

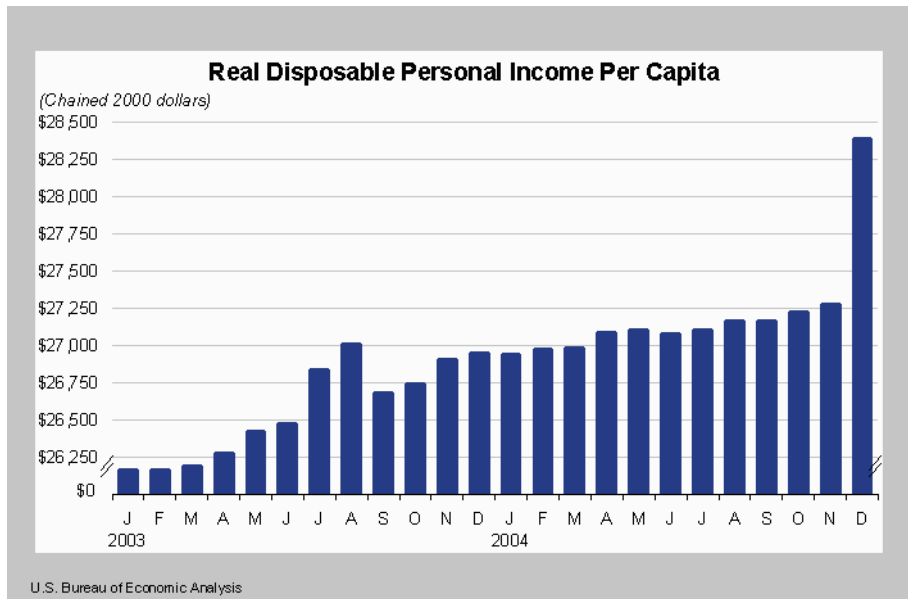
Your name _____

Midterm questions (43 points)

Time allotted: 90 minutes. Be concise!

General grading guideline: Don't be fussy about rounding error.

1. (8 points) The following graph is linked to the White House web site.



- a. (1 point) Despite its appearance, I would not call this graph a bar chart or histogram. What could it be called instead? (*Hint: What type of information does it contain?*)

Time series graph.

- b. (2 points) Taking the graph at face value, summarize the story it tells in a single sentence. Be specific but concise.

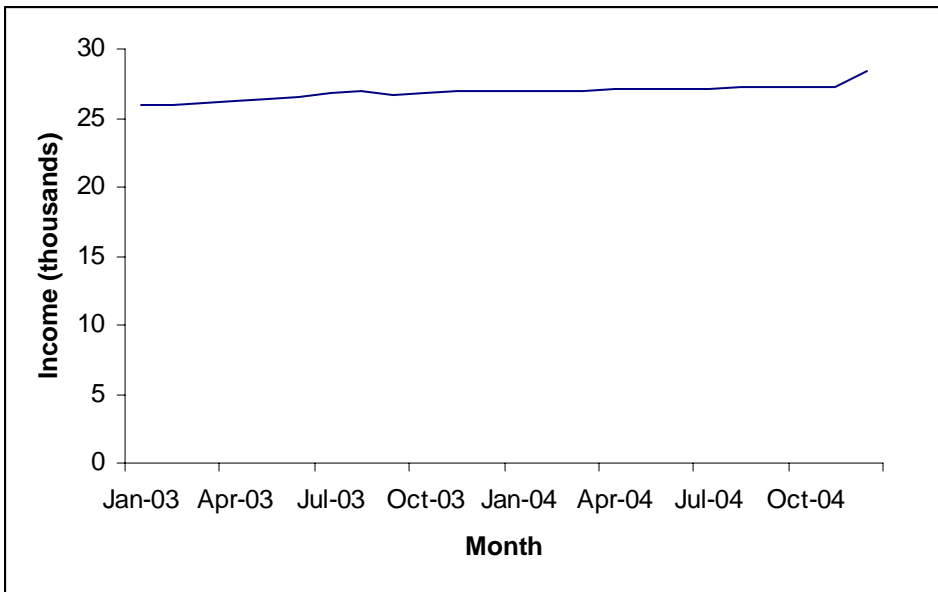
**Over the past two years, per capita income has risen by about \$2000 (or 7.5%).
(1 point for the length of time, 1 point for the amount of change, in either dollar or percentage terms.)**

- c. (2 points) In drawing the graph, did the artist make any choices that tend to emphasize or exaggerate changes? Point out the artist's one most important choice, and *explain its effect*.

