ASIL SHAAR

FINN3302

النمذجة المالية

# CHAPTER 3: A brief overview of the classical linear regression model

## $\boxed{\text{Chapter 3}}$ A brief overview of the classical linear regression model

* Regression analysis : قياس وتقدير العلاقة بين dependent variable واحد او مجموعة من ال independent variables

⇒ Slide 3: Some notations for x and y

→ Regression vs Correlation

ما بتدرس اتجاه العلاقة بين x , y ← بس انا بقيس قوة العلاقة بين متغيرين مثلاً بين A و B طلعت قوة العلاقة ولو عكستهم مثلاً (A,B) ح تطلع نفس قوة العلاقة

Measure the strength of a linear association between two variables

⇒ y is random variable (has a pdf) , while X is not random variable (given

→ Slide 6 example

X : excess return on market index

y : excess return

* there is a positive relationship between X and y

هاد كل شبيه ب الكودبودين ال market model ، يعني بمعنى اخر احنا هون بنحسب beta الي بتقيس ال systematic risk

Regression Analysis

Simple Regression analysis ↓

describing and evaluating a relationship between dependent variable y & a single independent variable x

wage → y, edu → x

multiple regression analysis ↓

describing and evaluating relationship between a single Y and a number of X's

wages → y, edu → $x_1$, gender → $x_2$, experince → $x_3$

y : dependent variable , regressand , effect variabl , Response Variable , predicted variable

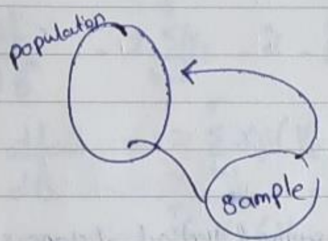
X : independent variable , regressors , Causal variable , explanatory variable , Control Variable , predict variable

Correlation
A B
-1
B A
-1

Correlation
$-1 \leq \text{Coefficient} \leq 1$

Regression
Y   X

y : Random "Stochastic" $\Rightarrow$ probability dis
x : non-Random "non-Stochastic" $\Rightarrow$ Fixed repeated sample




$$Y_t = \alpha + \beta X_t + U_t$$

$$wages_t = \alpha + \beta\, edu + U_t \rightarrow PRF$$

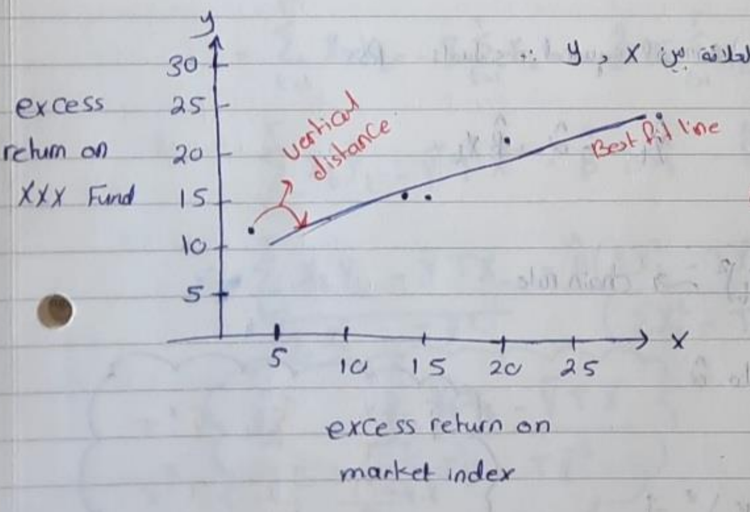
population Regression function

unobserved
unkowns

يعني بنهتم بدراسة الـ wages لكل الناس بفلسطين مثلاً

* نرجع لمثالنا بسلايد 6 :-

بنا خلال نرسم الـ scatter plot للعلاقة بين X , y :




(SRF)
simple regression function

* الخط يلي رسمناه هو خط بقلل المسافة بين كل نقطة والخط ، بحيث تكون كل النقاط اقرب ما يمكن للخط

$\Rightarrow$

* فلما استخدمنا ال SRF ( يعني اخذنا sample صغيرة من pop ) :-

SRF $\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$

fitted value (estimated value)

