A S S I G N M E N T 02

<u>Statistics for Business & Economics</u>
Course Code: STA 217

<u>Submitted To</u>

Dr. Abdus Sattar
Course Instructor,
Statistics for Business & Economics
Department of Economics
East-West University, Dhaka

<u>Submitted by</u>

S. M. Benzir Ahmed
Department of Business Administration
East-West University, Dhaka

East-West University, Dhaka
December 14, 2006

Exercise 1(a)

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Selling Price	102	125.00	345.30	220.6529	47.29017
Valid N (listwise)	102				

Let's find the Class Interval

2^k

$2^9 = 512$

If we set k = 9 it exceed the highest value 345.30

i ≥ H – L / k = 345.30 - 125.00 / 9 = 24.47 = 25 (Approx)

Statistics

P

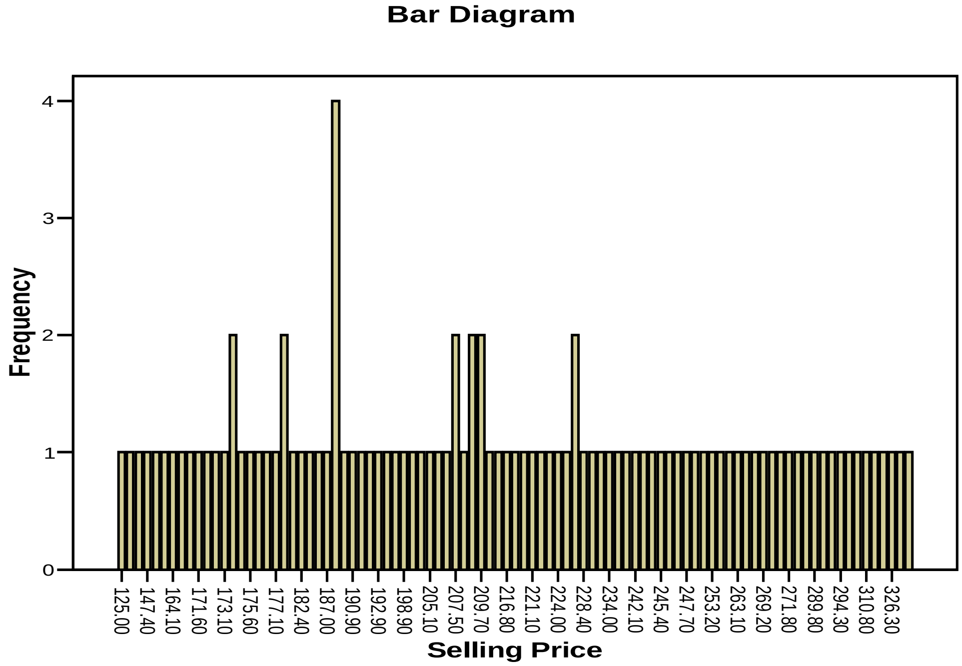
N	Valid	102
	Missing	0

