Example

A distributor of cakes wants to locate factors that could influence demand and, therefore, sales

- Dependent variable → Pie sales (per week)
- Explanatory variables → Price ($)
  → advertising (per $100)

Data collection for 15 weeks

Example

We have to verify the correlations among variables:

- dependent vs explanatory variables
- among explanatory variables (multicollinearity)

r(pie sales&price) = - 0.44

r(price&advertising) = -0.03 (no multicollinearity)

r(pie sales&advertising) = 0.56

	Pie sales	Price	Advertising
Pie sales	1,00	-0,44	0,56
Price	-0,44	1,00	-0,03
Advertising	0,56	-0,03	1,00

Example

Week	Pie Sales	Price ($)	Advertising ($100s)
1	350	5.50	3.3
2	460	7.50	3.3
3	350	8.00	3.0
4	430	8.00	4.5
5	350	6.80	3.0
6	380	7.50	4.0
7	430	4.50	3.0
8	470	6.40	3.7
9	450	7.00	3.5
10	490	5.00	4.0
11	340	7.20	3.5
12	300	7.90	3.2
13	440	5.90	4.0
14	450	5.00	3.5
15	300	7.00	2.7

$$\text{Sales} = \beta_0 + \beta_1(\text{Price}) + \beta_2(\text{Advertising}) + \varepsilon$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Output of the regression analysis (SPSS)

- Output
  - Log
  - Regression
    - Title
    - Notes
    - Active Dataset
    - Descriptive Statistics
    - Correlations
    - Variables Entered/Removed
    - Model Summary
    - ANOVA
    - Coefficients
    - Coefficient Correlations
    - Collinearity Diagnostics
    - Residuals Statistics
    - Charts
      - Title
      - *zresid Histogram
      - *zresid Normal P-P Plot
      - *zresid by *zpred Scatterplot

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	306,526	114,254		2,683	,020	57,588	555,464		
	Price ($)	-24,975	10,832	-,461	-2,306	,040	-48,576	-1,374	,999	1,001
	Advertising ('100$)	74,131	25,967	,570	2,855	,014	17,553	130,709	,999	1,001

a. Dependent Variable: Pie Sales

Coefficient Correlations[a]

Model			Advertising ('100$)	Price ($)
1	Correlations	Advertising ('100$)	1,000	-,030
		Price ($)	-,030	1,000
	Covariances	Advertising ('100$)	674,302	-8,562
		Price ($)	-8,562	117,335

a. Dependent Variable: Pie Sales

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	Price ($)	Advertising ('100$)
1	1	2,969	1,000	,00	,00	,00
	2	,023	11,314	,01	,73	,29
	3	,007	19,948	,99	,26	,71

a. Dependent Variable: Pie Sales

Coefficients of regression model

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	306,526	114,254		2,683	,020	57,588	555,464		
	Price ($)	-24,975	10,832	-,461	-2,306	,040	-48,576	-1,374	,999	1,001
	Advertising ('100$)	74,131	25,967	,570	2,855	,014	17,553	130,709	,999	1,001

a. Dependent Variable: Pie Sales

$$Y = 306.53 - 24.98(X_1) + 74.13(X_2)$$

$$Sales = 306.53 - 24.98(Price) + 74.13(Advertising)$$

Equation

$$\boxed{Y = 306.53 - 24.98(X_1) + 74.13(X_2)}$$

where
pie sales per week in $
Advertising in $ (x100)

b_1 = -24.98: pie sales decrease in mean of 24.98 cakes per week for each dollar of increase in price, not considering the effects due to the variation of advertising

b_2 = 74.13: pie sales increase in mean of 74.13 cakes per week for each 100$ of increase of advertising, not considering the effects due to price changes

R^2 and adjusted R^2

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df 1	df 2	Sig. F Change	
1	,722[a]	,521	,442	47,463	,521	6,539	2	12	,012	1,683

a. Predictors: (Constant), Advertising ('100$), Price ($)

