A S S I G N M E N T
02

Statistics for Business & Economics
Course Code: STA 217

Submitted To

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

Submitted by

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

East-West University, Dhaka
16 April 2008

Exercise 1(a)

Let's find the Class Interval

2^k

$2^7 = 128$

K number of classes = 7

i = Highest Value – Lowest Value / k = 83601 - 10000 / 7 = 10,515

Annual Wage's Frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10000-20515	51	52.0	52.0	52.0
	20515-31030	29	29.6	29.6	81.6
	31030-41545	12	12.2	12.2	93.9
	41545-52060	4	4.1	4.1	98.0
	52060-62575	1	1.0	1.0	99.0
	73090-83605	1	1.0	1.0	100.0
	Total	98	100.0	100.0	

Comment: The class interval is started from 10000 as the lowest value is 10000 and the upper limit is 83605 as the highest value is 83601. The class interval is 10515 in each & there are 7 class intervals.

Exercise 1(b)

Statistics

Annual Wages in Dollar

N	Valid	98
	Missing	0
Mean		22487.59
Median		20310.00
Mode		12000
Std. Deviation		11883.85
Variance		141225990.47
Range		73601

Interpretation:

Mean: the typical Annual Wages of the employees is: $22487.59

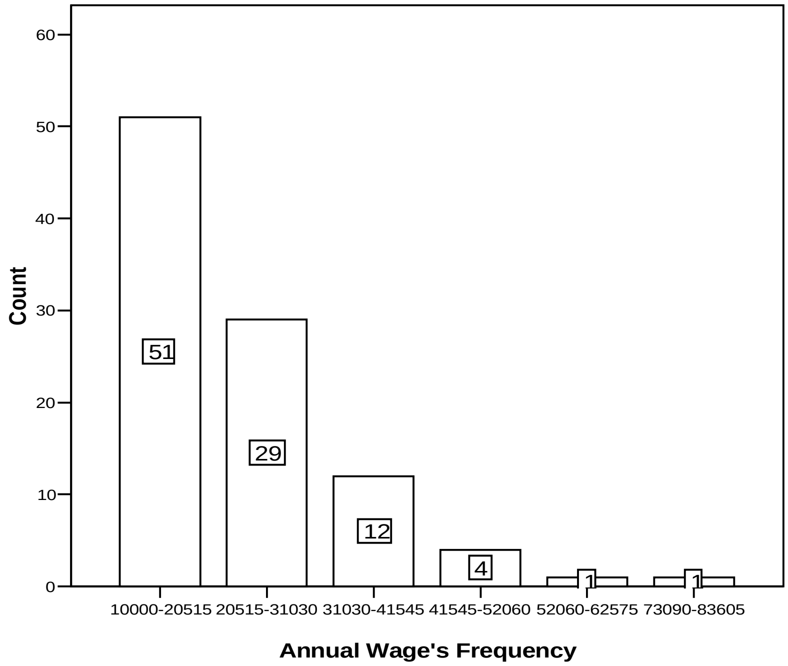
Median: 50% of Annual Wages are below or above the median $20310.00

Standard Deviation (SD): SD is the dispersion of actual data. Most of the Annual Wages are clustered from the mean value.

Variance: Variance is also measure of dispersion of actual data, and the value is: 141225990.47

