

## Analysis of Variance & Regression (ANova)

Regression model  $\rightarrow$  ANova helps us to estimate levels of variability within the regression model

$$Y_t = \beta_0 + \beta_1 X_t + U_t$$

$$\underbrace{(Y_t - \bar{Y})}_{\text{total variation in response to } Y} = \underbrace{(Y_t - \bar{Y})}_{\text{variation in mean}} + \underbrace{(Y_t - \hat{Y}_t)}_{\text{residual}}$$

$$\underbrace{\sum (Y_t - \bar{Y})^2}_{\text{SST total}} = \underbrace{\sum (\hat{Y}_t - \bar{Y})^2}_{\text{SSM model}} + \underbrace{\sum (Y_t - \hat{Y}_t)^2}_{\text{SSE error}}$$

| Source of Variation | df     | Sum of Squares | Mean Square | F       |
|---------------------|--------|----------------|-------------|---------|
| Treatment           | I-1    | SSM            | MSM         | MST/MSE |
| Error               | I(J-1) | SSE            | MSE         |         |
| Total               | IJ-1   | SST            |             |         |

F statistics can be used to test

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

| Source of Variation    | df | Sum of Squares | Mean Square | F      |
|------------------------|----|----------------|-------------|--------|
| Treatment (regression) | 1  | 8654.7         | 8654.7      | 102.35 |
| Error                  | 75 | 6342.1         | 84.6        |        |
| Total                  | 76 | 14996.8        |             |        |

38 observations for each group  
2 groups

$$R^2 = SSR / SST = 8654.7 / 14996.8$$

$$\approx 0.58$$

hoca sinavda  
boşluk bırakarak  
bazılarını biz  
hesaplayacağız

Exam:

- 1- Calculation on Regression (50p) a-e
- 2- Hypothesis Testing & Power of Tests (25p)
- 3- Regression - Theoretical & CLT 2 parts (25p)
- 4- ANOVA (en son gözüğü) (10p)

120 - 150 minutes

1- Consider the following model between average GDP and average employment growth for 25 OECD countries for the period (1988-2012)

$$\lambda_t = \beta_0 + \beta_1 G_t + U_t, \quad t = 1988, 2012$$

$\lambda$  → denotes average rate of employment growth

$G$  → average rate of gdp growth

# of obs. is 25

You are provided

$$\sum_{t=1988}^{2012} \lambda_t = 20.84, \quad \sum_{t=1988}^{2012} G_t = 70.42, \quad \sum_{t=1988}^{2012} G_t L_t = 88.4636$$

$$\sum_{t=1988}^{2012} \lambda_t^2 = 42.0742, \quad \sum_{t=1988}^{2012} G_t^2 = 259.1254$$

$$\sum_{t=1988}^{2012} g_t l_t = 29.7615, \quad \sum_{t=1988}^{2012} g_t^2 = 60.7703, \quad \sum_{t=1988}^{2012} l_t^2 = 24.7020$$

$$\text{where } g_t = (G_t - \bar{G})$$

$$l_t = (L_t - \bar{L})$$

Let  $\alpha = 0.05$  for all tests of significance unless otherwise stated

a) obtain the least-squares estimates of  $\beta_0$  &  $\beta_1$  (6p)

$$\hat{\beta}_1 = \frac{\sum_t x_t y_t}{\sum_t x_t^2} = \frac{29.7615}{60.7703}$$

$$\hat{\beta}_1 = 0.4897$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \left( \frac{20.84}{25} \right) - 0.4897 \left( \frac{70.42}{25} \right)$$

$$\hat{\beta}_0 = -0.5458$$

$$\hat{L}_t = -0.5458 + 0.4897 G_t, \quad t = 1988, 2012$$

% change in gdp growth increases 0.4897

b) calculate the coefficient of determination,  $R^2$ , sample correlation coefficient between  $L_t$  &  $\hat{L}_t$  ( $r_{y\hat{y}}$ ) & sample correlation coefficient bw.  $G_t$  &  $L_t$  ( $r_{xy}$ ) then intercept them

$$R^2 = \frac{(\sum_t x_t y_t)^2}{\sum_t x_t^2 \sum_t y_t^2} = \frac{(29.7615)^2}{(60.7703)(24.7000)} = 0.59$$

$$r_{y\hat{y}} = r_{xy} = \sqrt{0.59} = 0.7681$$

c) obtain the least squares estimate of the variance of model

$$\sum_t \hat{U}_t^2 = \sum_t y_t^2 - \hat{\beta}_0 \sum_t y_t - \hat{\beta}_1 \sum_t x_t y_t$$

$$= 42.0742 - (0.5458)(20.84) - (0.4897)(88.4636)$$

$$= 10.1281$$

$$\hat{\sigma}^2 = \frac{\sum_t \hat{U}_t^2}{T-2} = \frac{10.1281}{25-2} = \boxed{0.4404}$$

d) obtain the least square estimates of variances of coefficients

$$\hat{\sigma}_{\hat{\beta}_0}^2 = \frac{\hat{\sigma}^2 \sum_t x_t^2}{T \sum_t x_t^2} = \frac{(0.4404)(259.1294)}{(25)(60.7703)} = \boxed{0.0751}$$

$$\hat{\sigma}_{\hat{\beta}_1}^2 = \frac{\hat{\sigma}^2}{\sum_t x_t^2} = \frac{0.4404}{60.7703} = \boxed{0.0072}$$

