Multiple Regression in SPSS **STAT 314**

I. The accompanying data is on y = profit margin of savings and loan companies in a given year, x_1 = net revenues in that year, and x_2 = number of savings and loan branches offices.

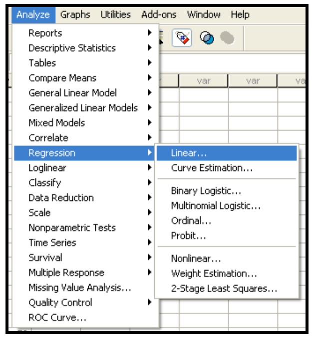
x_1	x_2	y	x_1	x_2	y	x_1	x_2	y
3.92	7298	0.75	3.42	6352	0.82	4.25	7546	0.72
3.61	6855	0.71	3.45	6361	0.75	4.41	7931	0.55
3.32	6636	0.66	3.58	6369	0.77	4.49	8097	0.63
3.07	6506	0.61	3.66	6546	0.78	4.70	8468	0.56
3.06	6450	0.70	3.78	6672	0.84	4.58	8717	0.41
3.11	6402	0.72	3.82	6890	0.79	4.69	8991	0.51
3.21	6368	0.77	3.97	7115	0.70	4.71	9179	0.47
3.26	6340	0.74	4.07	7327	0.68	4.78	9318	0.32
3.42	6349	0.90						

a. Determine the multiple regression equation for the data.
b. Compute and interpret the coefficient of multiple determination, R^2.
c. At the 5% significance level, determine if the model is useful for predicting the response.
d. Create scatterplots to check Assumption 1 as well as to identify potential outliers and potential influential observations.
e. Obtain the residuals and studentized residuals, and create residual plots. Decide whether or not it is reasonable to consider that the assumptions for multiple regression analysis are met by the variables in questions.
f. At the 5% significance level, does it appear that any of the predictor variables can be removed from the full model as unnecessary?
g. Obtain and interpret 95% confidence intervals for the slopes, β_i, of the population regression line that relates net revenues and number of branches to profit margin.
h. Are there any multicollinearity problems (i.e., are net revenues and number of branches collinear [estimating similar relationships/quantities])?
i. Obtain a point estimate for the mean profit margin with 3.5 net revenues and 6500 branches.
j. Test the alternative hypothesis that the mean profit margin with 3.5 net revenues and 6500 branches is greater than 0.70. Test at the 5% significance level.
k. Determine a 95% confidence interval for the mean profit margin with 3.5 net revenues and 6500 branches.
l. Find the predicted profit margin for my bank with 3.5 net revenues and 6500 branches.
m. Determine a 95% prediction interval for the profit margin for Dr. Street's bank with 3.5 net revenues and 6500 branches.

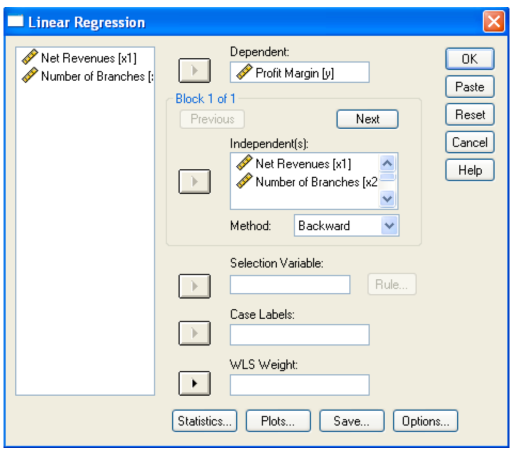
1. Enter the values of the three variables into SPSS.
2. Since we want to predict the profit margin for a bank with 3.5 net revenues and 6500 branches, enter the number "3.5" in the "x_1" variable column and "6500" in the "x_2" variable column of the data window after the last row. Enter a "." for the corresponding "y" variable value (this lets SPSS know that we want a prediction for this value and not to include the value in any other computations). *(See figure, below.)*

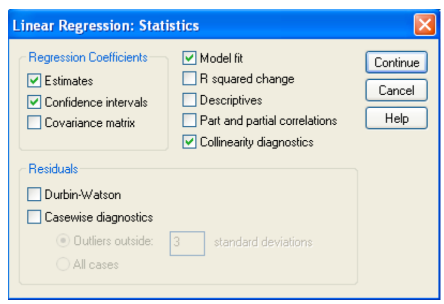
	x1	x2	y
24	4.71	9179	.47
25	4.78	9318	.32
26	3.50	6500	.

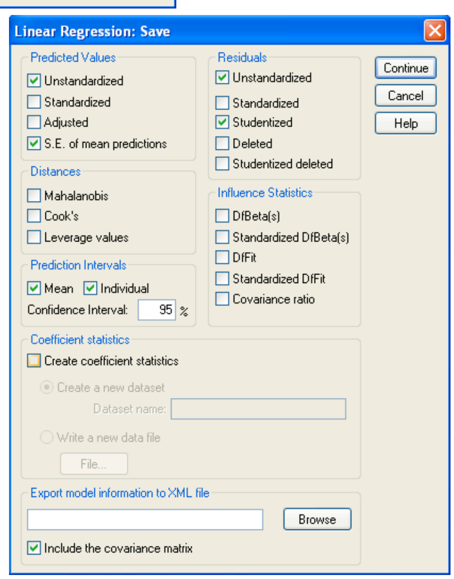
3. Select Analyze → Regression → Linear… *(see figure, below)*.

