Political Science 150B/350B
Final Exam - Answer Guide
Winter 2006

Question 1: A political scientist performed an experiment in which he showed 20 subjects a negative ad against President Bush (treatment) and another set of subjects no ad (control). Then, he asked each respondent to report how much they liked President Bush on a scale (feeling thermometer) ranging from 0 (complete dislike) to 100 (complete liking). He also collected data on respondents party identification and their score on a ten-item political knowledge test:

- Negative Ad: 1=respondent saw negative ad; 0=respondent didn't see ad
- Republican: 1=Republican; 0=Democrat or Independent
- Democrat: 1=Democrat; 0=Republican or Independent
- Knowledge: 0-10 (number of questions correct on knowledge test)

	Min	Max	Mean
Negative Ad	0	1	.65
Bush Like Score	0	100	45.6
Republican	0	1	.32
Democrat	0	1	.41
Knowledge	0	10	2.6

The researcher used least squares regression, producing the following results:

	Estimate	Std Error
Negative Ad	-4.6	2.2
Republican	15.2	10.6
Democrat	-13.7	9.1
Knowledge	1.4	.5
Knowledge Squared	-.4	.2
Negative Ad × Knowledge	.5	.1
Negative Ad × Knowledge Squared	-.1	.1
Intercept	51.3	18.6

(a): (5 points) How many subjects are in the control group?

Answer: The mean of the negative ad dummy variable is .65, meaning that the 20 subjects who saw the negative ad constitute 65% of the total number of subjects: i.e., $20 = .65n$ or $n = 20/.65 \approx 31$ (to the nearest integer). Thus the control group has 11 respondents.

Grading notes: partial credit for answers with arithmetic errors.

(b): (5 points) Interpret the coefficient on "Negative Ad"

Answer: The coefficient on "Negative Ad" is the estimated effect of the negative ad treatment conditional on knowledge being zero (its minimum value). At this lowest level of political knowledge, the treated group had thermometer scores that were, on average, 4.6 units lower than those in the control group; this difference is statistically significant at conventional levels of statistical significance (i.e., t = -4.6/2.2 = -2.1, $p \approx .05$, two-tailed test with 31 - 8 = 23 degrees of freedom). This does not seem like a particularly large difference in substantive terms, given that thermometer ratings vary between 0 and 100.

Grading notes: 1 point for correct discussion of the magnitude of the coefficient, 1 for discussion of statistical significance; 1 point for correctly mentioning that the -4.6 effect was for knowledge=0, 2 for substantive interpretation.

(c): (10 points) Evaluate the impact of the treatment on the dependent variable, both in terms of statistical and substantive significance. How does the treatment effect vary with knowledge?

Answer: The total effect of the negative ad treatment requires looking at the way the estimated treatment varies as a function of the knowledge variable, via the interactions with knowledge and the square of knowledge. Note that the coefficient on the knowledge-squared/treatment interaction is the same size as its standard error ($t \approx 1.0$) and thus not statistically significant, while the linear interaction has an estimated coefficient of .5 with $t \approx 5.0$. This is key, because interpretations of the effect of the treatment differ dramatically depending on whether we include the quadratic interaction with knowledge or not; see Figure 1 (not necessary for answer). Ignoring the (statistically insignificant) quadratic term, we see that as one moves from low to high levels of political knowledge, the effect of the treatment is effectively washed out, and is zero (or close to it) at high levels of political information.

Grading notes: 2 points for correct interpretation of the linear interaction; 2 for correctly dispatching with the squared interaction; 4 for correctly interpreting the linear interaction as an attenuating effect of knowledge on the ad; 2 for correct substantive interpretation of the magnitude of the effect. A common error here was to say that those with high levels of political knowledge reacted positively to the attack ad (looking only at the 0.5 coefficient) rather than looking for the net effect of the treatment on those with 1 more point of political knowledge.

(d): (8 points) Evaluate the impact of party identification on the dependent variable, both in terms of statistical and substantive significance.

Answer: Party identification enters the analysis via two dummy variables, for Republican and Democrat identifiers, respectively. The coefficients on each dummy variable are large in substantive terms (15.2 and -13.7, respectively), but not statistically significant at conventional levels of significance ($t \approx 1.4$ and $t \approx -1.5$, respectively); note that these coefficients tell us how different Republicans and Democrats are from independents. The difference between Democrats and Republicans is estimated to be 28.9 units on the thermometer

