COB 191 VERSION AAA Name______________________________

Statistical Methods Dr. Ping Wang

Exam 3 Fall 2004

The following exam consists of 40 questions, each worth 2.5 points. You will have 75 minutes to complete the test. This means that you have, on average, about 1 minute, 50 seconds per question.

1. Record your answer to each question on the scantron sheet provided. You are welcome to write on this exam, but your scantron will record your graded answer.
2. Read carefully, and check your answers. Don't let yourself write nonsense.
3. Keep your eyes on your own paper. If you believe that someone sitting near you is cheating, raise your hand and quietly inform me or the instructor of this. I'll keep an eye peeled, and your anonymity will be respected.
4. If any question seems unclear or ambiguous to you, do your best to answer the question and write down your comments, and I will look into it when I grade the test.
5. Be sure your correctly record your student number on your scantron, and blacken in the corresponding digits. **Failure to do so will cost you 10 points on this exam!**

Pledge: On my honor as a JMU student, I pledge that I have neither given nor received unauthorized assistance on this examination.

Signature ______________________________________

1. Which of the following statement about the standard error of the mean is NOT true?
   a). It is never larger than the standard deviation of the population.
   b). It decreases as the sample size increases and the population standard deviation keeps constant.
   c). It increases as the sample size increases and the population standard deviation keeps constant.
   d). It increases as the population standard deviation increases and the sample size keeps constant.

C because $\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right)$, as the sample size n ↑, $\sigma_{\bar{x}}$ will ↓, therefore, b) is correct and (c is NOT correct.

2. The Central Limit Theorem is important in statistics because
   a). for a large n, it says the sampling distribution of the sample mean is approximately normal, regardless of the shape of the population.
   b). for a large n, it says the population is approximately normal.
   c). for any population, it says the sampling distribution of the sample mean is approximately normal, regardless of the sample size.
   d). for any sized sample, it says the sampling distribution of the sample mean is approximately normal.

A the sample mean $\bar{x}$ is approximately normal if either the population is normal or the sample size n is large enough. Thus, c) cannot be correct.

3. Which of the following statements about the sampling distribution of the sample mean is incorrect?
   a). The sampling distribution of the sample mean is approximately normal whenever the sample size is sufficiently large ($n \geq 30$).
   b). The standard deviation of the sampling distribution of the sample mean is equal to the population standard deviation.
   c). The sampling distribution of the sample mean is generated by repeatedly taking samples of size n and computing the sample means.
   d). The mean of the sampling distribution of the sample mean is equal to the population mean.

B the standard deviation of the mean is given as $\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right)$, and does not equal to the population std σ.

The optioin d) is $E(\bar{x}) = \mu$ and it is correct.

4. Why is the Central Limit Theorem so important to the study of sampling distributions?
   a). It allows us to disregard the size of the sample selected when the population is not normal.
   b). It allows us to disregard the shape of the population when the sample size n is large.
   c). It allows us to disregard the shape of the sampling distribution when the size of the population is large.
   d). It allows us to disregard the size of the population from which we draw samples.

B This is the reason we need CLT. Could you see how c) is wrong?

5. The dollar amount of gasoline purchased per car at Sheeze Gas station has a population mean of \$25 and a population standard deviation of \$6. What is the probability that the sample mean dollar amount of gasoline purchased for a sample of 4 cars to be between \$19 and \$31?
   a). The probability is 68.26% because the sample mean amount is ± 1 standard deviation away from the population mean.
   b). The probability is 95.45% because the sample mean amount is ± 2 standard deviations away from the population mean.
   c). The probability is 86.64% because the sample mean amount is ± 1.5 standard deviations away from the population mean.
   d). The probability is neither 68.26%, nor 86.64%, and nor 95.45%.

D - the population distribution is unknown and the sample size n = 4 is too small to use central limit theorem or normal distribution. This is a situation that the central limit theorem cannot help. Any attempt to solve the problem with a normal distribution is a waste of time.

<u>The following problem is associated with the questions 6 to 9</u>:
Procter and Gamble is in the process to evaluate the performance of a recent advertisement to sale its Pantene Proof products in Harrisonburg. A survey was conducted. On the basis of this survey, the following events were identified:

SA = individual recalls seeing the advertisement
NSA = individual recalls not seeing the advertisement
P = individual purchased the product
NP = individual not purchased the product

Among the 100 people returned this survey, 40 people recalled seeing the advertisement, 20 people purchased the product, and 12 people recalled seeing the advertisement and purchased the product. (Hint: you may want to create a contingency table for the problem before answer the questions)

Given a contingency table below.

