1. On March 8, 1993 *Newsweek* announced "A really bad hair day: Researchers link baldness and heart attacks". The article reported that "men with typical male pattern baldness are anywhere from 30 to 300 percent more likely to suffer a heart attack than men with little or no hair loss at all.". The report was based on a study conducted at the Boston University School of Medicine. They compared 665 men who had been admitted to the hospital with their first heart attack to 772 men who had been admitted for other reasons. They found that 42% of the heart attack victims suffered from male pattern baldness, while only 34% of the men admitted for other reasons suffered from male pattern baldness.

   (a) (3 points) Is this an example of a controlled experiment or an observational study?

   Observational study

   (b) (8 points) Suggest a confounding factor and clearly explain why you think it is a confounding factor.

   age - baldness increases with age, and perhaps men who are admitted with heart attacks would be older, on average, than those admitted for other reasons.

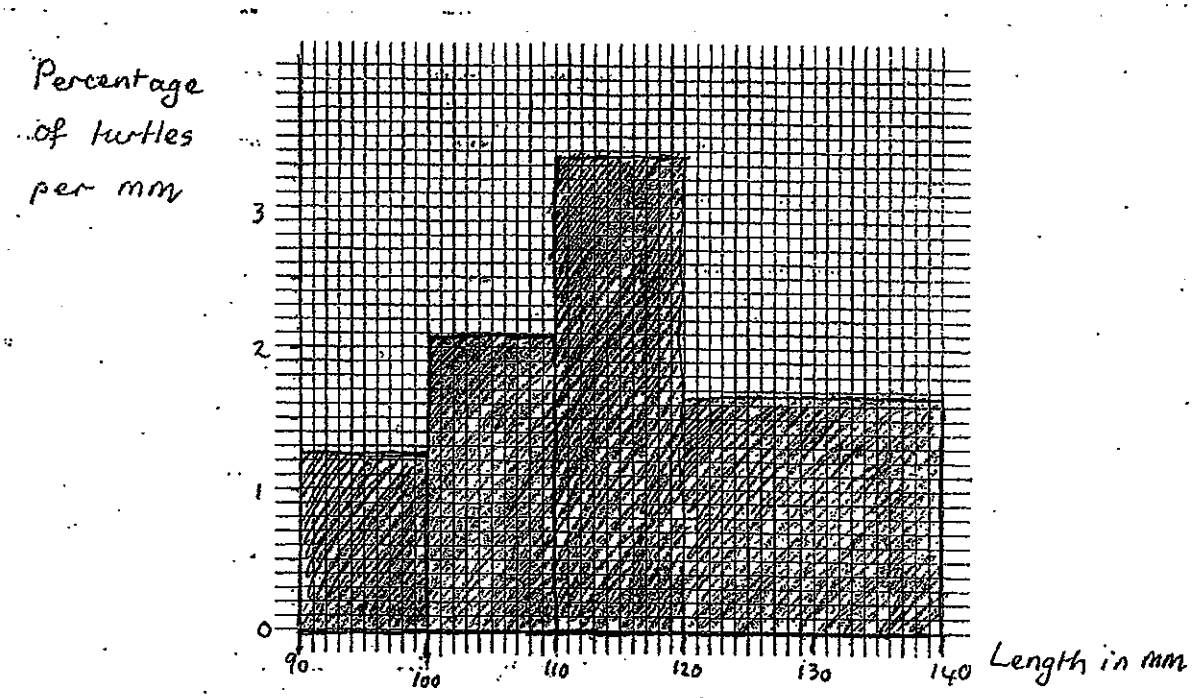
   (c) (4 points) If you had access to all the hospital records for these men, what would you do to help alleviate the confounding?

   Look at men in narrow age groups separately. eg. compare men 45-50 admitted with heart attacks to men 45-50 admitted for other reasons.

2. The following table summarizes the lengths of 24 male painted turtles. Class intervals include the left endpoint but not the right.

Width	Length (mm)	Number of turtles	Percentage of turtles	Height = Percentage / Width
10	90 to 100	3	$\frac{3}{24} \times 100 = 12.5$	$12.5/10 = 1.25$
10	100 to 110	5	$\frac{5}{24} \times 100 = 20.8$	$20.8/10 = 2.08$
10	110 to 120	8	$\frac{8}{24} \times 100 = 33.3$	$33.3/10 = 3.33$
20	120 to 140	8	$\frac{8}{24} \times 100 = 33.3$	$33.3/20 = 1.665$
		24	$99.9 \approx 100$ ✓	

(a) (10 points) Draw a histogram for the data, being careful to label the axes correctly.

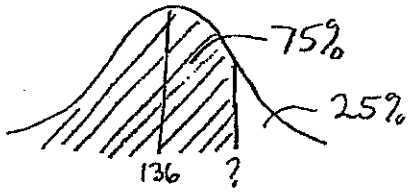


(b) (4 points) In which interval does the 25th percentile fall?

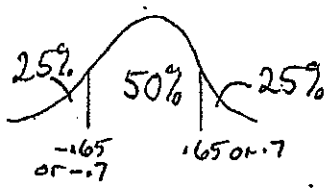
The 25th percentile has 25% to the left. There is 12.5% in the first block and 20.8% in the second, so the 25th percentile must fall in the second interval. (a little over half was across).

3. The length of female painted turtles follows the normal curve with an average of 136 mm and an SD of 21 mm.

(a) (10 points) If one of these turtles is at the 75th percentile in length, how long is she?

