CEP 932 Quantitative Methods in Educational Research I
Summer 2010
Homework #3

(Total: 45 points)
Due: July 12, 2010 12:00 AM Eastern Time

1. (6 points) Assume that a set of scores for a sample of 200 students on a mathematics assessment is normally distributed with a mean of 60 and a standard deviation of 11.

$$z = \frac{y-\mu}{\sigma}$$

a. What is the z score corresponding to the raw score of 75?
z= 75-60/11
z=15/11
z=1.36

b. What is the z score corresponding to the raw score of 59?
z= 59-60/11
z= -1/11= -0.09

c. What is the z score corresponding to the raw score of 62?
z=62-60/11
z=2/11=0.18

d. What proportion of students in the population had scores between the values of 58 and 73?
z=58-60/11= -2/11=-.018
z=73-60/11=13/11=1.18
convert both numbers to the area under the curve using the z-table:

$$P(-0.18 \le z \le 1.18) \Rightarrow 0.8818 - 0.4286 = .4532 \approx 45/100$$

The proportion of students in the population who had scores between the values of 58 and 73 is 45/100 (or 90/200).

e. What proportion of students in the population had scores between the values of 65 and 72?
z=65-60/11=5/11= 0.45
z= 72-60/11=12/11=1.09
convert both numbers to the area under the curve using the z-table:

$$P(0.45 \le z \le 1.09) \Rightarrow 0.7357 - 0.6736 = .0621 \approx 6\%$$

The proportion of the students in the population had scores between values of 65 and 75 is 6/100 or 12/200.

f. What proportion of students in the population had scores between the values of
34 and 52?
z=34-60/11= -2.36
z=52-60=-8/11=-0.727
convert both numbers to the area under the curve using the z-table:

$$P(-2.36 \leq z \leq -0.727) \Rightarrow 0.2358 - .0094 = .2264 \approx 23\,\%$$

The proportion of students in the population who had scores between the values of 34 and 52 is 23/100 (or 42/200).

2. (3 points) The distribution of IQ scores of a large group of high school students in a highly selective private school is normal with $\mu = 110$ and $\sigma = 15$. Which of the following are true?

My process: I converted each IQ to a z-score and then used the z-table to find standar nomal curve areas discussed below to consider the percentage of each.
Area=percentage=probability

a. About 10% of the students will have an IQ over 140.
No, about 1 % of the students will have an IQ over 140.
b. Probably fewer than 5% of the students will have an IQ below 100.
No, about 50 % will have an IQ below 100 (considering the mean and the standard deviation, this makes sense. Quite close to the central tendency).
c. At least half of the students can be expected to have an IQ of less than 120.
Yes, this is true. It is actually more like 90% of the students will have an IQ of less than 120.

3. (5 points) Random samples of size 15, 25, and 50 are drawn from a single population that has a mean $\mu = 100$ and standard deviation $\sigma = 12$.

a. (1 point) Give the mean of the sampling distribution of *y* for each of the sample sizes.
For each of the sample sizes of infinite samples of those sizes, the mean would always be 100. The means of the means would always hover around 100. This is at the core of the Central Limit Theorem.

b. (3 points) Give the standard deviation of the sampling distribution of *y* for each of the sample sizes.

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_x = 3.098 = \frac{12}{\sqrt{15}}$$ (sample size of 15)

$$\sigma_x = 2.4 = \frac{12}{\sqrt{25}}$$ (sample size of 25)

$$\sigma_x = 1.697 = \frac{12}{\sqrt{50}}$$ (sample size of 50)

c. (1 point) Based on the results obtained in parts (a) and (b), what do you conclude about the accuracy of using the sample mean *y* as an estimate of population mean μ?

Considering the mean of the means for an infinite number of samples remains the same, and considering that as the sample size increases, the standard deviation decreases (as we saw in the computations of the standard error the sample size 15, 25, and 50). This tells using the sample mean as the estimate of the population mean is accurate.

4. (6 points) For each of the following, find the best estimate of the mean *y* and the standard error of the mean *y* based on a random sample of size *n* drawn from a population with mean μ and standard deviation σ .

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

$$\mu_y = \mu$$

a. n = 30, μ = 10, σ = 10

μ_y =10

$$\sigma_y = \frac{10}{\sqrt{30}} = 1.83$$

b. n = 70, μ = 10, σ = 10

μ_y =10

$$\sigma_y = \frac{10}{\sqrt{70}}$$ = 1.19

c. n = 80, μ = 10, σ = 20

μ_y =10

$$\sigma_y = \frac{20}{\sqrt{10}}$$ = 2.24

d. n = 50, μ = 10, σ = 20

μ_y =10

$$\sigma_y = \frac{20}{\sqrt{50}} = 2.82$$

5. (4 points) A study of the effect of family SES (socio-economic status) level on the students' performance in a high school produced the following contingency table. (6 points)

SES Performance	Low	Middle	High	Total
Poor	11	6	2	19
Middle	16	20	8	44
Good	3	17	17	37
Total	30	43	27	100

Suppose a student is randomly selected from this population.

a. What is the probability that the student is from low SES family and does not perform poorly (middle or good)?
Middle + Good performances/Total # of students

16+3/100=19/100=.19 or 19%

b. Given the student's performance is good, what is the probability that the student comes from a high SES family?
good performances from High SES/Total Good performances
17/37=0.46 =46%

c. Given the student is from a middle SES family, what is the probability that the student performs poorly?

poor perfromance of Middle SES family/Total # of low SES

6/43=0.13 =13%

d. What is the probability that the student is from low SES family or the student performs poorly in his class?

