Sum of Squares Decomposition &

the F-test

An alternative way to test whether a predictor X is "significant" for explaining Y is through a phenomenon called the "sum of squares decomposition".

The decomposition gives us a break down of the total variance of Y $(\{y_i\}_{i=1}^n)$ ← SST into 2 parts:

1. the sum of squared errors (SSE)
2. the regression sum of squares. (SSR)

Mathematically, we say:

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_i e_i^2$$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

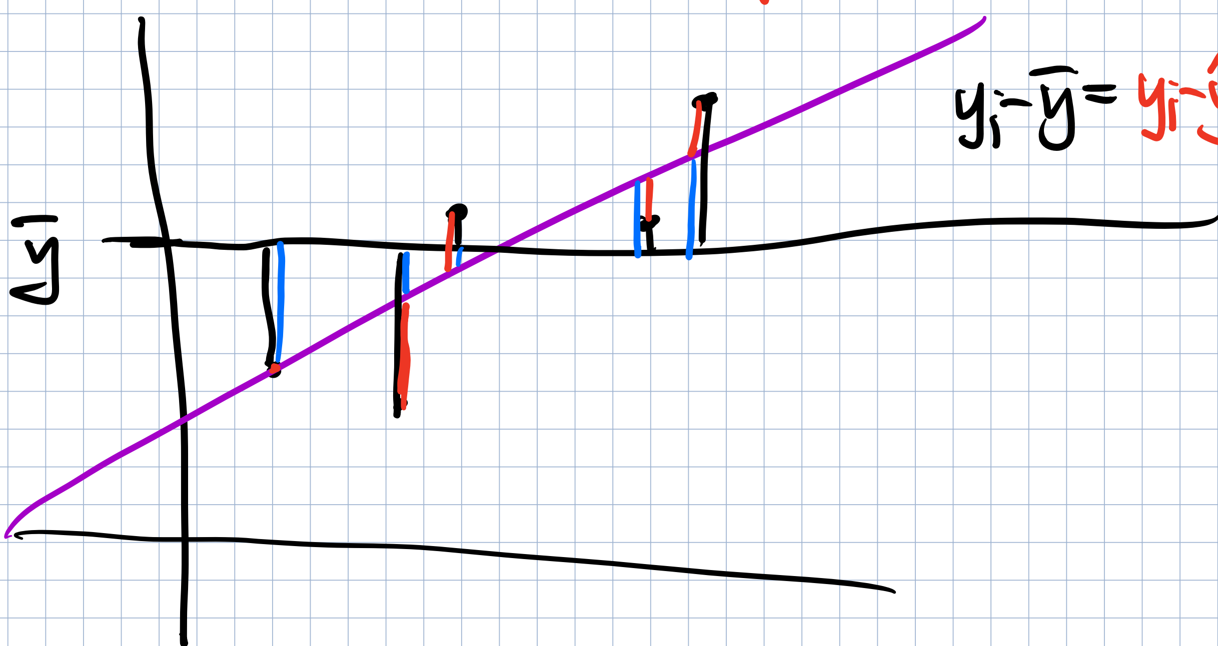
It turns out:

$$SST = SSE + SSR$$

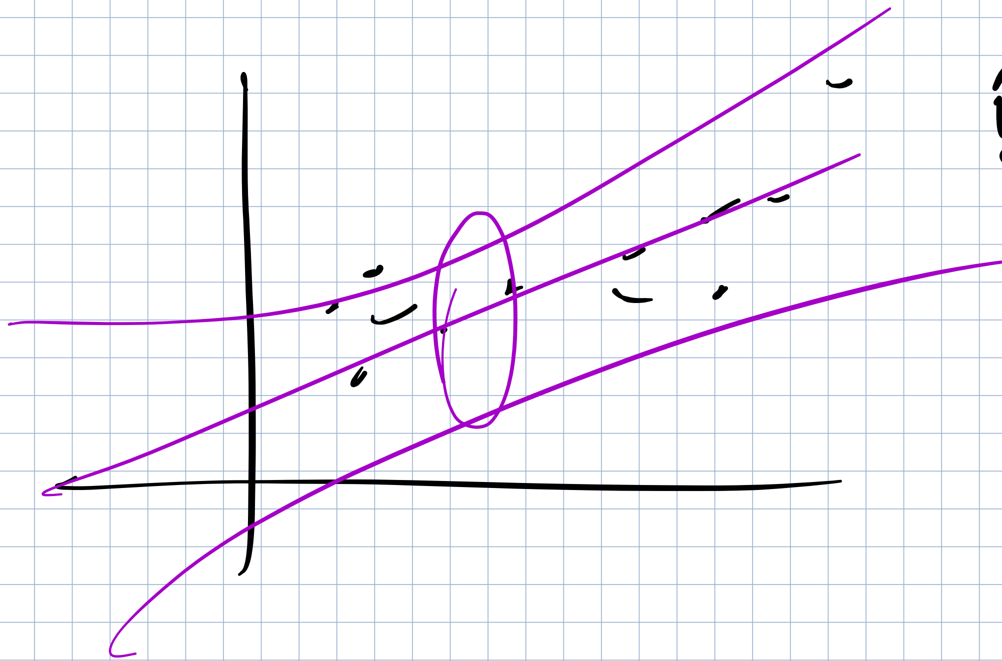
$$\Leftrightarrow \left[Var(y_i) = Var(e_i) + Var(\hat{y}_i) \right]$$

$$\underbrace{\sum_i (y_i - \bar{y})^2}_{\text{black}} = \underbrace{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}_{\text{red}} + \underbrace{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}_{\text{blue}}$$

(prove this later)



$$y_i - \bar{y} = y_i - \hat{y}_i + \hat{y}_i - \bar{y}$$



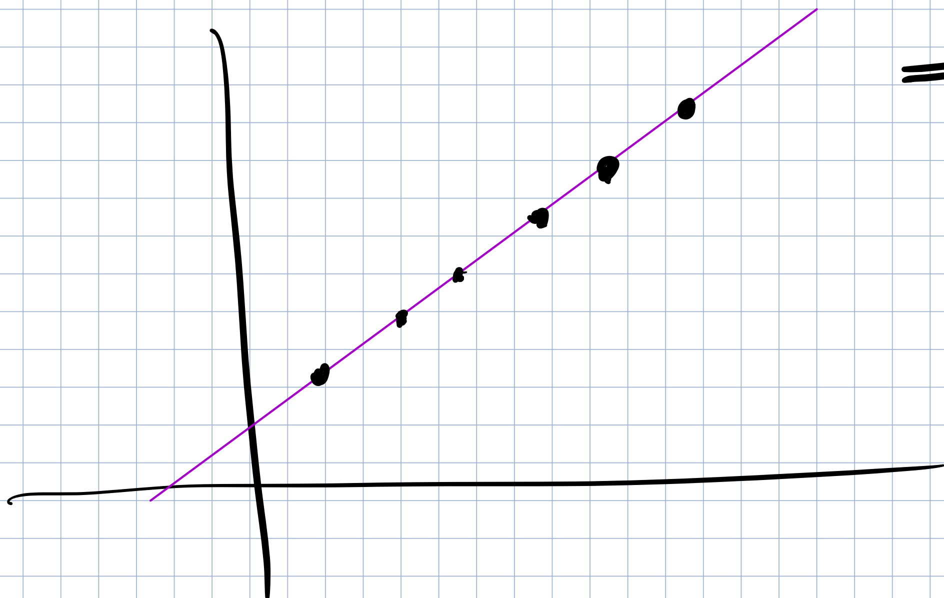
$$\hat{y}_i \sim N\left(\beta_0+\beta_1 x_i,\ \underbrace{\sigma^2\left(\frac{1}{n}+\frac{(x_i-\bar{x})^2}{SSx}\right)}_{Var(\hat{y}_i)}\right)$$

Let's construct a statistical test which decides whether X is a sig. predictor of y:

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

To build intuition let me think about extreme cases of the SS decomp.