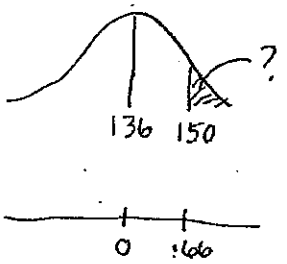


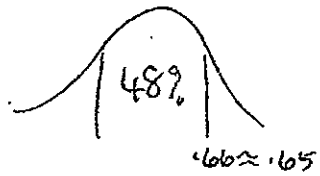
So



$X = (.7)(21) + 136 = 150.7$

(b) (10 points) What percentage of the turtles were more than 150 mm in length?





$\frac{150-136}{21} = .66$

So $100 - 48 \approx 52\%$ and the percentage is approx 26%.

(c) A new turtle is included and she is longer than any of the others.

i. (3 points) The average will now be (circle the correct answer, no explanation is required):

- smaller than before
- exactly the same as before
- larger than before

ii. (3 points) The SD will now be (circle the correct answer, no explanation is required):

- smaller than before
- exactly the same as before
- larger than before

more variability because she is extreme

Note: if she had been shorter than the others, part (i) would change to smaller but part ii would still be larger because variability would go up. (she would still be extreme).

4. (7 points) In a special "health awareness" program, all university employees are invited to participate in a wellness program in which they get screened for cholesterol, blood pressure, hearing loss, etc. They define high blood pressure as systolic blood pressure over 140 mm and they suggest that any participant with high blood pressure should return for a followup measurement. Other things being equal (stress levels, nerves, etc) do you expect their second systolic blood pressure measurement to be higher, lower, or the about same as their first measurement? Explain.

We expect it to be lower due to the regression effect (these people probably got unusually high measurements, for them, on the first day). It's just chance error at work. We should still expect their measurements to be high, but not quite as high as before.

extra

(7 points). Among male faculty, it is noticed that high salary is positively correlated with wearing a necktie: Does this tell you that if a male faculty member starts to wear a necktie then he can expect his salary to increase? If not, what does it tell you? Explain carefully what you can, and cannot, conclude about neckties and salary.

No, you can't conclude that wearing a necktie causes your salary to increase. It's more likely that male faculty in administrative positions (department heads, Deans, Provosts, Presidents) are expected to wear neckties and they have higher salaries because of their positions, not because of how they dress. Association ≠ causation.

5. The length and width of 24 male painted turtles have the following summary statistics:

Length: average$_x$ = 115 mm SD$_x$ = 12 mm r = 0.95
Width: average$_y$ = 90 mm SD$_y$ = 8 mm

The scatter-diagram is football-shaped.

(a) (10 points) Predict the width of a turtle that is 95 mm in length.

95 mm is 20 mm below ave.
That is $\frac{20}{12}$ = 1.67 SD's " "
Expect them to be (.95)(1.67) = 1.58 SDs below ave in width
(1.58)(8) = 12.67 below 90 is 77.3

$z = \frac{95-115}{12} = -1.67$

(.95)(-1.67) = -1.58

$x = (-1.58)(8)+90 = 77.3$

or:

$\text{slope} = r\frac{SD_y}{SD_x} = .95\left(\frac{8}{12}\right) = .633$

$\text{intercept} = \text{ave}_y - \text{slope}(\text{ave}_x)$
$= 90 - .633(115)$
$= 17.2$

$y = 17.2 + .633x$
$y = 17.2 + (.633)(95) = 77.3$

(b) (8 points) Would you be surprised to find that the turtle in part (a) was 110 mm in width? Explain.

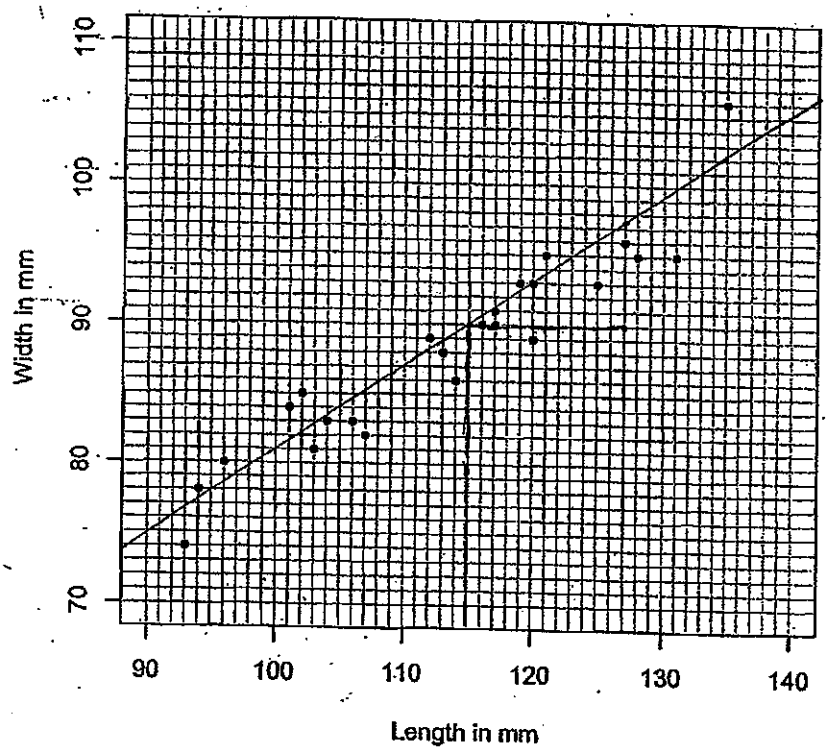
rms error = $\sqrt{1-(.95)^2}\ 8 = 2.5$

$77.3 - 2(2.5) = 72.3$

$77.3 + 2(2.5) \doteq 82.3$

Yes, I'd be amazed - it's way outside my interval!

(c) (8 points) Draw the regression line.



$\frac{\cdot 95(8)}{12} = 7.6$