

Lack of Fit Test

• **General ANOVA Table for Testing Lack of Fit of Simple Linear Regression**

Source	DF	Sum of Squares	Mean Square	F
Regression	1	$SSR = \sum_j \sum_i (\hat{Y}_{ij} - \bar{Y})^2$	$MSR = \frac{SSR}{1}$	$F_{\text{obs}} = \frac{MSG}{MSE}$
Error	$n - 2$	$SSE = \sum_j \sum_i (Y_{ij} - \hat{Y}_{ij})^2$	$MSE = \frac{SSE}{n - 2}$	
Lack of Fit	$c - 2$	$SSLF = \sum_j \sum_i (\bar{Y}_j - \hat{Y}_{ij})^2$	$MSLF = \frac{SSLF}{c - 2}$	$F^* = \frac{MSLF}{MSPE}$
Pure Error	$n - c$	$SSPE = \sum_j \sum_i (Y_{ij} - \bar{Y}_j)^2$	$MSPE = \frac{SSPE}{n - c}$	
Total	$n - 1$	$SSTO = \sum_j \sum_i (Y_{ij} - \bar{Y})^2$		

Note: $SSTO = SSR + SSE$ and $SSE = SSLF + SSPE$.

• **Full Model.** That is,

$$Y_{ij} = \mu_j + \epsilon_{ij}, \quad i = 1, 2, \dots, n.$$

where:

1. μ_j are parameters, $j = 1, \dots, c$.
2. ϵ_{ij} are independent $N(0, \sigma^2)$.

• **Reduced Model.** That is,

$$Y_{ij} = (\beta_0 + \beta_1 x_i) + \epsilon_i, \quad i = 1, 2, \dots, n.$$

where:

1. β_0 and β_1 are parameters.
2. x_i is a known constant.
3. ϵ_i are independent $N(0, \sigma^2)$.

• Test Statistic:

$$\begin{aligned}
 F^* &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\
 &= \frac{SSE - SSPE}{(n - 2) - (n - c)} \div \frac{SSPE}{n - c} \\
 &= \frac{SSLF}{c - 2} \div \frac{SSPE}{n - c} \\
 &= \frac{MSLF}{MSPE} \sim f_{(c-2, n-c)}
 \end{aligned}$$

• Expected Mean Squares

1. $E(MSPE) = \sigma^2$
2. $E(MSLF) = \sigma^2 + \frac{\sum n_j [\mu_j - (\beta_0 + \beta_1 x_j)]^2}{c - 2}$

- Example: Bank data (page 120)

```
x=c(125,100,200,75,150,175,75,175,125,200,100)
y=c(160,112,124,28,152,156,42,124,150,104,136)
```

```
results=lm(y~x)
anova(results)
```

```
reduced.mod=lm(y~x)
full.mod=lm(y~factor(x))
anova(reduced.mod,full.mod)
```

```
>anova(results)
Analysis of Variance Table
```

```
Response: y
          Df Sum Sq Mean Sq F value Pr(>F)
x           1  5141.3   5141.3    3.1389 0.1102
Residuals    9 14741.6   1638.0
```

```
> anova(reduced.mod,full.mod)
Analysis of Variance Table
```

```
Model 1: y ~ x
Model 2: y ~ factor(x)
      Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1          9 14742
2          5  1148  4     13594 14.801 0.005594 **
```

Source	DF	Sum of Squares	Mean Square	<i>F</i>
Regression				
Error				
Lack of Fit				
Pure Error				
Total				

ANOVA Table for Testing Lack of Fit of Simple Linear Regression

- **Homework # 7 (due: Friday, Sept. 30):** Answer the following questions:
On page 150, # 3.14 (part a) and # 3.15 (all parts).