BUEC 333: Assignment 3

8 points for this section

1) [Midterm, Spring 2009] (**1 point)** The power of a test is the probability that you:
   a) reject the null when it is true
   b) fail to reject the null when it is false
   c) reject the null when it is false
   d) fail to reject the null when it is true
   e) none of the above

2) [Midterm, Spring 2009] (**1 point)** Suppose you compute a sample statistic q to estimate a population quantity Q. Which of the following four statements is/are **false**?
   [1] the variance of Q is zero
   [2] if q is an unbiased estimator of Q, then $q = Q$
   [3] if q is an unbiased estimator of Q, then q is the mean of the sampling distribution of Q
   [4] a 95% confidence interval for q contains Q with 95% probability
   a) 2 only
   b) 3 only
   c) 2 and 3
   d) 2, 3, and 4
   e) 1, 2, 3, and 4

3) [Midterm, Spring 2009] **(1 point)** Suppose you want to test the following hypothesis at the 10% level of significance:

$$H_0: \mu = \mu_0$$
$$H_1: \mu \neq \mu_0$$

   Which of the following statements is/are **true**?
   a) the probability of a Type I error is 0.05
   b) the probability of a Type I error is 0.10
   c) the probability of a Type II error is 0.90
   d) the probability of a Type II error is 0.10
   e) none of the above

4) (**4 points)** Suppose you collect the following data that you **know** are a random sample from a $N(\mu,\sigma^2)$ population:

   4.37 6.99 7.85 2.60 3.34 5.94 4.21 5.99 8.53 4.92

   a) Compute the t-statistic for testing the hypothesis:

$$H_0 : \mu = 4$$
$$H_1 : \mu \neq 4$$

      i) mu-hat=5.474; E[V(mu-hat)]=s2=V(x)/(n-1)=3.74/9=0.41; mu-hat-4/sqrt(V(mu-hat))=1.474/sqrt(0.41)=2.28 (2 points)
   b) What is the sampling distribution of the test statistic you computed in part a? **t with 9 df (1 point)**
   c) Can you reject the null hypothesis of part a at the 5% level of significance? **5% crit with 9 df is 2.26. So, we reject because 2.28>2.26. (1 point)**

5) (**1 point)** In the linear regression model, R^2 measures
   a) The proportion of variation in *Y* explained by *X*

b) The proportion of variation in *X* explained by *Y*
c) The proportion of variation in *Y* explained by *X*, adjusted for the number of independent variables
d) The proportion of variation in *X* explained by *Y*, adjusted for the number of independent variables
e) None of the above

6) **[10 points for this section]** EVIEWS PART: Read Pendakur, Krishna and Ravi Pendakur, 2010, "Colour By Numbers: Minority Earnings Disparity 1995-2005", forthcoming, Journal of Intenational Migration and Integration, linked on the course website.

   a) In your sample of data from people living in Greater Vancouver, select a sample which corresponds to the sample selected for regression estimation by Pendakur and Pendakur (2010). Note that while they use the entire sample of the long-form census file, you use only about one in seven observations from that file: the long-forms cover 20% of the population, but the public-use data that you have cover only 3% of the population. You will use this sample for all the questions below, too. How many observations do you have? Make a table showing the average earnings of white, visible minority and Aboriginal men and women (6 types) in this sample.

   1 point: 13199 cases (if they're close, they get 1 point) THE KEY IS TO DROP MISSINGS

   these numbers are from descriptive statistics, manipulating the sample for each number

   b) Make a table showing the standard errors of the sample means you reported in part b. These standard errors equal the square-root of the variance of sampling distributions of the sample means in part b.

7) **these numbers are from descriptive statistics, manipulating the sample for each number**

   a) **these numbers are std dev/root-n, both of which are given in the descriptive statistics.**

   b) **1 point for sample means that are close; 1 point for std errs that are close.**

	mean	std dev	n	std dev/rootn
wt-wom	38534.84	30043.27	5652	399.619
wt-men	64084.8	78456.46	5847	1026.035
vm-wom	39734.27	31343.37	626	1252.733
vm-men	53484.13	65537.1	669	2533.811
ab-wom	40185.79	29223.81	183	2160.286
ab-men	31705.19	30444.25	222	2043.284

   c) Make a table showing the difference between the sample mean of white and visible minority and white and Aboriginal earnings for men and women (4 comparisons). Include the standard errors of these differences in this table.