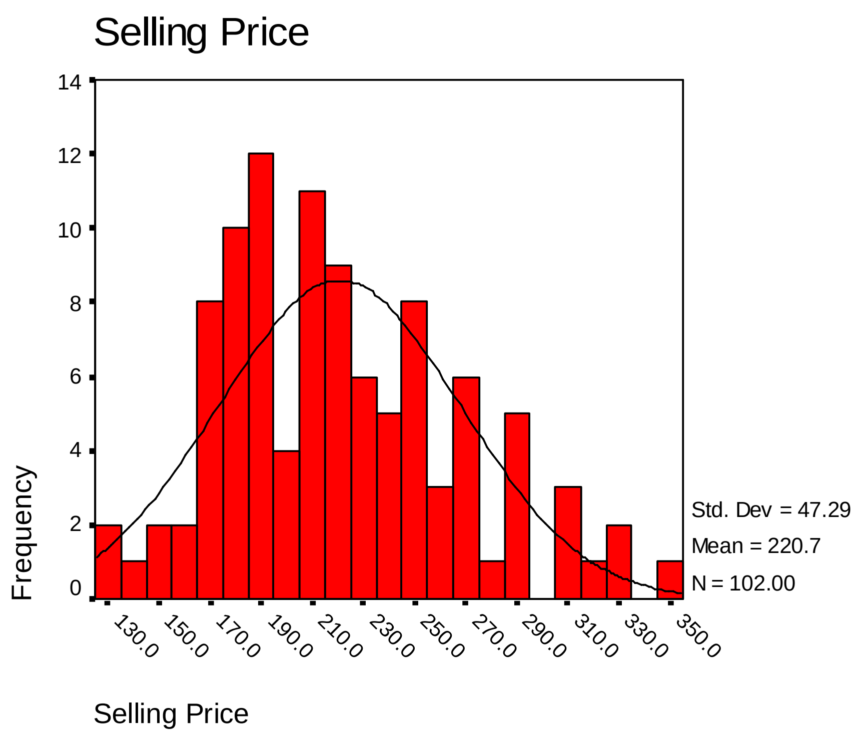
P

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	4	3.9	3.9	3.9
	2.00	12	11.8	11.8	15.7
	3.00	25	24.5	24.5	40.2
	4.00	20	19.6	19.6	59.8
	5.00	15	14.7	14.7	74.5
	6.00	13	12.7	12.7	87.3
	7.00	6	5.9	5.9	93.1
	8.00	4	3.9	3.9	97.1
	9.00	3	2.9	2.9	100.0
	Total	102	100.0	100.0	

Comment: The class interval is started from 125 as the lowest value is 125.00 and the upper limit is 350 as the highest value is 345.30. The class interval is 25 in each & there are 9 class intervals.

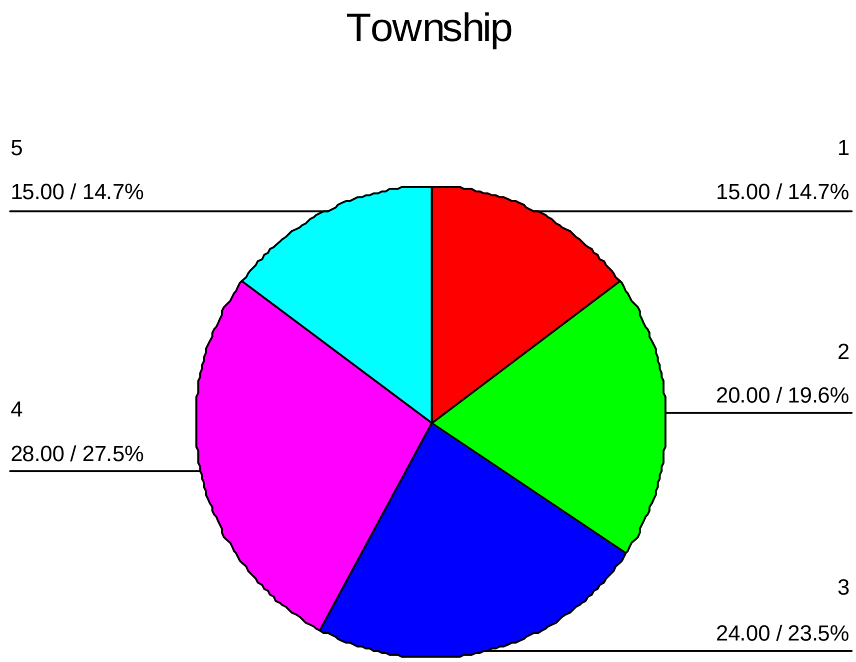
Exercise 1(b)





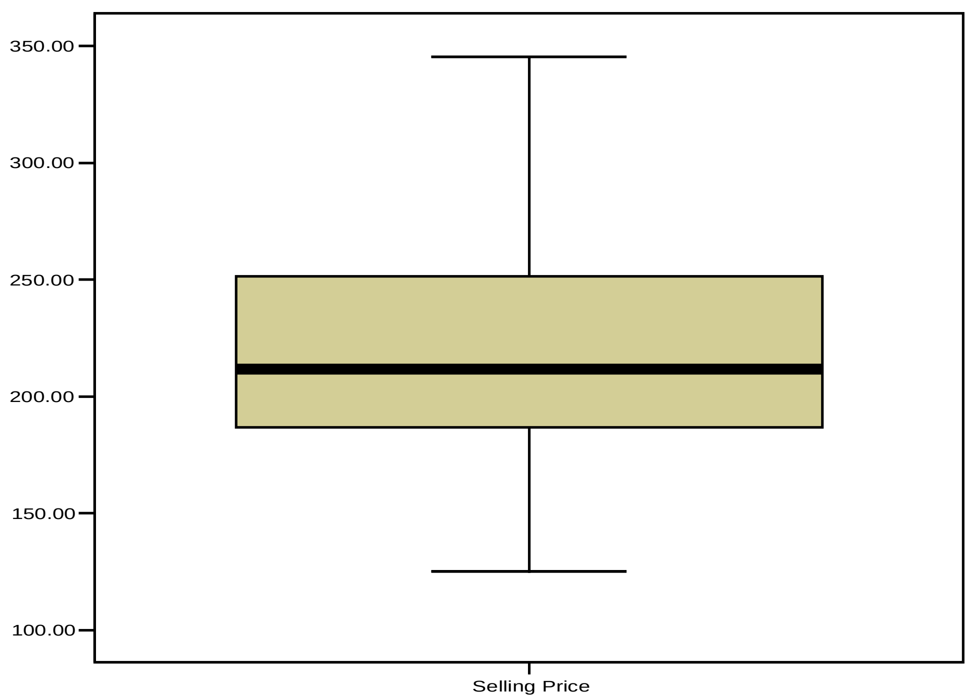
Comment: From the bar diagram we can see that the maximum number of selling price lies between 187.00 to 190.90. And the histogram is positively skewed.

