1) Before making a complete switch from BlackBoard to Canvas, the usage rate of BlackBoard was reported to be 56.43% in a recent month. Suppose that you decide to select a sample of 100 students at your university and you find that 60 use the BlackBoard platform. (12 points)

(a) Use the critical value approach to determine whether there is evidence that the usage rate for the BlackBoard platform at your university is greater than the reported usage rate of 56.43%. (Use the 0.05 level of significance.)

To tests if there is difference in the two portions;
Then hypothesis statement will be;
**$H_0$**: P=*p* (Population portion is not significantly different from sample proportion for the usage off Blackboard))
**$H_1$**: p < *p* (sample proportion is significantly greater from population proportion for the usage of Blackboard).

Population proportion (p) = 56.435 = 0.5643
Sample proportion (p^) = 60/100 = 0.6000
**Critical value at 95% (0**.05)
Z = (*p* -p)/√p(1-p)/n
Z= (0.6-0.5643)/√0.4357*0.5643/100
**Z= 0.7200**
**$Z_{0.05}$ = 1.645.**
**We therefore reject H0, since our test statistic is less than critical value.**
**In conclusion, there is sufficient evidence that the sample proportion is greater than the population for the usage of Blackborad.**

(b) Suppose that the sample is n = 500, and you find that 60% of the sample of students at your university (300 out of 500) use the BlackBoard platform. Use the p-value approach to try to determine whether there is evidence that the usage rate for the BlackBoard platform at your university is greater than the reported usage rate of 56.43%. (Use the 0.05 level of significance.) (Use the 0.05 level of significance.)

To tests if there is difference in the two portions;
Then hypothesis statement will be;
**$H_0$**: P=*p* (Population portion is not significantly different from sample proportion for the usage off Blackboard))
**$H_1$**: p < *p* (sample proportion is significantly greater from population proportion for the usage of Blackboard).

Population proportion (p) = 56.435 = 0.5643

Sample proportion (p^) = 3000/500 = 0.6000

**Critical value at 95% (0**.05)
Z = (*p* -p)/√p(1-p)/n
Z= (0.6-0.5643)/√0.4357*0.5643/500
**Z= 0.2397,**
**P-value corresponding is 0.59483**

**We therefore fail to reject H0, since our p-value is greater than critical value,0.05.**

**In conclusion, there is insufficient evidence that the sample proportion is greater than the population for the usage of Blackborad.**

2) A local pizzeria wants to reduce the waiting time for customers to receive their pizza orders. Until now, the population mean waiting time to receive the pizza orders was observed to be at least 4 minutes. However, in an effort to reduce the mean waiting time, the pizzeria has experimented with a new baking system with enhanced technology. A sample of 81 customers was selected, and their mean waiting time to receive the pizza orders was found to be 3.25 minutes, with a sample standard deviation of 130 seconds. (12 points)

(a) Using the critical value approach is there any evidence that the new baking system is successful to reduce the mean waiting time?

**H0:** The new baking system is not successful.

**H1:** T he new baking system is successful.

Calculating t-test,

t= (x-bar - μ)/ s/√n

t= (3.25-4)/2.167/√81

t= -3.112

t critical value is approximately Z critical value since our n is greater than 30, therefore,

$t_{ā}$ = 1.645.

Comparing test statistic, t = -3.112, absolute value, 3.112, with critical value, $t_{ā}$ = 1.645, then we fail to reject **H0.**

**In conclusion, there is no sufficient evidence that the new asking system is successful.**

(b) Determine the p-value and interpret its meaning.

p-value =1.9982

3) An auditor for a government agency is assigned the task of evaluating reimbursement for office visits to physicians paid by Medicare. The audit was conducted on a sample of 72 of the reimbursements, with the following results: (12 points)

In 12 of the office visits, an incorrect amount of reimbursement was provided. The mean amount of reimbursement was $93.70 with a standard deviation of $34.55.

(a) At the 0.05 level of significance, is there evidence that the population mean reimbursement was more than $100? (Use the critical value approach)

The hypothesis for the one sample *T* test can be expressed as:

$H_0$: $\mu = \mu_0$

$H_1$: $\mu \neq \mu_0$

where μ is the "true" population mean and $\mu_0$ is the proposed value of the population mean.

t= (x-bar - μ)/ s/√n

t= (93.70-100)/34.55/√72

t= -1.54726527003

at n= 72, t-value approaches z value, hence $t_{critical}$ =1.667.

