3Y03 - 3J04 PROBABILITY & STATISTICS FOR (CIVIL) ENGINEERING

C01 - Lecture 32

Last time: HYPOTHESIS TESTING IN LINEAR REGRESSION

Testing $H_0 : \beta_1 = (\beta_1)_0$ ($=0$ usually)
$H_1 : \beta_1 \neq (\beta_1)_0$.

Under H_0,

① $T_0 = \dfrac{\hat{\beta}_1 - (\beta_1)_0}{\sqrt{\hat{\sigma}^2 / S_{xx}}} \sim t_{n-2}$-distribution so carry out a t-test.

② [ANOVA] $\quad F_0 = \dfrac{SS_R/1}{SS_E/(n-2)} = \dfrac{\hat{\beta}_1 S_{xy}}{\hat{\sigma}^2} \sim F_{1,n-2}$-distribution...

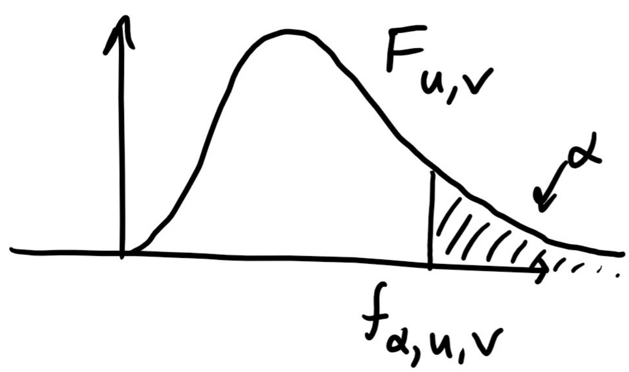
Analysis of Variance Approach

messy!

F-test: Under H_0 $\quad F_0 \sim F_{1,n-2}$ - distribution

So reject H_0 if $f_0 > f_{\alpha,1,n-2}$

plug ~~into~~ info. from data into F_0



Use the f-table corresponding to significance level α.

<u>Example</u> (ctd) $\quad H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$

Use ANOVA to test this at level $\alpha = 0.05$.

$(\hat{\beta}_1 = -0.49,\ S_{xy} = -16,\ \hat{\sigma}^2 = 0.72)$

Solution First compute $f_0 = \dfrac{SS_R/1}{SS_E/(n-2)} = \dfrac{\hat{\beta}_1 S_{xy}}{\hat{\sigma}^2}$

$$= \frac{(-0.49)(-16)}{0.72} = 10.8.$$

Look up $f_{\alpha,1,n-2} = f_{0.05,1,3} = 10.13$ ($n = 5$)

$10.8 > 10.13$ so reject H_0 in favour of H_1.

Note The t-test & f-test (ANOVA) are equivalent for 2-sided tests. (With 2 r.v.s X & Y — ANOVA comes into its own with 3+ r.v.s)

$$T_0 = \frac{\hat{\beta}_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} \Rightarrow T_0^2 = \frac{\hat{\beta}_1^2}{\hat{\sigma}^2/S_{xx}} = \frac{\hat{\beta}_1\left(\frac{S_{xy}}{\cancel{S_{xx}}}\right)}{\left(SS_E/(n-2)\right)/\cancel{S_{xx}}} = \frac{SS_R/1}{SS_E/n-2} = F_0.$$

If a 1-sided test, need to use t-test.

11.5 Confidence Intervals

Yesterday:
$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} \text{ and } \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim t_{n-2} \text{ distr.}$$

So a $100(1-\alpha)\%$ C.I. for β_1 is given by:

$$\hat{\beta}_1 \pm t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 / S_{xx}}.$$

Could clearly do something analagous for β_0 but not in this course.

<u>Example</u> (ctd) Give a 95% C.I. for β_1.

<u>Solution</u>

$$(-0.49) \pm \underbrace{t_{0.025,3}}_{3.182} \sqrt{\frac{0.72}{32.8}}$$

$$= -0.49 \pm 0.47 \quad \text{i.e. } (-0.96, -0.02)$$ //

Remember, if we used C.I. to do hypothesis test, we'd again reject H_0 since H_0 value ($\beta_1 = 0$) does not lie in this C.I.

Interested in using regression line to understand Y — both its underlying distribution and as a predictor.

