SOCI40003

SOCIAL RESEARCH DESIGN AND EVALUATION

STAGE 2

STATA ANALYSIS

GENDER DIFFERENCES IN THE LABOUR MARKET OUTCOMES IN AUSTRALIA

SHOUZHENG YU

UNIVERSITY OF MELBOURNE

DUE DATE:09/2020

1 First, we assert that some variables are relevant, and my main objective is to find aims to identify the most female education underutilization effects and the drivers behind it.

```
. sum wscei, detail

     WS: Current weekly gross wages & salary - all jobs
                (AUD) [imputed] [weighted top
-------------------------------------------------------------

      Percentiles      Smallest
 1%           0              0
 5%           0              0
10%           0              0       Obs                2,030
25%         330              0       Sum of Wgt.        2,030

50%         800                      Mean            848.7773
                        Largest      Std. Dev.       787.8042
75%        1150           5500
90%        1638           6904       Variance        620635.5
95%        2014           8490       Skewness        5.347448
99%        3350          16456       Kurtosis        85.99773
```

Using summarize, d command to get the average and standard division of income per week. As we can see, the skewness is more than 5 which means the distribution curve of the week's salary is right skew, and the kurtosis is high. The reason there are four 0 values in the smallest distribution is there are many people who lost their jobs and get zero payment at the current time. And in the 2030 samples, the mean of weekly gross inputs is 848.

By using the command hist wscei, freq by(sex), we get the histogram

As considered, that the number of females is higher than males in this sample, we only compare the relative data. Through the graph above, we can see males' relative wages are higher than females. Especially for the higher income part.

It may cause by the higher qualifications present reached by sex being different, thus

```
. tab ed5 sex, col
```

Key
frequency *column percentage*

Highest Education Level Achieved 5	Sex [1] Male	[2] Femal	Total
PhD/Masters	167 20.62	235 19.26	402 19.80
GradDip/Cert	135 16.67	207 16.97	342 16.85
Bachelor/Honours	508 62.72	778 63.77	1,286 63.35
Total	810 100.00	1,220 100.00	2,030 100.00

Using *tabed5 sex, col* to see the presence of each level that a male reached.

As compared to the per cent, a male has a higher presence rather than a female to get Ph.D./master's level.20.6>19.2. That may illustrate why males' average income is higher than females But, due to the asymmetry in the ratio of males and females in this data, I need to consider whether the unemployment rate is different for males and females.

```
. tab sex esbrd, row
```

Key
frequency *row percentage*

Sex	ES: Current labour force status - broad [1] Emplo	[2] Unemp	[3] Not i	Total
[1] Male	712	45	53	810
	87.90	5.56	6.54	100.00
[2] Female	1,039	48	133	1,220
	85.16	3.93	10.90	100.00
Total	1,751	93	186	2,030
	86.26	4.58	9.16	100.00

As a result, saying, the male unemployment rate is higher than females, however, a lot of females are not belonging to the labour force which leads to the employment rate being lower. It may result from the specific social roles of females and mothers; a mother will give more time to take care of a child or have more family units compared to males. I will be looking for the data later. The underutilization of female education in the HILDA survey is because a very large proportion of the female non-labour force population is adequately educated and can get full-time jobs. I will elaborate on this in part 2, as I realized the table attached with good variable controls about the time female sacrifices to find a full-time job under variable children or family units. I cannot do a better job with this section because the control for multiple variables is chaos if only using the technique of regression correlation, which would be making conclusions inaccurate. So, I will talk about some descriptive statistics that I consider relevant and interesting tin his part.