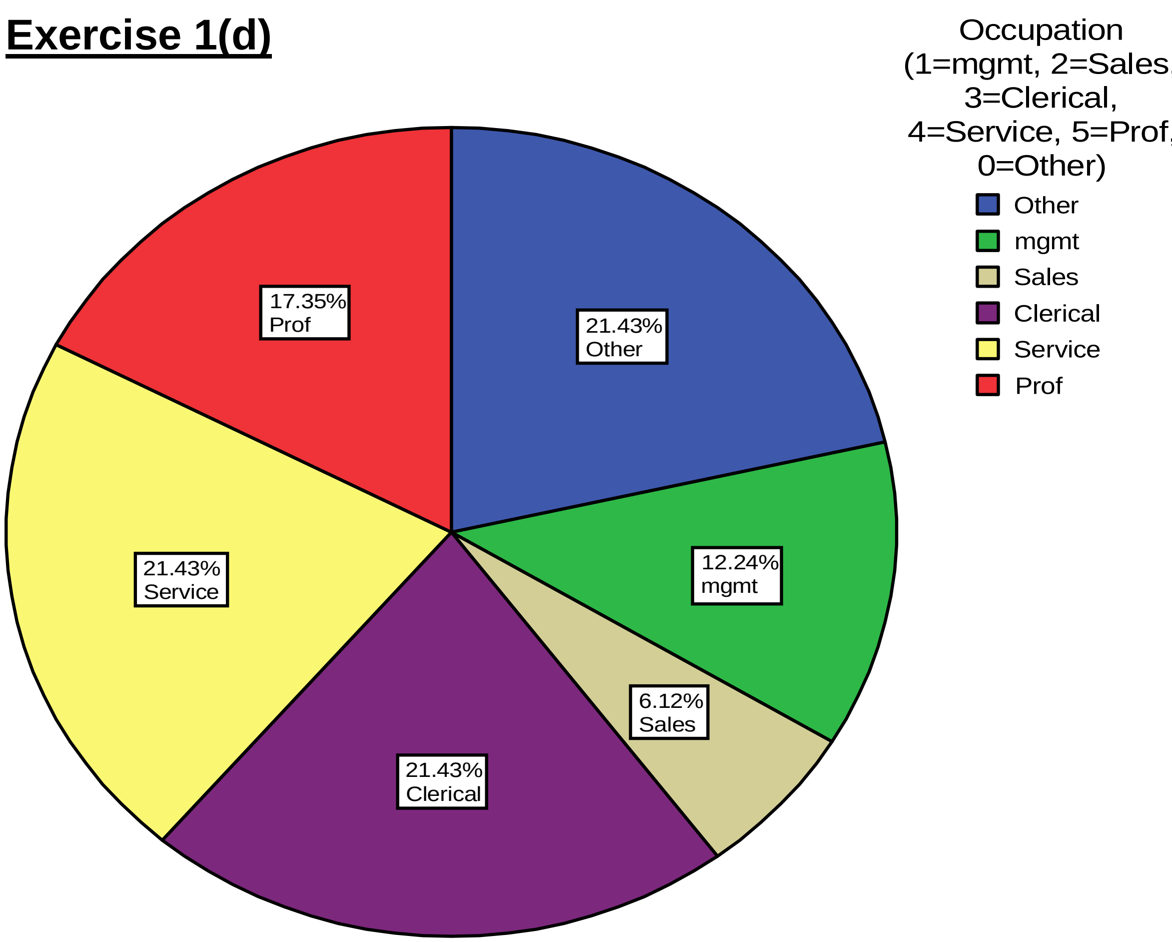
Range: The difference between the highest annual wage and lowest annual wage is $73601.

Exercise 1(c)



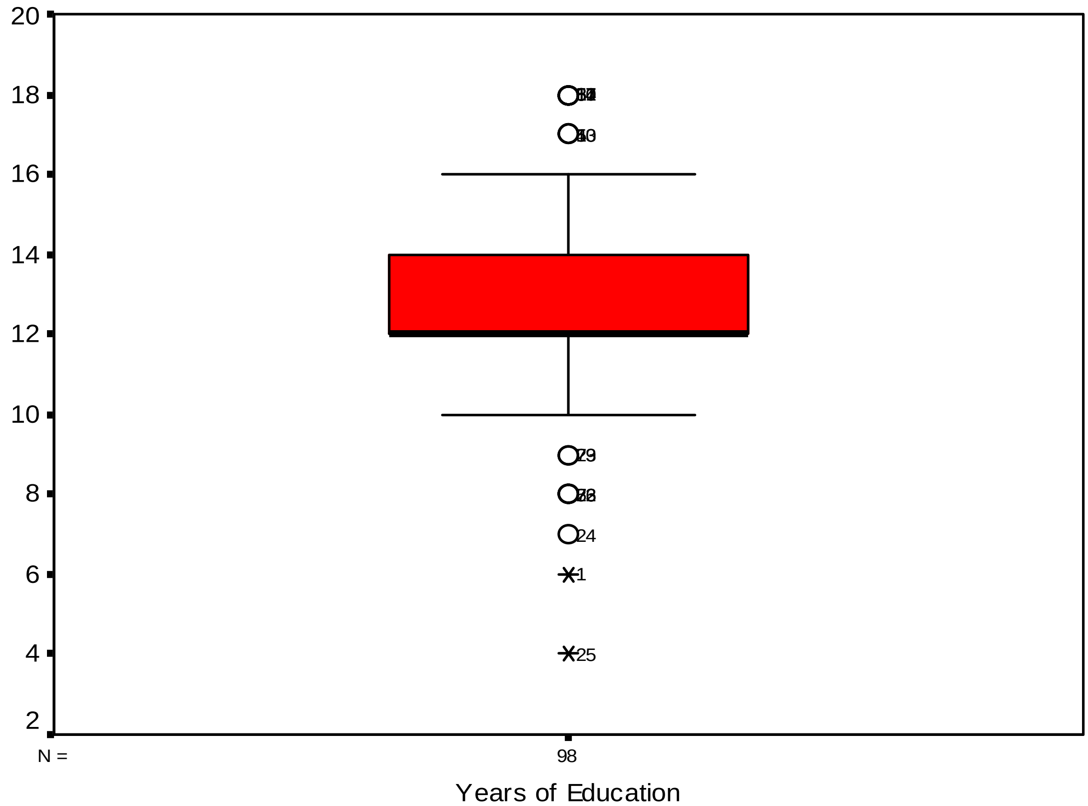
Comment: From the bar diagram we can see that the maximum annual wages lies between 10000 to 20515 range (51). The lowest is in 73090 to 83605 (1).