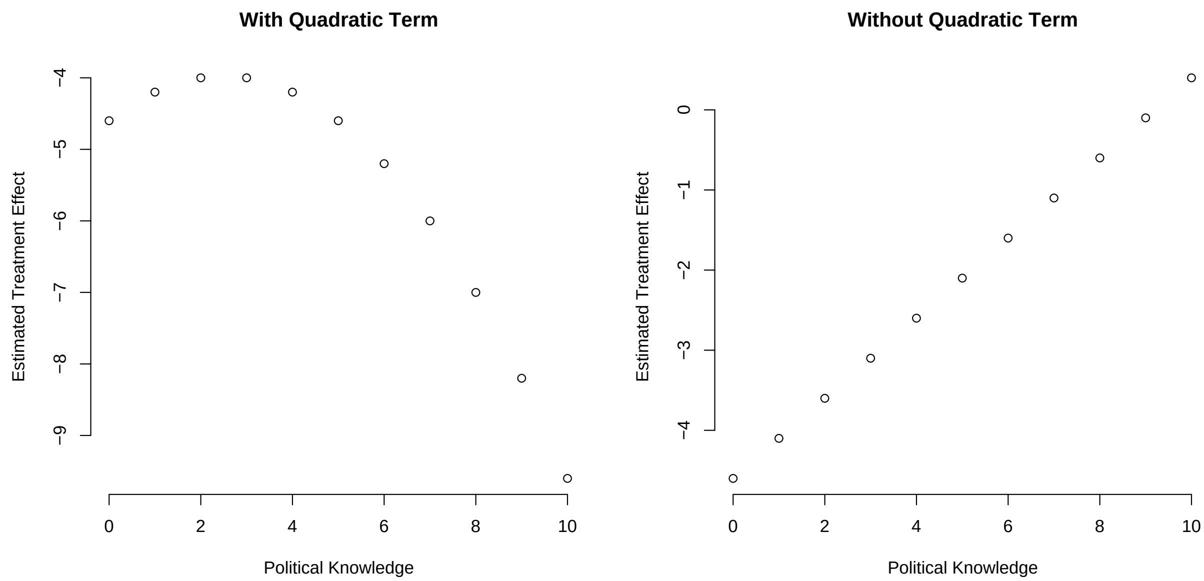

Figure 1: Queston 1c, negative ad effects, as a function of political knowledge

scale, but we don't have enough information to estimate whether that difference is statistically significant (see below).

Grading notes: 2 points of partial credit for knowing that the answer vaguely involves the square root of the sample size, three for having the correct formula for sigma hat squarted written down. Answers without a calculation of the magnitude of the effect received a maximum of 4 out of 5

(e): (8 points) The researcher is interested in a partisan symmetry hypothesis. That is, are Republicans as "pro-Bush" to the same extent that Democrats are "anti-Bush"? Can you test this hypothesis with the information given here? Why or why not?

Answer: This hypothesis could be tested with a t test of the null hypothesis the *sum* of the two coefficients is zero (i.e., the positive boost from Republican party identification is offset by the negative offset associated with Democratic party identification). The form of the test statistic is

$$t = \frac{\hat{\beta}_R + \hat{\beta}_D}{\sqrt{\text{var}(\hat{\beta}_R + \hat{\beta}_D)}}$$

where

$$\text{var}(\hat{\beta}_R + \hat{\beta}_D) = \text{var}(\hat{\beta}_R) + \text{var}(\hat{\beta}_D) + 2\text{cov}(\hat{\beta}_R, \hat{\beta}_D).$$

We have the first two terms of this expression: the estimated standard errors of $\hat{\beta}_R$ and $\hat{\beta}_D$ are (respectively) the square root of the corresponding variance term. We haven't been given the entire estimated variance-covariance matrix of $\hat{\boldsymbol{\beta}} = \hat{\sigma^2}(\mathbf{X}'\mathbf{X})^{-1}$ and so don't know the covariance term in the preceeding formula, and thus we can't implement this t-test.

Question 2: (5 points) A researcher estimates a regression with n = 100 iid observations. The researcher comes to you with a methodological question: if she were to go out and collect 5 times as much data, how much smaller would the standard errors of her regression be? What is your answer?

Answer: With independent data, statistical precision increases with the square root of sample size. So, on average (i.e., modulo sampling error), the research will obtain $\sqrt{5} \approx 2.23$ as much precision, or standard errors about $1/\sqrt{5} \approx 45\%$ the size of what the standard errors obtained with $n = 100$.

Grading notes: points of partial credit for knowing that the answer vaguely involves the square root of the sample size, three for having the correct formula for $\hat{\sigma}^2$ written down. Answers without a calculation of the magnitude of the effect received a maximum of 4 out of 5.

Question 3: Consider the regression model $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 W_i + \varepsilon_i$, where $W_i = X_{i1} + X_{i2}$.

(a): (5 points) What is *really* being asserted by the null hypothesis $H_0 : \beta_1 = 0$?
Answer: Substituting for W_i, the regression model is

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \\ &= \beta_0 + (\beta_1 + \beta_2) X_{i1} + \beta_2 X_{i2} + \varepsilon_i \end{aligned}$$

and under $H_0 : \beta_1 = 0$

$$Y_i = \beta_0 + \beta_2 X_{i1} + \beta_2 X_{i2} + \varepsilon_i.$$

Thus, the implication of the null hypothesis is that the partial effect of X_{i1} on Y_i is the same as the partial effect of X_{i2} on Y_i.