By the way, sex is not influenced people's emotions and behaviour to work. So, it's not a driver behind underutilization.

```
. corre sex jomcd jomcsb
(obs=2,030)
```

	sex
sex	1.0000
jomcd	0.0311
jomcsb	0.0588
jomfd	0.0107
jomls	0.0054
jomms	0.0618
jomns	0.0473
jompf	0.0465
jompi	0.0577
jomsf	0.0499
jomus	0.0579
jomwf	0.0613

Although the commanded codebook tells me those variables go with -1 selection or missing value. But it can still represent a weak relationship. The reason I hint that is I saw a lot of people say that.

```
. tab sex
```

Sex	Freq.	Percent	Cum.
[1] Male	810	39.90	39.90
[2] Female	1,220	60.10	100.00
Total	2,030	100.00	

```
. tab urban
```

Urbanity	Freq.	Percent	Cum.
Rural	419	20.64	20.64
Urban	1,611	79.36	100.00
Total	2,030	100.00	

```
. tab sex urban, row
```

Key
frequency *row percentage*

	Urbanity		
Sex	Rural	Urban	Total
[1] Male	150	660	810
	18.52	81.48	100.00
[2] Female	269	951	1,220
	22.05	77.95	100.00
Total	419	1,611	2,030
	20.64	79.36	100.00

```
.
```

Statistically speaking, males prefer to live in urban compared to females. And the reason I tab ‘urban’ is for the final part.

2

```
. tab age_g5
```

Age Group (5)	Freq.	Percent	Cum.
< 25	935	46.06	46.06
25-34	633	31.18	77.24
35-44	328	16.16	93.40
45-54	123	6.06	99.46
55+	11	0.54	100.00
Total	2,030	100.00	

```
. codebook age_g5
```

---

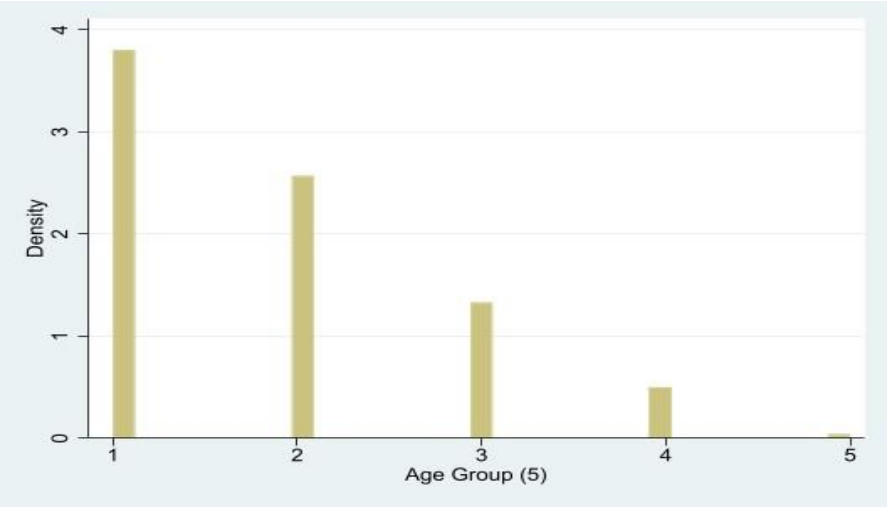
age_g5 — Age Group (5)

```
             type:  numeric (float)
            label:  ag5

            range:  [1,5]                        units:  1
      unique values:  5                      missing .:  0/2,030

     tabulation:  Freq.   Numeric  Label
                    935         1  < 25
                    633         2  25-34
                    328         3  35-44
                    123         4  45-54
                     11         5  55+
```

The reason to use age group 5 is I consider it could be easy to make a graph with wscei (income per week), and the age group in 5 are more averaged and the arrangement of age is more reasonable (10) rather than age7 (5 years as one rank).



Then, we see, in our sample, younger people who are less than 25 are a major percentage, and the sample is more focused on young people. (Less than 40).

```
. tab hlth5
```

Self Reported Health (5)	Freq.	Percent	Cum.
Excellent	394	22.04	22.04
Very Good	822	45.97	68.01
Good	467	26.12	94.13
Fair	95	5.31	99.44
Poor	10	0.56	100.00
Total	1,788	100.00	

```
. codebook hlth5
```

```
hlth5                                                    Self Reported Health (5)
```

```
                type:  numeric (float)
               label:  h5

               range:  [1,5]                        units:  1
        unique values:  5                    missing .:  242/2,030

           tabulation:  Freq.   Numeric  Label
                          394         1  Excellent
                          822         2  Very Good
                          467         3  Good
                           95         4  Fair
                           10         5  Poor
                          242         .
```

Self-reported health could be relevant and correlate to wscei, most people seem to have good health, and it is influenced by the younger sample collected. The very good and good ggroupshave the highest presence.

