Political Science 150B/350B
Winter 2006
Midterm Examination

This is a closed book, in-class examination. You may use a calculator. Attempt all questions. Show working for partial credit. The total number of points appears at the end of the exam.

Question 1: Consider the linear regression model

$$Y_i = \alpha + X_i\beta + U_i$$

where Y_i and X_i are observed data and α and β are unknown parameters to be estimated, U_i is an unobserved disturbance, and $i = 1, \ldots, n$. The least squares estimator of β is

$$\hat{\beta} = \sum_{i=1}^{n} w_i Y_i$$

where

$$w_i = \frac{X_i}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Q1.1 (3 points) In what sense is the least squares estimate of β a *linear* estimator?

Q1.2 (3 points) If Y_i and X_i are positively correlated, then what do we know about $\hat{\beta}$?

Q1.3 (3 points) "β is a random variable." True or false, and why?

Q1.4 (3 points) "If $E(\hat{\beta}) = \beta$ then $\hat{\beta}$ is a ... estimator" (Complete the sentence).

Q1.5 (3 points) Carefully state the conditions necessary to show that $E(\hat{\beta}) = \beta$.

Q1.6 (3 points) Why is the property $E(\hat{\beta}) = \beta$ desirable?

Q1.7 (3 points) What features of one's data and/or model can generate the circumstances under which $E(\hat{\beta}) \neq \beta$?

Q1.8 (5 points) Suppose we find that we cannot reject the null hypothesis $H_0 : \beta = 0$. Does this mean that X and Y are independent? Explain your answer.

Q1.9 (3 points) What is the difference between a "disturbance" and a "residual"?

Q1.10 (3 points) In deriving the sampling variability of the least squares estimators of α and β, we assume that the data (Y_i, X_i) are generated by random sampling. Explain what is meant by this assumption using terms that would make sense to a colleague who has not taken a statistics class.

Q1.11 (3 points) In addition to the assumptions of "strict exogeneity", and random sampling, what other assumption is required in order to establish that the least squares estimators of α and β have smallest sampling variability of linear unbiased estimators?

Q1.12 (3 points) Consider the assumption you provided in answering the previous question. Explain what is meant by this assumption using terms that would make sense to a colleague who has not taken a statistics class.

Question 2: Using the model defined in the previous question, consider the quantity

$$\hat{\beta}/\sqrt{\widehat{\text{var}(\hat{\beta})}}.$$

Q2.1 (3 points) Under $H_0 : \beta = 0$, what is the distribution of this quantity in repeated sampling?

Q2.2 (3 points) Suppose $n = 16$ and $q = -1.85$. How plausible is H_0?

Q2.3 (5 points) Suppose $\hat{\beta} = 2.2$ and $q = 3.45$ with $n = 22$. Construct a 99% confidence interval for β.

Q2.4 (3 points) Suppose you are talking to a colleague who has not taken a statistics class. Using language this colleague could understand, interpret the confidence interval you constructed in the previous question.

Question 3: Choose the best answer (no need to state reasoning) to each of the following propositions:

Q3.1 (3 points) "We use t-statistics to test hypotheses about regression coefficients because..."

(a) we never have infinite sample sizes.

(b) we can't be sure that our residuals have a normal distribution over repeated samples.

(c) the sampling variability of a regression coefficient is itself estimated from the data

(d) we need to correct for the fact that we are estimating β *and* α

Q3.2 (3 points) Consider the model $Q_i = \alpha K_i^{\beta} L_i^{\theta} e^{\varepsilon_i}$. Least squares regression analysis can be used to estimate

(a) none of the parameters

(b) α, β and θ

(c) the logs of α, β and θ

(d) β and θ

Q3.3 Consider the regression model $y_i = \beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\beta_3 D_i+u_i$, where D_i is a "dummy variable". The parameter β_3 can be interpreted as

(a) $E(y|x_1 = \bar{x}_1, x_2 = \bar{x}_2, D = 1)$

(b) $E(y|D = 1)$ - $E(y|D = 0)$.

