December 12, 2006

Final Exam — Math 102 / Core 143 CX

Points are in parentheses. Show your work to receive partial credit; an answer like $[1 - (6/7)^3]/4$ is worth more than 0.09257, because it displays your reasoning more clearly.

1. *(20 points)* A "Magic 8-Ball" is a toy that is almost spherical but with one flat window into its interior. A dodecahedron (12-sided solid) floats in the black fluid inside the 8-ball; when the 8-ball is held with the window up, one side of the dodecahedron floats up to the window, and the user can read what is printed on that side. Four sides of the dodecahedron say "Yes", four say "No", three say "Maybe", and one says "Ask again later". The manufacturer tests the the 8-ball by shaking it 240 times and noting results. She gets 85 "Yes", 70 "No", 63 "Maybe" and 22 "Ask again later". Is the 8-ball fair?

2. *(20 points).* From a new batch at the Utica Club Brewery, 16 bottles of Erie Canal Hard Water are sampled for levels of duclamine (an ingredient necessary for that authentic rock-strewn flavor). The specifications for ECHW say it should contain 3 mg of duclamine per ounce; the 16 bottles averaged 2.8 mg per ounce, with a standard deviation of 0.7 mg. Decide whether the sample's duclamine level means (with 95% certainty) that the whole batch's duclamine level is too low.

3. *(16 points)* Recall that a "straight" deck of cards (i.e., one used for bridge or poker) has 52 cards, in 13 ranks and 4 suits.

   (a) If two cards are chosen at random <u>without</u> replacement, what is the probability that both of them are clubs?

   (b) If a card is chosen at random, what is the probability that it is either a king or a club?

   (c) What is the probability of getting a king at least once if a card is drawn 8 times <u>with</u> replacement?

   (d) What is the probability of getting a <u>club</u> exactly 5 times if a card is drawn 8 times <u>with</u> replacement?

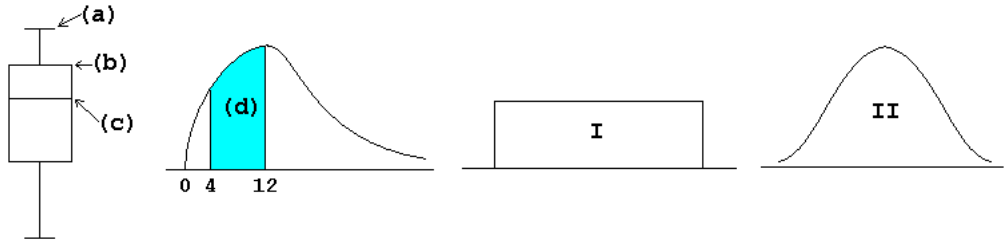
4. *(15 points)* The employment rate (fraction of the work force that is employed) averages 87% in New York villages, with a standard deviation of 10%, and the welfare rate (fraction of the population on welfare) averages 8% with a standard deviation of 6%. The correlation between the rates is -0.6.

   (a) What should we guess is the welfare rate in a village with an employment rate of 92%?

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

   (c) If we found that the welfare rate in that village was really 7%, what is the corresponding residual?

5. *(10 points)* For (a), (b) and (c), what do the indicated parts of the box-and-whisker plot indicate? For (d), what does the indicated region in the histogram tell you? For (e), which histogram is the best approximation of the numbers that come up if we roll a fair die 6000 times?



6. *(20 points)* A survey is taken of 1600 households in a large city; it is found that 900 of them heat with electricity, and that the average monthly electric bill (for all 1600) is $250, with a standard deviation of $120.

   (a) Assuming that the households surveyed were simple random sample of the city, what is a 95% confidence interval for the average monthly electric bill in the city?

   (b) How many households would the survey have had to include to get a 95% confidence interval only half as wide (assuming similar average and standard deviation)?

   (c) The power company claims that 60% of the city's households use electric heat. Is the survey's result significant evidence that the company's estimate is too high?

   (d) Suppose the survey sample was conducted by e-mail, using the client list of an internet service provider. How might that have biased the results?

7. *(12 points)* [The point of this exercise is to explore the mantra "Correlation is not causation" — what causes what?] In a small community, an unusually large number of people seem to have colds, and many of the sufferers are found to frequent the community's only restaurant. Answer each of the following in a sentence or two:

   (a) How might going to the restaurant have caused its customers to have colds?