① our predictions are perfect:



$\Rightarrow$ $y_i = \hat{y}_i \quad \forall i$

$\forall$ "for all"

$$\sum_i (y_i-\bar{y})^2 = \sum_{i=1}^{n} (y_i-\hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i-\bar{y})^2$$

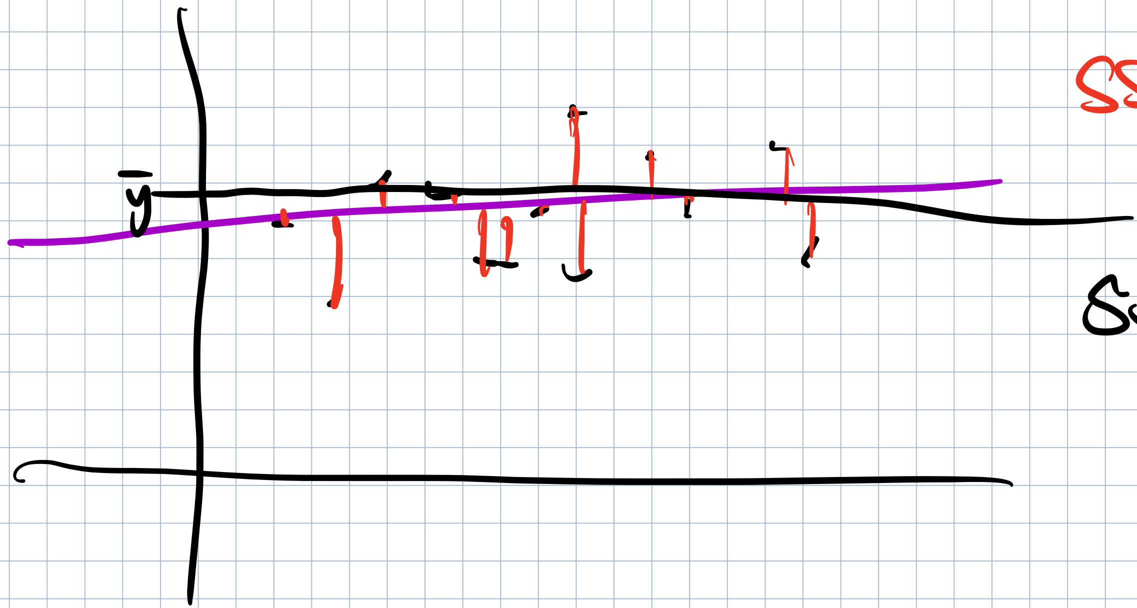
$$SST = 0 + SSR$$

$\Longleftrightarrow$ X predicts Y extremely well

$\Updownarrow$

we want to reject H_0

② Our predictions are terrible



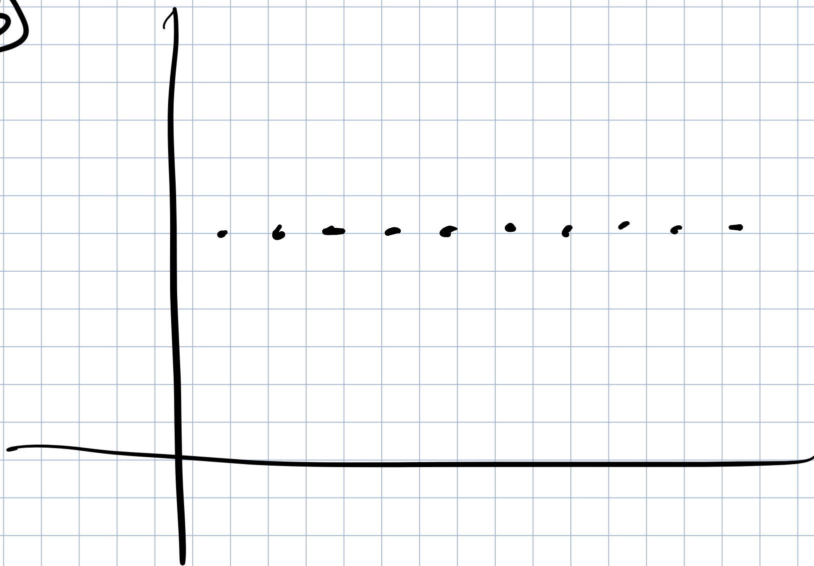
SSE ↑↑

$SSR \approx 0$

$\iff$ X is not useful as a predictor

$\iff$ hyp. test would fail to reject H_0.

③



With this intuition now we can try to develop a rigorous test:

I will tell you:

① $\frac{SSE}{\sigma^2} \sim \chi^2_{n-2}$

② $\frac{SSR}{\sigma^2} \sim \chi^2_1$

③ $\frac{SST}{\sigma^2} \sim \chi^2_{n-1}$

④ $SSE \perp\!\!\!\perp SSR$

Aside: F-dist.

If ① $U \sim \chi^2_n$ & ② $V \sim \chi^2_m$ &

③ $U \perp\!\!\!\perp V$ then

$\perp\!\!\!\perp$ means independence

$$F = \frac{U/n}{V/m} \sim F_{n,m}$$

lets define our statistic as:

$$F = \frac{SSR/1}{SSE/n-2} \sim F_{1,n-2}$$

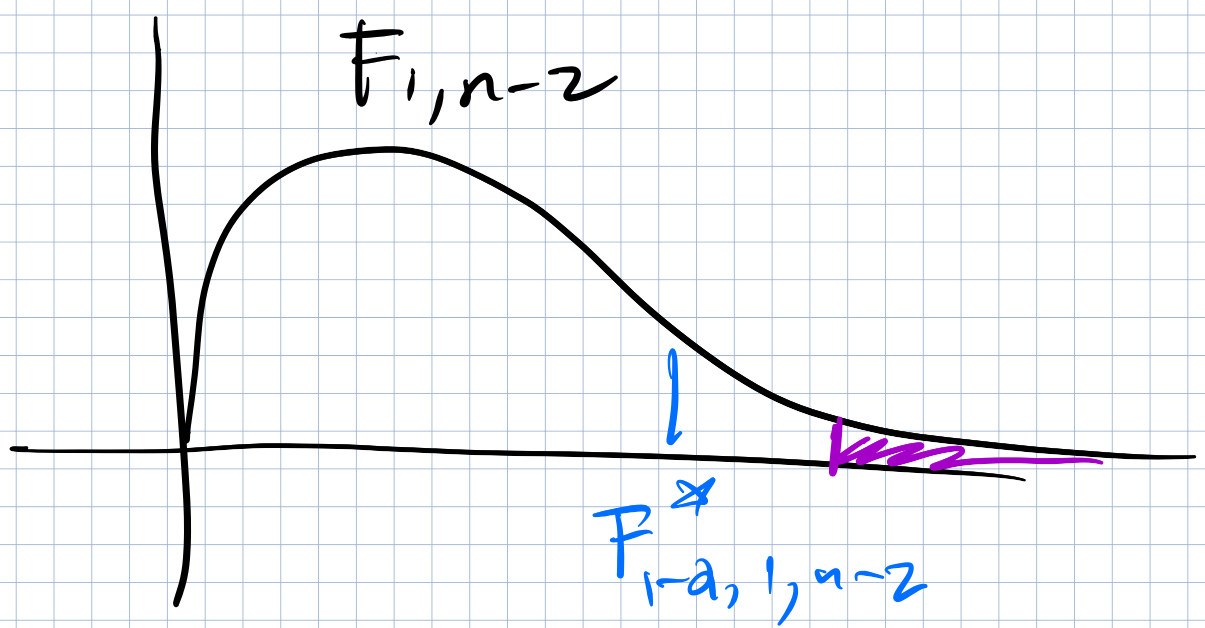
$$= \frac{MSR}{MSE}$$ (an alternative naming convention)

Decision Rule:

e.g. $F^*_{.95, 1, 98}$

If $F > F^*_{1-\alpha, 1, n-2}$ then

reject $H_0 \iff X$ is a sig. predictor of y.



