STAT 201

Assignment 3

Final version posted 12 October 2004; unchanged from preliminary version

1. From the text: # 8.14 on page 212.

   A: This is a blocked (by ultramarathoner or not) randomized controlled design.

   The diagram looks like that on page 211 in Figures 8.4. Instead of the 'Women' / 'Men' split you have the 'ultramarathoner' / 'not ultramarathoner' split. There are only two groups after 'random assignment' namely Vitamin C or Placebo.

2. From the text: # 8.20 on page 215.

   A: You will have the same sort of diagram as in Figure 8.11 again but this time the three groups will correspond to the three levels: 'Wine', 'Beer', 'Spirits'. There will be 3 groups of 60 men and 3 groups of 40 women. In each sex one of the 3 groups will be wine and so on.

3. From the text: # 8.25 on page 216.

   A: See the answer in the back of the book.

4. From the text: # 8.31 on page 217. NOTE: you do not need to do the randomization mentioned in either part.

   A: See the answer in the back of the book.

5. From the text: # 9.7 on page 228.

   A: See the answer in the back of the book.

6. From the text: # 9.9 on page 230.

   A: Here is my list with H for hit and M for miss.

HHHH	HHHM	HHMH	HHMM	HMHH	HMHM	HMMH	HMMM
MHHH	MHHM	MHMH	MHMM	MMHH	MMHM	MMMH	MMMM

   For b) it is natural to use $\{0, 1, 2, 3, 4\}$. (Often we would use the answer from (a) but keep in mind that you didn't actually observe the detailed outcome as in (a).

7. From the text: # 9.15 on page 234.

   A: For (a) the outcomes are $\{7, 8, 9\}$. See example 9.7 for S. To find the probability of this set of outcomes (this 'event') you add up the probabilities of the 3 possibilities: 0.058+0.051+0.046 = 155.

   For (b) the event is $\{1, 3, 5, 7, 9\}$ and you add up 5 numbers to get $P(B)$ from the table on page 231. You get 0.609.

   For (c) you put these two events together to get $\{1, 3, 5, 7, 8, 9\}$. The probability is the sum of the 6 corresponding probabilities. You get 0.660.

8. From the text: # 9.33 on page 243.

   A: For (a) the answer in the back of the book is fine. For (b) you need to explain that all 10 of the possibilities in (a) has chance 1/10 because there are 10 of them and they are equally likely. Of the outcomes listed in S a total of 4 have Mei-Ling in them. So the chance is 4/10 in (c). For (d) you must count how many of the 10 outcomes have neither an R nor an S in them. There are 3 such: $\{A, D\}, \{A, M\}, \{D, M\}$ so the chance is 3/10.

9. From the text: # 11.2 on page 284. NOTE: in doing this problem you are to assume that both parents carry the albinism gene but that neither is albino, that is, neither carries two copies of the gene.

   A: In order for the first child to be albino two events must both happen: the father passes on the gene and the mother passes on the gene. The chance that the father passes on the gene is 0.5; the chance that the mother passes on the gene is 0.5. These two events are independent since they depend on physically unconnected processes. So the chance they both happen is $0.5\times 0.5 = 0.25$.

   The chance that both of the first two are albino is similarly $0.25\times 0.25 = 0.0625$. The chance that the first is not and the second is not is also found by an application of the product rule. The chance the first is not albino is 0.75 (=1-0.25). The chance the second is not albino is 0.75. These two events are again independent so the chance that they both happen, that is, that neither child is albino, is $0.75\times 0.75 = 0.5625$.

10. From the text # 12.1, page 306.

    A: To answer any question of this type you need to think through a check-list: is a basic experiment repeated a fixed number of times? can the results of one trial influence the results of another? are we counting up the number of times something either happens or doesn't in that fixed number of trials? is the chance of the thing happening the same each time we repeat the experiment?

    For 12.1 the answers to all these are yes. There are 40 trials and $p = 0.40$. Unless the factory carefully makes sure you get the same number of reds in each package each time then it is likely that the packages are filled by picking 40 M&Ms from a very large group of M&Ms at random. This will make the results of the different picks independent.

11. From the text # 12.2, page 306.

    A: This time the number of trials is not fixed in advance so X is **not** Binomial.

12. From the text # 12.3, page 306.

    A: You may either say p changes (because the computer tries to make the probability of a correct answer increase by teaching) or that the trials are not independent: you get extra teaching after a wrong answer increasing the likelihood of a correct answer next time.