Exercise 1(c)



Comment: The pie chard drawn on variable Township says the highest no of frequency is 28 and it's 27.5% and the lowest no. of frequency is 15 which is 14.7%, we found it 2 times here.

Exercise 1(d)



Comment: There is no outlier in the box plot. First quartiles are 180.00 and third quartiles are 250.00 & the median is 210.

Exercise 2(a)

Statistics

Selling Price

N	Valid	102
	Missing	0
Mean		220.6529
Median		211.6500

Interpretation: The mean number of selling price in a typical house is 220652.9.The median number of selling price in a typical house is 211650.

Exercise 2(b)

Statistics

Number of bedrooms

N	Valid	102
	Missing	0
Mean		3.78
Median		4.00

Interpretation: The mean number of bedrooms in a typical house is 4 (Approx). The median number of bathrooms in a typical house is 4.

Exercise 2 (c)

Statistics

Number of bathrooms

N	Valid	102
	Missing	0
Mean		2.0784
Median		2.0000

Interpretation: The mean number of bathrooms in a typical house is 2.0784.The median number of bathrooms in a typical house is 2.0000.

Exercise 2(d)

Statistics

Distance from the center of the city

N	Valid	102
	Missing	0
Mean		14.68
Median		15.00

Interpretation: The mean distance from the center of the city from a typical house is 14.68.The median distance from the center of the city from a typical house is 15.

Selling Price Part

Exercise 3(a)

Statistics

Selling Price

N	Valid	102
	Missing	0
Mean		220.6529
Median		211.6500
Std. Deviation		47.29017

Exercise 3(b)

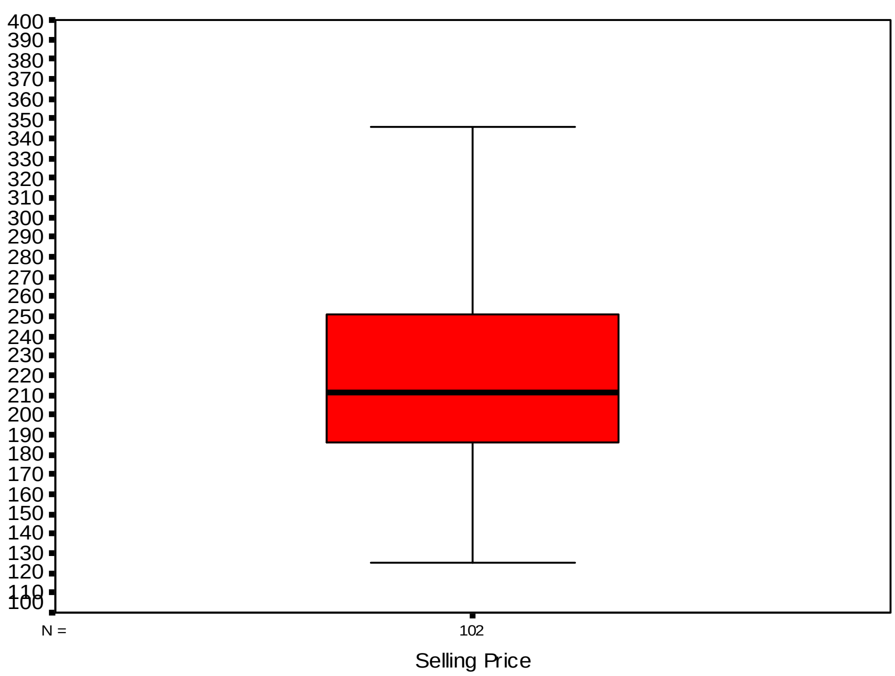
Statistics

Selling Price

N	Valid	102
	Missing	0
Skewness		.502
Std. Error of Skewness		.239

Comment: The coefficient of skewness is 0.502 which is positively skewed.

