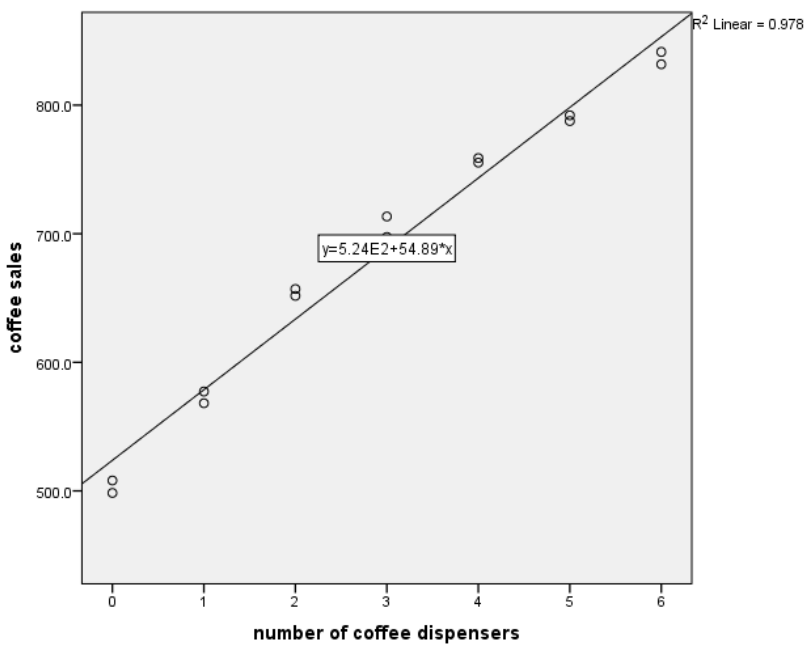
Kankanamge Harsha

PROJECT 3

A staff analyst for a cafeteria chain wishes to investigate the relation between the number of self-service coffee dispensers (X) in a cafeteria line and sales of coffee (Y). Fourteen cafeterias that are similar in such respects as volume of business, type of clientele and location are chosen for the experiment in **PROJ3-COFFEE SALES**.sav.

1. Estimate the linear regression line. Graph the line on the scatter plot.



Scatterplot shows strong positive linear association between "number of coffee dispensers" and "coffee sales". So there is a strong positive linear relationship between "number of coffee dispensers" and "coffee sales". When number of coffee dispenser increases coffee sales also increases strong positive linearly.

Descriptive Statistics

	Mean	Std. Deviation	N
coffee sales	688.479	115.1763	14
number of coffee dispensers	3.00	2.075	14

Average sales per cafeteria is $688.479. Sample size is 14.
Mean number of coffee dispensers per cafeteria is 3

Correlations

		coffee sales	number of coffee dispensers
Pearson Correlation	coffee sales	1.000	.989
	number of coffee dispensers	.989	1.000
Sig. (1-tailed)	coffee sales	.	.000
	number of coffee dispensers	.000	.
N	coffee sales	14	14
	number of coffee dispensers	14	14

Pearson Correlation is significant between “number of coffee dispensers” and “coffee sales”.

There is strong linear association between “number of coffee dispensers” and “coffee sales”.

Because Pearson correlation is 0.989. It is very closer to 1.

Model Summary

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.989[a]	.978	.977	17.5879	.978	545.495	1	12	.000

a. Predictors: (Constant), number of coffee dispensers

R Square = 0.978 » 98% of variability in “coffee sales” is explained by “number of coffee dispensers”

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	168740.643	1	168740.643	545.495	.000[b]
	Residual	3712.021	12	309.335		
	Total	172452.664	13			

a. Dependent Variable: coffee sales

b. Predictors: (Constant), number of coffee dispensers

SSR= 168740.643

SSE= 3712.021

S^2=309.335

F=545.495 F is significant.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	523.800	8.474		61.812	.000	505.337	542.263
	number of coffee dispensers	54.893	2.350	.989	23.356	.000	49.772	60.014

a. Dependent Variable: coffee sales

t = 23.356 t is significant. So there is an association between coffee sales and number of coffee dispensers.

