# Sum of Squares Decomposition & the F-test

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

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

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

(SSR)

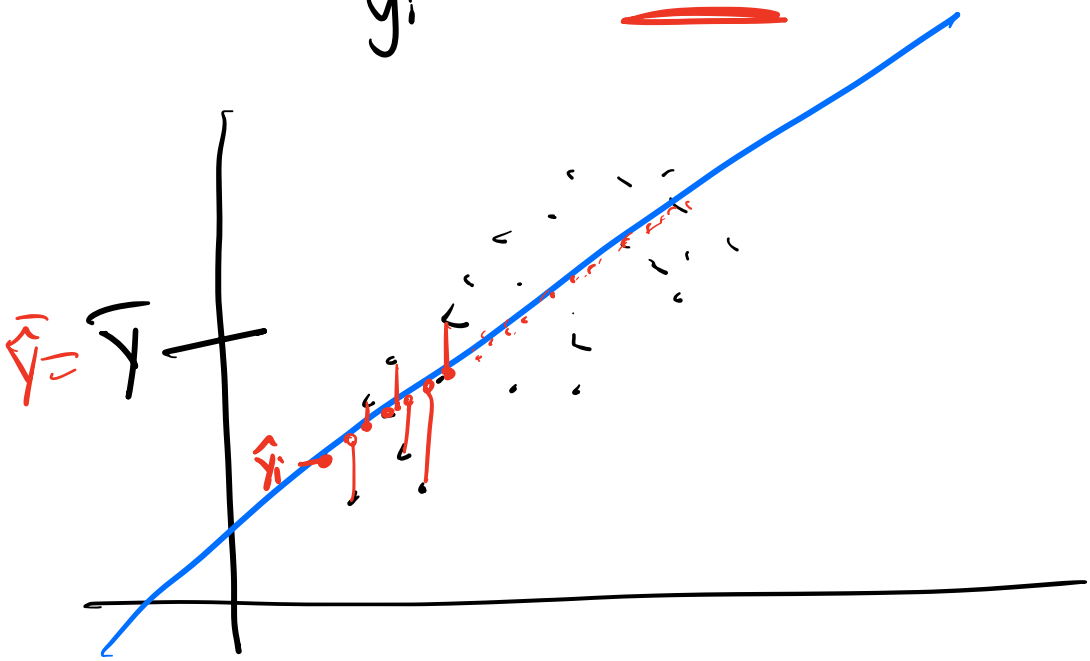
Mathematically:

Total sum of squares in Y = Regression Sum of sq. + Sum of squared error

SST = SSR + SSE

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

$y_i$ ↑ $\hat{y}_i$ — SSR — SSE



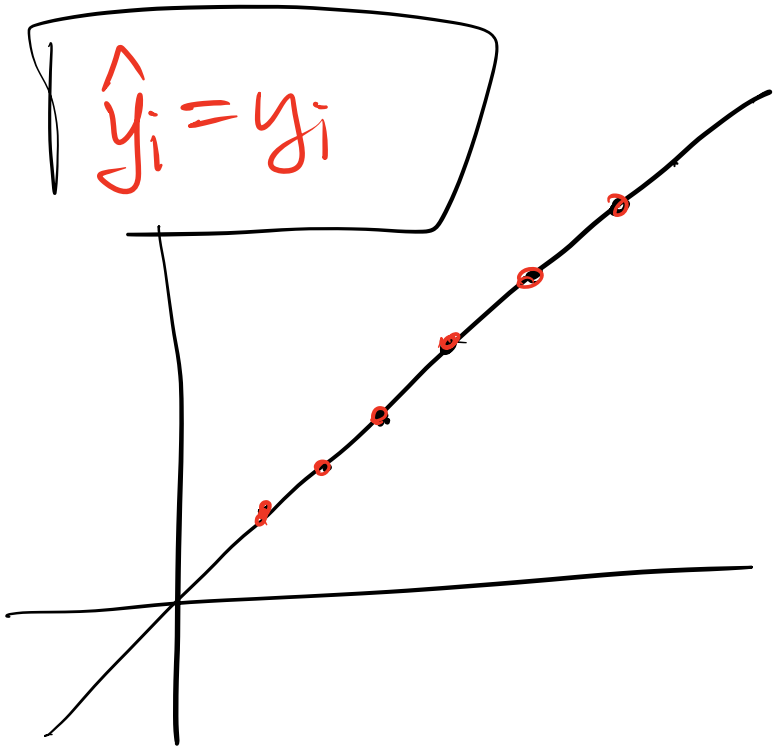
Let's construct a statistic which tests whether $X$ is a significant predictor of $y$:

$$H_0: \beta_1 = 0 \quad \text{vs.} \quad H_1: \beta_1 \neq 0.$$

To build intuition think about the extremes:

(1) our predictions are perfect

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



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

$\hat{y}_i$ ↑ — SSR — SSE ↑

If $SSE = 0$

$SSR = SST$ ⇔

Intuition:
X predicts Y well

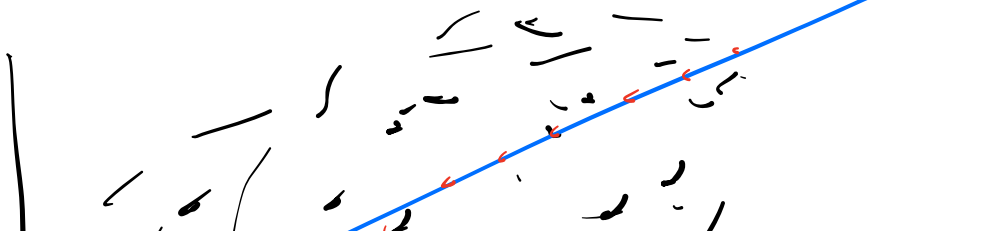
⇑

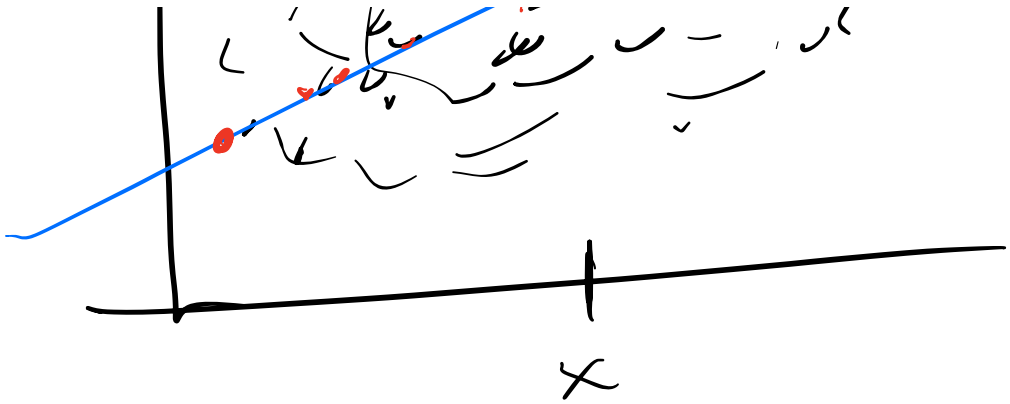
Reject $H_0$ in the t-test.

Other Extreme:

(2) If X was a really bad predictor of y.....

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






$SST = SSR + SSE$

$SSR \approx 0$ ; $SSE$ large

$\Rightarrow$ $X$ is not useful as a predictor

$\Leftrightarrow$ t-test would fail to reject

theoretical: $Var(y_i) = Var(\hat{y}_i) + Var(e_i)$

LSRL

cf. sample: $SST = SSR + SSE$

Exercise!

Argue that $SST = SSR + SSE$.

WTS: $\sum_i d_i^2 + \sum_i e_i^2$

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

Step 1

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

$$= \sum_{i=1}^n (e_i^2 + 2e_i d_i + d_i^2)$$

$$= \sum_{i=1}^n e_i^2 + 2\sum_{i=1}^n e_i d_i + \sum_{i=1}^n d_i^2$$

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

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

$$= \sum_{i=1}^{n} e_i \hat{y}_i - \sum_{i=1}^{n} e_i \bar{y}$$

$$= \sum_{i=1}^{n} e_i \hat{y}_i - \bar{y} \sum_{i=1}^{n} e_i = 0$$

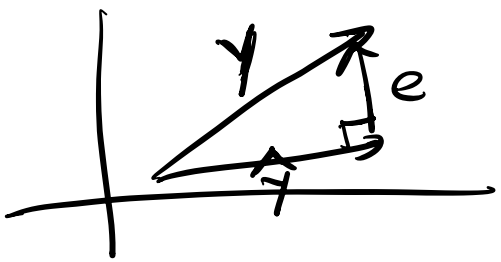
$$= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