Then, using the *codebook* command to see the numeric label, which I will use to correlate some variables, in addition, there are 242 missing values. But still can be used in this sample. By using the codebook, we could evaluate the relationship with other variables. Interpret here, the more income, the poorer health may have.

```
. tab hhiu
```

HH: Income unit	Freq.	Percent	Cum.
1	1,501	73.94	73.94
2	430	21.18	95.12
3	85	4.19	99.31
4	13	0.64	99.95
5	1	0.05	100.00
Total	2,030	100.00	

How many income units in one family may influence the income?

```
. tab hhiu sex, col
```

Key
frequency *column percentage*

HH: Income unit	Sex [1] Male	[2] Femal	Total
1	580 71.60	921 75.49	1,501 73.94
2	194 23.95	236 19.34	430 21.18
3	31 3.83	54 4.43	85 4.19
4	5 0.62	8 0.66	13 0.64
5	0 0.00	1 0.08	1 0.05
Total	810 100.00	1,220 100.00	2,030 100.00

Here I am trying to see if females have fewer units of income at home than males. This would prove that females have more elasticity for working compared to males, to reduce the number of working hours. It turns out that this idea is correct. 24% of males have another unit to share, compared to 19% of females. But still, a lot of variables to influence it. And it cannot perform something exactly.

```
. tab edhigh1
```

ED: Highest education level achieved	Freq.	Percent	Cum.
[1] Postgrad - masters or doctorate	402	19.80	19.80
[2] Grad diploma, grad certificate	342	16.85	36.65
[3] Bachelor or honours	1,286	63.35	100.00
Total	2,030	100.00	

```
. codebook edhigh1

-------------------------------------------------------------------------------
edhigh1                                          ED: Highest education level achieved
-------------------------------------------------------------------------------

                  type:  numeric (byte)
                 label:  REDHIGHB

                 range:  [1,3]                        units:  1
         unique values:  3                        missing .:  0/2,030

            tabulation:  Freq.   Numeric  Label
                           402         1  [1] Postgrad - masters or
                                          doctorate
                           342         2  [2] Grad diploma, grad
                                          certificate
                         1,286         3  [3] Bachelor or honours
```

And the same operation for the ‘edhigh1’, which is the highest-level people reach their qualification oforschool. Also, the codebook command is for the correlation later.

```
. tab edhigh1 age_g5, row
```

Key
frequency *row percentage*

ED: Highest education level achieved	Age Group (5)					
	< 25	25-34	35-44	45-54	55+	Total
[1] Postgrad - master	53	173	120	53	3	402
	13.18	43.03	29.85	13.18	0.75	100.00
[2] Grad diploma, gra	37	148	113	39	5	342
	10.82	43.27	33.04	11.40	1.46	100.00
[3] Bachelor or honou	845	312	95	31	3	1,286
	65.71	24.26	7.39	2.41	0.23	100.00
Total	935	633	328	123	11	2,030
	46.06	31.18	16.16	6.06	0.54	100.00

Using tab edhigh1 age_g5, row to see the Proportion of qualifications obtained by age group. People in groups 25-34 occupy 31.18% of finishing their qualifications. And people who are less than 25, in 46.6%. In the data, we find that as people complete their higher education. In their corresponding age group, the higher the proportion of the educated part of the population. So, based on the previous, the younger the healthier the hypothesis, we believe that the older you are, the worse your health is.

```
. corre hlth5 age_g5 edhigh1
(obs=1,788)
```

	hlth5	age_g5	edhigh1
hlth5	1.0000		
age_g5	0.1048	1.0000	
edhigh1	-0.0326	-0.4690	1.0000

Here, I just wonder how health is related to age and education, thus, using the command, *corre hlth5 age_g5 edhigh1.*

As we see the coefficient of hlth5 and age is 0.104 which indicates that there is a weak relationship exists (below 0.3). Much healthier, at a young age.

