STAT40720 Intro. to Data Analytics

Assignment 2

Name	Kevin O'Connell
Student ID	#########
Submission Date	19th. October 2015

Question 1

Statement **(a)** is more probable. The standard deviation of the sample distribution of x̄ (called the standard error, $\sigma_{\bar{x}}$) is inversely proportional to the square root of the sample size. This means a smaller sample size will produce a larger standard error, i.e. a wider sample distribution, increasing the probability a particular random sample will have a sample mean x̄ lying away from the population mean.

Question 2

(a) Find P(3 yrs. < x < 4 yrs.)

The NCST tables for the Normal Distribution shows values only for the positive side of the mean. The Normal Distribution is symmetric about the mean, yielding:

$$\begin{aligned} P(3\,yrs. < x < 4\,yrs.) &= P(3\,yrs. < x < 3.25\,yrs.) + \\ &\qquad P(3.25\,yrs. < x < 4\,yrs.) \\ &= P(3.25\,yrs. < x < 3.5\,yrs.) + \\ &\qquad P(3.25\,yrs. < x < 4\,yrs.) \end{aligned}$$

Now, converting the above values for x into normalised z values.

$$\begin{aligned} 3.25\,yrs &= 0 \\ 3.5\,yrs &= \tfrac{3.5-3.25}{0.5} = \tfrac{0.25}{0.5} = 0.5 \\ 4\,yrs &= \tfrac{4-3.25}{0.5} = \tfrac{0.75}{0.5} = 1.5 \end{aligned}$$

This yields a normalised representation of the probability interval equivalent to:

$$\begin{aligned} P(3\,yrs. < x < 4\,yrs.) &= P(z < 0.5) - P(z < 0.0) + \\ &\qquad P(z < 1.5) - P(z < 0.0) \\ &= (0.6915 - 0.5) + (0.9332 - 0.5) \\ &= 0.6247 \end{aligned}$$

(b) P(3 successes) given B(4, 0.6247)

$$\begin{aligned} P(3\,successes) &= \tbinom{4}{3} \times 0.6247^3 \times (1 - 0.6247)^{4-3} \\ &= \tfrac{4!}{3!(4-3)!} \times 0.6247^3 \times 0.3753 \\ &= 4 \times 0.2438 \times 0.3753 \\ &= 0.366 \end{aligned}$$

(c) P(3 yrs. < x̄ < 4 yrs.)

The standard error, $\sigma_{\bar{x}}$, of the sample distribution of x̄ is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.5}{\sqrt{4}} = \frac{0.5}{2} = 0.25$$

$$\mu_{\bar{x}} = \mu$$

Using the equivalences calculated in (a) above, the new normalised z values are:

$$3.25\ yrs = 0$$

$$3.5\ yrs = \frac{3.5-3.25}{0.25} = \frac{0.25}{0.25} = 1$$

$$4\ yrs = \frac{4-3.25}{0.25} = \frac{0.75}{0.25} = 3$$

Now, the probablility the sample mean lies in the interval 3 years to 4 years is given by:

$$\begin{aligned} P(3\ yrs. < \bar{x} < 4\ yrs.) &= P(z < 1) - P(z < 0.0) + \\ &\qquad P(z < 3) - P(z < 0.0) \\ &= (0.8413 - 0.5) + (0.99865 - 0.5) \\ &= 0.83995 \end{aligned}$$

Question 3

(a) Find the 95% CI

Since the sample size is large (N>30), and in lieu of the true population standard deviation, the sample standard deviation can be used to construct the 95% Confidence Interval of the population mean, given the sample mean.

$$\alpha = \frac{1-0.95}{2} = 0.025 = 2.5\%$$

Using the NCST tables, the z-critical value is:

$$z - critical_{\alpha=2.5\%} = 1.96$$

Now, the 95% Confidence Interval for the population mean of concurrent gym users is given by:

$$\begin{aligned} 95\%CI &= \bar{x} \pm 1.96 \times \frac{s}{\sqrt{n}} \\ &= 47.7 \pm 1.96 \times \frac{9.2}{\sqrt{120}} \\ &= 47.7 \pm 1.64609 \\ &= (46.0539, 49.3461) \end{aligned}$$

(b) Comment on 90% CI

Given the z-critical value for the 90% Confidence Interval is lower than of the 95% CI, this would yield a narrower range.

However, this question can be answered more intuitively by realising that estimating the location of the population mean to a lower confidence level, allows you to place the outer bounds closer to the sample mean.