NORTHWESTERN UNIVERSITY

SCHOOL OF CONTINUING STUDIES

Handout: Problem Set #3 Solutions

PREDICT 401: Introduction to Statistical Analysis

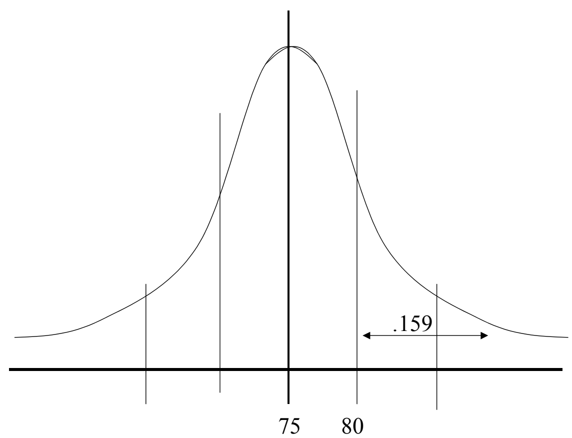
These problem sets are meant to allow you to practice and check the accuracy of your work. Please do not review the solutions until you have finalized your work. Although these problem sets are not submitted and graded, treat them as if they were. It is to your great benefit to work on and even struggle with the problem sets. Looking at the solutions before finalizing your work will, quite simply, make for a less meaningful learning experience.

1. A manufacturing survey reveals that an average worker produces an item in 75 seconds, with a standard deviation of 5 seconds.

a. What is the probability that a worker chosen at random will take 80 or more seconds to make an item?

$z = (x - \mu) / \sigma = (80-75) / 5 = 1.00$

The table says that 1.00 corresponds to .159. Thus, the probability that it will take 80 or more seconds = .159. Or, 15.9% of the normal curve falls to the right of 1.00 std devs.



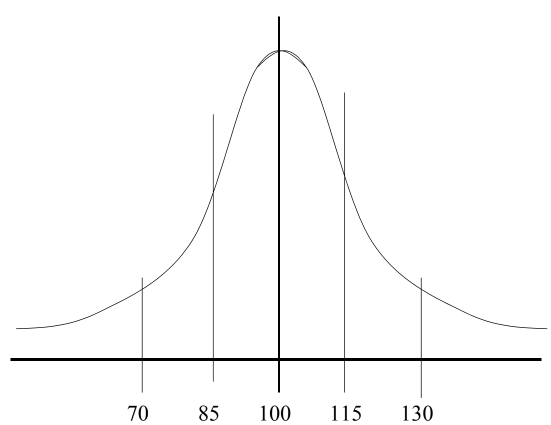
b. What is the probability that it will take a worker chosen at random between 75 and 80 seconds?

Well, we know that the right half of the curve = .500. So, the area between the mean and 80 = .500 - .159 = .341.

c. What is the probability that it will take 70 of fewer seconds?

The curve is symmetrical, so it is the same as taking 80 or more → .159.
Or, more formally,
$z = 70-75 / 5 = -1.00$.
Since z is negative, look up 1.00 (it = .159). This is the area under the curve to the left of 70 seconds.

2. Assume that for all Americans, IQ scores are normally distributed with a mean of 100 ($\mu = 100$) and a standard deviation of 15 ($\sigma = 15$).



a. What is the probability that a person selected at random will have an IQ below 90?

z = (90-100) / 15 = -10 / 15 = -0.67.
The table shows us a corresponding probability of .251. So, the probability that a person selected at random will have an IQ below 90 is .251 (or 25.1%).

b. What is the probability that a person selected at random will have an IQ between 90 and 120?

Let's break this up into 3 steps.

1) Find the probability it falls between 90 and 100.
Since the entire left hand part of the curve is .500, and since the amount to the left of 90 = .251 (found above), then the amount between 90 and 100 is .500-.251 = .249.

2) Find the probability it falls between 100 and 120.
z = (120-100) / 15 = 20/15 = 1.33
The corresponding probability is .092. Remember, though, this is the percentage falling <u>to the right</u> of 120. So, the amount between the mean and 120 = .500 - .092 = .408.

3) Combine them: The probability it falls between 90 and 120 = the probability it falls between 90 and 100 plus the probability it falls between 100 and 120 = .249+.408 = .657

c. What is the cutoff point for the top 5% of Americans? (In other words, what is the lowest IQ score someone can have and still be in the top 5%?)

Here, we have to work backwards. The table shows us that 5% (.05) corresponds to a z of about 1.64 (actually, .051 corresponds to this, but nothing fits exactly). So, what is the value of "x" when z = 1.64?

$z = (x - \mu) / \sigma$
$1.64 = (x - 100) / 15.$
Solve : x = 124.6

Answer: 5% of Americans have a 124.6 IQ or above.