Event	Purch	NotP		Row Sum
SeeAd	12	28	0	40
NotSA	8	52	0	60
Col Sum	20	80	0	100

Once you have this contingency table, you should see the answers are straight forward.

6. What is the probability of an individual's purchasing the product given that the individual recalls seeing the advertisement?
   a). 0.200 b). 0.300 c). 0.600 d). 0.080

B P(P|SA) = 12/40=0.30=30%

7. What is the probability of an individual's purchasing the product given that the individual don't recall seeing the advertisement?
   a). 0.200 b). 0.080 c). 0.400 d). 0.133

D P(P|NotSA)=8/60=0.133

8. Does seeing the advertisement increase the probability that an individual will purchase the product? As a decision maker, would you recommend continuing the advertisement (assuming that the cost is reasonable)?
   a). The probability of an individual's purchasing the product given that the individual recalls seeing the advertisement is the same as the probability of an individual's purchasing the product given that the individual don't recall seeing the advertisement. Therefore, the advertisement would not make any difference.
   b). The probability of an individual's purchasing the product given that the individual recalls seeing the advertisement is smaller than the probability of an individual's purchasing the product given that the individual don't recall seeing the advertisement. Therefore, the advertisement should not continue.
   c). The probability of an individual's purchasing the product given that the individual recalls seeing the advertisement is greater than the probability of an individual's purchasing the product given that the individual don't recall seeing the advertisement. Therefore, the advertisement should continue.
   d). The probability of an individual's purchasing the product given that the individual recalls seeing the advertisement is smaller than the probability of an individual's purchasing the product given that the individual don't recall seeing the advertisement. Therefore, the advertisement should continue.

C Your reasoning for this one should based on your computations for P(P|SA) and P(P|NotSA) above.

9. What is the probability of an individual's purchasing the product or the individual recalls seeing the advertisement?

a). 0.600 b). 0.300 c). 0.480 d). 0.120

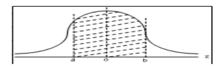
C P(P U SA)=P(P OR SA) = P(P) + P(SA) – P(P & SA) = 0.20 + 0.40 – 0.12 = 0.480

<u>The following problem is associated with the questions 10 to 15:</u>

Suppose a random variable Z has a standardized normal distribution with a mean of 0 and a standard deviation of 1.

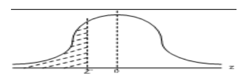
10. What is the probability that Z is between -2.5 and +2.5?

a). 0.9876 b). 0.4876 c). 0.0124 d). 0.9938

A P(-2.5 < Z < 2.5) = 2 * P(0<2.5) = 2*0.4938 = P(Z<2.5) – P(Z<-2.5)=0.9938 – 0.0062 

11. What is the probability that Z is greater than -1.5?

a). 0.5668 b). 0.4332 c). 0.0668 d). 0.9332



D P(Z>-1.5) is the white area =0.5+P(0<Z<1.5)=0.5+0.4332=0.9332=1-0.0668

12. Which one(s) of the following statements about the random variable Z is correct?

I. 3 of every 4 observations would fall between ±1 standard deviations around the mean.

II. 9 of every 10 observations would fall between ±1.645 standard deviations around the mean.

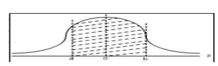
III. 99 of every 100 observations would fall between ± 2.576 standard deviations around the mean.

a). II and III only b). I and II only c). I and III only d). I, II and III

A ¾=0.75, not 68, thus option I is not correct and any choice has I is also not correct.
9/10=0.90 and P(-1.645<Z<+1.645)=0.90, thus II is correct.
99/100 = 0.99 and P(-2.576<Z<2.576) =0.99, thus III is correct.

13. For some positive value of Z, the probability that a standardized normal variable is between –Z and +Z is 0.95. The value of Z is:

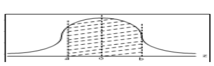
a). 1.645 b). 1.960 c). 0.1257 d). 2.576

B P(-1.96<Z<1.96)=0.95 as given in the graph. 

14. For some positive value of Z, the probability that a standardized normal variable is between –Z and + Z is 68.27%. The value of Z can be found by which of the following Excel functions?

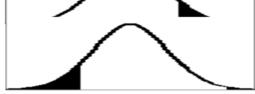
a). =NORMSINV(0.6827 / 2) b). =NORMSINV(0.5+0.6827 / 2)

c). =NORMSDIST(0.5+0.6827 / 2) d). =NORMSINV(0.6827)

B  This is the case 4, given the probability to find out the Z values. We should use Excel function =NORMSINV(prob*) that returns the Z* value corresponding to P(Z<Z*)=0.5+0.6827 / 2 as in the graph above.