Exercise 1(d)



Comment: The pie chard drawn on variable Occupation says the highest no of frequency is 21 and it's 21.4% we found it 3 times here. The lowest no. of frequency is 6 which is 6.1%.

Exercise 1(e)



Comment: There are 16 outliers in the box plot with 2 extreme values of 4 in 25th observation and 6 in 1st observation. 4 outliers are 7 in 24th observation, 8 in 5, 23, 72, 76th observation, 9 in 29 & 73rd observation and 17 in 4, 13, 50th and 18 in 10, 11, 54, 57, 81, 87th observation. First quartiles are 11 and third quartiles are 14 & the median is 12. The box plot is positively skewed.

Exercise 2(a)

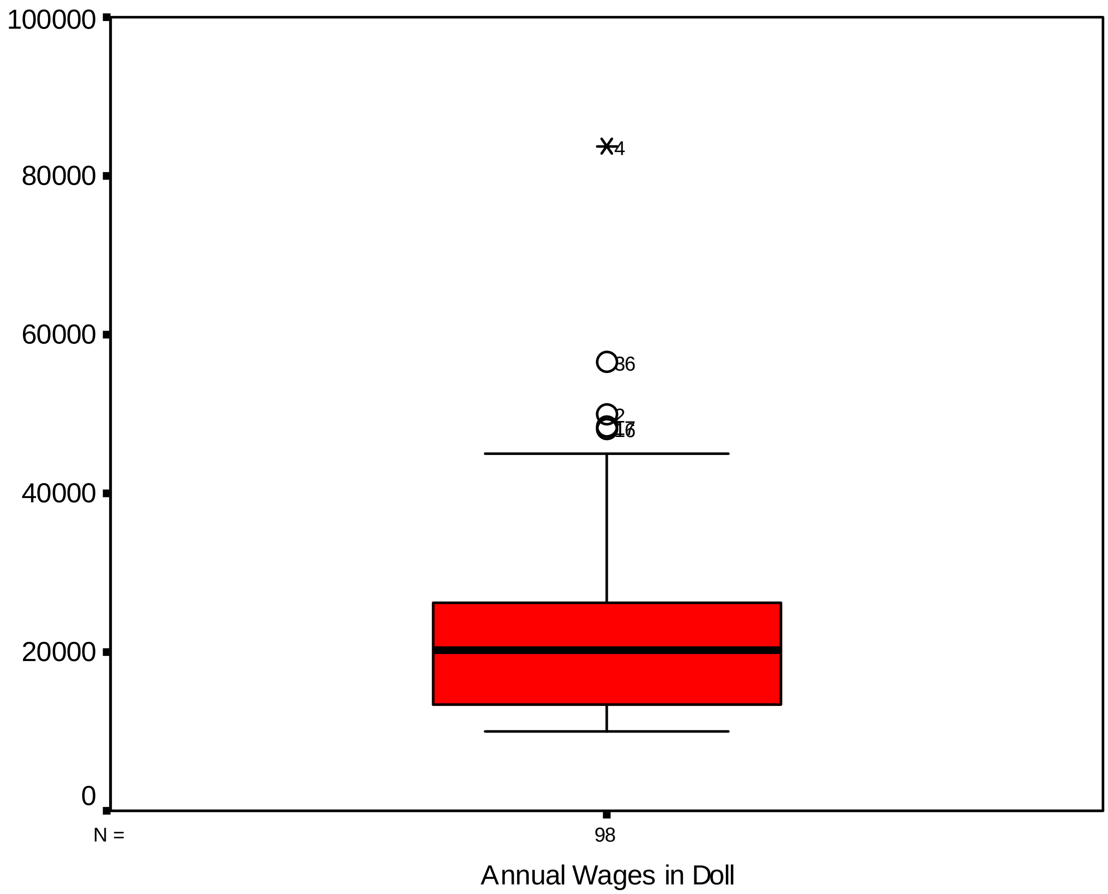
Statistics

Annual Wages in Dollar

N	Valid	98
	Missing	0
Skewness		2.046
Std. Error of Skewness		.244

Comment: The coefficient of skewness is 2.046 which is positively skewed.

