

**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 17<sup>th</sup>, **you may not submit the assignment after the due date.** The answer key will be released at 2:00 P.M. on October 15<sup>th</sup>.
- You MAY NOT discuss with other classmates when answering questions.

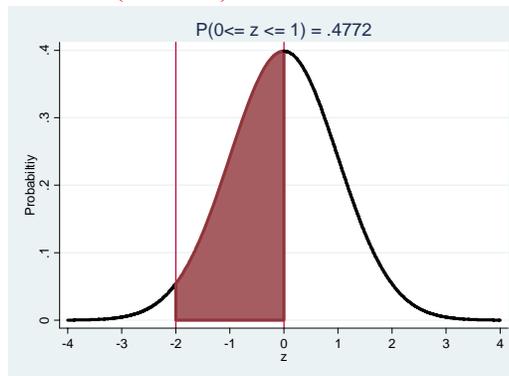
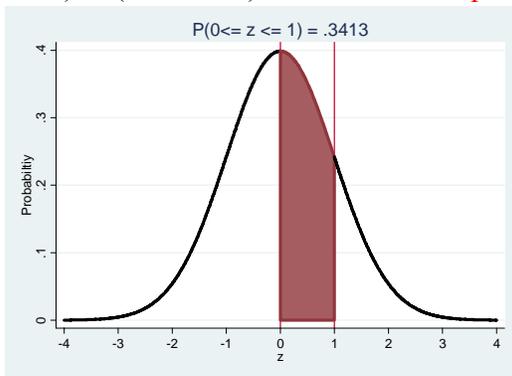
**1. (12 points)** True/False questions. *See page 284-285.*

- 1) A normal probability distribution is a distribution for a continuous random variable. **True**
- 2) A normal probability distribution is defined by its mean and variance. **True**
- 3) The mean, 50<sup>th</sup> percentile, and mode of a normal probability distribution are equal in a normal distribution. **True**
- 4) The standard normal distribution always has a mean 0 and a variance 1. **True**
- 5) In a normal distribution of a random variable  $x$ , the probability that  $x$  is -1 billion is numerically zero,  $P(x=-1,000,000,000)=0$ . **False. Any probability of  $x$  in a normal distribution is always greater than 0. Yes, the probability must be virtually (not numerically) zero.**
- 6) All normal probability distributions can be transformed into the standard normal distribution. **True**
- 7) If a random variable  $x$  is normally distributed,  $x \sim N(5, 9)$ , the probability distribution is not skewed to the left or right. **True**
- 8) Any random variable follows a normal probability distribution when  $N$  is large. **False. A random variable may or may not be normally distributed regardless of  $N$ .**
- 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. **False. When  $N$  is large, a binomial probability distribution is approximated to a normal distribution. If  $p$  is small or extremely large, more  $N$  is needed to reach the normal distribution. See the powerpoint slides I showed you.**
- 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  $\sigma^2$ . **False. You can get probabilities plug in some numbers in a normal probability distribution function.**
- 11) In a normal probability distribution, the probability of mean  $\mu$  is larger than those of any other points of  $x$ ,  $P(x=\mu) > P(x=x_i)$  where  $i$  is any number except  $\mu$ . **True**

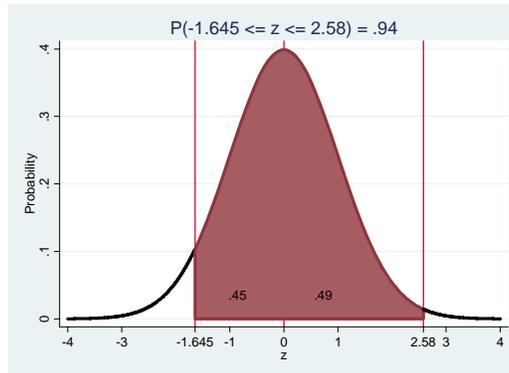
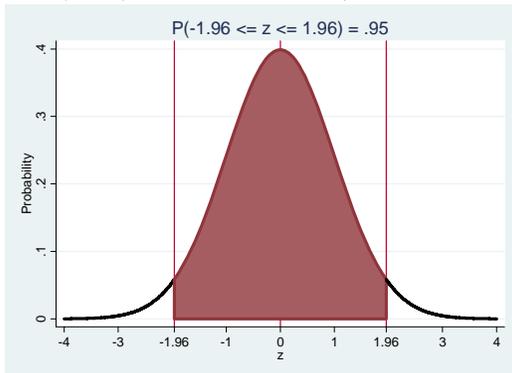
- 12) In the standard normal distribution, the probability that  $z$  is greater than  $-2.58$  and smaller than  $2.58$  is  $.99$ ,  $P(-2.58 \leq z \leq 2.58) = .99$ . Thus,  $P(-100 \leq z \leq 100)$  is obviously larger than  $1$ . **False.  $P(-100 \leq z \leq 100)$  is greater than zero but smaller 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.

