Confidence Intervals for Regression Lines

- Recall that a **regression line** is a line that describes the relationship between the explanatory variable x and the response variable y.
- The least squares regression line formula is $\hat{y} = a + bx$ where a is the y-intercept and b is the slope.
- The slope of a regression line can help determine if a relationship exists between two variables. When the slope of a regression line is zero, no relationship exists.

- The assumptions for regression inference are:
  1. If we hold x fixed and take many observations on y, the normal pattern will eventually appear in a stem-plot or histogram. Repeated values of y are independent of each other.
  2. The mean response of y has a straight-line relationship with x: $\mu_y = \alpha + \beta x$ (The slope, β, and y-intercept, α, are unknown parameters.)
  3. The standard deviation, σ, of y is the same for all values of x and is unknown.

- Confidence Interval for β is found with the formula:

  $$b \pm t^*(SE_b)$$

- Where t^* has $n - 2$ degrees of freedom, $s = \sqrt{\frac{1}{n-2}\Sigma(y - \hat{y})^2}$ and $SE_b = \frac{s}{\sqrt{\Sigma(x - \bar{x})^2}}$.

Example:

1. How well do golfers' scores on the first round of a two-round tournament predict their scores for a second round? Twelve golfers recorded scores for each round of a two-round tournament. Give the least squares regression line for the data, report on *r* and *r*-squared and compute the 90% confidence interval for the slope of the regression line.

Golfer	1	2	3	4	5	6	7	8	9	10	11	12
Round 1 x	89	90	87	95	86	81	102	105	83	88	91	79
Round 2 y	94	85	89	89	81	76	107	89	87	91	88	80

$n=12$, $df=10$

```
y=a+bx
a=26.33201581
b=.6877470356
r²=.472063054
r=.6870684493
```

$\hat{y} = 26.33 + .6877x$

$b \pm t^* SE_b$

$s = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}} = 5.974$

$SE_b = \frac{s}{\sqrt{\Sigma(x-\bar{x})^2}} = \frac{5.974}{\sqrt{674.667}} \approx .23$

$t^* = \text{invT}\left(.9 + \frac{(1-.9)}{2}, 10\right) = 1.812$

$.6877 \pm (1.812)(.23)$

$[.27, 1.105]$

Example:

2. Below is the computer output for the appraised value (in thousands of dollars) and number of rooms for 73 houses in South Springs, Texas.

df = 71

Predictor		Coef	StDev	T	
→ Constant	a	74.800	19.04	3.93	s=29.05
Rooms	b	19.718	2.631	7.49	R-sq=43.8%
					R-sq(adj)=43.0%

Using this information, find the equation of the LSRL and calculate the 95% confidence interval of the slope of the regression line for all houses in South Springs.

$\hat{y} = 74.8 + 19.718x$

$b \pm t^* SE_b$

$t^* = \text{invT}\left(.95 + \frac{(1-.95)}{2}, 71\right) = 1.99$

$SE_b = 2.631$

$19.718 \pm (1.99)(2.631)$

$[14.47, 24.96]$