Since t absolute value is 1.5473, is less than $t_{critical}$ =1.667, we reject $H_{0.}$

In conclusion, there is no sufficient evidence at 0.05 significance level that the population was more than 100.

(b) At the 0.02 level of significance, is there evidence that the proportion of incorrect reimbursements in the population was below 0.10? (Use the p-value approach)

Hypothesis statement will be;
**$H_0$**: P=*p* (Population portion is not significantly different from sample proportion)
**$H_1$**: p > *p* (sample proportion is significantly smaller from population proportion ).

Population proportion (p) = 0.10
Sample proportion (*p*) = 12/72 = 0.1667
**Critical value at 95% (0**.05)

P = 0.10

*P* = 12/72 = 0.1667

Z = (*p* -p)/√p(1-p)/n
Z= (0.1667-0.10)/√(0.10*0.90)/12
**Z= 0.2267**
**p-value corresponding to this test statistic is 0.59095.**
**We therefore fail to reject H0, since our p-value is much greater than the significance level of 0.02.**
**In conclusion, there is insufficient evidence that the sample proportion is smaller than the population.**

4) There is a dataset with nineteen observations on the yearly Gross Domestic Product (GDP) of a country. There are three predictor variables that affect the GDP of the country – Education Spend in $million, the Unemployment rateas %, Employee Compensation in $million. You can find this data in the Linear-Regression-Exam-Question-4.xlsx file. (12 points)

(a) State the multiple regression equation.

Let Y rep GDP,

$X_1$ rep Education Spend, (X variable 1)

$X_2$ rep Unemployment rate, (X variable 2)

$X_3$ rep Employee Compensation, (X variable 3)

And $\beta_{0,}$ $\beta_1$, $\beta_2$ and $\beta_3$ are the coefficients.

Therefore multiple regression is defined as:

$$\mathbf{Y = \beta_0 + \beta_1X_1 + \beta_2X_1 + \beta_3X_3}$$

**Substituting;**

$$\mathbf{Y(GDP) = 29500.43 + 3.983382X_1 - 2189.43X_2 + 1.434038X_3}$$

---

(b) Interpret the meaning of β1, β2, and β3, in this problem.

**$\beta_1$, $\beta_2$ and $\beta_3$** are the coefficients of the independent variables, that is, X variable 1, X variable 2 and X variable 3 respectively. This coefficient tells us whether there is a negative or positive correlation between each independent variable (Education Spend, Unemployment rate, and Employee Compensation) and the dependent variable, GDP.

From the equation above, **$\beta_1$ and $\beta_3$** (**3.983382** and **1.434038 respectively),** are positive meaning, Education Spend, (X variable 1) and Employee Compensation, (X variable 3) respectively are each positive correlated to GDP in a manner that, a unit increase in Education Spend will result to an increase in GDP by **3.983382 units** and a unit increase in Employee Compensation will mean **1.434038 increase** in GDP.

**$\beta_2$** is a negative value, implying that Unemployment rate, (X variable 2), is negatively correlated to GDP hence a unit increase in Unemployment rate will result to decrease of GDP by **2189.43.**

---

(c) What is the significance of$\beta$0?

**$\beta_0$** is a constant coefficient. The significance of this coefficient is that it tells us that without all these other factor or rather when all these predictor variable are zeros, then our GDP will be **$\beta_0$** , (**29500.43) .**

---

(d) Predict the GDP when Education Spend is $13,260 million, the Unemployment rate is 8.13%, and Employee Compensation is $199,999 million.

**Y(GDP) = 29500.43 + 3.983382$X_1$ - 2189.43$X_2$ + 1.434038$X_3$**

**Substituting;**

**Y(GDP) = 29500.43 + 3.983382(13,260) - 2189.43(8.13) + 1.434038(199999)**

**Y(GDP) = 351326. 148862**

---

(e) How much percent of variation in the GDP is explained by the regression model?

This is explained by adjusted square R, which is 0.985007. This implies, the variation explained by the regression model **is 98.5%.**

(f) Does the regression model overestimate or underestimatethe GDP for the year 2015?

Estimating 2015 GDP;

**Y(GDP), 2015 = 29500.43 + 3.983382(26282) - 2189.43(8.5) + 1.434038(208128)**

**Y(GDP), 2015 =** 414044.981588

From the table, 2015 GDP is 416,701.

