| | | |
|---|---|---|
| $n = 46$ | $\sum_{i=1}^{n} X_i = 3{,}959.80$ | $\sum_{i=1}^{n} Y_i = 3{,}180.80$ |
| $\bar{X} = 86.0826$ | $\bar{Y} = 69.1478$ | |
| $\sum_{i=1}^{n}(X_i)^2 = 364{,}023.30$ | | $\sum_{i=1}^{n} X_iY_i = 319{,}943.18$ |
| $\sum_{i=1}^{n}(X_i-\bar{X})^2 = 23{,}153.3861$ | | $\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = 94{,}525.1748$ |
| $\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) = 46{,}131.6183$ | | $\sum_{i=1}^{n}\hat{u}_i^2 = 2{,}610.9211$ |

HELLO
MY NAME IS
6304641894
fernc

DO NOT FORGET THIS

$X \neq x$
uuin x matters !

a) (4 points) From regression model: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $u_i \sim NIID(0, \sigma^2)$, find the estimators of $\beta_1$ and $\beta_2$ with OLS method and explain the meaning of the model.

find $\hat{\beta}_2$ : $\hat{\beta}_2 = \dfrac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2} = \dfrac{46{,}131.6183}{23{,}153.3861}$

$\therefore \hat{\beta}_2 = 1.9924$

find $\hat{\beta}_1$ : $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}$

$= 69.1478 - 1.9924(86.0826)$

$\therefore \hat{\beta}_1 = -102.3632$

$\therefore$ SRF : $\hat{Y}_i = -102.3632 + 1.9924 X_i$

The sample regression function has intercept at $-102.3632$ and the slope is $1.9924$. Hence, when $X_i$ increase 1 unit, $\hat{Y}_i$ will increase 1.9924 units #

b) (2 points) Find $R^2$ and explain its meaning.

$r^2 : 1 - \dfrac{\Sigma\hat{u}_i^2}{\Sigma(Y_i-\bar{Y})^2} = 1 - \dfrac{2{,}610.9211}{94{,}525.1748}$

$r^2 = 0.9724 = 97.24\%$

$\therefore$ $X_i$ explain about 97.24% of the variation in $Y_i$ #

c) (1 points) If $X_i = 60$, estimate the value of $\hat{Y}_i$ and explain its meaning.

From $\hat{Y}_i = -102.3632 + 1.9924 X_i$

$= -102.3632 + 1.9924(60)$

$\therefore \hat{Y}_i = 17.1808$

$\therefore$ when $X_i$ is equal to 60, $\hat{Y}_i$ is equal to 17.1808 #

d) (3 points) Find the estimators of $\text{var}(u_i)$, $\text{var}(\hat{\beta}_1)$ and $\text{var}(\hat{\beta}_2)$ ; $k = 2$

$\text{var}(u_i) = \dfrac{\Sigma\hat{u}_i^2}{n-k} = \dfrac{2{,}610.9211}{46-2} = 59.3391$

$\text{var}(\hat{\beta}_1) = \dfrac{\Sigma X_i^2}{n\Sigma(X_i-\bar{X})^2}\sigma^2 = \dfrac{364{,}023.30\,(59.3391)}{46(23{,}153.3861)} = 20.2814$

$\text{var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\Sigma x_i^2} = \dfrac{\sigma^2}{\Sigma(X_i-\bar{X})^2} = \dfrac{59.3391}{23{,}153.3861} = 0.0026$

e) (2.5 points) What are the 95-percent confident intervals for $\beta_2$? Interpret the meaning.

d.f. $= n-k = 46-2 = 44$

find $\alpha$ : $1 - 0.95 = \alpha$

$\alpha = 0.05$

find $t_{\frac{\alpha}{2}}$ : $t_{\frac{0.05}{2}} = t_{0.025}$

$t_{0.025} = \pm 2.021$

* use S.D. not var *

$\sigma_{\hat{\beta}_2} = \sqrt{\text{var}(\hat{\beta}_2)} = \sqrt{0.0026} = 0.051$

find CI : $\hat{\beta}_2 \pm t_{\frac{\alpha}{2}}\sigma_{\hat{\beta}_2}$

$1.9924 - 2.021(0.051) = 1.8893$

$1.9924 + 2.021(0.051) = 2.0955$

$P(1.8893 \leq \hat{\beta}_2 \leq 2.0955) = 0.95$

The 95% confidence intervals for $\beta_2$ is within [ 1.8893, 2.0955 ]

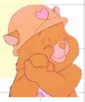
* use S.D. not var *

f) (2.5 points) Test the hypothesis whether coefficients (both $\beta_1$ and $\beta_2$ ) are different from zero at 0.05 level of significance.

$\sigma_{\hat{\beta}_1} = \sqrt{var(\sigma_{\hat{\beta}_1})} = \sqrt{20.2814} = 4.5035$

DO NOT forget to $\sqrt{var}$ !

we use S.D.



$\beta_1$ : ① $H_0$ : $\beta_1 = 0$

$H_a$ : $\beta_1 \neq 0$

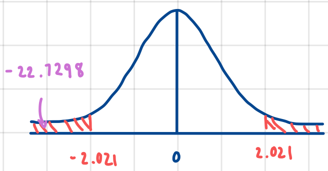
② $t_{calc} = \dfrac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} = \dfrac{-102.3632 - 0}{4.5035} = -22.7298$

③ decision rule :

d.f. = n - k = 46 - 2 = 44

$\alpha$ = 0.05

$t_{\frac{\alpha}{2}} = t_{0.025} = \pm 2.021$




$\therefore$ $t_{calc} < -2.021$, can reject null hypothesis, at the significance level 95%.
We are sure that $\beta_1$ is not zero 95 out of 100 times when we sample.