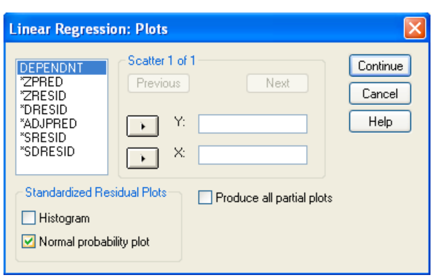


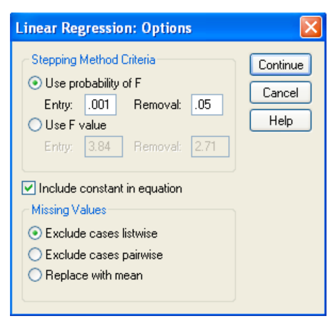
4. Select "Profit Margin" as the dependent variable and "Net Revenues" and "Number of Branches" as the independent variables, and select the "Backward" method. Click "Statistics", select "Estimates" and "Confidence intervals" for the regression coefficients, select "Model fit", select "Collinearity diagnostics", and click "Continue". Click "Plots…", select "Normal Probability Plot", and click "Continue". Click "Save…", select "Unstandardized" predicted values and "S.E. of mean predictions" ($s_{\hat{y}}$), select "Unstandardized" and "Studentized" residuals, select "Mean" and "Individual" prediction intervals at the 95% level (or whatever level the problem requires), and click "Continue". Click "Options…", enter .001 in the Entry box, enter the significance value (alpha) in the Removal box, and click "Continue". Click "OK". *(See the following figures.)*











The output from this procedure is extensive and will be shown in parts in the following answers.

a. Determine the multiple regression equation for the data.

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)	1.5645	.0794		19.7050	.0000	1.3998	1.7292		
	Net Revenues	.2372	.0556	.9872	4.2693	.0003	.1220	.3524	.1145	8.7321
	Number of Branches	-.0002491	.0000	-1.7971	-7.7719	.0000	-.0003155	-.0001826	.1145	8.7321

a. Dependent Variable: Profit Margin

From the above output, the regression equation is: $\hat{y} = 1.5645 + 0.2372x_1 - 0.0002491x_2$.

b. Compute and interpret the coefficient of multiple determination, R^2.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.9302[a]	.8653	.8531	.0533

a. Predictors: (Constant), Number of Branches, Net Revenues

b. Dependent Variable: Profit Margin

The coefficient of multiple determination is 0.8653; therefore, about 86.53% of the variation in the profit margin is explained by net revenues and number of branches for the savings and loan banks. The regression equation appears to be very useful for making predictions since the value of R^2 is close to 1.

c. At the 5% significance level, determine if the model is useful for predicting the response.

ANOVA[b]

Model		df	Sum of Squares	Mean Square	F	Sig.
1	Regression	2	.4015	.2008	70.6606	.0000[a]
	Residual	22	.0625	.0028		
	Total	24	.4640			

a. Predictors: (Constant), Number of Branches, Net Revenues

b. Dependent Variable: Profit Margin

Step 1: **Hypotheses**

$H_0: \ \beta_1 = \beta_2 = 0$

$H_a:$ at least one $\beta_i \neq 0$

Step 2: **Significance Level**

$\alpha = 0.05$

Step 3: **Rejection Region**

Reject the null hypothesis if p-value ≤ 0.05.

Step 4: **ANOVA Table (Test Statistic and p-value)**

(see above) $F = 70.6606$, p-value < 0.0001

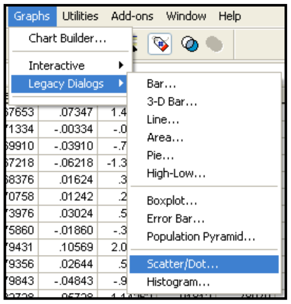
Step 5: **Conclusion**

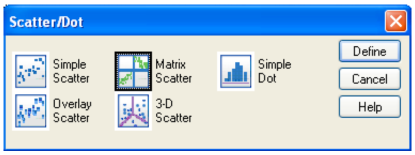
Since p-value $< 0.001 \leq 0.05$, we shall reject the null hypothesis.