(c) both (a) and (b)

(d) neither (a) nor (b)

Q3.4 Using the notation of Q1, the quantity $\sum_{i=1}^{n} \hat{U}_i^2/(n - 2)$ is

(a) an unbiased estimate of the sampling variability of each residual

(b) a biased estimate of the sampling variability of each residual, but with the bias vanishing as $n \rightarrow \infty$.

(c) an unbiased estimate of the sampling variability of $\hat{\beta}$.

(d) an unbiased estimate of the sampling variability of $\hat{Y}_i$.

Question 4: A measure of the perceived prestige of 102 occupations was analyzed via ordinary least squares multiple regression; the prestige measure ranges from 14.8 to 87.2 with a mean of 46.8. Two predictors were used: (1) the average number of years of education within each occupation (mean 10.7, minimum 6.4, maximum 16.0); (2) the type of occupation: blue collar (44 occupations), professional (31), white collar (23). Four occupations were not coded into any type, and are dropped from the analysis, leaving 98 observations. The regression analysis is summarized in the following table (table entries are regression estimates, standard errors in parentheses):

	Model		
	1	2	3
Intercept	-10.8	-2.7	-4.3
	(3.5)	(5.7)	(8.6)
Education	5.4	4.6	4.8
	(0.3)	(0.7)	(1.0)
Professional		6.1	18.9
		(4.3)	(16.9)
White Collar		-5.5	-24.4
		(2.7)	(21.8)
Education × Professional			-0.9
			(1.5)
Education × White Collar			1.7
			(2.1)
r^2	.75	.80	.80
$\hat{\sigma}$	8.6	7.8	7.8

Table 1: Regression Results for Q4

Q4.1 (4 points) Provide a brief interpretation of the coefficients for Professional and White Collar in Model 2.

Q4.2 (3 points) Provide a brief interpretation of the coefficient for Education in Model 3.

Question 5: (9 points) What is the interpretation of β_1, β_2 and β_3 in the following model?:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 (X_i \times D_i) + U_i$$

where Y_i is vote share for the Democratic candidate for Congress in district i, X_i is vote share for the Democratic presidential candidate in district i, and

$$D_i = \begin{cases} 1 & \text{if the Democratic candidate for Congress is an incumbent} \\ 0 & \text{if no incumbent running} \\ -1 & \text{if the Republican candidate for Congress is an incumbent} \end{cases}$$

Question 6: A researcher is interested in the effects of education on earnings, and has data on wage levels and educational attainment from a sample of workers. Of course, wages increase over the course of a career, and the researcher suspects wage trajectories to be linear in time, at least on average; however, the researcher suspects that the trajectories of wages over time differ depending on level of education. The researcher's measure of education is discrete, coded as "less than high school", "high school", "some college", "college graduate or higher". Career stage is measured as years since first entering the labor time market.

Q6.1 (8 points) Specify a regression model that will allow the researcher to test her hypothesis/hypotheses. Carefully define all the terms in your model (all variables, parameters, etc).

Q6.2 (4 points) Using the terms/notation you introduced in the answer to the previous question, provide a formal statement of the null hypothesis "the rate of increase in wages over time is the same for college graduates and those with only some college".

Q6.3 (4 points) Using the terms you introduced in the answer to the previous question, formally state the form of the *joint* or *compound* null hypothesis "the rates of increase in wages over time do not vary across levels of education".

Q6.4 (4 points) Instead of linear rates of growth in wages, suppose that over the course of a career, wages increase in constant *proportional* terms (e.g, 5% per, each year). How would you change the regression model you specified above (i.e., new variables, new parameters, or what)?