وبالتالي * error term $(U_t)$ ح يروح

Vertical distance $= Y_t - \hat{Y}_t$

$$\underbrace{\hat{U}_t}_{\text{Residual}} = \underbrace{Y_t}_{\text{actual value}} - \underbrace{\hat{Y}_t}_{\text{fitted value}}$$

* when estimating $\hat{\alpha}$ and $\hat{\beta}$ we will be minimicing sum of squared vertical distances (sum of squred residuals) → OLS method هاي الطريقة اللي
ordinary least squares method

→ $\min \sum \hat{U}_t^2$

when estimating $\hat{\alpha}$ and $\hat{\beta}$ sum of squared residuals = Rss

$$\underbrace{L}_{\text{less function}} = \sum_{t=1}^{T} (\underbrace{Y_t - \hat{Y}_t}_{\hat{u}_t})^2 \qquad \hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$$

$$L = \sum_{t=1}^{T} (Y_t - \hat{\alpha} - \hat{\beta} X_t)^2 \rightsquigarrow \text{chain rule.}$$

(1) differentiate with respect to $\hat{\alpha}$
(2) derivative $= 0$

$\bar{Y} = \frac{\sum Y_t}{T}$ $\quad \sum_{t=1}^{T} Y_t = \bar{Y} * T$

$\bar{X} = \frac{\sum X_t}{T}$ $\quad \sum_{t=1}^{T} X_t = T * \bar{X}$

$$\frac{dL}{d\hat{\alpha}} = 2 \sum_{t=1}^{T} (Y_t - \hat{\alpha} - \hat{\beta} X_t) * -1$$

$$\frac{dL}{d\hat{\alpha}} = \frac{-2}{-2} \sum_{t=1}^{T} (Y_t - \hat{\alpha} - \hat{\beta} X_t) = \frac{0}{-2}$$

$$\sum_{t=1}^{T} (Y_t - \hat{\alpha} - \hat{\beta} X_t) = 0$$

$$\sum_{t=1}^{T} Y_t - \sum_{t=1}^{T} \hat{\alpha} - \hat{\beta} \sum_{t=1}^{T} X_t = 0$$

→

$$\frac{\bar{Y}T}{T} - \frac{T\hat{\alpha}}{T} - \frac{\hat{\beta}T\bar{X}}{T} = \frac{0}{T}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$L = \sum_{t=1}^{T}(Y_t - \hat{\alpha} - \hat{\beta}X_t)^2$$

$$\frac{dL}{d\hat{\beta}} = 2\sum_{t=1}^{T}(Y_t - \hat{\alpha} - \hat{\beta}X_t) \cdot -X_t$$

$$\frac{dL}{d\hat{\beta}} = \frac{-2}{-2}\sum_{t=1}^{T}X_t(Y_t - \hat{\alpha} - \hat{\beta}X_t) = \frac{0}{-2}$$

$$= \sum_{t=1}^{T}X_t(Y_t - \hat{\alpha} - \hat{\beta}X_t) = 0$$

$$= \sum X_t(Y_t - \bar{Y} + \hat{\beta}\bar{X} - \hat{\beta}X_t) = 0$$

$$= \sum_{t=1}^{T}X_tY_t - \bar{Y}\sum_{t=1}^{T}X_t + \hat{\beta}\bar{X}\sum_{t=1}^{T}X_t - \hat{\beta}\sum X_t^2 = 0$$

$$\sum_{t=1}^{T}X_tY_t - \bar{Y}T\bar{X} = \hat{\beta}\sum X_t^2 - \hat{\beta}T\bar{X}^2$$

$$\frac{\sum_{t=1}^{T}X_tY_t - \bar{Y}T\bar{X}}{\sum X_t^2 - T\bar{X}^2} = \frac{\hat{\beta}(\sum X_t^2 - T\bar{X}^2)}{(\sum X_t^2 - T\bar{X}^2)}$$

$$\hat{\beta} = \frac{\sum_{t=1}^{T}X_tY_t - \bar{Y}T\bar{X}}{\sum_{t=1}^{T}X_t^2 - T\bar{X}^2}$$

$$= \frac{Cov(x,y)}{Variance(x)}$$

$$= \frac{\sum(X_t - \bar{X})(Y_t - \bar{Y})}{\sum(X_t - \bar{X})^2}$$

#

ols estimations. جدول المعارلش اسمه

Q slide 19

If an analyst tells you that she expects the market to yield a return 20% higher than the Risk free rate next year, what would you expect the return on fund XXX to be?

