

Multivariate Statistical Techniques Assignment 2 2024

Mashiane Tshepiso 22692037

July 10, 2024

Question 1

a.

Box's Test of Equality of Covariance Matrices ^a	
Box's M	520.920
F	3.317
df1	138
df2	10088.224
Sig.	<.001
Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.	
a. Design: savngs + prpownr + emptye	

Multivariate Tests ^a						
Effect		Value	F	Hypothesis df	Error df	Sig.
savngs	Pillai's Trace	.016	1.985	8.000	1978.000	.045
	Wilks' Lambda	.984	1.988 ^b	8.000	1976.000	.044
	Hotelling's Trace	.016	1.990	8.000	1974.000	.044
	Roy's Largest Root	.014	3.460 ^c	4.000	989.000	.008
prpownr	Pillai's Trace	.080	13.672	6.000	1978.000	<.001
	Wilks' Lambda	.921	13.922 ^b	6.000	1976.000	<.001
	Hotelling's Trace	.086	14.172	6.000	1974.000	<.001
	Roy's Largest Root	.084	27.683 ^c	3.000	989.000	<.001
emptye	Pillai's Trace	.091	15.661	6.000	1978.000	<.001
	Wilks' Lambda	.910	15.842 ^b	6.000	1976.000	<.001
	Hotelling's Trace	.097	16.022	6.000	1974.000	<.001
	Roy's Largest Root	.084	27.718 ^c	3.000	989.000	<.001

a. Design: savngs + prpownr + emptye

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Levene's Test of Equality of Error Variances^a

	F	df1	df2	Sig.
Duration in months	1.990	63	936	<.001
Credit amount	4.220	63	936	<.001

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: savngs + prpownr + emptytype

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	Duration in months	454084.023 ^a	11	41280.366	318.654	<.001
	Credit amount	1.212E+10 ^b	11	1102045112	166.693	<.001
savngs	Duration in months	1026.736	4	256.684	1.981	.095
	Credit amount	89551280.2	4	22387820.1	3.386	.005
prpownr	Duration in months	8791.476	3	2930.492	22.621	<.001
	Credit amount	425680848	3	141893616	21.463	<.001
emptytype	Duration in months	2595.055	3	865.018	6.677	<.001
	Credit amount	508476051	3	169492017	25.637	<.001
Error	Duration in months	128120.977	989	129.546		
	Credit amount	6538508295	989	6611231.845		
Total	Duration in months	582205.000	1000			
	Credit amount	1.866E+10	1000			

a. R Squared = .780 (Adjusted R Squared = .777)

b. R Squared = .650 (Adjusted R Squared = .646)