y= $\boldsymbol{\beta 0 + \beta 1 x + \varepsilon}$

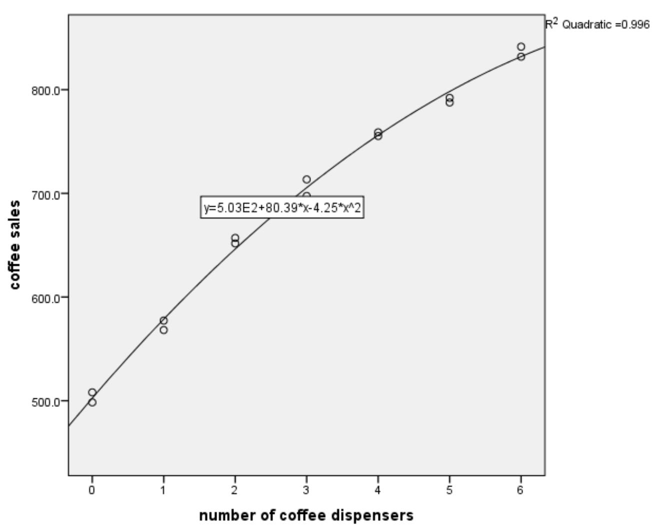
coffee sales^ (estimated)= b0 + b1x

b1=$\boldsymbol{\widehat{\beta}}$**1=54.893**

Linear regression line is:

coffee sales^ (estimated) = 54.893* number of dispensers + 523.800

2. Estimate the quadratic regression equation. Graph the curve on the scatter plot.



Scatterplot shows strong positive association between “coffee sales” and “number of coffee dispensers”.

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.998	.996	.995	7.858

The independent variable is number of coffee dispensers.

R Square = 0.996 » 99% of variability in “coffee sales” is explained by “number of coffee dispensers”

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	171773.443	2	85886.722	1390.939	.000
Residual	679.220	11	61.747		
Total	172452.664	13			

The independent variable is number of coffee dispensers.

SSR= 171773.443

SSE= 679.220

S^2=61.747

F=1390.939 F is significant.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
number of coffee dispensers	80.386	3.786	1.449	21.232	.000
number of coffee dispensers ** 2	-4.249	.606	-.478	-7.008	.000
(Constant)	502.556	4.850		103.619	.000

y= $\boldsymbol{\beta 0 + \beta 1 x + \beta 2 x^2 + \varepsilon}$

x = number of dispensers

coffee sales^ (estimated)= b0 + b1x +b2x^2

b1=$\boldsymbol{\hat{\beta}1}$**=80.386**

b2=$\boldsymbol{\hat{\beta}2}$**= -4.249**

Quadratic regression equation: coffee sales^ (estimated) = 80.386* x+ -4.249*x^2 + 502.556

3. Does the second order model provide significantly more predictive power than that provided by the straight line model at $\alpha = 0.05$? Explain.

(Show calculations)

H0: $\beta 2=0$

H1: $\beta 2\neq 0$

$R(\beta 2|\beta 1)$ =SSR(x,x^2) – SSR(x)

=171773.443- 168740.643 =3032.8

F= $R(\beta 2|\beta 1)/s^2$ = 3032.8/61.747 = 49.12

$F_{0.05,1,11}$ =4.84

F (49.12)>F table value (4.84)   Reject H0

The addition of the x^2 term to the linear model does significantly improve the prediction of coffee sales over and above that achieved by the linear model.