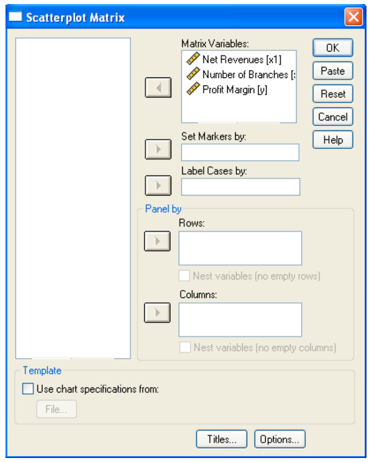
Step 6: **State conclusion in words**

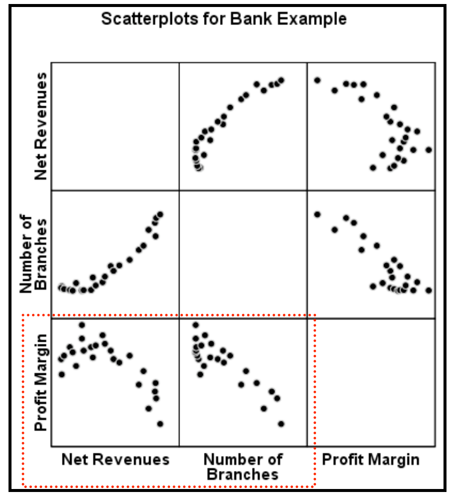
At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that at least one of the predictors is useful for predicting profit margin; therefore the model us useful.

d. Create scatterplots to check Assumption 1 as well as to identify potential outliers and potential influential observations.





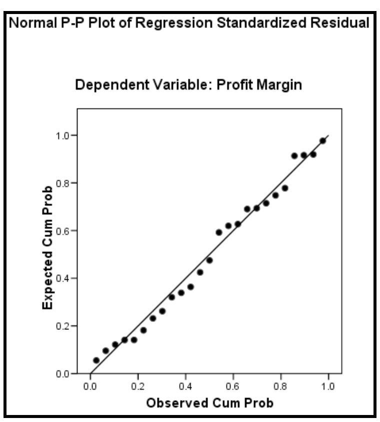


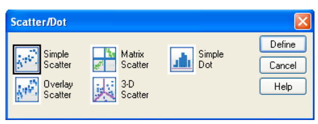


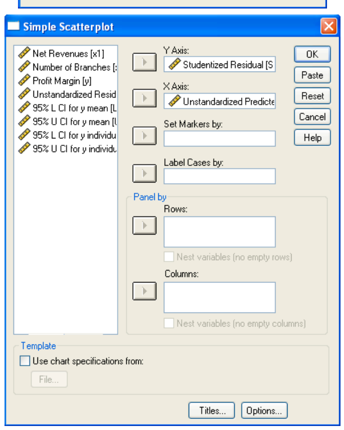
Profit Margin appears to be linearly related to each of the predictor variables with no visible potential outliers or influential observations (no points away from the main cluster of points); thus, Assumption 1 appears to be satisfied.

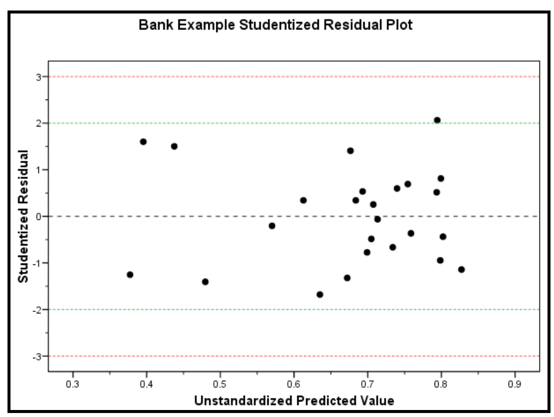
e. Obtain the residuals and studentized residuals, and create residual plots (normal probability plot of residuals and scatterplots for the predicted values versus the residuals and/or studentized residuals). Decide whether or not it is reasonable to consider that the assumptions for multiple regression analysis are met by the variables in questions.

The residuals [`res`] and standardized values [`sre`] (as well as the predicted values [`pre`], the standard errors of each prediction [`sep`], the prediction interval endpoints [`lici` & `uici`], and the confidence interval endpoints [`lmci` & `umci`]) can be found in the data window.









The normal plot of the residuals shows the points close to a diagonal line; thus, Assumption 2 is satisfied. The studentized residual plot shows a random scatter of points with constant variability and no definite outliers (although, there is one very slight potential outlier); thus, Assumption 3 is met.

f. At the 5% significance level, does it appear that any of the predictor variables can be removed from the full model as unnecessary?

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)	1.5645	.0794		19.7050	.0000	1.3998	1.7292		
	Net Revenues	.2372	.0556	.9872	4.2693	.0003	.1220	.3524	.1145	8.7321
	Number of Branches	-.0002491	.0000	-1.7971	-7.7719	.0000	-.0003155	-.0001826	.1145	8.7321

a. Dependent Variable: Profit Margin

<u>Step 1</u>: **Hypotheses**

$H_0: \beta_1 = 0$ (net revenue is not useful for predicting profit margin)

$H_a: \beta_1 \neq 0$ (net revenue is useful for predicting profit margin)

assuming that number of branches is included in the model

<u>Step 2</u>: **Significance Level**

$\alpha = 0.05$

<u>Step 3</u>: **Rejection Region**

Reject the null hypothesis if p-value ≤ 0.05.

<u>Step 4</u>: **Test Statistic and *p*-value**

(see above) $T = 4.2693$, p-value = 0.0003

<u>Step 5</u>: **Conclusion**

Since p-value = 0.0003 $\leq$ 0.05, we shall reject the null hypothesis.