b. Dependent Variable: Pie Sales

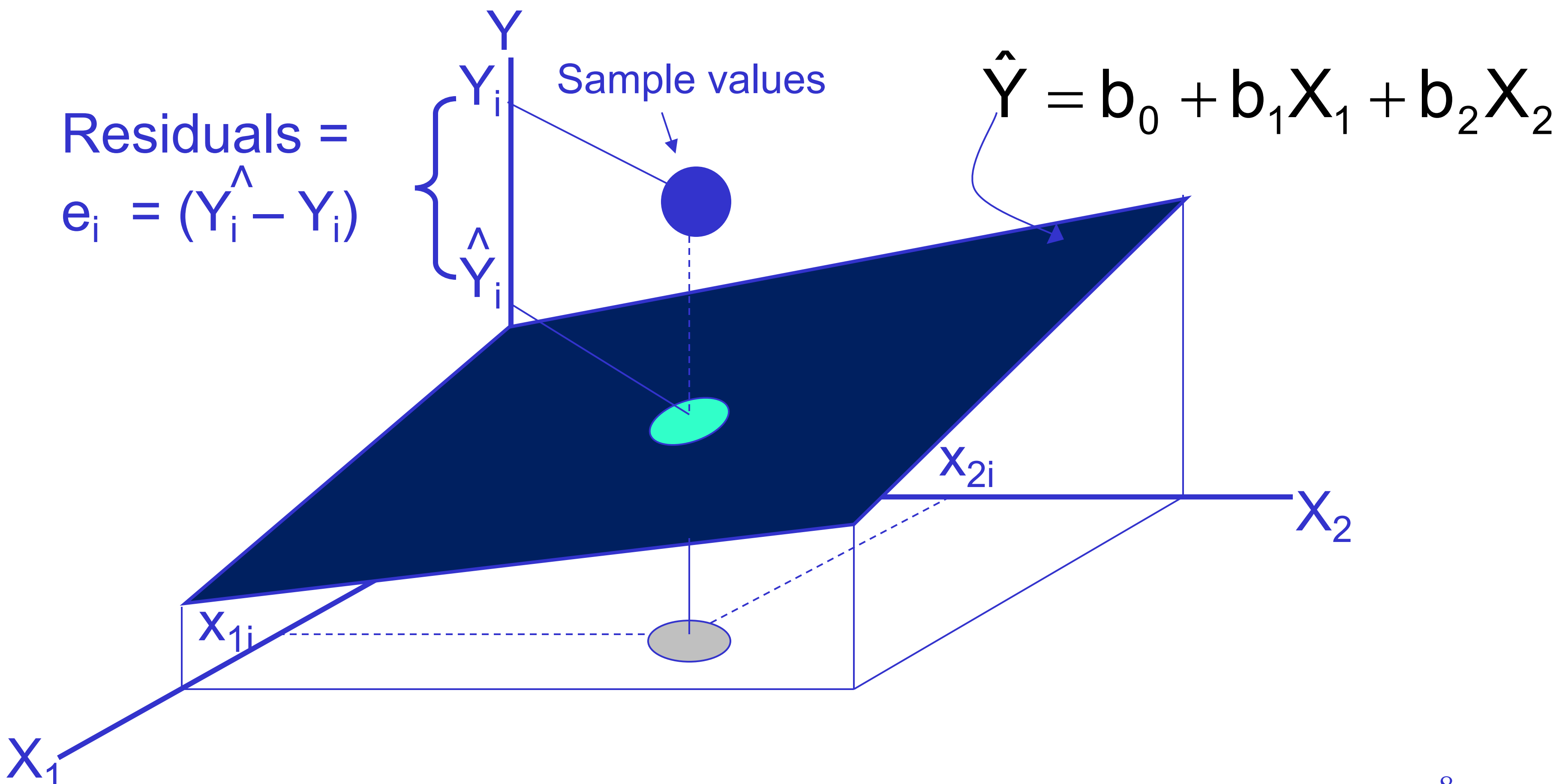
$$R^2 = \frac{SSR}{SST} = \frac{29460.0}{56493.3} = .52148$$

$$\overline{R}^2 = 0.44$$

52.1% of pie sale variability is explained by the variation of the product price and of advertising

44 % of pie sale variability is explained by the variation of the product price and of advertising considering the sample size and the number of regressors.

Residuals



Y
Y_i
Sample values
$\hat{Y} = b_0 + b_1X_1 + b_2X_2$
Residuals =
$e_i = (\hat{Y}_i - Y_i)$
$\hat{Y}_i$
x_{2i}
X_2
x_{1i}
X_1

ANOVA TEST

P Value: Probability of rejecting H0, when H0 is true

In general alpha = 0.05

H0 is rejected, then there exists a linear relationship with respect at least to one

$$F = \frac{MSR}{MSE} = \frac{14730.0}{2252.8} = 6.539$$

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	29460,027	2	14730,013	6,539	,012[a]
	Residual	27033,306	12	2252,776		
	Total	56493,333	14			

a. Predictors: (Constant), Advertising ('100$), Price ($)

b. Dependent Variable: Pie Sales

T- test

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	306,526	114,254		2,683	,020	57,588	555,464		
	Price ($)	-24,975	10,832	-,461	-2,306	,040	-48,576	-1,374	,999	1,001
	Advertising ('100$)	74,131	25,967	,570	2,855	,014	17,553	130,709	,999	1,001

a. Dependent Variable: Pie Sales

$$t = \frac{b_1}{S_{b_1}} = \frac{-24.975}{10.832} = -2.306$$

$$t = \frac{b_2}{S_{b_2}} = \frac{74.131}{25.967} = 2.855$$

Sig. is the probability of accepting H0, then, we affirm that both coefficients are significantly different from zero