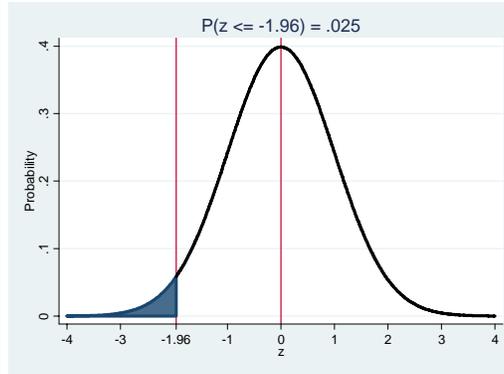
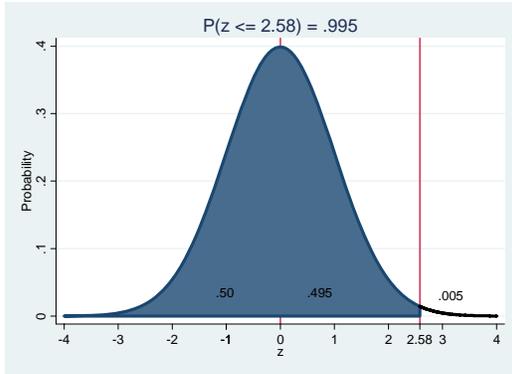
- 1)  $P(0 \leq z \leq 1)$  **.3413 from the standard normal distribution table.**
- 2)  $P(-2 \leq z \leq 0)$  **.4772. This is equivalent to  $P(0 \leq z \leq 2)$**



- 3)  $P(-1.96 \leq z \leq 1.96)$  **.95 = .475 + .475**
- 4)  $P(-1.645 \leq z \leq 2.58)$  **.94 = .45 + .49**



- 5)  $P(-\infty \leq z \leq 2.58)$  **.995 = .5 + .495. Note  $P(-\infty \leq z \leq 0) = P(0 \leq z \leq \infty) = .5$**
- 6)  $P(-\infty \leq z \leq -1.96)$  **.025 = .5 - .475.**

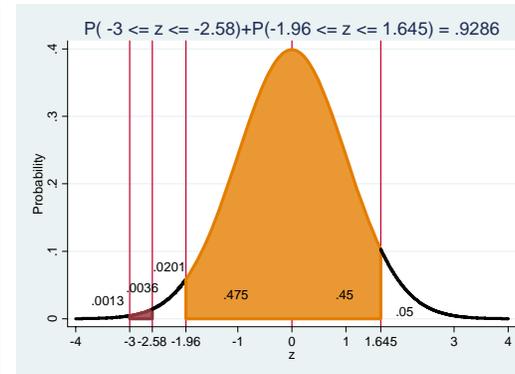
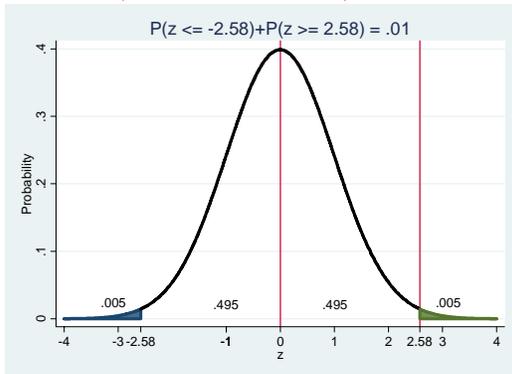


7)  $P(-\infty \leq z \leq -2.58 \text{ or } 2.58 \leq z \leq \infty) .01 = (.50 - .495) + (.50 - .495) = .005 \times 2$

8)  $P(-3 \leq z \leq -2.58 \text{ or } -1.96 \leq z \leq 1.645) .9286 = .0036 + .925$

$P(-3 \leq z \leq -2.58) = P(0 <= z <= 3) - P(0 <= z <= 2.58) = .4987 - .4951 = .0036$

$P(-1.96 \leq z \leq 1.645) = .925 = .475 + .45$



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).  $0 = P(z=0) - P(z=0)$

**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.

Information:  $\mu = 618,319$ ,  $\sigma = 50,200$

a.

$$z_x = \frac{x - \mu}{\sigma} = \frac{700,000 - 618,319}{50,200} = 1.6271$$

. di (700000-618319)/50200  
1.6271116

. di 1-normal(1.627116)  
.05185623

b.

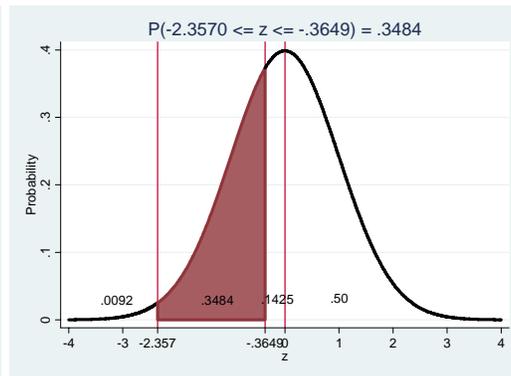
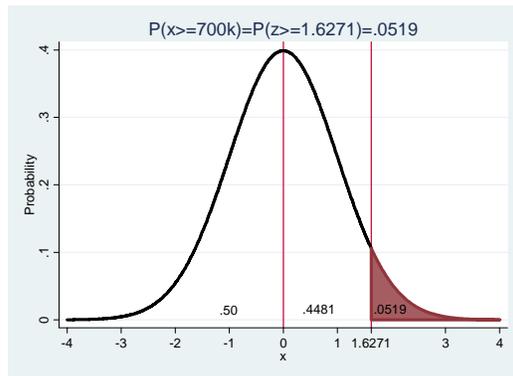
$$z_x = \frac{x - \mu}{\sigma} = \frac{500,000 - 618,319}{50,200} = -2.3570$$

$$z_x = \frac{x - \mu}{\sigma} = \frac{600,000 - 618,319}{50,200} = -.3649$$

```
. di (500000-618319)/50200
-2.3569522

. di (600000-618319)/50200
-.36492032

. di normal(-.36492032)-normal(-2.3569522)
.34837263
```