<u>Step 6</u>: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that the slope of the net revenue variable is not zero and, hence, that net revenues is useful (with number of branches) as a predictor of profit margin for savings and loan banks.

<u>Step 1</u>: **Hypotheses**

$H_0: \beta_2 = 0$ (number of branches is not useful for predicting profit margin)

$H_a: \beta_2 \neq 0$ (number of branches is useful for predicting profit margin)

assuming that net revenue is included in the model

<u>Step 2</u>: **Significance Level**

$\alpha = 0.05$

<u>Step 3</u>: **Rejection Region**

Reject the null hypothesis if p-value ≤ 0.05.

<u>Step 4</u>: **Test Statistic and *p*-value**

(see above) $T = -7.7719$, p-value < 0.0001

<u>Step 5</u>: **Conclusion**

Since p-value $< 0.0001 \leq 0.05$, we shall reject the null hypothesis.

<u>Step 6</u>: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that the slope of the number of branches variable is not zero and, hence, that number of branches is useful (with net revenues) as a predictor of profit margin for savings and loan banks.

g. Obtain and interpret 95% confidence intervals for the slopes, β_i, of the population regression line that relates net revenues and number of branches to profit margin.

We are 95% confident that the slope for net revenues is somewhere between 0.1220 and 0.3524. In other words, we are 95% confident that for every single-unit increase in net revenue, the average profit margin increases between 0.1220 and 0.3524.

We are 95% confident that the slope for number of branches is somewhere between –0.0003155 and –0.0001826. In other words, we are 95% confident that for every additional branch, the average profit margin decreases between 0.0003155 and 0.0001826.

h. **Are there any multicollinearity problems (i.e., are net revenues and number of branches collinear [estimating similar relationships/quantities])?**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	1.5645	.0794		19.7050	.0000	1.3998	1.7292		
	Net Revenues	.2372	.0556	.9872	4.2693	.0003	.1220	.3524	.1145	8.7321
	Number of Branches	-.0002491	.0000	-1.7971	-7.7719	.0000	-.0003155	-.0001826	.1145	8.7321

a. Dependent Variable: Profit Margin

Since neither of the predictor variables has a variance inflation factor (VIF) greater than ten (both VIFs are 8.7321), there are no apparent multicollinearity problems; in other words, there is no variable in the model that is measuring the same relationship/quantity as is measured by another variable or group of variables.

The remaining parts will be completed using output from the Data window.

	x1	x2	y	pre_1	res_1	sre_1	sep_1	lmci_1	umci_1	lici_1	uici_1
26	3.50	6500	.	.77567	.	.	.01365	.74736	.80398	.66156	.88978

i. **Obtain a point estimate for the mean profit margin with 3.5 net revenues and 6500 branches.**

The point estimate (`pre_1`) is 0.77567.

j. **Test the alternative hypothesis that the mean profit margin with 3.5 net revenues and 6500 branches is greater than 0.70. Test at the 5% significance level.**

Step 1: **Hypotheses**

$$H_0: \ \hat{y} = 0.70 \quad \left(when \ \ x_1 = 3.5 \ \& \ \ x_2 = 6500\right)$$

$$H_a: \ \hat{y} > 0.70 \quad \left(when \ \ x_1 = 3.5 \ \& \ \ x_2 = 6500\right)$$

Step 2: **Significance Level**

$\alpha = 0.05$

Step 3: **Critical Value(s) and Rejection Region(s)**

Critical Value: $t_{\alpha, df=n-(k+1)} = t_{0.05, df=22} = t_{90\% CI, df=22} = 1.72$

Reject the null hypothesis if $T \geq 1.72$ (or if p-value ≤ 0.05).

Step 4: **Test Statistic**

$$T = \frac{\hat{y} - y_0}{s_{\hat{y}}} = \frac{\texttt{pre\_1} - y_0}{\texttt{sep\_1}} = \frac{0.77567 - 0.70}{0.01365} = 5.5436$$

p-value < 0.001

Step 5: **Conclusion**

Since $5.5436 \geq 1.72$ (p-value $< 0.001 \leq 0.05$), we shall reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that the mean profit margin with 3.5 net revenues and 6500 branches is greater than 0.70.

k. **Determine a 95% confidence interval for the mean profit margin with 3.5 net revenues and 6500 branches.**

We are 95% confident that the mean profit margin with 3.5 net revenues and 6500 branches is somewhere between 0.74736 (`lmci_1`) and 0.80398 (`umci_1`).

l. **Find the predicted profit margin for my bank with 3.5 net revenues and 6500 branches.**