15. For some positive value of Z, the probability that a standard normal variable Z is greater than 2 can be given in the following ways:

I.  the dark right tail equals to P(Z>2)=0.0228

II.  The dark left tail equals to P(Z>2)=0.0228

III.  The whole white area on the chart equals to P(Z>2)=0.9772

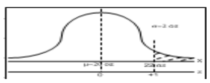
a). III only. b). II only c). I only d). II and III only

C only the option I is correct here.

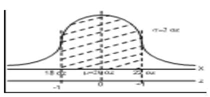
<u>The following problem is associated with the questions 16 to 24:</u>

Suppose that the length of time it takes a JMU student to find a parking spot follows a normal distribution with a mean of 15 minutes and a standard deviation of 5 minutes.

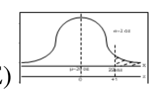
16. What is the probability that a randomly selected JMU student will find a parking spot in more than 18.5 minutes?
    a). 0.242 b) 0.700 c). 0.758 d). 0.2580

A  P(X>18.5) = P{Z>(18.5-15)/5=0.70}=0.242

17. What is the probability that a randomly selected JMU student will find a parking spot between 10 minutes and 25 minutes?
    a). 0.5228 b) 0.8186 c). 0.1814 d). 0.4772

B  P(10<X<25)=P(-1<Z<+2)=0.3413+0.4772=0.8186

18. Which Excel function is correct for you to find the probability that a randomly selected JMU student will find a parking spot in more than 18.5 minutes?
    a). =1-NORMDIST(18.5,15,5,TRUE) b). =NORMDIST(18.5,15,5,TRUE)
    c). =0.5-NORMDIST(18.5,15,5,TRUE) d). =NORMINV(18.5,15,5)

A Because Excel function =NORMDIST is cumulative, thus the use of =1-NORMDSIT(18.5,15,5,TRUE) 

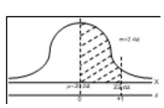
19. Which statement is correct about the probability that a randomly selected JMU student will find a parking spot between X minutes to Y minutes?
    I. About 1 out of 4 JMU students spent less than 11.6 minutes to find a parking spot.
    II. About 1 out of 4 JMU students spent more than 18.4 minutes to find a parking spot.
    III. About 4 out of 5 JMU students spent more than 8.6 minutes and less than 21.4 minutes to find a parking spot.
    a). I and II only b). I and III only c). I, II and III d). II and III only

C ¼ = 0.25, P(X<11.6)=P{Z<(11.6-15)/5=-0.70}=0.2483 ≈ 0.25, thus the options I is correct.

¼ = 0.25, P(X>18.4)=P(X<11.6) ≈ 0.25, could you see 15-11.6 = 18.4-15 ? thus the options II is correct.

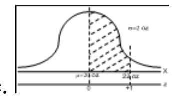
4/5= 0.80, P(8.6<X<21.4) = 0.7995 ≈ 0.80, thus the option III is correct.

20. What is the number of minutes X such that a randomly selected JMU student has a probability 0.40 to find a parking spot between the mean 15 minutes and X minutes?
    a). 1.2816 b). 21.4078 c). 13.7333 d). 16.2667



B P(15<X<X*)=0.40 z*=1.28, X* = 15 + 1.28 * 5 = 21.4078

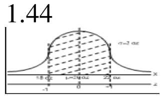
21. What is the Excel function to find the number of minutes X such that a randomly selected JMU student has a probability 0.40 to find a parking spot between the mean 15 minutes and X minutes?
    a). =NORMINV(0.40,15,5) b). =NORMINV(0.5-0.40,15,5)
    c). =NORMDIST(0.40,15,5) d). =NORMINV(0.5+0.40,15,5)



D Excel =NORMINV should be used to find X value given its probability and =NORMINV is cumulative.

The total probability up to the cut off point is 0.5+0.40=0.90, so is the answer d).

22. What is the number of minutes X such that a randomly selected JMU student has a probability 0.85 to spend between 15 – X minutes and 15 + X minutes to find a parking spot?
    a). 1.44 b). 5.18 c). 7.80 d). 7.20

D  what we really looking for is the distance away from the mean 15 on the graph. The point 0.85/2 gives the z* =1.44 value and X = (5) (z*) = 7.20=15-7.8=22.2-15

23. What is the Excel functions to find the number of minutes Y=15-X and Z=15+X such that a randomly selected JMU student has a probability 0.85 to spend between Y=15-X and Z=15+X minutes to find a parking spot? (where X is a positive number)
    I. =NORMINV(0.85,15,5)
    II. =NORMINV(1-0.85,15,5)
    III =NORMINV(0.5-0.85/2,15, 5)
    IV. =NORMINV(0.5+0.85/2,15, 5)

    a). III and IV only b). II and III only c). I and II only d). II and IV only

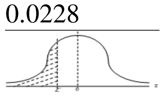
A look at the reasoning for Question 21 and keep in mind = NORMINV is cumulative.