The vertical axis has a break between \$0 and \$26,250, so that a 7% change in income covers almost 100% of the height of the graph.

(1 point for noticing the gap, 1 point for explaining how it affects perception.)

- d. (3 points) Suppose you were promoting a gloomier view of the economy. Redraw the graph, presenting the same information but changing the layout in *one* way to make the changes in the graph look smaller. Point out the change you have made, and *explain its effect*.
(Don't spend a lot of time drawing; a rough sketch is fine.)



The key is to start the vertical axis at zero and to avoid breaking it.

(1 point for saying this, 1 point if the drawing does it.)

This way a 7% change takes up closer to 7% of the vertical distance.

(1 point for the explanation.)

A second-best answer would be to stretch the horizontal axis, so that the slope of the increase is flattened. This is not nearly as effective, though, so it gets half credit for each part of the answer.

Zero credit for an unimportant change like relabeling the axes or rendering the graph in three dimensions.

2. (3 points) During the 2004 Presidential campaign, candidates debated the recent federal tax cuts. Democrats claimed that only the very rich received substantial cuts. President Bush claimed that the cuts also helped the poor and middle class.
- a. (1 point) According to the nonpartisan Tax Policy Center, across all filers the median tax cut was \$470, which is close to the figures given by Democrats. Interpret this median in a sentence.

At most half of all filers received a cut of more than \$470.

Or: at least half of all filers received a cut of less than \$470.

- b. (2 points) Also according to the Tax Policy Center, the average tax cut was \$1217, which is close to the figure given by President Bush. Why might the average be so much larger than the median? Explain, using a concept from class.

The average is sensitive to extreme values. (1 point)

If a few taxpayers received very large cuts, this would pull up the average but not the median. (1 point)

3. (6 points) Each month, Ohio's government estimates the state unemployment rate. The government's estimate is based on a simple random sample of 2,000 households.
- (2 points) If the true unemployment rate (in the population) is about 6%, what is the standard error of the government's estimate? Give your answer as a percentage. What is the "margin of error", again as a percentage?

The standard error is $[(.06)(1-.06)/2000]^{1/2}=.053$, or 0.53%.

The margin of error is about twice the standard error, or 1.06%.

- (2 points) Over the past month, the government's estimate of unemployment declined by 0.6%. Does this convince you that the employment situation has improved? Why or why not?

It doesn't convince me (1 point),

because the decline is smaller than the margin of error (1 point).

Note: If their margin of error was too big, then they might be convinced of improvement. Give them both points if their reasoning is consistent.

- (2 points) Suppose the employment situation has not in fact changed. Then why did the estimate change last month? Explain, using a key concept from class.

If the employment situation has not improved, then the change in the estimate could be due to *sampling variation* (1 point).

That is, if you take two samples from a population with 6% unemployment, the percentage unemployed will vary by 1% or so from one sample to another. (1 point)

4. (2 points) In the 1994-1998 Continuing Survey of Food Intakes by Individuals (CSFII), 21,662 people reported what they ate and drank on a given day. The surveyors used this information to estimate the number of calories consumed by each person.
- a. (2 points) The number of calories consumed has a minimum possible value (zero), but it doesn't have a maximum possible value. Bearing this in mind, what do you suspect is the shape of the distribution for calories consumed? Make a quick sketch. What is the name for this type of shape?

1 point for drawing a right-skewed distribution.

1 point for saying its right (or positively) skewed.

($\frac{1}{2}$ point for saying it's skewed but not saying right or left)

5. (6 points) The CSFII includes information on 13 black nine-year-old boys who were visited in 1996. Among these boys, the average number of calories consumed was 1887, and the standard deviation was 481.
- (2 points) Interpret the standard deviation in a sentence.