Grading notes: 3 points for correctly writing some version of the formula with $\beta_1 = 0$ and 2 points for discussing what it meant in words. The most common mistake was to think of the problem as positing an interaction between X_1 and X_2, rather than noticing that the equation would still be a simple linear model.

(b): (5 points) The previous question implies that H_0 can be tested via t-test. What is another way to test the assertion of H_0? Provide some brief detail as to how you would implement this test.

Answer: Another way to test this null hypothesis would be to fit a regression of Y_i on X_{i1} and X_{i2}, say,

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

and test the linear restriction implied by the null hypothesis $H_O : \beta_1 = \beta_2 \Rightarrow \beta_1 - \beta_2 = 0$ either with a t test or an F test.

Grading notes: 2 points for demonstrating some idea how an F-test works, 3 points for correctly stating the restrictions.

Question 4: (12 points) A resercher is interested in the effect on test scores of computer usage, and has relevent data aggregated to the level of school districts. The researcher regresses district average test scores on the number of computers per student. Will the estimated coefficient on the number of computers per student be an unbiased estimator of the effect of this variable on test scores? Why or why not? If you think the estimator is biased, is it biased up or down? Why?

Answer: Here is one plausible story: suppose socio-economic status has a positive effect on district test scores (net of computer usage). But SES will be positively correlated with computer usage. Thus, we're omitting something that has a positive impact on y and is positively correlated with the included regressor, and so we'll overstate the true effect of computer usage (since it is proxying for the omitted variable).

Grading notes: 3 points for noting omitted variable bias, 3 for telling a plausible story and 6 for being able to explain how we determine the sign on the bias. Most people did very well on this question.

Question 5: Consider the regression model: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ where $E(\varepsilon_i|X_i) = 0$, and (X_i, y_i) are iid draws from their joint distribution. Suppose that y_i is measured with error, so that the data are $\tilde{y}_i = y_i + w_i$, where w_i is the measurement error which is iid and independent of y_i and X_i. Consider the regression $\tilde{y}_i = \beta_0 + \beta_1 X_i + v_i$.

(a): (5 points) What is $E(v_i|X_i)$?

Answer: $E(v_i|X_i) = E[(w_i + \varepsilon_i)|X_i] = E(w_i|X_i)$ since $E(\varepsilon|X_i) = 0$. Note that we don't have enough information to conclude that $E(v_i|X_i) = 0$ because we haven't been told about $E(w_i|X_i)$. We know that w_i and X_i are independent, and thus $E(w_i|X_i) = E(w_i)$ but we don't know this quantity.

Grading notes: No one got this entirely correct. Partial credit for writing correct formulas for getting to $E(v_i|X_i)$.

(b): (5 points) Is the ordinary least squares estimate of β_1 in the regression of $\tilde{y}_i$ on X_i an unbiased estimate? Why or why not?

Answer: Measurement error in y_i that is uncorrelated with X_i and y_i does not violate any of the assumptions required to show that $\hat{\beta}_1$ is an unbiased estimator.

Not necc for answer: suppose $\hat{\beta}_1 = \sum q_i Y_i$ where

$$q_i = \frac{X_i - \bar{X}}{\sum (X_i - \bar{X})^2}$$

Taking expectations:

$$\begin{aligned}
E(\hat{\beta}_1) &= E(\sum q_i Y_i) \\
&= E[\sum q_i(\beta_0 + \beta_1 X_i + \varepsilon_i + w_i)] \\
&= \beta_0 E(\sum q_i) + \beta_1 E(\sum q_i X_i) + E(\sum q_i \varepsilon_i) + E(\sum q_i w_i) \\
&= \beta_1
\end{aligned}$$

since

(a) by construction, $\sum q_i = 0$ and $\sum q_i X_i = 1$

(b) $E(\varepsilon_i|X_i) = 0 \Rightarrow E(q_i \varepsilon_i) = 0$

(c) the assumption that w_i and X_i are independent implies that $E(q_i w_i) = 0$.

and these properties hold $\forall\, i$.

Grading notes: for full credit, had to mention that the measurement error in y was uncorrelated with X and that it would be absorbed in the usual error terms.

Question 6: Suppose that y and X are related by the regression $y = 1.0 + 2.0X + \varepsilon$. A researcher has observations on y and X, where $0 \leq X \leq 20$, where the conditional variance is $\text{var}(\varepsilon_i|X_i) = 1$ for $0 \leq X \leq 10$ and $\text{var}(\varepsilon_i|X_i) = 16$ for $10 < X \leq 20$.