Exercise 3(c)



Comment: No there are no outliers in the box plot. The first quartile is 185 and third quartile is 253. Here Median is 210.

Exercise 3(d)

Summary: The mean selling price is 220652.9 and the median & Standard Deviation is 211650 & 47290.17. The coefficient of skewness is 0.502 which is positively skewed. There are no outliers in the box plot. The first quartile is 185 and third quartile is 253.

Area of homes in Square Feet

Exercise 3(a)

Statistics

Size of the home in square feet

N	Valid	102
	Missing	0
Mean		2225.49
Median		2200.00
Std. Deviation		249.628

Exercise 3(b)

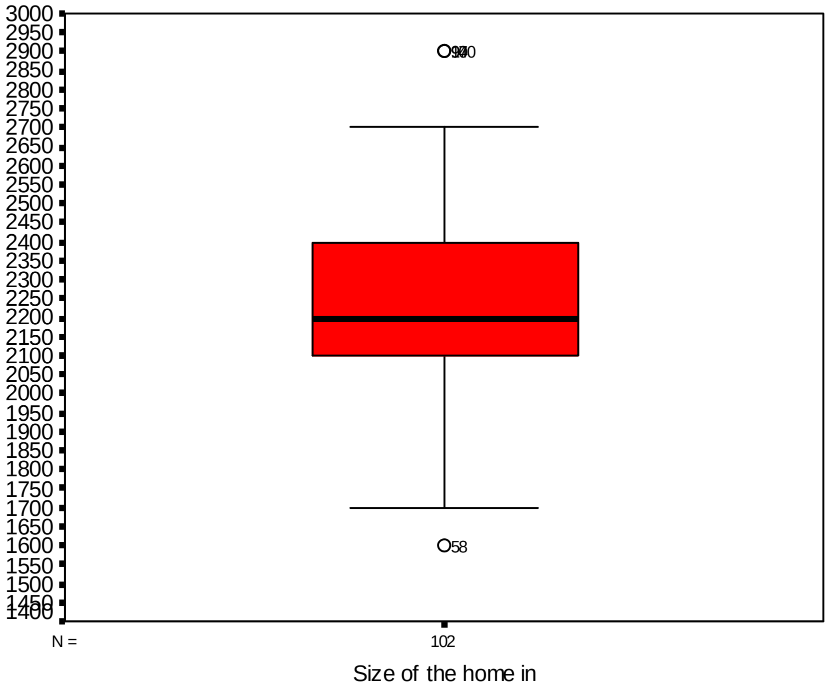
Statistics

Size of the home in square feet

N	Valid	102
	Missing	0
Skewness		.327
Std. Error of Skewness		.239

Comment: The coefficient of skewness is 0.502 which is positively skewed.

Exercise 3(c)



Comment: There are outliers in the box plot1600 & 2900. The first quartile is 2100 and third quartile is 2400. Here Median is 2200.

Exercise 3(d)

