Section 3.2 Assessing a Multiple Regression Model

Load needed packages.

```
library(Stat2Data)
```

Create a dataframe for **NFLStandings2016** and look at the structure of the data.

```
data("NFLStandings2016")
str(NFLStandings2016)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ Team         : Factor w/ 32 levels "Arizona Cardinals",..: 20 9 16 24 2 22 26 29 12 18 ...
##  $ Wins         : int  14 13 12 12 11 11 11 10 10 10 ...
##  $ Losses       : int  2 3 4 4 5 5 5 5 6 6 ...
##  $ Ties         : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ WinPct       : num  0.875 0.813 0.75 0.75 0.688 0.688 0.688 0.656 0.625 0.625 ...
##  $ PointsFor    : int  441 421 389 416 540 310 399 354 432 363 ...
##  $ PointsAgainst: int  250 306 311 385 406 284 327 292 388 380 ...
##  $ NetPts       : int  191 115 78 31 134 26 72 62 44 -17 ...
##  $ YardsFor     : int  6179 6027 5488 5973 6653 5291 5962 5715 5901 5325 ...
##  $ YardsAgainst : int  5222 5502 5896 6002 5939 5435 5482 5099 5822 6122 ...
##  $ TDs          : int  51 49 42 47 63 36 47 37 51 45 ...
```

EXAMPLE 3.3 NFL winning percentage: checking conditions

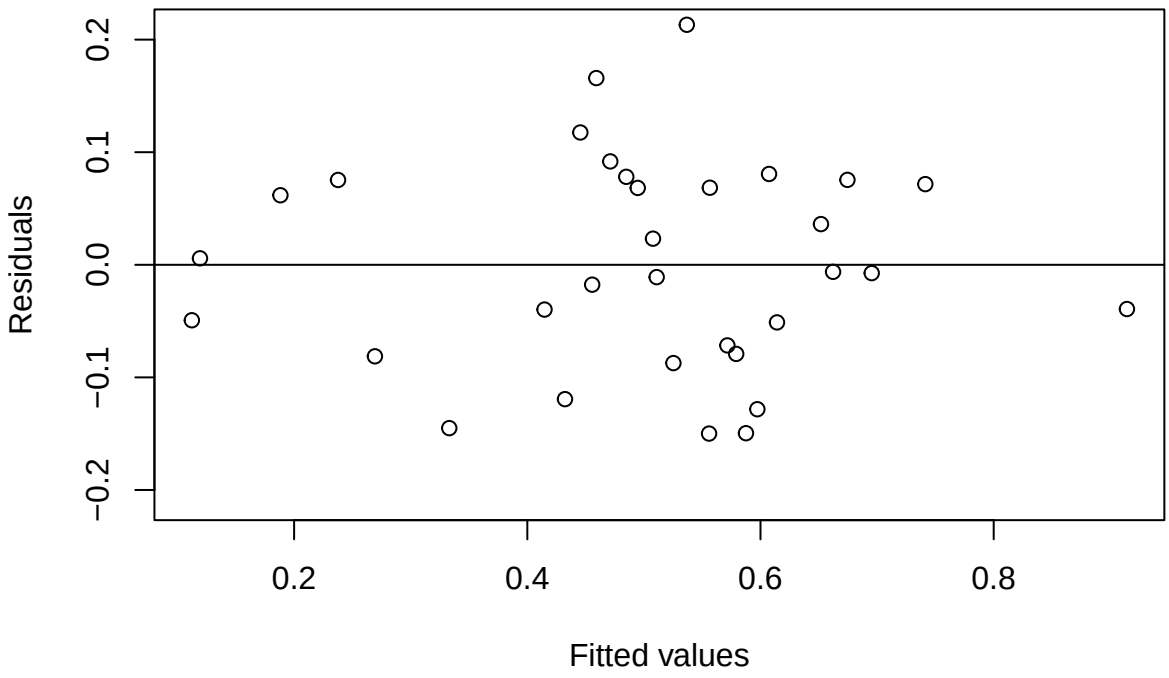
Fit the multiple regression model.

```
NFLmodel2016mr <- lm(WinPct ~ PointsFor + PointsAgainst , data=NFLStandings2016)
```