And a moderate relationship (above 0.3) between age and education which confirm the speculation above.

From the graph, we can see the average income in the higher age group is higher. Which expresses the same result compared to correlation.

```
. tab hhura
```

HH: ABS unemployment rate in major statistical region (October of interview year	Freq.	Percent	Cum.
1.9	3	0.15	0.15
2.2	17	0.84	0.99
2.3	7	0.34	1.33
2.4	2	0.10	1.43
2.6	10	0.49	1.92
2.9	2	0.10	2.02
3.1	5	0.25	2.27
3.2	11	0.54	2.81
3.3	19	0.94	3.74
3.4	43	2.12	5.86
3.5	27	1.33	7.19
3.6	33	1.63	8.82
3.8	11	0.54	9.36
3.9	57	2.81	12.17
4	90	4.43	16.60
4.1	51	2.51	19.11
4.2	34	1.67	20.79
4.3	53	2.61	23.40
4.4	53	2.61	26.01
4.5	14	0.69	26.70
4.6	163	8.03	34.73
4.7	27	1.33	36.06
4.8	79	3.89	39.95
4.9	9	0.44	40.39
5	122	6.01	46.40
5.1	64	3.15	49.56
5.2	57	2.81	52.36
5.3	108	5.32	57.68
5.4	88	4.33	62.02
5.5	80	3.94	65.96
5.6	102	5.02	70.99
5.7	25	1.23	72.22
5.8	82	4.04	76.26
5.9	82	4.04	80.30
6	60	2.96	83.25
6.1	42	2.07	85.32
6.2	19	0.94	86.26
6.3	67	3.30	89.56
6.4	28	1.38	90.94
6.5	8	0.39	91.33
6.6	62	3.05	94.38
6.7	34	1.67	96.06
6.8	13	0.64	96.70
6.9	13	0.64	97.34
7	3	0.15	97.49
7.1	1	0.05	97.54
7.2	23	1.13	98.67
7.4	2	0.10	98.77
7.5	18	0.89	99.66
7.8	2	0.10	99.75
8.2	5	0.25	100.00
Total	2,030	100.00	

```
. sum hhura
```

Variable	Obs	Mean	Std. Dev.	Min	Max
hhura	2,030	5.105813	1.026075	1.9	8.2

```
.
```

Through *some* commands, we can see the average unemployment rate is 5.1, and the min one is 1.9. the standard division is 1.02.

```
. tab hhad10
```

HH: SEIFA 2001 Decile of Index of relative socio-economic advantage/disadvantage	Freq.	Percent	Cum.
[1] Lowest decile	78	3.84	3.84
[2] 2nd decile	130	6.41	10.25
[3] 3rd decile	144	7.10	17.35
[4] 4th decile	141	6.95	24.30
[5] 5th decile	123	6.06	30.36
[6] 6th decile	169	8.33	38.69
[7] 7th decile	226	11.14	49.83
[8] 8th decile	267	13.16	62.99
[9] 9th decile	347	17.10	80.09
[10] Highest decile	404	19.91	100.00
Total	2,029	100.00	

There should be a relationship go with the unemployment rate. Because the region with high ump has extreme advantages/disadvantages in common sense.

```
. corre wscei age_g5 yob hhura hlth5 hhad10 ed5 urban hhiu sex
(obs=1,787)
```

	wscei	age_g5	yob	hhura	hlth5	hhad10	ed5	urban
wscei	1.0000							
age_g5	0.3147	1.0000						
yob	-0.2343	-0.8120	1.0000					
hhura	-0.0453	-0.0005	0.0877	1.0000				
hlth5	-0.0010	0.1048	-0.1133	-0.0029	1.0000			
hhad10	0.0501	-0.0531	0.0118	-0.2066	-0.0668	1.0000		
ed5	-0.2709	-0.4691	0.3982	0.0552	-0.0330	-0.0862	1.0000	
urban	-0.0251	-0.1169	0.1070	-0.2005	-0.0423	0.3312	-0.0297	1.0000
hhiu	-0.1097	-0.3470	0.2971	-0.0710	-0.0263	0.0441	0.2292	0.1380
sex	-0.1661	-0.0073	0.0229	0.0145	0.0465	-0.0663	0.0189	-0.0323

Using *correlation* to see the variance of each variable.