(a): (5 points) Draw a hypothetical scatterplot of the observations (X_i, y_i), $i = 1, \ldots, n$.

Answer: See Figure 2 for some examples. Something resembling this will suffice.

Grading notes: maximum of three points if failed to label the y-axis, maximum of 4 if miscalculated the y-axis.

(b): (5 points) Does EGLS put more weight on observations with $X \leq 10$ or $X > 10$? Why?

Answer: EGLS will give proportionally less weight to the "noisier" observations, in this case those with $X > 10$.

Grading notes: partial credit for a discussion of data weighting.

Question 7: In November 1993, the state of Pennsylvania conducted elections for its state legislature. The result in the Senate election in the 2nd district (based in Philadelphia) was challenged in court. The facts of the matter are that the Democratic candidate won 19,127 of the votes cast by voting machines on election day, while the Republican won 19,691 votes cast by voting machines, giving the Republican a lead of 564 votes. However, the Democrat won 1,396 absentee ballots, while the Republican won just 371 absentee ballots, more than offsetting the Republican lead based on the votes recorded by machines on election day. The Republican candidate sued, claiming that many of the absentee ballots were fraudulent. The judge in the case solicited expert analysis from Orley Ashenfelter, an economist at Princeton University. Ashenfelter examined the relationship between absentee vote margins and machine vote margins in 21 *previous* Pennsylvania Senate elections in seven districts in the Philadelphia area over the preceding decade, yielding a data set with the following summary statistics:

Variable	Mean	Min	Max
x, Democratic Margin in Machine Balloting (percentage point lead among two-party votes)	40.68	-13.16	89.32
y, Democratic Margin in Absentee Ballots (percentage point lead among two-party votes)	29.06	-34.67	72.97

We now use linear regression to examine the relationship between y and x (as defined in the previous table), i.e.,