Exercise 2(b)



Comment: There are 4 outliers in the box plot with an extreme value of 83601 in 4^{th} observation. 4 outliers are 56560 in 36^{th} observation, 49898 in 2^{nd} observation, 48400 in 17^{th} observation and 48055 in 16^{th} observation. The first quartile is 10000 and third quartile is 25000. Here Median is 20000. The box plot is positively skewed.

Exercise 3(a)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Annual Wages in Dollar	98	22487.59	11883.85	1200.45

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Annual Wages in Dollar	18.733	97	.000	22487.59	20105.03	24870.15

The upper value is 24870.15 and the lower value is 20105.03.

$$20105.03 \leq \mu \leq 24870.15$$

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 3(b)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Age in Years	98	33.60	6.54	.66

One-Sample Test

	Test Value = 0					
					99% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Age in Years	50.871	97	.000	33.60	31.87	35.34

The upper value is 35.34 and the lower value is 31.87.

$$31.87 \leq \mu \leq 35.34$$

Interpretation: If 100 sample of the same size could be taken and similar confidence interval were constructed, 99 would contain population parameter, mean, µ.

Exercise 4(a)

Step 1
H_0: $\mu \geq 31300$
H_1: $\mu < 31300$

Step 2
$\alpha = 0.05$

Step 3
t statistics is to be used.

Step 4:
Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

Step 5

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Annual Wages in Dollar	98	22487.59	11883.854	1200.451

One-Sample Test

	Test Value = 31300					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Annual Wages in Dollar	-7.341	97	.000	-8812.41	-11194.97	-6429.85

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 Annual Wage is less than \$31300.

Exercise 4(b)

Step 1
H_0: $\mu \leq 40$
H_1: $\mu > 40$

Step 2
$\alpha = 0.05$

Step 3
t statistics is to be used.

Step 4:
Decision Rule: If p value < α value (0.05), we reject H_0 otherwise it is accepted.

Step 5

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Age in Years	98	33.60	6.539	.661

One-Sample Test

	Test Value = 40					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Age in Years	-9.686	97	.000	-6.40	-7.71	-5.09

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 Age is greater than 40.

Exercise 5(a)

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

$\alpha = 0.05$

t statistics is to be used.

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

Group Statistics

	Gender (Male=0, Female=1)	N	Mean	Std. Deviation	Std. Error Mean
Annual Wages in Dollar	0	51	23508.33	14173.916	1984.745
	1	47	21379.98	8776.363	1280.164

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Annual Wages in Dollar	Equal variances assumed	4.011	.048	.885	96	.379	2128.35	2405.599	-2646.721	6903.430
	Equal variances not assumed			.901	84.382	.370	2128.35	2361.786	-2568.005	6824.715

Decision: Since p value (0.379) is more than α value (0.05) so H_0 is accepted. No, there is no difference in the mean annual wages of male and female wage earnings. So H_0 is accepted.

Exercise 5(b)

H_0: $\mu_1=\mu_2$

H_1: $\mu_1 \neq \mu_2$

α =0.05

t statistics is to be used.

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

Group Statistics

	Marital Status (1=Married, 0=Single)	N	Mean	Std. Deviation	Std. Error Mean
Annual Wages in Dollar	0	41	21392.02	9892.946	1545.019
	1	57	23275.63	13160.749	1743.184

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Annual Wages in Dollar	Equal variances assumed	.802	.373	-.772	96	.442	-1883.61	2438.630	-6724.250	2957.035
	Equal variances not assumed			-.809	95.786	.421	-1883.61	2329.329	-6507.420	2740.205

Decision: Since p value (0.442) is more than α value (0.05) so H_0 is accepted. No, there is no difference in the mean annual wages of married and unmarried wage earnings. So H_0 is accepted.

Exercise 6(a)

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

α =0.05

F statistics is to be used.