Post Hoc Tests

Saving plus bonds

Multiple Comparisons

Tukey HSD

Dependent Variable	(i) Saving plus bonds	(j) Saving plus bonds	Mean Difference (i-j)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Duration in months	LT 100DM	100 - 500DM	-2.2967	1.21349	.322	-5.6130	1.0195
		500-IK DM	1.4094	1.50702	.883	-2.7090	5.5278
		> 1K DM	2.1286	1.70696	.724	-2.5362	6.7934
		UNK/NA	-2.2747	.96059	.125	-4.8998	.3504
	100 - 500DM	LT 100DM	2.2967	1.21349	.322	-1.0195	5.6130
		500-IK DM	3.7061	1.82044	.250	-1.2688	8.6811
		> 1K DM	4.4254	1.98912	.171	-1.0105	9.8613
		UNK/NA	.0220	1.40201	1.000	-3.8094	3.8535
	500-IK DM	LT 100DM	-1.4094	1.50702	.883	-5.5278	2.7090
		100 - 500DM	-3.7061	1.82044	.250	-8.6811	1.2688
		> 1K DM	.7192	2.18063	.997	-5.2400	6.6785
		UNK/NA	-3.6841	1.66258	.175	-8.2276	.8594
	> 1K DM	LT 100DM	-2.1286	1.70696	.724	-6.7934	2.5362
		100 - 500DM	-4.4254	1.98912	.171	-9.8613	1.0105
		500-IK DM	-.7192	2.18063	.997	-6.6785	5.2400
		UNK/NA	-4.4033	1.84575	.120	-9.4474	.6407
	UNK/NA	LT 100DM	2.2747	.96059	.125	-.3504	4.8998
		100 - 500DM	-.0220	1.40201	1.000	-3.8535	3.8094
		500-IK DM	3.6841	1.66258	.175	-.8594	8.2276
		> 1K DM	4.4033	1.84575	.120	-.6407	9.4474
Credit amount	LT 100DM	100 - 500DM	-196.2063	274.13613	.953	-945.3699	552.9573
		500-IK DM	615.7214	340.44689	.369	-314.6573	1546.1001
		> 1K DM	614.4367	385.61366	.502	-439.3745	1668.2479
		UNK/NA	-718.5773 [*]	217.00426	.009	-1311.6101	-125.5446
	100 - 500DM	LT 100DM	196.2063	274.13613	.953	-552.9573	945.3699
		500-IK DM	811.9277	411.25042	.279	-311.9440	1935.7994
		> 1K DM	810.6430	449.35589	.372	-417.3640	2038.6500
		UNK/NA	-522.3710	316.72333	.466	-1387.9176	343.1756
	500-IK DM	LT 100DM	-615.7214	340.44689	.369	-1546.1001	314.6573
		100 - 500DM	-811.9277	411.25042	.279	-1935.7994	311.9440
		> 1K DM	-1.2847	492.61972	1.000	-1347.5238	1344.9543
		UNK/NA	-1334.2987 [*]	375.58905	.004	-2360.7145	-307.8829
	> 1K DM	LT 100DM	-614.4367	385.61366	.502	-1668.2479	439.3745
		100 - 500DM	-810.6430	449.35589	.372	-2038.6500	417.3640
		500-IK DM	1.2847	492.61972	1.000	-1344.9543	1347.5238
		UNK/NA	-1333.0140 [*]	416.96636	.012	-2472.5063	-193.5217
	UNK/NA	LT 100DM	718.5773 [*]	217.00426	.009	125.5446	1311.6101
		100 - 500DM	522.3710	316.72333	.466	-343.1756	1387.9176
		500-IK DM	1334.2987 [*]	375.58905	.004	307.8829	2360.7145
		> 1K DM	1333.0140 [*]	416.96636	.012	193.5217	2472.5063

Based on observed means.
The error term is Mean Square(Error) = 6611231.845.

*. The mean difference is significant at the .05 level.

Homogeneous Subsets

Duration in months

Tukey HSD^{a,b,c}

Saving plus bonds	N	Subset 1
> 1K DM	48	18.3125
500-1K DM	63	19.0317
LT 100DM	603	20.4411
UNK/NA	183	22.7158
100 - 500DM	103	22.7379
Sig.		.061

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 129.546.

- Uses Harmonic Mean Sample Size = 93.392.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- Alpha = .05.

Credit amount

Tukey HSD^{a,b,c}

Saving plus bonds	N	1	2
500-1K DM	63	2572.1111	
> 1K DM	48	2573.3958	
LT 100DM	603	3187.8325	3187.8325
100 - 500DM	103	3384.0388	3384.0388
UNK/NA	183		3906.4098
Sig.		.197	.313

Means for groups in homogeneous subsets are displayed.
Based on observed means.
The error term is Mean Square(Error) = 6611231.845.

- Uses Harmonic Mean Sample Size = 93.392.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- Alpha = .05.

Property owner

Multiple Comparisons

Tukey HSD

Dependent Variable	(I) Property owner	(J) Property owner	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
Duration in months	Real Estate	BLD SAV/LIFE INS	-3.4219*	1.00885	.004	-6.0181	-.8257
		CAR OR OTHER	-6.7082*	.92173	<.001	-9.0802	-4.3362
		Unknown/None	-10.6780*	1.14043	<.001	-13.6128	-7.7432
	BLD SAV/LIFE INS	Real Estate	3.4219*	1.00885	.004	.8257	6.0181
		CAR OR OTHER	-3.2863*	.97395	.004	-5.7927	-.7799
		Unknown/None	-7.2561*	1.18305	<.001	-10.3006	-4.2116
	CAR OR OTHER	Real Estate	6.7082*	.92173	<.001	4.3362	9.0802
		BLD SAV/LIFE INS	3.2863*	.97395	.004	.7799	5.7927
		Unknown/None	-3.9698*	1.10969	.002	-6.8255	-1.1141
	Unknown/None	Real Estate	10.6780*	1.14043	<.001	7.7432	13.6128
		BLD SAV/LIFE INS	7.2561*	1.18305	<.001	4.2116	10.3006
		CAR OR OTHER	3.9698*	1.10969	.002	1.1141	6.8255
Credit amount	Real Estate	BLD SAV/LIFE INS	-950.8621*	227.90519	<.001	-1537.3543	-364.3699
		CAR OR OTHER	-1420.8403*	208.22448	<.001	-1956.6861	-884.9946
		Unknown/None	-2764.0186*	257.63183	<.001	-3427.0094	-2101.0277
	BLD SAV/LIFE INS	Real Estate	950.8621*	227.90519	<.001	364.3699	1537.3543
		CAR OR OTHER	-469.9782	220.02281	.142	-1036.1859	96.2294
		Unknown/None	-1813.1565*	267.25786	<.001	-2500.9190	-1125.3939
	CAR OR OTHER	Real Estate	1420.8403*	208.22448	<.001	884.9946	1956.6861
		BLD SAV/LIFE INS	469.9782	220.02281	.142	-96.2294	1036.1859
		Unknown/None	-1343.1782*	250.68591	<.001	-1988.2944	-698.0620
	Unknown/None	Real Estate	2764.0186*	257.63183	<.001	2101.0277	3427.0094
		BLD SAV/LIFE INS	1813.1565*	267.25786	<.001	1125.3939	2500.9190
		CAR OR OTHER	1343.1782*	250.68591	<.001	698.0620	1988.2944