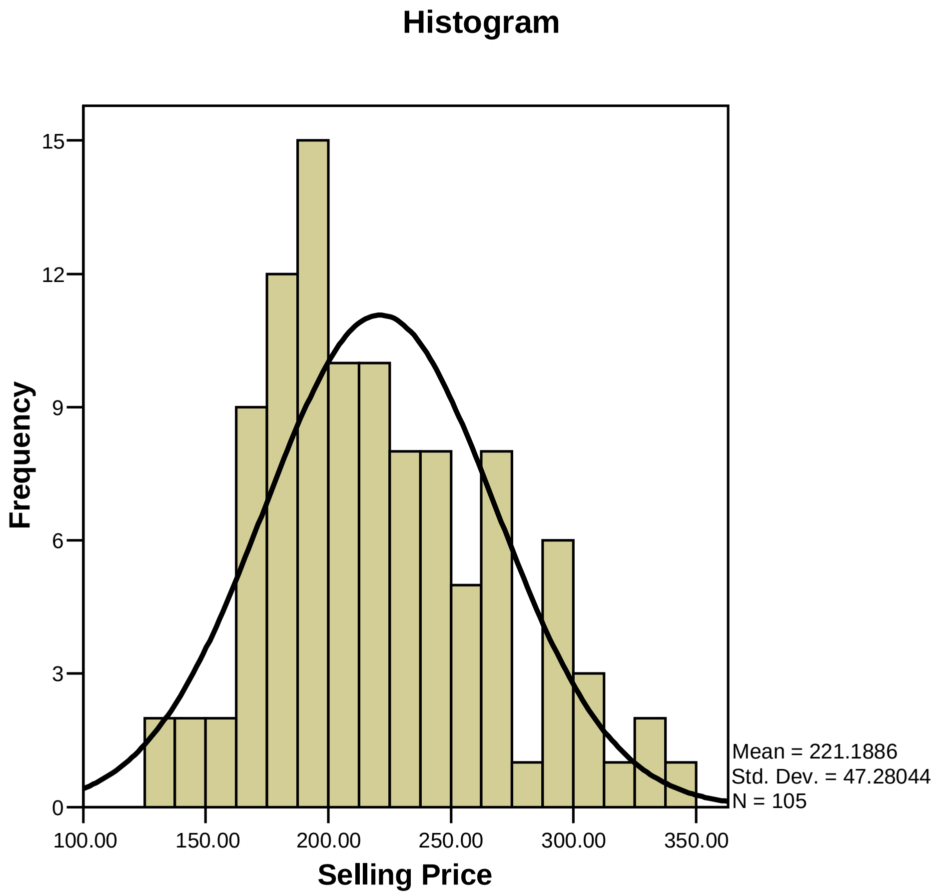
Summary: The mean area of the homes in square feet is 2225.49 and the median & Standard Deviation is 2200 & 249.628. The coefficient of skewness is 0.327 which is positively skewed. There are outliers in the box plot1600 & 2900. The first quartile is 2100 and third quartile is 2400.

Exercise 4(a)

Statistics

Selling Price

N	Valid	105
	Missing	0
Mean		221.1886
Std. Deviation		47.28044



Comment: From histogram we see it's positively skewed. It don't follow the normal distribution

Exercise 4(b)

Statistics

Selling Price

N	Valid	92
	Missing	0
Mean		219.9717
Std. Deviation		47.47787

Exercise 5(a)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Selling Price	102	220.6529	47.29017	4.68242

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Selling Price	47.124	101	.000	220.65294	211.3643	229.9416

The upper value is 229.9416 and the lower value is 211.3643.

$$211.3643 \leq \mu \leq 229.9416$$

Interpretation: If 100 sample of the same size could be taken and similar confidence interval were constructed, 95 would contain population parameter, mean, μ.

Exercise 5(b)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Distance from the center of the city	102	14.68	4.859	.481

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Distance from the center of the city	30.507	101	.000	14.676	13.72	15.63

The upper value is 15.63 and the lower value is 13.72

$$13.72 \leq \mu \leq 15.63$$

Interpretation: If 100 sample of the same size could be taken and similar confidence interval were constructed, 95 would contain population parameter, mean, μ.

Exercise 6(a)

Step 1

H_0: $\mu \geq 220000$

H_1: $\mu < 220000$

Step 2
$\alpha = 0.01$

Step 3
Z statistics is to be used.

$$Z = \frac{X - \mu}{\sigma / \sqrt{n}}$$

Step 4:
Decision Rule: If p value < α value (0.01), we reject H_0 otherwise it is accepted.

Step 5:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Selling Price	102	220.6529	47.29017	4.68242

One-Sample Test

	Test Value = 220000					
					99% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Selling Price	-46937.085	101	.000	-219779.34706	-219791.6402	-219767.0539

P value (two tailed) = .000
P value (one tailed) = 0

Decision: Since p < α value we reject H_0 . So H_1 is accepted. That means, the mean selling price in the Venice area is less than $220000.

Exercise 6(b)

Step 1
H_0: $\mu \geq 2100$
H_1: $\mu < 2100$

Step 2
$\alpha = 0.01$

Step 3
Z statistics is to be used.

$$Z = \frac{X - \mu}{\sigma / \sqrt{n}}$$

Step 4:
Decision Rule: If p value < α value (0.01), we reject H_0 otherwise it is accepted.