**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.*

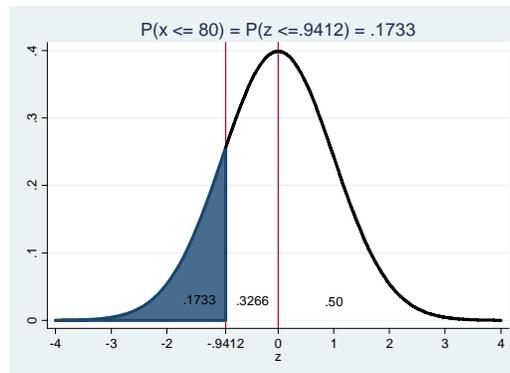
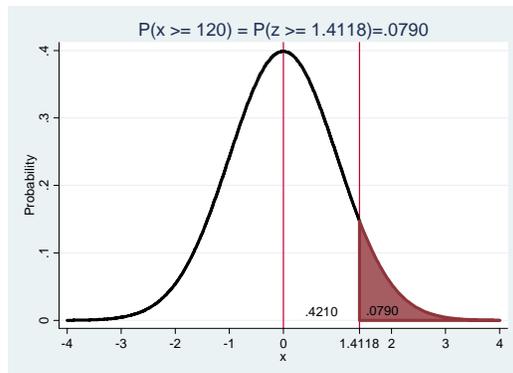
**Information:**  $\mu=96$   $\sigma=17$ . *See question 3 for computation.*

```
. di (120-96)/17
1.4117647

. di 1-normal(1.4117647)
.07900963

. di (80-96)/17
-.94117647

. di normal(-.94117647)
.17330722
```



**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.*

**Information:**  $\mu=949$   $\sigma=100$ .

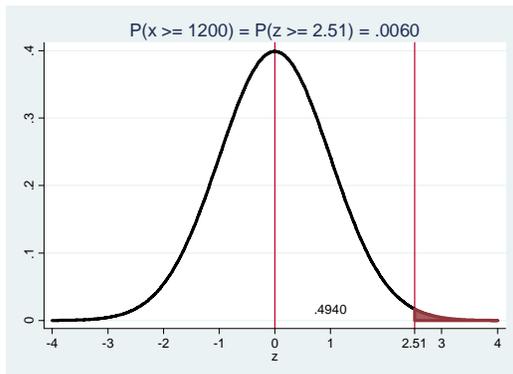
```
. di (1200-949)/100
2.51
```

```
. di 1-normal(2.51)
.00603656
```

$P(x \geq 1200)$  indicates a one-tailed test.

$$x = \mu + z_{\alpha}\sigma = 949 + (-1.28) \times 100 = 821$$

```
. di -1.28*100+949
821
```



**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.*

Information:  $\mu=5.9$   $\sigma=1.7$ .

From the standard normal distribution table, I found  $-1.036$  is critical value of the one-tailed test at the  $.15$  level. That is,  $P(z \leq -1.036) = .15$ . Now, I convert this  $z$  score (critical value) back to the corresponding  $x$  of the original normal distribution.

$$x = \mu + z_{\alpha}\sigma = 5.9 + (-1.0364) \times 1.7 = 4.1381$$

```
. di normal(-1.0364)
.15000779
```

```
. di -1.0364*1.7+5.9
4.13812
```

The critical  $z$  value of the one-tailed test at the  $.25$  level is  $.6745$ . See the table. That is  $(P(z \geq .6745) = .25)$ . Use the above formula to get  $7.0467$ .

$$x = \mu + z_{\alpha}\sigma = 5.9 + .6745 \times 1.7 = 7.0467$$

```
. di 1-normal(.6745)
.24999674
```

```
. di .6745*1.7+5.9
7.04665
```

**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 words,  $P(x_1 \leq x \leq \infty) \times 2 = .05$  or  $P(-\infty \leq x \leq -x_1) \times 2 = .01$ . 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 original distribution. *See pages 286-287.*

You need to identify mean and standard deviation from the plot. The critical values of the two-tailed test in standard normal distribution are 1.96 for the .05 level and 2.58 for the .01 level. Note that a normal distribution is symmetric and has the mean and median at the peak.

a. mean: 120, standard deviation: 20,  $2.5=(170-120)/20$ , 159.2, 171.6

b. mean: 15, standard deviation: 2.5,  $-4=(14-15)/2.5$ , 19.9, 21.45

c. mean: 30, standard deviation: 5,  $3.4=(47-30)/5$ , 39.8, 42.9

a. mean: 120, standard deviation: 20

$x = \mu + z_{\alpha/2}\sigma = 120 + 1.96 \times 20 = 159.2$  for the .05 significance level

```
. di 1.96*20+120
159.2
```

This critical value means that  $P(x \leq -159.2) + P(x \geq 159.2) = .05$ . You need to be comfortable going back and forth between the original normal distribution and the standard normal distribution.