Based on observed means.
The error term is Mean Square(Error) = 6611231.845.

*. The mean difference is significant at the .05 level.

Homogeneous Subsets

Duration in months

Tukey HSD^{a,b,c}

Property owner	N	Subset			
		1	2	3	4
Real Estate	282	16.2376			
BLD SAV/LIFE INS	232		19.6595		
CAR OR OTHER	332			22.9458	
Unknown/None	154				26.9156
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

Based on observed means.

The error term is Mean Square(Error) = 129.546.

- Uses Harmonic Mean Sample Size = 230.388.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- Alpha = .05.

Credit amount

Tukey HSD^{a,b,c}

Property owner	N	Subset		
		1	2	3
Real Estate	282	2153.2801		
BLD SAV/LIFE INS	232		3104.1422	
CAR OR OTHER	332		3574.1205	
Unknown/None	154			4917.2987
Sig.		1.000	.203	1.000

Means for groups in homogeneous subsets are displayed.

Based on observed means.

The error term is Mean Square(Error) = 6611231.845.

- Uses Harmonic Mean Sample Size = 230.388.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- Alpha = .05.

Employment type

Multiple Comparisons

Tukey HSD

Dependent Variable	(i) Employment type	(j) Employment type	Mean Difference (i-j)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Duration in months	UEMP / UNSKL NR	UNSKL RESIDENT	.8286	2.55660	.988	-5.7505	7.4078
		SKL EMP/OFFICIAL	-4.0475	2.46862	.357	-10.4002	2.3053
		MGT / SELF-EMP	-7.8053 [*]	2.60072	.015	-14.4980	-1.1126
	UNSKL RESIDENT	UEMP / UNSKL NR	-.8286	2.55660	.988	-7.4078	5.7505
		SKL EMP/OFFICIAL	-4.8761 [*]	.92377	<.001	-7.2534	-2.4989
		MGT / SELF-EMP	-8.6339 [*]	1.23412	<.001	-11.8098	-5.4580
	SKL EMP/OFFICIAL	UEMP / UNSKL NR	4.0475	2.46862	.357	-2.3053	10.4002
		UNSKL RESIDENT	4.8761 [*]	.92377	<.001	2.4989	7.2534
		MGT / SELF-EMP	-3.7578 [*]	1.03968	.002	-6.4333	-1.0823
	MGT / SELF-EMP	UEMP / UNSKL NR	7.8053 [*]	2.60072	.015	1.1126	14.4980
		UNSKL RESIDENT	8.6339 [*]	1.23412	<.001	5.4580	11.8098
		SKL EMP/OFFICIAL	3.7578 [*]	1.03968	.002	1.0823	6.4333
Credit amount	UEMP / UNSKL NR	UNSKL RESIDENT	386.6164	577.55233	.909	-1099.6593	1872.8921
		SKL EMP/OFFICIAL	-325.8287	557.67783	.937	-1760.9593	1109.3018
		MGT / SELF-EMP	-2690.3569 [*]	587.52108	<.001	-4202.2862	-1178.4276
	UNSKL RESIDENT	UEMP / UNSKL NR	-386.6164	577.55233	.909	-1872.8921	1099.6593
		SKL EMP/OFFICIAL	-712.4451 [*]	208.68679	.004	-1249.4806	-175.4096
		MGT / SELF-EMP	-3076.9732 [*]	278.79499	<.001	-3794.4255	-2359.5210
	SKL EMP/OFFICIAL	UEMP / UNSKL NR	325.8287	557.67783	.937	-1109.3018	1760.9593
		UNSKL RESIDENT	712.4451 [*]	208.68679	.004	175.4096	1249.4806
		MGT / SELF-EMP	-2364.5282 [*]	234.87125	<.001	-2968.9469	-1760.1095
	MGT / SELF-EMP	UEMP / UNSKL NR	2690.3569 [*]	587.52108	<.001	1178.4276	4202.2862
		UNSKL RESIDENT	3076.9732 [*]	278.79499	<.001	2359.5210	3794.4255
		SKL EMP/OFFICIAL	2364.5282 [*]	234.87125	<.001	1760.1095	2968.9469

Based on observed means.

