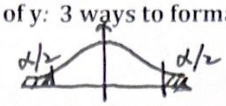
The following 3 questions are equivalent for $\underline{\underline{SLR}}$: (associated) → The line is useful for prediction.

1. Is x a good predictor of y? 2. Are x and y ~~independent~~? 3. Is ~~sb~~ the slope SIG diff from zero?

Tests to determine if x is a GOOD predictor of y: 3 ways to formally test this.

$(1-\alpha)$ two-sided



1. **Confidence Interval for β:**

$$\text{estimator} \pm t \cdot \text{stderror}$$

$$b \pm t_{n-2,\frac{\alpha}{2}} \cdot \cancel{Sd_b}\, s.e._b$$

$b \to r\frac{S_y}{S_x}$

$s.e._b$ → from computer output $\left(\frac{\sqrt{MSE}}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}}\right)$

$t_{n-2,\frac{\alpha}{2}}$ → # from t-table w/ desired confidence & df from ERROR.

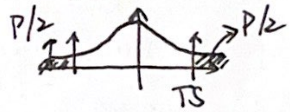
β sig diff from zero if CI does NOT include zero.

2. **t-test for β:**

$H_0: \beta = 0 \qquad H_a: \beta \neq 0$

$$TS: \quad t = \frac{\text{estimator} - \#}{\text{stderr}} = \frac{b-0}{s.e._b} = \frac{b}{s.e._b}$$

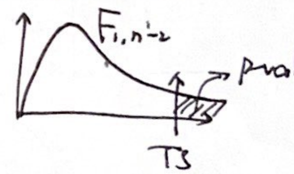
p-value: $TS \sim t_{n-2}$ when H_0 is True.



β sig diff from zero if p-value $\leq \alpha$.

3. **ANOVA Test:**

Source	df	SS	$MS(=\frac{SS}{df})$	F	p-val
Regression	1 (only one predictor)	SSR	MSR	$\frac{MSR}{MSE}$ ↓ test statistic.	$P(F_{1,n-2} \geq F)$ $(F \sim F_{1,n-2})$
Error	n-2	SSE	MSE		
Total	n-1	SST			



$H_0: \beta = 0 \qquad H_a: \beta \neq 0$

$TS: F = \frac{MSR}{MSE} \qquad$ p-val $= P(Y \geq \frac{MSR}{MSE})$ where $Y \sim F_{1,n-2}$

β sig diff from zero if p-val $\leq \alpha$.

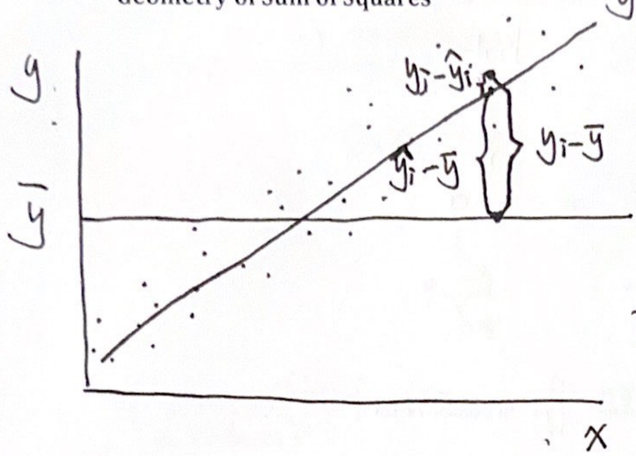
NOTE: Relationship between R^2 and SS:

$$R^2 = (r)^2 = \frac{SSR}{SST} = \%\text{ of total variability in } y \text{ explained by the regression model.}$$

* Note: If $X \sim t_N$ (N is some positive integer) then $X^2 \sim F_{1,N}$.

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

Geometry of Sum of Squares



Total variability		Random Error Variability		Variability Explained by the Reg model
$\sum (y_i - \bar{y})^2$		$\sum (y_i - \hat{y}_i)^2$		$\sum (\hat{y}_i - \bar{y})^2$
SST	=	SSE	+	SSR

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

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

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

$= SSE + SSR + 2 \underbrace{\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})}_{=0}$

Example: Can we predict final grades (y) in STA 3024 from Exam 1 scores (x)? Suppose we have the following data from previous semesters with n = 436 students.

Exam 1	Final Grades	Correlation
$\bar{x} = 79.183$	$\bar{y} = 85.412$	$r = 0.819$
$S_x = 13.418$	$S_y = 10.047$	

- Compute the LSR equation.

Slope $\quad b = r \frac{S_y}{S_x} = 0.819 \times \frac{10.047}{13.418} = 0.6132$ $\qquad$ SLR: $\hat{y} = 36.85 + 0.6132x$

y-int $\quad a = \bar{y} - b\bar{x} = 85.412 - \cancel{0.6132}\ 0.819 \times \frac{10.047}{13.418} \times 79.183 = 36.85$

- Interpret the slope, the y-intercept, the correlation, and R^2.

Slope: 0.6132 As exam 1 score (x) increases by 1pt, we predict that the final score in class (y) will increase by 0.6132 pts.

y-int: 36.85 Mathematically, the line predicts a final score of 36.85 for someone with exam1 score 0.
Statistically, we would only interpret it if Ex1=0 is close to values of Ex1 scores observed. — Unlikely
Looking at the graph — Not so — We do NOT interpret it.

corr: $r = 0.819$ positive, very strong (pretty) correlation btwn Ex1 score and Final grade.

R^2: $(.819)^2 = .6707 = 67.07\% \rightarrow \%$ of variability in Final score explained by regression on Ex1 score

First note that $\hat{y}_i = \bar{y} + b(x_i - \bar{x})$ and

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

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

Now we can prove the crossproduct term is zero:

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) = \sum_{i=1}^{n}(y_i-\bar{y}-b(x_i-\bar{x}))\cdot b(x_i-\bar{x})$$

$$= b\left[\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x}) - b\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$$

$$= b\cdot 0 = 0$$