Geometric Concept

$Y = (y_1, \ldots, y_n)$ ← observed data

$Y = \hat{Y} + e \Longrightarrow$

$(\hat{y}_1 \cdots \hat{y}_n) \quad (e_1 \cdots e_n)$




I will tell you that

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

(2) $\frac{SSR}{\sigma^2} \sim \chi^2_1$, &

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

To construct a statistic which follows an F-dist, we can take:

$$F = \frac{SSR/1}{SSE/n-2} = \frac{MSR}{MSE}$$

# F Table:

| SS | notation | df | mean square |
|---|---|---|---|
| Reg. | SSR | 1 | $MSR = SSR/1$ |
| Err. | SSE | $n-2$ | $MSE = SSE/n-2$ |
| Tot. | SST | $n-1$ | |

$$F = \frac{MSR}{MSE} \overset{H_0}{\sim} F_{1,n-2}$$

high F-vals $\Rightarrow$ reject $H_0$

Decision Rule:

If $F > F^{*}_{1,n-2}(1-\alpha)$ then reject $H_0$.

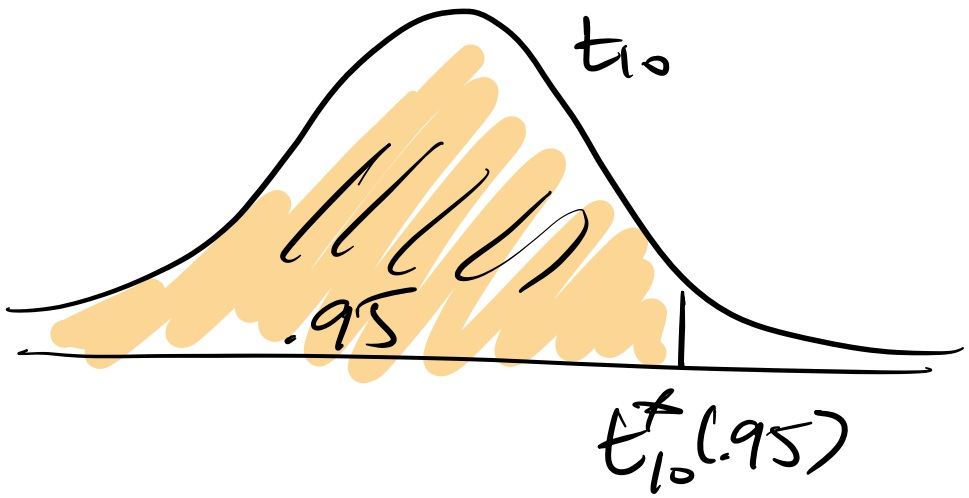
$\Rightarrow$ X is a significant predictor of Y.

What are the $^*$s denoting?

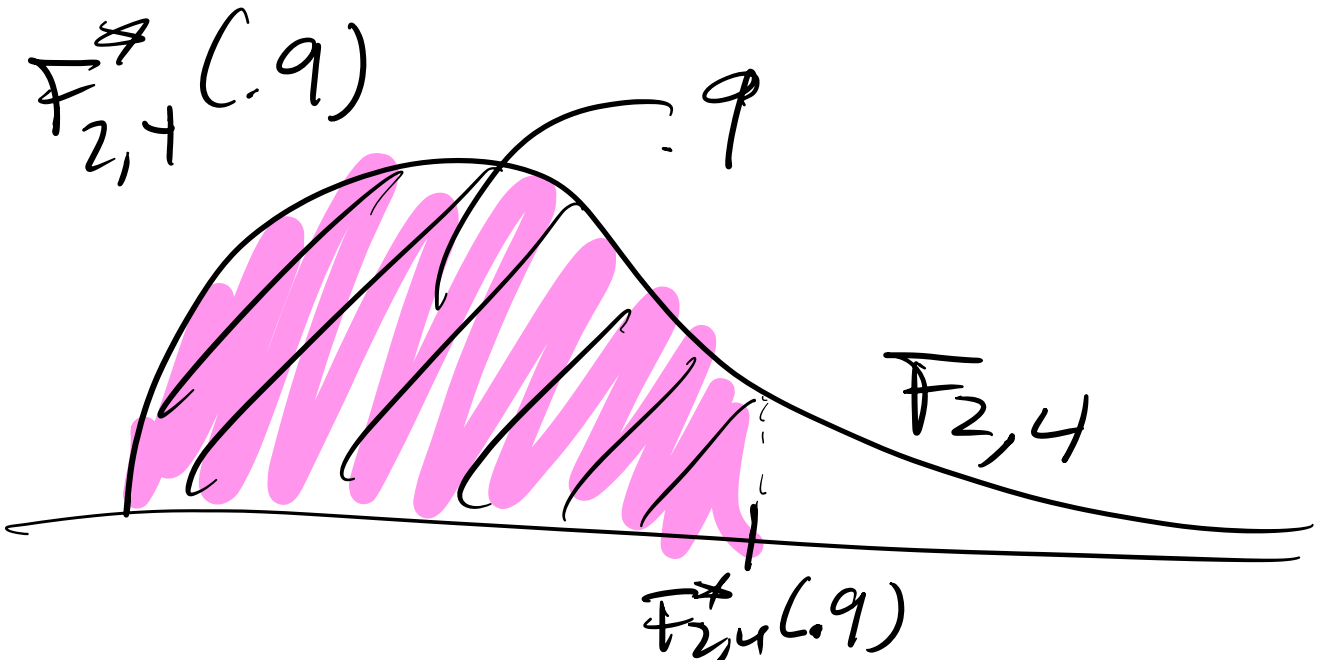
"critical values" or "quantiles."

