Question 4.1

- Clearly state chosen Variables.

  Dependent Variable — Independent Variables (5)

- Build a paragraph for each Variable saying why it was chosen, using appropriate literature (Reports, Journal Articles).

- Discuss the Descriptive Statistics

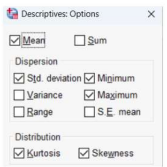
  i.e describe the nature of the Variable.

Descriptive Statistic table

Example

Variables	Mean	Std Dev	Min	Max
Price	6165.26	2949.50	...	....
Weight	3019.46	777.19	...	....

Analyse
-> Descriptive Statistics
-> Descriptives
-> Options Button
-> Check Boxes.



Q4.2 Step 1 Define the model

Write the model based on chosen variables.

Example: $\text{mthly sales} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

where: β_0 is the intercept

β_1 β_2 β_3 are the coefficients of ....

ε is the error term

Step 2 - Check correlation of Chosen Variables.

- Analyse → Correlate
  - → Select variable
  - → OK.
- check for Significant Relationship; Write up Findings

step 3 - Run Regression Analysis.

- Analyse → Regression
  - Select Dep Var
  - Select Indep Variables
  - → Choose [Enter] Method
  - → Click [Statistics]
    - → Check Collinearity Diagnostics.

Step 4 Interpret Results

① Goodness of Fit using R^2

% of variation in the dependent variable explained by the independent variables.

Interpret R^2 value, write up findings

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.811[a]	.658	.615	72167.801
2	.915[b]	.837	.790	53250.533
3	.962[c]	.926	.889	38735.512

a. Predictors: (Constant), X2
b. Predictors: (Constant), X2, X4
c. Predictors: (Constant), X2, X4, X6

② Test for Significance of the Model

- ANOVA F-test.

$H_0: \beta_0 = \beta_1 = \beta_2 \cdots \beta_k = 0$

$H_1: \beta_j \neq 0$ for at least one j

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.014E+10	1	8.014E+10	15.387	.004[b]
	Residual	4.167E+10	8	5208191543		
	Total	1.218E+11	9			
2	Regression	1.020E+11	2	5.098E+10	17.978	.002[c]
	Residual	1.985E+10	7	2835619270		
	Total	1.218E+11	9			
3	Regression	1.128E+11	3	3.760E+10	25.060	<.001[d]
	Residual	9002639413	6	1500439902		
	Total	1.218E+11	9			

a. Dependent Variable: Monthly Sales
b. Predictors: (Constant), X2
c. Predictors: (Constant), X2, X4
d. Predictors: (Constant), X2, X4, X6

If $p < 0.05$, we Reject H_0

∴ The model is significant as it explains some variation of the dependent variable.

write up findings.

③ Test for Significance for each Independent Variable

$H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

• t Test

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	100814.413	74881.737		1.346	.215		
	X2	1804.920	460.125	.811	3.923	.004	1.000	1.000
2	(Constant)	-59540.320	79969.378		-.745	.481		
	X2	1385.280	371.696	.623	3.727	.007	.834	1.199
	X4	15.582	5.589	.463	2.774	.028	.834	1.199
3	(Constant)	-27102.134	59409.276		-.456	.664		
	X2	2111.925	382.290	.949	5.524	.001	.417	2.396
	X4	26.685	5.816	.798	4.588	.004	.408	2.453
	X6	-5.385	2.003	-.629	-2.689	.036	.225	4.450

a. Dependent Variable: Monthly Sales

If $|P < 0.05|$, we Reject H_0

∴ the Independent Variable has a statistically Significant effect on the dependent Variable.

Write up findings for each independent variable.

④ Check for Multicollinearity

Variance Inflation Factor (VIF)

If VIF > 5

∴ Remove Variable from Model.

Step 5 - State Final Model

Example:

$$y\,(\text{monthlySales}) = -27102.13 + \underline{2111.93}\,X_1 + \underline{26.69}\,X_2 - \underline{5.39}\,X_3$$

Step 6 – Create Table to Report Regression Results

Example:

Variables	Dependent Var Monthly Sales
X_2	2111.93* ← coefficient (382.29) ← std error
X_4	26.69* (5.82)
X_6	−5.39** (2.00)
Constant	−27102.13
observations	10
R^2	0.93

* $p < 5\%$ (0.05)
** $p < 10\%$ (0.10)

Step 7 – Interpret Coefficients

Example:
If No discounts are applied then Mthly sales = −27102.

- For one unit increase in disc 2 (X2) Monthly Sales will increase by $2111.93
- For one unit increase in disc 4 (X4) Monthly sales will increase by $26.69
- For one unit increase in disc 6 (X6) Monthly sales will decrease by $5.39

Q4.3 Diagnostics.

