→ Inference in Linear Regression

Recall that for the Least Square Estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ we have:

$$E(\hat{\beta}_0) = \beta_0 \quad , \quad var(\hat{\beta}_0) = \left(\frac{1}{n} + \frac{\bar{x}^2}{\Sigma(x_i-\bar{x})^2}\right)\sigma^2 \quad \text{and} \quad E(\hat{\beta}_1) = \beta_1 \ , \ var(\hat{\beta}_1) = \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2}$$

where $\Sigma_{i=1}^{n}(x_i-\bar{x})^2$ is commonly denoted as S_{xx} ("sum of squares")

Since the errors are normally distributed (i.e. $\varepsilon_i \sim N(0,\sigma^2)$) and $\hat{\beta}_0, \hat{\beta}_1$ are both linear combinations of normally distributed RV's, then $\hat{\beta}_0, \hat{\beta}_1$ are also normally dist.

i.e $\hat{\beta}_0 \sim N(\beta_0, var(\hat{\beta}_0))$ and $\hat{\beta}_1 \sim N(\beta_1, var(\hat{\beta}_1))$

For instance, $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$ and $se(\hat{\beta}_1) = \sqrt{\sigma^2/S_{xx}}$ (standard deviation)

Then

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sigma^2/S_{xx}}} \sim N(0,1) \quad ⊛ \ \text{(standardization)}$$

Turns out:

(1) $$\overset{\text{MSE}}{\hat{\sigma}^2} = \frac{SSE}{n-2} = \frac{\Sigma(y_i-\hat{y}_i)^2}{n-2} = \Sigma(y_i)^2 - \hat{\beta}_0\Sigma y_i - \hat{\beta}_1\Sigma x_i y_i \sim \frac{\sigma^2 \chi^2_{n-2}}{n-2}$$

(2) $\hat{\sigma}^2$ is independent of $\hat{\beta}_1$.

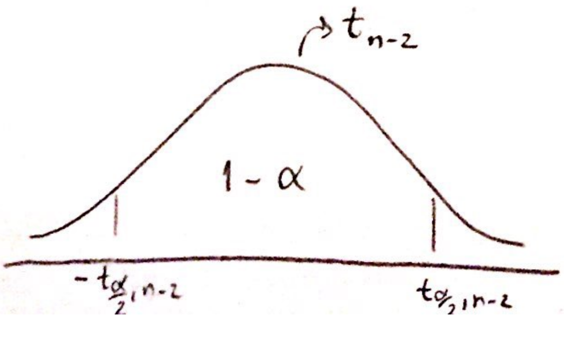
unknown estimate

⊛ $$\frac{\hat{\beta}_1-\beta_1}{\sigma/\sqrt{S_{xx}}} = (\hat{\beta}_1-\beta_1)\cdot\frac{\sqrt{S_{xx}}}{\sigma} \sim N(0,1) \text{ and we want to replace } \sigma \text{ by } \hat{\sigma}$$

$$\Rightarrow (\hat{\beta}_1-\beta_1)\cdot\frac{\sqrt{S_{xx}}}{\sigma}\cdot\frac{\sigma}{\hat{\sigma}} \left(\text{we did } \frac{\frac{\hat{\beta}_1-\beta_1}{\sigma/\sqrt{S_{xx}}}}{\hat{\sigma}/\sigma} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2(n-2)}{n-2}}}\right) \sim t_{(n-2)}$$

$$\Rightarrow \frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}/\sqrt{S_{xx}}} \sim t_{n-2}$$

⇒ Therefore, we will use the statistic $T = \frac{\hat{\beta}_i - \beta_i}{se(\hat{\beta}_i)} \sim t_{n-2}$ for Regression inference.



- 2-sided CI for β_i:

$$1-\alpha = P\left(-t_{\frac{\alpha}{2},n-2} \le \frac{\hat{\beta}_i-\beta_i}{se(\hat{\beta}_i)} \le t_{\frac{\alpha}{2},n-2}\right)$$
$$= P\left(\hat{\beta}_i - t_{\frac{\alpha}{2},n-2}\, se(\hat{\beta}_i) \le \beta_i \le \hat{\beta}_i + t_{\frac{\alpha}{2},n-2}\, se(\hat{\beta}_i)\right)$$

- 1-sided CI for β_i:

$\beta_i \in (-\infty, \hat{\beta}_i + t_{\alpha,n-2}\, se(\hat{\beta}_i))$ (upper)