END OF EXAM

Total Number of Points: 87

	One-Tailed Significance Level								
	0.001	0.005	0.01	0.025	0.05	0.1	0.15	0.2	0.25
	Two-Tailed Significance Level								
df	0.002	0.010	0.02	0.05	0.1	0.2	0.3	0.4	0.5
1	318.309	63.657	31.821	12.706	6.314	3.078	1.963	1.376	1.000
2	22.327	9.925	6.965	4.303	2.920	1.886	1.386	1.061	0.816
3	10.215	5.841	4.541	3.182	2.353	1.638	1.250	0.978	0.765
4	7.173	4.604	3.747	2.776	2.132	1.533	1.190	0.941	0.741
5	5.893	4.032	3.365	2.571	2.015	1.476	1.156	0.920	0.727
6	5.208	3.707	3.143	2.447	1.943	1.440	1.134	0.906	0.718
7	4.785	3.499	2.998	2.365	1.895	1.415	1.119	0.896	0.711
8	4.501	3.355	2.896	2.306	1.860	1.397	1.108	0.889	0.706
9	4.297	3.250	2.821	2.262	1.833	1.383	1.100	0.883	0.703
10	4.144	3.169	2.764	2.228	1.812	1.372	1.093	0.879	0.700
11	4.025	3.106	2.718	2.201	1.796	1.363	1.088	0.876	0.697
12	3.930	3.055	2.681	2.179	1.782	1.356	1.083	0.873	0.695
13	3.852	3.012	2.650	2.160	1.771	1.350	1.079	0.870	0.694
14	3.787	2.977	2.624	2.145	1.761	1.345	1.076	0.868	0.692
15	3.733	2.947	2.602	2.131	1.753	1.341	1.074	0.866	0.691
16	3.686	2.921	2.583	2.120	1.746	1.337	1.071	0.865	0.690
17	3.646	2.898	2.567	2.110	1.740	1.333	1.069	0.863	0.689
18	3.610	2.878	2.552	2.101	1.734	1.330	1.067	0.862	0.688
19	3.579	2.861	2.539	2.093	1.729	1.328	1.066	0.861	0.688
20	3.552	2.845	2.528	2.086	1.725	1.325	1.064	0.860	0.687
21	3.527	2.831	2.518	2.080	1.721	1.323	1.063	0.859	0.686
22	3.505	2.819	2.508	2.074	1.717	1.321	1.061	0.858	0.686
23	3.485	2.807	2.500	2.069	1.714	1.319	1.060	0.858	0.685
24	3.467	2.797	2.492	2.064	1.711	1.318	1.059	0.857	0.685
25	3.450	2.787	2.485	2.060	1.708	1.316	1.058	0.856	0.684
26	3.435	2.779	2.479	2.056	1.706	1.315	1.058	0.856	0.684
27	3.421	2.771	2.473	2.052	1.703	1.314	1.057	0.855	0.684
28	3.408	2.763	2.467	2.048	1.701	1.313	1.056	0.855	0.683
29	3.396	2.756	2.462	2.045	1.699	1.311	1.055	0.854	0.683
30	3.385	2.750	2.457	2.042	1.697	1.310	1.055	0.854	0.683
50	3.261	2.678	2.403	2.009	1.676	1.299	1.047	0.849	0.679
100	3.174	2.626	2.364	1.984	1.660	1.290	1.042	0.845	0.677
200	3.131	2.601	2.345	1.972	1.653	1.286	1.039	0.843	0.676
500	3.107	2.586	2.334	1.965	1.648	1.283	1.038	0.842	0.675
1000	3.098	2.581	2.330	1.962	1.646	1.282	1.037	0.842	0.675
3000	3.093	2.577	2.328	1.961	1.645	1.282	1.037	0.842	0.675
10000	3.091	2.576	2.327	1.960	1.645	1.282	1.036	0.842	0.675
∞	3.090	2.576	2.326	1.960	1.645	1.282	1.036	0.842	0.674

Table 2: Critical values of the *t* distribution.