Name_____ **Student #** _____

STAT 101
Midterm Examination: Key

Richard Lockhart 22 February 2012

Instructions:

1. This is an open book exam; you may use your text or another text and any notes you wish to bring.
2. You may use a calculator (with no wireless communications ability).
3. If you don't have normal tables the invigilators have some; just stick up your hand and ask.
4. Sometimes, to do the problem, you will need to make assumptions. You should be clear and explicit about what assumptions you need to make the technique you are using reasonable. You need to write these down in complete sentences; sentence fragments will get 0 marks.
5. The exam is out of 20.
6. **DON'T PANIC.**

1. In a study of husband-wife nutrition patterns a sample of 1482 married couples from the island of Oahu, Hawaii, were asked about their dietary habits. The husbands consumed an average of 11600 calories per week with a standard deviation of 4100 calories. The wives consumed an average of 9300 calories per week with a standard deviation of 3100 calories. The correlation between husband's weekly calorie consumption and wife's weekly calorie consumption was 0.35.

   (a) About what percentage of the husbands consume more than 20000 calories per week? Be clear about any assumption you must make to do the problem. (5 marks)

   You must assume that the histogram for the calories consumed by husbands follows the normal curve. [1 mark] *No marks for 1 word answers like "normal" or even "the distribution is normal without saying "the distribution of father's calorie consumption is normal" or some other phrase which makes clear what distribution is being discussed.*

   Now convert the number 20000 to standard units:

$$\frac{20000 - 11600}{4100} = 2.05.$$

   One mark for this. Perhaps half a mark off for using the wrong mean, half a mark for the wrong SD.

   Then look the area to the left of 2.05 up in the table and get 0.9798. [1 mark] *Finally the desired proportion is the area to the right which is 1-0.9798 = 0.0202=2.02%.* [2 marks]. *Dock half a mark for not converting to a percentage.*

   (b) If a husband consumes 20000 calories per week what would you predict for the wife's weekly calorie consumption? (5 marks)

   This is a regression problem in which the dependent variable is the wife's weekly calorie consumption and the predictor or independent variable is the husband's weekly calorie consumption. If you get those backward somehow you should not get more than 2.5 marks in total.

   Most students will compute the slope via

$$b = r\frac{s_y}{s_x} = 0.35\frac{3100}{4100} = 0.2646$$

   and

$$a = \bar{y} - b\bar{x} = 9300 - 0.2646 \times 11600 = 6230.$$

   I have promised not to mark them on the subject of how many digits they use. If they used only 3 digits or 2 digits for b there estimate of the intercept will be off and if they used more digits than I did their answer might also be slightly different. Give them 3 marks for finding the correct slope and intercept and don't be fussy about the number of decimal places.

   Then they need to put the number $x = 20000$ into the equation for the regression line to get the prediction

$$y = 6230 + 0.2646 \times 20000 = 11522 \textit{calories per week.}$$

   Deduct 1 mark for failing to have units for the final answer.

(c) Consider the families where the husband consumes about 20000 calories per week. Approximately what would be the standard deviation of the weekly caloric intake of the wives in these families? (1 mark)

The residual standard error, that is, the standard deviation of the y values for the cases with some particular x value (here $x = 20000$) is

$$\sqrt{1 - r^2} s_y = \sqrt{1 - .35^2} 3100 = 2904 \textit{calories per week.}$$

Again I don't care about the number of digits.

(d) The authors of this study were aware that larger people eat more so they also studied calories per kilogram of body weight: for each person they divided the weekly caloric intake by that person's weight. The husbands averaged 171 calories per kilogram of body weight each week while the wives averaged 168 calories per kilogram of body weight. The standard deviations of these variables were 63.7 calories per kilogram for the men and 63.2 calories per kilogram for the women. The correlation between these husband and wife for these variables was 0.48. The authors of the study were concerned about whether the difference between husband's intake and wife's intake was "consistent" so they looked at the group of couples where the husband's calorie consumption per kilogram was low – well below average. Why should they not have been surprised to learn that the majority of wives in this group had higher calorie per kilogram levels than their husbands? (1 mark)

This is the regression effect. For husbands below average on this variable we expect the wives to be below average but not by as many standard deviations as their husbands. In this case the standard deviations for men and women are similar so the women will be fewer calories per kilogram below average than their husbands. I want this one marked with a tough eye. Half a mark for regression effect but to get full marks the explanation needs to be good and clear.

2. A study of 1000 adults shows mean heights of 66 inches with a standard deviation of 4 inches. It also shows mean weights of 150 pounds and a standard deviation of 30 pounds. The correlation between height and weight is 0.45.

   The authors of the study decide to extend their work so they gathered a sample of 1000 children aged 6 to 8 and measured the heights and weights of these children.

   If the authors now put all 2000 people together in 1 group:

   (a) Will the mean height be more than, less than or about the same as 66 inches?
   The children are shorter than the adults so the mean height will go down.

   (b) Will the standard deviation of the heights be more than, less than or about the same as 4 inches?
   Addition of these shorter children will spread out the histogram driving the standard deviation up.

   (c) Will the correlation between weight and height be more than, less than, or about the same as 0.48?
   In a scatterplot the children will be to the left of, and below, the adults. This will mean that the overall picture is more stretched out and the correlation is higher than 0.48.

   Each part is worth 1 marks: 0.5 mark for the answer and 0.5 mark for a well written one sentence explanation or a good picture showing the point. (3 marks)

3. In class I tossed several thumbtacks on overhead projectors to demonstrate to you that thumbtacks can land either tipped over or point up. The demonstration is a success unless all the thumbtacks land the same way – that is, all of them land point up or all of them land tipped over. Suppose that the chance that the demonstration is a success is 0.9. Now that mandatory retirement has been abolished I may be blessed with the opportunity to do the demonstration another 400 times. Approximately what is the chance that I will have more than 380 successful demonstrations in those 400 trials? (5 marks)

This question asks for the probability of more than 380 successes in $n = 400$ *binomial trials with* $p = 0.9$. [1 mark for recognizing binomial trials. If they get all the rest of the problem right without using the word binomial you can give them this mark.] *The mean number of successes is* [1 mark]

$$\mu = np = 400 \times 0.9 = 360.$$

The standard deviation of the number of successes is [1 mark]

$$\sigma = \sqrt{np(1-p)} = \sqrt{400 \times 0.9 \times 0.1} = 6.$$

Now you use the mean and standard deviation to convert 380 or 380.5 (I will even take 379.5) to standard units

$$z = \frac{380.5 - 360}{6} = 3.42$$

or

$$z = \frac{380 - 360}{6} = 3.33$$

or even

$$z = \frac{379.5 - 360}{6} = 3.25.$$

[1 mark]

Finally you need the area to the right of this value under a normal curve, which is one of:

z	Lower Tail	Upper Tail
3.42	0.9997	0.0003
3.33	0.9996	0.0004
3.25	0.9994	0.0006

The desired chance is approximately 0.0003.[1 mark]