$x = \mu + z_{\alpha/2}\sigma = 120 + 2.58 \times 20 = 171.6$  for the .01 significance level

```
. di 2.58*20+120
171.6
```

This critical value means that  $P(x \leq -171.6) + P(x \geq 171.6) = .01$ .

b. mean: 15, standard deviation: 2.5

```
. di 1.96*2.5+15
19.9
```

```
. di 2.58*2.5+15
21.45
```

c. mean: 30, standard deviation: 5

```
. di 1.96*5+30
39.8
```

```
. di 2.58*5+30
42.9
```

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

Information:  $\mu=215$   $\sigma=15$ .

a.  $P(x \geq 220) = P(z \geq .3333) = .3694$

$$z_x = \frac{x - \mu}{\sigma} = \frac{220 - 215}{15} = .3333$$

```
. di (220-215)/15
.33333333
```

```
. di 1-normal(.33333333)
.36944134
```

b.  $P(\bar{x} \geq 220) = P(z \geq 1.6667) = .0478$

$$z_{\bar{x}} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{220 - 215}{15/\sqrt{25}} = 1.6667$$

```
. di (220-215)/(15/sqrt(25))
1.6666667

. di 1-normal(1.6666667)
.04779035
```

Keep in mind that we are talking about sample mean, not a particular data point. The standard deviation of a sample mean is difference from those of particular data points. Since population variance (standard deviation) is known, you need to conduct z-test.

**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?

Let us collect necessary information first. I got N=100, x=15~20, p=.1, and q=.9.

1)

Step 1: the binomial distribution is approximated to the normal distribution since N is large.

Step 2: Mean:  $10=np=100 \times .1$ , variance:  $9=npq=100 \times .1 \times .9$ , standard deviation:  $3=\sqrt{9}$

Step 3:  $P(15 \leq x \leq 20)$

Step 4: Draw a figure to indicate the area you need to compute

Step 5: Compute z value.

$$z_x = \frac{x - \mu}{\sigma} = \frac{15 - 10}{3} = 1.6667$$

```
. di (15-10)/3
1.6666667
```

$$z_x = \frac{x - \mu}{\sigma} = \frac{20 - 10}{3} = 3.3333$$

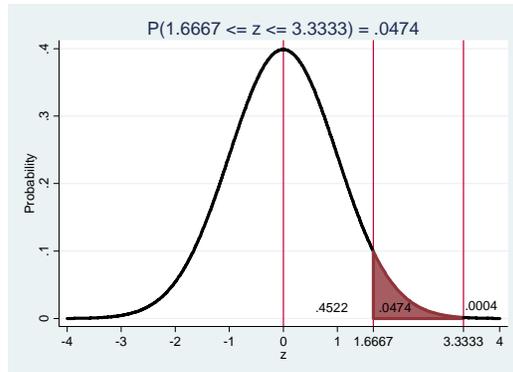
```
. di (20-10)/3
3.3333333
```

Step 6: Compute the area:  $P(15 \leq x \leq 20) = .0474$

```
. di normal(3.3333333)
.99957094

. di normal(1.6666667)
.95220965

. di normal(3.3333333)-normal(1.6666667)
.04736129
```



2) The probability .0474 is not likely. I would doubt  $p = .1$  if a sample mean is between 15 and 20. In other words, a player with .1 batting rate is not likely to hit 15-20 out of 100 chances. Therefore, the player may have a higher rate; I feel like changing my conjecture of the rate .1.

3)  $P(15 \leq x \leq 20) = .0718$  from a binomial distribution

$$.0327 = 100! / (15!85!) \times .1^{15} \times .9^{85}$$

$$.0193 = 100! / (16!84!) \times .1^{16} \times .9^{84}$$

$$.0106 = 100! / (17!83!) \times .1^{17} \times .9^{83}$$

$$.0054 = 100! / (18!82!) \times .1^{18} \times .9^{82}$$

$$.0026 = 100! / (19!81!) \times .1^{19} \times .9^{81}$$

$$.0012 = 100! / (20!80!) \times .1^{20} \times .9^{80}$$

4) The .0718 from a binomial distribution is larger than .0474 from the standard normal distribution. If  $p$  is extremely small or large, the binomial distribution requires more  $N$  to approach a normal distribution. If  $p$  is close .5,  $N=100$  may be sufficient for the normal approximation. Look at the slides again.

5) Therefore, .0718 appears more accurate in this case as long as  $x$  follows a binomial distribution. It is much easier to use a normal distribution when  $N$  is large enough. Computing six probabilities is painful. Don't you think so?

The following is the output of computation in Stata. Usage of factorial function in R is much easier than that in Stata.

```
. di exp(lnfactorial(100))/exp(lnfactorial(15))/exp(lnfactorial(85))* .1^15*.9^85
```

```
.03268244
. di exp(lnfactorial(100))/exp(lnfactorial(16))/exp(lnfactorial(84))* .1^16*.9^84
.01929172
. di exp(lnfactorial(100))/exp(lnfactorial(17))/exp(lnfactorial(83))* .1^17*.9^83
.01059153
. di exp(lnfactorial(100))/exp(lnfactorial(18))/exp(lnfactorial(82))* .1^18*.9^82
.00542653
. di exp(lnfactorial(100))/exp(lnfactorial(19))/exp(lnfactorial(81))* .1^19*.9^81
.00260219
. di exp(lnfactorial(100))/exp(lnfactorial(20))/exp(lnfactorial(80))* .1^20*.9^80
.00117099
. di .03268244+.01929172+.01059153+.00542653+.00260219+.00117099
.0717654
```