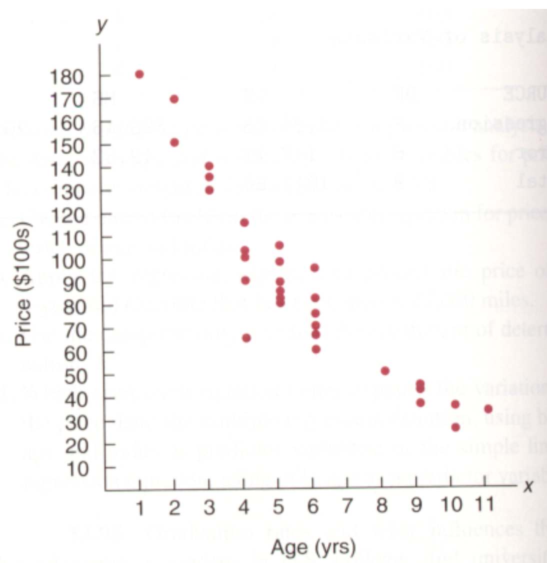
The predicted profit margin for my (individual) bank is 0.77567.

m. **Determine a 95% prediction interval for the profit margin for Dr. Street's bank with 3.5 net revenues and 6500 branches.**

We are 95% certain that the profit margin for Dr. Street's bank with 3.5 net revenues and 6500 branches will be somewhere between 0.66156 (`lici_1`) and 0.88978 (`uici_1`).

II. Although at first glance, the relationship between age and price of Nissan Zs appears to be linear in the age range from 2 to 7 years, it is definitely not so in the age range from 2 to 11 years (see graph, below). From *Auto Trader* we obtained the data on age and price for a sample of 31 Nissan Zs shown in the table below. (Ages are in years, prices are in hundreds of dollars.) Below the table is a scatterplot of the data.

Age x	Price y	Age x	Price y	Age x	Price y	Age x	Price y
5	85	4	103	10	25	3	135
6	70	4	100	5	82	9	44
4	90	6	75	10	35	9	36
2	150	3	140	5	89	11	33
5	98	6	66	6	95	1	180
6	95	2	169	4	65	5	80
3	129	6	60	6	82	5	105
4	115	8	50	9	42		



As you can see from the scatterplot, the data points are not clustered about a straight line but instead follow a curve. This means we should not determine a regression line but instead should try to fit a curve to the data. From the curvature of the scatter diagram, it appears that a parabola might be an appropriate curve to fit to the data. To fit a parabola to the data, we need a regression equation of the form

$$\hat{y} = a + b_1 x + b_2 x^2.$$

If we let $x_1 = x$ and $x_2 = x^2$, then the above equation becomes

$$\hat{y} = a + b_1 x + b_2 x_2,$$

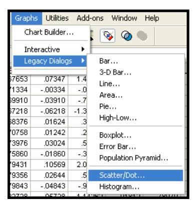
which is a multiple regression equation with two predictor variables, *age* and *age*^{2} (square of the age).

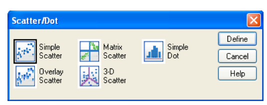
a. Create a scatterplot to check Assumption 1 as well as to identify outliers and potential influential observations.
b. Determine the quadratic regression equation for the data.
c. Compute and interpret the coefficient of multiple determination, R^2.
d. At the 10% significance level, determine if the model is useful for predicting the response.
e. Obtain the residuals and studentized residuals, and create residual plots. Decide whether or not it is reasonable to consider that the assumptions for quadratic regression analysis are met by the variables in questions.
f. At the 10% significance level, does it appear that the quadratic term can be removed from the full model as unnecessary (thus reducing the model to a linear model)?
g. Obtain a point estimate for the mean sales price of 8-year-old Nissan Zs.
h. Test the alternative hypothesis that the mean sales price of 8-year-old Nissan Zs is greater than \$4000 ($y = 40$ hundred dollars).
i. Determine a 90% confidence interval for the mean sales price of 8-year-old Nissan Zs.
j. Find the predicted sales price of an 8-year-old Nissan Z for sale by Sandy Beaches.
k. Determine a 90% prediction interval for the sales price of Sandy's 8-year-old Nissan Z.

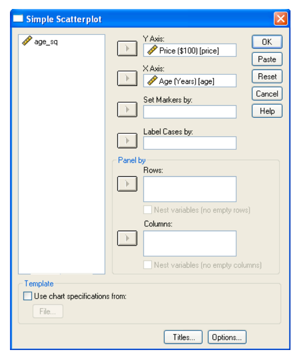
1. Enter the values of the two variables into SPSS and create a scatterplot. *(See figures, below.)*

2. Since we want to predict the price of 8-year-old Nissan Zs, enter the number "8" in the "age" variable column of the data window after the last row. Enter a "." for the corresponding "price" variable value (this lets SPSS know that we want a prediction for this value and not to include the value in any other computations). *(See left figure, below.)*