$$E(y_i|x_i) = \beta_0 + \beta_1 x_i$$

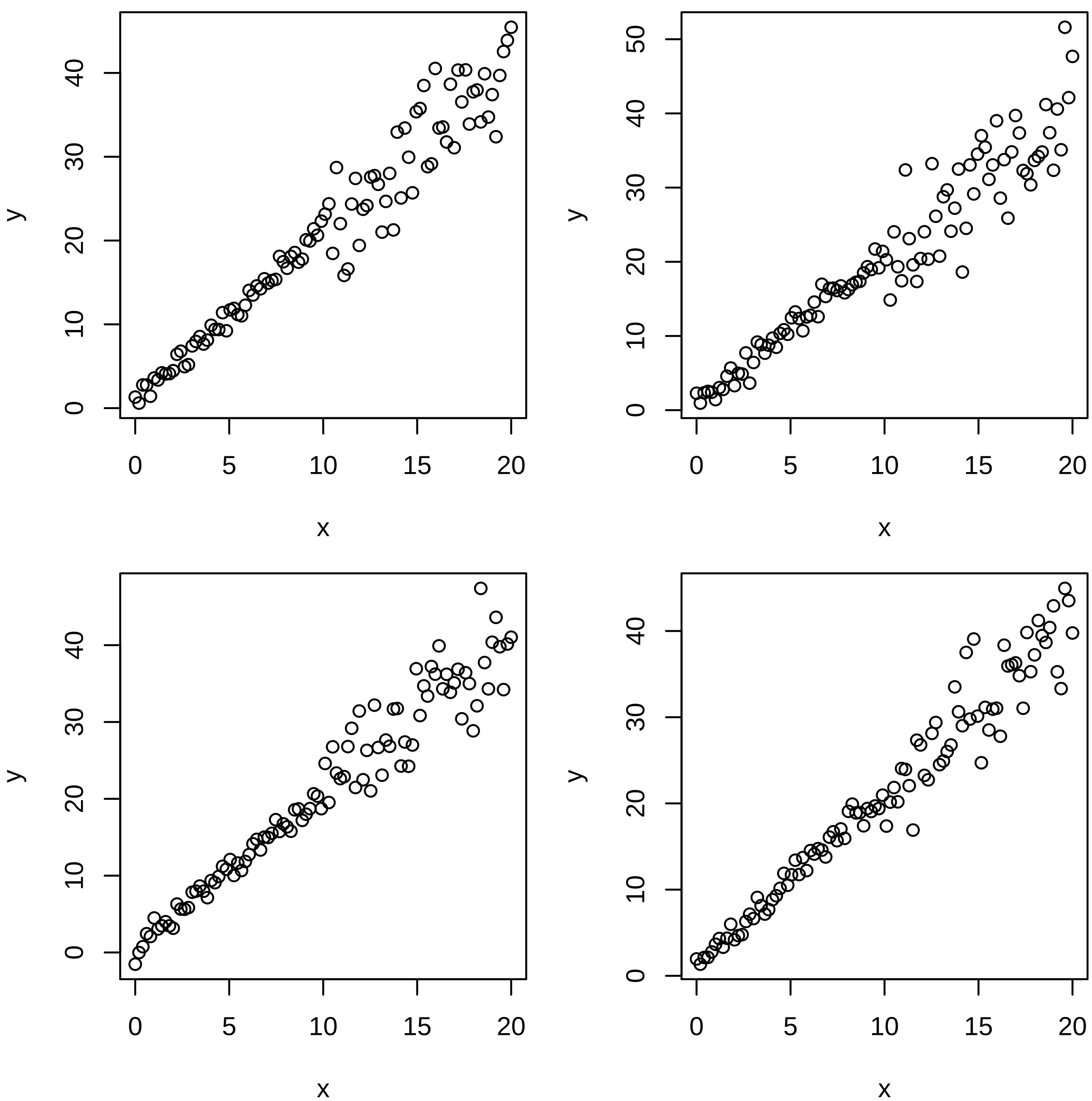

Figure 2: Hypothetical Heteroskedastic Data Sets, Q6a

and to assess the possibility that the disputed election (not included in the data set) is "suspect". For this regression, we have

$$\begin{aligned}
\mathbf{X}'\mathbf{X} &= \begin{bmatrix} 21 & 854.21 \\ 854.21 & 53530.38 \end{bmatrix} \\
(\mathbf{X}'\mathbf{X})^{-1} &= \begin{bmatrix} .14 & -.0022 \\ -.0022 & 5.3\times10^{-5} \end{bmatrix} \\
\mathbf{X}'\mathbf{y} &= \begin{bmatrix} 610.36 \\ 40,597.39 \end{bmatrix} \\
\mathbf{y}'\mathbf{y} &= 35,141.94
\end{aligned}$$

and the least squares estimate of σ^2 is 219.07.

(a): (5 points) What is the ordinary least squares estimate of $\boldsymbol{\beta} = (\beta_0, \beta_1)'$?

Answer:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= \begin{bmatrix} .14 & -.0022 \\ -.0022 & 5.3\times10^{-5} \end{bmatrix} \begin{bmatrix} 610.36 \\ 40,597.39 \end{bmatrix} \\
&= \begin{bmatrix} -3.86 \\ .81 \end{bmatrix}
\end{aligned}$$

Grading notes: partial credit for correctly stating the formula for $\hat{\boldsymbol{\beta}}$, and/or doing most of the algebra correctly.

(b): (8 points) What are the estimated standard errors of the least squares estimates?

Answer:

$$\begin{aligned}
\widehat{\text{var}(\hat{\boldsymbol{\beta}})} &= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \\
&= 219.07 \begin{bmatrix} .14 & -.0022 \\ -.0022 & 5.3\times10^{-5} \end{bmatrix} \\
&= \begin{bmatrix} 30.67 & -.4820 \\ -.4820 & .01161 \end{bmatrix}
\end{aligned}$$

(to four significant digits) and so $\widehat{\text{se}(\hat{\beta}_0)} = \sqrt{30.67} \approx 5.538$ and $\widehat{\text{se}(\hat{\beta}_1)} = \sqrt{.01161} \approx .1078$.

Grading notes: partial credit for correctly stating the formula for $\widehat{\text{var}(\hat{\boldsymbol{\beta}})}$, and/or doing most of the algebra correctly.

(c): (5 points) Provide an interpretation of the estimated *intercept* in the regression you just estimated.

Answer: $\beta_0 = E(y_i|x_i = 0)$ or in this case, the expected Democratic lead in absentee ballots conditional on an evenly matched two-party election as recorded with voting machines on election day. The estimate is -3.86, suggesting that Democrats suffer a comparative disadvantage in absentee ballots relative to machine ballots, but the estimate is not even close to being statistically significant at conventional levels ($t \approx$ -.70).

Grading notes: wrong answers to (a) and/or (b) did not automatically lead to deductions in the remainder of this question. Partial credit for stating that the estimate was not statistically significant.

(d): (5 points) For this regression model, consider the null hypothesis $H_0 : \beta_0 = 0, \beta_1 = 1$. Using terms that a non-specialist could understand, what is being assumed/asserted by model corresponding to H_0?

Answer: That on average Democratic margins among absentee ballots equals Demoratic margin among voting machines (or, put differently, the two populations of voters -- those turning to vote in person on election day and those voting absentee -- have the same partisan tendencies).

(e): (5 points) How would you test the null hypothesis given in the previous question?

Answer: *F* test of the joint null hypothesis.

(f): (6 points) Suppose that instead of least squares estimates, we wanted the maximum likelihood estimates of $\boldsymbol{\beta}$ and σ^2. How would your answers to the two questions change?

