MONT 107N – Understanding Randomness
Solutions for Sample Final Exam Questions
April 30, 2010

1. Briefly explain the following chance and statistical concepts.

   (a) Central Limit Theorem.

   Solution: The Central Limit Theorem says that the probability histogram for sums of independent draws with replacement from any box model will follow a shifted, scaled normal curve more and more closely as the number of draws increases (mathematically, as $n \to \infty$). Recall from one of our labs that there are cases where not every value of the sum can be achieved. In one of these cases, the way to interpret the statement is that the *area* of the boxes in the probability histogram in any range tends to the corresponding *area* under the shifted, scaled normal curve. See the statement on page 325 of the text. The way we said this in class was slightly different because you can do the shifting and scaling either with the histogram or with the normal curve.

   (b) Law of Averages.

   Solution: The most basic form is the statement that in tossing a fair coin, the number of heads will be one half of the number of tosses, plus some chance error. The chance error is likely to be large in absolute terms, but small in relation to the number of tosses. We later quantified this statement with our discussion of EV's and SE's. For the coin tossing chance process, we are talking about n draws with replacement from the box $[0, 1]$, so $EV = n \times \text{average of box} = n/2$ and $SE = \sqrt{n} \times \text{SD of box} = \sqrt{n} \times \sqrt{(1/2)(1/2)}$ (this is the SE for the number of heads).

   (c) 95% confidence interval.

   Solution: The 95% confidence interval is the interval observed $\pm\, 2SE$. Here "observed" might mean a sample percent or a sample average, etc. The idea is that if we repeat the sampling process many times, we would expect about 95% of the intervals constructed this way to contain the corresponding population percent or population average, etc.

   (d) χ^2 statistic.

   Solution: This is computed from a set of observed and expected values by the formula

   $$\chi^2 = \text{sum of } \frac{(\text{observed} - \text{expected})^2}{\text{expected}}.$$

   The χ^2 statistic is used in tests of "goodness of fit" of a collection of observed data with a given box model, and also for tests of independence of characteristics in tabulated data.

   (e) Null hypothesis of a test of significance.

   Solution: The null hypothesis of a test of significance is a "negative result" that says a claimed pattern or difference is explainable merely by chance variation in the sampling process.

(f) p-value of a test of significance.

Solution: The p-value is the chance that the observed test statistic, or a "more extreme" value, would be observed from a sample *if the null hypothesis is true*. The smaller the p-value is, the stronger the evidence is for rejecting the null hypothesis.

2. Short Answer:

(a) (Hypothetical) A college has an elective quantitative reasoning course for first-year students. Each year approximately one fifth of the first-year students elect to take this course. The college does a study of the grades of first-year students. The study shows that after the first- year of college, the students who elect to take this math course have an average GPA for their other three courses that is .1 units higher than the average GPA for the other students for all their courses. Based on this data, the Mathematics department argues that this course raises student GPAs and that every student should be required to take it. Does the Mathematics Department have a good argument or is it possibly flawed? If it is a good argument, explain why, if not, how would you correct the study?

Solution: This experimental design is flawed because the students who took the quantitative reasoning course were self-selected (they "elect" to take the course). There could be some characteristic that they share (for instance a higher than average interest in mathematics or its applications) that led them to elect this course and that also is associated with better performance in the other courses they take. To minimize the influence of possible *confounding factors* like this, a study of the effects of a course like this should randomly assign students to the course ("treatment group") or not ("control group"). Of course this would mean the students involved were not choosing their own courses, so this kind of study would be hard to do at a college!

(b) Why are we justified in using a normal curve to make estimates about draws from box models regardless of whether or not the data in the box is normally distributed?

Solution: The reason is that we are relying on the result called the Central Limit Theorem (see above). As long as there are enough draws, the probability histogram for the sum will follow a normal curve closely enough to make estimates.

(c) In major national polls, which guide politicians and political candidates in their decision making and play a prominent role in the national media, the sample size is usually 1000 or so. Why are samples of this size used? In particular, what are the benefits of using a sample size of 1000 as opposed to smaller sample sizes or larger sample sizes?

Solution: A sample size of $n = 1000$ or so is designed to produce a standard error (SE) for the percent that is small enough to make the projections usually accurate to within a few percentage points. In numerical terms, the SE for the percent with $n = 1000$ is always bounded above by

$$\sqrt{\frac{(.5)(.5)}{1000}} \times 100\% \doteq 1.6\%$$

Then $2SE$ is about 3%, so the results would be reported as the observed percent $\pm 3\%$ or so. (Note that this is a 95% confidence interval for the percent.)

(d) If you want to get polling results with comparable accuracy, do you need a bigger sample size if the polling is done in California than you do if the polling is done in Rhode Island? Explain.

Solution: The answer is NO. The previous part shows why this is true – the SE for the percent depends only on the sample size, not on the size of the whole population. (Now of course, this *is assuming* that the size of the population is enough larger than the size of the sample that the "correction factor" for sampling without replacement can be ignored!)

(e) True or False and explain: If the p-value of a test of significance works out to $p = .05$ then we can say that the null hypothesis was false.

