

Note: This is an edited exam whose answer key differs slightly from the one used in Autumn 2005

Midterm (40 points): ANSWER KEY

Time allotted: 90 minutes. Be concise. Stick to concepts that were covered in class.

General grading guidelines: Don't be fussy about rounding error. If answer A is wrong, but answer B is consistent with answer A, only deduct points for answer A.

1. (7½ points) Recently the *New York Times* ran a story on apartment prices in Manhattan.
 - a. (3½ points) If you had to guess, what do you think the distribution of Manhattan apartment prices would look like? Sketch the distribution and describe its shape. Why do you think it has this shape?

1 point for drawing a right-skewed distribution.

½ point for labeling the axes (¼ point for each axis).

½ point for saying it's skewed; another ½ point for saying the skew is positive. (Full credit for the description if it agrees with the sketch, even if the sketch is wrong.)

1 point for saying the skew comes from the fact that there's a minimum possible price (maybe zero) but no maximum possible price.

- b. (2 points) In the *Times* story, the headline said that the average price was a million dollars. But the fifth paragraph said that the median price was \$500,000. Why might the average be so much larger than the median?

The average is sensitive to extreme values (½ point), but the median is not (½ point). If a few apartments had very high prices—and the skew suggests they do—this would pull up the average but not the median. (1 point)

- c. (2 points) Suppose you wanted to buy an apartment in Manhattan. Your realtor shows you an apartment priced at \$750,000. Are most of the apartments in Manhattan cheaper than this, or more expensive? How do you know? Be explicit.

Most of the apartments are cheaper than \$750,000. (1 point)

The median price is \$500,000 ($\frac{1}{2}$ point), so at least half of the apartment are priced at \$500,000 or less ($\frac{1}{2}$ point)

2. (7½ points) Over the past month, the University of Akron polled Ohio voters about Issue 4, a constitutional amendment that would change the way Ohio draws legislative boundaries. In a simple random sample of 409 likely voters, 178 said that they favored Issue 4; the other respondents opposed it.
- (1 points) On November 8, about 2.5 million Ohioans will vote on Issue 4. Predict the outcome of the upcoming election. Give a point estimate for the percentage of voters who will vote for Issue 4.

1 point: A point estimate is that $178/409=43.5\%$ will vote for Issue 4.

- (3½ points) Give a 95% confidence interval for the percentage of voters who will vote for Issue 4. Interpret this confidence interval.

1 point: The standard error of the point estimate is $\text{sqrt}(.435*(1-.435)/409)=.0245$, or 2.45%

1 point: So a 95% confidence interval is $43.5\% \pm 1.96*2.45\%$, or 38.7% to 48.3%.

1½ points: I am 95% confident (½ point) that between 38.7% and 48.3% (½ point) of all voters (½ point) will vote for Issue 4.

- c. (3 points) Assuming that the poll was done properly and the voters won't change their minds, do you predict that Issue 4 will win or lose? How sure are you of your prediction? Why are you sure or unsure?

1 point: I predict that Issue 4 will lose.

1 point: I feel sure (or at least 95% sure) of this prediction.

1 point: because even if the population's support level is at the top of the confidence interval, that's still less than 50%.

3. (6½ points) Suppose you weigh a simple random sample of kindergarteners and calculate a 95% confidence interval for the mean weight. The interval runs from 32 to 38 pounds. Indicate whether each of the statements below is a true or false interpretation of the confidence interval. If a statement is false, explain why.¹
- a. (1½ points) 95% of *all* kindergarteners weigh between 32 and 38 pounds.

½ point: False.

½ point: This is a confidence interval for the population *mean*

½ point: it does not apply to individual weights.

- b. (1½ points) 95% of the *sampled* kindergarteners weigh between 32 and 38 pounds.

¹ This question is adapted from Watkins, Scheaffer, and Cobb, *Statistics In Action*.

½ point: False.

½ point: The confidence interval applies to the population, not to the sample.

½ point: and the confidence interval concerns the average weight, not individual weights.

- c. (1 points) We can be 95% sure that the confidence interval includes the population mean.

1 point: True.

- d. (1½ points) The confidence interval may not include the sample mean.

½ point: False.

½ point: The confidence interval is centered around the sample mean.

½ point: thought it may not include the population mean.