نعوض 20٪ في المعادلة

$$\hat{Y} = -1.74 + 1.64 * 20 = 31.06$$

* Forms of a regression function:

In order to use the ols method, a model that is linear is required. The relationship between x and y must be graphically using a straight line. Moreover, the model must be linear in parameters $(\alpha, \beta)$ and not necessarily in variables.

* Linear in prameters: the parameters are not cubed, squared, multiple together..

* Models that are not linear in variable can be made to take a linear form by applying a suitable transformation or manipulation

1. linear (Level - level model) — بنسميه level-level لأنه في X,Y

$$Y = \underline{a} + bX$$

* intercept interpretation:

when X equals zero the average predicted y would equal a (intercept) units

$$b = \text{slope} = \frac{\Delta y}{\Delta x} \qquad \Delta y = b \Delta x$$

* Slope interpretation

If X increase by one unit the averages predicted y would increase / decrease by b unit.

Example 1:

$$\hat{Y}_t = 23651 + 30.533 X_t$$

intercept interpretation:

when X equals zero units then averages predicted y would equal 23,651 units

Slope interpretation:

If X increase by one unit the averages predicted Y would increase by 30.533 units

→

$\boxed{2}$ logarithmic (level_log model)

$Y = a + b\ln(x)$

* intercept interpretation:  $\boxed{\ln 1 = 0}$

when x equal 1 unit then average predicted y would equal a unit.

* Slope interpretation:

$\frac{dy}{dx} = \frac{b}{x}$  $dy \frac{\cancel{x}}{\cancel{x}} = \frac{b\,dx}{x}$  $100 \times dy = b\frac{dx}{x} \times 100$

$\frac{\cancel{100}}{100}{}^{*} \Delta y = \frac{b}{100} \% \Delta X$  $\Delta y = \frac{b}{100} \% \Delta X$

If x increases by one percent (1%) then average predicted would increase/decrease by $\frac{b}{100}$ units.

Example 2:

power Cost = -63,993 + 16,654 ln(units)

* intercept interpretation

when unit of power equal 1 unit then average predicted power cost would equal \$ -63,993

* slope interpretation

If units power increase by 1% then average predicted power cost would increase by $\left(\frac{16,654}{100}\right)$ 166.54\$

$\boxed{3}$ Exponential (log-level)

$y = ae^{bx}$

$\boxed{\ln e = 1}$

بنضرب ln للطرفين

$\ln y = \ln a + bx\ln e$

$\ln y = \ln a + bx$

$d\ln y \times \frac{\cancel{y}}{\cancel{y}} = \frac{dy}{y}$

* Slope interpretation

$\frac{d\ln y}{dx} = 0 + b$  $\frac{d\ln y}{y} = \frac{1}{y}$  $d\ln y = b\,dx$  $\boxed{d\ln y = \frac{dy}{y}}$

$\frac{d\ln y}{dx} = b$  $100 \times \frac{dy}{y} = b\,dx \times 100$

$\boxed{\% \Delta y = (100 \times b)\Delta X}$

$\Rightarrow$

If X increase by one unit then average predicted would increase/decrease by $(100 \times b)$ %

Example 3:

$\ln(wage) = 0.584 + 0.083$ education

* intercept interpretation

when years of education equals 0 year then average predicted hourly wage would equal $e^{0.584} = \$1.79$

* Slope interperlation

If years of education increases by one year then average predicted hourly wage would increase by 8.3 %

[4] power (log-log model)

$$y = ax^{b}$$

$$\ln y = \ln a + b \ln x$$

* Slope interpretation:

$$\frac{d\ln y}{dx} = 0 + \frac{b}{x} \qquad \frac{d\ln y}{dx} = \frac{b}{x} \qquad \frac{d\ln y}{dy} = \frac{1}{y}$$

$$d\ln y * \frac{\cancel{y}}{\cancel{y}} = \frac{dy}{y} \qquad \boxed{d\ln y = \frac{dy}{y}}$$

$$d\ln y * \frac{\cancel{x}}{\cancel{x}} = \frac{b\,dx}{x} \qquad d\ln y = \frac{b\,dx}{x}$$

$$100 \times \frac{dy}{y} = b\frac{dx}{x} \times 100$$

$$\%\Delta y = b\,\%\Delta X$$

If ~~X~~ increases by 1% then average predicted y would increase/decrease by b%.