The error term is Mean Square(Error) = 6611231.845.

*. The mean difference is significant at the .05 level.

Homogeneous Subsets

Duration in months

Tukey HSD^{a,b,c}

Employment type	N	Subset	
		1	2
UNSKL RESIDENT	200	16.5350	
UEMP / UNSKL NR	22	17.3636	
SKL EMP/OFFICIAL	630	21.4111	21.4111
MGT / SELF-EMP	148		25.1689
Sig.		.061	.218

Means for groups in homogeneous subsets are displayed.

Based on observed means.

The error term is Mean Square(Error) = 129.546.

- Uses Harmonic Mean Sample Size = 68.029.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- Alpha = .05.

Credit amount

Tukey HSD^{a,b,c}

Employment type	N	Subset	
		1	2
UNSKL RESIDENT	200	2358.5200	
UEMP / UNSKL NR	22	2745.1364	
SKL EMP/OFFICIAL	630	3070.9651	
MGT / SELF-EMP	148		5435.4932
Sig.		.370	1.000

Means for groups in homogeneous subsets are displayed.

Based on observed means.

The error term is Mean Square(Error) = 6611231.845.

- Uses Harmonic Mean Sample Size = 68.029.
- The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- Alpha = .05.

b. Interpretations

Box's test

Since the associated p-value for the Box's M Test is <0.001 , which is less than the desired significance level of 0.05, this suggests that homogeneity of covariances is not met

Multivariate Tests:

For the "savings plus bond" effect, the Pillai's Trace statistic is 0.016, the F-statistic is 1.985, and the p-value is 0.045. This indicates that the "savings plus bond" effect is statistically significant at the 5% Sig.level.

For the "property owner" effect, the Pillai's Trace statistic is 0.080, the F-statistic is 13.672, and the p-value is less than 0.001. This indicates that the "property owner" effect is statistically significant at the 5% Sig.level.

For the "employment type" effect, the Pillai's Trace statistic is 0.091, the F-statistic is 15.661, and the p-value is less than 0.001. This indicates that the "employment type" effect is statistically significant at the 5% Sig.level.

Levene's Test of Equality of Error Variances.

For the "Duration in months" variable, the F-statistic is 1.990 with 63 and 936 degrees of freedom, and the p-value is less than 0.001. For the "Credit amount" variable, the F-statistic is 4.220 with 63 and 936 degrees of freedom, and the p-value is less than 0.001. The low p-values (less than 0.05) for both variables indicate that we reject the null hypothesis of equal error variances across groups and conclude that the error variances are not equal across the groups defined by the independent variables("Duration" and "Credit amount").

Tests between subjects effects:

For the "Duration in months" dependent variable the p-values (Sig. column) for the Model, savings, property owner, and employment type effects are all less than 0.001, indicating that these effects are statistically significant at the 0.05 level of significance

For the "Credit amount" dependent variable the p-values (Sig. column) for the Model, savings, property owner, and employment type effects are all less than 0.001, indicating that these effects are statistically significant at the 0.05 level of significance

The low p-values (less than 0.05) for the majority of the effects suggest that the independent variables("savings", "property owner", and "employment type") included in the model have a significant impact on both dependent variables ("Duration in months" and "Credit amount").

Post Hoc Test: Saving Plus Bond

a.Multiple comparison:saving plus bond

For the "Duration in months" dependent variable the mean differences between various groups of "Saving plus bonds" are statistically insignificant, as indicated by the $p\text{-values} > 0.05$ (Sig. column). The 95% confidence intervals for the mean differences include zero, further confirming the statistical insignificance of the differences.

