ANOVA-F test in Regression

An ANOVA-F test can be constructed to test overall (*global*) fit of the linear regression model.

The decomposition of sums of squares for regression takes the form

$$SS = SSR + SSE$$

where

- SS: overall or total sum of squares
- SSR: sum of squares due to Regression
- SSE: sum of squares due to Error

$$\begin{aligned} SS &= \sum_{i=1}^{n}(y_i - \overline{y})^2 \\ SSR &= \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \\ SSE &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \end{aligned}$$

where

$$\hat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i \qquad i = 1, \ldots, n$$

Degrees of Freedom

- TOTAL: $n - 1$
- REGRESSION: 1
- ERROR: $n - 2$

(error d.f. is $n - p$, here $p = 2$).

The ANOVA Table

SOURCE	DF	SS	MS	F
REGRESSION	1	SSR	$MSR = \frac{SSR}{1}$	$F = \frac{MSR}{MSE}$
ERROR	$n-2$	SSE	$MSE = \frac{SSE}{(n-2)}$	
TOTAL	$n-1$	SS		

The test of the hypothesis

$$\begin{aligned} H_0 &: E[Y] = \beta_0 \\ H_a &: E[Y] = \beta_0 + \beta_1 x \end{aligned}$$

can be completed by using the test statistic

$$F = \frac{MSR}{MSE}$$

If H_0 is true

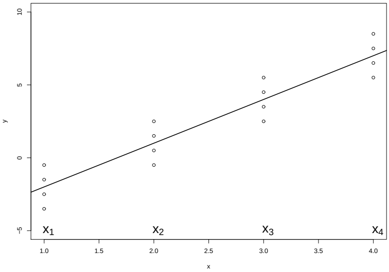
$$F \sim \text{Fisher-F}(1, n-2)$$

This is just like the ANOVA in the one-way layout (CRD) with n groups, but where

$$\mu_i = \beta_0 + \beta_1 x_i$$

That is, the group means are **structured**, that is, we have a formula relating the μ_i quantities.

Consider four replicates at x values (x_1, x_2, x_3, x_4) in a regression;



Then for group i, $\mu_i = \beta_0 + \beta_1 x_i$, $i = 1, 2, 3, 4$.

Checking the Local Fit

A plot of the *residuals*

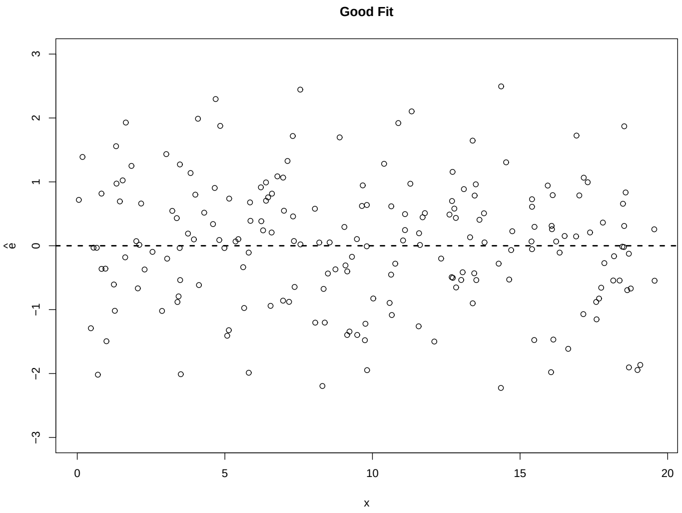
$$\hat{e}_i = y_i - \hat{y}_i$$

can reveal model inadequacies. We should observe that in plots of

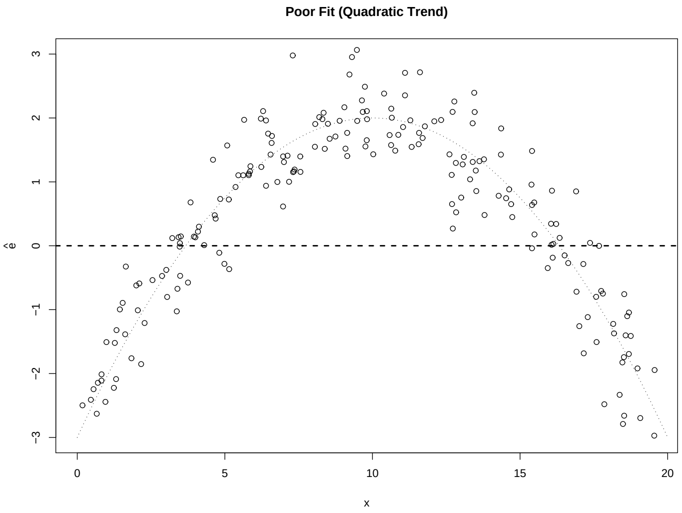
- x vs $\hat{e}$
- y vs $\hat{e}$
- $\hat{y}$ vs $\hat{e}$

there is no discernible pattern

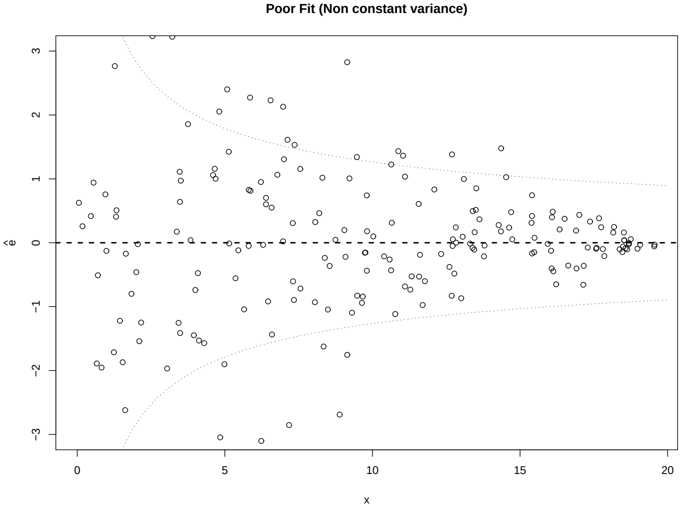
Checking the Local Fit: Good Fit



Checking the Local Fit: Poor Fit



Checking the Local Fit: Poor Fit



R^2 and adjusted R^2

SPSS reports both the R^2 statistic

$$R^2 = 1 - \frac{SSE}{SS}$$

and the **adjusted** R^2 statistic

$$R^2 = 1 - \frac{SSE/EDF}{SS/TDF}$$

where

- EDF = error degrees of freedom = $n - 2$
- TDF = total degrees of freedom = $n - 1$