The model therefore underestimated the 2015 GDP by **2656.018412.**

Paste your Excel output and working:

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.993733 |
| R Square | 0.987506 |
| Adjusted R Square | 0.985007 |
| Standard Error | 7659.401 |
| Observations | 19 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 3 | 6.96E+10 | 2.32E+10 | 395.1845 | 1.71E-14 |
| Residual | 15 | 8.8E+08 | 58666431 | | |
| Total | 18 | 7.04E+10 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 29500.43 | 31570.15 | 0.93444 | 0.364877 | -37789.8 | 96790.62 | -37789. | 96790.62 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | 8 | |
| X Variable 1 | 3.983382 | 3.651881 | 1.090775 | 0.292584 | -3.80042 | 11.76718 | -3.80042 | 11.76718 |
| X Variable 2 | -2189.43 | 2466.116 | -0.887816 | 0.38866 | -7445.83 | 3066.971 | -7445.83 | 3066.971 |
| X Variable 3 | 1.434038 | 0.556454 | 2.577099 | 0.021036 | 0.247984 | 2.620093 | 0.247984 | 2.620093 |

5) There is a dataset with some information for each of the ten states in the United States. We are interested in predicting the per-pupil student expenditures in a state. To do so, there are three predictor variables: average income of that state, the % of the population under 18, and the region in which the state falls. You can find this data in the Linear-Regression-Exam-Question-5.xlsx file. (12 points)

(a) State the multiple regression equation.

Let Y rep Expendture,

$X_1$ rep Av-income, (X variable 1)

$X_2$ rep % under_18, (X variable 2)

$X_3$ rep Region , (X variable 3), ( Since these are categorical variables, we rewrite them assigning nuericals as;

| Region | Region |
|---|---|
| Southeast | 1 |
| Southeast | 1 |
| Northeast | 2 |
| Southeast | 1 |
| West | 3 |
| Midwest | 4 |
| Southeast | 1 |
| Midwest | 4 |
| Northeast | 2 |
| Midwest | 4 |

Respectively)

And $\beta_{0,}$ $\beta_1$, $\beta_2$ and $\beta_3$ are the coefficients.

Therefore multiple regression is defined as:

$\mathbf{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}$

**Substituting;**

$\mathbf{Y}$ = -1143.9+ 0.061739$\mathbf{X_1}$ + 3504.168$\mathbf{X_2}$ - 9.41984$\mathbf{X_3}$

---

(b) Predict the per-pupil student expenditures for a statewith an average income of 5889, %_under_18 of 31.8%, and is in the Northeast region.

**Y** = -1143.9+ 0.061739**(5889)** + 3504.168**(0.318) -** 9.41984**(2)**

**Y = 315.166715**

---

(c) Predict the per-pupil student expenditures for a state with an average income of 5132, %_under_18 of 33.4%, and is in the West region.

**Y** = -1143.9+ 0.061739**(5132)** + 3504.168**(0.334) -** 9.41984**(3)**

**Y = 315.07714**

---

(d) How much percent of variation in the per-pupil student expenditures is explained by the regression model?

This is explained by adjusted square R, which is 0.891769. This implies, the variation explained by the regression model is **89.18%.**

---

(e) Does the regression model overestimate or underestimate the per-pupil student expenditures for the state “Ohio”?

Ohio has an average income of 5012, %_under_18 of 32.4%, and is in the Midwest region.

**Y** = -1143.9+ 0.061739**(5012)** + 3504.168**(0.324) -** 9.41984**(4)**

**Y = 263.2069**

**This has overestimated Ohio from 221 to 263.2069.**

---

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.963248 |
| R Square | 0.927846 |
| Adjusted R Square | 0.891769 |
| Standard Error | 33.75533 |
| Observations | 10 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 3 | 87912.36 | 29304.12 | 25.7184 | 0.000799 |
| Residual | 6 | 6836.535 | 1139.423 | | |
| Total | 9 | 94748.9 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | -1143.9 | 184.757 | -6.19137 | 0.000818 | -1595.98 | -691.814 | -1595.98 | -691.814 |
| X Variable 1 | 0.061739 | 0.037523 | 1.645348 | 0.151 | -0.03008 | 0.153554 | -0.03008 | 0.153554 |
| X Variable 2 | 3504.168 | 802.3129 | 4.367582 | 0.004731 | 1540.979 | 5467.356 | 1540.979 | 5467.356 |
| X Variable 3 | -9.41984 | 13.85042 | -0.68011 | 0.521796 | -43.3106 | 24.47092 | -43.3106 | 24.47092 |