Handout 10.1. Analysis of variance

Two-way ANOVA

Consider the model which decomposes observations into a general level, a row effect, a column effect and a noise term.

$$x_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}, \epsilon_{ij} \sim N(0,\ \sigma^2)$$

in which $\sum \alpha_i = 0$, and $\sum \beta_j = 0$. Let x_{ij} denote the *j*th observation in the *i*th row, $\bar{x}_{i.}$ the mean of the ith row, $\bar{x}_{.j}$ the mean of the jth column, and $\bar{x}_{..}$ the overall mean.

```
data(heart.rate)
attach(heart.rate)
heart.rate
```

```
> heart.rate <- data.frame(hr = c(96,110,89,95,128,100,72,79,100,
     92,106,86,78,124,98,68,75,106,86,108,85,78,118,100,67,74,
     104, 92,114,83,83,118,94,71,74,102),
     subj=gl(9,1,36),   time=gl(4,9,36,labels=c(0,30,60,120)))
```

The gl (generate levels) function is specially designed for generating patterned factors for balanced experimental designs. It has three arguments: the number of levels, the block length (how many times each level should repeat), and the total length of the result. The two patterns in the data frame are thus

```
> gl(9,1,36)
[1] 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3
[31] 4 5 6 7 8 9
Levels: 1 2 3 4 5 6 7 8 9

> gl(4,9,36,labels=c(0,30,60,120))
[1] 0 0 0 0 0 0 0 0 0 30 30 30 30 30 30
[16] 30 30 30 60 60 60 60 60 60 60 60 60 120 120 120
[31] 120 120 120 120 120 120
Levels: 0 30 60 120
> attach(heart.rate)

> anova(lm(hr~subj+time))
Analysis of Variance Table
Response: hr
          Df Sum Sq Mean Sq F value     Pr(>F)
subj       8 8966.6 1120.82 90.6391 4.863e-16
***
time       3  151.0   50.32  4.0696   0.01802 *
Residuals 24  296.8   12.37
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.1 graphics for repeated measurements

interaction.plot(time,subj,hr)

1. ANOVA table in the regression analysis

The variation between and within groups for a one-way analysis of variance generalizes to model variation

$$\text{SSmodel}= \sum_i (\hat{y}_i - \hat{y}_.)^2$$

and residual variation

$$\text{SSres}= \sum_i (y_i - \hat{y}_i)^2$$

which partition the total variation

$$\text{SStotal}= \sum_i (y_i - \hat{y}_.)^2$$

This applies only when the model contains an intercept. The role of the group means in the one-way classification is taken over by the fitted values in the more general linear model.

The analysis of variance table corresponding to a regression analysis can be extracted with the function anova, just as for one- and two- way analyses of variance.

```
data(thuesen)
attach(thuesen)
lm.velo<-lm(short.velocity~blood.glucose)
summary(lm.velo)

Call:
lm(formula = short.velocity ~ blood.glucose)

Residuals:
     Min       1Q   Median       3Q      Max
-0.40141 -0.14760 -0.02202  0.03001  0.43490

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.09781    0.11748   9.345 6.26e-09 ***
blood.glucose  0.02196    0.01045   2.101   0.0479 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2167 on 21 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.1737, Adjusted R-squared:  0.1343
F-statistic: 4.414 on 1 and 21 DF,  p-value: 0.0479
```

```
anova(lm.velo)

Analysis of Variance Table

Response: short.velocity
               Df  Sum Sq  Mean Sq F value Pr(>F)
blood.glucose   1 0.20727 0.207269   4.414 0.0479 *
Residuals      21 0.98610 0.046957
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Notice that the F test gives the same p-value as the t test for a zero slope. It is the same F test that gets printed at the end of the summary output Residual standard error is the square root of residual mean squares, namely 0.2167=sqrt(0.04696). R^2 is the proportion of the total sum of squares explained by the regression line, that is 0.1737=0.2073/(0.2073+0.9861)

```
logret<- read.table("~/Desktop/d_logret_6stocks.txt",  header=T)
attach(logret)
fit1<-lm(Pfizer~Intel)
anova(fit1)
Analysis of Variance Table

Response: Pfizer
          Df   Sum Sq    Mean Sq F value Pr(>F)
Intel      1 0.000154 0.00015441  0.2865 0.5944
Residuals 62 0.033409 0.00053885

> fit2<-lm(Pfizer~Intel+AmerExp)
> anova(fit2)
```