The key here is to note (from the formula for variance) that the variance of a difference of two sample means is the sum of their variances, and the std err is the square root of that. see next table

   d) Construct test statistics for 4 hypotheses: the average earnings of visible minority men in the population is lower than that of white men; the average earnings of Aboriginal men in the population is lower than that of white men; the average earnings of visible minority women in the population is lower than that of white women; and the average earnings of Aboriginal women in the population is lower than that of white women.

8) **1 point for std errs that are close; 1 point for test stats that are close; 1 point for correct interpretation of tests**

	Diff	std err diff	test stat
vm-wm	1199.43	1314.928	0.912164
vm-men	-10600.7	2733.669	-3.87782

Abwom	**1650.95**	**2196.937**	**0.751478**
ab-men	**-32379.6**	**2286.429**	**-14.1617**

for men, these minorities have lower average earnings.

a) Run regressions like those in Table 2, corresponding to regressions controlling for personal characteristics only in Vancouver in 2005 (2006 microdata are about 2005 earnings). Be sure that you get the dependent variable correct. Report the output for the visible minority and Aboriginal coefficients only.

MEN

Dependent Variable: LOG(WAGES)
Method: Least Squares
Date: 11/12/10 Time: 14:41
Sample: 1 56529 IF AGEGRP>8 AND AGEGRP<17 AND WAGES>100
AND SEX=2 AND CITIZEN=1 AND WAGES<8000000
Included observations: 6708

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.941972	0.065480	151.8332	0.0000
VISMIN<13	-0.039790	0.038440	-1.035122	0.3006
ABOID<6	-0.230604	0.068768	-3.353367	0.0008
AGEGRP=10	0.309813	0.041488	7.467624	0.0000
AGEGRP=11	0.386448	0.042412	9.111749	0.0000
AGEGRP=12	0.485940	0.043395	11.19814	0.0000
AGEGRP=13	0.566416	0.044362	12.76799	0.0000
AGEGRP=14	0.473529	0.046511	10.18092	0.0000
AGEGRP=15	0.375384	0.049936	7.517266	0.0000
AGEGRP=16	-0.020625	0.061228	-0.336847	0.7362
HDGREE=2	0.227064	0.044155	5.142462	0.0000
HDGREE=3	0.144260	0.057971	2.488471	0.0129
HDGREE=4	0.326140	0.055970	5.826997	0.0000
HDGREE=5	0.183697	0.075001	2.449266	0.0143
HDGREE=6	0.358519	0.050980	7.032498	0.0000
HDGREE=7	0.376088	0.061504	6.114836	0.0000
HDGREE=8	0.311281	0.064569	4.820913	0.0000
HDGREE=9	0.510465	0.047439	10.76043	0.0000
HDGREE=10	0.476429	0.083294	5.719855	0.0000
HDGREE=11	1.102464	0.137275	8.031053	0.0000
HDGREE=12	0.623808	0.063306	9.853824	0.0000
HDGREE=13	0.774046	0.115171	6.720865	0.0000
HDGREE=88	0.411254	0.347135	1.184709	0.2362
MARST=2	0.216753	0.043658	4.964759	0.0000
MARST=3	0.005277	0.074057	0.071258	0.9432
MARST=4	-0.191438	0.047174	-4.058091	0.0001
MARST=5	-0.097890	0.161492	-0.606162	0.5444
KOL=2	-0.169922	0.525802	-0.323167	0.7466
KOL=3	-0.077730	0.038501	-2.018911	0.0435