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

Information:  $P=.53$ ,  $q=.47$ ,  $n=250$ ,  $x=120$

Mean:  $132.5=np=250 \times .53$

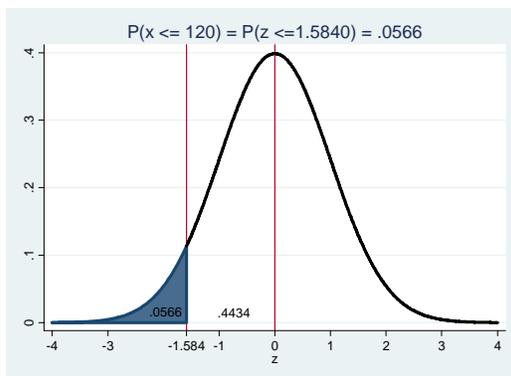
Variance:  $62.275=npq=250 \times .53 \times .47$

Standard deviation:  $7.8915=\text{sqrt}(62.275)$

$$z_x = \frac{x - \mu}{\sigma} = \frac{120 - 132.5}{7.8915} = -1.5840$$

$P(x \leq 120) = P(z \leq -1.5840) = .0566$ . In this case, it will be painful to use a binomial distribution. Why? You have to compute 120 factorials!

```
. di (120-132.5)/7.8914511
-1.5839926
. di normal(-1.5839926)
.0565977
```



Some of you compute using 119.5 instead of 120. The textbook may be misleading. This is a continuous variable. Even when we use 119.5, there are so many values between 119.5 and 120. The effort to make it more sense is intuitive and understandable but useless at best. Whether 120 inclusive or exclusive does not make any difference. See 9) of Q2. In fact,  $p(z < 120)$  and  $p(z \leq 120)$  are same from the standpoint of statistical inference.

**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<sub>2</sub>O. **Neither interesting nor informative.**
- 2) An increase in Indiana property tax rate will improve the gap between the haves and have-nots. **Yes (but I would use the present tense to make it more specific).**
- 3) Construction of I-69 to Evansville may revitalize the economy there, worsen the air pollution, and/or threaten endangered species. **Not specific.**
- 4) H<sub>0</sub>:  $s^2 = 1$  **H<sub>0</sub> should be about population, not about sample.**
- 5) H<sub>0</sub>:  $\mu \leq 3.141592$  **Why not? This is a H<sub>0</sub> for one-tailed test.**

**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.” **Not always.**
- 2) The null hypothesis and its alternative hypothesis are mutually exclusive. **True.**
- 3) Sample size does not matter at all in statistical inferences. **False. Yes, it does matter.**
- 4) We may not use a significant level other than the conventional .05 and .01 level. **False. You may use such levels as .11, .6, .2. It is up to you! But conventional significance levels are necessary in order to communicate with others successfully.**
- 5) Once the significance level is determined, the critical value, rejection region, and confidence interval lead to the exactly same conclusion. **True.**
- 6) If you reject the null hypothesis, the hypothesis is not actually true. **False. Not necessarily.**
- 7) A two-tailed test is always better (or superior to) than a one-tail test. **False. Not necessarily.**
- 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.” **False. Not necessarily. Some researchers may want to show (his claim) there is no substantial effect. For example, the Bush administration may insist that their war against terrorism has improved U.S. security substantially. But some scholars may argue against by illustrating negligible or negative effect. In this case, H<sub>0</sub> is “no effect” and H<sub>a</sub> is “significant effect.” Scholars do not want to reject the null hypothesis because the null hypothesis is their claim.**

**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.*

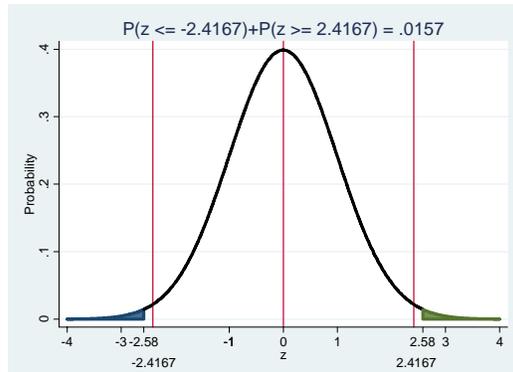
**Necessary information:**  $\mu=24,672$ ,  $\sigma=3251$ , sample mean: 26,000,  $N=35$ ,  $\alpha=.01$

- 1) H<sub>0</sub>:  $\mu=\$24,672$ , H<sub>a</sub>:  $\mu\neq\$24,672$
- 2) The critical value of the two-tailed test at the .01 significance level is  $\pm 2.58$ .
- 3) Test statistic is  $z_{\bar{x}} = 2.4167 = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{26000 - 24672}{3251/\sqrt{35}} \sim N(0,1)$

- 4) The test statistic 2.4167 is smaller than 2.58; do not reject the null hypothesis. (The p-value .0157 is larger than the significance level .01, indicating that the risk you would take is larger than the .01 level, your subjective criterion)
- 5) The average cost of rehabilitation appears to be \$24,672 at the .01 level.

```
. di (26000-24672)/(3251/sqrt(35))
2.4166576

. di (1-normal(2.4166576))*2
.01566374
```



If you use the .05 significance level, the test statistic is compared to the critical value for the .05 level, which is 1.96. The test statistic (2.4167) is larger than the critical value (1.96). The p-value .0157 is smaller than the .05 significance level; it is less risky than you thought to reject the null hypothesis. Therefore, you reject the null hypothesis at the .05 level. The average cost may not be \$24,672 at the new significance level.