EX:

$t^*_{10}(.95)$




$F^*_{2,4}(.9)$



# Coefficient of Determination

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

To do this we define the coefficient of determination, $R^2$:

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

Facts about $R^2$:

(1) $0 \leq R^2 \leq 1$ for SLR

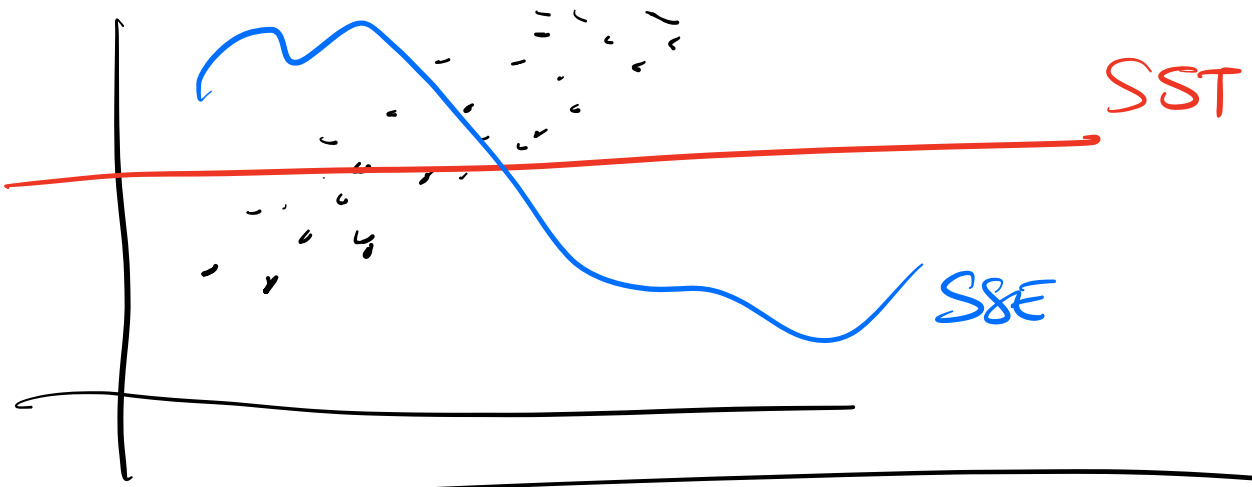
In general, $R^2$ doesn't have to be positive if we take

$$R^2 = 1 - SSE/SST.$$

---

When would $SSE > SST$

$(\Rightarrow R^2 < 0)$




---

② In SLR, we have

$R^2 = r^2$ where $r$ = corr. coeff.

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Consequence:

EX: If I give you $SSE = 4$, $SST = 10$ & $\text{sign}(\hat{\beta}_1) < 0$, you can figure out $r$.

EX: $r = \pm\sqrt{R^2}$ ?

$R^2 = 1 - SSE/SST$

$= 1 - 4/10 = .6$

$r = -\sqrt{.6}$

Interpreting $R^2$:

"$(R^2)$% of variation in $y$ (observed) is explained by the LSRL" (fitted)

the variation in $x$