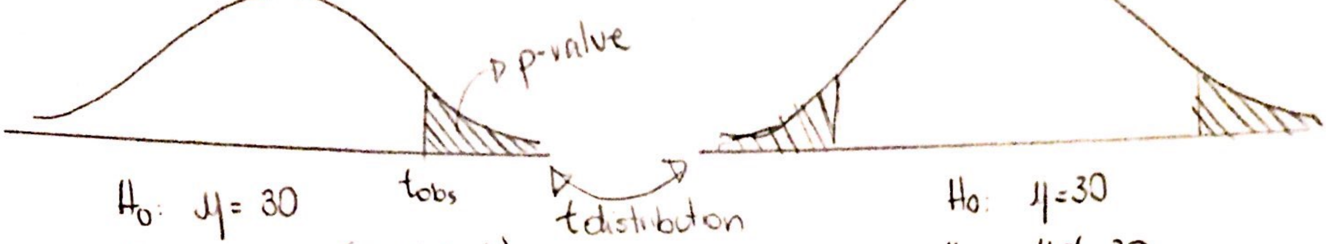
$\beta_i \in (\hat{\beta}_i - t_{\alpha,n-2}\, se(\hat{\beta}_i), \infty)$ (lower)

Understanding P-value

1. P-value is just a "trigger" to decide whether you reject the null hypothesis or not.

Before continuing, understand the difference between a one-tailed and two-tailed hypothesis test.

One-tailed t-test (Usually right tail) | Two-tailed t-test



$H_0: \mu = 30$

$H_1: \mu > 30$ (right tail)

or $H_1: \mu < 30$ (left tail)

$H_0: \mu = 30$

$H_1: \mu \neq 30$

(ie $\mu > 30$ or $\mu < 30$)

For a Confidence level of:

C	α	One-tailed t test $(t_{\alpha,1})$	Two-tailed t test $(t_{\alpha/2,1})$	dfs
0.90	0.10	3.078	$\pm$ 6.314	1
0.95	0.05	6.314	$\pm$ 12.706	1
0.99	0.01	31.821	$\pm$ 63.657	1

notice the ≠?

most t-tables will bring upper tail prob. of the dist. (i.e. t s.t. $P(X > t) = 1-\alpha$)

So, you construct your hypothesis testing first. Then, you calculate, based on your data, the probability of obtaining a result more "extreme" than those observed in your data assuming Ho is true. This probability $(P(|t| > |t_{obs}|))$ is called the p-value. When it's very small, it means that the prob. of getting a result that confirms the null hypothesis is also very small, so you do not have enough evidence in the data to accept the null hypothesis (roughly speaking, you "reject" the null hypothesis based on your data).

In our previous example, p-value was 0.175.

Confidence Level	Accept/Reject Ho?
0.99	Accept
0.95	Accept
0.9	Accept
[illegible]	Reject

Notice that for a C of 0.99, I want strong evidence that Ho is not true, whereas for a C of 0.8, I'm willing to reject Ho if I have "weak" evidence that Ho is not true.

Multiple Regression Example:

After running a multiple regression with 2 covariates (X1 and X2) in R, the following table was obtained ($n=20$).

Predictor	Coef.	SE coef.	T
constant	3.324	3.111	1.07
X1	3.7681	0.6142	$\boxed{}$
X2	5.0796	0.6655	$\boxed{}$

(1) Compute t-value corresponding to X1 and X2.

Answer: $t_{X1} = \frac{3.7681}{0.6142} = 6.13$ $t_{X2} = \frac{5.0796}{0.6655} = 7.63$

(2) Test whether the predictor variables significantly affect Y at level 0.05.

$n = 20$ so $t_{n-\overset{2}{p}-1} = t_{20-2-1} = t_{17}$

$t_{17, 1-\frac{0.05}{2}} = t_{17, 0.975} = 2.109816$ (in R qt(0.975,17))

since both t_{X1} and t_{X2} are $> t_{CRITICAL} = 2.109816$, both t_{X1} and t_{X2} are significant at 0.05 level

(3) Using the model, estimate the value of Y at $X_1 = 5$ and $X_2 = 16$.

$$\hat{Y} = 3.324 + 3.7681(\underset{5}{X1}) + 5.0796(\underset{16}{X2}) = 103.44$$

... Coming next class (10/06) { • CI for estimated average • Prediction interval