	price	age
26	44	9
27	36	9
28	33	11
29	180	1
30	80	5
31	105	5
32	.	8

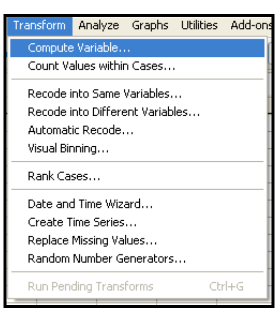


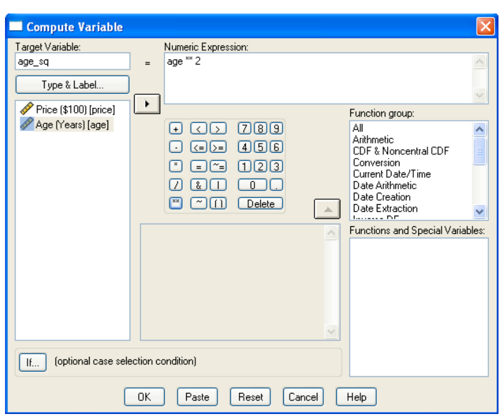




3. Create a variable for the age^2 values.

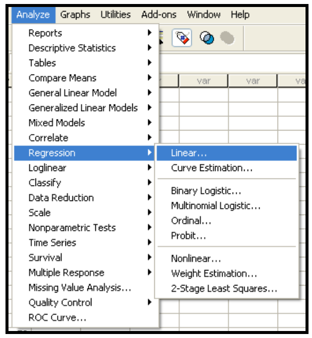
Select Transform → Compute… *(see left figure, below)*. For "Target Variable" enter `age_sq`. For "Numeric Expression" enter `age ** 2`. Now press "OK" to create the new variable.



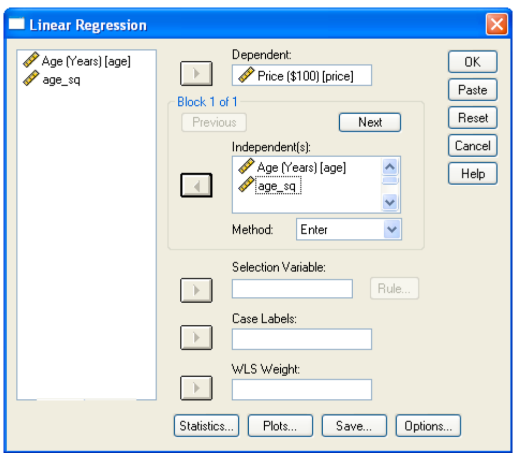


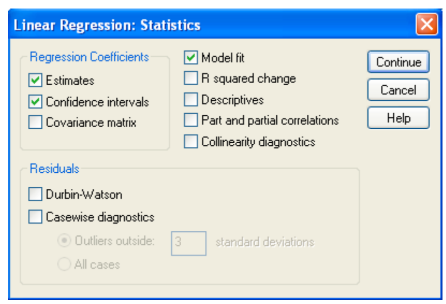
	price	age	age_sq
26	44	9	81
27	36	9	81
28	33	11	121
29	180	1	1
30	80	5	25
31	105	5	25
32	.	8	64

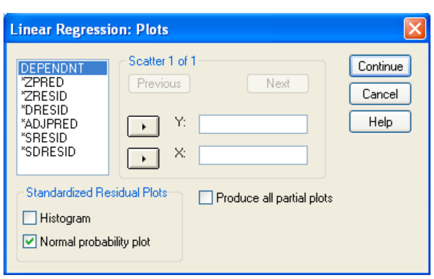
4. Select Analyze → Regression → Linear… *(see figure, below)*.

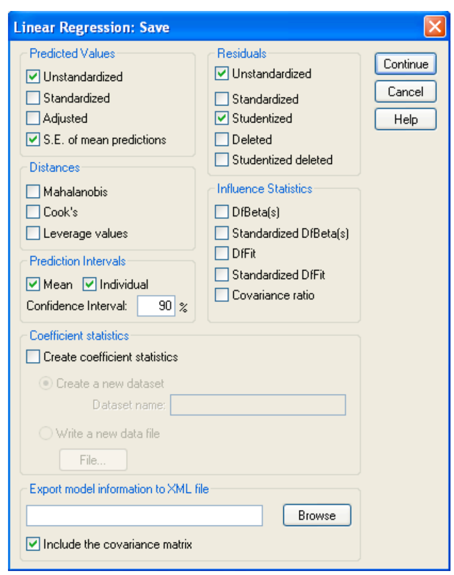


5. Select "Price" as the dependent variable and "Age" and "Age-sq" as the independent variables, and select the "Enter" method. Click "Statistics", select "Estimates" and "Confidence intervals" for the regression coefficients, select "Model fit", and click "Continue". Click "Plots…", select "Normal Probability Plot", and click "Continue". Click "Save…", select "Unstandardized" predicted values and "S.E. of mean predictions" ($s_{\hat{y}}$), select "Unstandardized" and "Studentized" residuals, select "Mean" and "Individual" prediction intervals at the 90% level (or whatever level the problem requires), and click "Continue". Click "OK". *(See the following figures.)*



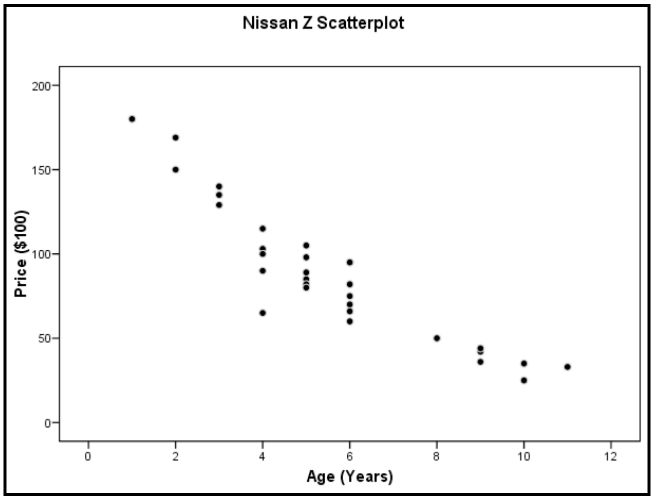






The output from this procedure is extensive and will be shown in parts in the following answers.

a. Create a scatterplot to check Assumption 1 as well as to identify outliers and potential influential observations.