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

F-Test Two-Sample for Variances

	Female	*Male*
Mean	21379.97872	23508.33
Variance	77024540.11	2.01E+08
Observations	47	51
df	46	50
F	0.383397613	
P(F<=f) one-tail	0.000648224	
F Critical one-tail	0.617070839	

Decision: F_{comp} (0.383397613) < F_{table} (0.617070839), we accept H_0. That means there is no difference in the variability between the annual wages of Male and Female at 05 significance level.

Exercise 6(b)

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

α =0.01

F statistics is to be used.

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

F-Test Two-Sample for Variances

	Single	*Married*
Mean	21392.02439	23275.63158
Variance	97870376.52	173205315.1
Observations	41	57
df	40	56
F	0.565054118	
P(F<=f) one-tail	0.029972066	
F Critical one-tail	0.607730755	

Decision: F_{comp} (0.565054118) < F_{table} (0.607730755), we accept H_0. That means there is no difference in the variability between the annual wages of married and unmarried earners at .01 significance level.

Exercise 6(c)

H_0: $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5=\mu_6$
H_1: Not all the means are equal.

$\alpha = 0.05$

F statistics is to be used.

ANOVA

Annual Wages in Dollar

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1209521305.554	5	241904261.111	1.782	.124
Within Groups	12489399770.119	92	135754345.327		
Total	13698921075.673	97			

Decision: Since F_{calc} (1.782) < $F_{critical}$ (2.25) H_0 is not rejected.
That means all the treatment means are equal.

Exercise 7(a)

Coefficients(a)

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	14337.495	8986.742		1.595	.114
	Industry (1=Manufacturing, 2=Construction, 0=Other)	2732.314	3060.301	.106	.893	.374
	Occupation (1=mgmt, 2=Sales, 3=Clerical, 4=Service, 5=Prof, 0=Other)	1413.367	810.151	.215	1.745	.084
	Years of Education	-340.620	474.984	-.078	-.717	.475
	Gender (Male=0, Female=1)	-2247.135	2506.750	-.095	-.896	.372
	Years of Work Experience	464.164	355.765	.138	1.305	.195
	Marital Status (1=Married, 0=Single)	461.151	2589.086	.019	.178	.859
	Age in Years	186.233	193.383	.102	.963	.338

a Dependent Variable: Annual Wages in Dollar

$Y_{\text{Annual Wages}} = 14337.495 + 2732.314X_1 + 1413.367X_2 - 340.620X_3 - 2247.135X_4 + 464.164X_5 + 461.151X_6 + 186.233X_7$.

Interpret: 14337.495 is the Y intercept.

$2732.314X_1$ means if number of industry (X_1) is increased by 1 unit, total annual wages will be increased by $2732.314 keeping all other variables constant.

$1413.367X_2$ means if occupation changes by 1 unit total annual wages will be increased by $1413.367.

$-340.620X_3$ means if a year of education is increased by 1 unit, total annual wages will be decreased by $340.620.

$-2247.135X_4$ means, for females, total annual wages will be decreased by $2247.135 and there will no change for males.

$464.164X_5$ means if a year of work experience is increased by 1 unit, total annual wages will be increased by $464.164.

$461.151X_6$ means, for married persons, total annual wages will be increased by 461.151 and there will be no change for single (unmarried) persons.

$186.233X_7$ means if age increased by 1 unit, total annual wages will be increased by $186.233.

Exercise 7(b)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.271(a)	.073	.001	11876.33

a Predictors: (Constant), Age in Years, Years of Education, Gender (Male=0, Female=1), Industry (1=Manufacturing, 2=Construction, 0=Other), Years of Work Experience, Marital Status (1=Married, 0=Single), Occupation (1=mgmt, 2=Sales, 3=Clerical, 4=Service, 5=Prof, 0=Other)

Interpretation 7.3% of the total variation in (Annual Wages) is explained by regression (Independent Variables: Age of Years, Years of Education, Industry, Years of Work Experience, Marital Status, Occupation).

