Lecture 33

Confidence intervals for regression parameters

A level C confidence interval for β_0 ← intercept is

$$b_0 \pm t^* SE_{b_0}$$

where $P(-t^* < T < t^*) = C$ and T has $t(n-2)$ distribution.

A level C confidence interval for β_1 ↙ slope is given by

$$b_1 \pm t^* SE_{b_1}$$

where $P(-t^* < T < t^*) = C$ and T has $t(n-2)$ distribution.

Example

Suppose I fit a linear regression to some data using excel and part of the output is

$b_0 = -2.2761$, $b_1 = 2.0680$, $SE_{b_0} = 0.1647$, $SE_{b_1} = 0.1054$
with $n = 52$

What are the 95% CI for β_0, β_1?

for β_0 the 95% CI is

$-2.2761 \pm (2.009)(0.1647)$

-2.2761 ± 0.3309

$(-2.6070, -1.9452)$

A 95% CI for β_1 is

$2.0680 \pm (2.009)(.1054)$

2.0680 ± 0.2117

$(1.8562, 2.2798)$

Note I do not expect you to know how SE_{b_0}, SE_{b_1} are calculated. Only how to use them for CI and testing questions.

Hypothesis test for the slope

to test

$H_0: \beta_1 = 0$

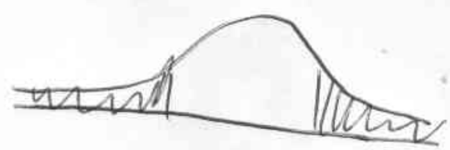
vs

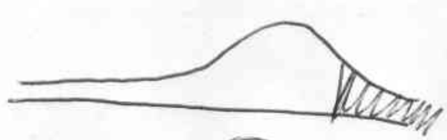
$H_A: \beta_1 \neq 0$

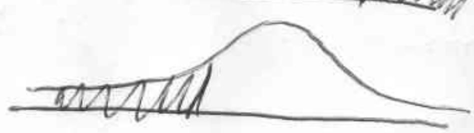
use $t = \dfrac{b_1}{SE_{b_1}}$ ← test statistic

which has t distribution with $n-2$ df.

As discussed previously the alternative tells you which tail to use for your P-value

$H_A: \beta_1 \neq 0$ $\quad 2P(T > |t|)$ 

$H_A: \beta_1 > 0$ $\quad P(T > t)$ 

$H_A: \beta_1 < 0$ $\quad P(T < t)$ 

Note This test can be used to test whether the "x" variable is useful or not at predicting the "y" variable. If $\beta_1 = 0$ is true then

(ie if your prefer H_0)
the "x" variable is of no use. Otherwise it is useful (ie if you prefer H_A).

Example Computer output for the regression line fitted to a particular dataset gave the following results,

$b_0 = 2.436 \qquad SE_{b_0} = 2.973$

$b_1 = -10.431 \quad SE_{b_1} = 6.473$

$n = 18$

test whether or not the "x" variable is useful for determining the "y" value.

$H_0 : \beta_1 = 0$

$H_A : \beta_1 \neq 0$

$t = \dfrac{-10.431}{6.473} = -1.611 \qquad df = 18-2 = 16$

so $\text{Pvalue} = 2P(T > |-1.611|)$

from table

$$1.337 < 1.611 < 1.746$$

$$.1 > P(T > 1.611) > .05$$

$$\Rightarrow .1 < 2P(T > 1.611) < .2$$

Since Pvalue is large cannot reject H_0.
Therefore conclude "x" variable is of no use in predicting "y" variable.

Residuals

Recall that the residual is the difference between the observed value of the response variable and that predicted by the regression line ie

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

plots of residuals against the explanatory variables can be useful for assessing whether the

Linear regression line fit or not.

Plots of residuals can also be highly informative visually about the regression line and how well it fit the observed data.