NFL

Chris Parrish

January 8, 2016

NFL

reference:
- Cannon, et al., Stat2, chapter 03, examples 3.1-3.7

Import the data.

```
data <- read.csv("NFL2007Standings.csv", header=TRUE)
head(data)
```

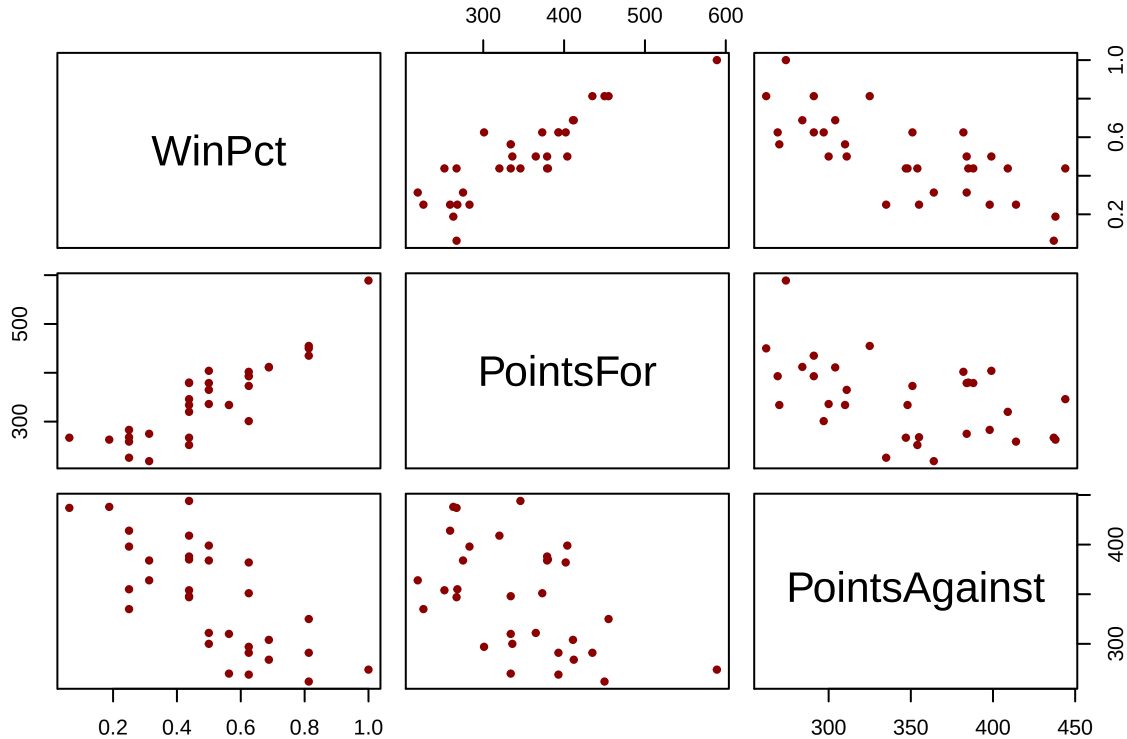
```
##                  Team Conference Division Wins Losses WinPct PointsFor
## 1 New England Patriots        AFC      ACE   16      0  1.000       589
## 2       Dallas Cowboys        NFC      NCE   13      3  0.813       455
## 3     Green Bay Packers       NFC      NCN   13      3  0.813       435
## 4   Indianapolis Colts        AFC      ACS   13      3  0.813       450
## 5 Jacksonville Jaguars        AFC      ACS   11      5  0.688       411
## 6   San Diego Chargers        AFC      ACW   11      5  0.688       412
##   PointsAgainst NetPts TDs
## 1           274    315  75
## 2           325    130  54
## 3           291    144  49
## 4           262    188  54
## 5           304    107  50
## 6           284    128  49
```

```
dim(data)
```

```
## [1] 32 10
```

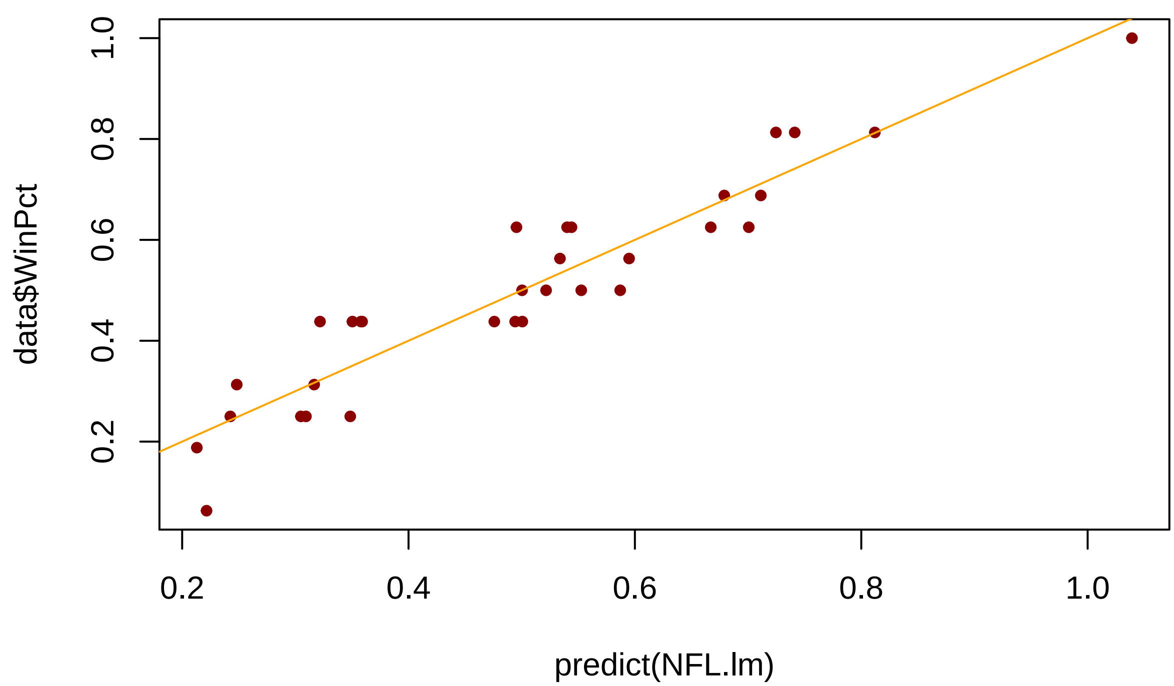
Scatterplot matrix.

```
pairs(~ WinPct + PointsFor + PointsAgainst, data=data,
      pch=20, col="darkred")
```



Multiple regression.

```r
NFL.lm <- lm(WinPct ~ PointsFor + PointsAgainst, data=data)
plot(predict(NFL.lm), data$WinPct,
     pch=20, col="darkred")
abline(a=0, b=1, col="orange")
```



Linear model.

$\widehat{WinPct} = 0.417 + 0.002\ PointsFor + \text{-}0.002\ PointsAgainst$

```r
options(show.signif.stars=FALSE)
summary(NFL.lm)
```

```
##
## Call:
## lm(formula = WinPct ~ PointsFor + PointsAgainst, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.15857 -0.05318 -0.01259  0.07360  0.12962
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4172230  0.1394480   2.992  0.00561
## PointsFor      0.0017662  0.0001870   9.445 2.37e-10
## PointsAgainst -0.0015268  0.0002751  -5.551 5.50e-06
##
## Residual standard error: 0.07298 on 29 degrees of freedom
## Multiple R-squared:  0.8844, Adjusted R-squared:  0.8764
## F-statistic: 110.9 on 2 and 29 DF,  p-value: 2.598e-14
```

```r
confint(NFL.lm)
```

```
##                      2.5 %        97.5 %
## (Intercept)    0.132019770  0.7024263043
## PointsFor      0.001383757  0.0021486491
## PointsAgainst -0.002089413 -0.0009642705
```

```r
anova(NFL.lm)
```

```
## Analysis of Variance Table
##
## Response: WinPct
##               Df  Sum Sq Mean Sq F value    Pr(>F)
## PointsFor      1 1.01724 1.01724 190.983 2.727e-14
## PointsAgainst  1 0.16411 0.16411  30.812 5.496e-06
## Residuals     29 0.15446 0.00533
```

Standard error of the multiple regression model.

$\hat{\sigma}_\epsilon = \sqrt{MSE}$

```r
sqrt(0.00533)
```

```
## [1] 0.07300685
```

Coefficient of multiple determination.

$$R^2 = \frac{SSModel}{SSTotal}$$

```r
SSModel <- 1.01724 + 0.16411
SSE <- 0.15446
SSTotal <- SSModel + SSE
R.sq <- SSModel / SSTotal
R.sq
```

```
## [1] 0.8843698
```

Adjusted coefficient of multiple determination.

$$R^2_{Adj} = 1 - \frac{SSE/(n-k-1)}{SSTotal/(n-1)}$$

```r
n <- 32
k <- 2
Adj.R.sq <- 1 - (SSE / (n - k - 1)) / (SSTotal / (n - 1))
Adj.R.sq
```

```
## [1] 0.8763953
```

Correlation of y and $\hat{y}$

```r
y <- data$WinPct
y.hat <- predict(NFL.lm)
r <- cor(y, y.hat)
r
```

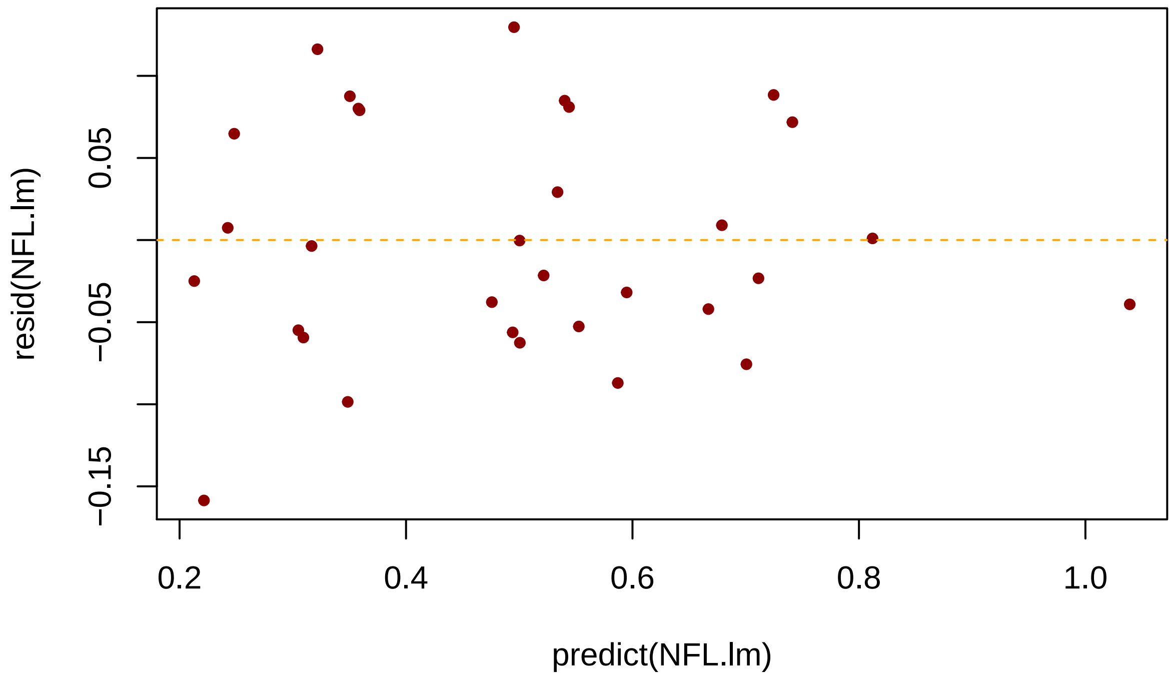
```
## [1] 0.9404083
```

```r
r^2
```

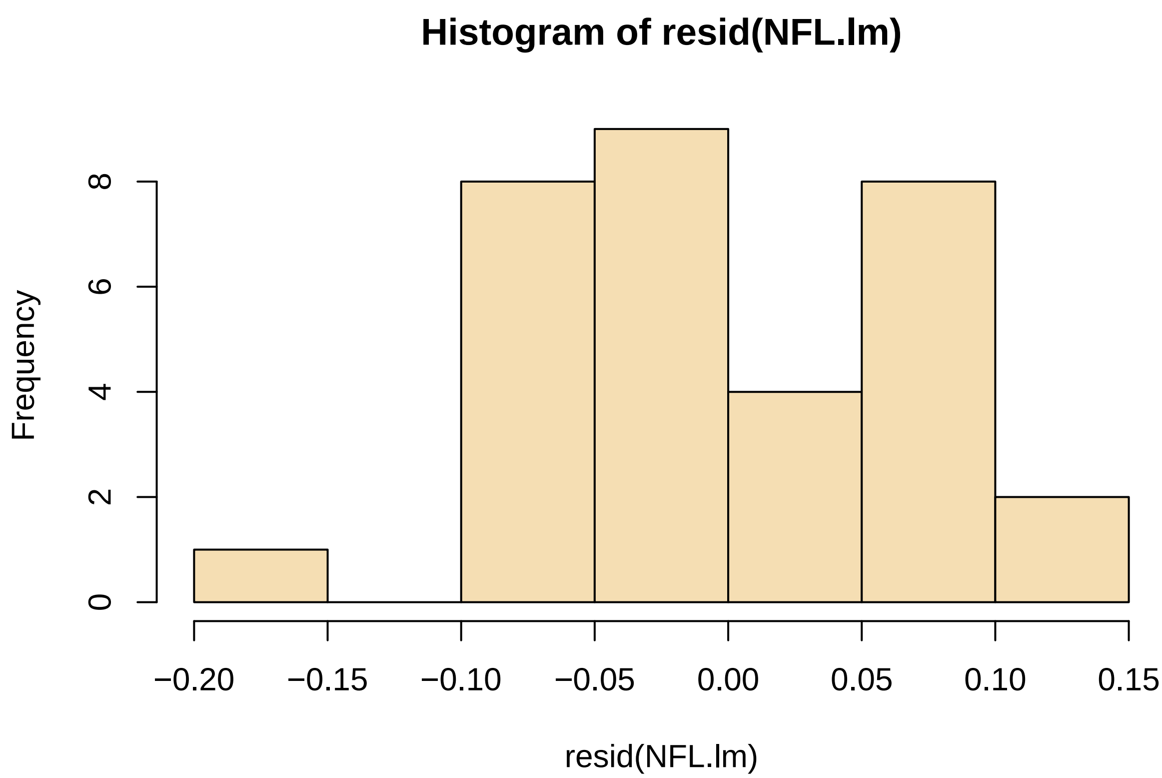
```
## [1] 0.8843678
```

Residuals.

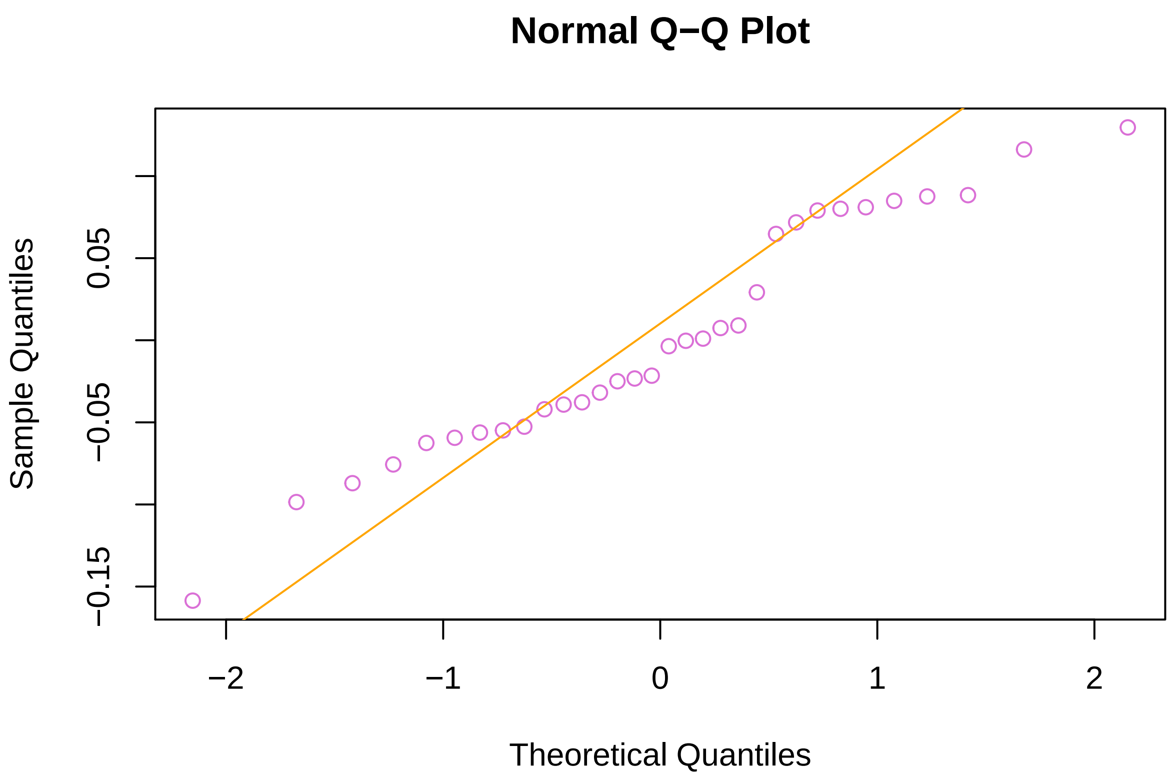
```r
plot(predict(NFL.lm), resid(NFL.lm),
     pch=20, col="darkred")
abline(h=0, col="orange", lty="dashed")
```




```
hist(resid(NFL.lm),
     col="wheat")
```

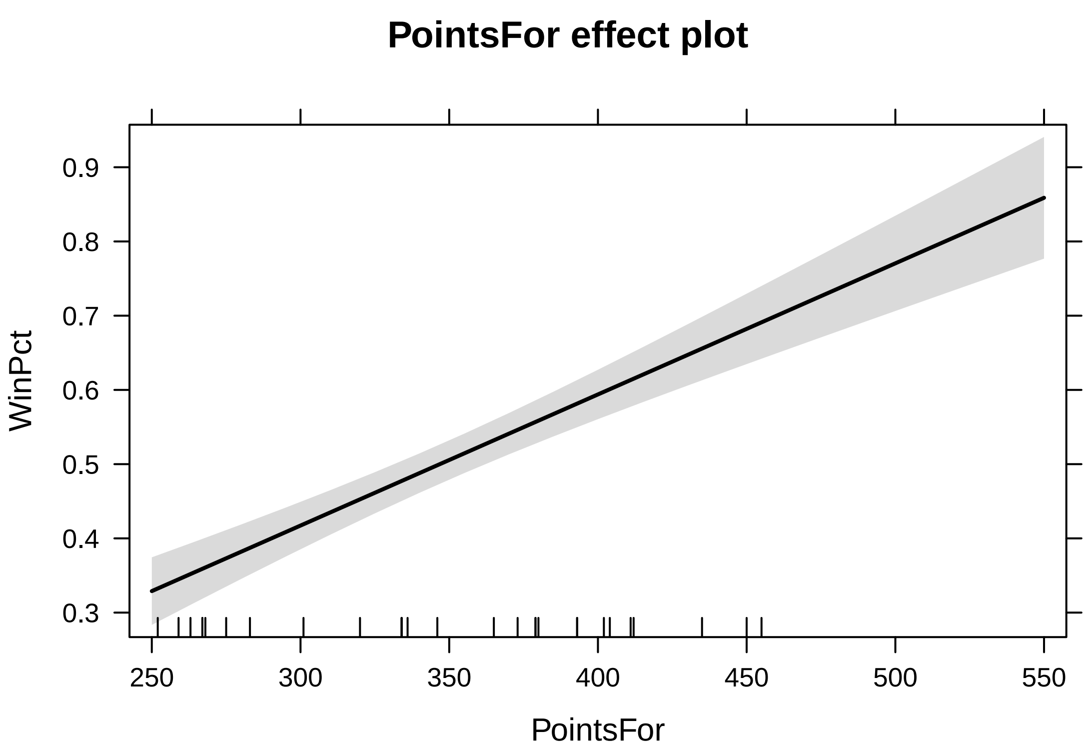



```
qqnorm(resid(NFL.lm),
       col="orchid")
qqline(resid(NFL.lm),
       col="orange")
```

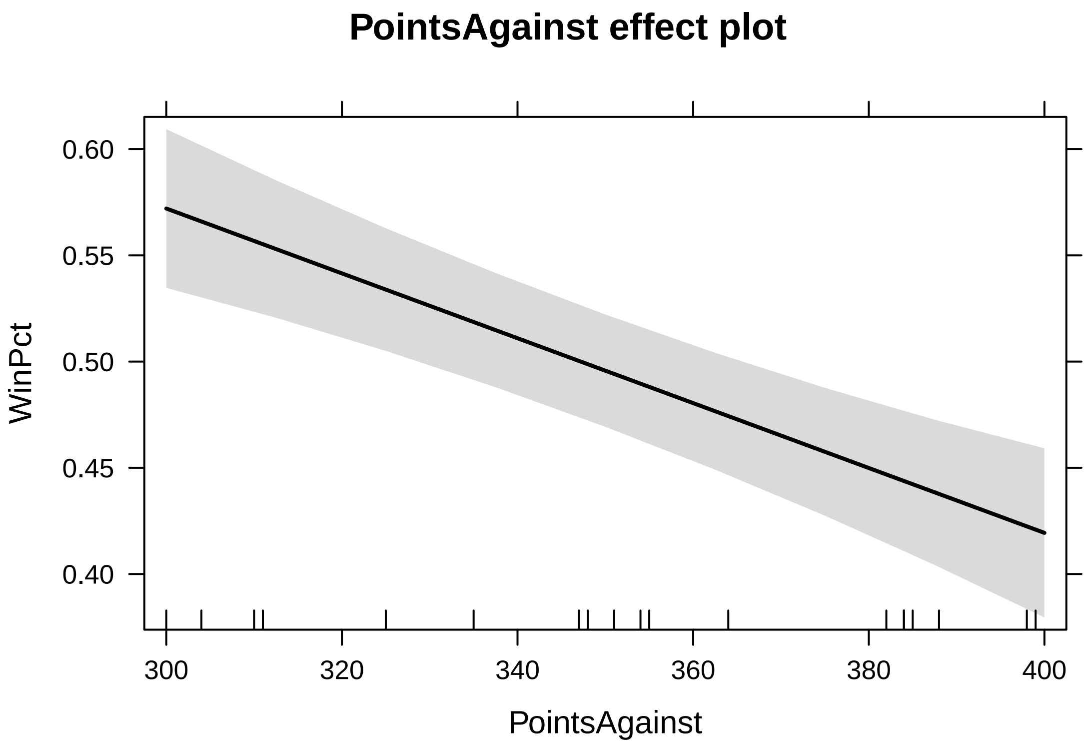


Effect plots.

```r
library(alr4)
plot(effect("PointsFor", NFL.lm))
```



```r
plot(effect("PointsAgainst", NFL.lm))
```



Prediction.

```r
new.data <- data.frame(PointsFor=393, PointsAgainst=260)  # Pittsburgh Steelers, 2007
(y.hat <- predict(NFL.lm, new.data))
```

```
##         1
## 0.7143618
```

```r
y <- 0.625
(residual <- y - y.hat)
```

```
##           1
## -0.08936183
```

CI and PI.

```r
new.data <- data.frame(PointsFor=400, PointsAgainst=350)  # hypothetical
conf <- predict(NFL.lm, new.data, interval="confidence")
pred <- predict(NFL.lm, new.data, interval="prediction")
intervals <- rbind(conf, pred)
row.names(intervals) <- c("CI", "PI")
intervals
```

```
##          fit       lwr       upr
## CI 0.5893095 0.5555032 0.6231158
## PI 0.5893095 0.4362648 0.7423541
```