Answer: The only difference that MLE would make is with respect to the estimate of σ^2 and hence the estimated standard errors of $\hat{\boldsymbol{\beta}}$. The MLE of σ^2 is $\hat{\mathbf{u}}'\hat{\mathbf{u}}/n$, while the estimate reported is $\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n - k)$, and so we recover the MLE as

$$\hat{\sigma}^2_{\text{MLE}} = 219.07 \times \frac{n - k}{n} = 219.07 \times \frac{21 - 2}{21} = 198.21.$$

In turn, this implies that

$$\begin{aligned} \widehat{\text{var}(\hat{\boldsymbol{\beta}})}_{\text{MLE}} &= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \\ &= 198.21 \begin{bmatrix} .14 & -.0022 \\ -.0022 & 5.3\times10^{-5} \end{bmatrix} \\ &= \begin{bmatrix} 27.72 & -.4356 \\ -.4356 & .01049 \end{bmatrix} \end{aligned}$$

(to four significant digits) and so $\widehat{\text{se}(\hat{\beta}_0)}_{\text{MLE}} = \sqrt{27.72} \approx 5.265$ and $\widehat{\text{se}(\hat{\beta}_1)}_{\text{MLE}} = \sqrt{.01049} \approx .1024$.

Grading notes: partial credit given for stating the formulas correctly but not claculating new standard errors. More partial given for saying that the standard errors under MLE should be smaller than under LS.

(g): (5 points) With the information provided (and whatever you can deduce from that information) say what you can about how well the linear regression fits the data.

Answer: $\hat{\sigma} = \sqrt{219.07}$ or about 14.8, which is interpretable as the standard deviation of the data about the fitted regression surface. This is a relatively small amount of dispersion when one notes that the actual y data vary between -35 and 73 (a range of over 100 percentage points). Two standard errors above and below the regression line span about 60 percentage points, which suggests a moderate fit to the data.

And if you are really good, you'll see that you actually have enough information to compute r^2. Recall that

$$r^2 = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\sum(y_i - \bar{y})^2}.$$

We can recover $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ from $\hat{\sigma}^2$: i.e., since

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n - k} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{19} = 219.07$$

it follows that

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = 219.07 \times 19 = 4,162.39.$$

Also note that

$$\begin{aligned} \sum(y_i - \bar{y})^2 &= \sum\left[y_i^2 + \bar{y}^2 - 2y_i\bar{y}\right] \\ &= \mathbf{y}'\mathbf{y} + n\bar{y}^2 - 2\bar{y}\sum y_i \\ &= 35,151.94 + 21\times(29.06)^2 - 29.06\times 2\times 610.36 \\ &= 17,401.97, \end{aligned}$$

exploiting the fact that we are given $\bar{y}$ in the problem, or alternatively, noting that the first element of $\mathbf{X}'\mathbf{y}$ is actually $\sum y_i = 610.36$. We can then compute

$$r^2 = 1 - \frac{4,162.39}{17,401.97} = .761$$

which indicates quite a good fit to the data.

Grading notes: Partial credit given for a good attempt at calculating r^2. Partial credit for either calculating $\hat{\sigma}$ correctly but not interpreting it correctly.

(h): (5 points) The disputed election has $x = -1.45$ and $y = 58.01$. Carefully make a sketch with y on the vertical axis, and x on the horizontal axis, showing where the disputed data point lies in relation to the mean of the data and least squares regression line.

Answer: See Figure 3. Obviously you didn't have access to the entire data set. But it should be obvious that the disputed election lies a long way from where the regression line is; i.e., $E(y_i|x_i = -1.45) = -5.04$, versus the actual result of 58.01!

(i): (8 points) If the disputed election had been included in the data set used to estimate $\boldsymbol{\beta}$ and σ^2, do you think the estimates of these parameters would change much? Why or why not? In particular, what would happen to the estimate of the slope parameter: would it get bigger or smaller, or stay about the same?

Answer: For the disputed election, $x = -1.45$ is closer to the minimum of the x data than the mean of the x data, so one would imagine it would be a high leverage