Intuition ~ df:

estimated 1 parameter

by

hand

$\Downarrow$

lose 1 df

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$df = n-1$$

2 parameters

estimated

$\Downarrow$

lose

2 df

$$SSE = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$df = n-2$$

$$SST = SSE + SSR$$

$$\chi^2_{n-1} = \chi^2_{n-2} + \chi^2_1$$

Coefficient of Determination

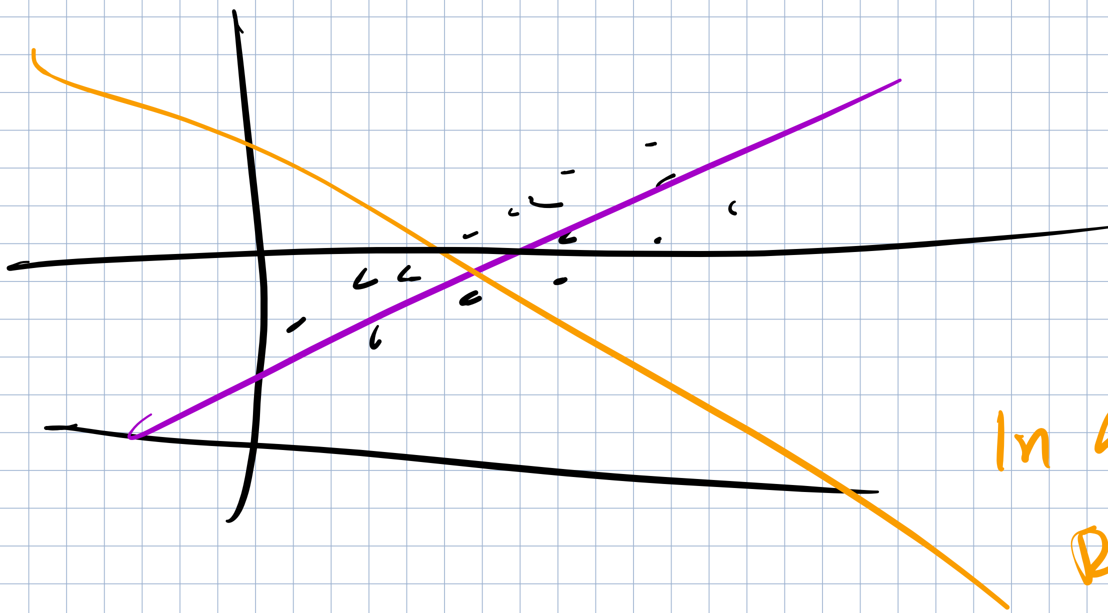
The sum of squares decomposition can also be used to evaluate the "goodness-of-fit" of a model.

To do this, define the coef. of determination, R^2 as:

$$R^2 := \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Facts about R^2:

1. For SLR, $0 \leq R^2 \leq 1$, if we use LS estimators.



In general R^2 doesn't have to be positive if our model is worse than just guessing $\hat{y}_i = \bar{y}$ across the board.

② In SLR, it is true that

$R^2 = r^2$ where r is sample correlation b/w $\{(x_i, y_i)\}_{i=1}^n$

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

Ex: If I tell you

① $SSE = 4$

② $SST = 10$

③ $\text{sign}(\hat{\beta}_1) < 0$, you can

figure out r:

$$r = \pm\sqrt{R^2} = \pm\sqrt{1 - SSE/SST}$$

$$= \pm\sqrt{1 - 4/10}$$

$$= -\sqrt{.6}$$

$\text{sign}(\hat{\beta}_1) < 0$

Interpreting R^2:

" (R^2) of the variation in [y-variable] is explained by the least squares regression line."