However for the "Credit amount" dependent variable to the "Duration in months" variable, the mean differences between various groups of "Saving plus bonds" are statistically significant, as indicated by the low p-values. And some of the 95% confidence intervals for the mean differences do not include zero, further confirming the statistical significance of the differences.

b.Homogenous subset/duration in months: saving plus bonds

The group "> 1K DM" has a mean "Duration in months" of 18.3125, which is significantly different from the other groups. The groups "500-1K DM", "LT 100DM", and "UNK/NA" form a homogeneous subset, meaning they are not significantly different from each other. The group "100 - 500DM" has a mean "Duration in months" of 22.7379, which is significantly different from the other groups.

c.Homogenous subset/credit amount: saving plus bonds

The group "> 1K DM" and the group "500-1K DM" form Subset 1, with a mean "Credit amount" of 2572.1111 and 2573.3958, respectively. The group "LT 100DM", the group "100 - 500DM", and the group "UNK/NA" form Subset 2, with a mean "Credit amount" of 3187.8325, 3384.0388, and 3906.4098, respectively.

Post Hoc Test: Property owner

a.Multiple comparison:Property owner

For the "Duration in months" dependent variable the mean differences in "Duration in months" between the "Real Estate" group and the "BLD SAV/LIFE INS" and "CAR OR OTHER" groups are statistically significant, as indicated by the low p-values (Sig. column). As some of the 95% confidence intervals for these mean differences do not include zero, further confirming the statistical significance of the differences.

For the "Credit amount" dependent variable the mean differences in "Credit amount" between the "Real Estate" group and the "BLD SAV/LIFE INS" and "CAR OR OTHER" groups are statistically significant, as indicated by the low p-values. The 95% confidence intervals for these mean differences do not include zero, further confirming the statistical significance of the differences.

b.Homogenous subset/duration in months:property owner

The "Real Estate" group has the highest mean "Duration in months" of 16.2376 and is in Subset 1 on its own. The "BLD SAV/LIFE INS" group has a mean "Duration in months" of 19.6595 and is in Subset 2/

The "CAR OR OTHER" group has a mean "Duration in months" of 22.9458 and is in Subset 3. The "Unknown/None" group has the highest mean "Duration in months" of 26.9156 and is in Subset 4.

These results indicate that there are statistically significant differences in the "Duration in months" variable between the different "Property owner" groups.

c.Homogenous subset/credit amount: property owner

The "Real Estate" group has the lowest mean "Credit amount" of 2153.2801 and is in Subset 1.

The "BLD SAV/LIFE INS" group has a mean "Credit amount" of 3104.1422 and the "CAR OR OTHER" group has a mean "Credit amount" of 3574.1205, both of which are in Subset 2. The "Unknown/None" group has the highest mean "Credit amount" of 4917.2987 and is also in Subset 2.

These results indicate that there are statistically significant differences in the "Credit amount" variable between the "Real Estate" group and the other groups ("BLD SAV/LIFE INS", "CAR OR OTHER", and "Unknown/None").

Post Hoc Test: Employment type

a. Multiple comparison: Employment type

For the "Duration in months" dependent variable: The mean differences in "Duration in months" between several pairs of "Employment type" groups are statistically significant, as indicated by the low p-values (Sig. column). The 95% confidence intervals for these mean differences do not include zero, further confirming the statistical significance of the differences.

For the "Credit amount" dependent variable: The mean differences in "Credit amount" between multiple pairs of "Employment type" groups are also statistically significant, as indicated by the low p-values. The 95% confidence intervals for these mean differences do not include zero, further confirming the statistical significance of the differences. These results suggest that there are significant differences in both "Duration in months" and "Credit amount" between various "Employment type" groups.

b. Homogenous subset/duration in months: Employment type

The "UNEMP/ UNSKL NR" group has the lowest mean "Duration in months" of 16.5350 and is in Subset 1. The "UNEMP/ UNSKL NR" group and the "MGT / SELF-EMP" group are in Subset 1.

The "UNSKL RESIDENT" group and the "SKL EMP/OFFICIAL" group are in Subset 2, with higher mean "Duration in months" of 21.4111 and 25.1689, respectively. These results indicate that there are statistically significant differences in the "Duration in months" variable between the "UNEMP/ UNSKL NR" group and the "UNSKL RESIDENT" and "SKL EMP/OFFICIAL" groups.

c. Homogenous subset/credit amount: Employment type

The "UNEMP/ UNSKL NR" group has the lowest mean "Credit amount" of 2745.1364 and is in Subset 1. The "UNEMP/ UNSKL NR" group and the "MGT / SELF-EMP" group are in Subset 1.

The "UNSKL RESIDENT" group and the "SKL EMP/OFFICIAL" group are in Subset 2, with higher mean "Credit amount" of 3070.9651 and 5435.4932, respectively. These results indicate that there are statistically significant differences in the "Credit amount" variable between the "UNEMP/ UNSKL NR" group and the "UNSKL RESIDENT" and "SKL EMP/OFFICIAL" groups.