Checking of linear regression model assumptions

- MULICOLLINEARITY → VIF and tollerance
- LINEARITY → scatter plot
- INDEPENDENCE → Durbin-Watson test
- NORMALITY → BOXPLOT, histogram, Q-Q plot, P-P plot, tests
- HOMOSKEDASTICITY → scatter-plot of residuals vs Xi

MULTICOLLINEARITY: VIF

Coefficients[a]

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
)	306,526	114,254		2,683	,020	57,588	555,464		
	-24,975	10,832	-,461	-2,306	,040	-48,576	-1,374	,999	1,001
ng ('100$)	74,131	25,967	,570	2,855	,014	17,553	130,709	,999	1,001

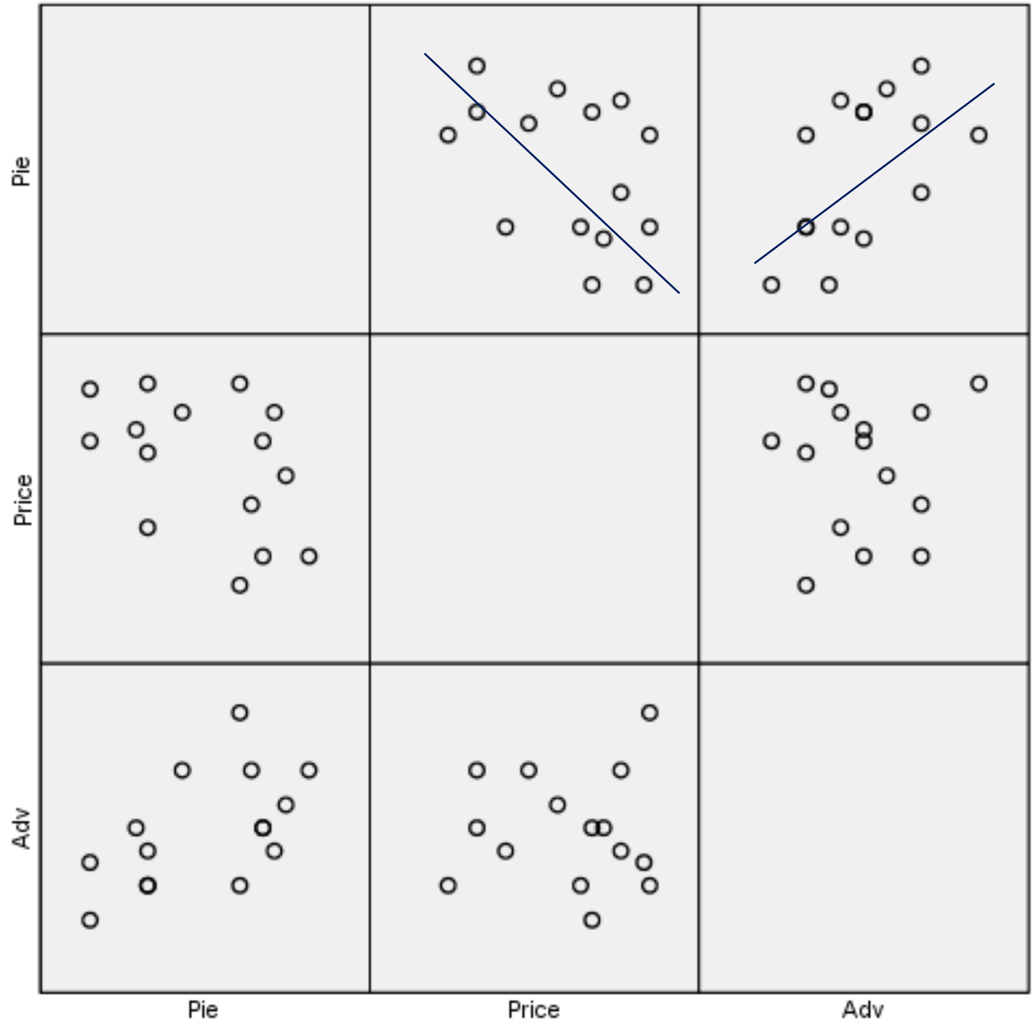
ariable: Pie Sales

$$\mathrm{VIF} = \frac{1}{1 - R_J^2} = 1.001$$

ABSENCE OF MULTICOLLINEARITY

We can include both variables in the model

LINEARITY: multiple scatter-plot



Linear relationship among Pie(Y) vs Price (X1) and Advertising (X2), as showed by the correlation coefficients.

