# K300 (4392) Statistical Techniques (Fall 2007)

## Assignment 4: Statistical Inferences (150 points, Due October 15)

Instructor: Hun Myoung Park
kucc625@indiana.edu, (317) 274-0573

Please read the following instructions carefully. The pages indicated in each question provide good guidelines (sometimes answers) for solving the question. If you have any problem with any of the questions, please contact the instructor.

- **Do not use any wordprocessor; Use separate sheets instead.**
- Answer all questions. Do not skip any one.
- Try to show your work clearly. Do not simply copy answers on the textbook.
- Due to the midterm exam on October 17th, **you may not submit the assignment after the due date**. The answer key will be released at 2:00 P.M. on October 15th.
- You MAY NOT discuss with other classmates when answering questions.

**1. (12 points)** True/False questions. *See page 284-285*.

1) A normal probability distribution is a distribution for a continuous random variable.
2) A normal probability distribution is defined by its mean and variance.
3) The mean, 50th percentile, and mode of a normal probability distribution are equal in a normal distribution.
4) The standard normal distribution always has a mean 0 and a variance 1.
5) In a normal distribution of a random variable x, the probability that x is -1 billion is numerically zero, P(x=-1,000,000,000)=0.
6) All normal probability distributions can be transformed into the standard normal distribution.
7) If a random variable x is normally distributed, x~N(5, 9), the probability distribution is not skewed to the left or right.
8) Any random variable follows a normal probability distribution when N is large.
9) Even if N is sufficiently large, a binomial probability distribution with an extremely large or small p (e.g., p=.9999 or p=.0001) is not approximated to a normal probability distribution.
10) Without the standard normal distribution, we cannot get a probability of a random variable that follows a normal distribution with mean $\mu$ and variance $\sigma^2$.
11) In a normal probability distribution, the probability of mean μ is larger than those of any other points of $x$, $P(x=\mu)>P(x=x_i)$ where $i$ is any number except μ.
12) In the standard normal distribution, the probability that z is greater than -2.58 and smaller than 2.58 is .99, $P(-2.58 \leq z \leq 2.58)=.99$. Thus, $P(-100 \leq z \leq 100)$ is obviously larger than 1.

**2. (10 points)** Find the following probabilities using the standard normal distribution table on page 634. Then, draw the standard normal distribution and represent the area with the mean, a particular value $z_x$, and the probability indicated in the figure as shown in Figure 6-16 on page 292. Keep in mind a probability in the table is $P(0 \leq z \leq z_x)$. Since

a normal distribution is symmetric, $P(-z_x \leq z \leq 0) = P(0 \leq z \leq z_x)$. *See page 287-288 and question 26-45 on page 299*.

1) $P(0 \leq z \leq 1)$
2) $P(-2 \leq z \leq 0)$
3) $P(-1.96 \leq z \leq 1.96)$
4) $P(-1.645 \leq z \leq 2.58)$
5) $P(-\infty \leq z \leq 2.58)$
6) $P(-\infty \leq z \leq -1.96)$
7) $P(-\infty \leq z \leq -2.58 \text{ or } 2.58 \leq z \leq \infty)$
8) $P(-3 \leq z \leq -2.58 \text{ or } -1.96 \leq z \leq 1.645)$
9) $P(0 \leq z \leq 0)$ DO NOT use the standard normal probability function to compute this probability. Use the standard normal probability table instead. This process involves an integration of the function, which you do not need to know to solve this question (2 points).

**3. (5 points)** Solve question 3 on page 312. Represent the proper area on the standard normal distribution as you did in question 2. *See page 303-307*.

**4. (5 points)** Solve question 12 on page 313. Represent the proper area on the standard normal distribution as you did in question 2. Skip question c. *See page 303-307*.

**5. (5 points)** Solve question 21 on page 313. Represent the proper area on the standard normal distribution as you did in question 2. Note that $P(z \leq -1.28)=.10$. *See page 307-308 with special attention to example 6-17*.

**6. (5 points)** Solve question 25 on page 314. Represent the proper area on the standard normal distribution as you did in question 2. *See page 307-308 with special attention to example 6-17*.

**7. (10 points)** Solve question 31 on page 314. In addition, 1) draw the standard normal distribution and then indicate *z* values for 170 in question *a*, 14 in question *b*, and 47 in question *c*. 2) Find out the (critical) values of *x* that form the .05 and .01 significance levels for the two-tailed test. In other word, $P(x \leq \infty)\times 2=.05$ or $P(-\infty \leq x)\times 2 = .01$. You need to find *z* values and then solve equation to get critical values of *x*. 3) Represent rejection areas at the .05 level (ignore rejection areas at the .01 level here) on the standard normal distribution. *See pages 286-287*.

