Show your work. The test is out of 100 points and you have 50 minutes.

1. The following information came from **FOX NEWS**, October 10 2008:

> Drinking red wine not only reduces your risk for cardiovascular disease, but it may also reduce your risk for lung cancer, especially if you are a current or ex-smoker, Reuters reported Thursday.
>
> People who do or have smoked and drink at least one glass of wine each day are 60 percent less likely to develop lung cancer than those who have smoked and don't drink red wine, said Dr. Chun Chao, of the Kaiser Permanente Southern California in Pasadena.
>
> Chao said it's the resveratrol and flavonoids in red wine that are protective – something white wine does not have.
>
> The reduction seen with red wine "lends support to a causal association for red wine and suggests that compounds that are present at high concentrations in red wine but not in white wine, beer or liquors may be protective against lung carcinogenesis," Chao wrote in her study.
>
> However, previous studies examining the correlation between alcohol consumption and lung cancer haven't always had the same results, Chao and her team noted in the journal Cancer Epidemiology, Biomarkers and Prevention.

(a) (4 point) Was the study a controlled experiment or an observational study? How do you know?

It was an observational study because they did not control how much wine the people drank - they just recorded what people chose to do.

(b) (8 points) Clearly explain why socioeconomic status could be a confounding factor in this study and why this might make you doubt their conclusion.

high Socioeconomic status

more red wine   lower lung cancer rates
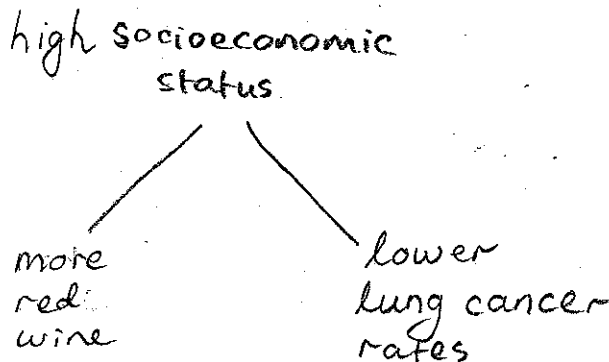
Red wine is expensive. The people who can afford red wine are probably wealthier, on average. People who are better-off may tend to have lower exposure to harmful chemicals (eg asbestos, factory pollutants) and therefore less susceptible to lung cancer.

2. The wingspan of 58 tundra swans gave the following information:

Width	Wingspan	Number of swans	percentage	height = percentage/width
10	175–185	5	$\frac{5}{58} \times 100 = 8.6$	.9
10	185–195	18	$\frac{18}{58} \times 100 = 31.0$	3.1
10	195–205	21	$\frac{21}{58} \times 100 = 36.2$	3.6
15	205–220	14	$\frac{14}{58} \times 100 = 24.1$	1.6
		$\overline{58}$	$\overline{99.9 \approx 100\%}$	