Question 2

a.

After performing the main effect only multivariate analysis of variance (MANOVA) using the Income before the program and Income after the program as dependent variables, and program status and level of education as independent variables, here are the outputs that includes the Homogeneity test.

General Linear Model

Between-Subjects Factors

	Value	Label	N
Program status	0		517
	1		483
Level of education	1	Did not complete high school	459
	2	High school degree	348
	3	Some college	193

Box's Test of Equality of Covariance Matrices^a

Box's M	78.420
F	5.197
df1	15
df2	1798542.139
Sig.	<.001

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: ed + prog

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.
ed	Pillai's Trace	.743	294.417	4.000	1992.000	<.001
	Wilks' Lambda	.258	482.595 ^b	4.000	1990.000	<.001
	Hotelling's Trace	2.878	715.222	4.000	1988.000	<.001
	Roy's Largest Root	2.877	1432.820 ^c	2.000	996.000	<.001
prog	Pillai's Trace	.333	247.819 ^b	2.000	995.000	<.001
	Wilks' Lambda	.667	247.819 ^b	2.000	995.000	<.001
	Hotelling's Trace	.498	247.819 ^b	2.000	995.000	<.001
	Roy's Largest Root	.498	247.819 ^b	2.000	995.000	<.001

a. Design: ed + prog

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Levene's Test of Equality of Error Variances^a

	F	df1	df2	Sig.
Income before the program	8.108	5	994	<.001
Income after the program	4.492	5	994	<.001

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: ed + prog

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	Income before the program	82160.596 ^a	4	20540.149	29674.908	<.001
	Income after the program	285776.166 ^b	4	71444.042	6272.306	<.001
ed	Income before the program	1983.514	2	991.757	1432.818	<.001
	Income after the program	5311.620	2	2655.810	233.162	<.001
prog	Income before the program	.001	1	.001	.001	.971
	Income after the program	4721.929	1	4721.929	414.554	<.001
Error	Income before the program	689.404	996	.692		
	Income after the program	11344.834	996	11.390		
Total	Income before the program	82850.000	1000			
	Income after the program	297121.000	1000			

a. R Squared = .992 (Adjusted R Squared = .992)

b. R Squared = .962 (Adjusted R Squared = .962)

b.

Since the Box's Test of Equality of Covariance Matrices shows that we reject the null hypothesis (Sig. < 0.001), and conclude that homogeneity of covariance matrices is not met, it is necessary do the appropriate transformation(s) of the dependent variables.

Log Transformation

Between-Subjects Factors

		Value Label	N
Program status	0		517
	1		483
Level of education	1	Did not complete high school	459
	2	High school degree	348
	3	Some college	193

Box's Test of Equality of Covariance Matrices^a

Box's M	76.473
F	5.068
df1	15
df2	1798542.139
Sig.	<.001

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: ed + prog

Multivariate Tests ^a						
Effect		Value	F	Hypothesis df	Error df	Sig.
ed	Pillai's Trace	.722	281.353	4.000	1992.000	<.001
	Wilks' Lambda	.279	443.667 ^b	4.000	1990.000	<.001
	Hotelling's Trace	2.574	639.576	4.000	1988.000	<.001
	Roy's Largest Root	2.572	1280.733 ^c	2.000	996.000	<.001
prog	Pillai's Trace	.354	272.660 ^b	2.000	995.000	<.001
	Wilks' Lambda	.646	272.660 ^b	2.000	995.000	<.001
	Hotelling's Trace	.548	272.660 ^b	2.000	995.000	<.001
	Roy's Largest Root	.548	272.660 ^b	2.000	995.000	<.001
a. Design: ed + prog b. Exact statistic c. The statistic is an upper bound on F that yields a lower bound on the significance level.						