4. Which of the two models do you recommend and why?

x = number of dispensers

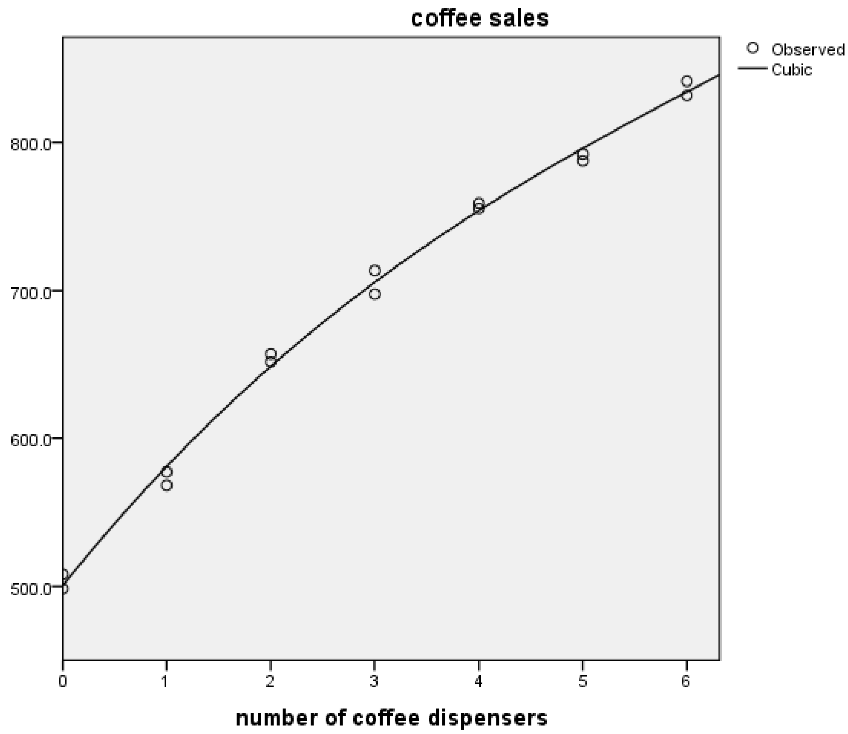
Quadratic model is the recommend model:

coffee sales^ (estimated) = 80.386* x - 4.249*x^2 + 502.556

	Linear	Quadratic	
R^2	0.978	0.996	
S^2	309.335	61.747	
F	545.495	1390.939	
ß	54.893	ß1=80.386	ß2=-4.249

Quadratic model has higher R^2 and lower S^2 (61.747 versus 309.335) compare with the linear model. Also F statistic has a higher value in the quadratic model than the linear model (F=1390.939 versus 545.495). The scatterplot with the fitted linen of the two model show improvement in the quadratic model over linear model. So the second order model (quadratic model) appear to be better than linear model.

5. Would the third order model be any better? Explain showing calculations and computer output.



Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.998	.996	.995	7.864

The independent variable is number of coffee dispensers.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	171834.193	3	57278.064	926.124	.000
Residual	618.470	10	61.847		
Total	172452.664	13			

The independent variable is number of coffee dispensers.

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
number of coffee dispensers	87.886	8.463	1.584	10.385	.000
number of coffee dispensers ** 2	-7.624	3.459	-.858	-2.204	.052
number of coffee dispensers ** 3	.375	.378	.253	.991	.345
(Constant)	500.306	5.359		93.365	.000

Third order model:

coffee sales^ (estimated) = 87.886* x - 7.624*x^2 +0.375*x^3 +500.306

H0: $\beta 3$=0

H1: $\beta 3 \neq 0$

R($\beta 3|\beta 1, \beta 2$) =SSR(x,x^2,x^3) – SSR(x,x^2) =171834.193 -171773.443=60.75

F= R($\beta 3|\beta 1, \beta 2$)/s^2 = 60.75/61.847= 0.98

F 0.05,1,10 =4.96

F (0.98) <F table value (4.96)   Do not reject H0

The addition of the x^3 term to the quadratic model does not significantly improve the prediction of coffee sales over and above that achieved by the quadratic model.

	cubic			Quadratic	
R^2	0.996			0.996	
S^2	61.847			61.747	
F	926.124			1390.939	
ß	87.886	-7.624	0.375	ß1=80.386	ß2=-4.249

The R2 and S2 values are very similar in the quadratic and cubic model. However, F statistic has a higher value in the quadratic model than the cubic model (F=1390.939 versus 926.124). The scatterplot with the fitted curve of the two model do not show any improvement in the cubic model over quadratic model. So the second order model (quadratic model) appear to be better than cubic model