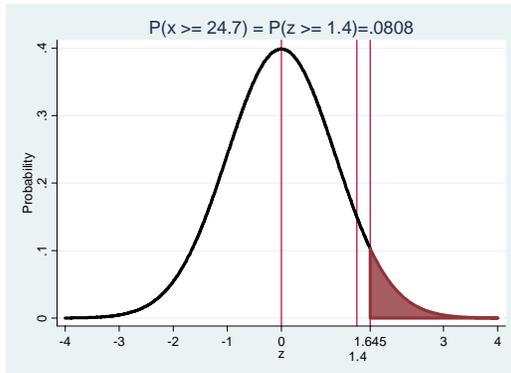
**14. (5 points)** Solve example 8-6 by on page 407 by changing the standard deviation of population from 2 years to 3 years. *See pages 406-409.*

Necessary information:  $\mu=24$ ,  $s=3$ , sample mean: 24.7,  $N=36$ ,  $\alpha=.05$ . According to the textbook, the standard normal distribution can be used when  $N$  is larger than 30 (although I prefer the t distribution when  $\sigma$  is unknown).

- 1)  $H_0: \mu \leq 24$ ,  $H_a: \mu > 24$
- 2) At the .05 significance level (the critical value of the one-tailed test at the .05 level is  $\pm 1.645$ ).
- 3) Test statistic is  $1.4 = (24.7 - 24) / (3 / \sqrt{36})$ . Its p-value is .0808.
- 4) The p-value (.0808) is larger than the significance level of .05 (The test statistic 1.4 is smaller than 1.645). Thus, we do not reject the null hypothesis at the .05 level.
- 5) The average age of lifeguards appears to be 24.

```
. di =(24.7-24)/(3/sqrt(36))
1.4

. di 1-normal(1.4)
.08075666
```



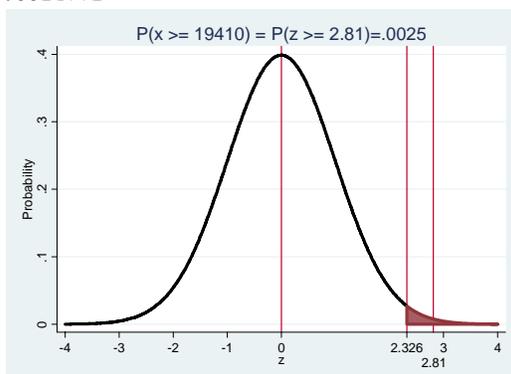
**15. (5 points)** Solve question 9 on page 411. Show your work clearly. *See pages 401-404.*

Necessary information:  $\mu=19,410$ ,  $s=6,050$ , sample mean: 22,098,  $N=40$ ,  $\alpha=.01$ . Since  $N$  is larger than 40, let us conduct a z-test. Either the test statistic or p-value approach will do.

- 1)  $H_0: \mu \geq \$19,410$ ,  $H_a: \mu < \$19,410$
- 2) The critical value of the one-tailed test at the .01 significance level is 2.326. Keep in mind that this is a one-tailed test (whether or not average cost has increased)
- 3) Test statistic is  $2.8100 = (22098 - 19410) / (6050 / \sqrt{40})$ . Its p-value is .0025.
- 4) The test statistic 2.8100 is larger than 2.326 and the p-value .0025 is smaller than the significance level of .01. Thus, we reject the null hypothesis at the .01 level.
- 5) The average undergraduate cost for tuition does not appear to have increased.

```
. di (22098-19410)/(6050/sqrt(40))
2.8099842
```

```
. di 1-normal(2.8099842)
.0024772
```



**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.*

Necessary information: hypothesized (claimed) mean =24,000,  $s=400$ , sample mean: 23,450,  $N=29$ ,  $\alpha=.05$ . Since population variance is unknown and  $N$  is smaller than 30, let us try the t-test. Either the test statistic or p-value approach will do.

- 1)  $H_0: \mu = \$24,000$ ,  $H_a: \mu \neq \$24,000$

- 2)  $N=29$ . The degrees of freedom are  $28=29-1$ . The critical value of the two-tailed test at the .05 significance level is  $\pm 2.048$ .
- 3) Test statistic is  $-7.4046=(23450-24000)/(400/\sqrt{29})$ . (Its p-value is .0025).
- 4) The test statistic  $-7.4046$  is smaller than  $-2.048$  (the p-value .0000 is smaller than the significance level of .05). If the average starting salaries are \$24,000, the probability that we get this sample mean of \$23,450 is almost zero. It is extremely unlikely. Thus, we reject the null hypothesis at the .05 level.
- 5) The average starting salaries for nurses does not appear to be \$24,000 at the .05 significance level.

```
. di ttail(28, 2.048)
.02502126

. di (23450-24000)/(400/sqrt(29))
-7.4046016

. di (1-ttail(28, -7.4046016))*2
4.599e-08
```

**17. (10 points)** Solve question 20 on page 423. In addition, solve the question with an assumption that **the sample standard deviation is the population standard deviation**. 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.*

Necessary information: hypothesized (claimed) mean  $=3.18$ ,  $s=1.4346$ , sample mean:  $3.8333$ ,  $N=24$ ,  $\alpha=.05$ . Since  $\sigma$  is unknown and  $N$  is smaller than 30, you need to conduct the t-test. Let us employ the test statistic approach.