Levene's Test of Equality of Error Variances^a

	F	df1	df2	Sig.
LogTrans_Income before the program	11.918	5	994	<.001
LogTrans_Income after the program	3.871	5	994	.002

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: ed + prog

Tests of Between-Subjects Effects						
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	LogTrans_Income before the program	897.361 ^a	4	224.340	132158.496	<.001
	LogTrans_Income after the program	1454.145 ^b	4	363.536	46478.740	<.001
ed	LogTrans_Income before the program	4.348	2	2.174	1280.711	<.001
	LogTrans_Income after the program	3.402	2	1.701	217.507	<.001
prog	LogTrans_Income before the program	2.787E-5	1	2.787E-5	.016	.898
	LogTrans_Income after the program	3.516	1	3.516	449.548	<.001
Error	LogTrans_Income before the program	1.691	996	.002		
	LogTrans_Income after the program	7.790	996	.008		
Total	LogTrans_Income before the program	899.052	1000			
	LogTrans_Income after the program	1461.935	1000			
a. R Squared = .998 (Adjusted R Squared = .998) b. R Squared = .995 (Adjusted R Squared = .995)						

Post Hoc Tests

Level of education

Multiple Comparisons							
Tukey HSD							
Dependent Variable	(I) Level of education	(J) Level of education	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LogTrans_Income before the program	Did not complete high school	High school degree	-.0821 [*]	.00293	< .001	-.0890	-.0752
		Some college	-.1746 [*]	.00353	< .001	-.1829	-.1663
	High school degree	Did not complete high school	.0821 [*]	.00293	< .001	.0752	.0890
		Some college	-.0925 [*]	.00370	< .001	-.1012	-.0839
	Some college	Did not complete high school	.1746 [*]	.00353	< .001	.1663	.1829
		High school degree	.0925 [*]	.00370	< .001	.0839	.1012
LogTrans_Income after the program	Did not complete high school	High school degree	-.0671 [*]	.00629	< .001	-.0819	-.0524
		Some college	-.1629 [*]	.00759	< .001	-.1807	-.1451
	High school degree	Did not complete high school	.0671 [*]	.00629	< .001	.0524	.0819
		Some college	-.0957 [*]	.00794	< .001	-.1144	-.0771
	Some college	Did not complete high school	.1629 [*]	.00759	< .001	.1451	.1807
		High school degree	.0957 [*]	.00794	< .001	.0771	.1144
Based on observed means.							
The error term is Mean Square(Error) = .008.							
*. The mean difference is significant at the .05 level.							

LogTrans_Income after the program

Tukey HSD^{a,b,c}

Level of education	N	Subset		
		1	2	3
Did not complete high school	459	1.1481		
High school degree	348		1.2152	
Some college	193			1.3110
Sig.		1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

Based on observed means.

The error term is Mean Square(Error) = .008.

a. Uses Harmonic Mean Sample Size = 293.153.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

c. Alpha = .05.

LogTrans_Income before the program				
Tukey HSD ^{a,b,c}				
Level of education	N	Subset		
		1	2	3
Did not complete high school	459	.8827		
High school degree	348		.9648	
Some college	193			1.0574
Sig.		1.000	1.000	1.000
Means for groups in homogeneous subsets are displayed.				
Based on observed means.				
The error term is Mean Square(Error) = .002.				
a. Uses Harmonic Mean Sample Size = 293.153.				
b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.				
c. Alpha = .05.				

c.

Box's Test of Equality of covariance matrices:

Observing the Box's Test of Equality of covariance matrices we notice the we notice that though there is a decrease on the value of Box's M and its F-value, our results still indicate that homogeneity of covariance matrices is not met.

Multivariate Tests:

Though there is a slight difference in the multivariate value, the result still remain significant(sig. <.001)

Tests between subjects effects:

In observation of both not transformed and transformed data("Log-Trans Income before the program" and "LogTrans-Income after the program"), the p-values (Sig. column) are all less than 0.001, indicating that the between-subjects effects are statistically significant at the 0.001 level. The high F-statistics and low p-values suggest that the independent variables included in the model have a significant impact on the dependent variables though there is a decrease in the p-value of a transformed "Income before the program" associate with "program status" and the Type III Sum of Squares have be transformed from larges values to small reasonable values though that does no change the results entirely

In overall

Though results of our analysis has changed the conclusion have not. Thefore the data violates homogeneity of covariance matrices.

d.

Box's test

Since the associated p-value for the Box's M Test is < 0.001 , which is less than the desired significance level of 0.05, this suggests that homogeneity of covariances is not met

Multivariate Tests:

For both the "educationa level" and "program status" independent variables, all the multivariate test statistics (Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root) have very low p-values ($p < 0.001$), indicating that the effects of these variables on the dependent variables are statistically significant. The high F-statistics and low p-values suggest that the independent variables included in the model have a significant impact on the overall set of dependent variables.

Levene's Test of Equality of Error Variances.