e) Construct a 95% confidence interval for  $\beta_0$

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{0.0351} = 0.2740, \quad t_{T-2, \alpha/2} = t_{25-2, \alpha/2} = t_{23, \alpha/2} \\ \downarrow \\ = 2.069$$

$$P[\hat{\beta}_0 - \hat{\sigma}_{\hat{\beta}_0} t_{T-2, \alpha/2} \leq \beta_0 \leq \hat{\beta}_0 + \hat{\sigma}_{\hat{\beta}_0} t_{T-2, \alpha/2}] = 1 - \alpha$$

$$P[-0.5458 - (0.2740)(2.069) \leq \beta_0 \leq -0.5458 + (0.2740)(2.069)] \\ = 0.95$$

$$P[-1.1127 \leq \beta_0 \leq -0.0217] = 0.95$$

f) evaluate whether slope coefficient is both statistically & economically significant

$$H_0: \beta_1 = 0 \quad \hat{\sigma}_{\hat{\beta}_1} = \sqrt{0.0072} = 0.0849$$

$$H_1: \beta_1 \neq 0 \quad t_{23, 0.025} = 2.069$$

$$t = \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{0.4897}{0.0849} = 5.7680 > 2.069 > \text{Reject } H_0 \\ \hat{\beta}_1 \text{ is stat significant}$$

yazma 5.7680 puan yok

$\hat{\beta}_1 > 0 \rightarrow$  theory consistent  $\rightarrow$  economically significant

g) test whether the regression model as a whole significant

$$H_0: \beta_0 = \beta_1 = 0$$

$$Q = \left( \frac{R^2}{1-R^2} \right) \left( \frac{T-k-1}{P} \right) \Rightarrow Q = \left( \frac{0.59}{1-0.59} \right) \left( \frac{25-2}{2} \right) = 16.585$$

$Q > F_{2,23}^{0.05} \rightarrow$  Reject  $H_0$ : The model as a whole is statistically significant

h) Test if  $\beta_0 + \beta_1 = 1$

$$H_0: \beta_0 + \beta_1 = 1$$

$$H_1: \beta_0 + \beta_1 \neq 1$$

$$t_{\hat{\beta}_0 + \hat{\beta}_1} = \frac{(\hat{\beta}_0 + \hat{\beta}_1) - 1}{\hat{\sigma}_{\hat{\beta}_0 + \hat{\beta}_1}} \quad t_{23, 0.025} = 2.069$$

$$\begin{aligned} \hat{\sigma}_{\hat{\beta}_0 + \hat{\beta}_1}^2 &= \hat{\sigma}_{\hat{\beta}_0}^2 + \hat{\sigma}_{\hat{\beta}_1}^2 - 2 \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = 0.075 + 0.072 \\ &\quad \sim 0.000001 \\ &= \sqrt{0.0823} \\ &= \hat{\sigma}_{\hat{\beta}_0 + \hat{\beta}_1} \\ &= 0.2869 \end{aligned}$$

$$t = \frac{(0.4897 - 0.5458) - 1}{0.2869} = -3.681$$

i) predict the level of  $L$  for 2013 when average GDP growth is 2%.

$$\hat{L}_{2013} = -0.5458 + 0.4897(2) = 0.4336$$

j) obtain the least square estimate of prediction error variance

$$\hat{\sigma}_{e_p}^2 = \hat{\sigma}^2 \left[ 1 + T^{-1} + \left( \frac{x_p^2}{\sum x_i^2} \right) \right]$$

$$= 0.4404 \left[ 1 + 1/25 + \left( \frac{-0.8168}{60.7703} \right)^2 \right] = 0.4629$$

$$\text{where } x_p = 2 - \bar{x} = 2 - 2.8168 = -0.8168$$

k) construct a 95% confidence interval for  $\lambda_{2023}$  when  $G_{2013} = 2$

$$P[\hat{Y}_{2013} - (\hat{\sigma}_{ep})(t_{+2, \alpha/2}) \leq Y_p \leq \hat{Y}_{2013} + (\hat{\sigma}_{ep})(t_{+2, \alpha/2})] = 1 - \alpha$$

$$P[0.4336 - (0.6804)(2.069) \leq Y_p \leq 0.4336 + (0.6804)(2.069) = 0.95]$$

$$P[-0.9741 \leq Y_p \leq 1.8413] = 0.95$$

l) test whether the prediction error is statistically significant when  $\lambda_{2013} = 1.5\%$

$$H_0: E(e_p) = 0 \quad e_p = 1.5 - 0.4336 = 1.0664$$

$$H_1: E(e_p) \neq 0 \quad t_{23, 0.025} = 2.069$$

$$t = e_p / \hat{\sigma}_{ep} = \frac{1.0664}{0.6804} = 1.5673$$

$t < t_{23, 0.025} \Rightarrow \text{Reject } H_0 \Rightarrow \text{prediction error is not statistically significant}$

m) interpret the model you have estimated & explain in detail whether the model should be reformulated

→ there is no need to reformulate the model since the slope term is statistically & economically significant

→ existing model explains 59% of relation b/w  $G_t$  &  $G_{t+1}$  as evident from  $R^2$

→ model as a whole is significant

→ predictive performance is acceptable

→ only problem is the with the intercept term  
intercept term is insignificant

the model with intercept term improve performance of model