Example 4 :

$\ln(\text{Salary}) = 4.822 + 0.257 \ln(\text{Sales})$

* intercept interpretation

When Sales equal \$1 then average predicted salary would equal $e^{4.822} = \$124.21$

* Slope interpretation

If sales increases by 1% then average predicted salary would increase by 0.257%

PRF → $y_t = \alpha + \beta x_t + u_t$

SRF $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$ → Best fit line equation.

Classical linear regression model assumptions :

① The error terms have a zero mean

$E(u_t) = 0$

② The error terms have a constant variance $= \sigma^2$

Homoscedasticity assumption $Var(u_t) = \sigma^2 \neq f(x)$

③ NO serial autocorrelation / No autocorrelation ⇒ The error terms are statistically independent of one another $Cov(u_i, u_j) = 0 \quad i \neq j$

④ The error terms and the independent variables are independent of one another

$Cov(u_t, x_t) = 0$

⑤ $u_t$ is normally distributed.

$A_1 \rightarrow A_4$ must hold for OLS estimator to be blue.

| | | |
|---|---|---|
| B | Best | minimum variance أقل |
| L | linear | $\hat{B}$ is a linear estimator |
| U | unbiased | |
| E | Estimator. | true value of B لـ estimate $\hat{B}$ تعني أن |

* Estimators : are the formulas used to calculate the coefficient.
* Estimates : are the actual numerical values for the coefficient.

$A_1 \rightarrow A_5$ must hold to be able to make inference about population parameters.

SE → tells us how likely is our estimate varies from one sample to another using the one sample of information we got.

SE is a function of
1. Total number of observations T.
2. S → estimate of the standard deviation of the error terms
3. actual observations on the explanatory variables $X_t$

كل ما كان اقل كل ما كان افضل

PRF

$$y_t = \alpha + \beta X_t + \underbrace{u_t}_{\text{random variable}}$$

$$E(u_t) = 0$$

$$Var(u_t) = E(u_t^2) - \underbrace{E(u_t)}_{0}{}^2$$

$$Var(u_t) = E(u_t)^2 = \frac{\sum u_t^2}{T}$$

$\hat{u}_t$ ال counter

$$S^2 = \frac{\sum \hat{u}_t^2}{T} = \frac{RSS}{T}$$

$$S^2 = \frac{RSS}{T-2}$$

degrees of freedom

$$S = \sqrt{\frac{RSS}{T-2}}$$

بنرمز لها كذا SER

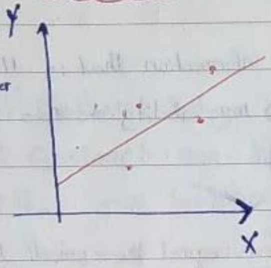
↓ standard error of the regression.

1. $SE(\hat{\alpha})$ and $SE(\hat{\beta})$ depend on S

$$S = \sqrt{\frac{RSS}{T-2}} \Rightarrow$$

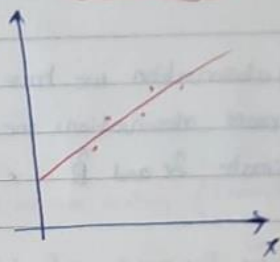
Rss large

we can draw another line
النقاط ابعد الى الخط



Rss small

النقاط اقرب الى الخط!

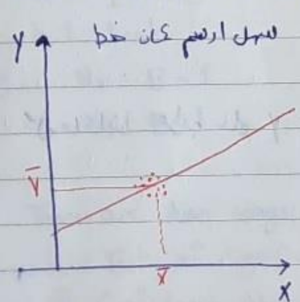


If Rss ↑ , SE ↑
→ lack of Precision of the Coefficent estimates

* we want SE to be small

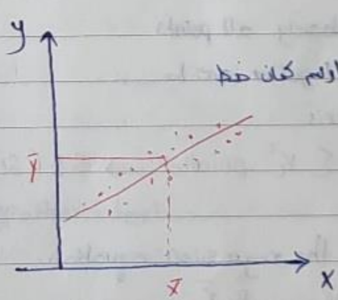
[2] $SE(\hat{\alpha})$ and $SE(\hat{\beta})$ depend on the variability of the explanatory variables about their mean values.

