ASHUTOSH GERA

ECONOMETRICS

QUIZ 3 | REPORT

QUESTION 1

We ran the following regression model:

RL(ID) = β0 + β1trend + u(ID),
where uID is a random error.

The results are shown below:

```
Call:
lm(formula = road_sum ~ trend, data = df)

Residuals:
   Min     1Q Median     3Q    Max
-115.0  -93.4  -55.4   25.6 3932.9

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  107.835      8.156  13.221   <2e-16 ***
trend          1.071      1.613   0.664    0.507
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 186.4 on 2122 degrees of freedom
  (301 observations deleted due to missingness)
Multiple R-squared:  0.0002077, Adjusted R-squared:  -0.0002635
F-statistic: 0.4408 on 1 and 2122 DF,  p-value: 0.5068
```

INTERPRETATION

β0 represents the road length when trend = 0, which has come out to be ~ 107.83 kms.

β1 represents the avg change in road length (in km) due to a unit increase In trend (i.e when moving from any one year to the next).

Thus, between the years 2011-2021, the road length increased on avg ~ 1.07 kilometres each year

β0 -> the intercept
β1 -> the slope
RL -> Response variable
trend -> predictor variable

QUESTION 3

We ran the following regression model:

RL(ID) = γ0 + γ1DSouth + ε(ID),
where ε(ID) is a random error.

The results are shown below:

```
> summary(road_lm_DSouth)

Call:
lm(formula = road_sum ~ DSouth, data = df)

Residuals:
   Min     1Q Median     3Q    Max
-117.8  -94.6  -48.7   27.0 3930.9

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  118.363      4.229   27.99  < 2e-16 ***
DSouth       -62.179     13.817   -4.50 7.16e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 185.6 on 2122 degrees of freedom
  (301 observations deleted due to missingness)
Multiple R-squared:  0.009453,  Adjusted R-squared:  0.008986
F-statistic: 20.25 on 1 and 2122 DF,  p-value: 7.158e-06
```

The model is a linear regression of road length (in kilometers) on the dummy variable DSouth, which takes the value 1 if the district belongs to any of the following states: Andhra Pradesh, Tamil Nadu, Kerala, or Telangana, and 0 otherwise.

INTERPRETATION

γ0 represents the road length when DSouth = 0, which has come out to be ~ 118.36 kms.

γ1 is the slope, is equal to ~ -62.18kms, which means that when DSouth is equal to 1 (i.e., when the district is in the southern states), the expected road length is ~62.18 kilometers LESS than when DSouth is equal to 0 (i.e., when the district is not in the southern states).

γ0 -> the intercept
γ1 -> the slope
RL -> Response variable
DSouth -> predictor variable

QUESTION 4

We ran the following regression model:

$$RL(ID) = \eta 0 + \eta 1 DSouth + \eta 2 DNorth + \eta 3 DEast + \eta 4 DWest + \theta(ID),$$

where $\theta(ID)$ is a random error.

The results are shown below:

```
Call:
lm(formula = road_sum ~ DSouth + DNorth + DEast + DWest, data = df)

Residuals:
   Min     1Q Median     3Q    Max
-149.0  -87.3  -43.6   24.2 3950.2

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   99.119      6.004  16.509  < 2e-16 ***
DSouth       -42.935     14.339  -2.994  0.00278 **
DNorth        43.658     11.194   3.900 9.92e-05 ***
DEast         51.298     10.168   5.045 4.93e-07 ***
DWest        -46.897     18.514  -2.533  0.01138 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 183.7 on 2119 degrees of freedom
  (301 observations deleted due to missingness)
Multiple R-squared:  0.03071,   Adjusted R-squared:  0.02888
F-statistic: 16.78 on 4 and 2119 DF,  p-value: 1.492e-13
```

The model predicts the road sum based on the district belonging to any of the south, north, east, or west regions. The coefficients indicate the average effect on the road sum of belonging to any of the respective regions, while holding other factors constant

INTERPRETATION

η0, The intercept of ~99.12 kms represents the estimated average road sum for districts that do not belong to any of the four regions.

