March 23, 1999

**Math 102 / Core 143, Sections B and C — Exam II**

Show all work clearly for partial credit — an unevaluated expression is worth more than the numerical answer.

1. *(25 points)* A psychologist studies people's ability to memorize a list of words. He varies the length of time allowed for study and records the number of words memorized. Here is the summary data:

| | $AV$ | $SD$ | |
|---|---|---|---|
| minutes allowed | 8 | 2 | $r = 0.6$ |
| words memorized | 20 | 6 | |

   (a) Not knowing how long a randomly chosen subject was allowed to study, how many words should we guess she learned?

   (b) How far should we expect our guess in (a) to be off?

   (c) If we are told that the subject was given 11 minutes to study, how many words should we guess she learned?

   (d) How far should we expect our guess in (c) to be off?

2. *(15 points)* A study is made of people who take primary responsibility for the homemaking chores in their home. It is found that 75% are women, that 30% have completed a bachelor's degree, and that only 2% have completed a doctorate. If it is possible to answer the following questions on the basis of this data, do so; if not, explain why not.

   (a) What is the probability that a person randomly chosen from this list will be a woman with a bachelor's degree?

   (b) What is the probability that a person randomly chosen from this list will have either a bachelor's degree or a doctorate?

   (c) It is also found that 40% of the people in the list earn at least $30,000 annually. Assuming (though it probably is not true) that sex and earnings are independent, what is the probability that a person randomly chosen from this list is a woman earning at least $30,000 annually?

3. *(15 points)* A die is rolled 12 times.

   (a) How many times would you expect to get a 4?

   (b) What is the probability of getting exactly three 4's in the 12 rolls?

   (c) If no 4's appear in the 12 rolls, should you conclude that the die is unfair? (Your answer should include both the computation needed to decide the probability of that outcome and a short, subjective discussion of your conclusion from the evidence.)

4. *(10 points)* Short answers:

   (a) When probability experts were called in to decide an election in Pennsylvania, the counting of what kind of votes was in dispute?

   (b) Craps, since it is decided rolling dice, would seem to be a game of pure chance; but Linsalata, in his article on riverboat gambling in Missouri, says otherwise. Why is craps not a game of pure chance?

5. *(25 points)* In "Repealing the Law of Averages," Stewart asserts that, if a coin is flipped more and more often, the difference between the number of heads and the number of tails will grow larger rather than smaller.

   (a) We can model the computation of this difference as a ________ (sum? average?) of many draws (with replacement) from a box with tickets labelled ________ .

   (b) The <u>expected</u> value of the difference (number of heads minus number of tails) is ________ , no matter how many times we flip the coin.

   (c) But if we actually flip a coin 400 times, we should expect that our actual difference should be about how much off from the expected value in (b)? And the same question for 10,000 flips. (DON'T FORGET THE SECOND PART.)

   (d) In view of (c), is our expected error (our difference from the expected difference) growing as a <u>linear</u> function of the number of flips? Explain briefly.

   (e) Describe how you might use a spreadsheet program like Excel to test Stewart's assertion.

6. *(10 points)* Explain how the discussion in Chapter 17 of the text, on sums of repeated draws from a probability distribution, relates to the article by Miriam Lipschütz-Yevick.

---

$$y - AV_y = (\text{sign of } r)\frac{SD_y}{SD_x}(x - AV_x)$$

$$y - AV_y = (r)\frac{SD_y}{SD_x}(x - AV_x)$$

$$\text{average of products}(x_i \text{ in std units})(y_i \text{ in std units})$$

$$\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

$$EV \text{ of sum} = n \cdot (AV \text{ of box}) \qquad SE \text{ of sum} = \sqrt{n} \cdot (SD \text{ of box})$$

$$SD = (\text{larger} - \text{smaller})\sqrt{(\text{fraction with larger})(\text{fraction with smaller})}$$

$$SD_y\sqrt{1-r^2}$$

**Math 102 / Core 143, Sections B and C — Solutions to Exam II**

1. (a) 20 words (the average)
   (b) 6 words (the $SD$)
   (c) Using linear regression, we see that she was given $(11-8)/2 = 1.5$ $SD$'s above the average amount of time to study, so we should expect that she learned $(.6)(1.5) = .9$ $SD$'s above the average amount of words, or $20 + 6(.9) = 25.4 \approx 26$ words.
   (d) $6\sqrt{1-(.6)^2} = 4.8$, the RMS error for regression.

2. (a) We cannot answer, because we do not know whether sex and degree are independent.
   (b) 30% — not 32%, because those who have doctorates must also have bachelor's degrees.
   (c) $(.75)(.4) = .3 = 30\%$.

3. (a) Two — about once every 6 rolls.
   (b) $C(12,3)(\frac{1}{6})^3(\frac{5}{6})^9 \approx 20\%$.
   (c) $(\frac{5}{6})^{12} \approx 11\%$, which is probably not enough to conclude that the die is unfair — that result would happen more than one time out of 10 just by chance.

4. (a) Absentee ballots.
   (b) Because there are so many different ways to bet, which affects chances of winnning.

5. (a) As the sum of many draws from a box with two tickets labelled 1 and $-1$.
   (b) 0, the average of the box.
   (c) The $SD$ of the box is $(1-(-1))\sqrt{(.5)(.5)} = 1$. So for 400 flips, the standard error of the sum is $\sqrt{400}(1) = 20$, while for 10,000 flips, the $SE$ is $\sqrt{10,000} = 100$.
   (d) No, it grows as the square root of the number of flips.
   (e) Make a column of, say, 300 random 1's and $-1$'s (one possible command would be $= if(rand() < .5, -1, 1)$, copied down column A) and next to it make a column of the sum up to that point (maybe $= a1$ in cell B1 and then $= b1 + a2$ in cell B2, copied down column B). Make a column graph of column B to see if it moves toward or away from 0. It should have a general tendency away from 0, though not consistently. (I would have suggested more than 300 points, but there is a limit — about 400, I believe — on the number of data points that Excel will put on a graph.)

6. Lipschütz-Yevick says that a distribution has a normal curve if it is the sum of a sample from many distributions (perhaps the same distribution many times), each of which contribute only a small fraction of the value of the total. Chapter 17 illustrates this by showing that, no matter what distribution one starts with (and graphs of several appear in the text and were displayed in class), if one takes increasingly larger samples, the distribution of the sum of the sample will more closely approximate the normal curve.