13. From the text # 12.11, page 314.

    A We are dealing with a Binomial situation with $n = 90$ trials and $p = 0.477$. If X is the number of 1s or 2s then we want $P(X \leq 29)$. I want you to convert 29.5 to standard units and look up the area (but I will accept, grudgingly, use of 29 instead of 29.5). The mean of the distribution of X is $\mu = np = 90 \times 0.477 = 42.93$. The standard deviation is $\sigma = \sqrt{np(1-p)} = 4.74$. Convert 29.5 to Standard Units and get $(29.5 - 42.93)/4.74 = -2.83$. The area to the left of -2.83 is 0.0023. Using 29 gives $z = -2.94$ and an area of 0.0016. The correct answer using the Binomial formula is 0.0021. My approximation is better but not so much better as to make a hugely important difference.

14. From the text # 12.19, page 317. NOTE: you do not need to draw the probability histogram.

    A: The answer in the back of the book is satisfactory.

15. From the text # 12.23, page 318.

    A The answers to a, b and d in the book are ok. For (c) I would prefer that you convert 70.5 to standard units using the mean $\mu = 75$ and the standard deviation $\sigma = \sqrt{np(1-p)} = 6.12$. You get $z = (70.5 - 75)/6.12 = -0.74$ for which the area to the left is 0.2297. The book's anser used 70 to convert to standard units.

16. From the text # 12.28, page 319.

    A: The number of heads has a Binomial distribution with $n = 10000$, $p = 0.5$, $\mu = np = 5000$ and $\sigma = \sqrt{np(1-p)} = 50$. Convert 5066.5 to standard units to get 1.33 for which the area to the left is 0.9082. The chance you want is the area to the right which is 1-0.9082=0.0918. If you do it without the continuity correction you convert 5067 to get $z = 1.34$ and a chance of 0.901.

17. From the text # 10.7, page 258.

    A: There are $n = 3$ measurements so the standard deviation (or standard error) of $\bar{x}$ is $\sigma/\sqrt{3} = 5/\sqrt{3}$. The book's answer has more digits than I would have reported personally.

    To get $\sigma/\sqrt{n} = 10/\sqrt{n} = 5$ you need $\sqrt{n} = 2$ or $n = 4$. Smaller standard deviations correspond to smaller probable deviations from the mean and small deviations are better.

18. From the text # 10.9, page 258.

    A: For a) the values 300 and 335 are converted to Standard units to give (300-300)/35 =0 and (335-300)/35=1. From the tables the areas to the left of 0 and 1 are 0.5000 and 0.8413. The areas to the right are then 0.500 and 0.1587. Some students may be tempted to use continuity corrections. Those are not really called for here since we are using continuity corrections only for the Binomial distribution and this is not a Binomial problem.

For (b) the average of a sample of 4 has mean 300 and standard deviation $\sigma/\sqrt{4} = 17.5$. Then convert 300 and 335 to Standard Units to get 0 and 2. The areas to the right of 0 and 2 are 0.5000 and 0.0228. the text has 0.0250 which is wrong.

If you tried to use a continuity correction in (a) the limits would have been 299.5 and 334.5. In (b) because you are averaging 4 numbers you would have used 299.875 and 334.875. (You go half way between two adjacent possible values of the sample mean. I am assuming that the possible scores are all integers. If you average 4 integers you get numbers that are integer numbers of quarters. The closest possible average to 300 less than 300 would then be 299.75. Halfway between 299.75 and 300 is 299.875.)

I hope this suggests that using the continuity correction here is not worth while!

19. From the text # 10.12, page 262.

    A: We are dealing with the average of a sample of $n = 200$ square yards. The sample average has a mean $\mu = 1.6$ and a standard deviation of $\sigma/\sqrt{200} = 1.2/14.14 = 0.0849$. Convert 2 to standard units to get $z = (2-1.6)/0.0849 = 4.71$. This number is beyond the end of the table. The area to the left of 3.49 is 0.9998 and so the area to the left of 4.71 must be more than this. The area to the right of 4.71 must be much less than 0.0002. You can't do better than this from the tables. In fact the area is 1.2×10^{-6} but I do not expect you to find this number.

20. From the text # 10.16, page 270.

    A: Each of these numbers is a statistic – a numerical summary of sample data.

21. From the text # 10.30, page 274.

    A: In order for the total of 12 eggs' weights to be in the range 750 to 825 the average weight of the 12 eggs must be between $750/12 = 62.5$ and 825/12=68.75 grams. We convert these figures to standard units using the mean $\mu = 65$ and the standard deviation for the average of a sample of 12 which is $5/\sqrt{12} = 1.44$. We get the range

    $$\frac{62.5-65}{1.44} = -1.74 \text{ to } \frac{68.75-65}{1.44} = 2.60$$

    The area is then 0.9953-0.0409 = 0.9544.