Solution: False - this p-value is often taken as meaning that the evidence indicates that we should reject the null hypothesis, but we cannot definitely say it was false. There is always the chance that the sample used to compute the test statistic was unrepresentative.

(f) True or False and explain: Increasing the sample size will always decrease the width of a 95% confidence interval for an average.

Solution: False – in computing the confidence interval, we have to use the SD of the sample to estimate the SD of the box corresponding to the null hypothesis in the formula for the SE. The formula is

$$SE = \frac{\text{SD of sample}}{\sqrt{n}}.$$

Even if n increases, the SD of the sample could also increase enough to produce a larger estimated SE.

3. A gambler is playing roulette. On each play, he bets \$1 on four adjacent numbers, which pays 8 to 1. That is, if one of his 4 numbers comes up, he will win \$8 on a \$1 bet. Otherwise, he loses \$1.

(a) Construct a box model for this bet. (Hint: Recall there are a total of 38 slots on a roulette wheel.)

Solution: The box model will be $[8, 8, 8, 8, -1, \ldots, -1]$ with 34 -1's. The 8's are the payoffs if one of the four selected numbers comes up.

(b) Compute the average and SD of the box.

Solution: The average is $-2/38 \doteq -0.053$, and the SD can be computed by the shortcut rule since there are only two different types of tickets:

$$SD = (8 - (-1))\sqrt{(4/38)(34/38)} \doteq 2.76.$$

(c) If the bet is played 50 times, compute the EV and the SE for the bet.

Solution: The EV is $50 \times -0.053 = -2.65$ and the SE is $\sqrt{50} \times 2.76 \doteq 19.5$.

(d) What are the gambler's chances of winning \$40 or more on 50 spins of the wheel?

Solution: To compute this chance, we convert to standard units and consult the normal curve area table:

$$z = \frac{40 - (-2.65)}{19.5} \doteq 2.2.$$

So the chance that the gambler will win at least 40 dollars is the same as the area in the upper tail of the normal curve starting at $z = 2.2$. This is

$$\frac{1}{2}(100\% - A(2.2)) = \frac{1}{2}(100\% - 97.22\%) = 1.39\%$$

(not very good odds!)

4. A CBS News poll released on November 24, 2008 surveyed a random sample of 706 adults nationwide. It reported that 61% of Americans say it is too soon for movies to be made about the current Iraq war.

(a) What box model would you use to represent this poll?

Solution: This could be a 0/1 box with one ticket for each person in the population – (0 if the person does not feel this way, 1 if the person does).

(b) Use your model to determine a 95% confidence interval for the percent of Americans who actually feel this way.

Solution: Recall that this is the case where we need to use the "bootstrap" method to estimate the SE for the percent – we use the usual formula but with the $p = .61$:

$$\text{estimated SE for percent} \doteq \sqrt{\frac{(.61)(.39)}{706}} \times 100\% \doteq 1.84\%.$$

Then the confidence interval is $61\% \pm 2 \times (1.84\%) \doteq 61 \pm 3.68\%$.

(c) True or False and explain: There is a 95% chance that the actual percent of Americans who feel this way is in the interval you computed in part b.

Solution: False – the chance is in the sampling process not in the location of the population percent. Remember that the idea is: if we did this sampling process repeatedly with $n = 706$, then about 95% of the computed intervals would contain the population percentage.

5. Before a person gives blood, the Red Cross requires that the hemoglobin in their blood measures 12.0 or higher on an electronic scan. There is typically some error in the measurements. (The following is hypothetical.) Suppose that such a device is being calibrated by ve readings on a standardized sample known to have a hemoglobin content of 12.0. The ve readings are 11.5, 11.9, 12.0, 12.1, 12.1 for an average of 11.92. Based on these readings, we want to do a test of significance to determine whether the device is calibrated correctly or not.

(a) What box model would you use to model this calibration process?

Solution: The box should have a large number of tickets with an average of 12.0 and a *normal* probability histogram for the values.

(b) Formulate a null and alternative hypothesis for the difference between 12.0 and 11.92.

Solution: The null hypothesis is that the average reading is 12.0 (and the lower sample average 11.92 is just due to chance). The alternative could be that the average reading is less than 12.0.

(c) What significance test should you use? Why?

Solution: Since we only have $n = 5$ samples, we are in the *small sample case* and we need to use a t-test.

(d) Carry out your test of significance. What do you get for a a test statistic and observed significance level?

Solution: For the SE, we need to have the SD of the sample values. Computing as usual,

$$SD = \sqrt{\frac{1}{5}\left((11.5 - 11.92)^2 + (11.9 - 11.92)^2 + (12.0 - 11.92)^2 + 2 \times (12.1 - 11.92)^2\right)} \doteq .22$$

In the small sample case, recall that the SE actually uses the thing we called SD^+:

$$SD^+ = \sqrt{5/4} \times SD \doteq .25,$$

so the SE for the average is

$$SE \doteq \frac{.25}{\sqrt{5}}.$$

Then the test statistic is

$$t = \frac{11.92 - 12.0}{\frac{.25}{\sqrt{5}}} = -0.72.$$

From the t-table with $n - 1 = 4$ degrees of freedom, we see that this gives $p > .25$. (How to see this: The t-curves are symmetric like the normal curve. So the area in the lower tail here is the same as the area in the upper tail starting at .72. But the smallest entry on the row for 4 d.f. is .75 for 25%. So the area in the upper tail starting at .72 must be slightly larger than 25%.)

