STAT 1302 ; Lecture 19 ; March 21, '24

In § 13.3 (Inferences about B) :

Last Class : We jotted down the test stat. and rejection regions for testing $H_0 : B = 0$ against $H_1 : B > 0$, $H_1 : B < 0$, or $H_1 : B \neq 0$.

Ex. Y = food expenditure (monthly), X = monthly income (in hundreds of \$).

Data : $n = 7$, $S_e = 0.9922$; $SS_{xx} = 801.4286$

(See Previous lectures).

Q'n : Test at the 1% Significance level whether the slope of the population regression line (ie. $Y = A + BX + \varepsilon$) is positive.

(Or, test at 1% Significance level, whether there is a positive relationship between monthly food expenditure and monthly income.)

$H_0: B = 0$ vs. $H_1: B > 0$, $\alpha = 0.01$.

See Lecture 17

$$t = \frac{b - 0}{S_e / \sqrt{SS_{xx}}} = \frac{0.2642 - 0}{0.9922/\sqrt{801.4286}} = 7.549$$

$df = n - 2 = 7 - 2 = 5$ (Memory aid for $n-2$: look at expression for S_e).

Reject H_0 if $t_{obs.} > t_{5;0.01} = 3.365$

Since $t_{obs.} = 7.549 > t_{5;0.01} = 3.365$, we reject H_0.

Conclusion: We are 99% Confident that the monthly food expenditure increases as monthly income increases.

Result: A $100(1-\alpha)\%$ Confidence interval for B is

$$b \pm t_{n-2;\alpha/2} \cdot \frac{S_e}{\sqrt{SS_{xx}}}$$

Point estimate of B

Standard deviation of b

See $$t = \frac{b - B_0}{s_e / \sqrt{SS_{xx}}}$$

§ 13.4 Linear Correlation

Suppose Y and X are two quantitative random variables (that are jointly normally distributed).

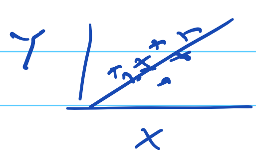
We can measure the linear association between X and Y through the Correlation Coefficient, ρ.

ρ → "rho"

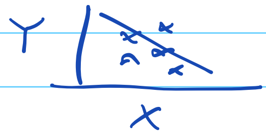
Note: It does not matter if we speak of the Correlation between X and Y or the correlation between Y and X; they are identical.

Remark: $-1 \le \rho \le 1$

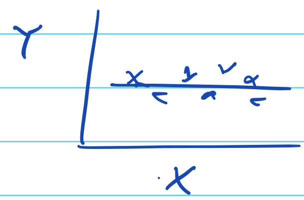
where $\rho > 0 \Rightarrow$ a positive linear association between X and Y,

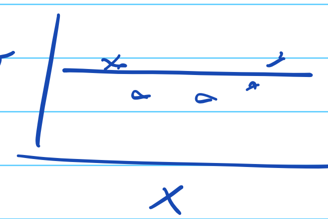


$\rho < 0 \Rightarrow$ a negative linear association between X and Y,



$\rho = 0 \Rightarrow$ no linear association between X and Y.





Interpretation:

i) $\rho > 0 \Rightarrow$ As x increases, Y increases.

ii) $\rho < 0 \Rightarrow$ As x increases, Y decreases.

iii) $\rho = 0 \Rightarrow$ As x increases, Y does not change.

The correlation parameter, ρ, is estimated using the sample correlation coefficient, r,

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} \cdot SS_{yy}}} \leftarrow \text{Given}$$

Note: Coefficient of Determination

$$= \left(\text{Sample Correlation coefficient} \right)^2$$

Ex. Y = monthly food expenditure,

X = monthly income (See Lecture 17).

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 211.7143$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 801.4286$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 60.8571.$$

Qn: What is the Sample correlation coefficient between monthly income and monthly food expenditure?

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}} = \frac{211.7143}{\sqrt{801.4286 \times 60.8571}} = 0.96$$

There is a strong positive linear association between monthly income and monthly food expenditure.

Observe : $[r]^2 = [0.96]^2 = \underbrace{0.92}_{r^2}$

where r^2 is the coefficient of determination, meaning 92% of the variation in monthly food expenditure is explained by monthly income.

Ex. See previous example. Compute a 95% confidence interval for the slope parameter of the model regressing Y (monthly food expenditure) against X (monthly income).

$$b \pm t_{n-2;\alpha/2} \cdot \frac{S_e}{\sqrt{SS_{xx}}}$$

$b = 0.2642$ (Lecture 17).

$1 - \alpha = 0.95$; $\alpha = 0.05$; $\alpha/2 = 0.025$

Table V, $t_{5;0.025} = 2.571$.

95% CI for B:

$$0.2642 \pm 2.571 \times \frac{0.9922}{\sqrt{801.4286}}$$

$$= 0.2642 \pm 0.09011$$

$$= (0.17409, 0.35431)$$

$$\approx (0.174, 0.354).$$

Interpretation: We are 95% confident that the true slope parameter in the regression line regressing monthly food expenditure against monthly income varies between 0.174 and 0.354. (i.e \$17.4 and \$35.4).

back to r:

Hypothesis Tests about the Linear Correlation Coefficient (ρ):

Test	Rejection Region
$H_0: \rho = 0, \quad H_1: \rho > 0$	$t > t_{n-2;\alpha}$
$H_0: \rho = 0, \quad H_1: \rho < 0$	$t < -t_{n-2;\alpha}$
$H_0: \rho = 0, \quad H_1: \rho \neq 0$	$\lvert t \rvert > t_{n-2;\alpha/2}$

Test Statistic :

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$ ← Given

Under H_0, $t = r\sqrt{\frac{n-2}{1-r^2}} \sim t_{n-2}$

Ex. Y = monthly food Expenditure,

X = monthly income.

$r = 0.96$.

Test whether the correlation between X and Y is positive. Let $\alpha = 0.01$. (Assume X and Y are jointly normally distributed.).

Parameter: ρ

$H_0: \rho = 0$ vs. $H_1: \rho > 0$.

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0.96\sqrt{\frac{7-2}{1-0.96^2}} = 7.667$$

Reject H_0 if $t_{obs.} > t_{5;0.01} = 3.365$.

Decision Rule:

Since $7.667 > 3.365$, we reject H_0.

Conclusion: We are 99% confident that the correlation between monthly income and monthly food expenditure is positive.