Step 5:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Size of the home in square feet	102	2225.49	249.628	24.717

One-Sample Test

	Test Value = 2100					
					99% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Size of the home in square feet	5.077	101	.000	125.490	60.60	190.38

P value (two tailed) = .000
P value (one tailed) = 0

Decision: Since p < α value we reject H_0 . So H_1 is accepted. That means, the mean size of homes sold in the Venice area is less than 2100 square feet.

Exercise 7(a)

H_0: $\mu_1=\mu_2$
H_1: $\mu_1\neq\mu_2$

α =0.05

Z statistics is to be used.

$$Z = \frac{X_1 - X_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Decision Rule: If p value < α Value (0.05) we reject H_0 otherwise we accept it.

Group Statistics

	Pool	N	Mean	Std. Deviation	Std. Error Mean
Selling Price	0	38	202.7974	33.70506	5.46768
	Yes	64	231.2547	51.10378	6.38797

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Selling Price	Equal variances assumed	8.509	.004	-3.057	100	.003	-28.45732	9.30776	-46.92366	-9.99098
	Equal variances not assumed			-3.384	98.816	.001	-28.45732	8.40843	-45.14186	-11.77278

Decision: Since p value (0.003) is less than α value (0.05) so H_0 is rejected. Yes there is a difference in the mean selling price of homes with a pool and without a pool. So H_1 is accepted.

Exercise 7(b)

H_0: $\mu_1=\mu_2$
H_1: $\mu_1\neq\mu_2$
α =0.05

Z statistics is to be used.

$$Z = \frac{X_1 - X_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Decision Rule: If p value < α Value (0.05) we reject H_0 otherwise we accept it.

Group Statistics

	Garage attached	N	Mean	Std. Deviation	Std. Error Mean
Selling Price	0	33	184.8545	28.21938	4.91236
	Yes	69	237.7739	45.02835	5.42078

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Selling Price	Equal variances assumed	9.554	.003	-6.186	100	.000	-52.91937	8.55434	-69.89093	-35.94780
	Equal variances not assumed			-7.234	92.699	.000	-52.91937	7.31547	-67.44707	-38.39167

Decision: Since p value (.000) is less then α value (0.05), H_0 is rejected so H_1 is accepted. Yes there is difference in the mean selling price of homes with attached garage and without a garage.

Exercise 7(c)

H_0: $\mu_1=\mu_2$
H_1: $\mu_1\neq\mu_2$

α =0.05

Z statistics is to be used.

$$Z = \frac{X_1 - X_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Decision Rule: If p value < α Value (0.05) we reject H_0 otherwise we accept it.

Group Statistics

	Township	N	Mean	Std. Deviation	Std. Error Mean
Selling Price	1	15	196.9133	35.78405	9.23940
	2	20	227.4500	44.19337	9.88194

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Selling Price	Equal variances assumed	1.835	.185	-2.189	33	.036	-30.53667	13.94877	-58.91565	-2.15768
	Equal variances not assumed			-2.257	32.761	.031	-30.53667	13.52846	-58.06815	-3.00518

Decision: Since p value (.036) is greater than α value (0.05), H_0 is rejected H_1 is accepted. So there is difference in the mean selling price of township 1 and township 2.

Exercise 8(a)

H_0: $\sigma_1^2=\sigma_2^2$
H_1: $\sigma_1^2\neq\sigma_2^2$

α =0.02

F statistics is to be used.

Decision Rule: If F_{comp} > 2.3449, we reject H_0 otherwise accept it.

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	220.6529412	0.62745098
Variance	2236.359942	0.236070666
Observations	102	102
Pooled Variance	1118.298006	
Hypothesized Mean Difference	0	
df	202	
t Stat	46.98717231	
P(T<=t) one-tail	5.3336E-111	
t Critical one-tail	2.067095502	
P(T<=t) two-tail	1.0667E-110	
t Critical two-tail	2.344949608	

Decision: Since t_{comp} value (46.9871) > 2.3449, H_0 is rejected. H_1 is accepted. Yes there is a difference in the variability of the selling prices of the homes that have a pool and those that have no pool.

Exercise 8(b)

H_0: $\sigma_1^2=\sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$

α =0.02
F statistics is to be used.

Decision Rule: If t computed value > 2.3449, H_0 is rejected otherwise accepted.