24. Suppose a random sample of 16 students is taken. What is the standard error of the mean number of minutes to find a parking spot?
    a). 5.0000 b). 0.3125 c). 1.5625 d). 1.2500

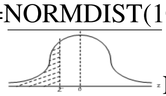
D the standard error of the mean is $\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right) = \frac{5}{\sqrt{16}} = \frac{5}{4} = 1.25$

<u>The following problem is associated with the questions 25 to 32:</u>
Suppose that the length of time it takes a JMU student to find a parking spot follows a normal distribution with a mean of 15 minutes and a standard deviation of 5 minutes. A random sample of 4 students is taken and the standard deviation of the mean is thus 2.5 minutes.

25. What is the probability that the sample mean length of time to find a parking spot is less than 10 minutes?
    a). 0.0228 b). 0.1587 c). 0.9772 d). 0.4772

A  z* = (10-15)/2.5 = -2 so P(Xbar < 10) = P(Z<-2) = 0.5 – 0.4772 = 0.0228

26. What is the Excel function to compute the probability that the sample mean length of time to find a parking spot is greater than 10 minutes?
    a). =1-NORMDIST(10,15,2.5,TRUE) b). =1-NORMDIST(10,15,5,TRUE)
    c). =NORMDIST(10,15,2.5,TRUE) d). =NORMDIST(10,15,5,TRUE)

A  Now we look for the white area in the graph and thus to use 1-NORMDIST. 2.5 should be used.

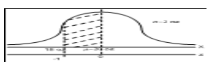
27. What is the mean $\mu_{\bar{x}}$ of the sampling distribution of the mean length of time to find a parking spot?
    a). 15 / 4 b). 15 / 2 c). $15/\sqrt{2}$ d). 15

D This is by the unbiased ness of the sampling distribution of the mean or $E(\bar{x}) = \mu$.

28. Which of the following is necessarily true about the sampling distribution of the mean length time?
    a). It is normally distributed because the sample size is small.
    b). It is normally distributed because the sample size is large.
    c). It is normally distributed, since the population of time to find a parking spot is normally distributed.
    d). It is normally distributed because sampling distribution of the mean is normally distributed regardless the sample size.

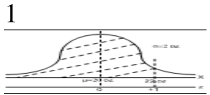
C Look at the Central Limit Theorem, only c) is correct.

29. What is the probability that the sample mean length time to find a parking spot is more than 10 minutes and less than 15 minutes?
    a). 0.9772 b). 0.3413 c). 0.5228 d). 0.4772

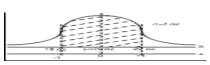
D 
The standard deviation of the mean is 2.5, not 5. z*=(10-15)/2.5=-2,

$P(10<Xbar<15)=P(-2<Z<0)=0.4772$

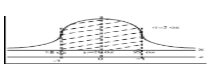
30. What is the z score for the sample mean length time to find a parking spot to be less than 20 minutes?
    a). 1 b). 0.9772 c).2 d). 0.8413

C 
z* = (20-15)/2.5 = 2

31. What are the number of minutes X and Y such that 90% of all the sample mean length times to find parking spots are between X and Y?
    a). -1.645 and + 1.645 b). 6.7757 and 23.2243
    c). 10.8879 and 19.1121 d). 11.7961 and 18.2039

C 
z* =±1.645 for the two limits. $\bar{x} = 15 \pm 1.645 \times 2.5$ =(10.888, 19.112)

32. What are the Excel functions to compute the minutes X and Y such that 95% of all the sample mean length times to find parking spots are between X and Y?
    a). =NORMINV(0.5-0.95/2,15, 2.5) and =NORMINV(0.5+0.95/2,15, 2.5)
    b). =NORMINV(0.95,15,5) and =NORMINV(1-0.95,15,5)
    c). =NORMINV(0.95,15,2.5) and =NORMINV(0.95,15,2.5)
    d). =NORMINV(0.5-0.95/2,15, 5) and =NORMINV(0.5+0.95/2,15, 5)

A 
use 2.5 as SE of the mean, and 0.5±0.95/2 as the probabilities as given on the graph.

<u>The following problem is associated with questions 33 to 40:</u>
Although most people believe breakfast is the most important meal of the day, 25% of adults skip breakfast. Assume the population proportion is p = 0.25 and p_s is the sample proportion of adults who skip breakfast based on a sample of 400 adults.