η1, the coefficient of ~ −42.94 for DSouth indicates that the mean outcome variable for districts in the southern region (represented by DSouth=1) is estimated to be 42.94 km lower than the mean outcome for the reference region (represented by DSouth=0) while holding all other variables constant.

η0 -> the intercept term
η1 -> the coefficient of DSouth
η2 -> the coeff. of DNorth
η3 -> the coeff. of DEast
η4 -> the coeff. of DWest

INTERPRETATION (CONTD)

η2, the coefficient of ~ 43.66 for DNouth indicates that the mean outcome variable for districts in the Northern region (represented by DNorth = 1) is estimated to be ~ 43.66 km higher than the mean outcome for the reference region (represented by DNorth=0) while holding all other variables constant.

η3, the coefficient of ~ 51.3 for DEast indicates that the mean outcome variable for districts in the eastern region (represented by DEast = 1) is estimated to be ~ 51.3 km higher than the mean outcome for the reference region (represented by DEast = 0) while holding all other variables constant.

η4, the coefficient of ~ -46.9 for DWest indicates that the mean outcome variable for districts in the western region (represented by DWest = 1) is estimated to be ~ 46.9 km lower than the mean outcome for the reference region (represented by DWest = 0) while holding all other variables constant.

NOTE: In this model with four dummy variables representing different regions of a country and a continuous outcome variable, the coefficients represent the difference in the mean outcome between the reference region (not included as a dummy variable, i.e **DCentre** in our case) and each of the other regions, while holding all other variables constant.

QUESTION 5

We ran the following regression model:

$$RL(ID) = \alpha 0 + \alpha 1 trend + \alpha 2 DSouth + \alpha 3 DSouth * trend + \delta(ID),$$

where δID is a random error.

The results are shown below:

```
Call:
lm(formula = road_sum ~ trend + DSouth + (DSouth * trend), data = df)

Residuals:
   Min     1Q Median     3Q    Max
-127.2  -94.0  -45.6   25.1 3920.7

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   106.350      8.404  12.654   <2e-16 ***
trend           2.779      1.681   1.653   0.0984 .
DSouth          2.341     32.287   0.073   0.9422
trend:DSouth  -13.187      5.850  -2.254   0.0243 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 185.4 on 2120 degrees of freedom
  (301 observations deleted due to missingness)
Multiple R-squared:  0.01233,   Adjusted R-squared:  0.01094
F-statistic: 8.825 on 3 and 2120 DF,  p-value: 8.183e-06
```

The above model is a multiple linear regression model that aims to explain the variation in the dependent variable "road_sum" based on three independent variables: "trend", "DSouth", and the interaction between "trend" and "DSouth".

INTERPRETATION

The intercept (α0) is 106.350, which means that when trend and DSouth are equal to zero, the estimated mean of road_sum is 106.350 km.

The coefficient estimate for the trend variable (α1) is 2.779, which means that on average, for each unit increase in the trend, the estimated mean of road_sum increases by 2.779, holding DSouth constant.

α0 -> the intercept term
α1 -> the coefficient of trend
α2 -> the coeff. of DSouth
α3 -> the coeff. of (trend*DSouth), it captures the interaction effect between tID and DSouth.

INTERPRETATION (CONTD)

The coefficient estimate for the DSouth (α2) is ~ 2.34, which means that, on average, the estimated mean of road_sum is 2.341 km higher for roads in the southern region, holding the trend (i.e. year) constant.

The coefficient estimate for the interaction between trend and DSouth (α3) is ~ -13.19, which means that the effect of trend on road_sum is different for roads in the southern region than for roads in other regions. In particular, for each unit increase in the trend, the estimated mean of road_sum decreases by 13.19 km for roads in the southern region.

THANK YOU

ASHUTOSH GERA
2021026
CSE