FIGURE 3.4 Residual plots for NFL winning percentage model

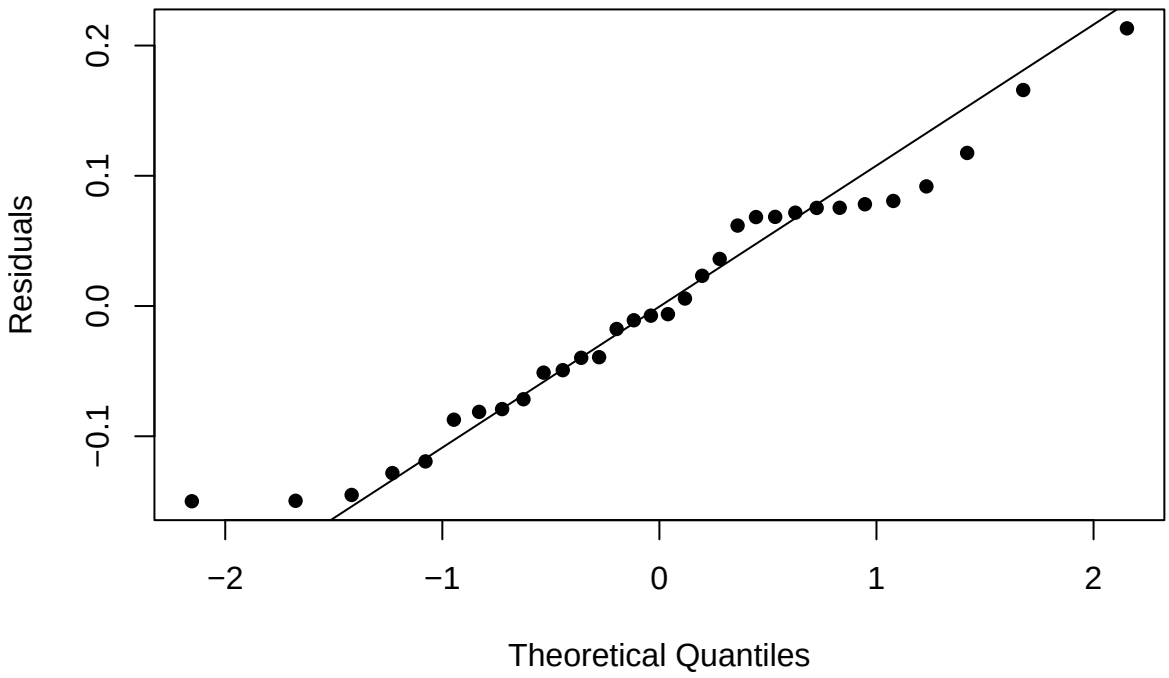
(a) Residuals versus fits

```
plot(NFLmodel2016mr$residuals~NFLmodel2016mr$fitted,ylab="Residuals",xlab="Fitted values",ylim=c(-0.21,
abline(h=0)
```



(b) Normal quantile plot

```
qqnorm(NFLmodel2016mr$residuals, xlab="Theoretical Quantiles", ylab="Residuals",main="", pch=16)
qqline(NFLmodel2016mr$residuals)
```



EXAMPLE 3.4 NFL winning percentage: testing individual terms

```
summary(NFLmodel2016mr)
```

```
##
## Call:
## lm(formula = WinPct ~ PointsFor + PointsAgainst, data = NFLStandings2016)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.149898 -0.073482 -0.006821  0.072569  0.213189
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7853698  0.1537422   5.108 1.88e-05 ***
## PointsFor      0.0016992  0.0002628   6.466 4.48e-07 ***
## PointsAgainst -0.0024816  0.0003204  -7.744 1.54e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09653 on 29 degrees of freedom
## Multiple R-squared:  0.7824, Adjusted R-squared:  0.7674
## F-statistic: 52.13 on 2 and 29 DF,  p-value: 2.495e-10
```

ExAMPLE 3.5 NFL winning percentage: CI for coefficients

```
confint(NFLmodel2016mr)
```

```
##                      2.5 %       97.5 %
## (Intercept)    0.470931804  1.099807882
## PointsFor      0.001161680  0.002236626
## PointsAgainst -0.003136962 -0.001826191
```

