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?

Answer: $\hat{\beta}$ is a linear function of the Y_i (in this case, a weighted sum, with weights w_i.

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

Answer: $\hat{\beta} > 0$.

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

Answer: False, in the standard setup: β is a constant, an unknown feature of a population. (Not necc for answer: contrast Bayesian approaches).

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

Answer: Unbiased.

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

Answer: Strict exogeneity. $E(U_i|X_1, \ldots, X_n) = 0, \forall\, i$. Equivalently, weak exogeneity $E(U_i|X_i) = 0, \forall\, i$, plus random sampling.

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

Answer: The sampling distribution for $\hat{\beta}$ is centered over the population parameter β. This means that on average (in the repeated sampling sense), the estimates of $\hat{\beta}$ equal the population parameter we're trying to estimate (although this does remind one of the old joke about statisticians deer-hunting etc, of course not necessary for answer).

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

Answer: Omitted variable bias, a form of model misspecification. That is, consigning something to the error term that *even conditional on the variables that are in the model*, does correlate with Y.

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.

Answer: No. We are given that $H_0 : \beta = 0$ is not rejected. Because of sampling uncertainty (i.e., n is finite), this could happen because either H_0 is true or because H_0 is false. So consider case (a), H_0 is true, which is only a statement about linear association and not sufficient to show that X and Y are statistically independent (i.e., correlation and regression test linear hypotheses about the relationship between Y and X; two variables can have zero correlation and still be statistically dependent via a non-linear pattern of association). On the other hand H_0 could actually be false, which implies linear association and hence X and Y are not independent. In any event, simply because we fail to reject H_0 (at a given significance level etc), we can't say which of these is true, and so the conclusion "X and Y are independent" does not follow from the premise.

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

Answer: U_i is a disturbance, $\hat{U}_i$ is 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.

Answer: Random sampling means that values of Y_i and X_i do not provide information about the values of Y_j and X_j, for all $i \neq j$, where i and j index the observations in the data set. An appropriate analogy is flipping a fair coin: seeing that toss i was a head tells you nothing about the likelihood of a head on any other toss.

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?

Answer: Conditional homoskedasticity, or

$$V(U_i|X_1, \ldots, X_n) = \sigma^2, \forall\ i = 1, \ldots, n$$

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.

Answer: That the errors I make in predicting Y_i as a linear function of X_i have the same spread over the data set.

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

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

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

Answer: t with n - 2 degrees of freedom.

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

Answer: If H_0 were true, we'd see q as large or larger as 1.85 in absolute value in less than 10% repeated samples, but more than 5% of repeated samples. We report the p-value for H_0 as lying between .05 and .1; .05 is a conventional benchmark for "statistical significance" and so we fall short of that benchmark in this instance.

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

Answer: If $q = 3.45$ and $\hat{\beta} = 2.2$, then the estimated standard error of $\hat{\beta}$ is $2.2/3.45 = .638$. Then a 99% interval is $\hat{\beta} \pm 2.845 \times .638$ or $[.38, 4.02]$ (up to rounding error).

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.

Answer: In repeated sampling, 99% of the 99% confidence intervals we form will include β. Thus, in a repeated sampling sense, there is a "99% chance" that $\beta \in [.38, 4.02]$.

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

Answer: Option (c).

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

Answer: (d). Take logs of both sides to yield:

$$\ln Q_i = \ln \alpha + \beta \ln K_i + \theta \ln L_i + \varepsilon_i$$

The parameters $\ln \alpha$, β and θ are estimable by least squares.

Note: option (b) is "almost correct" in the sense that we could exponentiate the least squares estimate of $\ln \alpha$, $\widehat{\ln \alpha}$, to yield $\hat{\alpha}$. Please note that we can't apply the same transformation to the standard error of $\widehat{\ln \alpha}$ to obtain a standard error for $\hat{\alpha}$, and so it is not immediately clear what properties that estimate has (this kind of thing gets easier if we were to estimate the model via maximum likelihood or adopt a Bayesian approach, coming soon).

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)

Answer: (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 $n \to \infty$.
- **(c)** an unbiased estimate of the sampling variability of $\hat{\beta}$.
- **(d)** an unbiased estimate of the sampling variability of $\hat{Y}_i$.

Answer: (a).

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

	Model 1	Model 2	Model 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

from the analysis, leaving 98 observations. The regression analysis is summarized in the following table (table entries are regression estimates, standard errors in parentheses):

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

Answer: These coefficients tap the change in the average level of prestige across occupational types, holding education level constant (or "controlling for" education level). The blue collar type is absorbed in the intercept term, so the coefficients are offsets relative to the blue collar type. Thus, with the same mean level of education, professional occupations are rated as being 6.1 units more prestigious than blue collar occupations, while white collar occupations are rated 5.5 units less prestigious.

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

Answer: This is the slope coefficient for education in the baseline blue collar group. It implies that conditional on an occupation being in the blue collar group of occupations, each additional year of education is associated with a 4.8 unit increase on the prestige scale. This effect is distinguishable from zero at conventional levels of significance.