Independence: DW test

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,722[a]	,521	,442	47,46341	1,683

a. Predictors: (Constant), X2, X1

b. Dependent Variable: Y

Coefficient Correlations[a]

Model			Advertising ('100$)	Price ($)
1	Correlations	Advertising ('100$)	1,000	-,030
		Price ($)	-,030	1,000
	Covariances	Advertising ('100$)	674,302	-8,562
		Price ($)	-8,562	117,335

a. Dependent Variable: Pie Sales

TEST DURBIN-WATSON

e_i	e_{i-1}	$(e_i-e_{i-1})^2$	e_i^2	
-63.80			4069.85	
96.15	-63.80	25584.06	9245.75	
20.88	96.15	5666.05	436.04	
-10.31	20.88	973.22	106.39	
-9.09	-10.31	1.50	82.60	
-35.74	-9.09	710.14	1277.12	
13.47	-35.74	2421.20	181.41	
49.03	13.47	1264.58	2403.92	
58.84	49.03	96.26	3462.27	
11.83	58.84	2210.47	139.84	
-46.16	11.83	3362.76	2131.11	
-46.44	-46.16	0.08	2156.86	
-15.70	-46.44	945.26	246.40	
8.89	-15.70	604.56	79.05	
-31.85	8.89	1660.16	1014.69	
	-31.85	45500.31	27033.31	1.68