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	220.6529412	0.676470588
Variance	2236.359942	0.221025044
Observations	102	102
Pooled Variance	1118.290483	
Hypothesized Mean Difference	0	
df	202	
t Stat	46.97686202	
P(T<=t) one-tail	5.5548E-111	
t Critical one-tail	2.067095502	
P(T<=t) two-tail	1.111E-110	
t Critical two-tail	2.344949608	

Decision: Since t computed value (46.97)>2.3449 so H_0 is rejected. H_1 is accepted. Yes, there is a difference in the variability of the selling prices of the homes with an attached garage and those that do not have attached garage.

Exercise 8(c)

H_0: $\mu_1=\mu_2$
H_1: $\mu_1 \neq \mu_2$

α =0.05
F statistics is to be used.

Decision Rule: If p value < α Value (0.05) we reject H_0 otherwise we accept it.

ANOVA
Selling Price

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11825.696	4	2956.424	1.340	.261
Within Groups	214046.658	97	2206.667		
Total	225872.354	101			

Decision: Since p value (0.261) is less then α value (0.05), H_0 is rejected so H_1 is accepted. Yes there is difference in the mean selling price of homes among the 5 townships.

Exercise 9(a)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	47.211	41.442		1.139	.258
	Number of bedrooms	7.387	2.625	.233	2.814	.006
	Size of the home in square feet	.038	.015	.203	2.559	.012
	Pool	20.339	7.224	.209	2.815	.006
	Distance from the center of the city	-1.118	.756	-.115	-1.479	.143
	Township	-2.031	2.738	-.055	-.742	.460
	Garage attached	37.116	7.749	.369	4.790	.000
	Number of bathrooms	21.459	9.448	.174	2.271	.025

a Dependent Variable: Selling Price

$Y_{\text{Selling Price}} = 47.211 + 7.387X_1 + .038X_2 + 20.339X_3 - 1.118X_4 - 2.031X_5 + 37.116X_6 + 21.459X_7$.

Interpret: 47.211 is the Y intercept.

$7.387X_1$ means if number of bedrooms (X_1) is increased by 1 unit, total selling price will be increased by $7.387 keeping all other variables constant.

$0.038X_2$ means if size of the home in square feet increased by 1 unit total selling price will be increased by $.038.

$20.339X_3$ means if probability of pool is increased by 1 unit, total selling price will be increased by $20.339.

$- 1.118X_4$ means if distance from the center of the city is increased by 1 unit, total selling price will be decreased by $1.118.

$- 2.031X_5$ means if township is increased by 1 unit, total selling price will be decreased by $2.031.

$37.116X_6$ means if number of garage attached increased by 1 unit, total selling price will be increased by 37.116.

$21.459X_7$ means if number of bathrooms increased by 1 unit, total selling price will be increased by $21.459.

Exercise 9(b)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.729(a)	.532	.497	33.54996

a Predictors: (Constant), Number of bathrooms, Size of the home in square feet, Township, Garage attached, Pool, Distance from the center of the city, Number of bedrooms

Interpretation: 49.7% of the total variation in (Selling Price) is explained by regression (Independent Variable).

Exercise 9(c)

Correlations

		Selling Price	Number of bedrooms	Size of the home in square feet	Pool	Township	Distance from the center of the city	Garage attached	Number of bathrooms
Selling Price	Pearson Correlation	1	.452(**)	.363(**)	.292(**)	.117	**-.362(**)**	**.526(**)**	.357(**)
	Sig. (2-tailed)	.	.000	.000	.003	.243	.000	.000	.000
	N	102	102	102	102	102	102	102	102
Number of bedrooms	Pearson Correlation	.452(**)	1	.371(**)	-.003	.194	-.173	.224(*)	.298(**)
	Sig. (2-tailed)	.000	.	.000	.979	.050	.081	.024	.002
	N	102	102	102	102	102	102	102	102
Size of the home in square feet	Pearson Correlation	.363(**)	.371(**)	1	.210(*)	.188	-.132	.071	-.011
	Sig. (2-tailed)	.000	.000	.	.034	.059	.186	.478	.915
	N	102	102	102	102	102	102	102	102
Pool	Pearson Correlation	.292(**)	-.003	.210(*)	1	.190	-.135	.074	.052
	Sig. (2-tailed)	.003	.979	.034	.	.056	.175	.460	.603
	N	102	102	102	102	102	102	102	102
Township	Pearson Correlation	.117	.194	.188	.190	1	-.194	.059	.028
	Sig. (2-tailed)	.243	.050	.059	.056	.	.051	.558	.784
	N	102	102	102	102	102	102	102	102
Distance from the center of the city	Pearson Correlation	**-.362**(**)	-.173	-.132	-.135	-.194	1	-.341(**)	-.207(*)
	Sig. (2-tailed)	.000	.081	.186	.175	.051	.	.000	.037
	N	102	102	102	102	102	102	102	102
Garage attached	Pearson Correlation	**.526**(**)	.224(*)	.071	.074	.059	-.341(**)	1	.224(*)
	Sig. (2-tailed)	.000	.024	.478	.460	.558	.000	.	.023
	N	102	102	102	102	102	102	102	102
Number of bathrooms	Pearson Correlation	.357(**)	.298(**)	-.011	.052	.028	-.207(*)	.224(*)	1
	Sig. (2-tailed)	.000	.002	.915	.603	.784	.037	.023	.
	N	102	102	102	102	102	102	102	102