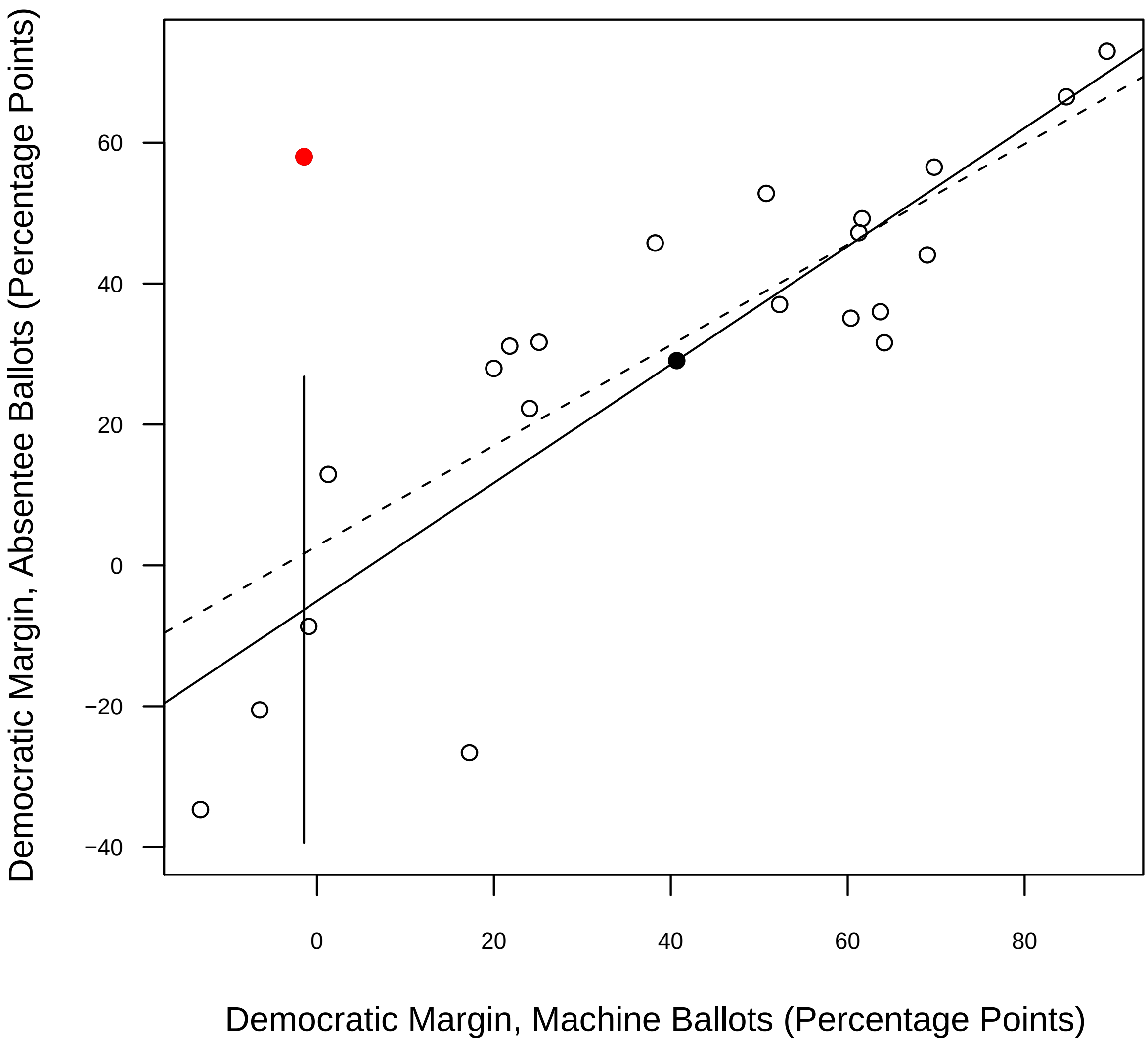

Figure 3: Pennsylvania State Senate Election Data. The disputed election is shown with a red dot. Solid line indicates least squares linear regression fit with 21 historical data points; dotted line shows how the regression line changes after including the disputed election. The solid vertical line indicates a 95% confidence interval for an observation with x equal to the Democratic voting machine margin observed in the disputed election (x = -1.45). The solid dot indicates the joint mean of the 21 historical data points.

data point. Including the disputed election as a data point in the analysis would cause the fitted regression to "tilt up" towards the disputed result, as indicated in Figure 3, making the slope estimate smaller (less negative, closer to zero).

The data point will not be fit particularly well, and so σ^2 will go up, the increase in the sum of the squared errors more than offsetting the increase from 19 to 20 in the denominator of the expression for $\hat{\sigma}^2$.

Grading notes: need three things for full credit (partial credit given for some): (1) saying point was far from mean of data; (2) saying point had high residual; (3) explaining that slope of line should flatten and making the parameter estimate smaller; and (4) saying that the fit should worsen with the inclusion of the point.

(j): (3 points) Based on the linear regression analysis of the historical data, what prediction would one make for the Democratic lead in postal ballots in the disputed election, given that the Democratic candidate trailed the Republican by -1.45 percentage points in machine ballots?

Answer: $E(y_i|x_i = -1.45) = -3.86 + (.81\times - 1.45) = -5.04$.

(k): (5 points) What is the standard error around the predicted value you computed in answering the previous question?

