Sample Problems

1. True or False

Note: If you find the following statements true, you should briefly prove them. If you find them false, you should correct them.

(a) The point $(\bar{x}, \bar{y})$ falls on the regression line of the following simple linear regression,

$$y = \beta_0 + \beta_1 x + u. \tag{1}$$

(b) The following model suffers from perfect collinearity,

$$\log(income) = \beta_0 + \beta_1 education + \beta_2 male + \beta_3 female + u,$$

where $male = 1$ for male individuals and 0 for female individuals, and $female = 1$ for female individuals and 0 for male individuals.

(c) Suppose we have collected n pairs of (x_i, y_i) for the simple regression in (1), The more variation in x, the less accurate $\hat{\beta}_1$ is, where $\hat{\beta}_1$ is the OLS estimator of β_1.

2. Maximum Likelihood Estimation Suppose $(y_i, x_i, i = 1, ..., n)$ follow the simple regression in (1). We assume $u \sim iidN(0, \sigma^2)$.

(a) Write the log likelihood function.

(b) Obtain the MLE estimator of β_1, say $\hat{\beta}_1$.

(c) Obtain the MLE estimator of σ^2.

(d) If it turns out that normality does not hold for u, is $\hat{\beta}_1$ still unbiased? Is it still consistent?

3. Weighted Least Square Consider the savings function,

$$sav_i = \beta_0 + \beta_1 inc_i + u_i, \; u_i = \sqrt{inc_i} \cdot \varepsilon_i, \tag{2}$$

where sav denotes "saving" and inc denotes "income", and where ε_i is iid with $\mathbb{E}\varepsilon_i = 0$ and $\text{var}(\varepsilon_i) = \sigma^2$. We assume ε is independent of inc.

(a) Show that $\mathbb{E}(u|inc) = 0$, so that the exogeneity assumption in CLR (Classical Linear Regression) Assumptions still holds.

(b) Show that $\text{Var}(u|inc) = \sigma^2 inc$. Which assumption in CLR assumptions is violated? And what consequences does it have for estimation and making inferences?

(c) Instead of running (2), we can run the following regression, which is called "weighted least square",

$$\frac{1}{\sqrt{inc_i}}sav_i = \beta_0\frac{1}{\sqrt{inc_i}} + \beta_1\frac{1}{\sqrt{inc_i}}inc_i + \varepsilon_i. \tag{3}$$

Now, does homoscedasticity hold for (3)?

(d) Provide a discussion that supports the assumption that the variance of savings increases with family income.

3. Inferences Consider the following regression,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_1 x_3 + \beta_4 x_4 + \beta_5 x_5 + u.$$

Show how you would test the following hypotheses by running two regressions and computing an F-statistic. Show explicitly the regressions you would run and the degrees of freedom of each test statistic.

(a) $\beta_1 = 0$.

(b) $\beta_1 = 0$ and $\beta_4 = \beta_5$.

(c) $\beta_1 = 0$, $\beta_3 = 0$, and $\beta_4/\beta_5 = 2$.

4. Empirical Problems To study how residence status () affect income, we run the following regression,

$$\log(INCOME) = \beta_0 + \beta_1 EDU + \beta_2 EDU^2 + \beta_3 RURAL + u, \tag{4}$$

where EDU denotes years of education and $RURAL$ is a binary variable defined as 1 for rural residence and 0 for urban residence. The regression results are reported in Figure 1.

(a) Interpret β_3, the coefficient on $RURAL$. Test the following hypothesis,

$$\mathrm{H}_0: \ \beta_3 = 0 \quad \mathrm{H}_1: \ \beta_3 < 0.$$

You should show explicitly the test statistic, the distribution under H_0 and the critical value.

(b) From the results, infer SST (total sum of squares).

We may conjecture that gender influences income. So we run another regression,

$$\begin{aligned}\log(INCOME) \quad = \quad & \beta_0 + \beta_1 EDU + \beta_2 EDU^2 + \beta_3 RURAL \\ & + \beta_4 FEMALE + \beta_5 FEMALE \cdot RURAL + u,\end{aligned} \tag{5}$$

where $FEMALE$ is a binary variable (1 for females and 0 for males). The regression results are shown in Figure 2.

(c) Interpret the economic meaning of β_5.

(d) Test whether gender plays any role in income determination. You should write down the null and the alternative hypotheses, show explicitly the test statistic, the distribution under H_0 and the critical value.

```
============================================================
Dependent Variable: LOG(INCOME)
Method: Least Squares
Date: 06/02/09   Time: 16:13
Sample: 1 5778
Included observations: 5778
============================================================
     Variable      CoefficienStd. Errort-Statistic  Prob.
============================================================
         C          8.195432   0.043328   189.1502   0.0000
        EDU         0.065987   0.008549   7.718704   0.0000
       EDU^2        0.002936   0.000526   5.584321   0.0000
       RURAL       -0.744145   0.027641  -26.92178   0.0000
============================================================
R-squared           0.452430    Mean dependent var 8.636962
Adjusted R-squared  0.452146    S.D. dependent var 1.130683
S.E. of regression  0.836900    Akaike info criteri2.482467
Sum squared resid   4044.115    Schwarz criterion  2.487079
Log likelihood     -7167.846    F-statistic        1590.257
Durbin-Watson stat  1.521717    Prob(F-statistic)  0.000000
============================================================
```

Figure 1: Model 1.

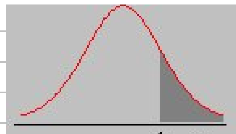
```
============================================================
Dependent Variable: LOG(INCOME)
Method: Least Squares
Date: 06/02/09   Time: 16:33
Sample: 1 5778
Included observations: 5778
============================================================
     Variable      CoefficienStd. Errort-Statistic  Prob.
============================================================
         C          8.430184   0.046132   182.7407   0.0000
        EDU         0.043737   0.008562   5.108331   0.0000
       EDU^2        0.003804   0.000521   7.307523   0.0000
       RURAL       -0.674596   0.033591  -20.08261   0.0000
      FEMALE       -0.227601   0.032273  -7.052272   0.0000
   FEMALE*RURAL    -0.168730   0.044147  -3.821994   0.0001
============================================================
R-squared           0.472691    Mean dependent var 8.636962
Adjusted R-squared  0.472234    S.D. dependent var 1.130683
S.E. of regression  0.821413    Akaike info criteri2.445456
Sum squared resid   3894.477    Schwarz criterion  2.452374
Log likelihood     -7058.922    F-statistic        1034.828
Durbin-Watson stat  1.503794    Prob(F-statistic)  0.000000
============================================================
```

Figure 2: Model 2.

				5% Critical Values of the F Distribution						
				Numer DF						
Denom DF	1	2	3	4	5	6	7	8	9	10
10	4.97	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.97	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.56	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.33	3.47	3.07	2.84	2.69	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.38	2.32	2.28
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.26
25	4.24	3.39	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.17
40	4.09	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
inf	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Figure 3: 5% Critical Values of F Distribution.

Right tail probability of t distribution



df\p	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	4.141
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.850
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.725
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.646
inf	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.291

Figure 4: Right-tail Probabilities of t Distribution