Notes: MS 204 Chapter 6

Overview

- Multiple linear regression

Multiple linear regression: book volume, weight, and cover type

Ex: $X_1 =$ volume, $X_2 = cover$, $Y =$ SAT

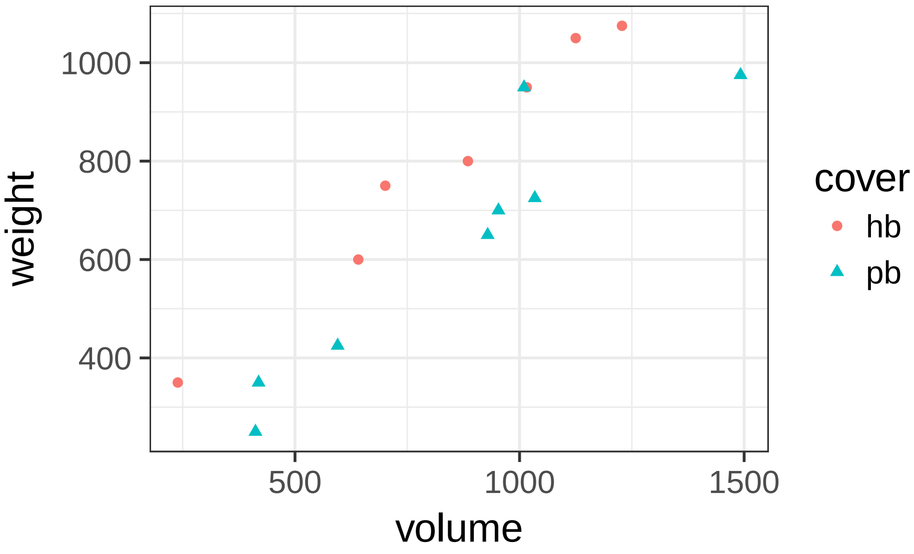
```
#install.packages("DAAG")
library(DAAG); library(mosaic); library(tidyverse)
set.seed(0)
allbacks %>% sample_n(3)
```

```
##    volume area weight cover
## 14    595    0    425    pb
## 4     239  371    350    hb
## 5     701  371    750    hb
```

```
dim(allbacks)
```

```
## [1] 15  4
```

```
qplot(x = volume, y = weight, data = allbacks, pch = cover, color = cover) +
  theme_bw(15)
```



1. Describe the overall association between volume and weight.

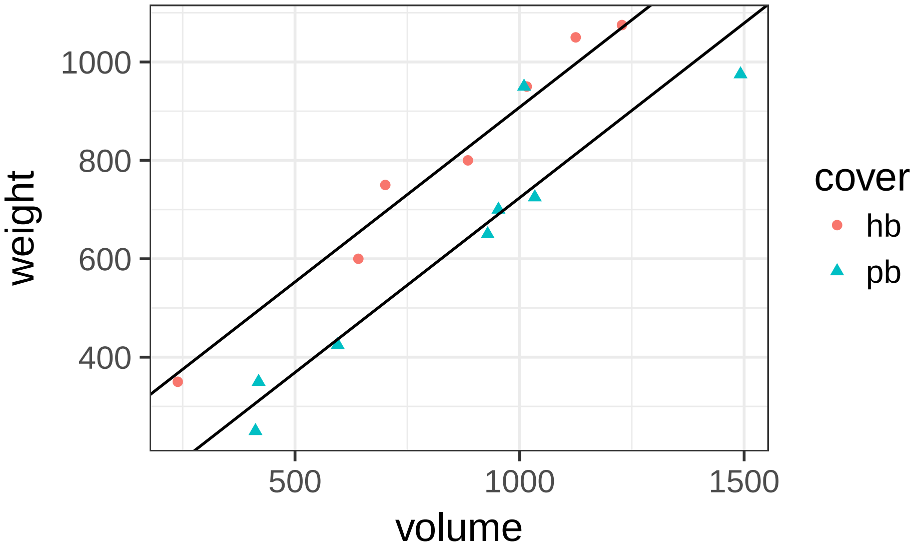
2. Describe the association between book cover type and weight

```r
fit <- lm(weight ~ volume + cover, data = allbacks)
msummary(fit)
```

```
##               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  197.96284   59.19274   3.344 0.005841 ** 
## volume         0.71795    0.06153  11.669  6.6e-08 ***
## coverpb     -184.04727   40.49420  -4.545 0.000672 ***
## 
## Residual standard error: 78.2 on 12 degrees of freedom
## Multiple R-squared:  0.9275, Adjusted R-squared:  0.9154 
## F-statistic: 76.73 on 2 and 12 DF,  p-value: 1.455e-07
```