Age-g5 and yob have the same attributes, just used to test which can best describe age. And the relationship can be ignored.

As age increases, we increase, moderate r

The unemployment rate and health data have a really weak relationship with us. Which does not match the previous interpretation. (More income, less unemployment rate)

SEIFA is the same as above, but with higher income, and a higher decile.

A higher degree of qualification reached can get more income, weak r

Less income unit, more income, weak r

Male get more income, weak r.

As age increases, school degrees increase, moderate r.

As age increases, income units decrease, moderate r.

As the area unemployment rate increases, the sea index decrease (more balance of a/dis ad), and more likely to live in rural, weak r.

High decile is more likely to live in urban, moderate r.

A high school degree has more income units, weak r.

```
. regress age_g5 wscei hhura hlth5 hhad10
```

Source	SS	df	MS		
Model	186.251442	4	46.5628605	Number of obs	= 1,787
Residual	1448.36076	1,782	.812772591	F(4, 1782)	= 57.29
				Prob > F	= 0.0000
Total	1634.6122	1,786	.915236394	R-squared	= 0.1139
				Adj R-squared	= 0.1120
				Root MSE	= .90154

age_g5	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
wscei	.0003764	.0000264	14.23	0.000	.0003245	.0004282
hhura	.0013493	.0213743	0.06	0.950	-.040572	.0432707
hlth5	.1139735	.0252365	4.52	0.000	.0644772	.1634697
hhad10	-.0213476	.0078786	-2.71	0.007	-.0368	-.0058953
_cons	1.425604	.1478092	9.64	0.000	1.135706	1.715501

As we use age as an independent vavariablethe regression model is like above. The reason I select is that r square is the biggest one, others I test all-around 0.05 and above 0.1. This one at least means something I guess.

We, hhad10, and hlth5 have a p-value below 1%, we claim that there is a statistically significant relationship. And under 99 confidence Significance Level.

95% confidence interval is between 1.135 and 1.715.

Ahura is not the true value in our test. Thus we may consider to re-make the regress model.

Table 1. Cox Regression Model [Dependent Variable: Time (years) spent in Unstable or Low Wage Employment]