**8. (5 points**) Solve question 23 on page 327. *See pages 320-323*.

**9. (10 points**) Solve example 6-26 on page 331-332 by changing the batting average from .320 to .100 and the number of hits from "at most 26" to "between 15 and 20." 1) Follow all steps and draw a figure similar to Figure 6-50. 2) Is the probability likely? How do you interpret this result? (Do you think the average of .1 is true?) 3) Now, solve the question using the binomial probability distribution. You may not use the probability table on page 626-631. Factorial and fraction will be accepted as answers. The probability is .0718. 4) Is this .0718 different from what you got from the normal

distribution? What do you think make such a difference, if any? Remember the (powerpoint) slides that illustrate how binomial distributions change as N increases. 5) Which probability is more accurate? Which approach do you think is more efficient (or easy to compute)? Why?

**Bonus 5 point** (you are not required to answer): Run R from the STC machines. At the R prompt > type in "`factorial(n)/(factorial(x)*factorial(n-x))*p^x*q^(n-x)`" and hit Enter to get the probability. You need to replace n, x, p, and q with appropriate numbers. Sum the six probabilities P(x=15), P(x=16)… P(x=20) up and compare it with one you compute from the normal distribution. Which one is larger?

**10. (5 points**) Solve question 32 on page 338. *See pages 330-332*.

**11. (5 points)** Are the following hypotheses appropriate for statistical inferences? Why? And why not? *See lecture note and pages 388-390*.

1) Water consists of $H_2O$.
2) An increase in Indiana property tax rate will improve the gap between the haves and have-nots.
3) Construction of I-69 to Evansville may revitalize the economy there, worsen the air pollution, and/or threaten endangered species.
4) $H_0$: $s^2 = 1$
5) $H_0$: $\mu \leq 3.141592$

**12. (8 points)** True and false questions. *See lecture note and pages 388-390*.

1) The null hypothesis always takes the form of "no effect" and "no difference."
2) The null hypothesis and its alternative hypothesis are mutually exclusive.
3) Sample size does not matter at all in statistical inferences.
4) We may not use a significant level other than the conventional .05 and .01 level.
5) Once the significance level is determined, the critical value, rejection region, and confidence interval lead to the exactly same conclusion.
6) If you reject the null hypothesis, the hypothesis is not actually true.
7) A two-tailed test is always better (or superior to) than a one-tail test.
8) An alternative hypothesis should be what you want to claim. Therefore, an alternative hypothesis must include such phrase as "significant effect," "not equal," and "substantial difference."

**13. (5 points**) Solve example 8-5 on page 404 by changing the sample mean from $25,226 to $26,000. Follow all steps shown on page 404. In addition, check if your conclusion is different when using the .05 significance level. *See pages 401-402*.

**14. (5 points**) Solve example 8-6 by on page 407 by changing the standard deviation of population from 2 years to 3 years. *See pages 406-409*.

**15. (5 points**) Solve question 9 on page 411. Show your work clearly. *See pages 401-404*.

**16. (5 points**) Solve example 8-12 by on page 417 by changing the number of sample from 10 to 29. See the t distribution table on page 635. *See pages 415-416*.

**17. (10 points**) Solve question 20 on page 423. In addition, solve the question with an assumption that the sample mean is the population mean. Is there any big difference in conclusion? Which approach do you think is more reliable? Show your work clearly. *See pages 415-416 and 401-402*.

**18. (10 points**) Solve question 19 on page 355. Do you think the population mean is 60 decibels? In addition, draw a conclusion on the basis of the p-value. Which approach do you like in terms of computation and interpretation? Show your work clearly. *See pages 346-347*.

**19. (5 points**) Solve question 16 on pages 362-363. Do you think the population mean is 45? Show your work clearly. *See pages 358-359*.

**20. (10 points**) Solve question 20 on page 363 by changing the confidence level from 98 percent to 99 percent. You have to find the critical value from the t distribution table. Do you think the population mean is 60? What if we assume that the sample variance is equal to the population variance? In order word, conduct the hypothesis test using the standard normal distribution (5 points). Which confidence interval is wider? Can you explain me why? You need to consider the amount of information used in each approach. Show your work clearly. *See pages 358-359 and 346-347*.

**21. (5 points**) Solve question 15 on page 431. Note that $H_0$: $\pi$=.18. *See pages 425-428*.

**22. (5 points**) Solve question 14 on page 371 by changing the confidence level from 95 percent to 99 percent. Do you think the population proportion is .6? Note that $H_0$: $\pi$=.6. *See pages 365-368*.

***"Do not use any statistical method that you do not understand."***