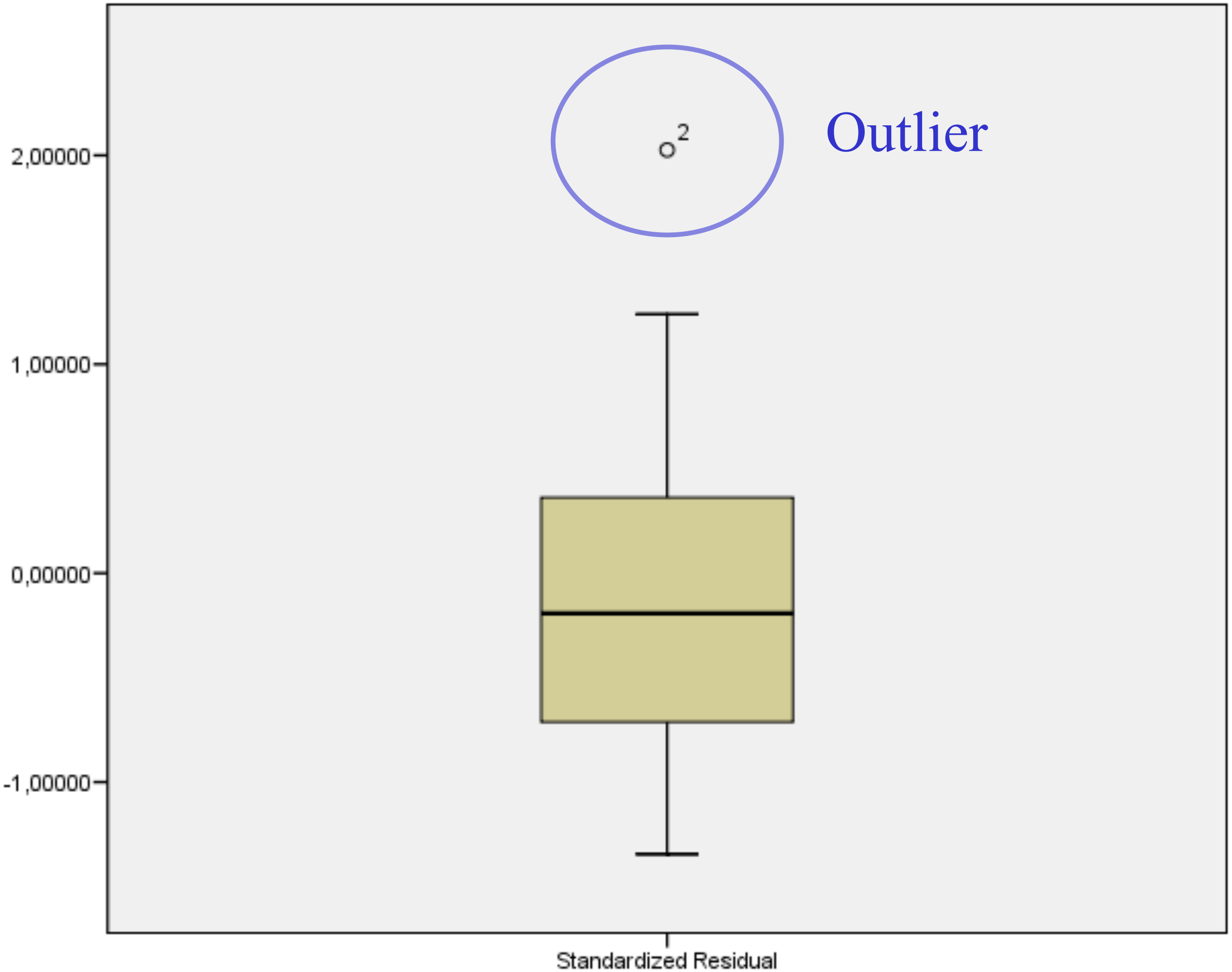
DW= 45500.31/27033.31=1.68

Tabulated values DL=0.945 e DU=1.543

Ho is true

The residuals are not correlated

NORMALITY: BOX PLOT



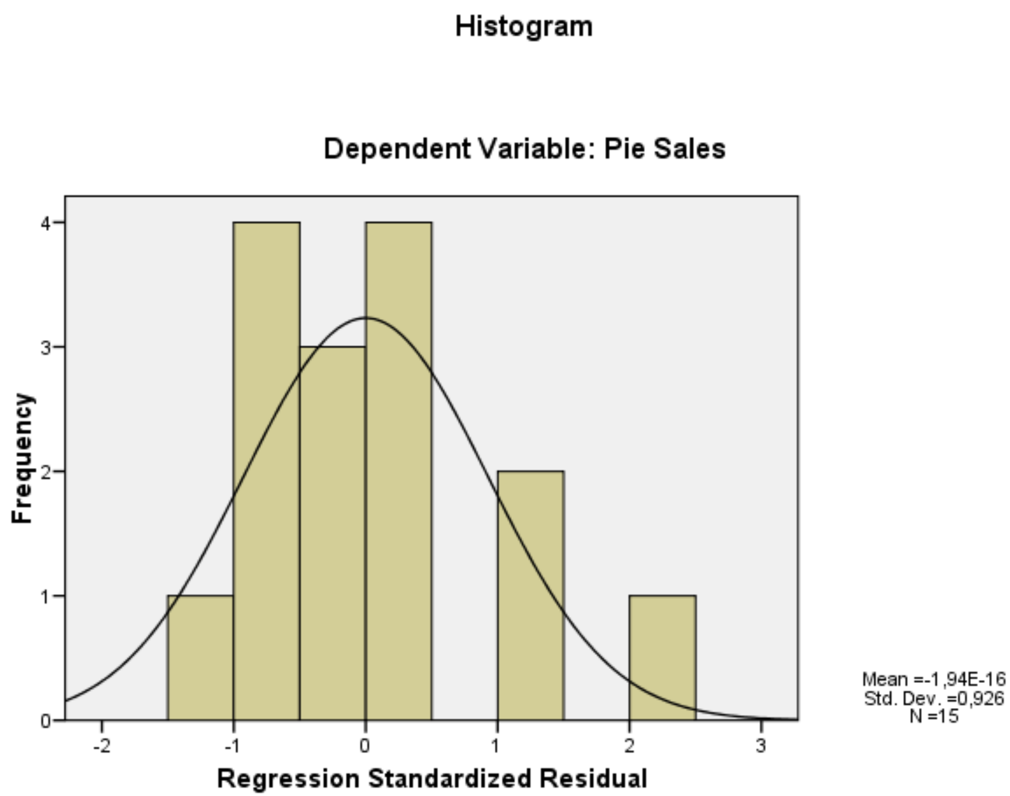
NORMALITY: histogram and tests