	Model 1: Finds Full Time Job			Model 2: Finds Permanent Job			Model 3: Total Earnings > Mean Earnings			Model 4: Total Earnings > Median Earnings		
Variables	**ALL**	**Male**	**Female**	**ALL**	**Male**	**Female**	**ALL**	**Male**	**Female**	**ALL**	**Male**	**Female**
Female[+]	**-0.260****			**-0.154*****			**-0.413*****			**-0.360*****		
	(0.000)			(0.005)			(0.000)			(0.000)		
Age	0.006	0.003	0.009	-0.001	0.005	-0.003	**0.029*****	**0.025*****	**0.035*****	**0.027*****	**0.029*****	**0.029*****
	(0.203)	(0.651)	(0.107)	(0.800)	(0.518)	(0.601)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Parent[+]	**-0.247*****	0.122	**-0.492*****	0.040	0.143	-0.027	**-0.168****	0.042	**-0.371*****	**-0.218****	-0.012	**-0.384*****
	(0.003)	(0.352)	(0.000)	(0.643)	(0.305)	(0.810)	(0.072)	(0.771)	(0.003)	(0.015)	(0.931)	(0.001)
Married/De Facto[+]	0.046	0.123	-0.028	-0.009	-0.067	-0.012	0.067	0.177	-0.029	0.040	0.133	-0.045
	(0.438)	(0.235)	(0.705)	(0.879)	(0.546)	(0.874)	(0.344)	(0.138)	(0.746)	(0.554)	(0.246)	(0.593)
Health Score[a]	-0.013	-0.002	-0.043	-0.003	0.029	-0.027	**-0.061****	-0.053	**-0.091***	**-0.073****	-0.058	**-0.105****
	(0.676)	(0.963)	(0.270)	(0.914)	(0.575)	(0.496)	(0.090)	(0.354)	(0.053)	(0.034)	(0.288)	(0.019)
Urban[+]	**-0.152****	-0.026	**-0.218*****	**-0.133****	-0.025	**-0.196****	-0.004	0.078	-0.046	-0.009	0.073	-0.050
	(0.021)	(0.811)	(0.009)	(0.050)	(0.824)	(0.021)	(0.964)	(0.531)	(0.649)	(0.908)	(0.539)	(0.602)
Highest Degree: PhD/Masters[++]	0.099	0.083	0.084	-0.006	-0.053	0.004	**0.480*****	**0.402*****	**0.530*****	**0.402*****	**0.280****	**0.476*****
	(0.185)	(0.482)	(0.385)	(0.944)	(0.677)	(0.971)	(0.000)	(0.003)	(0.000)	(0.000)	(0.031)	(0.000)
Highest Degree: Grad Dip/Cert[++]	-0.012	0.106	-0.112	**0.162****	**0.241***	0.108	**0.189****	**0.333****	0.049	**0.222*****	**0.272****	0.161
	(0.880)	(0.393)	(0.277)	(0.042)	(0.064)	(0.291)	(0.034)	(0.016)	(0.675)	(0.009)	(0.041)	(0.147)
COB: English-Speaking[+++]	0.026	0.040	-0.007	0.021	-0.152	0.122	0.113	0.043	0.158	0.100	0.122	0.072
	(0.821)	(0.811)	(0.965)	(0.856)	(0.394)	(0.449)	(0.368)	(0.813)	(0.368)	(0.410)	(0.489)	(0.667)
COB: Non-English Speaking[+++]	**-0.159****	-0.157	-0.127	**-0.189****	**-0.256****	-0.122	**-0.342*****	**-0.304****	**-0.353*****	**-0.356*****	**-0.309****	**-0.376*****
	(0.040)	(0.173)	(0.228)	(0.021)	(0.039)	(0.262)	(0.000)	(0.027)	(0.007)	(0.000)	(0.019)	(0.002)
Area Unemployment Rate	**-0.067*****	-0.061	**-0.077****	**-0.054****	-0.042	**-0.066***	-0.039	**-0.081***	0.001	**-0.052***	**-0.081***	-0.026
	(0.010)	(0.121)	(0.027)	(0.048)	(0.315)	(0.063)	(0.208)	(0.071)	(0.984)	(0.079)	(0.063)	(0.510)
Year of Graduation	**-0.016*****	**-0.016***	**-0.014****	**-0.012****	**-0.016***	-0.009	**-0.052*****	**-0.048*****	**-0.054*****	**-0.043*****	**-0.041*****	**-0.044*****
	(0.004)	(0.067)	(0.066)	(0.035)	(0.087)	(0.216)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

[+]Binary; Base variables are male, no children, single, rural. [++]Tertiary Degree Base: Bachelors/Honours. [+++]Country of Birth base is Australia.
[a]Health Score ranges from 1 (excellent) to 5 (poor), self-rated.