(a) (10 points) Sketch a histogram. Be sure to label the axes.
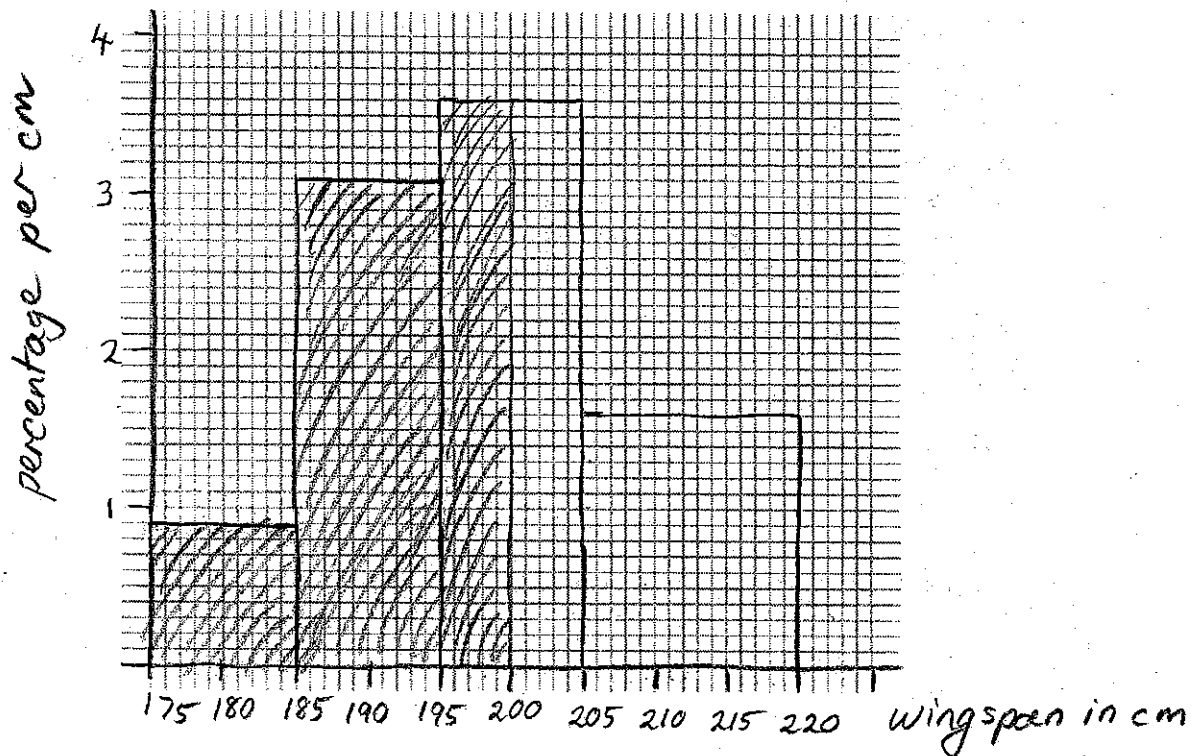
(b) (4 points) What percentage of the swans have a wingspan of less than 200 cm? (Use the histogram). See shaded area
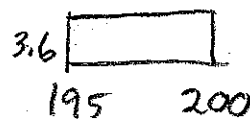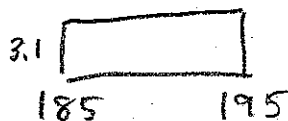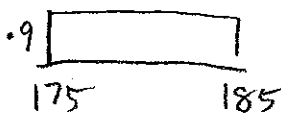
area $10(.9) = 9$ $10(3.1) = 31$ $5(3.6) = 18$

area = $9 + 31 + 18 = \underline{58\%}$

3. (10 points) A journalist plots the percentage of high-income people versus the percentage of foreign-born people in each state, giving a plot with 50 points. The correlation coefficient is 0.5. The journalist concludes that foreign-born people earn more, on average, than native-born people. Give TWO DIFFERENT mistakes that the researcher is making. Explain clearly.

1. This is an ecological correlation so the correlation coefficient does not tell us about individual people. (In fact, in this data, the foreign-born had lower salaries, on average, than the native-born).

2. This is observational data so there are confounding factors. Perhaps foreign-born are concentrated in states that have high incomes, but they may not be the people with those high incomes.
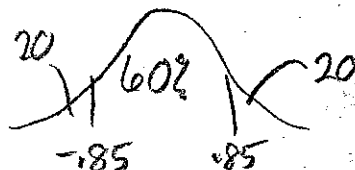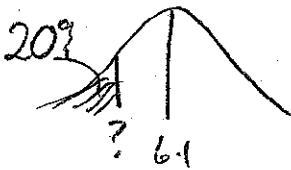
4. In November 2008, unemployment rates for the 50 States followed the normal curve with an average of 6.1 and an SD of 1.5.

(a) (10 points) What percentage of the States had unemployment rates higher than 4.0?