Tests of Normality

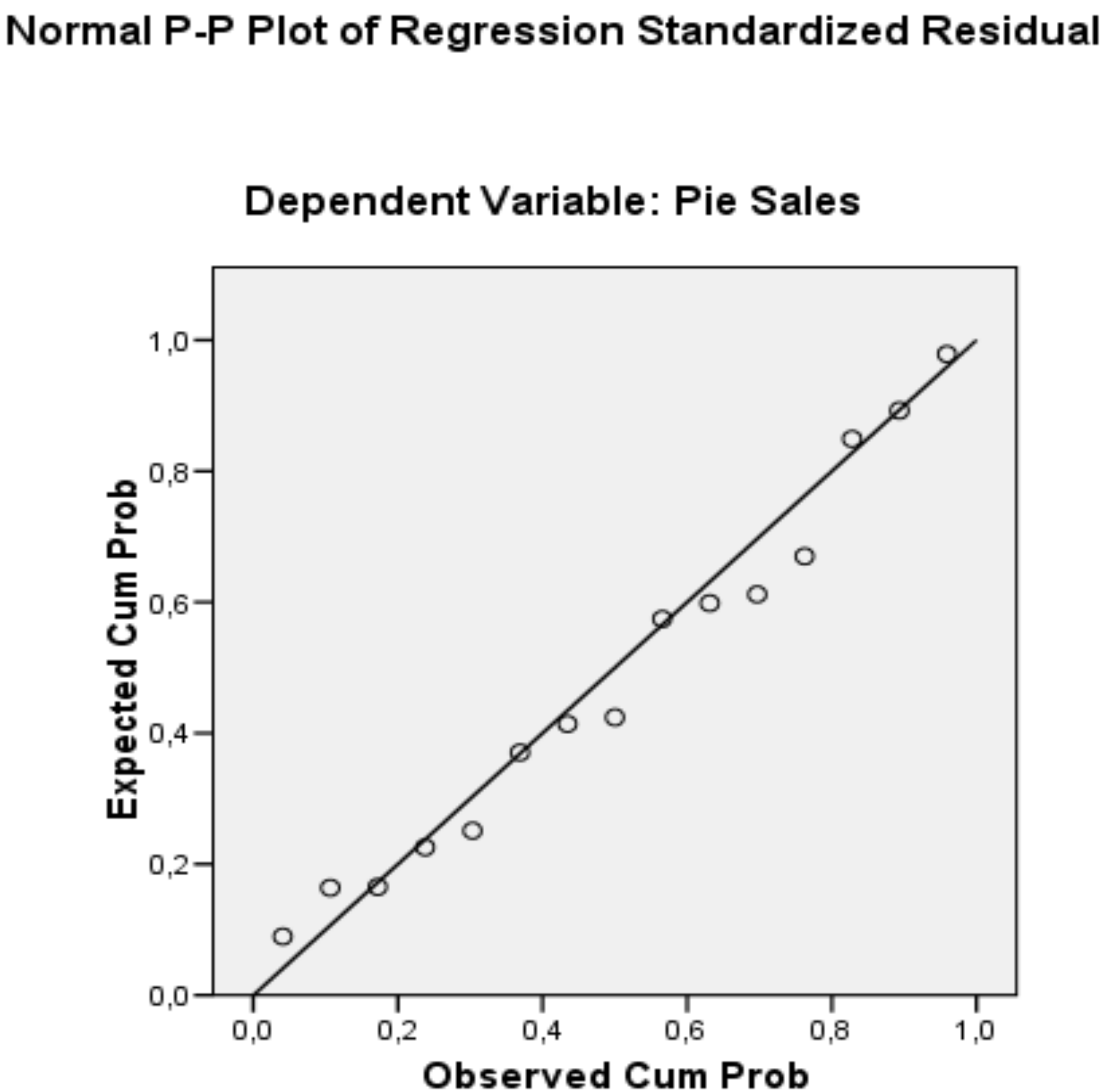
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual	,117	15	,200*	,958	15	,662

*. This is a lower bound of the true significance.

Residuals are normal!!!