33. What are the mean μ_{ps} and the standard error σ_{ps} of the sample proportion p_s?
    a). 0.25 and $p(1-p)/\sqrt{n} = 0.825$ b). 0.25 and $\sqrt{p(1-p)/n} = 0.0217$

b). np=100 and $\sqrt{p(1-p)}/n = 0.0011$ d). np=100 and $\sqrt{np(1-p)} = \sqrt{75}$

B the properties of the sampling distribution of the proportion provide that $E(p_s) = p$ and $\sigma_p = \sqrt{\frac{p(1-p)}{n}}$

34. Which statement is correct about the sampling distribution of the proportion p_s?

I. It is normal distributed because both np and n(1-p) are greater than or equal to 5.

II. It is binomial distributed because a person in the sample either skips or has breakfast.

III. σ_{ps} decreases when the sample proportion p_s is constant and the sample size increases.

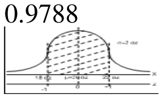
IV. σ_{ps} increases when the sample size n is constant and the sample proportion p_s increases.

a). I and II only b). II and IV only c). I and III only d).III and IV only

C the sampling distribution of proportion ps is not binomial even though each person is, thus II is not correct.

The $\sigma_p = \sqrt{p_s(1-p_s)/n}$ becomes large if ps is closer to 0.5, thus IV is not correct when ps>0.5.

35. What is the probability that the sample proportion p_s will be between 0.20 and 0.30?
a). 0.9788 b). 0.4894 c). 0.0212 d). 0.5106

A 

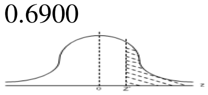
$$z_1 = (0.20 - 0.25)/\sqrt{0.25(1-0.25)/400} = -2.3094$$

$$z_2 = (0.30 - 0.25)/\sqrt{0.25(1-0.25)/400} = +2.3094$$

36. What is the Excel function to computer the probability that the sample proportion p_s will be between 0.22 and 0.28?
a). =NORMDIST(0.22,0.25,0.0217,TRUE)-NORMDIST(0.28,0.25,0.0217,TRUE)
b). =NORMINV(0.28,0.25,0.0217,TRUE)-NORMINV(0.22,0.25,0.0217,TRUE)
c). =NORMSDIST(0.28,0.25,0.0217,TRUE)-NORMSDIST(0.22,0.25,0.0217,TRUE)
d). =NORMDIST(0.28,0.25,0.0217,TRUE)-NORMDIST(0.22,0.25,0.0217,TRUE)

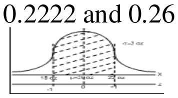
D the ps are 0.28 and 0.22 and SE of ps is 0.0217.

37. What is the probability that the sample proportion p_s will be greater than 0.265?
a). 0.6900 b). 0.2447 c). 0.7553 d). 0.7447

B 

z*=(0.265-0.25)/0.0217=0.690

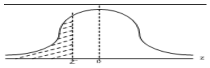
38. What are the limits p_x and p_y of the sample proportions such that 90% of all sample proportions will be between these two limits?
a). 0.2222 and 0.2778 b). 0.2143 and 0.2857
c). 0.2222 and 0.2614 d). –1.645 and +1.645

B 

z*=±1.645, $p_{xy} = 0.25 \pm 1.645 \times 0.0217$ = (0.2143, 0.2857)

39. What is the sample proportion p_z such that there is a 0.25 probability that a sample proportion p_s is less than p_z?

a). -0.6745 b). 0.2646 c). 0.2354 d). 0.225

C  z*=-0.6745 and pz = ps – (z) (0.0217) = 0.2354

40. Which statement is correct about the sample mean $\bar{x}$ as a point estimator for the population mean?

I. The expected value of the sample mean equals to the population mean.

II. The standard deviation of the sample mean is smaller than the sample standard deviation of the sample median.

III. The sample mean tends to become closer to the population mean as the sample size increases.

a). I and II only b). I and III only c). II and III only d). I, II and III

D the option I states the unbiased ness of the sample mean to be used to eatimate the population mean $E(\bar{x}) = \mu$,

the option II states $\sigma_{\bar{x}} < \sigma_{md}$ or the sample mean is more efficient than the sample median when used to estimate the population mean.

The option III states $\sigma_{\bar{x}} = \left(\frac{\sigma}{\sqrt{n}}\right)$, or as the sample size n increases, the standard error of the mean

$\sigma_{\bar{x}}$ decreases, and thus the sample mean $\bar{x}$ is moving closer to the population mean μ.