We call a y-value corresponding to a fixed x-value a "<u>response</u>".

We can estimate expected or "mean response" for a given x-value ($X = x_0$) with $\hat{\beta}_0 + \hat{\beta}_1 x_0$

This is an estimator for mean of Y at $X = x_0$:

$$E(Y \mid X = x_0) \left(= E(Y \mid x_0) \right) = \mu_{Y|x_0}$$

We rewrite $\hat{\beta}_0 + \hat{\beta}_1 x_0$ as $\hat{\mu}_{Y|x_0}$.

Our setup lets us find a C.I. for true mean of Y at $X = x_0$ i.e. $\mu_{Y|x_0}$.

We need mean & variance of $\hat{\mu}_{Y|x_0}$.

$$E(\hat{\mu}_{Y|x_0}) = E(\hat{\beta}_0 + \hat{\beta}_1 x_0) = E(\hat{\beta}_0) + x_0 E(\hat{\beta}_1)$$
$$= \beta_0 + \beta_1 x_0 = \mu_{Y|x_0}$$

(So $\hat{\mu}_{Y|x_0}$ unbiased)

Can also be shown that $V(\hat{\mu}_{Y|x_0}) = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)$.

Since $\hat{\beta}_0, \hat{\beta}_1$ are normally distributed & $\hat{\mu}_{Y|x_0}$ is a linear combination of $\hat{\beta}_0, \hat{\beta}_1$, $\hat{\mu}_{Y|x_0}$ is also normal

So!

$$\frac{\hat{\mu}_{Y|x_0} - \mu_{Y|x_0}}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}} \sim t\text{-distribution with } n-2 \text{ degrees of freedom}$$

Have to estimate σ^2 with $\hat{\sigma}^2$, hence t-distr.

So a $100(1-\alpha)\%$ C.I. for true value of mean of Y at $X = x_0$ i.e. for $\mu_{Y|x_0}$ is given by

$$\hat{\mu}_{Y|x_0} \pm t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(x_0-\bar{x})^2}{S_{xx}}\right)}$$

$$\uparrow$$

$$= \hat{\beta}_0 + \hat{\beta}_1 x_0$$

<u>Example</u> For our running example, we had $\bar{x} = 4.2$, $S_{xx} = 32.8$, ... find a 95% C.I. for the mean value of Y at $x = \underset{= x_0}{3}$.

<u>Solution</u>

$$(4.05 + (-0.49)(3)) \pm 3.182\sqrt{0.72\left(\frac{1}{5} + \frac{(3-4.2)^2}{32.8}\right)}$$

$$= 2.58 \pm 3.182(0.42)$$

$$= 2.58 \pm 1.33$$

i.e. $(1.25, 3.91)$

$\hookrightarrow$ So with 95% confidence, true value of mean of Y at $x=3$ is in this interval.

11.6 Prediction of New Observations

We have the point estimator

$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ for the future/new value of Y when $X = x_0$.

In predicting a <u>single</u> observation we have to take into account possible error $e_{\hat{p}} = Y_0 - \hat{Y}_0$

p : prediction (stands for prediction; not probability or proportion !!!)

Y_0: actual observation at $X = x_0$

$\hat{Y}_0$: prediction at $X = x_0$

$$E(e_{\hat{p}}) = 0 \quad \text{but} \quad V(e_{\hat{p}}) = V(Y_0 - \hat{Y}_0)$$

$$= V(Y_0) + V(\hat{Y}_0)$$

$$= \sigma^2 + \sigma^2\left(\frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)$$

$$= \sigma^2\left(1 + \frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)$$

So

$$\frac{\overbrace{(Y_0 - \hat{Y}_0)}^{e_{\hat{p}}}}{\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(\bar{x}-x_0)^2}{S_{xx}}\right)}} \sim t_{n-2}\text{-distribution.}$$

Estimate σ^2 with $\hat{\sigma}^2$ hence t-distribution.

This lets us define a so-called $100(1-\alpha)\%$ <u>prediction interval</u> for Y_0 (future observation at $X = x_0$):

$$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-2} \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right)}$$

$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$

Extra margin of error when trying to predict a single observation.

(Compare with the C.I. for underlying parameter the true mean value of the mean of Y at $X = x_0$ i.e. for $\mu_{Y|x_0}$.)