   (b) How might the fact that they have colds caused people to eat at the restaurant?

   (c) What other factor may have caused people both to get colds and to eat at the restaurant?

8. *(12 points)* Relative to the article, "Strong messages get girls to wait on motherhood; Several factors pushing down pregnancy rate", by Wendy Koch: What are some of the "several factors" mentioned in the article? Which can be regarded as numerical data, rather than categorical data?

Solutions to Final Exam

1. We expect counts of 80,80,60,20 of a fair 8-ball, and we got 85,70,63,22; so we can do a χ^2-test to decide whether there is a significant difference between the two.

$$\begin{aligned}\chi^2 &= \frac{(85-80)^2}{80}+\frac{(70-80)^2}{80}+\frac{(63-60)^2}{60}+\frac{(22-20)^2}{20}\\ &= \frac{25}{80}+\frac{100}{80}+\frac{9}{60}+\frac{4}{20}=\frac{25+100+12+16}{80}\approx 1.91\end{aligned}$$

and with $4-1=3$ degrees of freedom, we see from the table that a χ^2-value that high occurs by chance between 70% and 50% of the time, not less than 5%; so we do not reject the null hypothesis: The 8-ball is fair.

2. With the small sample, we conduct a t-test, with the null hypothesis that the batch's duclamine level is the required 3 mg per ounce: The SE is $.7(\sqrt{\frac{16}{15}})/\sqrt{16}\approx .18$, so $t=(2.8-3)/.18$, which is slightly less than -1. The degrees of freedom is $16-1=15$. We don't have a t-table, but it would probably show that the probability of a t-value that low or lower is not less than 5%, so we would not reject the null hypothesis and would accept the batch.

3. (a) $(13/52)(12/51)=1/17\approx 6\%$.
   (b) $(4/52)+(13/52)-(1/52)=16/52=4/13\approx 31\%$.
   (c) $1-(12/13)^8\approx 47\%$
   (d) $(8!/(5!3!))(1/4)^5(3/4)^3\approx 2\%$

4. (a) The employment rate in standard units is $(92\%-87\%)/10\%=.5$, so we should expect the welfare rate in standard units to be $(-.6)(.5)=-.3$, or in percent, $8\%+(-.3)(6\%)=6.2\%$.
   (b) By the RMS error for regression: $\sqrt{1-(-.6)^2}(6\%)=4.8\%$.
   (c) $7\%-6.2\%=0.8\%$.

5. (a) Maximum
   (b) 75th percentile
   (c) Median
   (d) The fraction of the data between 4 and 12 (to my eye, about 35%)
   (e) I (II would be for the expected sum or average of many rolls)

6. (a) $\$250\pm 2(\$120/\sqrt{1600})=\$250\pm\$6$.
   (b) Four times as many households, or 6400.
   (c) Taking the power company's estimate as the null hypothesis, we perform a z-test: Because the SE for percent is (using the figure from the null hypothesis)

$$(1-0)\sqrt{(.6)(.4)}/\sqrt{1600}\approx 1.2\%\ ,$$

   we have

$$P(\%\le 900/1600\approx 56\%)=P(z\le(56\%-60\%)/1.2\%\approx -3.1)<5\%\ ;$$

   so we reject H_0: the power company's estimate is too high.

(d) One possible answer: People with e-mail addresses are more likely to be technologically inclined, and hence to prefer electric heat to, say, oil heat.

7. (a) Someone in the restaurant, perhaps even an employee, might have spread the cold to those around him/her.

   (b) Cold sufferers might not feel up to cooking for themselves and decide to go out to eat.

   (c) One possibility: If it's a college town, students who are studying for finals late into the night are more susceptible to colds and less likely to feel like cooking for themselves.

8. Some factors mentioned were reduced benefits because of welfare reform (which probably counts as numerical data) and greater availability of abstinence- and sex-education programs. This availability is difficult to quantify, so the availability of such a program to a given teenager is probably a yes-no, categorical question. Also, fewer teens are having sex — or at least saying they don't have sex — and more who do have sex using — or claiming to use — condoms because of fear of AIDS and other STDs. Finally, there is less stigma attached to virginity among high school students. The last three may be regarded as effects of the programs mentioned above, as well as causes of lower teen pregnancy rates; but in any case the first two can be quantified, numerical data; while the level of stigma attached to virginity would be hard to quantify, so it might be categorical data (or better, ordinal data: ordered but not quantified).