The graph appears to have a curvilinear pattern that appears to be a simple curve; thus, a quadratic regression model will be fit to the data (Assumption 1 is met). There appear to be no visible potential outliers or influential observations (no points lie away from the main cluster of points).

b. Determine the quadratic regression equation for the data.

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
1	(Constant)	209.4405	11.4842		18.2372	.0000	185.9161	232.9649
	Age (Years)	-30.7759	4.0563	-1.9552	-7.5873	.0000	-39.0848	-22.4671
	age_sq	1.3297	.3219	1.0644	4.1305	.0003	.6703	1.9891

a. Dependent Variable: Price ($100)

The quadratic regression equation is: $\hat{y} = 209.4405 - 30.7759x + 1.3297x^2$.

c. Compute and interpret the coefficient of multiple determination, R^2.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.9504[a]	.9032	.8963	12.8102

a. Predictors: (Constant), age_sq, Age (Years)

b. Dependent Variable: Price ($100)

The coefficient of multiple determination is 0.9032; therefore, about 90.32% of the variation in the price of Nissan Zs is explained by its quadratic relationship with the age of the car. The regression equation appears to be very useful for making predictions since the value of R^2 is close to 1.

d. At the 10% significance level, determine if the model is useful for predicting the response.

ANOVA[b]

Model		df	Sum of Squares	Mean Square	F	Sig.
1	Regression	2	42895.36	21447.6788	130.6978	.0000[a]
	Residual	28	4594.8359	164.1013		
	Total	30	47490.19			

a. Predictors: (Constant), age_sq, Age (Years)

b. Dependent Variable: Price ($100)

Step 1: **Hypotheses**
$H_0: \ \beta_1 = \beta_2 = 0$
$H_a: \ \text{at least one } \beta_i \neq 0$

Step 2: **Significance Level**
$\alpha = 0.10$

Step 3: **Rejection Region**
Reject the null hypothesis if p-value ≤ 0.10.

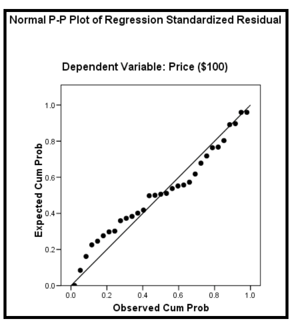
Step 4: **ANOVA Table (Test Statistic and p-value)**
(see above) $F = 130.6978$, p-value < 0.0001

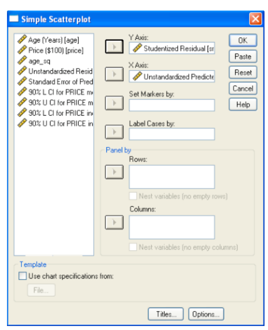
Step 5: **Conclusion**
Since p-value $< 0.0001 \leq 0.10$, we shall reject the null hypothesis.

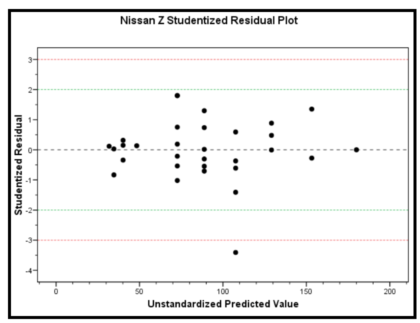
Step 6: **State conclusion in words**
At the α = 0.10 level of significance, there exists enough evidence to conclude that at least one of the terms of the quadratic model is useful for predicting the price of Nissan Zs; therefore the model us useful.

e. Obtain the residuals and studentized residuals, and create residual plots. Decide whether or not it is reasonable to consider that the assumptions for quadratic regression analysis are met by the variables in questions.

The residuals [`res`] and standardized values [`sre`] (as well as the predicted values [`pre`], the standard errors of each prediction [`sep`], the prediction interval endpoints [`lici` & `uici`], and the confidence interval endpoints [`lmci` & `umci`]) can be found in the data window.







The normal plot of the residuals shows the points close to a diagonal line; thus, Assumption 2 is satisfied. Each of the studentized residual plots shows a random scatter of points with constant variability; thus, Assumption 3 is met.

Note that observation #22 has an unusual residual in each of the studentized residual plots (the residual is more that three standard deviations from the mean residual of 0). This indicates that observation #22 is likely to be an outlier.

Also, at first glance one might think that the variability is less for the right half of the plots when compared to the left half. This is likely not the case, and any apparent decrease in variability is probably due to the fact that there are far fewer observations in the right half (having fewer values leaves less room for variability).

f. At the 10% significance level, does it appear that the quadratic term can be removed from the full model as unnecessary (thus reducing the model to a linear model)?