- e. (1 point) If we took 100 different samples of children and calculated 100 different confidence intervals, about 5 of those intervals would not include the population mean.

1 point: True.

4. (2 points) Over the last two days of September 2005, *Newsweek* magazine surveyed a simple random sample of US adults, and found that 40% approved of the job George W. Bush is doing as President. At about the same time, Fox News surveyed a simple random sample of US adults, and found that 45% approved of Bush's performance.
- a. (2 points) Some people accuse *Newsweek* of a liberal bias, and some accuse Fox News of a conservative bias. The difference between these poll results could be offered as evidence of either bias. Can you offer a more benign explanation for the difference? Explain.

This could be ordinary sampling variation. If you take two samples from the same population, you'll get two different sample proportions.

5. (13½ points) As part of the Early Childhood Longitudinal Study, 100 school principals were asked how many days of instruction were scheduled for the academic year. The following frequency table summarizes the answers.

Number of school days	Number of schools
175	21
176	6
177	5
178	3
179	2
180	48
181	15

(Notice that, since there are exactly 100 schools, the frequencies are also the percentages.)

- a. (1 point) Describe the shape of this distribution.

It is bimodal. (1 point)

- b. (4 point) In this sample, what is the average length of the school year? The median length? The modal length?

It's vital to handle this as a frequency table. If you don't, you'll get the wrong answers.

2 points: The average length of the school year is 178.65 days. (178 is wrong.)

1 point: The median length is 180 days. (178 is wrong.)

1 point: The modal length is 180 days.

- c. (1 point) Interpret the mode in a sentence.

The most common length for the school year is 180 days.

It's wrong to say "most schools have a 180-day school year." Actually less than half (48%) have a 180-day school year.

- d. (3 points) Stop the presses! Some of the schools that reported a 175-day school year actually have shorter years than that (e.g., 174, 173, or 172 days). If the values for those schools were corrected, what would happen to the mean? the median? What would happen to the standard deviation? Explain all your answers. (No calculations are needed.)

**1 point: The mean would decrease ($\frac{1}{2}$ point),
because it would be pulled toward the new low values. ($\frac{1}{2}$ point)**

**1 point: The median would be unaffected ($\frac{1}{2}$ point)
because it's robust to extreme values ($\frac{1}{2}$ point).**

**1 point: The standard deviation would increase ($\frac{1}{2}$ point),
as it always does when extreme values are added. ($\frac{1}{2}$ point)**

- e. ($4\frac{1}{2}$ points) What is the first quartile? the third quartile? the interquartile range? Interpret each of these values in a sentence.

1½ points: The first quartile is 176 days. (½ point)

At least 25% [exactly 27%] of the schools have a school year of 176 days or shorter. (1 point. Deduct ½ point for omitting “or shorter”, and ¼ point for omitting “at least.”)

1½ points: The third quartile is 180 days. (½ point)

At least 25% [exactly 63%] of the schools have a school year of 180 days or longer. (1 point. Deduct ½ point for omitting “or longer”, and ¼ point for omitting “at least.”)

1½ points: The interquartile range is 4 days. (½ point)

At least half [exactly 64%] of the school-year lengths would fit in a four-day range. (1 point. Deduct ¼ point for omitting “at least.” Half credit for giving the endpoints of the IQR without saying that they’re 4 days apart.)

6. (3 points) In January 2004, the New York *Times* asked three dozen Des Moines bar, restaurant, and hotel workers which Presidential candidate they favored in the upcoming Iowa caucus.² The *Times* used this survey to correctly predict that John Kerry would win the caucus (most other surveys predicted a win by Howard Dean). The *Times*' reasoning was that hospitality workers had extensive contact with the candidates when they stayed in Des Moines.

List three problems with this way of conducting a survey.

- a. (1 point) **The wrong population was sampled. The population of interest is all Democratic voters in the state who planned to vote in the caucus. The population that was sampled was hospitality workers in Des Moines.**

- b. (1 point) **The sample was not taken in a random fashion.**

- c. (1 point) **The sample size was just 36. More are usually required to estimate a candidate's popularity with much precision. (In Lecture 9, for example, we found that, even in a landslide election, a sample of 100 voters is often not adequate to predict the winner.)**

² A caucus is a sort of primary election, in this case run by the Democratic Party.