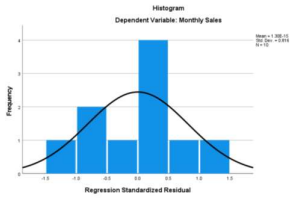
A healthy regression model needs to have its residuals:

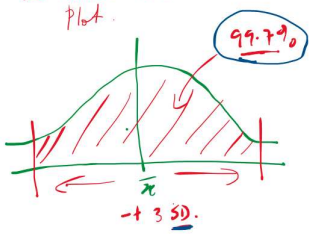
1. Normally Distributed
2. Not - Autocorrelated
3. Homoscedastic

① Test for Residuals being Normally Distributed

① Graphical - Histogram
- look for Bell shape curve
- check bars are under the curve
- Ensure SD not > 3

Analyse
→ Regression
→ Linear
→ choose Var
→ Plots
→ ☑ Histogram
☑ Normal Prob Plot





② Shapiro-Wilk Statistic

Whether the sample data originated from a Normally Dist Population

H_0 : The residuals are Normally Distributed

If $P < 0.05$, we reject H_0

If $P > 0.05$, we fail to reject H_0

∴ The residuals are normally distributed.

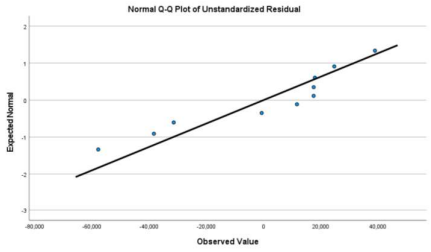
Regression
→ linear
→ Choose Var
→ Save
→ ☑ Unstd Residuals
☑ Std Residuals
→ Analyse
→ Descriptive
→ Explore
→ Select Unstd + Std Res to Dependent Factor
→ Plots
☑ Normality Plot with Test.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.244	10	.094	.886	10	.154
Standardized Residual	.244	10	.094	.886	10	.154

a. Lilliefors Significance Correction

Since $P > 0.05$, we conclude that the residuals are normally distributed

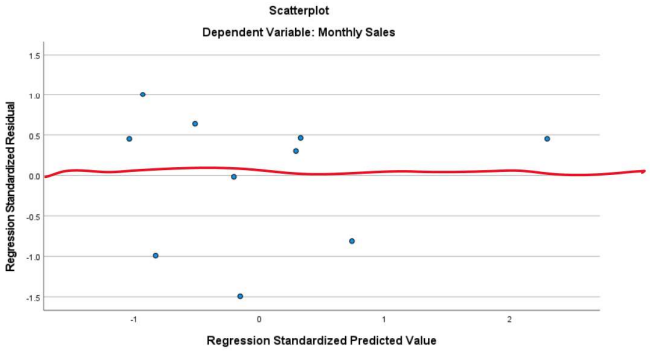


If values are close to the line, it means that the residuals are minimal Therefore it is Normally Distributed

② Test for Constant Variance
↳ the Error term
i.e Test for Homoscedasticity

- Scatter Plot
  - → Z Res (y)
  - → Z Pred (x)

Analyse
→ Regression
→ Linear
→ Choose Var
→ Plots
→ Select:
  Z Resid → Y
  Z Pred → X
→ Continue
→ Ok.



If the spread or variance of the residuals are pretty constant i.e. the distance from '0' is the same regardless of the Std Predicted values i.e. as the Std Pred values increase.
Then there is Homoscedasticity ✓

If the variance is Not the same (constant) as the Std Pred values increase, i.e. the distance from '0' is smaller or larger
Then there is Heteroscedasticity

③ Test for Auto Correlation of the Residuals

Durbin-Watson Test, checks whether residuals are independent over time.

If the residuals are independent or uncorrelated
Then there is No Auto correlation

Analyse
→ Regression
→ linear
→ Statistics
☑ Durbin Watson

Range of Durbin-Watson Test Values:

d-Value	Interpretation	
2	No Auto correlation	
0-2	Positive Auto correlation	Yesterday + Today same behaviour
2-4	Negative Autocorrelation	Yesterday + Today opp behaviour

Values under 1 or more than 3 are definite cause for concern.

Rule of Thumb: d value between 0-2, ∴ No Auto correlation

Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.811[a]	.658	.615	72167.801	
2	.915[b]	.837	.790	53250.533	
3	.962[c]	.926	.889	38735.512	1.204

a. Predictors: (Constant), X2
b. Predictors: (Constant), X2, X4
c. Predictors: (Constant), X2, X4, X6
d. Dependent Variable: Monthly Sales

- Based on all the test for residual problem write up a conclusion.
- Discuss how you can address them.