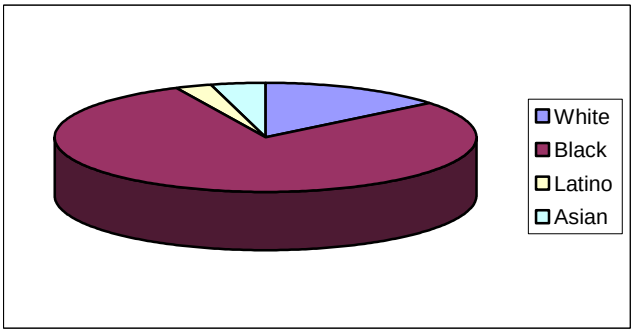
Midterm questions (53 points)
Time allotted: 90 minutes
Clock yourself!

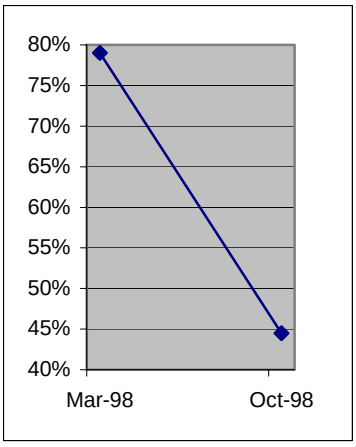
Please write your name and answers on the answer sheet provided—not on this question sheet.

1. (4½ points) During the week of March 7, 1997, Philadelphia police pulled over 262 drivers. Although Philadelphia is mostly white, whites made up only 14.1% of the drivers who were pulled over. 79.0% of the drivers pulled over were black, 2.7% were Latino, and 4.2% were Asian. (*Source*. Data from Philadelphia Police Department, compiled by Jeannine Sims.)
    a. (2½ points) Suppose you work for a civil rights advocacy group. Draw a pie chart that *emphasizes* how many of the drivers pulled over were black. Explain what you have done to emphasize black drivers.



1 point for a 3-D pie chart
1 point if blacks are in front.
½ point if the front wedge is highlighted using color or shading.

    b. (2 points) Under pressure, the police changed their behavior. By the week of October 6, 1997, black drivers were involved in just 44.5% of traffic stops. Suppose you work for the police department. Draw a time series that *emphasizes* the decline in black pullovers from March to October. Explain what you have done to emphasize this decline.



1 point if it's tall and skinny.
1 point if it starts above zero.

2. (26½ points) In assessing the impact of possible discrimination by traffic cops, it is important to know how often black drivers are pulled over. The table below describes a random sample of 9,415 black Americans who are old enough to drive, giving the number of times that each of them was pulled over (stopped) by police in 1999. (*Source*. Contacts Between Police and the Public: Findings from the 1999 National Survey.)
    a. (2 points) Complete the table by filling in the percentages and cumulative percentages.

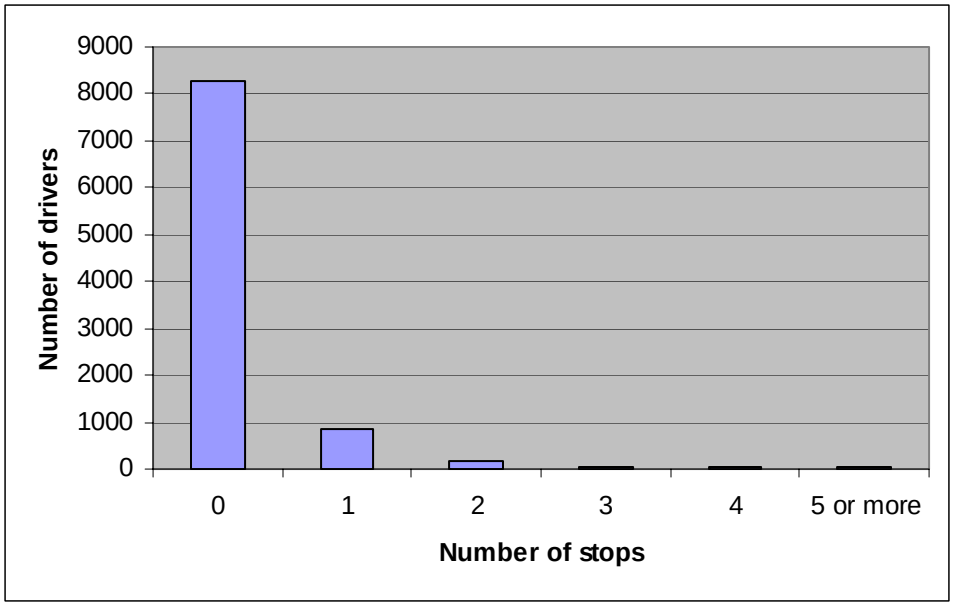
1 point for each column:

Number of stops	Number of drivers	%	c%
0	8257	87.70%	87.70%
1	867	9.21%	96.91%
2	173	1.83%	98.75%
3	51	0.54%	99.29%
4	28	0.30%	99.58%
5 or more	39	0.42%	100.00%

    b. (4 points) Interpret in a couple of sentences all of the numbers in the gray row, including the percentage and cumulative percentage.

2 points: 173 drivers, or 1.83% of the sample, were pulled over exactly twice.
2 points: 98.75% of the sample was pulled over less than three times.

    c. (2½ points) Draw a histogram of the data and describe its shape in technical language.



1 point for the histogram.
½ a point for saying its skewed
½ point for saying the skew is positive
½ a point for saying it's unimodal or has a prominent mode at zero.

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

1 point: The mode is zero.
1 point (interpretation): Most commonly, respondents hadn't been pulled over at all.