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

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

Answer: β_1 measures the association between Presidential and congressional election outcomes in *open seats* (where no incumbent is running, $D_i = 0$).

β_2 in itself doesn't have a meaningful interpretation: it is

$$E(Y_i|X_i = 0, D_i = 1) - E(Y_i|X_i = 0, D_i = 1)$$

i.e., the difference in expected vote share for an incumbent Democratic congressional candidate and expected vote share for a Democratic congressional candidate in an open seat *conditional* on Democratic presidential vote share being 0. Knowing the boost in vote share due to incumbency is important, but knowing it under these circumstances ($X_i = 0$) is not. Note also that given the definition of D_i, β_2 is also equal to

$$E(Y_i|X_i = 0, D_i = 0) - E(Y_i|X_i = 0, D_i = -1)$$

β_3 is interesting. It taps the way the mapping from Democratic share of the presidential vote maps into Democratic share of the congressional vote varies depending on whether we are in an open seat contest (no incumbent running, $D_i = 0$) or the more typical situation of an incumbent seeking reelection ($D_i = 1$). For a Democratic incumbent ($D_i = 1$),

$$E(Y_i|X_i, D_i = 1) = \beta_0 + \beta_2 + (\beta_1 + \beta_3)X_i,$$

while for a Republican incumbent ($D_i = -1$),

$$E(Y_i|X_i, D_i = -1) = \beta_0 - \beta_2 + (\beta_1 - \beta_3)X_i,$$

and in an open seat ($D_i = 0$),

$$E(Y_i|X_i, D_i = 0) = \beta_0 + \beta_1 X_i.$$

[Not necc for complete credit]: the hypothesis here (not necc for answer) is interesting to students of elections etc: if X_i is a proxy for the ideological or partisanship disposition of the i-th district, then taps the extent to having an incumbent in office alters the way congressional election outcomes reflect the a given district's ideological or partisan composition (say, via the incumbent's personal vote, his advantage in fund-raising, name-recognition, etc). This suggests a more plausible model is

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

where

$$F_i = \begin{cases} 1 & \text{incumbent running, so } D_i = 1 \text{ or } D_i = -1 \\ 0 & \text{otherwise} \end{cases}$$

i.e., incumbency provides a symmetric boost (intercept shift) the sign of which depends on the party of the incumbent, but then if the (linear) mapping from X_i to Y_i is different in open seat contests than in races where incumbents are running, the difference in the mappings should not depend on the party of the incumbent.

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 full-time labor 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).

Answer: Here is my stab at it (using conventional notation etc):

$$Y_i = \beta_0 + \beta_1 T_i + \sum_{j=2}^{k} D_{ij}(\delta_j + \gamma_j T_i) + U_i$$

where

- Y_i is wage level of worker i
- T_i is years since entering the full-time labor market
- D_{ij} is 1 if worker i belongs to education category j, and zero otherwise
- U_i is an unobserved disturbance
- β_0 is an intercept, in this case equal to expected earnings of a worker in education category $j = 1$ with $T_i = 0$ years in the full-time labor market (e.g., average starting salary in education category $j = 1$).
- β_1 is a slope coefficient, in this case equal to the slope of the regression of wages on T_i in education category $j = 1$.
- the δ_j are intercept shifts or offsets for education categories $j = 2, \ldots, k$, relative to the intercept in the "baseline" $j = 1$ education category (in this parameterization δ_1 is undefined, or if you like, constrained to be zero, but in any event, not required to be estimated).
- the γ_j are slope shifts or offsets for education categories $j = 2, \ldots, k$, relative to the slope in the "baseline" $j = 1$ education category (again, in this parameterization γ_1 is either undefined or zero).

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".

Answer: In my notation, let j index the set

$$\left\{ \begin{array}{c} \text{less than high school,} \\ \text{high school,} \\ \text{some college,} \\ \text{college graduate or higher} \end{array} \right\}.$$

Then the required null hypothesis is $H_0 : \gamma_3 = \gamma_4$ or $H_0 : \gamma_3 - \gamma_4 = 0$.

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".

Answer: In my notation, $H_0 : \gamma_2 = \gamma_3 = \gamma_4 = 0$ or,

$$H_0 : \left\{ \begin{array}{l} \gamma_2 = 0, \\ \gamma_3 = 0, \\ \gamma_4 = 0 \end{array} \right.$$

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 year, each year). How would you change the regression model you specified above (i.e., new variables, new parameters, or what)?

Answer: Re-run the analysis with $\log Y_i$ as the dependent variable. Then consider any of the slope coefficients in the model, which for the purposes of what follows, I will label as β.

Then we have

$$\beta = \frac{\partial \ln Y_i}{\partial T_i} = \frac{\partial \ln Y_i}{\partial Y_i}\frac{\partial Y_i}{\partial T_i} = \frac{1}{Y_i}\frac{\partial Y_i}{\partial T_i}$$

i.e., the rate of change in Y_i is being normalized by Y_i, giving us proportional change, or changes in Y_i as a proportion of Y_i.

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.