Low SES = 30/100=.3
Performs poorly= 19/100=.19
Both low SES AND performs poorly = 11/100=.11

Add the probabilities of the two conditions and then subtract the intersection to discover the probability that the student is from low SES or performs poorly=

.3 + .19 -.11=.38

6. (4 points) You are given the following z-scores. You would like to convert them to values that correspond to the original raw scale that have a mean equal to 75 and a standard deviation equal to 5. What are the raw values for each of the following z scores?

$$z = \frac{y - \mu}{\sigma}$$

So we are going to use algebra to solve for the raw score=

a. 2.40

$$2.40 = y - 75 / 5$$

y=87

b. 0.27
0.27=y-75/5
y=76.35

c. 0.00

0.00=y-75/5
y=75

d. -0.55

-0.55= y- 75/5
y=72.25

7. (3 points) Let *z* be a standard normal variable. Calculate each of the following probabilities:

Use the z-table to convert each z value to the area of the standard normal curve, then find the area between those two values.

a. P(-1.96 ≤ z ≤ 2.90) = .9731

b. P(-2.40 ≤ z ≤ 0.85) = .7491

c. P(-1.50 ≤ z ≤ 1.50) = .8664

8. (2 points) A math exam is given to a sample of students. The instructor determined that the exam had a mean of 80, with a standard deviation of 5 and that the data is normally

distributed.

a. What percentage of students scored below 80?

Since 80 is the mean, we know that 50% of the students score below 80.
(We could also compute the z-value, which would be 0. According to the z-table, the area is then .50 which translates to 50% of the students scoring below 80).

b. What percentage of students scored above 70?
70-80/5=-10/5=-2 leads to an area of .228

Since we are concerned with the area to the right of this point (rather than the left which would indicate the percentage of students scoring below 70), we subract .228 from 1 (the area of the whole curve equals 1). This lets us know that 77% of the studetns scored above a 70. This is also confirmed once we consider the mean and that a score of 70 is two standard deviations from the mean.

9. (1 point) Which of the following normal distributions will the least proportion of the area lie above a score of 100?
where s = standard deviation.

For each of these normal distributions, I calculated a z-score with the score value of 100, and then compared the percentages. (I wasn't sure if each of these needed to be shown in terms of the computation: please let me know in the future...it is getting tedious typing out all the equations.).

(a) mean = 90, s = 25;
Percentage lying above 100=1- 0.665 = .335
(b) mean = 50, s = 100
Percentage lying above 100=1- 0.5= .5
(c) mean = 90, s = 20
Percentage lying above 100=1- 0.665 = .335
(d) mean = 60, s = 40Percentage=.814
Percentage lying above 100=1- .814 = .186

Therefore, since answer (d) gives us the smallest percentage lying above 100, we know d to be the correct answer to the question.

10. (1 point) If a set of scores has a standard deviation of 8 and mean of 50, then a z-score of 1 would represent a score of….

1=x-50/8
x-58
B is the correct answer

(a) 54
C(b) 58

(c) 46
(d) 42

Professor Johnson assigns letter grades in a large class using the following criteria:
Top 10% = A
Next 20% = B
Next 50% = C
Next 20% = D or F
The points are normally distributed with a mean = 500 and a standard deviation, s = 100.

11. (1 point) John had 585 points, hence his grade would be

(a) A
(b) B
(c) C
(d) D or F

585-500/100=.85
z-table gives area as .80203, which would put John in the B range.

12. (1 point) Alex said "A couple of more points and I would have had a B." Which of these scores could be Alex's?

(a) 549
(b) 505
(c) 576
(d) 590

.700=x-500/100
x=507

507 is the raw score for a B, according to the situation describe above.

Therefore, 505 would put Alex just a few points away from a B. To test this, I figured out the z-score for a score of 505, which is .005, which translates in an area of .6915, a hair shy of the area that includes the B, putting him into the C grade.

13. (8 points) Assume that the following data represent a population with mean $\mu = 8$ and standard deviation $\sigma = 2.67$. The values of X are as follows:

2 4 4 6 6 6 8 8 8 8 10 10 10 12 12 14

i. Create a histogram of the distribution using SPSS (2 points)

FOR i, ii, and iii, I've attached the exported histogram document from SPSS rather than embedding it here.

ii. Convert the distribution to a distribution of $X - \mu$ and re-plot the results using SPSS. (2 points)

-6 -4 -4 -2 -2 -2 0 0 0 0 2 2 2 4 4 6

iii. Convert the distribution to a distribution of z-scores and re-plot the results using SPSS. (2 points)

-2.25 -1.50 -1.50 -0.75 -0.75 -0.75 0 0 0 0 0.75 0.75 0.75 1.5 1.5 2.25

iv. What do you notice about the distribution of X, X - μ, and the z-scores? Explain your answer. (2 points)

What I noticed is that the graphs, at first glance, look about the same. The width of each value in the histogram varied slightly, but the distribution remained the same. What I take away from this is that the concept of converting values into the standard normal curve through the z-score formula is a valid way to get to probability.