سهل ارسم كذا خط



narrowly dispersed about their mean value.

صعب ارسم كذا خط



widely dispersed about their mean value.

* The bigger $\Sigma(x_t - \bar{x})^2$ the smaller the SE.

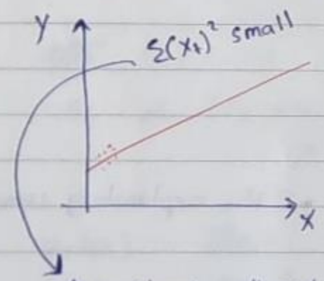
$\Rightarrow$

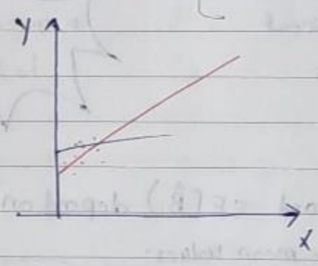
3) $SE(\hat{\alpha})$ and $SE(\hat{B})$ depend on the number of observations T

every observation we have represents a piece of information that is useful. The more observations we have more information → more ability to reliabily estimate $\hat{\alpha}$ and $\hat{B}$.

* The larger the number of observations all things being equal the small the SE.

4) $SE(\hat{\alpha})$ depends on $\sum X_t^2$




how closely all points taken together are to they axis

كل ما كانوا قراب على y بكون افضل

* $\sum X_t^2$ smaller → $SE(\hat{\alpha})\downarrow$

Example:

Estimate the regression equation.

$$\hat{\alpha} = \bar{y} - \hat{B}\bar{x}$$

$$\hat{B} = \frac{\sum x_t y_t - T\bar{x}\bar{y}}{\sum x_t^2 - T\bar{x}^2} = \frac{830{,}102 - (22 * 416.5 * 86.65)}{3{,}919{,}654 - 22*(416.5)^2} = 0.35$$

$$\hat{\alpha} = 86.65 - 0.35 * 416.5 = -59.12$$

$$\hat{Y}_t = -59.12 + 0.35\,X_t$$

$$\quad (3.35) \qquad (0.0079)$$

SE →

$$S = \sqrt{\frac{RSS}{T-2}} = \sqrt{\frac{130.6}{22-2}} = 2.55$$

Hypothesis testing

(1) Construct our hypotheses

null hypothesis ← $H_0$ : → hypothesis of interest → the one we are testing (from finance theory / Economic theory)

alternative hypothesis ← $H_1$ : → remaining outcomes

Type of test :

(a) one sided test

(b) two sided test

(a) one sided test

(1) upper tail test

e.g $H_0 : B = 1$

$H_1 : B > 1$

(2) lower tail test

e.g $H_0 : B = 1$

$H_1 : B < 1$

(b) two sided test

e.g $H_0 : B = 1$

$H_1 : B \neq 1$

There are two ways to conduct a hypothesis test

(a) Test of significance approach

(b) Confidence interval approach

PRF $y_t = \alpha + B x_t + u_t$

$y_t$ : normally distributed

$\hat{\alpha}\ \hat{B}$ : normally distributed

$u_t$ : normally distributed

→

$\hat{\alpha} \sim N(\alpha, var(\alpha))$ $\qquad E(\hat{\alpha}) = \alpha$

$\hat{\beta} \sim N(\beta, var(\beta))$ $\qquad E(\hat{\beta}) = \beta$

$$\frac{\hat{\alpha} - \alpha}{\sqrt{Var(\alpha)}} \to (0,1)$$

follow a standard normal distribution

$$\frac{\hat{\beta} - \beta}{\sqrt{Var(\beta)}} \to (0,1)$$

follows a standard normal distribution

test statistic $\dfrac{\hat{\alpha} - \alpha}{SE(\hat{\alpha})} \sim t$ distribution with T-2 degrees of freedom

$$\frac{\hat{\beta} - \beta}{SE(\hat{\beta})} \to t_{t-2}$$

Test of significance approach:

① estimate your regression $\hat{\alpha}, \hat{\beta}, SE(\hat{\alpha}), SE(\hat{\beta})$

