Chapter 6: Multiple Regression

- Multiple Regression Model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \qquad (i = 1, 2, \ldots, n; n > k)$$

where,

- ε_i*'s are uncorrelated with a mean of 0 and constant variance* σ^2_ε.
- ε_i*'s are normally distributed.* (*This is needed to do inferences about the coefficients.*)

- Least-Square Regression Equation:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

by choosing $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ to minimize $\text{SS(Residual)} = \Sigma_i (y_i - \hat{y}_i)^2$

Obtaining Model Parameter Estimates

```
data.health=read.csv("HealthExam.csv",header=T)
head(data.health)
Gender Age Height Weight Waist Pulse SysBP DiasBP Cholesterol BodyMass  Leg Elbow Wrist  Arm
1      F  12   63.3  156.3  81.4    64   104     41          89     27.5 41.0   6.8   5.5 33.0
2      F  16   57.0  100.7  68.7    64   106     64           2     21.9 33.8   5.6   4.6 26.4
3      M  17   63.0  156.3  86.7    96   109     65          78     27.8 44.2   7.1   5.3 31.7

attach(data.health)
result2=lm(Weight~Waist+Height)
summary(result2)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -201.5717    21.5686  -9.346 2.59e-14 ***
Waist          2.1565     0.1003  21.500  < 2e-16 ***
Height         2.5978     0.3438   7.557 7.22e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.2 on 77 degrees of freedom
Multiple R-squared:  0.8996,    Adjusted R-squared:  0.8969
F-statistic: 344.8 on 2 and 77 DF,  p-value: < 2.2e-16
```

*Weight = -201.5717 + 2.1565*Waist + 2.5978*Height.*

Both p-values are <0.05, therefore, both variables have significant effects on Weight.

Since R^2 will keep on increasing as we put more explanatory variables in the model, the adjusted R^2 is an alternative measure that can be used to help choose between two competing models. This measure gets a penalty for the number of explanatory variables in the model.

By including Height in the model we increased R^2 by 0.0745. Hence, our new model is able to explain 7.45% more of the variability in Weight.

2

Prediction and Model Assessments

Predict the weight of 2 individuals – one with waist size of 80 cm and height of 60 in and another with waist size of 90 cm and height of 70 inches.

*Since, Weight = -201.5717 + 2.1565*Waist + 2.5978*Height*

```
Weight1=-201.5717+2.1565*80+2.5978*60              # 126.8163
Weight2=-201.5717+2.1565*90+2.5978*70              # 174.3593

predict(result2,newdata=data.frame(Waist=c(80,90),Height=c(60,70)))
 1        2
126.8145 174.3573

confint(result2,level=.99)
                  0.5 %      99.5 %
(Intercept) -258.538718 -144.604756
Waist          1.891574    2.421405
Height         1.689847    3.505723

# Model Assessments
qqnorm(result2$res);qqline(result2$res)
shapiro.test(result2$res)
# W = 0.9884, p-value = 0.6898

plot(result2$fit,result2$res)
```