Z fixed.
$Y = Z\beta + \varepsilon$ and estimate β with . $\hat{\beta} = (Z'Z)^{-1}Z'y$ Can compute SSR, SSE.

Connection with Anova – p 357 – use of dummy variables.
Note: as example p 357, Z is not rank 4 and Z'Z is singular. In this case replace $(Z'Z)^{-1}$ by the generalized inverse $(Z'Z)^{-}$ (B is gen inv of A if BAB=B) p 418 in J-W
More detail at
Eric W. Weisstein. "Moore-Penrose Matrix Inverse." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/Moore-PenroseMatrixInverse.html

```
y=c(9,6,9,0,2,3,1,2)
a=c(rep(1,11),rep(0,8),1,1,rep(0,8),1,1,1)
z=matrix(a,ncol=4,nrow=8)
> z
     [,1] [,2] [,3] [,4]
[1,]    1    1    0    0
[2,]    1    1    0    0
[3,]    1    1    0    0
[4,]    1    0    1    0
[5,]    1    0    1    0
[6,]    1    0    0    1
[7,]    1    0    0    1
[8,]    1    0    0    1
> tzz=t(z)%*%z
> bhat=ginv(tzz)%*%t(z)%*%y
> bhat
      [,1]
[1,]  2.75
[2,]  5.25
[3,] -1.75
[4,] -0.75
```

Note 2.75+5.25 = 8 = mean in pop 1
2.75-1.75 = 1 = mean in pop 2
2.75-0.75 = 2 = mean in pop 3

Connection between regression (lm) and analysis of variance (aov)
g=cbind(d1,d2,d3)

```
> a=lm(y~g)
> summary(a)

Call:
lm(formula = y ~ g)

Residuals:
   Min     1Q Median     3Q    Max
 -4.00  -1.25   0.50   2.00   2.00

Coefficients:
            Estimate Std. Error t value
(Intercept)  10.0000     2.1180   4.721
g            -3.0000     0.9718  -3.087
            Pr(>|t|)
(Intercept)  0.00325 **
g            0.02147 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.38 on 6 degrees of freedom
Multiple R-Squared: 0.6136,  Adjusted R-squared: 0.5492
F-statistic: 9.529 on 1 and 6 DF,  p-value: 0.02147

> b=aov(y~g)
> summary(b)
            Df Sum Sq Mean Sq F value
g            1 54.000  54.000  9.5294
Residuals    6 34.000   5.667
             Pr(>F)
g           0.02147 *
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note 9.529 = (-3.087)^2

See last line of p 357: regression includes aov.

Sampling properties of $\hat{\beta}$ (pp 363-365)

Result 7.2 (p 363) mean and cov of $\hat{\beta}$.

Result 7.3 (Gauss Theorem) about minimum variance linear unbiased estimators.

(Do we need unbiasedness?) "Justifies" plugging in $\hat{\beta}$ for β in linear predictor.

More next time ....

.