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

1 point: The median is zero.
1 point (interpretation): <u>At least (or more than)</u> half the respondents were not pulled over at all.

f. (4 points) Let's pretend that all the drivers with "5 or more" stops had *exactly* 5 stops. If this were true, what would be the mean? What would be the standard deviation?

Y	f	Yf	Y-Ybar	(Y-Ybar)2	f (Y-Ybar)2
0	8257	0	-0.1777	0.031576	260.7195
1	867	867	0.822305	0.676185	586.2526
2	173	346	1.822305	3.320795	574.4975
3	51	153	2.822305	7.965405	406.2356
4	28	112	3.822305	14.61001	409.0804
5	39	195	4.822305	23.25462	906.9303
SUM	9415	1673		SUM	3143.716
	MEAN	0.177695		Var	0.333941
				SD	0.577876

2 point for the mean, 2 for the standard deviation
No points for using the raw-data formulas. (If they use the raw data formulas, they'll get a mean of 2.5 and a standard deviation of 1.87.)
No points for using the frequency table formulas but not knowing what stands for what.

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.

g. (4 points) In truth, some of the drivers had *more* than 5 stops—one was stopped 40 times! Knowing this, would you say that the true sample mean is larger or smaller than what you have just calculated? What about the standard deviation? Explain.

1 point: The real mean is larger than what I have calculated,
1 point: because it includes larger values.
1 point: The real standard deviation is larger, too,
1 point: because it includes values that are further from the mean.

h. (2 points) Is the mean larger or smaller than the median? Why?

1 point: The mean is larger
1 point: because it's influenced by a few large values.

i. (2 points) Would the trimmed mean be larger or smaller than the mean? Why? (Don't do any calculations.)

1 point: The trimmed mean would be smaller
1 point: because it's not influenced by the large trimmed values.

j. (2 points) What is the *Z* score for someone with 1 stop? Interpret this *Z* score in a sentence.

1 point: Z=(1-.1779)/.579=1.42, or whatever value is consistent with the mean and standard deviation calculated earlier.
1 point: Someone with one stop is 1.42 standard deviations above the average number of stops. (½ off for not saying "above")

3. (16 points) There are perhaps 25 million black Americans of driving age. Describe these people using information from the previous question. As before, pretend that all respondents with "5 or more" stops had *exactly* 5 stops.

a. (1 point) What percentage of black driving-age Americans was stopped *at least* once? Give a point estimate, and interpret it in an ordinary sentence.

100-87.7=12.3%

b. (4 points) Give a 95% confidence interval for the quantity that you have just estimated. Interpret this confidence interval in a sentence.

Est.	**12.30%**
N	**9,415**
SE	**0.34%**
Z	**1.96**
LCB	**11.64%**
UCB	**12.96%.**

We are 95% confident that 11.64-12.96% of driving-age black Americans were stopped by police in 1999. (1 point for the standard error, 1 for Z, 1 for the interval, 1 for the interpretation.)

c. (2 points) Would an 85% confidence interval be wider or narrower than the interval you have just calculated? Why? (Don't do any calculations.)

1 point (rationale): The wider the interval, the more confident we can be that it covers the population mean. Or: the higher the confidence, the larger the value for *t*, and the larger *t* is, the wider the confidence interval.
1 point (answer): 85% interval is narrower than a 95% interval.

d. (2 points) Give a point estimate for the average number of times a black driving-age American was stopped by police in 1999. Interpret this point estimate in an ordinary sentence (that doesn't use the phrase "point estimate").

1 point: The sample mean, .178, is a point estimate for the population mean.
1 point: For black Americans of driving age, the average number of stops in 1999 was about .178. (½ off for not saying "about")

e. (5 points) Give a 99% confidence interval for the quantity that you have just estimated. Interpret this confidence interval in an ordinary sentence (that doesn't use the phrase "confidence interval").

with such a large df, t=Z=2.58

SE	0.005967
LCB	0.162505
UCB	0.193295

(1 point for the right t, 1 point for the right SE, 1 point for a confidence interval that's consistent with whatever they got for t and SE.)
2 points: I'm 99% confident that black driving-age Americans were stopped, on average, no more than .163 times and no less than .193 times. (1 point for getting the right population, 1 point for saying this is an average.)

f. (2 points) Your confidence interval is very narrow. Does this mean you should collect more data, or that you could have gotten away with collecting less? Explain.

1 point: We could have gotten away with a smaller sample.
1 point: Confidence intervals are narrower in larger samples. Here we have a very large sample and so a very narrow confidence interval. Since we don't need such a narrow confidence interval, we could have gotten away with a smaller sample.

4. (6 points) In 1998, doctors in New Jersey extracted eggs from over a hundred 40-year-old infertility patients. 48% of the extracted eggs were chromosomally abnormal.

a. (2 points) From what population were these women sampled?

The population of 40-year-old female *infertility* patients.
(1 point for age, 1 point for infertility)

b. (2 points) In April 2002, *Time* magazine ran a cover story on the "hard facts" about fertility and age. Citing the New Jersey study, *Time* wrote: "At 40, half a woman's eggs are chromosomally abnormal." Is this a reasonable interpretation of the study results? Why or why not?

It's unreasonable because it refers implicitly to the whole population of women. But the sample only justifies inferences about women who go to infertility clinics.
(½ point for saying it's unreasonable and relating this to the population,

1½ points for pointing out that the population is infertile.)

c. (2 points) Do you think that abnormal eggs are more common, less common, or about as common as the quote from *Time* implies? Why?

I think women in infertility clinics probably have more abnormal eggs than other women. So in general abnormal eggs are probably less common than *Time* implied. (½ point for saying they're less common and relating this to the population. 1½ points for saying the general population has healthier eggs than infertility patients.)