$z = \frac{4 - 6.1}{1.5} = -1.4$

$100 - 84 = 16\%$

so 8% in each

$84 + 8 = 92\%$
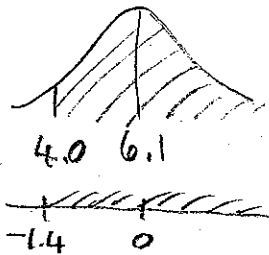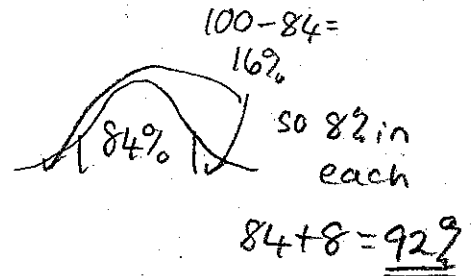
(b) (10 points) If I tell you that Oklahoma was at the 20th percentile, what was the unemployment rate in Oklahoma?

$x = 6.1 + (-.85)(1.5) = 4.825$

(c) (3 points) Wyoming had the lowest unemployment rate, while Michigan had the highest. If these two states were removed from the list, would the SD for the remaining 48 states be smaller than 1.5, equal to 1.5, or larger than 1.5? (No explanation required).

Variability goes down if the extremes are removed.

(d) (3 points) Would the median unemployment rate be quite a bit smaller than 6.1, about equal to 6.1 or quite a bit larger than 6.1? (Remember, the histogram looks like the normal curve). (No explanation required).

The median will be about the same as the average, because the normal curve is symmetric

ave = median.

5. The chest circumference and wingspan of 58 tundra swans have the following summary statistics:

X Circumference: average = 63.0 cm   SD = 3.4 cm   r = 0.5
Y Wingspan: average = 198.0 cm   SD = 9.8 cm

The scatter-diagram is football-shaped.

(a) (7 points) Find the regression equation for predicting wingspan from circumference.

$$\text{slope} = r\left(\frac{SD_y}{SD_x}\right) = .5\left(\frac{9.8}{3.4}\right) = 1.4412$$

$$\text{intercept} = \text{ave}_y - \text{slope}(\text{ave}_x) = 198 - (1.4412)(63) = 107.2$$

$$y = 107.2 + 1.4412\, x$$

$\uparrow$ wingspan $\uparrow$ circumference

(b) (7 points) Estimate the wingspan of a tundra swan that is 58 cm in circumference.

$$y = 107.2 + 1.4412(58) = 190.8$$

or: $z = \dfrac{58-63}{3.4} = -1.47$

new $z = (.5)(1.47) = -.74$

$$y = 198 + (-.74)(9.8) = 190.8$$

(c) (6 points) Would you be surprised if the swan from part (b) has a wingspan of 208 cm? Explain.

$$\text{rms error} = \left(\sqrt{1-.5^2}\right)(9.8) = 8.5$$

We expect them to be between

$$190.8 - 2(8.5) = 173.8$$

and $190.8 + 2(8.5) = 207.8$

So yes, I would be somewhat surprised, because it's outside the interval (but only just outside).

(d) (8 points) Sketch the scatter-diagram.

