Asha Bintang W M.
23/511656/EK/24267

Quiz #9

1- From the regression $\hat{Y} = 6.9 + 0.055X_1 + 0.107X_2 + 0.085X_3$

We know that:

- $\hat{Y}$ is the predicted GPA
- X_1 is the IQ score
- X_2 is the hours spent studying
- X_3 is the highschool average

a) With $X_1 = 108$, $X_2 = 32$, $X_3 = 82$. We get:

$\hat{Y} = 6.9 + 0.055(108) + 0.107(32) + 0.085(82)$

$= 6.9 + 5.94 + 3.424 + 6.97 = 23.234$ //

b) For difference of the two $X_2 = 36, 32$ we get:

$\Delta X_2 = 36 - 32 = 4$

since it is an addition of 4 on X_2.

$\Delta \hat{Y} = 0.107 \times 4 = 0.428$ //

Meaning that $\hat{Y}$ would be 23.862

c) The smallest coefficient is IQ score (X_1) with 0.055, meaning that it has the smallest effect //

d) The largest coefficient is time studying (X_2) with score of 0.107 //

2. a) Based on the scatterplot on Net Migration Vs Unemployment rate, there is no clear linear relationship as the points are widely dispersed with a high concentration around the origin, suggesting that changes in net migration do not have a straight correlation with unemployment. There are as well outliers that shows that countries have high Net migration /unemployment rate that occurs independently of each other. This is also the case for net migration Vs Annual Average temperature and net migration Vs Annual average daily precipitation. Meaning net migration varies widely within the temperature and precipitation levels, regardless of every points in the two variables. thus concluding that all of the variables versused net migration, is not a strong predictor of the net migration.

b) From the correlation matrix, We cans see that most of the variables versused with each other does not have strong correlations, meaning that there is no probable connection between the variables with the exception Unemployment rate Vs HPI. (that has negative correlation -0.27), Net Migration Vs Unemployed & Unemployment rate (that has negative correlation -0.21 respectively), Employed Vs Unemployed (r= 0.97), Labor rate Vs Unemployed & employed (r= 0.98 & 1.00 respectively). The negative correlation indicates an inverse relationship but not strong as it is relatively small, and The High positive correlation indicates strong relationship which means that for variable X Vs Y, we can expect an increase/decrease of X if Y is increased/decreased by value/number.

c) The document is in the Apendix

d) Since the data has mising values, We both imputed and dropped it prior to conducting the simple linear regressin model (OLS). Here is the summary of the findings:

- For Net Migration Vs Unemployment rate, With R^2= 0.00 and coefficient -19.25 & -18.17, P-Value 0.293 & 0.312 for Dropped & imputed respectively, We see that Unemplogment rate is not a significant predictor for net migration. And R^2=000 means that it virtually explains no variations

- Net Migration Vs Metro both have $R^2 = 0.099$ and coefficient of 9725.65 & 9608.34 for Dropped & Imputed respectively. P-Value scoring 0.00 means that being a metropolitan area significantly predicts higher Net migration/ Metropolitan area attracts more migrants.
- Net Migration Vs Annual average Temperature both have $R^2 = 0.01$ and coefficients larger than P-Value = 0.00 (Respectively), We can say that it is significant and that basically areas with higher temperature tend to have higher net migration (although similarly to the previous, the model did not explain much of the variation).
- Net Migration Vs Annual average Deity-precipitation have $R^2 = 0.001$, P-Value of 0.171 & 0.170 (Dropped & Imputed respectively), and Coefficient higher than P-Value at 74.65, indicates that it is not significant since R^2 is too small. Thus there is no impact on it.

e) The document is in the Apendix

f) From all the four regression, we found that Metro is the best Predictor since it is the most significant with results (coefficient) consistently above 700 in comparison to the others.

g) As a researcher I would gather data label on median household income, Cost of living (in dollars or in some index), employment growth (Probably the rates/percentage change), Crime rates, Healthcare access (dummy variable or some continuous value with whether there is Healthcare support or no, and/or How many Healthcare institution and it's mean distance to each other), Tax rates, Government service (Probably dummy variables on education support/no, Public transportation/ no. But it is also likely for some integer value because the number of schools or unit of transportation may be much more accurate.)

Data and Visualizations statistical results can be accessed via the github link.

THIS IS THE MENTIONED GITHUB LINK

https://github.com/Mrhemm/Tugas-Statek/blob/main/Quiz%204.ipynb