Model 1&2

Overall, it takes more time for females to find a job whether full-time or permanent compared to males. And whether the sample has children also affects the time to find a job, when the sample does not have children, it will take less time. By comparison, we can see that females rive at a value of -0.49, which is a strong relationship. When females have children, their time to find a job increases dramatically, the reason may about females' cost of investing in a child and the duration of responsible are longer compared to males. Males receive a very low impact in this area causing them to pull down the mean value. Similarly, females living in urban areas have a higher time to find full-time and part-time jobs, while males have no effect, either because of the invisible discrimination against females in urban jobs or because of the biological differences between males and females. Males are more suited to rational and logical work and to maintaining a stable emotional state to make analyses, and females, are more suited to emotional, detailed, and fluctuating jobs that require great insight and empathy. And the attributes and arrangement of the city make the front type of occupation more than the back. For cob(non-eng), hypnosis 1 for non-English speakers and 0 for English speakers. The data shows that non-English speakers tend to take more time to find a job. An interesting phenomenon here is that male non-English speaking ethnic groups tend to take more time to find part-time jobs, possibly since the fact that society

is less tolerant of minority males. The next two variables have negligible effects, the longer it takes to find a job the higher the unemployment rate, and the longer it takes to graduate, the longer it takes to find a job, with no significant difference between males and females. It is interesting to note that the more educated the person, the longer it takes to find a part-time job, which may be because highly educated people tend to be reluctant to work in fragmented jobs with fewer skills and lower wages, while this value is higher for female, probably because female are physical can perform some demanding tasks or the task go with the requirement of high intensity of strength. In addition, to the low production of testosterone y a female from the chemical mechanism, there is also a strong resistance to this type of work.

Model 3&4

Firstly, the median income in a society better reflects the actual income situation in the society. Females hurt both average and median incomes, and the reason for the high value of average incomes (-0.413>-0.36) is that there are more males than females in the higher income groups, as explained in part 1. Having children lead to lowers income, as explained earlier, the longer you care for your children, the more fragmented your work time becomes. And, the higher the school grades of females, the lower their income, which can be linked to the previous section on higher education leading to more time spent looking for permanent jobs. Because, females may be more pressured by public opinion, the social evaluation system in many cases frames the career and status that females should have, and good grades in school add to the psychological shackles and this kind of status to many people, making them reluctant to do jobs which look like so 'decent' 'not cool'. After completing their undergraduate education, the data shows that females have higher income levels than males if they go on to complete higher education. This may result in the fact that there is far fewer female than male in the highest levels of education, resulting in an imbalance between the supply and demand of the jobs that required a high knowledge structure and women having a specific advantage. We could get the same assumption by comparing the mean earning of 0.53 and the median of 0.476. The mean value is higher, cause of the high ceiling. On the language issue, non-English speaking females can find work quickly but still earn low incomes, while men cannot find work quickly but suddenly when getting laborioussskilled jobs, like moving bricks and fixing car sites. All of these have higher earnings which increase the mean.

This explanation is for the first assignment, I am apologized, due to my bad mental state, I knew it.

This assignment will hopefully be a little advanced.

15 Sep 2020 at 5:27 pm

1 Firstly, we assert that some variables are relevant, and my main objective is to find aims to identify the female education underutilization most affects, and the drivers behind it.

```
. sum wscei, detail
```

WS: Current weekly gross wages & salary - all jobs (AUD) [imputed] [weighted top

	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	2,030
25%	330	0	Sum of Wgt.	2,030
50%	800		Mean	848.7773
		Largest	Std. Dev.	787.8042
75%	1150	5500		
90%	1638	6904	Variance	620635.5
95%	2014	8490	Skewness	5.347448
99%	3350	16456	Kurtosis	85.99773

Using summarize, d command to get the average and standard division of income per week. As we can see, the skewness is more than 5 which means the distribution curve of the week salary is right skew, and the kurtosis is high. The reason there are four 0 value in smallest distribution is there are many people who lost job and get zero paym[illegible]et at the current

Comments | Files (1) | Rubric

Good critical analysis., Saar.
Excellent Work.

Part 1. Demonstrates strong quantitative-analytic skills. Relationships highlighted were relevant to issue and presented in a logical and very structured way. Good comprehensive analysis but a bit too long. You have a lot of room to consolidate findings and weave them into a coherent narrative.

Part 2. Demonstrates strong grasp of mechanics of regression models. Strong inferential skills evident, very much on the right track. Points need to be more solidly consolidated in the close, but very good job overall.

Comment