Economics 224-001/002
Assignment 4
ANSWER KEY

1. The file has 3 immigrant categories for each individual: permanent residents who are non-immigrants, permanent residents who are immigrants, and non-permanent residents. I think we do not want the non-permanent residents in our data set. Please clean up the data set to remove these individuals from the data set. Explain carefully how you did it, and how many individuals are left in the data set. Please provide the full set of summary statistics on the remaining individual for all 6 variables in the data set.

You need to use the sort function, and sort via the IMMPOPP column. Then, the category IMMPOPP=3 (nonpermanent residents) will be the last set of observations, and you can just delete these. This leaves you with 1254 observations, with the following summary stats:

AGEP		*SEXP*		*IMMPOPP*	
Mean	41.00638	Mean	1.522329	Mean	1.107656
Standard Error	0.286008	Standard Error	0.014111	Standard Error	0.008756
Median	41	Median	2	Median	1
Mode	45	Mode	2	Mode	1
Standard Deviation	10.12807	Standard Deviation	0.4997	Standard Deviation	0.310068
Sample Variance	102.5778	Sample Variance	0.249701	Sample Variance	0.096142
Kurtosis	-0.99148	Kurtosis	-1.99517	Kurtosis	4.431965
Skewness	0.017055	Skewness	-0.08951	Skewness	2.534738
Range	42	Range	1	Range	1
Minimum	22	Minimum	1	Minimum	1
Maximum	64	Maximum	2	Maximum	2
Sum	51422	Sum	1909	Sum	1389
Count	1254	Count	1254	Count	1254

YRIMMIG		DGREEP		WAGESP	
Mean	8.405104	Mean	6.597289	Mean	47269.27
Standard Error	0.05033	Standard Error	0.033749	Standard Error	813.6614
Median	9	Median	6	Median	45000
Mode	9	Mode	6	Mode	50000
Standard Deviation	1.782286	Standard Deviation	1.195105	Standard Deviation	28813.26
Sample Variance	3.176542	Sample Variance	1.428277	Sample Variance	8.3E+08
Kurtosis	6.825142	Kurtosis	1.502853	Kurtosis	4.552645
Skewness	-2.86967	Skewness	1.76128	Skewness	1.431326
Range	8	Range	4	Range	199994
Minimum	1	Minimum	6	Minimum	6
Maximum	9	Maximum	10	Maximum	200000
Sum	10540	Sum	8273	Sum	59275661
Count	1254	Count	1254	Count	1254

(5 marks in total)

2. The Minister has speculated that the average person in Saskatchewan with a university degree gets paid at least $45,000 per year. You need to test this for him with your sample.

a) Write out your hypothesis statement.

This is a research hypothesis, so it goes in the alternative:

H_0: $\mu \leq 45{,}000$
H_A: $\mu > 45{,}000$

(3 marks)

b) Using the individuals you have extracted, and using Excel and the appropriate test (see Appendix 9.2 of the text), test your hypothesis using the probability value approach. Please be sure to attach a printout of your Excel results.

The Excel results are below:

Sample Size:	1254
Sample Mean:	47269.27
Sample Std. Dev.:	28813.26
Hypothesized Value:	45000
St. Error	813.6614
Test Statistic t	2.788958
Degrees of Freedom	1253
p-value (lower tail):	0.002684
p-value (upper tail):	0.997316

We can reject the null hypothesis at the 0.3% level. Therefore, the average person in the sample gets paid more than $45,000.

(5 marks)

c) Be sure to write out a simple paragraph summarizing your results for the Minister in non-technical language.

The data in the sample supports the hypothesis that people in Saskatchewan with a bachelor's degree or higher gets paid at least $45,000 on average.

(2 marks) – (10marks in total)

3. In class, we have examined some information on the labour income of immigrants versus nonimmigrants using the SLID data for Alberta. The Minister was interested by those results, but want to see some results for Saskatchewan. Specifically, he wants to see if immigrants to Saskatchewan get paid more than nonimmigrants. Using your cleaned-up dataset from question 1, test whether nonimmigrants make more than immigrants, using the probability value approach.

a) You will need to sort out the immigrants and nonimmigrants first. Explain how you did this.

We have already used the sort function in question 1, and sorted via the IMMPOPP column and removed the nonpermanent residents. This leaves us with 1254 observations, sorted by IMMPOPP. We can now fine the breakpoint between IMMPOPP = 1 (non-immigrants) and IMMPOPP = 2 (immigrants). We can cut and paste the second group (all 6 variables) into a different set of columns.

(3 marks)

b) Write out your hypothesis statement.

We are testing if $\mu_{IMM} > \mu_{NONIMM}$, or if $\mu_{IMM} - \mu_{NONIMM} > 0$. This is a research hypothesis, so it goes in the alternative:

H_0: $\mu_{IMM} - \mu_{NONIMM} \leq 0$
H_A: $\mu_{IMM} - \mu_{NONIMM} > 0$

(3 marks)

c) Using the individuals you have sorted, and using Excel and the appropriate test (see Appendix 9.2 of the text), test your hypothesis using the probability value approach. Please be sure to attach a printout of your Excel results.

