	8.3,	Principal	16-
Saman	8.4	Components	Nov
Simon		Principal Componentss – Large	16-
Goring	8.5	Sample	Nov

Ch 7 : Multivariate Regression (Continued)

Skip to Multivariate Regression (section 7.7) p 383ff

Suppose # response variables = m, vector **Y**
        # predictors = r
Design matrix **Z** includes a column of 1's so is m x (r+1)
        # number of rows in data matrix is n.
Data matrix is n x (m+r) but we don't operate on whole matrix so unlabelled.

Model is

$$\mathbf{Y} \quad = \quad \mathbf{Z} \qquad \qquad \beta \qquad \qquad + \varepsilon$$

$$\mathbf{n \ x \ m} \qquad \mathbf{n \ x \ (r+1)} \quad \mathbf{(r+1) \ x \ m} \qquad \mathbf{n \ x \ m}$$

with usual covariance assumptions.   (p 384)

With this notation, the equations of multiple regression are extended directly to
multivariate regression.

Normal Theory: Least squares -> MLE ests (Result 7.10 p 390)

$\hat{\beta}$ is indept of $\hat{\Sigma}$

Inference:

Likelihood ratio test to see if addition of (r-q) new predictors adds predictive information
to the model using only q predictors. Result 7.11 (p 393) says can use chi-square
distribution to assess. Normality assumption is important.
Note Likelihood Ratio test is same as test using Wilks Lambda.

Can get confidence ellipsoids, and simultaneous CIs, for the mean multivariate response
for a given selection of z-values (a row of the design matrix or else an externally defined
row of this type). There are also formulas for the prediction ellipsoids, and simultaneous
prediction intervals, for the actual multivariate response for a given selection of z-values
(a row of the design matrix or else an externally defined row of this type). p 396.

Section 7.8 Z fixed versus Z multivariate normal. Bottom line is no diff for linear preds.
See the two sentences following (7-70) on p 409.

Example based on Ex 7.21 (p 421)

```
> T1.5.df
   Wind SolRad CO NO NO2 O3 HC                  21   10   62  5  3  14   4  4
1     8     98  7  2  12  8  2                  22    9   88  4  2   7   6  3
2     7    107  4  3   9  5  3                  23    8   80  4  2  13  11  4
3     7    103  4  3   5  6  3                  24    5   30  3  3   5   2  3
4    10     88  5  2   8 15  4                  25    6   83  5  1  10  23  4
5     6     91  4  2   8 10  3                  26    8   84  3  2   7   6  3
6     8     90  5  2  12 12  4                  27    6   78  4  2  11  11  3
7     9     84  7  4  12 15  5                  28    8   79  2  1   7  10  3
8     5     72  6  4  21 14  4                  29    6   62  4  3   9   8  3
9     7     82  5  1  11 11  3                  30   10   37  3  1   7   2  3
10    8     64  5  2  13  9  4                  31    8   71  4  1  10   7  3
11    6     71  5  4  10  3  3                  32    7   52  4  1  12   8  4
12    6     91  4  2  12  7  3                  33    5   48  6  5   8   4  3
13    7     72  7  4  18 10  3                  34    6   75  4  1  10  24  3
14   10     70  4  2  11  7  3                  35   10   35  4  1   6   9  2
15   10     72  4  1   8 10  3                  36    8   85  4  1   9  10  2
16    9     77  4  1   9 10  3                  37    5   86  3  1   6  12  2
17    8     76  4  1   7  7  3                  38    5   86  7  2  13  18  2
18    8     71  5  3  16  4  4                  39    7   79  7  4   9  25  3
19    9     67  4  2  13  2  3                  40    7   79  5  2   8   6  2
20    9     69  3  3   9  5  3                  41    6   68  6  2  11  14  3
                                               42    8   40  4  3   6   5  2
> a=lm(NO2~Wind*SolRad)

> summary(a)

Call:
lm(formula = NO2 ~ Wind * SolRad)

Residuals:
    Min      1Q  Median      3Q
-5.8059 -2.2182 -0.6222  1.7497
    Max
10.4253

Coefficients:
            Estimate Std. Error
(Intercept)  8.107481   9.694219
Wind         0.057170   1.248271
SolRad       0.050258   0.136447
Wind:SolRad -0.003992   0.017846
            t value Pr(>|t|)
(Intercept)   0.836    0.408
Wind          0.046    0.964
SolRad        0.368    0.715
Wind:SolRad  -0.224    0.824


Residual standard error: 3.459 on 38 degrees of freedom
Multiple R-Squared: 0.0244,      Adjusted R-squared: -0.05262
F-statistic: 0.3168 on 3 and 38 DF,  p-value: 0.8131
```
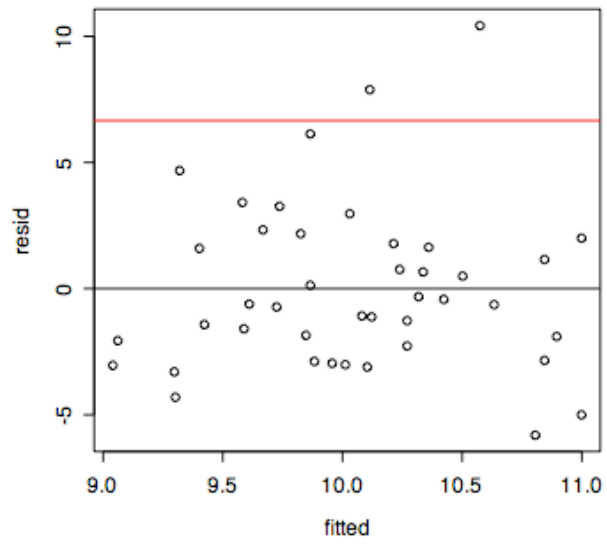
```
>resid=a[[2]]
>fitted=a[[5]]
> plot(fitted,resid)
> abline(h=0)
> abline(h=2*sd(resid),col="red")
```


prediction interval?