3. You are an inspector for Coke. Your current job requires you to examine the volume of their 2-liter bottles. The standard deviation for Coke's 2-liter bottles is .05 liters. You take a random sample of 100 bottles and get a sample mean of 1.99 liters. What is the 90% confidence interval estimate of the true mean of all of Coke's 2-liter bottles?

Since we know the population standard deviation, we can use the formula for the confidence with the z-distribution.

$\bar{x} - (z_{\alpha/2} * \sigma / \sqrt{n}) \leq \mu \leq \bar{x} + (z_{\alpha/2} * \sigma / \sqrt{n})$
$1.99 - (1.65 * .05 / \sqrt{100}) \leq \mu \leq 1.99 + (1.65 * .05 / \sqrt{100})$
$1.982 \leq \mu \leq 1.998$
We are 90% confident that the true mean of the population of 2-liter bottles is between 1.982 and 1.998 liters.

4. The designers of an SAT math review course claim that their enrollees have higher SAT math scores than the average student accepted at Harvard. You randomly select 25 enrollees from and find, on average, their math SAT score is 2 points higher than the Harvard average. You know from previously collected data that the standard deviation for all enrollees is 10 points.

a. What is the 95% confidence interval of the enrollees' score (relative to Harvard's average)?

Again, we know the population standard deviation, so:

$\bar{x} - (z_{\alpha/2} * \sigma / \sqrt{n}) \leq \mu \leq \bar{x} + (z_{\alpha/2} * \sigma / \sqrt{n})$

2 – (1.96 * 10 / √25) ≤ μ ≤ 2 + (1.96 * 10 / √25)
2 - (3.92) ≤ μ ≤ 2 + (3.92)
-1.92 ≤ μ ≤ 5.92
We are 95% confident that the true mean of the population is between 1.92 less and 5.92 more than Harvard's average.

b. If Harvard's average math SAT score is 750, what is the 95% confidence interval of the enrollees' score?

750 – 1.92 = 748.08
750 + 5.92 = 755.92.
We are 95% confident that the true mean of the population is between 748.08 and 755.92.

c. What do you think of the SAT course designers' claim?

Not much. After all, our 95% confidence interval contains "0," meaning that there might be no real difference between enrollees and Harvard's students. The difference might even be negative, according to our confidence interval.

5. Nine months ago, the director of a local health clinic implemented a new prenatal outreach program designed to increase the number of prenatal clinic visits among pregnant women served by the clinic. Before the program, the average number of prenatal visits for women was 2.9. The director wants to evaluate the success of the new program, so she obtained data on 34 women who delivered babies in the most recent month. The distribution of number of visits for this group was:

# Prenatal Visits	% of women with # of visits
1	11.8%
2	11.8%
3	23.5%
4	29.4%
5	17.6%
6	5.9%

a. Find the sample average and sample standard deviation of the number of visits for the 34 women.

Mean = Σ x *p(x) = .118*1 + .118*2 + .235*3 + .294*4 + .176*5 + .059*6 = 3.47.

Std deviation = √[(Σ (x-$\bar{\bar{x}}$)^2 * p(x)] √[.118*(1 - 3.47)^2 + .118*(2 - 3.47)^2 + .235*(3 - 3.47)^2 +.294*(4-3.47)^2 +.176*(5-3.47)^2 + .059*(6-3.47)^2] = 1.40.

b. Find the 90% confidence interval for the mean number of visits for women under the new program. In one sentence, what does this interval mean? Find the 95% confidence interval.

Since we don't know the population standard deviation, we use the t-distribution. (Since it is 34 people, it is nearly equivalent to the z-distribution.)

$$\bar{x} - t_{(\frac{\alpha}{2}, n-1)} * (\frac{S}{\sqrt{n}}) \le \mu \le \bar{x} + t_{(\frac{\alpha}{2}, n-1)} * (\frac{S}{\sqrt{n}})$$

3.47 – (1.70*1.4 / √34) ≤ μ ≤ 3.47 + 1.70 *1.4/√34)
3.47 - .41 ≤ μ ≤ 3.47 + .41
3.06 ≤ μ ≤ 3.88
We are 90% confident that the mean number of visits for all women in this outreach program ranges from 3.06 to 3.88.

For the 95% CI, same equation, except that $t_{.05/2}$ = 1.96
3.47 – (2.03*1.4 / √34) ≤ μ ≤ 3.47 + 2.03 *1.4/√34)

$3.47 - .48 \leq \mu \leq 3.47 + .48$
$2.99 \leq \mu \leq 3.95$
We are 95% confident that the true mean number of visits ranges from 2.99 to 3.95.

c. Based on your calculations above, is the program effective at increasing the number of visits? Explain.

Before the program, the mean number of visits was 2.90. For both the 90% and 95% confidence intervals, even the lowest bound is higher than 2.90. Thus, we are pretty confident that the program is resulting in more visits.