② Calculate t-stat:

$\dfrac{\hat{\alpha} - \alpha^*}{SE(\alpha)}$ → value of $\alpha$ in $H_0$

$\dfrac{\hat{\beta} - \beta^*}{SE(\hat{\beta})}$ → value of $\beta$ in $H_0$

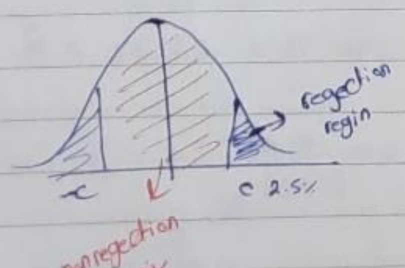
③ choose the significance level → probability of rejection $H_0$ null when it is True.

$\alpha = 1\%, 5\%, 10\%$

④ get the critical values

⑤ perform the test

ⓐ Two sided test




$\alpha = 5\%$

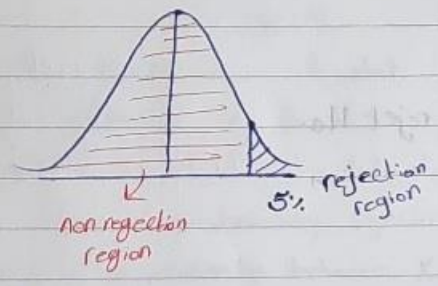
$\dfrac{5\%}{2}$

If $|t\text{-stat}| >$ critical value then reject $H_0$

If $|t\text{-stat}| <$ critical then fail to reject $H_0$

b) one sided test (upper tail test)

$\alpha = 5\%$

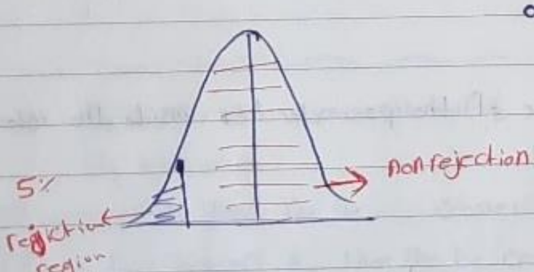



If t-state > critical value then Reject $H_0$

If t-state < critical value the fail to Reject $H_0$

c) one sided test (lower tail test)

$\alpha = 5\%$




If t-stat < critical value then Reject $H_0$

If t-stat > critical value then fail to reject $H_0$.

Example:

$$\hat{Y}_t = 20.3 + 0.5091 X_t \qquad T = 22 \qquad \text{two sided}$$
$$(14.38) \qquad (0.2561)$$

$H_0: B = 1$ $\qquad \alpha = 5\%$

$H_1: B \neq 1$

$$t\text{-stat} = \frac{\hat{B} - B^*}{SE(\hat{B})} = \frac{0.5091 - 1}{0.2561} = -1.917$$

$\Rightarrow$

$\frac{\alpha}{2} = \frac{5\%}{2} = 2.5\%$

$df = T - 2 = 22 - 2 = 20$

$t\text{-critical} = 2.086$

$|-1.917| = 1.917 < 2.086$ then fail to rejet $H_0$.

Confidence interval approach:

① Estimate $\hat{\alpha}$, $\hat{B}$, $SE(\hat{\alpha})$, $SE(\hat{B})$

② choose a significance level $\alpha = 5\%$ equivalent to choosing $(1-\alpha)100\%$ confidence interval

Example:

$\alpha = 5\%$

95% confidence interval

→ we are 95% confident that the true value of the parameter lies inside the interval.

③ get the critical values from the t-table.

④ Construct the confidence interval:

$$\left[\hat{B} - t\text{-critical} * SE(\hat{B}),\ \hat{B} + t\text{-critical} * SE(\hat{B})\right]$$

$$\left[\hat{\alpha} - t\text{-critical} * SE(\hat{\alpha}),\ \hat{\alpha} + t\text{-critical} * SE(\hat{\alpha})\right]$$

⑤ perform the test

rejection rule: If the hypohesized value $(\alpha^*, B^*)$ lies outside the interval the reject $H_0$ otherwise fail to reject $H_0$

Example: $\hat{Y}_t = 20.3 + 0.5091 X_t$ $\quad H_0: B = 1$

$\qquad (14.38) \quad (0.2561) \qquad H_1: B \neq 1$

critical value $= 2.086$ $\qquad$ 95% Confidence interval

$\frac{\alpha}{2} = \frac{5\%}{2} = 2.5\%$

$T - 2 = 22 - 2 = 20$ $\quad \left[0.5091 - 2.086 * 0.2561,\ 0.5091 + 2.086 * 0.2561\right]$

$= [-0.0251, 1.0433]$

$1 \in [-0.0251, 1.0433]$ the fail to reject $H_0$.