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Comment: The strong correlation is found with Garage attached & Selling Price. Weak correlation is found with Distance from the center of the city & Selling Price. No there is no multicolenearity problem.

Exercise 9(d)

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$
H_1: They are not all same
α =0.05

F statistics is to be used.

Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	120065.957	7	17152.280	15.238	.000(a)
	Residual	105806.397	94	1125.600		
	Total	225872.354	101			

a Predictors: (Constant), Number of bathrooms, Size of the home in square feet, Township, Garage attached, Pool, Distance from the center of the city, Number of bedrooms
b Dependent Variable: Selling Price

Decision: Since p value (.000) is less than α value (0.05), H_0 is rejected H_1 is accepted. So, all the values are not same.

Exercise 9(e)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	47.211	41.442		1.139	.258
	Number of bedrooms	7.387	2.625	.233	2.814	.006
	Size of the home in square feet	.038	.015	.203	2.559	.012
	Pool	20.339	7.224	.209	2.815	.006
	Distance from the center of the city	-1.118	.756	-.115	-1.479	.143
	Township	-2.031	2.738	-.055	-.742	.460
	Garage attached	37.116	7.749	.369	4.790	.000
	Number of bathrooms	21.459	9.448	.174	2.271	.025

a Dependent Variable: Selling Price

For number of bedrooms:

H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$
$\alpha = 0.05$

F statistics is to be used.

Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

Decision: Since p value (.006) is less than α value (0.05), H_0 is rejected & H_1 is accepted.

For size of the home:

H_0: $\beta_2 = 0$
H_1: $\beta_2 \neq 0$
$\alpha = 0.05$

F statistics is to be used.

Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

Decision: Since p value (.012) is less than α value (0.05), H_0 is rejected & H_1 is accepted.

For pool:

H_0: $\beta_3 = 0$
H_1: $\beta_3 \neq 0$
$\alpha = 0.05$

F statistics is to be used.

Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

Decision: Since p value (.006) is less than α value (0.05), H_0 is rejected & H_1 is accepted.

For distance from the center of the city:

H_0: $\beta_4 = 0$
H_1: $\beta_4 \neq 0$
$\alpha = 0.05$

F statistics is to be used.

Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

Decision: Since p value (.143) is more than α value (0.05), H_0 is accepted.

For township:

H_0: $\beta_5 = 0$
H_1: $\beta_5 \neq 0$
$\alpha = 0.05$

F statistics is to be used.

Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

Decision: Since p value (.460) is more than α value (0.05), H_0 is accepted.

For Garage attached:

H_0: $\beta_6 = 0$
H_1: $\beta_6 \neq 0$
$\alpha = 0.05$

F statistics is to be used.

Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

Decision: Since p value (.000) is less than α value (0.05), H_0 is rejected & H_1 is accepted.

For Number of bathrooms:

H_0: $\beta_7 = 0$
H_1: $\beta_7 \neq 0$
$\alpha = 0.05$

F statistics is to be used.

Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

Decision: Since p value (.025) is less than α value (0.05), H_0 is rejected & H_1 is accepted.

Here we can consider delete 2 variable, they are $\beta_{5\ \&}\ \beta_6$

So, the new equation will be
$Y_{\text{Selling Price}} = 47.211 + 7.387X_1 + .038X_2 + 20.339X_3 - 1.118X_4 + 21.459X_7$.