MONT 101N – Analyzing Environmental Data
Solutions for Sample Final Exam Questions
April 27, 2018

1. Briefly explain the following chance and statistical concepts.

   (a) Central Limit Theorem.

   Solution: The Central Limit Theorem says that if independent and identically distributed values Y_i, $i = 1, \ldots, N$ are sampled from any distribution with finite mean μ and SD σ, then the distribution of the *sample means*

   $$\overline{Y} = \frac{1}{N}(Y_1 + \cdots + Y_N)$$

   becomes more and more normal as $N \to \infty$. Precisely, the distribution of

   $$\frac{\sqrt{N}(\overline{Y} - \mu)}{\sigma}$$

   approaches a standard normal distribution as $N \to \infty$.

   (b) 95% confidence interval.

   Solution: The 95% confidence interval is the interval *observed value* $\pm\, 2SE$. 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.

   (c) t-distribution

   Solution: When sampling from a normal population with mean μ, the t distribution describes the "standardized sample means" where the population σ is estimated by the sample S: $\dfrac{\overline{Y} - \mu}{S/\sqrt{N}}$. This gives the t-distribution with $N - 1$ degrees of freedom.

   (d) standard normal distribution

   Solution: The normal distribution with $\mu = 0$ and $\sigma = 1$.

   (e) Null hypothesis of a test.

   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 hypotheses.

   Solution: The p-value is the *smallest value of the Type I error probability α for which the null hypothesis would be rejected with the observed data.* In more intuitive terms, it 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), so they do not form a random sample. 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 sample means with sample sizes $N > 30$ even if the population the individual Y_i are sampled from is not 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 > 30 sampled values used to compute the means, the probability histogram for the sample means 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 approximately 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. (This *is assuming* that the size of the population is enough larger than the size of the sample that the difference between sampling with replacement and 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 S from the sample. The formula is

$$\overline{Y} \pm 1.96 \frac{S}{\sqrt{N}}.$$

Even if N increases, the S of the sample could also conceivably also increase enough to produce a larger estimated confidence interval.

3. Let Z represent a standard normal random variable and X represent a normal random variable with $\mu = 4.5$ and $\sigma = 1.3$.

(a) Find $P(1.24 < Z < 2.0)$

Solution: From the z-table, this is $.4772 - .3925 = .0847$.

(b) Find $P(-0.4 < Z < 1.33)$

Solution: From the z-table this is $.1554 + .4083 = .5637$.

(c) Find $P(X > 5)$

Solution: We "standardize":

$$X > 5 \Leftrightarrow \frac{X - 4.5}{1.3} > \frac{5 - 4.5}{1.3} \Leftrightarrow Z > .38$$

From the z-table this is $.5000 - .1480 = .3520$.

(d) Find $P(4 < X < 5.4)$

Solution: We "standardize" again

$$4 < X < 5.4 \Leftrightarrow \frac{4 - 4.5}{1.3} < Z < \frac{5.4 - 4.5}{1.3} \Leftrightarrow -.38 < Z < .69$$

From the z-table, this is $.3520 + .2549 = .6069$

(e) Suppose that values of X are sampled $n = 12$ times independently. What is the probability that 4 out of the 12 values satisfy $4 < X < 5.4$.

Solution: This is a binomial situation. The probability is

$$\binom{12}{4}(.6069)^4(.3931)^8$$

This is approximately .0383, but it's fine to leave these in the first form!

(f) Now suppose you have a (very) large urn containing 10000 balls, 6069 of them blue and 3031 of them red. Suppose you picked 12 of them at random *without replacement.* What is the probability that you get exactly 4 blue balls?

Solution: This would be given by the hypergeometric formula:

$$\frac{\binom{6069}{4}\binom{3031}{8}}{\binom{10000}{12}}$$

You could definitely leave this one in this form. This is approximately .0382 (very close to the answer to the previous part, which is equivalent to doing the selection *with replacement*).

4. A Gallup poll released on April 3, 2017 surveyed a random sample of 706 adults nationwide. It reported "that 59% of Americans say the environment should be prioritized over energy production."

(a) Determine a 95% confidence interval for the percent of Americans who actually feel this way. Use the "conservative" method of computing the standard error.

Solution: The "conservative method" for the standard error says estimate SE by

$$SE = \sqrt{\frac{pq}{706}} \leq \frac{1/2}{\sqrt{706}} \doteq .019$$

or 1.9%. Then the confidence interval is $59\% \pm 2 \times (1.9\%) \doteq 59 \pm 3.8\%$.

(b) 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) 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.

(b) 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.

(c) Carry out your test using $\alpha = .05$. What do you get for a test statistic and what is the conclusion?

Solution: We need to have the SD of the sample values. Computing as usual,

$$S = \sqrt{\frac{1}{4}\left((11.5 - 11.92)^2 + (11.9 - 11.92)^2 + (12.0 - 11.92)^2 + 2 \times (12.1 - 11.92)^2\right)} \doteq .25.$$

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 for a lower tail test we would want $t < -2.131847$ to reject H_0. However, that is not true, so 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.