→

* B is not statistically significant and is not different from one.

$H_0 : B = 0$ $\quad$ $t\_stat = \dfrac{\hat{B} - 0}{SE(\hat{B})} = \dfrac{\hat{B}}{SE(\hat{B})} \rightarrow$ t-ratio

$H_1 : B \neq 0$

* If $|t\_stat| > t\_critical$ then reject $H_0$
  Beta is statistically significant and different from zero.

There is a relationship between $X$ and $Y$

* If $|t\_stat| <$ Critical value then fail to reject $H_0$
  B is not statistically significant and is not different from zero.

There is no relationship between $X$ and $Y$.

---

$H_0 : \alpha = 0$ $\quad$ intercept يعني $\rightarrow$ قطع $y$ axis

$H_1 : \alpha \neq 0$

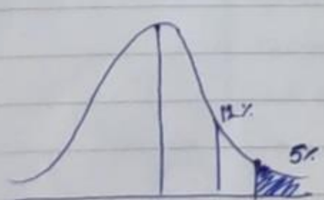
If we reject $H_0$ then the lie crosses the $Y$-axis.

If we fail to reject $H_0$ then the lie crosses the origin $(0,0)$.

$\boxed{3}$ p_value:

The probability of obtaining test results at least as extreme as the actually observed during the test assuming $H_0$ is true.

If p_value $< \alpha$ then reject $H_0$
$\alpha$ ↳ significance level.



Problem [6] Page 132

$\hat{B} = 1.147$

$SE(\hat{B}) = 0.0548$

$H_0 : B = 1$

$H_1 : B > 1$

$T = 62$

$df = 62 - 2 = 60$

$$t\text{-stat} = \frac{\hat{B} - B^*}{SE(\hat{B})} = \frac{1.147 - 1}{0.0548} = 2.68$$

$t\text{-critical} = 1.67$

$2.68 > 1.67$ then reject $H_0$

Beta is statistically significant and different from one.

problem [7] page 133

$\hat{B} = 0.214$

$SE(\hat{B}) = 0.186$

$T = 38$

$H_0 : B = 0$

$H_1 : B \neq 0$

$$t\text{-stat} = \frac{0.214 - 0}{0.186} = 1.1505$$

$t\text{-critical}$ $\frac{\alpha}{2} = \frac{5\%}{2} = 2.5\%$ $df = T - 2 = 38 - 2 = 36$

$= 2.03$

$|1.1505| = 1.1505 < 2.03$ then fail to reject $H_0$.

Beta is not statistically significant and not different from zero.

$\Rightarrow$

problem $\boxed{8}$ page 133

$$\left[\hat{\beta} - t_{critical} * SE(\hat{\beta}) \quad , \quad \hat{\beta} + t_{critical} * SE(\hat{\beta})\right]$$

ⓐ 95%

$$[-0.163 \ , \ 0.59] \qquad H_0: B = 0$$
$$H_1: B \neq 0$$

$0 \in [-0.16 \ , 0.59]$ then fail to reject Ho at 5% level

ⓑ 99% critical value = 2.71

$$[-0.29 \ , \ 0.72]$$

$0 \in [-0.29, 0.72]$ then fail to reject $H_0$ at 1% level.

~~Exalte~~ Excel مثال

Y = Sales
X = Advertising spending.

$H_0 : \alpha = 0$ $\qquad 1.6 < 2.3 \qquad$ fail Reject
$H_1 : \alpha \neq 0$

| Result of test | | $H_0$ is true | $H_0$ is false |
|---|---|---|---|
| | Significant (reject Ho) | Type one error | |
| | Insignificant (don't reject Ho) | | Type two error |

بالتقليل من قيمة α% بيقلل من Type one error ولكن رح يزيد من Type two error