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	209.4405	11.4842		18.2372	.0000	185.9161	232.9649
	Age (Years)	-30.7759	4.0563	-1.9552	-7.5873	.0000	-39.0848	-22.4671
	age_sq	1.3297	.3219	1.0644	4.1305	.0003	.6703	1.9891

a. Dependent Variable: Price ($100)

Step 1: **Hypotheses**

$H_0 : \beta_2 = 0$ (quadratic term is not useful for predicting price)

$H_a : \beta_2 \neq 0$ (quadratic term is useful for predicting price)

assuming that the linear term is included in the model

Step 2: **Significance Level**

$\alpha = 0.10$

Step 3: **Rejection Region**

Reject the null hypothesis if *p*-value ≤ 0.10.

Step 4: **Test Statistic and *p*-value**

(see above) $T = 4.1305$, *p*-value = 0.0003

Step 5: **Conclusion**

Since *p*-value = 0.0003 ≤ 0.10, we shall reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.10$ level of significance, there exists enough evidence to conclude that the slope of the quadratic term is not zero and, hence, that the quadratic term is useful (when included with the linear term) as a predictor of price for Nissan Zs.

Since we rejected the null hypothesis, we should construct a 90% confidence interval for β_2. Note that we cannot compute this directly in SPSS—SPSS provides only 95% confidence intervals for the slopes; however, it does provide values in the "Unstandardized Coefficients" columns that can be used in with our manual calculations.

$$b_2 \pm t_{\alpha/2, df=n-k-1}\left(s_{b_2}\right)$$

$$b_2 \pm t_{0.05, df=28}\left(s_{b_2}\right)$$

$$b_2 \pm t_{90\%CI, df=28}\left(s_{b_2}\right)$$

$$1.3297 \pm (1.70)(0.3219)$$

$$1.3297 \pm 0.5472$$

We are 90% confident that the true average change in price due to each single-unit increase in age-squared (β_2) is somewhere between 0.7825 hundred dollars and 1.8769 hundred dollars (between $78.25 and $187.69).

The remaining parts will be completed using output from the Data window.

	price	age	age_sq	PRE_1	RES_1	SRE_1	SEP_1	LMCI_1	UMCI_1	LICI_1	UICI_1
32	.	8	64	48.33213	.	.	3.37639	42.58846	54.07581	25.79608	70.86819

g. Obtain a point estimate for the mean sales price of 8-year-old Nissan Zs.

The point estimate (`pre_1`) is 48.33213 hundred dollars (about $4833.21).

h. Test the alternative hypothesis that the mean sales price of 8-year-old Nissan Zs is greater than $4000 ($y$ = 40 hundred dollars). Test at the 10% significance level.

<u>Step 1</u>: **Hypotheses** | <u>Step 2</u>: **Significance Level**

$H_0:\ \hat{y} = 40 \quad \left(when \ \ x = 8 \ \ \& \ \ x^2 = 64\right)$ | $\alpha = 0.10$

$H_a:\ \hat{y} > 40 \quad \left(when \ \ x = 8 \ \ \& \ \ x^2 = 64\right)$

<u>Step 3</u>: **Critical Value(s) and Rejection Region(s)**

Critical Value: $t_{\alpha, df = n-(k+1)} = t_{0.10, df=28} = t_{80\%CI, df=28} = 1.31$

Reject the null hypothesis if $T \geq 1.31$ (or if p-value ≤ 0.10).

<u>Step 4</u>: **Test Statistic**

$$T = \frac{\hat{y} - y_0}{s_{\hat{y}}} = \frac{\texttt{pre\_1} - y_0}{\texttt{sep\_1}} = \frac{48.33213 - 40}{3.37639} = 2.4678$$

p-value = 0.009

<u>Step 5</u>: **Conclusion**

Since 2.4678 ≥ 1.31 (p-value = 0.009 ≤ 0.10), we shall reject the null hypothesis.

<u>Step 6</u>: **State conclusion in words**

At the $\alpha = 0.10$ level of significance, there exists enough evidence to conclude that the mean sales price of 8-year-old Nissan Zs is greater than $4000.

i. Determine a 90% confidence interval for the mean sales price of 8-year-old Nissan Zs.

We are 90% confident that the mean sales price of 8-year-old Nissan Zs is somewhere between 42.58846 (`lmci_1`) and 54.07581(`umci_1`) hundred dollars (or somewhere between about $4258.85 and about $5407.58).

j. Find the predicted sales price of an 8-year-old Nissan Z for sale by Sandy Beaches.

The predicted sales price (`pre_1`) of the 8-year-old Nissan Z for sale by Sandy Beaches is 48.33213 hundred dollars ($4833.21).

k. Determine a 90% prediction interval for the sales price of Sandy's 8-year-old Nissan Z.

We are 90% certain that the sales price of Sandy's 8-year-old Nissan Z will be somewhere between 25.79608 (`lici_1`) and 70.86819 (`uici_1`) hundred dollars (or somewhere between about $2579.61 about and $7086.82).