$\beta_2$ : ① $H_0$ : $\beta_2 = 0$

$H_a$ : $\beta_2 \neq 0$

② $t_{calc} = \dfrac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} = \dfrac{1.9924 - 0}{0.051} = 39.0667$

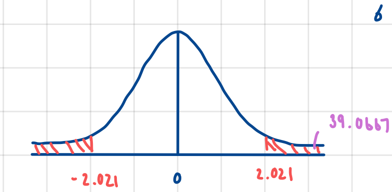
$\sigma_{\hat{\beta}_2} = \sqrt{var(\hat{\beta}_2)} = \sqrt{0.0026} = 0.051$

③ decision rule :

d.f. = n - k = 44

$\alpha$ = 0.05

$t_{\frac{\alpha}{2}} = t_{0.025} = \pm 2.021$

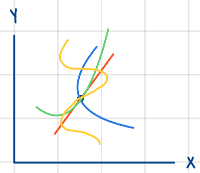



$\therefore$ $t_{calc} > 2.021$, can reject the null hypothesis, at the significance level of 95%.
We are sure that $\beta_2$ is not zero 95 out of 100 times when we sample.

**2. (8 points)** Answer the following question without any mathematical proof. A 3-6-line paragraph for a question is sufficient.

a) (2 points) If we have only one data point, can we create a sample regression function? Why?

No, if we only have one data point, a sample regression function can be any line that pass through this data. To conclude, one data point is not sufficient for creating a sample regression function.




b) (2 points) Does a significant $\beta_2$ sufficient for us to believe that $X$ and $Y$ are causally related? Provide an example to support your answer.

Yes, when $\beta_2$ is not zero, X and Y are said to be related.
For instance, $Y_i = 7.65 + 0.32 X_i$ : In this case $\beta_2$ is equal 0.32
Hence, when $X_i$ increases by 1 unit, we expect $Y_i$ will increase by 0.32 unit

c) (2 points) When we test a hypothesis and find that $\beta_2$ is significantly different from zero, what does the result actually suggest? Answer with specific wording from statistical perspective.

When $\beta_2$ is significantly different from zero, it can interpret that $\beta_2$ is different from zero at particular confidence interval, and implies we expect that changes in X are associated with changes in Y.

d) (2 points) What is (are) (an) advantage(s) of an interval estimation over point estimate?

Interval estimation can set up different confidence level and can get the range of the values. To conclude, the advantages of the interval estimation over point estimation are the pecision, indication of estimate's uncertainty, and lower chance of error.

??? 

YOU CAN DO IT

$\ln \hat{y}$ = nominal wage

x = # hour work

| Source | SS | df | MS | | n | | |
|---|---|---|---|---|---|---|---|
| Model | 50.060869 | 1 | 50.060869 | | Number of obs | = | 308 |
| Residual | 166.152715 | 306 | .542982728 | | F(1, 306) | = | 92.20 |
| | | | | | Prob > F | = | 0.0000 |
| | | | | | R-squared $r^2$ | = | 0.2315 |
| | | | | | Adj R-squared | = | 0.2290 |
| Total | 216.213584 | 307 | .704278775 | | Root MSE | = | .73687 |

| lwage | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| main_hr | .0318017 | .003312 | 9.60 | 0.000 | .0252844 | .0383189 |
| _cons | 7.658082 | .1256392 | 60.95 | 0.000 | 7.410856 | 7.905308 |

Answer the following questions. Show your work.

a) (2 points) On average, how much is the nominal wage for a person who works 0 hour a week? (Note that this is a point estimation, not a prediction)

$\ln \hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \rightarrow \hat{y}_1 = 7.658082 + 0.318017 X_i$

when $x_i = 0$ , $\hat{y}_i = 7.658082$

∴ A person who works 0 hour a week, will have 7.658082 nominal wage on average. #

b) (2 points) If a person works an hour more, how much, on average, wage change do we expect? ( mean prediction )

$\ln \hat{y}_1 = 7.658082 + 0.318017 X_i$

when $X_i = 1$, $\hat{y}_i = 7.976099$

The difference of nominal wage is $7.976099 - 7.658082 = 0.318017$

or

from $\hat{\beta}_2 = 0.318017$, it can interpret that

when the worker works an hour more, we expect nominal wage will increase by 0.318017 #

c) (3 points) If you want to change the reading of unit from hours worked to days worked, what values in the main_hr row will differ? Calculate the changes to all the values in that row, **disregarding the constant row**. You might want to impose an assumption here. State that clearly before calculation. - data scaling ? or lin-log ?

Assumption :

Assume that each worker works 8 hours per day

hour : $\ln \hat{wage} = 7.658082 + 0.318017 \text{ (main\_hour)}$

se = (0.003312) (0.1256392)

day : $\ln \hat{wage} = 7.658082 + 0.318017 (8) \text{ (main\_day)}$

se = (0.003312) (0.1256392)(8)

∴ $\ln \hat{wage} = 7.658082 + 2.544136 \text{ (main\_day)}$

se = (0.003312) (1.0051136)