EXAMPLE 3.6 NFL winning percentage: ANOVA for overall fit

Note: R does not provide the same ANOVA table as in the example, but the F statistic and P-value are provided in the last line of this output.

```
summary(NFLmodel2016mr)
```

```
##
## Call:
## lm(formula = WinPct ~ PointsFor + PointsAgainst, data = NFLStandings2016)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.149898 -0.073482 -0.006821  0.072569  0.213189
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.7853698  0.1537422   5.108 1.88e-05 ***
## PointsFor      0.0016992  0.0002628   6.466 4.48e-07 ***
## PointsAgainst -0.0024816  0.0003204  -7.744 1.54e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09653 on 29 degrees of freedom
## Multiple R-squared:  0.7824, Adjusted R-squared:  0.7674
## F-statistic: 52.13 on 2 and 29 DF,  p-value: 2.495e-10
```

The default anova( ) command for multiple regression model shows the amount of new variability explained as each predictor is added to the model.

```
anova(NFLmodel2016mr)
```

```
## Analysis of Variance Table
##
## Response: WinPct
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## PointsFor      1 0.41262 0.41262  44.280 2.692e-07 ***
## PointsAgainst  1 0.55884 0.55884  59.972 1.537e-08 ***
## Residuals     29 0.27023 0.00932
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If you add the df for the indiviudal predictors (1+1), you get the df for the model (2).

If you add the SS for the individual predictors (0.41262+0.55884), you get the SS for the model (0.97146).

The information for the residuals (Error df, SSE, and MSE) matches the ANOVA table in the text, but R does not provide the information for the Total.

The coefficeint of multiple determination R^2 is shown in the regression summary( ). Multiple R-squared: 0.7824.

But we can also find it from the correlation between actual and predicted response values.

Find the Pearson correlation between winning percentage and the fitted values, then square it to get R^2.

```
cor(NFLStandings2016$WinPct, NFLmodel2016mr$fitted.values)
```

```
## [1] 0.8845147
```

```
cor(NFLStandings2016$WinPct, NFLmodel2016mr$fitted.values)^2
```

```
## [1] 0.7823662
```

EXAMPLE 3.7 NFL winning percentage: adjusted R^2

Even though adjusted R^2 is always part of the summary(NFLmodel2016mr) output, we illustrate how to compute it here.

```
anovaNFL2016=anova(NFLmodel2016mr)
SSE=anovaNFL2016$'Sum Sq'[3]
SSE
```

```
## [1] 0.2702337
```

```
SSTotal=anovaNFL2016$'Sum Sq'[1]+anovaNFL2016$'Sum Sq'[2]+anovaNFL2016$'Sum Sq'[3]
SSTotal
```

```
## [1] 1.24169
```

```
dfe=anovaNFL2016$'Df'[3]
dfe
```

```
## [1] 29
```

```
dfTotal=length(NFLStandings2016$WinPct)-1
dfTotal
```

```
## [1] 31
```

```
Rsq_adj=1-(SSE/dfe)/(SSTotal/dfTotal)
Rsq_adj
```

```
## [1] 0.767357
```

EXAMPLE 3.8 NFL winning percentage: prediction interval

```
newx=data.frame(PointsFor=400, PointsAgainst=350)
predict(NFLmodel2016mr,newdata=newx,interval="prediction")
```

```
##         fit       lwr       upr
## 1 0.5964794 0.3948782 0.7980807
```

Alternative Solution

You can use the nested F test, which we formally introduce in Section 3.6, by fitting a model with only an intercept parameter. The anova command is then used to compare the residual sum of squares for the two models. Notice that the F statistic matches the F statistic provided in the last line of the summary(NFLmodel2016mr) command.

```
reduced = lm(WinPct ~ 1 , data=NFLStandings2016)
anova(reduced, NFLmodel2016mr)
```

```
## Analysis of Variance Table
##
## Model 1: WinPct ~ 1
## Model 2: WinPct ~ PointsFor + PointsAgainst
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     31 1.24169
## 2     29 0.27023  2   0.97146 52.126 2.495e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```