B,

Exam # 2 - Group Assignment

Members of Group - Thylia Watson

Ugochukwu Achara

Olaoluwa Ogunsola

Question One

Alcohol would be the most appropriate to use to model the number of hours per week because it is highly correlated with the work hours at 0.461 compared to age which is −0.110 and Sphrs at −0.153.

Please See Diagram Below

Correlations

		WkHrs	Age	SpHrs	Alcohol
WkHrs	Pearson Correlation	1	-.110**	-.153**	.461**
	Sig. (2-tailed)		<.001	<.001	<.001
	N	1655	1655	1655	1655
Age	Pearson Correlation	-.110**	1	-.105**	-.040
	Sig. (2-tailed)	<.001		<.001	.103
	N	1655	1655	1655	1655
SpHrs	Pearson Correlation	-.153**	-.105**	1	-.060*
	Sig. (2-tailed)	<.001	<.001		.014
	N	1655	1655	1655	1655
Alcohol	Pearson Correlation	.461**	-.040	-.060*	1
	Sig. (2-tailed)	<.001	.103	.014	
	N	1655	1655	1655	1655

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

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

Question TWO

Description of Model

$y = B_0 + B_1 x_1 + B_2 x_2 + E$

y = Annual Income for the household, in USD

x_1 = Education level for the head of household

x_2 = Average hours worked per week by the head of the household

At 95% CI, f-statistic is < 0.001 is statistically significant

The average annual income for the household where the head of household has an education level of 12 and works 40 hours per week is $55,892.00

The average annual income for the household where the head of household has an education level level of 16 and works 30 hours per week is $65,900.00

Yes, the prediction is significant difference because Education and Work hours are both highly correlated with income by 0.491 and 0.545 respectively with a difference of 0.054.

Please See Diagrams Used to Answer Question Two

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6.697E+11	2	3.349E+11	921.445	<.001[b]
	Residual	6.004E+11	1652	363408602.54		
	Total	1.270E+12	1654			

a. Dependent Variable: Income

b. Predictors: (Constant), WkHrs, Educ

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-31812.139	2520.284		-12.622	<.001	-36755.427	-26868.852		
	Educ	4562.153	160.714	.480	28.387	<.001	4246.928	4877.377	1.000	1.000
	WkHrs	824.066	26.064	.535	31.617	<.001	772.945	875.188	1.000	1.000

a. Dependent Variable: Income

Question Three

H_0: No difference between models. Keep reduced model

H_1: There is a difference. Full model is a significant improvement. Keep full model.

General Linear F-Test

Full Model		Reduced Model	
SSE (Residuals)	df(F)	SSE (Residuals)	df(R)
5.92814E+11	1651	6.00351E+11	1652
F-statistic	20.9901501		
p-value	4.96295E-06		

Based on the F-test conducted, we will use the full model because p-value of $4.96 \times 10^{-6} < \alpha\ 0.05$, which will result in rejecting the null hypothesis and conclude that there is a significant improvement

when we use the full model.

Question Four

4a Yes, the gender is statistically significant since p-value 0.037 < α 0.05.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-30659.408	2577.688		-11.894	<.001	-35715.290	-25603.526		
	Educ	4566.345	160.564	.481	28.439	<.001	4251.415	4881.276	.999	1.001
	WkHrs	826.350	26.060	.536	31.709	<.001	775.236	877.465	.998	1.002
	Head=Male	-2035.040	975.774	-.035	-2.086	.037	-3948.926	-121.155	.998	1.002

a. Dependent Variable: Income

4b When you include the sector of the economy, the head of the household works in, gender is no longer significant with p-value of 0.601 > α 0.05

(4c) When gender is significant it means that gender of the head of household is a contributing factor in

how much they earn. However, gender is no longer a meaningful contributing factor when the sector of the economy is included.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-30071.789	2489.397		-12.080	<.001	-34954.501	-25189.076		
	Educ	4673.946	155.737	.492	30.012	<.001	4368.483	4979.409	.989	1.011
	WkHrs	867.891	25.473	.563	34.071	<.001	817.928	917.854	.973	1.028
	Head=Male	507.151	970.607	.009	.523	.601	-1396.600	2410.902	.939	1.064
	Sector=2.0	-9224.161	990.671	-.164	-9.311	<.001	-11167.267	-7281.055	.855	1.169
	Sector=3.0	-13709.634	1556.219	-.155	-8.810	<.001	-16762.007	-10657.260	.861	1.161

a. Dependent Variable: Income

Question Five

Before

Classification Table[a,b]

			Predicted		
			LoanForgive		Percentage Correct
	Observed		0	1	
Step 0	LoanForgive	0	0	574	.0
		1	0	1081	100.0
	Overall Percentage				65.3

a. Constant is included in the model.

b. The cut value is .500

After

Classification Table[a]

Observed			Predicted		
			LoanForgive		Percentage Correct
			0	1	
Step 1	LoanForgive	0	322	252	56.1
		1	136	945	87.4
	Overall Percentage				76.6

a. The cut value is .500

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	516.588	1	<.001
	Block	516.588	1	<.001
	Model	516.588	1	<.001

Yes, a model that uses household income is a good model to predict the likelihood that the household has applied.

The predictive probability that the household has applied for loan forgiveness if the

household has an annual income of $50,000 equals 83%, $75,000 equals 58% and $100,000 is 27%.

Interpretation – For every 1unit increase in annual income, there is a decrease in the likelihood for loan forgiveness

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Income	-.0000518	.000	341.270	1	<.001	1.000
	Constant	4.205	.210	402.403	1	<.001	67.031

a. Variable(s) entered on step 1: Income.