R-squared	0.139391	Mean dependent var	10.63560
Adjusted R-squared	0.135783	S.D. dependent var	0.976286
S.E. of regression	0.907588	Akaike info criterion	2.648261
Sum squared resid	5501.596	Schwarz criterion	2.677706
Log likelihood	-8853.267	Hannan-Quinn criter.	2.658429
F-statistic	38.63503	Durbin-Watson stat	2.023390
Prob(F-statistic)	0.000000		

WOMEN

Dependent Variable: LOG(WAGES)
Method: Least Squares
Date: 11/12/10 Time: 14:42
Sample: 1 56529 IF AGEGRP>8 AND AGEGRP<17 AND WAGES>100
AND SEX=1 AND CITIZEN=1 AND WAGES<8000000
Included observations: 6508

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.427923	0.068987	136.6629	0.0000
VISMIN<13	0.088903	0.040238	2.209420	0.0272
ABOID<6	-0.151119	0.064010	-2.360854	0.0183
AGEGRP=10	0.205175	0.042687	4.806477	0.0000
AGEGRP=11	0.294197	0.044133	6.666189	0.0000
AGEGRP=12	0.483487	0.044757	10.80258	0.0000
AGEGRP=13	0.527377	0.044440	11.86718	0.0000
AGEGRP=14	0.544731	0.047376	11.49810	0.0000
AGEGRP=15	0.408922	0.051016	8.015549	0.0000
AGEGRP=16	0.020519	0.062757	0.326963	0.7437
HDGREE=2	0.335415	0.053664	6.250289	0.0000
HDGREE=3	0.283880	0.070474	4.028155	0.0001
HDGREE=4	0.345029	0.109560	3.149238	0.0016
HDGREE=5	0.320319	0.064221	4.987739	0.0000
HDGREE=6	0.405661	0.058203	6.969763	0.0000
HDGREE=7	0.583164	0.068969	8.455416	0.0000
HDGREE=8	0.517791	0.067027	7.725157	0.0000
HDGREE=9	0.696710	0.056237	12.38877	0.0000
HDGREE=10	0.648735	0.079914	8.117956	0.0000
HDGREE=11	0.762487	0.176601	4.317575	0.0000
HDGREE=12	0.851479	0.068188	12.48724	0.0000
HDGREE=13	1.209336	0.185073	6.534382	0.0000
HDGREE=88	0.246933	0.331299	0.745347	0.4561
MARST=2	-0.010021	0.037827	-0.264930	0.7911
MARST=3	-0.061969	0.063042	-0.982982	0.3257
MARST=4	0.036099	0.042848	0.842492	0.3995
MARST=5	-0.087644	0.094282	-0.929585	0.3526
KOL=2	0.699086	0.930578	0.751239	0.4525
KOL=3	-0.054751	0.037589	-1.456558	0.1453
KOL=4	-0.613353	0.927668	-0.661178	0.5085

R-squared	0.083489	Mean dependent var	10.21663
Adjusted R-squared	0.079386	S.D. dependent var	0.964859
S.E. of regression	0.925769	Akaike info criterion	2.688215
Sum squared resid	5551.956	Schwarz criterion	2.719472
Log likelihood	-8717.450	Hannan-Quinn criter.	2.699025
F-statistic	20.34866	Durbin-Watson stat	2.101746
Prob(F-statistic)	0.000000		

1 point for regression output

1 point for tests

b) Test the hypothesis that the mean log-earnings of visible minority men is lower than that of white men, and similarly for Aboriginal men, and similarly for women.

c) vm men not significantly lower; vm women not significantly lower; aboriginal men significantly lower; aboriginal women significantly lower. not that vm women earn significantly more.

d) How do your results differ from Pendakur and Pendakur (2010)? Why do they differ?

e) 1 point: not identical numbers because different sample, with 1/7 the number of cases. they could see the negative for vm men that is too small to see with only 1/7 the data.

f) Table 2 also considers regressions that control for work characteristics. How does their argument for the difference between these two variables connect with missing variables bias?

1 point: work characteristics are missing variables in the regression above. if they're correlated with vm or aboriginal status, then that correlation biases the coefficient.