Answer:

$$\begin{aligned}\text{var}(\hat{y}|\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, x_0 = -1.45) &= \hat{\sigma}^2 \left[1 + \mathbf{x}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0'\right] \\ &= 219.07 \left(1 + [1, -1.45] \begin{bmatrix} .14 & -.0022 \\ -.0022 & 5.3\times10^{-5} \end{bmatrix} \begin{bmatrix} 1 \\ -1.45 \end{bmatrix}\right) \\ &= 219.07\times(1 + .1465) = 251.16,\end{aligned}$$

and so

$$\text{se}(\hat{y}|\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, x_0 = -1.45) = \sqrt{251.16} = 15.85.$$

Grading notes: Partial credit for correctly stating the formula for the variance of an individual prediction, but not correctly doing the matrix algebra.

(l): (8 points) Test the null hypothesis that the disputed election was generated by the same linear process that we have assumed for the 21 historical data points. That is, if the model fitted to the historical data is the correct, how plausible is the result observed in the disputed election? What would be your advice to the judge in the case?

Answer: Suppose the disputed election was generated by the same linear process that generated the 21 historical data points. Then, as shown above, given that the Democrat trailed by -1.45 percentage points on the votes cast by voting machine, we would expect a Democratic absentee ballot margin of -5.04, where the standard error around this prediction is 15.85 percentage points. We use a t distribution with $n - k = 21 - 2 = 19$ degrees of freedom to assess the plausability of the null hypothesis given the observed results. That is, if the regression model is true, then, in repeated sampling, conditional on seeing $x_0 = -1.45$, we would see absentee margins at least as extreme as the one obtained (i.e., $y = 58.01$) with probability given by the area under a t distribution with 19 degrees of freedom above and below $\pm|t|$, where

$$|t| = \frac{|\hat{y} - 58.01|}{\mathrm{se}(\hat{y})} = \frac{|-5.04 - 58.01|}{15.85} = 3.98$$

This value of t, 3.98, is a long way into the tail of the t_{19} distribution, with just 0.0004 of the probability mass lying above this value, which is the one-tailed probability of seeing a result such as the one obtained (or more extreme) under the null hypothesis.

Thus, we infer that the disputed election was almost certainly generated by a process other than the one generated the 21 historical election results. Put differently, in about 4 elections out of 10,000 would we see a result at least as anomalous as the disputed election, given the historical data. This is not evidence of voter fraud *per sé*, but clearly evidence that the result is *highly* irregular [and the judge agreed!].

Grading notes: Partial credit for stating that a t-test is appropriate, making a good attempt at calculating the t-stat (even if answer to (k) was wrong), getting the correct degrees of freedom, or stating the intuition that the influential observation follows a different linear process.

Question 8: Consider the first order autoregressive process, $z_t = \rho z_{t-1} + e_t$, $t = 1, 2, \ldots$

(a): (5 points) Describe the dynamics of this process if $\rho > 1$.

Answer: "Explosive" growth results, similar to exponential growth or a compounding process (like compound interest).

(b): (5 points) Assume $\rho = .9$. Calculate the $t = 4$ value of a "once-only" $t = 0$ 10 point shock (i.e., $e_0 = 10$ and $z_t = 0, \forall\, t < 0$).

Answer:

t	e_t	z_t	=	$\rho z_{t-1} + e_t$
0	10	10	=	.90 × 0 + 10
1	0	9.0	=	.90 × 10 + 0
2	0	8.1	=	.90 × 9.0 + 0
3	0	7.3	=	.90 × 8.1 + 0
4	0	6.6	=	.90 × 7.3 + 0

Or, $z_{t+j} = \rho^j z_t + e_t$, and so $z_4 = .9^4 \times 10 = 6.6$.

Grading notes: Partial credit for calculating value at t=3, or summing the values from t=0 to t=4.

Question 9: (8 points) What does it mean for an estimator to be consistent? If you can give a formal definition, be sure to also give an explanation in words that could be understood be a colleague who has not taken a class in statistics.

Answer: As sample size n tends to infinity, an estimator $\hat{\theta}_n$ converges on its value in the population θ. Formally, $\hat{\theta}_n$ is a **consistent** estimator of θ if

$$\lim_{n \to \infty} Pr(|\hat{\theta}_n - \theta| < \varepsilon) = 1$$

for any arbitrarily small positive ε. This definition of consistency is usually more compactly written as

$$\text{plim}\ \hat{\theta}_n = \theta.$$

Informally, as we throw more and more data at an estimation problem, the particular statistical procedure we are using (e.g., OLS, GLS, EGLS, MLE) gives answers that "get closer" to the population parameter. This is a weak property of an estimator, and indeed, if a statistical procedure should do anything, it should do this.

[Not necc for answer]: Two conditions are sufficient for establishing the consistency of an estimator:

(a): asymptotic unbiasedness: $\lim_{n \to \infty} E(\hat{\theta}_n) = \theta$

(b): asymptotic variance is 0: $\lim_{n \to \infty} var(\hat{\theta}_n) = 0$.

Grading notes: partial credit for imprecise statement of intuition (full credit for clear statement of intuition or formal statement).