**For most of these boys, the amount consumed was within 481 calories of the mean.
Or: for nearly all of these boys, the amount consumed was within $2(481)=962$ calories of the mean.**

- (2 points) What is the Z score for a black nine-year-old boy who eats 2000 calories a day? Interpret this Z score in a sentence.

1 point: $Z=(2000-1887)/481=.23$.

1 point: 2000 calories would be .23 standard deviations above the average daily consumption of black 9-year-old boys. (½ off for not saying “above”)

- (2 points) Would 2000 calories be an unusual amount for a black nine-year-old boy to consume? Why or why not?

**It's not unusual, (1 point)
since it's only .23 standard deviations from the mean (1 point).**

6. (8 points) To repeat: in a simple random sample of 13 nine-year-old black boys, the average number of calories consumed was 1887, and the standard deviation was 481.
- (1 point) There are about 200,000 nine-year-old black boys in the United States. Give a point estimate for the average number of calories consumed in this population.

The sample mean, 1887, is a point estimate for the population mean. (1 point)

- (3 point) Calculate a 95% confidence interval for the same quantity.

These are the steps in calculating the confidence interval.

| | | <u>Points</u> |
|-----------------|-----------------|---------------|
| | N | 13 |
| | pt. est. | 1887 |
| | st. dev. | 481 |
| | std. err. | 133.4054 |
| (12 df, 95% CI) | t | 2.18 |
| | margin of error | 291 |
| | CI | (1596, 2178) |

Give a point for any CI that's consistent with t and SE, even if t and SE are wrong.

- (2 points) Interpret your confidence interval in a sentence.

Interpreting the confidence interval (2 points): “I’m 95% confident that the average number of calories consumed by nine-year-old black boys is between 1596 and 2178.” (1 point for saying it’s an average, 1 point for identifying the population.)

- d. (2 points) Suppose there were hundreds of boys in the sample, instead of 13. What would happen to your confidence interval? Give one or two reasons for your answer.

**The confidence interval would get narrower, (1 point)
because the standard error would be smaller (1 point)
or because the t statistic would be smaller (½ point).
1½ points for mentioning both the standard error and the t statistic. So a total of 2½ points are possible on this problem—of which ½ point is extra credit.**

7. (10 points) In the 2002 General Social Survey, respondents are asked how many people older than 18 live in their household. The table below summarizes the responses.

a. (2 points) Complete the table by filling in *all* the percentages and the cumulative percentages.

1 point for each column:

| Number of adults | Number of households | % | c% |
|------------------|----------------------|--------|---------|
| 1 | 1047 | 37.77% | 37.77% |
| 2 | 1364 | 49.21% | 86.98% |
| 3 | 258 | 9.31% | 96.28% |
| 4 | 75 | 2.71% | 98.99% |
| 5 | 28 | 1.01% | 100.00% |
| TOTAL | 2772 | | |

b. (2 points) Interpret in a couple of sentences the numbers in the two gray cells.

1 points: 9.31% of the sampled households had three adults living in them.

1 point: 96.28% of the sampled households had three adults or fewer living in them.

c. (2 points) What is the mean?

The mean is 1.80. (2 points, because it's a lot of work)

1 point on each for using the right formula, knowing what stands for what, and just miscalculating. 1½ points on each if the miscalculations are not serious.

No points for using the raw-data formula, which gives a mean of 3.

No points for using the frequency table formulas but not knowing what stands for what.

d. (2 point) What is the mode? Interpret the mode in a sentence.

1 point: The mode is two.

1 point (interpretation): More households have two adults than any other number.

e. (2 points) What is the median? Interpret the median in a sentence.

1 point: The median is also two.

1 point (interpretation): At least half the households (more precisely, 87%) have two adults or fewer. (½ point off a less precise answer)

f. (3 points) What is the variance? What is the standard deviation?

The variance is 0.63. (2 points, same guidelines as for the mean)

The standard deviation is 0.79. (1 point if it's the square root of their variance.)