For the "LogTrans-Income before the program" variable, the F-statistic is 11.918 with 5 and 994 degrees of freedom, and the p-value is less than 0.001. For the "LogTrans-Income after the program" variable, the F-statistic is 3.871 with 5 and 994 degrees of freedom, and the p-value is 0.002. The low p-values (less than 0.05) for both variables indicate that we reject the null hypothesis of equal error variances across groups. In other words, the error variances are not equal across the groups defined by the independent variables (transformed Income before the program and Income after the program).

Tests between subjects effects:

For both "LogTrans-Income before the program" and "LogTrans-Income after the program", the p-values (Sig. column) are all less than 0.001, indicating that the between-subjects effects are statistically significant at the 0.001 level. The high F-statistics and low p-values suggest that the independent variables included in the model ("level of education" and "program status") have a significant impact on the dependent variables (LogTrans-Income before and after the program).

post Hoc test:

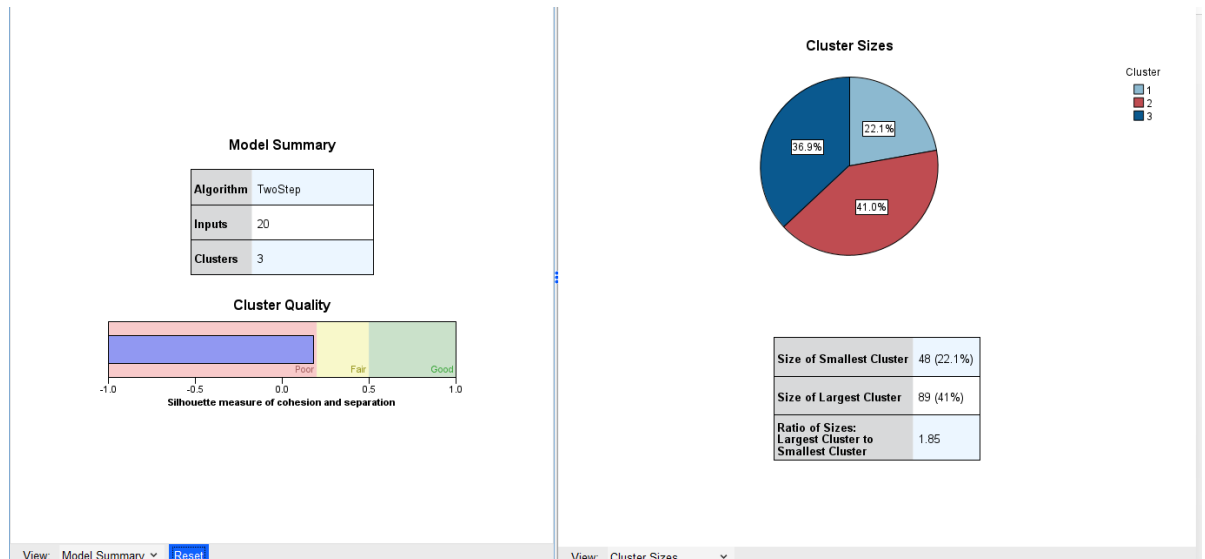
The Tukey HSD test indicates that there are three homogeneous subsets (Subset 1, 2, and 3) for the "Level of education" variable. The "Did not complete high school" group has a mean of 7.6776, and it is in Subset 1, which is significantly different from the other two groups. The "High school degree" group has a mean of 9.2500, and it is in Subset 2, which is significantly different from the "Did not complete high school" group (Subset 1) but not significantly different from the "Some college" group (Subset 3). The "Some college" group has a mean of 11.4560, and it is in Subset 3, which is significantly different from the "Did not complete high school" group (Subset 1) but not significantly different from the "High school degree" group (Subset 2).

For "Income before the program" the mean difference between "Did not complete high school" and "High school degree" is statistically significant ($p < 0.001$). The mean difference between "High school degree" and "Some college" is also statistically significant ($p < 0.001$). The mean difference between "Did not complete high school" and "Some college" is statistically significant ($p < 0.001$).

For "Income after the program" the mean differences between all pairwise comparisons of the "Level of education" groups are statistically significant ($p < 0.001$).

Question 3

(i)



Number of Clusters: 3 clusters

Cluster Quality poor, based on the Silhouette measure.

(ii)

➔ Frequencies

Statistics		
TwoStep Cluster Number		
N	Valid	217
	Missing	0

TwoStep Cluster Number					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	48	22.1	22.1	22.1
	2	89	41.0	41.0	63.1
	3	80	36.9	36.9	100.0
	Total	217	100.0	100.0	

Cluster Distribution:

Cluster 1: 48 cases (22.1%)

Cluster 2: 89 cases (41.0%)

Cluster 3: 80 cases (39.9%)