Visualizing the linear model

```r
qplot(x = volume, y = weight, data = allbacks, pch = cover,
      color = cover, group = cover) +
  geom_abline(aes(intercept = 197.9, slope = 0.71)) +
  geom_abline(aes(intercept = 197.9 - 184, slope = 0.71)) +
  theme_bw(15)
```



Interpreting coefficients in a multiple regression model

Predictions and residuals

Predict the weight of the book in first row of the data, and use that prediction to find that book's residual.

```
allbacks %>% head(1)
```

```
##   volume area weight cover
## 1    885  382    800    hb
```

R-squared

```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: weight
##           Df Sum Sq Mean Sq F value    Pr(>F)
## volume     1 812132  812132 132.809  7.58e-08 ***
## cover      1 126320  126320  20.657 0.0006719 ***
## Residuals 12  73381    6115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Adjusted R-squared

```
set.seed(0)
fit0 <- lm(weight ~ volume, data = allbacks)
fit1 <- lm(weight ~ volume + cover, data = allbacks)
allbacks <- allbacks %>% mutate(noise = rnorm(15))
fit2 <- lm(weight ~ volume + cover + noise, data = allbacks)
c(msummary(fit0)$r.squared, msummary(fit1)$r.squared, msummary(fit2)$r.squared)
```

```
## [1] 0.8026346 0.9274776 0.9327799
```

```
c(msummary(fit0)$adj.r.squared, msummary(fit1)$adj.r.squared, msummary(fit2)$adj.r.squared)
```

```
## [1] 0.7874526 0.9153905 0.9144471
```

Collinearity

```r
fit.sat0 <- lm(salary ~ verbal + ratio, data = SAT)
msummary(fit.sat0)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 70.50197   11.12959   6.335 8.37e-08 ***
## verbal      -0.08088    0.02169  -3.729 0.000516 ***
## ratio        0.07704    0.33659   0.229 0.819949
##
## Residual standard error: 5.329 on 47 degrees of freedom
## Multiple R-squared:  0.2284, Adjusted R-squared:  0.1955
## F-statistic: 6.954 on 2 and 47 DF,  p-value: 0.002261
```

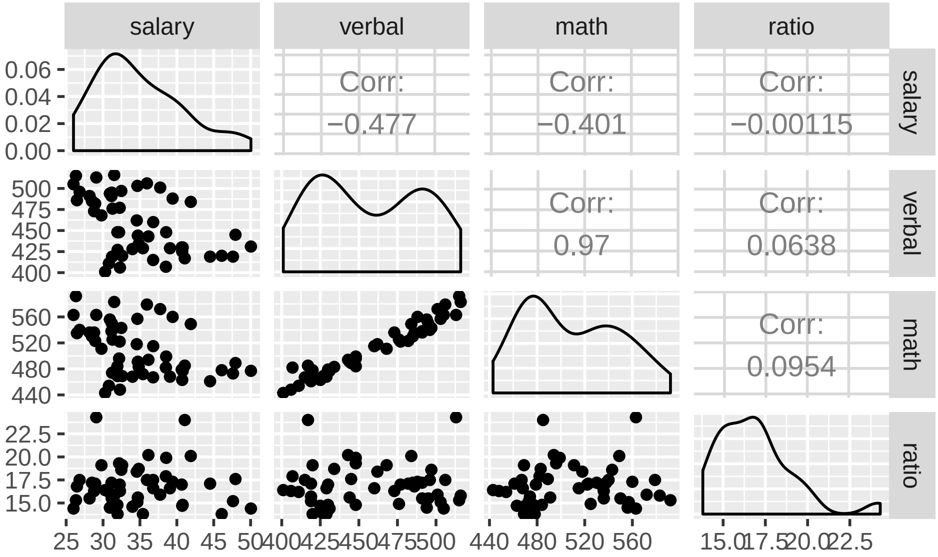
Interpretations:

```r
fit.sat1 <- lm(salary ~ verbal + math + ratio, data = SAT)
msummary(fit.sat1)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 71.60985   10.78979   6.637  3.2e-08 ***
## verbal      -0.25294    0.08718  -2.901  0.00569 **
## math         0.15550    0.07647   2.033  0.04780 *
## ratio       -0.01589    0.32908  -0.048  0.96169
##
## Residual standard error: 5.16 on 46 degrees of freedom
## Multiple R-squared:  0.292,  Adjusted R-squared:  0.2458
## F-statistic: 6.324 on 3 and 46 DF,  p-value: 0.001107
```

Interpretations:

```r
library(GGally)
ggpairs(select(SAT, salary, verbal, math, ratio))
```


salary
verbal
math
ratio
Corr:
Corr:
Corr:
Corr:
Corr:
Corr: