# Week 12 Multiple Linear Regression: Extra Sum of Squares and Lack of Fit Tests

STAT 111

Zhengfei Li (Alex) Instructed by Prof. Phil Everson

*I would like to acknowledge that this handout is created with reference to Mathematical Statistics and Data Analysis by J. A. Rice (2007). I would also like to acknowledge the instructions from Prof. Everson.*

## 1 Preliminaries

**Properties 1** *The projection matrix $H = X(X^TX)^{-1}X^T$ has a few helpful properties for our topic today:*

*a. $H$ is symmetric: $H = H^T$*

*b. $H$ is idempotent: $H^2 = H$*

*c. $X$ is invariant under $H$: $HX = X$*

*d. Let there be a projection matrix of a full model $H_f$ and that of a reduced model $H_r$.*

$$H_fH_r = H_rH_f = H_r$$

## 2 Introduction to Extra Sum of Square with Soccer Player Data (5b)

We are given the soccer player data, and we are interested in predicting the weights.

```
  Division Pos GK Weight Height
1        1   F  N    158     71
2        1   M  N    145     71
3        1   M  N    150     67
4        1   D  N    147     68
5        1   F  N    160     68
```

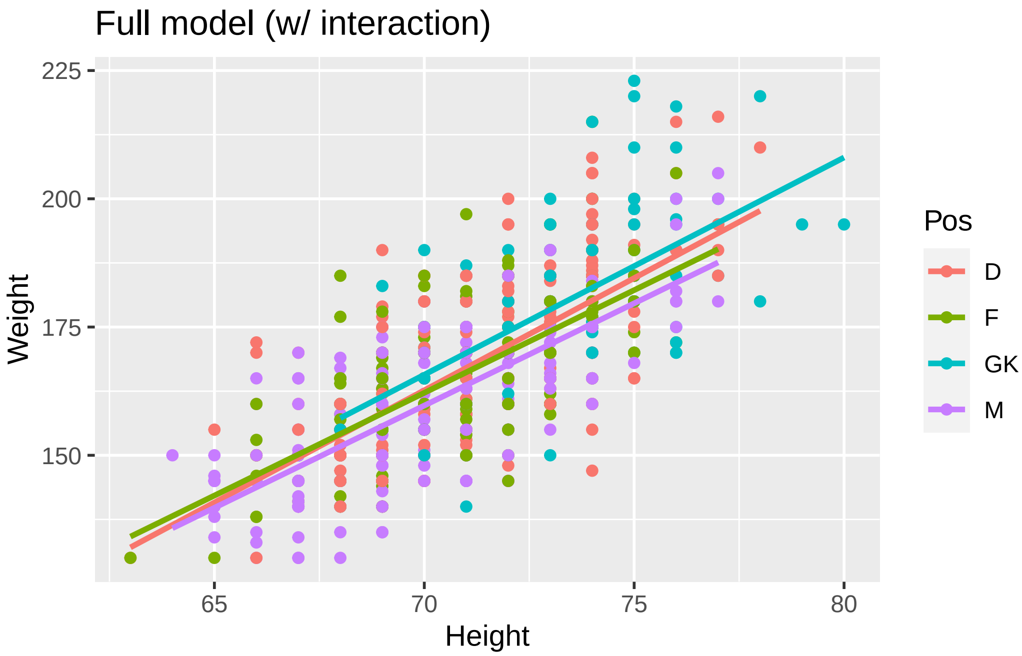
We may have the following full model and be wondering about the effect of the interaction term `Height * Pos`. Essentially, our reduced model only allows different weight against height intercept, while our full model also allows different weight against height slope.

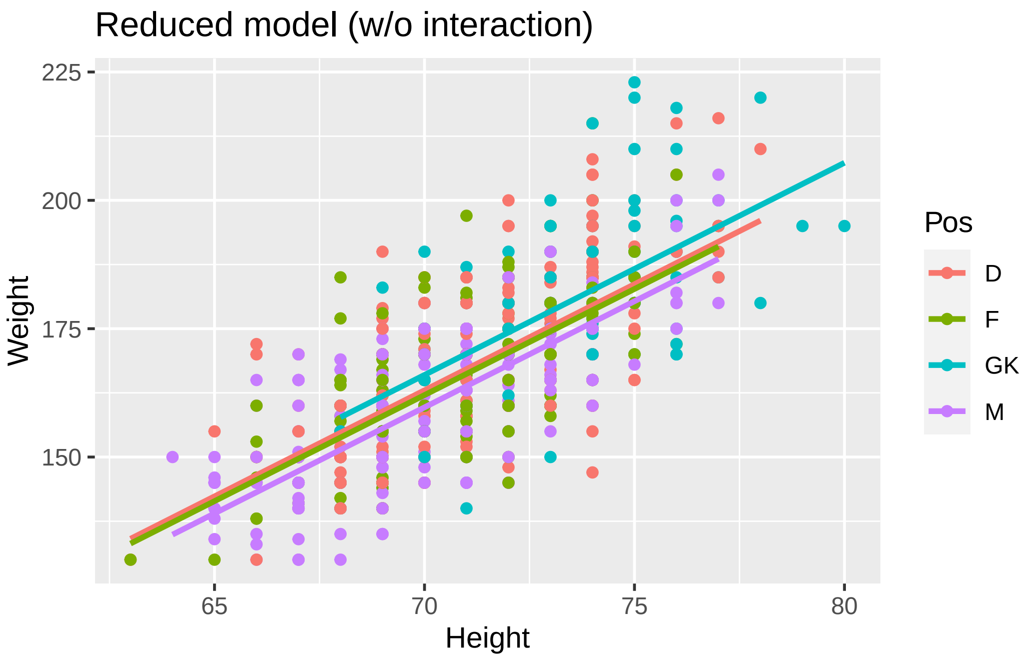
```
soccer_full = lm(Weight ~ Height + Pos + Height * Pos,
                 data = soccerplayers)
soccer_reduced = lm(Weight ~ Height + Pos,
                    data = soccerplayers)
anovaSoccer = anova(soccer_reduced, soccer_full)
print(anovaSoccer)
```

```
Analysis of Variance Table

Model 1: Weight ~ Height + Pos
Model 2: Weight ~ Height + Pos + Height * Pos
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1   1035 110509
2   1032 110329  3    180.13 0.5616 0.6404
```

```
ssFull = anovaSoccer$RSS[2]
extrass = anovaSoccer$RSS[1] - anovaSoccer$RSS[2]
dfReduced = anovaSoccer$Res.Df[1]
dfFull = anovaSoccer$Res.Df[2]
dfExtra = dfReduced - dfFull
fStat = (extrass/dfExtra)/(ssFull/dfFull)
p_value = 1-pf(fStat, dfExtra, dfReduced)

cat(sprintf("\nESS Test with ExtraSS %.1f on df %d and ssLof %.1f on df %d\n
            extrass, dfExtra, ssFull, dfFull, fStat, p_value))
```

```
ESS Test with ExtraSS 180.1 on df 3 and ssLof 110328.7 on df 1032
yields f-stats 0.5616 and p-value 0.6404
```






A hypothesis test could be constructed as following:

$H_0 : \beta_{h*pos} = 0$

$H_A : \beta_{h*pos} \neq 0$

According to the anova table, the Extra Sum of Square provided by the interaction term is 180.13.

Conducting Extra Sum of Square F-test, the `p-value = 0.6404` suggests that the extra interaction term does not explain a significant addition amount of variations in weight.

With reference to the graph, we may also observe that the difference in intercept is not significant.

## 3 Extra Sum of Squares (5a)

The Extra Sum of Squares F test allows tests of hypotheses that involve multiple β's, such as testing for a set of indicators defining multiple groups, or interaction effects. The lack of fit test is a special case of extra sum of squares, comparing the fitted model to a saturated model.

### 3.1 Setup the full vs. reduced model

Assume that we have the following linear model:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k_f} x_{k_f}$$

Let the projection matrix of the full model be $H_f$, so the error of the full model is given by

$$\hat{\epsilon} = (I - H_f)Y$$

Without loss of generality, we are interested in whether $\beta_{k_r+1} = \beta_{k_r+2} = \cdots = \beta_{k_f} = 0 \quad (k_r < k_f)$ and suggest a reduced model:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k_r} x_{k_r}$$

Let the projection matrix of the reduced model be $H_r$, so the error of the reduced model is given by

$$\hat{\epsilon} = (I - H_r)Y$$

### 3.2 Extra Sum of Squares is the sum of squares in error difference

We know that:

$$\begin{aligned}
SST_f &= SSM_f + SSE_f \\
SST_r &= SSM_r + SSE_r \\
SST_f &= SST_r = SD(y) \\
SSM_f + SSE_f &= SSM_r + SSE_r \\
SSM_f - SSM_r &= SSE_r - SSE_f \\
&= \|\hat{\epsilon}_r\|^2 - \|\hat{\epsilon}_f\|^2
\end{aligned}$$

And that:

$$\begin{aligned}
\|\hat{\epsilon}_r - \hat{\epsilon}_f\|^2 &= \|(I - H_r)Y - (I - H_f)Y\|^2 \\
&= \|(H_f - H_r)Y\|^2 \\
&= ((H_f - H_r)Y)^T((H_f - H_r)Y) \\
&= Y^T(H_f - H_r)^T(H_f - H_r)Y \\
&= Y^T(H_f^T - H_r^T)(H_f - H_r)Y \\
&= Y^T(H_f - H_r)(H_f - H_r)Y && \text{by symmetry property} \\
&= Y^T(H_f^2 - H_r H_f - H_f H_r + H_r^2)Y \\
&= Y^T(H_f - H_r - H_r + H_r)Y && \text{by idempotent property and property 1d} \\
&= Y^T(H_f - H_r)Y
\end{aligned}$$

ESS is the extra sum of squares resulting from adding one or more predictors to the existing model.

$$
\begin{aligned}
ESS &= SSM_f - SSM_r \\
&= SSE_r - SSE_f \\
&= \|\hat{\epsilon}_r\|^2 - \|\hat{\epsilon}_f\|^2 \\
&= ((I - H_r)Y)^T((I - H_r)Y) - ((I - H_f)Y)^T((I - H_f)Y) \\
&= Y^T(I - H_r)^T(I - H_r)Y - Y^T(I - H_f)^T(I - H_f)Y \\
&= Y^T(I^T - H_r^T)(I - H_r)Y - Y^T(I^T - H_f^T)(I - H_f)Y \\
&= Y^T(I - H_r)(I - H_r)Y - Y^T(I - H_f)(I - H_f)Y && \text{by symmetry property} \\
&= Y^T(I - H_rI - IH_r + H_r^2)Y - Y^T(I - H_fI - IH_f + H_f^2)Y \\
&= Y^T(I - H_r - H_r + H_r)Y - Y^T(I - H_f - H_f + H_f)Y && \text{by idempotent property} \\
&= Y^T(I - H_r)Y - Y^T(I - H_f)Y \\
&= Y^T(I - H_r - I + H_f)Y \\
&= Y^T(H_f - H_r)Y \\
&= \|\hat{\epsilon}_r - \hat{\epsilon}_f\|^2
\end{aligned}
$$

Therefore, the “extra sum of squares” is the sum of squares in error difference.

### 3.3 Independence of $ESS$ and $SSE_f$

Showing $ESS$ $(= \|\hat{\epsilon}_r - \hat{\epsilon}_f\|^2)$ is independent of $SSE_f$ $(= \|\hat{\epsilon}_f\|^2)$ is equivalent to showing that $\hat{\epsilon}_r - \hat{\epsilon}_f$ and $\hat{\epsilon}_f$ are orthogonal.

$$
\begin{aligned}
(\hat{\epsilon}_r - \hat{\epsilon}_f) \cdot \hat{\epsilon}_f &= ((H_f - H_r)Y)^T(I - H_f)Y \\
&= Y^T(H_f^T - H_r^T)(I - H_f)Y \\
&= Y^T(H_f - H_r)(I - H_f)Y && \text{by symmetry property} \\
&= Y^T(H_f - H_r - H_fH_f + H_rH_f)Y \\
&= Y^T(H_f - H_r - H_f + H_r)Y && \text{by idempotent and property 1d} \\
&= 0
\end{aligned}
$$

Therefore, $ESS$ and $SSE_f$ are independent.

### 3.4 Construction of F-statistics for ESS test

We want to perform the following hypothesis test:

$H_0 : \beta_{k_r+1} = \beta_{k_r+2} = \cdots = \beta_{k_f} = 0$

$H_A$ : at least one of $\beta_j \neq 0, \quad j \in \{k_r + 1, k_r + 2, \cdots, k_f\}$

Construct

$$F = \frac{ESS/(\mathrm{df}_f - \mathrm{df}_r)}{SSE_f/\mathrm{df}_f}$$

and we want to show that this statistics follows F-distribution under null hypothesis.

Under our assumption that error follows normal distribution with known $\sigma$:

$$\frac{SSE_f}{\sigma^2} \sim \chi^2_{(n-k_f)}$$

Under our null hypothesis $\beta_{k_r+1} = \beta_{k_r+2} = \cdots = \beta_{k_f} = 0$, the new predictors $x_{k_r+1} = x_{k_r+2} = \cdots = x_{k_f}$ carries no new information at all, suggesting that the error of the reduced model also follows the same normal distribution.

$$\frac{SSE_r}{\sigma^2} \sim \chi^2_{(n-k_r)}$$

$$\underbrace{\frac{SSE_r}{\sigma^2}}_{\sim\chi^2_{(n-k_r)}} = \underbrace{\frac{SSE_f}{\sigma^2}}_{\sim\chi^2_{(n-k_f)}} + \frac{ESS}{\sigma^2}$$

Using MGF (similar to Min's presentation on wk11-5c), $\frac{ESS}{\sigma^2} \sim \chi^2_{(k_f-k_r)}$.

Finally, by the characteristic of F-distribution:

$$\frac{\frac{ESS}{\sigma^2}/(k_f - k_r)}{\frac{SSE_r}{\sigma^2}/(n - k_r)} \sim F_{(k_f-k_r, n-k_r)}$$

# 4 Lack of Fit and ESS

## 4.1 Lack of Fit and ESS connection

In this section, we will set up Lack of Fit test as a special case of the Extra Sum of Squares test.

Regardless of our original data type of our predictors (continuous or discrete/categorical), we may treat them all as categorical. In the case of soccer player dataset, `Position` is readily a categorical variable; though our predictor `Height` is continuous integer with min 63 and max 80, we may treat each integer between 63 and 80 as their own category.

Then, for each combination of predictor values (for example `(height, Position) = (71, F)`), we will estimate:

$$\beta_{X=\vec{x}} = \hat{y} = \overline{y|X = \vec{x}}$$

Credit Sunny's presentation wk11-3.2, we can setup Lack of Fit:

$H_0$ : The linear model is correct.

$H_A$ : The data does not follow linear trend: $y = \overline{y|X = \vec{x}} = \sum_{\vec{x}} \beta_{X=\vec{x}} I(X = \vec{x})$, where $I(X = \vec{x})$ is an indicator variable which evaluates 1 if $X$ is exactly equal to $\vec{x}$ category, and 0 if otherwise.

In this case, our alternative model $y = \overline{y|X = \vec{x}}$ can either be interpreted as a saturated model, or a full model $y = \sum_{\vec{x}} \beta_{X=\vec{x}} I(X = \vec{x})$

```
soccerplayers$HeightCat = as.character(soccerplayers$Height)
soccerplayers$HeightPosCat = paste(soccerplayers$HeightCat,
                                   soccerplayers$Pos, sep = "_")
soccer_saturated = lm(Weight ~ HeightPosCat,
                      data = soccerplayers)
cat(sprintf("In this data with %d entries, \nthere are %d unique height and
        nrow(soccerplayers),
        length(unique(soccerplayers$HeightPosCat))))
```

```
In this data with 1040 entries,
there are 56 unique height and position combinations.
```

```
print(anova(soccer_reduced, soccer_saturated))
```

```
Analysis of Variance Table

Model 1: Weight ~ Height + Pos
Model 2: Weight ~ HeightPosCat
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1   1035 110509
2    984 104787 51    5721.5 1.0535 0.3748
```

Briefly, the result shows that considering height value as their own category explain a lot of additional variability. Therefore, we would consider a linear model as more appropriate.

## 5 Lack of Fit and Projection Matrix

Suppose for a simple regression there are $n$ individuals with m distinct $x$ values $(1 < m < n)$. Define $\bar{Y}_j$ to be the average $Y$ value for the $n_j$ individuals with covariate $x_j$ , $j = 1, ..., m$. Define $H$ to be the usual regression projection matrix that projects the $n \times 1$ vector $Y$ to the p dimensional subspace spanned by the columns of the $X$ matrix. Define $H_s$ to be the projection matrix for the saturated model, projecting $Y$ to the vector of averages $\bar{Y}_j$, with each $\bar{Y}_j$ is replicated $n_j$ times, for $j = 1, ..., m$.

$$Y^T(I-H)Y/\sigma^2 = Y^T(I-H_s+H_s-H)Y/\sigma^2 = \underbrace{Y^T(I-H_s)Y/\sigma^2}_{\text{SS Pure}} + \underbrace{Y^T(H_s-H)Y/\sigma^2}_{\text{SS Lack of Fit}}$$

Assume that the null hypothesis (the simple regression) is true, then according to Soph's presentation wk12-4

$$\frac{Y^T(I-H)Y}{\sigma^2} \sim \chi^2_{(n-p)}$$
$$\frac{Y^T(I-H_s)Y}{\sigma^2} \sim \chi^2_{(n-m)}$$

By equation above and MGF expansion, we know that $Y^T(H_s - H)Y/\sigma^2 \sim \chi^2(m-p)$

Because we know that

$$(I - H_s)(H_s - H) = H_s - H - H_s^2 + H_sH = H_s - H - H_s + H = 0$$

suggesting $Y^T(I - H_s)Y/\sigma^2 \perp Y^T(H_s - H)Y/\sigma^2$

Therefore, we can set up the statisitcs:

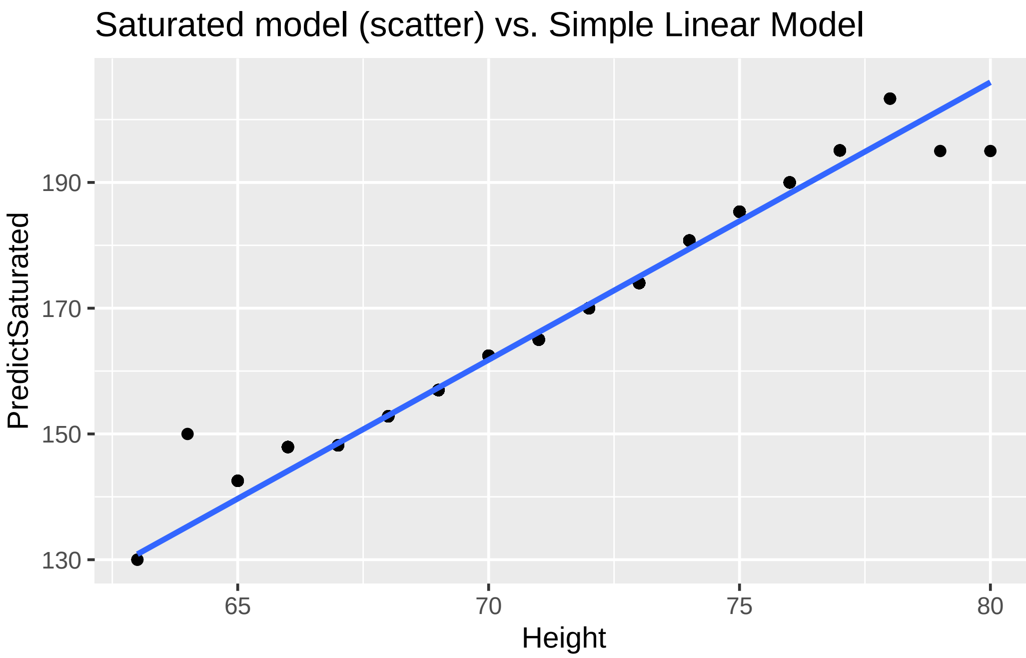
$$\begin{aligned}
F &= \frac{MS_{\text{LackOfFit}}}{MS_{\text{pure}}} \\
&= \frac{Y^T(H_s - H)Y/(\sigma^2 * (m-p))}{Y^T(I - H_s)Y/(\sigma^2 * (n-m))} \\
&= \frac{Y^T(H_s - H)Y/(m-p)}{Y^T(I - H_s)Y/(n-m)} \sim F_{((m-p),(n-m))}
\end{aligned}$$

```
soccerWH_linear = lm(Weight ~ Height, data = soccerplayers)
soccerWH_saturate = lm(Weight ~ HeightCat, data = soccerplayers)
anovaLinearSaturate = anova(soccerWH_linear, soccerWH_saturate)
print(anovaLinearSaturate)
```

```
Analysis of Variance Table

Model 1: Weight ~ Height
Model 2: Weight ~ HeightCat
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1   1038 114596
2   1022 112799 16    1796.9 1.0175 0.4348
```

```
ssPure = anovaLinearSaturate$RSS[2]
ssLoF = anovaLinearSaturate$RSS[1] - anovaLinearSaturate$RSS[2]
dfPure = anovaLinearSaturate$Res.Df[2]
dfExtra = anovaLinearSaturate$Res.Df[1] - anovaLinearSaturate$Res.Df[2]
fStat = (ssLoF/dfExtra)/(ssPure/dfPure)
p_value = 1-pf(fStat, dfExtra, dfPure)

cat(sprintf("\nLoF Test with ssPure %.1f on df %d and ssLof %.1f on df %d\ny
            ssPure, dfPure, ssLoF, dfExtra, fStat, p_value))
```

```
LoF Test with ssPure 112799.3 on df 1022 and ssLof 1796.9 on df 16
yields f-stats 1.0175 and p-value 0.4348
```


Saturated model (scatter) vs. Simple Linear Model
PredictSaturated
190
170
150
130
65
70
75
80
Height

# 6 Appendix

```r
print(anova(soccer_full))
```

```
Analysis of Variance Table

Response: Weight
            Df Sum Sq Mean Sq   F value    Pr(>F)    
Height       1 133284  133284 1246.7172 < 2.2e-16 ***
Pos          3   4087    1362   12.7442 3.503e-08 ***
Height:Pos   3    180      60    0.5616    0.6404    
Residuals 1032 110329     107                        
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
print(anova(soccer_reduced))
```

```
Analysis of Variance Table

Response: Weight
            Df Sum Sq Mean Sq F value    Pr(>F)    
Height       1 133284  133284 1248.30 < 2.2e-16 ***
Pos          3   4087    1362   12.76 3.421e-08 ***
Residuals 1035 110509     107                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
print(anova(soccerWH_linear))
```

```
Analysis of Variance Table

Response: Weight
            Df Sum Sq Mean Sq F value    Pr(>F)    
Height       1 133284  133284  1207.3 < 2.2e-16 ***
Residuals 1038 114596     110                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
print(anova(soccerWH_saturate))
```

```
Analysis of Variance Table

Response: Weight
            Df Sum Sq Mean Sq F value    Pr(>F)    
HeightCat   17 135080  7945.9  71.993 < 2.2e-16 ***
Residuals 1022 112799   110.4                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```