We can now use the Excel Data Analysis function "t-Test: Two-Sample Assuming Unequal Variances" on the (renamed) wage variables:

t-Test: Two-Sample Assuming Unequal Variances

	IMMWAGESP	*WAGESP*
Mean	55754.35	46245.6
Variance	1.48E+09	7.43E+08
Observations	135	1119
Hypothesized Mean Difference	0	
df	151	
t Stat	2.788633	
P(T<=t) one-tail	0.002987	
t Critical one-tail	1.655008	
P(T<=t) two-tail	0.005975	
t Critical two-tail	1.9758	

Since our probability value is 0.3%, we reject the null hypothesis and accept the alternative hypothesis that immigrants get paid more.

(7 marks)

d) Be sure to write out a simple paragraph summarizing your results for the Minister in non-technical language.

The data in the sample supports the hypothesis that immigrants to Saskatchewan with at least a bachelor's degree get paid more than non-immigrants on average.

(2 marks) – (15 marks in total)

4. The Minister has heard that immigrants are different than nonimmigrants, specifically that they are more educated than nonimmigrants. Using your sorted data from question 3, test whether immigrants are more educated than nonimmigrants using the probability value approach.

a) Write out your hypothesis statement.

We are testing if $\mu_{EDUCIMM} > \mu_{EDUCNONIMM}$, or if $\mu_{EDUCIMM} - \mu_{EDUCNONIMM} > 0$. This is a research hypothesis, so it goes in the alternative:

H_0: $\mu_{EDUCIMM} - \mu_{EDUCNONIMM} \leq 0$
H_A: $\mu_{EDUCIMM} - \mu_{EDUCNONIMM} > 0$

(3 marks)

b) Explain how you might test this hypothesis with the data at hand. What problems exist with using the method you suggest?

We have some categorical data on the numbers with different university level degrees etc. However, we have 2 problems:
1) We have excluded people with less than a university degree, so we have a biased sample.
2) This is not really data that we can properly take averages with (but I am going to anyways!), we should instead do some more complex analysis. However, using the numerical codes (with implies that I am assuming a linear relationship), we can calculate the average levels of education.
(Up to 10 BONUS Marks if a student comes up with a way of adjusting the data, e.g. giving 4 years of schooling for a Bachelors, 5 for post-degree cert, 6 for medical or masters, 8 for PHD, and then taking average years of schooling.)

(3 marks)

c) Using the individuals you have sorted, and using Excel and the appropriate test (see Appendix 9.2 of the text), test your hypothesis using the probability value approach and the method you suggested in b). Please be sure to attach a printout of your Excel results.

We can now use the Excel Data Analysis function "t-Test: Two-Sample Assuming Unequal Variances" on the two DGREEP variables:

t-Test: Two-Sample Assuming Unequal Variances

	IMMDGREEP	*DGREEP*
Mean	7.362963	6.504915
Variance	2.621006	1.207266
Observations	135	1119
Hypothesized Mean Difference	0	
df	149	
t Stat	5.993782	
P(T<=t) one-tail	7.4E-09	
t Critical one-tail	1.655144	
P(T<=t) two-tail	1.48E-08	
t Critical two-tail	1.976014	

Since our probability value is 0.000000007%, we reject the null hypothesis and accept the alternative hypothesis that immigrants have higher average education levels.

(7 marks)

d) Be sure to write out a simple paragraph summarizing your results for the Minister in non-technical language.

The data in the sample supports the hypothesis that immigrants to Saskatchewan with a university education have on average higher levels of university education compared to non-immigrants with a university education.

(2 marks) – (15 marks in total)

5. One of the factors we have discussed in class is how men make more than women among all full-time workers. The Minister wants to know if similar results hold for male and female immigrants.

a) Go back to the original cleaned-up data set. Separate out the immigrants. Sort the resulting immigrants into males and females Explain how you did both of these steps.

You need to use the sort function, and sort via the SEXP column. We can now fine the breakpoint between SEXP = 1 (females) and SEXP= 2 (males). We can cut and paste the second group (all 6 variables) into a different set of columns.

(3 marks)

b) Write out your hypothesis statement.

We are testing if $\mu_{MALEIMM} > \mu_{FEMIMM}$, or if $\mu_{MALEIMM} - \mu_{FEMIMM} > 0$. This is a research hypothesis, so it goes in the alternative:

H_0: $\mu_{MALEIMM} - \mu_{FEMIMM} \leq 0$
H_A: $\mu_{MALEIMM} - \mu_{FEMIMM} > 0$

(3 marks)

c) Using your cleaned, sorted data, test whether immigrant men make more than immigrant women using the probability value approach. Please be sure to attach a printout of your Excel results.

We can now use the Excel Data Analysis function "t-Test: Two-Sample Assuming Unequal Variances" on the two IMMWAGE variables:

t-Test: Two-Sample Assuming Unequal Variances

	MALEIMMWAGE	*FEMALEIMMWAGESP*
Mean	65683.62	38327.47
Variance	1.68E+09	6.61E+08
Observations	86	49
Hypothesized Mean Difference	0	
df	132	
t Stat	4.755853	
P(T<=t) one-tail	2.55E-06	
t Critical one-tail	1.656479	
P(T<=t) two-tail	5.1E-06	

Since our probability value is 0.000003%, we reject the null hypothesis and accept the alternative hypothesis that male immigrants have higher average wage levels.

(7 marks)

d) Be sure to write out a simple paragraph summarizing your results for the Minister in non-technical language.

The data in the sample supports the hypothesis that male immigrants to Saskatchewan with university education on average have higher wage levels than female immigrants to Saskatchewan.

(2 marks) – (15 marks in total)