Exercise 7(c)

		Annual Wages in Dollar	Industry (1=Manufacturing, 2=Construction, 0=Other)	Occupation (1=mgmt, 2=Sales, 3=Clerical, 4=Service, 5=Prof, 0=Other)	Years of Education	Gender (Male=0, Female=1)	Years of Work Experience	Marital Status (1=Married, 0=Single)	Age in Years
Annual Wages in Dollar	Pearson Correlation	1.000	.037	.111	-.009	-.090	.141	.079	.150
	Sig. (2-tailed)	.	.715	.277	.933	.379	.166	.442	.140
	N	98	98	98	98	98	98	98	98
Industry (1=Manufacturing, 2=Construction, 0=Other)	Pearson Correlation	.037	1.000	**-.482**(**)	-.084	-.093	.075	-.145	.121
	Sig. (2-tailed)	.715	.	.000	.411	.365	.463	.153	.234
	N	98	98	98	98	98	98	98	98
Occupation (1=mgmt, 2=Sales, 3=Clerical, 4=Service, 5=Prof, 0=Other)	Pearson Correlation	.111	**-.482**(**)	1.000	**.289**(**)	.150	-.104	.070	-.035
	Sig. (2-tailed)	.277	.000	.	.004	.140	.306	.492	.732
	N	98	98	98	98	98	98	98	98
Years of Education	Pearson Correlation	-.009	-.084	**.289**(**)	1.000	-.063	.093	-.019	-.026
	Sig. (2-tailed)	.933	.411	.004	.	.535	.362	.849	.801
	N	98	98	98	98	98	98	98	98
Gender (Male=0, Female=1)	Pearson Correlation	-.090	-.093	.150	-.063	1.000	-.044	-.180	-.127
	Sig. (2-tailed)	.379	.365	.140	.535	.	.670	.077	.214
	N	98	98	98	98	98	98	98	98
Years of Work Experience	Pearson Correlation	.141	.075	-.104	.093	-.044	1.000	.150	.177
	Sig. (2-tailed)	.166	.463	.306	.362	.670	.	.141	.081
	N	98	98	98	98	98	98	98	98
Marital Status (1=Married, 0=Single)	Pearson Correlation	.079	-.145	.070	-.019	-.180	.150	1.000	.199(*)
	Sig. (2-tailed)	.442	.153	.492	.849	.077	.141	.	.049
	N	98	98	98	98	98	98	98	98
Age in Years	Pearson Correlation	.150	.121	-.035	-.026	-.127	.177	.199(*)	1.000
	Sig. (2-tailed)	.140	.234	.732	.801	.214	.081	.049	.
	N	98	98	98	98	98	98	98	98

Correlations

** 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 Years of Education & Occupation. Weak correlation is found with Industry & Occupation. No there is no multicolenearity problem.

Exercise 7(d)

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$
H_1: Not all the β's are same

$\alpha = 0.05$

F statistics is to be used.

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

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1004679712.604	7	143525673.229	1.018	.424(a)
	Residual	12694241363.070	90	141047126.256		
	Total	13698921075.673	97			

a Predictors: (Constant), Age in Years, Years of Education, Gender (Male=0, Female=1), Industry (1=Manufacturing, 2=Construction, 0=Other), Years of Work Experience, Marital Status (1=Married, 0=Single), Occupation (1=mgmt, 2=Sales, 3=Clerical, 4=Service, 5=Prof, 0=Other)
b Dependent Variable: Annual Wages in Dollar

Decision: Since p value (.424) is more than α value (0.05), H_0 is accepted H_1 is rejected. So, all the β's are same.

Exercise 7(e)

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	14337.495	8986.742		1.595	.114
	Industry (1=Manufacturing, 2=Construction, 0=Other)	2732.314	3060.301	.106	.893	.374
	Occupation (1=mgmt, 2=Sales, 3=Clerical, 4=Service, 5=Prof, 0=Other)	1413.367	810.151	.215	1.745	.084
	Years of Education	-340.620	474.984	-.078	-.717	.475
	Gender (Male=0, Female=1)	-2247.135	2506.750	-.095	-.896	.372
	Years of Work Experience	464.164	355.765	.138	1.305	.195
	Marital Status (1=Married, 0=Single)	461.151	2589.086	.019	.178	.859
	Age in Years	186.233	193.383	.102	.963	.338

a Dependent Variable: Annual Wages in Dollar

For number of Industry:

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 (.374) is more than α value (0.05), H_0 is accepted.

For Change in Occupation:

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

t 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 (.084) is more than α value (0.05), H_0 is accepted.

For Years of Education:

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 (.475) is more than α value (0.05), H_0 is accepted.

For Change in Gender:

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 (.372) is more than α value (0.05), H_0 is accepted.

For Years of Work Experience:

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 (.195) is more than α value (0.05), H_0 is accepted.

For Change in Marital Status:

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 (.859) is more than α value (0.05), H_0 is accepted.

For Age in Years:

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 (.338) is more than α value (0.05), H_0 is accepted.

Exercise 7(f)

The regression equation is not valid, because all values have been proofed to be reduced from main equation.