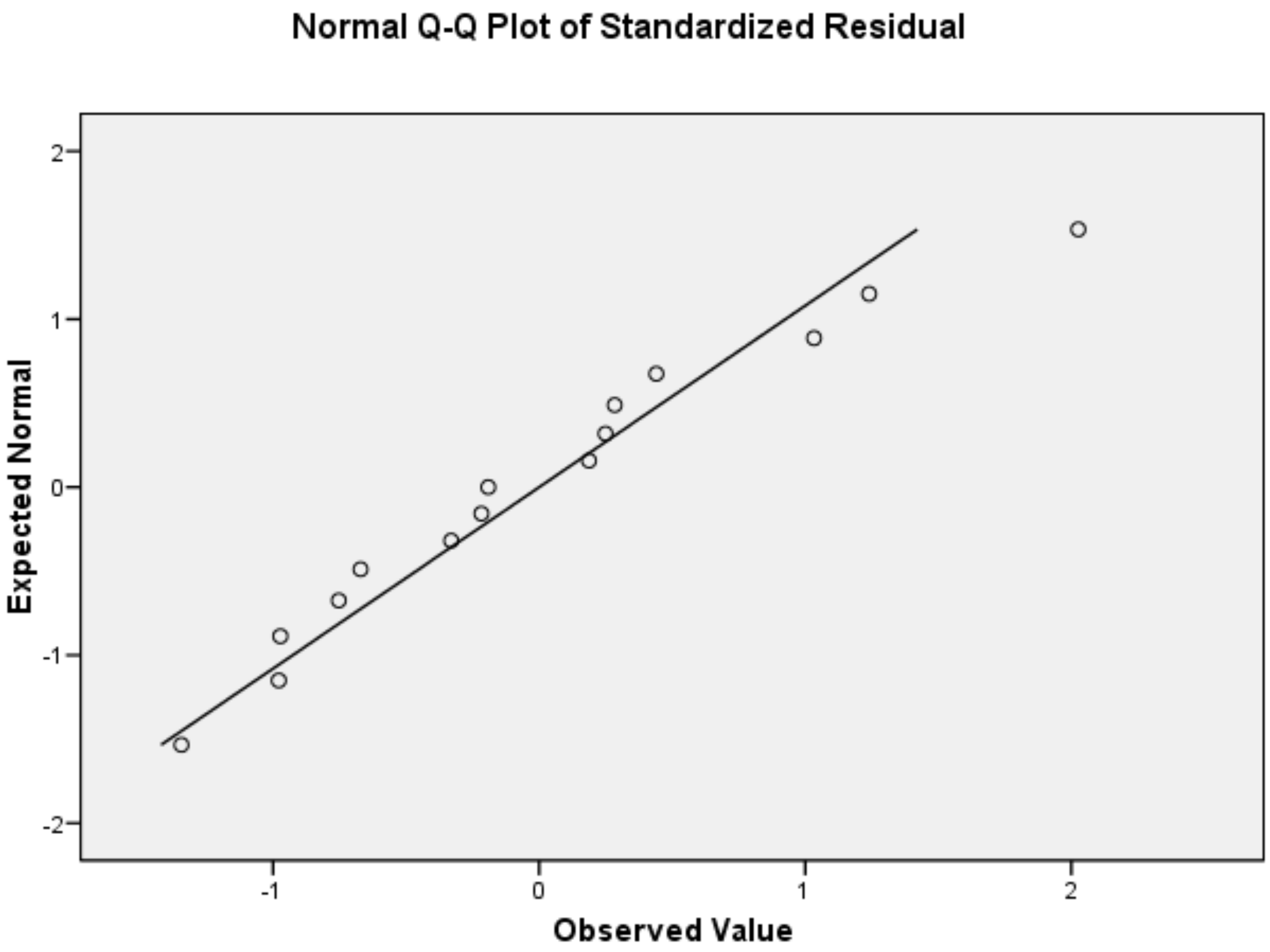


NORMALITY: PP - PLOT



ARE NORMAL!!!