$ave_x = 63$ $SD_x = 3.4$

$ave_y = 198$ $SD_y = 9.8$

$r = .15$

$198 - 2(9.8) = 178.4$
$198 + 2(9.8) = 217.6$

Wingspan in cm

Circumference in cm

$63 - 2(3.4) = 56.2$
$63 + 2(3.4) = 69.8$
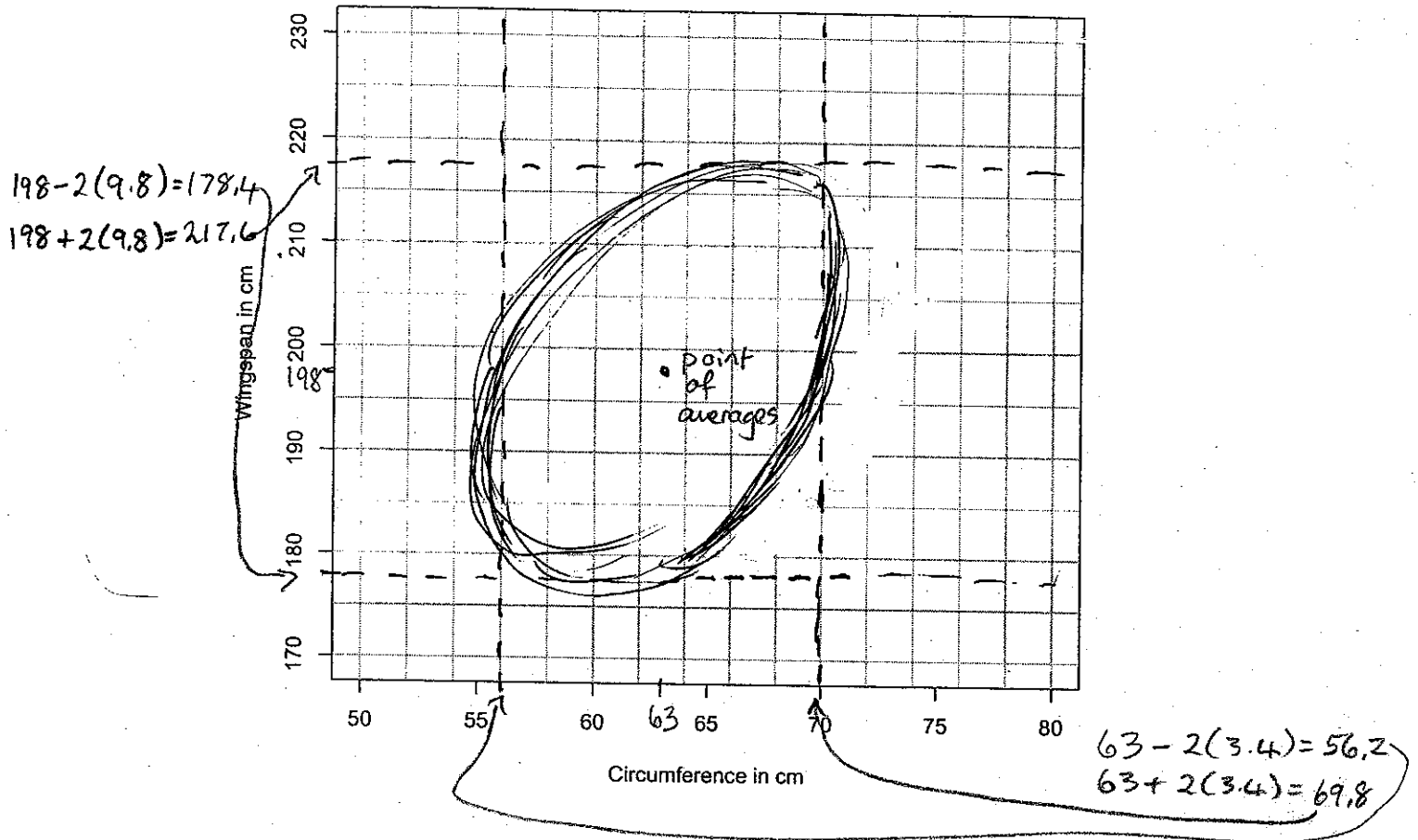
6. (10 points) A market researcher takes a sample of women aged 20 to 70 and finds a negative correlation between age and the number of movies seen per year. She concludes that as people age, they go to the movies less often. Is this necessarily true? If not, what could account for the negative correlation?

This is a cross-sectional study. To draw conclusions about what happens as people age, you need to follow them as they age (longitudinal). It could be that the older women never went to very many movies and their behavior has not changed since they were 20. Similarly, perhaps these 20 year olds will continue to go to the same number of movies as they grow older.