- 1)  $H_0: \mu=3.18$ ,  $H_a: \mu \neq 3.18$
- 2) The .05 significance level. The degrees of freedom are  $23=24-1$ . The critical value of the two-tailed test is  $\pm 2.0685$ .
- 3) mean:  $3.8333$ , standard deviation:  $1.4346$ . Test statistic is  $2.2311=(3.8333-3.18)/(1.4346/\sqrt{24})$ . (Its p-value is .0357) The degree of freedom is  $23=24-1$ .
- 4) The test statistic  $2.2311$  is larger than the critical value  $2.0685$ . (The p-value .0357 is smaller than the .05 significance level.) The sample mean  $3.8333$  is not likely at the .05 level if the population average is  $3.18$ . Therefore, we reject the null hypothesis at the .05 level.
- 5) The average family size does not appear to be  $3.18$ .

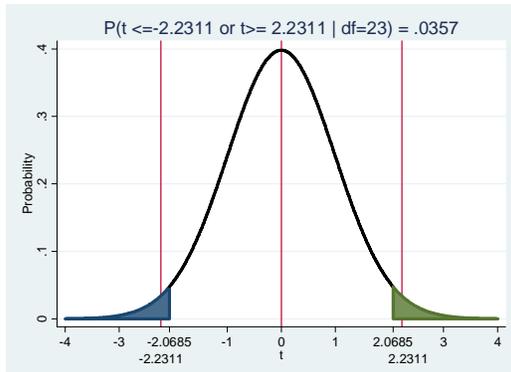
```
. di ttail(23, 2.0685)
.02500803

. di (3.83333-3.18)/(1.434563/sqrt(24))
2.2310977

. di ttail(23, 2.2310977)*2
.03571174
```

If  $\sigma$  is known as  $3.8333$ , you need to conduct the z-test. The only difference is the critical value and p-value because of the standard distribution used. The critical value of the two tailed test at the .05 level is  $1.96$ . The test statistic  $2.2311$  (remains unchanged) is larger than  $1.96$ . Thus, reject the null hypothesis at the .05 level. The p-value is .0257.

Therefore, there is no difference in conclusion even when we assume the population standard deviation is known.



```
. di (1-normal(2.2311))*2
.02567451
```

**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.*

Information: hypothesized mean=60, sample mean=61.2, sample standard deviation=7.9, N=84,  $\alpha=.05$  (95 percent confidence interval). The population standard deviation is unknown, but N is larger than 30; you need to conduct the z-test. Again, I prefer the t-test for this case since the population mean is not known.

Confidence interval approach

- 1)  $H_0: \mu=60, H_a: \mu \neq 60$
- 2) The critical value of the two-tailed test at the .05 significance level is  $\pm 1.96$ .
- 3) The 95 percent confidence interval is

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 61.2 \pm 1.96 \times \frac{7.9}{\sqrt{84}} = 61.2 \pm 1.6894. \text{ It ranges from } 59.5106 \text{ to } 62.8894.$$

- 4) Since the hypothesized population mean 60 falls into the 95 percent confidence interval, we do not reject the null hypothesis. The sample mean 61.2 is likely at the .05 level if population mean is 60.
- 5) The average noise level in hospital appears to be 60 decibels.

```
. di 61.2+1.96*7.9/sqrt(84)
62.889443
```

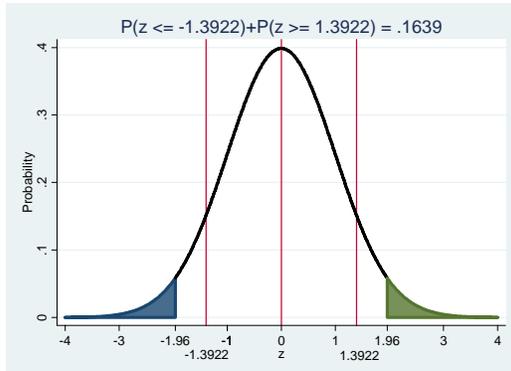
```
. di 61.2-1.96*7.9/sqrt(84)
59.510557
```

P-value approach

- 1)  $H_0: \mu=60, H_a: \mu \neq 60$
- 2) At the .05 significance level
- 3) N=84, mean: 61.2, standard deviation: 7.9. Test statistic is  $1.3922=(61.2-60)/(7.9/\sqrt{84})$ . P-value is .1639.

4) Since p-value is greater than the significance level, we do not reject the null hypothesis. You have to take 16 percent risk to be wrong when you reject the null hypothesis.

5) The average noise level in hospital appears to be 60 decibels.



```
. di (61.2-60)/(7.9/sqrt(84))
1.3921749
```

```
. di (1-normal(1.3921749))*2
.16386944
```

The p-value approach is generally used. But it depends.

**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.*

Information: hypothesized (population) mean=45, sample mean=41.6, sample standard deviation=7.5, N=24,  $\alpha=.05$  (95 percent confidence interval). Since  $\sigma$  is unknown and N is smaller than 30, you need to conduct the t-test.

Confidence interval approach

1)  $H_0: \mu=45$ ,  $H_a: \mu \neq 45$

2) N=24, the degrees of freedom are  $23=24-1$ . The critical value of the two-tailed test at the .05 significance level is  $\pm 2.0685$ .

3) mean: 41.6, standard deviation: 7.5. The 95 percent confidence interval is

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 41.6 \pm 2.0685 \times \frac{7.5}{\sqrt{24}} = 41.6 \pm 3.1667. \text{ It ranges from } 38.4333 \text{ to } 44.7667.$$

4) Since the hypothesized population mean 45 falls beyond the 95 percent confidence interval, we reject the null hypothesis.