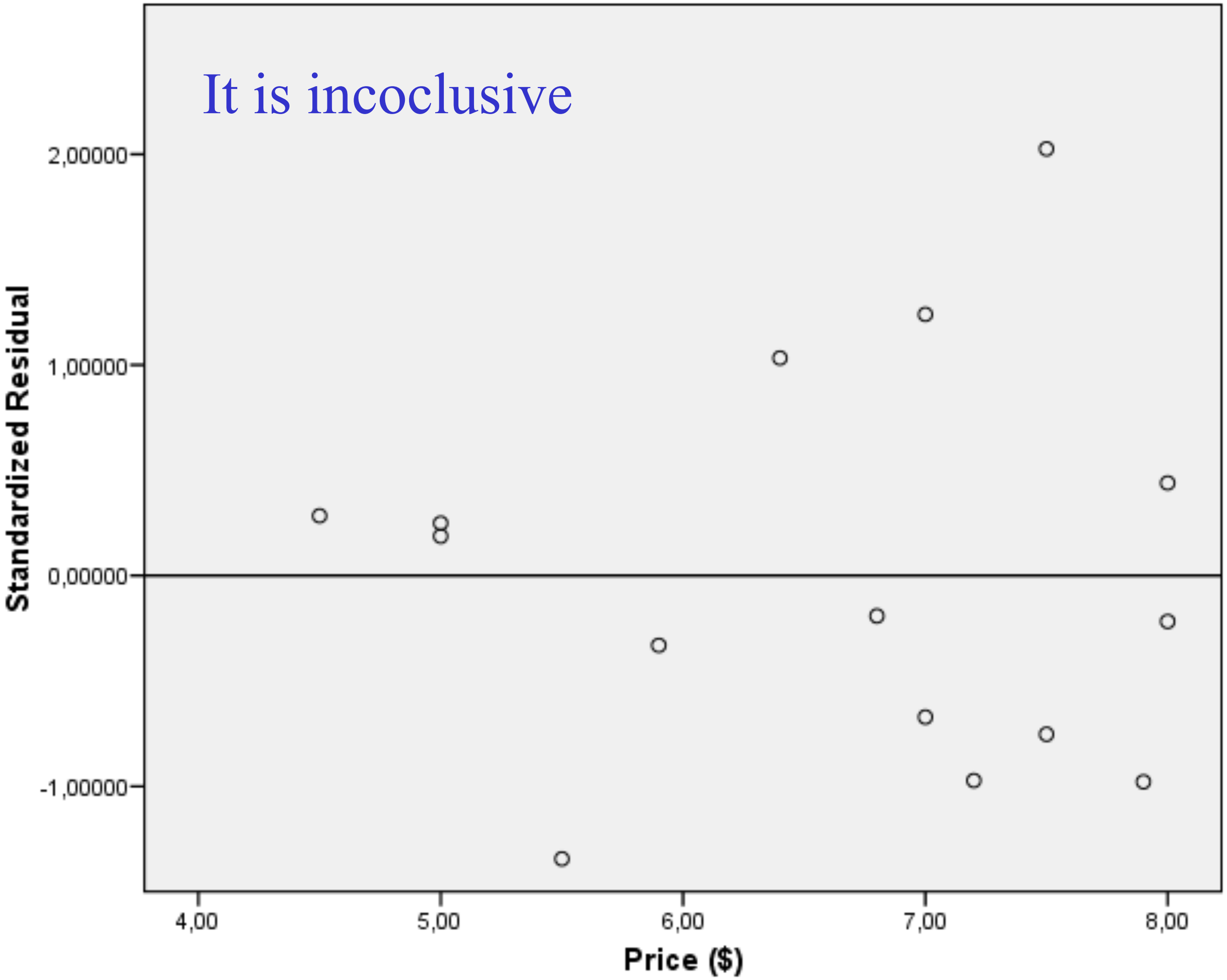
NORMALITY : Q-Q- PLOT



Homoskedasticity

- TO VERIFY HOMOSCHEDASTUCITY OF RESIDUALS WE REPRESENT THE SCATTER-PLOT OF RESIDUALS VS EACH EXPLANATORY VARIABLE
- IN THE CASE OF MULTIPLE REGRESSION WE CAN ALSO USE **Harvey-Godfrey LM test**

Homoskedasticity
scatter plot st. residuals vs price



Homoskedasticity: LM test

Harvey-Godfrey LM Test

- Residuals e_i
- Squared residuals e_i^2 and their logarithm
- Regression of this new variable and regressors Xi

$$\ln(e_i^2) = a_0 + a_1 X_1 + a_2 X_2 + ... + a_k X_k$$

LM TEST

$$\begin{cases} H_0 : \mathrm{a}_1 = \mathrm{a}_2 = \ldots. = \mathrm{a}_\mathrm{k} = 0 \\ H_1 : \mathrm{a}_1 \neq \mathrm{a}_2 \neq \ldots. \neq \mathrm{a}_\mathrm{k} \neq 0 \end{cases}$$

Homoskedasticity

Heteroskedasticity

Calculate $LM = nR^2$ and compare it with chi-square critical value

$$\boxed{LM \approx \chi^2_{k-1}}$$

$$LM > \chi^2_{\alpha,k-1} \qquad \text{REJECT } \mathrm{H}_0$$

where k= explanatory variables

LM TEST

e	e^2	ln e^2	X1	X2
-63.80	4069.85	8.31	5.5	3.3
96.15	9245.75	9.13	7.5	3.3
20.88	436.04	6.08	8	3
-10.31	106.39	4.67	8	4.5
-9.09	82.60	4.41	6.8	3
-35.74	1277.12	7.15	7.5	4
13.47	181.41	5.20	4.5	3
49.03	2403.92	7.78	6.4	3.7
58.84	3462.27	8.15	7	3.5
11.83	139.84	4.94	5	4
-46.16	2131.11	7.66	7.2	3.5
-46.44	2156.86	7.68	7.9	3.2
-15.70	246.40	5.51	5.9	4
8.89	79.05	4.37	5	3.5
-31.85	1014.69	6.92	7	2.7

LM TEST

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df 1	df 2	Sig. F
1	,422[a]	,178	,041	1,55816	,178	1,299	2	12	

a. Predictors: (Constant), Advertising ('100$), Price ($)

b. Dependent Variable: lne2

$$\boxed{\begin{aligned}LM&=nR^2\\&=15*0.178=2.67\end{aligned}}$$

$$\boxed{\chi^2_{\alpha,k-1}=\chi^2_{0.05,1}=3.94}$$

$$\boxed{LM<\chi^2}$$

We conclude for homoskedasticity, accepting H_0!