Inference for Regression: Slope

When we fit a regression line to X and Y, the values of X and Y are subject to sampling variability.

As a consequence, the regression equation

$\hat{y} = \hat{b}_0 + \hat{b}_1 \times X$ is an approximation of the true relationship between X and Y

$$\underbrace{Y_i = b_0 + b_1 X_i + \varepsilon_i}$$

This is random noise $\sim N(0, \sigma^2)$

Where $\hat{b}_0$ and $\hat{b}_1$ are approximations of b_0 and b_1

Similar to $\hat{p}$ approximating p

or $\bar{X}$ approximating μ

or S approximating σ

If $\hat{b}_1$ is an approximation b_1, do we know its distribution?

Yes! Based on the assumptions of regression

$$\frac{\hat{b}_1 - b_1}{SE_{\hat{b}_1}} \sim T_{df=n-2}$$

sample size

We can use this to....

Construct CIs

and

Do hypothesis testing for b_1

Similar...

$$\frac{\bar{X} - \mu}{SE_{\bar{X}}} \sim T \quad \text{or} \quad \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{SE_{\bar{X}_1 - \bar{X}_2}} \sim T$$

$\rightarrow$ We can do exactly the same inference on $\hat{b}_1$

Why? Generally....

If we can show evidence that $b_1 > 0$, or $b_1 \neq 0$ or $b_1 < 0$ $\Rightarrow$ This informs us about the relationship between X and Y

same inference on $\hat{b}_1$ as we do for other statistics.

Similarly, we may want to report CIs for the value of b_1

CIs are largely the same...
We construct a CI for b_1 from $\hat{b}_1$, and we are (ex.) 95% confident that b_1 is inside range...

Hypothesis Testing

$H_0: b_1 = 0$

$H_a: b_1 \neq 0$, or... $b_1 > 0$, or... $b_1 < 0$

Test statistic: $\dfrac{\hat{b}_1 - 0}{SE_{\hat{b}_1}} \sim T_{df=n-2}$ ← from H_0

How do we get this? R

Remember:

p-value not significant
Not strong evidence against $b_1 = 0$, it is possible that the linear relationship observed is purely due to random chance.

p-value is significant.
We reject H_0 and conclude that there is a relationship between X and Y. Note: Relationship is not causal.