5) The average noise level in hospital does not appear to be 45 decibels.

```
. di 2.0685*7.5/sqrt(24)
3.166731
```

```
. di 41.6+2.0685*7.5/sqrt(24)
44.766731
```

```
. di 41.6-2.0685*7.5/sqrt(24)
38.433269
```

**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 other words, 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.*

**Information:** hypothesized (population) mean=60, sample mean=33.5, sample standard deviation=27.6777, N=10,  $\alpha=.01$  (99 percent confidence interval). Since  $\sigma$  is unknown and N is smaller than 30, you need to conduct the t-test.

- 1)  $H_0: \mu=60, H_a: \mu \neq 60$
- 2)  $N=10$ , the degrees of freedom are  $9=10-1$ . The critical value of the two-tailed test at the .01 significance level is  $\pm 3.2495$ .
- 3) mean: 33.5, standard deviation: 27.6777. The 99 percent confidence interval is  $\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = 33.5 \pm 3.2495 \times \frac{27.6777}{\sqrt{10}} = 33.5 \pm 28.4411$ . It ranges from 5.0589 to 61.9411.
- 4) Since the hypothesized population mean 45 falls into the 99 percent confidence interval, we do not reject the null hypothesis.
- 5) The average unhealthy day appears to be 60 days.

```
. di ttail(9,3.2495)
.00500269

. di 3.2495*27.6777/sqrt(10)
28.44111

. di 33.5+3.2495*27.6777/sqrt(10)
61.94112

. di 33.5-3.2495*27.6777/sqrt(10)
5.0588799
```

If the population mean is equal to the sample mean, you need to conduct the z-test. The only difference is the critical value.

- 1)  $H_0: \mu=60, H_a: \mu \neq 60$
- 2) The critical value of the two-tailed test at the .01 significance level is  $\pm 2.58$ .
- 3) mean: 33.5, standard deviation: 27.6777. The 99 percent confidence interval is  $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} = 33.5 \pm 2.58 \times \frac{27.6777}{\sqrt{10}} = 33.5 \pm 22.5813$ . It ranges from 10.9187 to 56.0813.
- 4) Since the hypothesized population mean 45 falls into the 99 percent confidence interval, we do not reject the null hypothesis.
- 5) The average unhealthy day appear to be 60 days.

```
. di 2.58*27.6777/sqrt(10)
22.58134

. di 33.5-2.58*27.6777/sqrt(10)
10.918652

. di 33.5+2.58*27.6777/sqrt(10)
```

56.081348

The first approach (t-test) has a wider confidence interval because the population variance is unknown. The sample variance is used instead. Therefore, estimation of the t-test is less accurate than the z-test.

**21. (5 points)** Solve question 15 on page 431. Note that  $H_0: \pi=.18$ . See pages 425-428. Information: hypothesized (population) mean=.18, sample p ( $\hat{p}$ )=50/300,  $q=250/300$ ,  $N=300$ ,  $\alpha=.05$

1)  $H_0: \pi=.18$ ,  $H_a: \pi \neq .18$

2) The critical value of the two-tailed test at the .05 significance level is  $\pm 1.96$ .

3)  $N=300$ ,  $p=50/300$ ,  $q=250/300$ , test statistics is

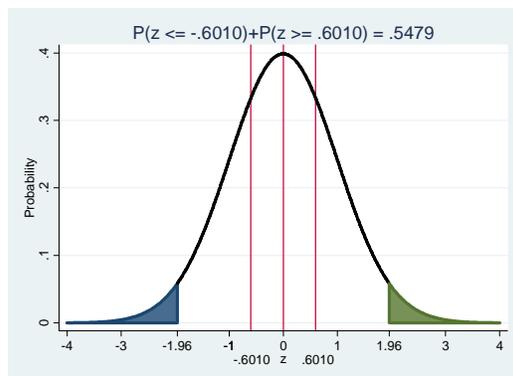
$$z = \frac{\hat{p} - p}{\sqrt{pq/n}} = \frac{.1667 - .18}{\sqrt{.18 \times .82/300}} = -.6010. \text{ P-value is } .5479.$$

4) Since the test static  $|- .6010|$  is smaller than the critical value of 1.96, we do not reject the null hypothesis. (the p-value is larger than the .05 level, do not reject the null hypothesis)

5) The average you smoking rate appears to be .18.

```
. di (.16667-.18)/sqrt(.18*.82/300)
-.60096281
```

```
. di normal(-.60096281)*2
.54786476
```



**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.

Information: hypothesized (population) mean=.6, sample  $p=.56=560/1000$ ,  $q=.44$ ,  $N=1000$ ,  $\alpha=.01$

1)  $H_0: \pi=.6$ ,  $H_a: \pi \neq .6$

2) The critical value of the two-tailed test at the .01 significance level is  $\pm 2.58$ .

3)  $N=1000$ ,  $\hat{p}=.56=560/1000$ ,  $\hat{q}=.44=440/1000$ . The 99 percent confidence

interval is  $\hat{p} \pm z_{\alpha/2} \frac{pq}{\sqrt{n}} = .56 \pm 2.58 \times \sqrt{\frac{.56 \times .44}{1000}} = .56 \pm .0405$ . It ranges from .5195

to .6005.

4) Since the hypothesized proportion falls into the confidence interval, we do not reject the null hypothesis at the .01 significance level.

5) The proportion of voters who say that the U.S. spends too little on fighting hunger at home is .6.

```
. di .56+ 2.58*sqrt(.56*.44/1000)  
.6004986
```

```
. di .56- 2.58*sqrt(.56*.44/1000)  
.5195014
```