Qualitative Predictors

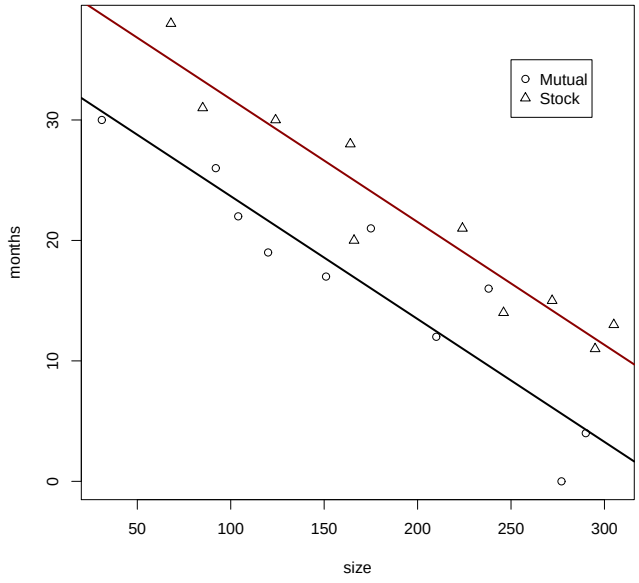
- **Insurance Example.** In a study of innovation in the insurance industry, an economist wished to relate the speed with which a particular insurance innovation is adopted (Y) to the size of the insurance firm (X_1) and the type of the firm (stock company or mutual fund). The response variable is measured by the number of months elapsed between the time the first firm adopted innovation and the time the given firm adopted the innovation. The data are stored in the file "Insurance.csv".

```
data=read.csv("Insurance.csv",header=T)   # Example on page 316
attach(data)
type=factor(type)   # 0=Mutual; 1=Stock     # Declaring that type is a qualitative variable
results=lm(months~size+type)
summary(results)$coef
              Estimate  Std. Error    t value     Pr(>|t|)
(Intercept) 33.8740690 1.813858297  18.675146 9.145269e-13
size        -0.1017421 0.008891218 -11.442990 2.074687e-09
type1        8.0554692 1.459105700   5.520826 3.741874e-05

tapply(months,type,mean) # Computes the mean of 'months' for each firm type
   0    1
16.7 22.1
tapply(size,type,mean) # Computes the mean of 'size' for each firm type
    0     1
168.8 194.9

# Note:
# Mutual: 16.7=33.8741-0.10174*(168.8)
# Stock: 22.1=33.8741+8.0555-0.10174*(194.9)

plot(size,months,pch=as.numeric(type))      # Will create scatterplots for the 2 type of firms
abline(33.874,-0.102,lwd=2) # For mutual firms
abline(33.874+8.055,-0.102,lwd=2,col="darkred") # For stock firms
legend(250,35,c("Mutual","Stock"),pch=1:2)
```

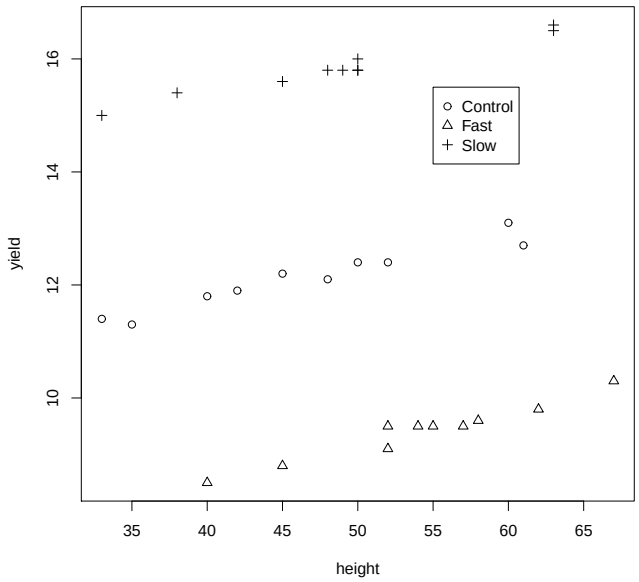



```
# Interactions between quantitative and qualitative predictors
results2=lm(months~size+type+size*type)
summary(results2)$coef
                Estimate Std. Error     t value     Pr(>|t|)
(Intercept) 33.8383694765 2.44064985 13.86449166 2.472768e-10
size        -0.1015306249 0.01305254 -7.77861250 7.965637e-07
type1        8.1312501223 3.65405169  2.22526959 4.079375e-02
size:type1  -0.0004171412 0.01833121 -0.02275578 9.821265e-01
```

- **Fertilizer Example.** In this study, the effects of two treatments, a slow-release fertilizer (S) and a fast-releasing fertilizer (F), on seed yield (grams) of peanut plants were compared with a control (C), a standard fertilizer. Ten replications of each treatment were to be grown in a greenhouse study. When setting up the experiment, the researcher recognized that the 30 peanut plants were not exactly at the same level of development or health. Consequently, the researcher recorded the height (cm) of the plant, a measure of plant development and health, at the start of the experiment. The data are stored in the file "

```
data.fertilizer=read.csv("Fertilizer.csv",header=T)
attach(data.fertilizer)

plot(height,yield,pch=as.numeric(treatment))
legend(55, 15.5, c("Control","Fast","Slow"), pch=1:3)
```




```
summary(lm(yield~treatment+height+treatment:height))$coef
summary(lm(yield~treatment*height))$coef
                     Estimate  Std. Error    t value     Pr(>|t|)
(Intercept)        9.491768741 0.202459041  46.882415 4.084943e-25
treatmentf        -3.519620102 0.350334683 -10.046451 4.489809e-10
treatments         3.906558043 0.295789635  13.207217 1.675966e-12
height             0.056614405 0.004265179  13.273628 1.506750e-12
treatmentf:height  0.006814587 0.006746322   1.010119 3.225144e-01
treatments:height -0.006886936 0.006084213  -1.131935 2.688450e-01

summary(lm(yield~treatment+height))$coef
               Estimate  Std. Error   t value     Pr(>|t|)
(Intercept)  9.52925636 0.133573491  71.34092 2.346301e-31
treatmentf  -3.14415561 0.060373900 -52.07806 7.942625e-28
treatments   3.57163712 0.057032666  62.62441 6.817469e-30
height       0.05580995 0.002734287  20.41115 1.580422e-17

tapply(yield,treatment,mean) # Computes the mean of yield for each treatment type
    c     f     s
12.13  9.41 15.83
tapply(height,treatment,mean) # Computes the mean of height for each treatment type
   c    f    s
46.6 54.2 48.9
# Note:
# Control: 12.13=9.529+0.0558*(46.6)
# Fast: 9.41=9.529-3.144+0.0558*(54.2)
# Slow: 15.83=9.529+3.572+0.0558*(48.9)
```