(e) Is the device calibrated correctly, or is it biased?

Solution: There is no reason to conclude the device is biased. The deviation from 12.0 to 11.92 could easily be caused by chance variation.

6. (This problem is adapted from "Heart rate in yoga asana practice: A comparison of styles," *Journal of Bodywork and Movement Therapies* Volume 11, Issue 1, January 2007, Pages 91-95.) Yoga is often recommended for stress relief, yet some of the more fitness-oriented styles of yoga can be vigorous forms of exercise. The purpose of this study was to investigate differences in heart rate during the physical practice of yoga postures, breathing exercises,

and relaxation. The study led a group of participants through three different styles of yoga: astanga yoga, hatha yoga, and "gentle" yoga. Participants wore heart rate monitors during the sessions and their heart rates were monitored repeatedly throughout the sessions. Assume there were three independent groups of 50 participants each in the study. The average and standard deviation for the heart rates (in beats per minute) of the participants for each style of yoga are given below:

Yoga style	Average heart rate (bpm)	SD
astanga	95	12.84
hatha	80	9.32
"gentle"	74	7.41

The researchers then applied two-sample z-tests to each pair of styles and concluded that there may be different fitness benefits for different styles of yoga practice.

(a) Apply a two-sample z-test to the astanga and hatha yoga data. What is your p-value?

Solution: The null hypothesis would be that there is no difference between the average heart rates, and the alternative is that astanga yoga produces higher average heart rates. The test statistic is

$$z = \frac{95 - 80}{\sqrt{\frac{(12.84)^2}{50} + \frac{(9.32)^2}{50}}} \doteq 6.68.$$

The p value is virtually zero. So there is strong evidence to reject the null hypothesis

(b) Apply a two-sample z-test to the hatha and gentle yoga data. What is this p-value?

Solution: Similarly,

$$z = \frac{80 - 74}{\sqrt{\frac{(9.32)^2}{50} + \frac{(7.41)^2}{50}}} \doteq 3.56.$$

Again the p-value is very small, about $p = .001$. So there is strong evidence to reject the null hypothesis that there is no difference between hatha yoga and gentle yoga.

(c) Were the researchers' conclusions justified? (Note – the higher the average heart rate attained, the higher the fitness benefit.)

Solution: Yes, the data indicate that astanga yoga produces higher heart rates, hence more health benefits, than hatha yoga, and similarly for hatha yoga versus gentle yoga.

7. The December 7, 2007 Gallup Poll repeated a number of questions about how people perceive the relative ability of the Democratic Party and the Republican Party to handle ten important issues. The poll was of adults 18 years old or older. On the issue of health care, the poll asked the following question: *Do you think the Republican Party or Democratic Party would do a better job of dealing with each of the following issues and problems? How about health care policy?* Here are the responses for more recent poll and the poll of January 2004:

	Republicans	Democrats	No difference/no opinion	Totals
2004 Jan. 9-11	357	578	115	1050
2007 Nov. 30-Dec. 2	291	592	120	1003
Totals	648	1170	235	2053

(Data was adapted from the Gallup web site.) The natural question to ask is whether or not this represented a real shift in public opinion.

(a) Formulate a null and alternative hypothesis for this question.

The null hypothesis would be that there was no difference between the two polls, or in other words, that that the proportion of people who thought each party would do a better job was *independent of the year the poll was taken.* The alternative would be that there was a definite change from 2004 to 2007 (with more people thinking the Democrats would do a better job in the later year).

(b) What significance test should you use to determine the answer?

Solution: This calls for a χ^2-test for independence.

(c) Carry out your test of signficance. What do you get for a test statistic and observed significance level?

Solution: Under the assumption that the opinions did not change, the expected values are:

	Republicans	Democrats	No difference/no opinion	Total
2004 Jan. 9-11	331.4	598.4	120.2	1050
2007 Nov. 30-Dec. 2	316.6	571.6	114.8	1003
Totals	648	1170	235	2053

Then the χ^2-statistic is computed as the sum of

$$\frac{(\text{observed} - \text{expected})^2}{\text{expected}}.$$

This comes out:

$$\chi^2 \doteq 1.98 + .70 + .22 + 2.07 + .73 + .24 \doteq 5.94.$$

Since the table (excluding the row and column totals) has 2 rows and 3 columns, the total number of degrees of freedom is $(2-1) \times (3-1) = 2$. From the χ^2 table we see that this gives a p-value of slightly larger than $p = .05$, since the 5% value is 5.99.

(d) Did this data represent a real shift in public opinion?

Solution: This is a matter of interpretation. The p-value does not reach the "magic" .05 significance level. However, it is very close. There is at least some evidence to believe there was a change.