

Stat 400, section 5.4 supplement: The Central Limit Theorem

Solutions to homework exercises

First answer the questions for yourself.

Second, use the summary of answers (part A) to check yourself. If you did not get the same result, see whether you can find the error.

Only as a final resort, go to the solutions worked out (part B) to determine what went wrong.

A. Homework Exercises solutions summary (answers only)

These answers were calculated using exact values and the exact value answers are given. Decimal approximations are given in part B below. If you used rounded-off decimal approximations to do calculations, your answers may differ in the third or fourth decimal place.

1.a) 1509; b) 1509, 33.9; c) 0.0409; d) 0.9554; e) 0.0037

2.a) 2.3, $\sqrt{1.41}$; b) 2.3, $\sqrt{\frac{1.41}{50}}$; c) 0.2743; d) 0.6087

3.a) 0.5, 0.5; b) 0.5, 0.05; c) 0.8185; d) 0.0228

4.a) $-\frac{1}{3}$, $\frac{\sqrt{11}}{3}$; b) $-\frac{1}{3}$, $\frac{\sqrt{11}}{30}$; c) 0.0013; d) $1.28 \left(\frac{\sqrt{11}}{30} \right) - \frac{1}{3}$

5.a) $e - 1$, $\sqrt{-\frac{e^2}{2} + 2e - \frac{3}{2}}$; b) 0.5423; c) $e - 1$, $\sqrt{\frac{-e^2 + 2e - \frac{3}{2}}{30}}$; d) 0.5398; e) $1.645 \left(\sqrt{\frac{-e^2 + 2e - \frac{3}{2}}{30}} \right) + e - 1$

See notes on the answers to parts b) and d).

6.a) 3.5, $\sqrt{6.75}$; b) 3.5, $0.01\sqrt{6.75}$; c) 0.0274

7.a) 45; b) $e^{-\frac{47}{45}}$; c) 0.0793; d) $-1.28 \left(\frac{45}{\sqrt{1000}} \right) + 45$

B. Homework Exercises solutions worked out

1. In 2010, the mean score for the three sections of the SAT was $\mu = 1509$, with a standard deviation $\sigma = 339$. (source: *College Entrance Examination Board, College-Bound Seniors: Total Group Profile (National) Report, 1966-67 through 2009-10*)

a) One student takes the SAT. What is her expected score (on the three parts)?

The scores of students taking the SAT exhibit some variability—we know this because the standard deviation does not equal 0. While we cannot say with absolute certainty that this student *will have* any specific score, the *expected value* of her score is the population mean 1509.

b) One hundred students take the SAT. What is the expected value of their average score (i.e. expected value of the sample mean)? What is the standard error for the sampling distribution?

By the Central Limit Theorem, for sample size $n = 100$,

$$\text{expected value } E(\bar{X}) = \mu_X = 1509 \text{ and standard error } = \frac{\sigma_X}{\sqrt{n}} = \frac{339}{\sqrt{100}} = 33.9.$$

c) One hundred students take the SAT. What is the probability that their average score is less than 1450?

$$Z = \frac{\bar{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \Rightarrow \frac{1450 - 1509}{\frac{339}{\sqrt{100}}} \approx -1.74, P(\bar{X} < 1450) = P(Z < -1.74) = 0.0409$$

d) One hundred students take the SAT. What is the probability that their average score is between 1450 and 1600?

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow \frac{1600 - 1509}{339 / \sqrt{100}} \approx 2.68$$

$$P(1450 < \bar{X} < 1600) = P(Z < 2.68) - P(Z < -1.74) = 0.9963 - 0.0409 = 0.9554$$

e) One hundred students take the SAT. What is the probability that their average score is above 1600?

$$P(\bar{X} > 1600) = 1 - P(\bar{X} < 1600) = 1 - P(Z < 2.68) = 1 - 0.9963 = 0.0037$$

2. An experiment has the following probability distribution:

X	1	2	3	4
$P(X)$	0.4	0.1	0.3	0.2

a) What are the expected value and standard deviation for a single randomly-chosen value of X ?

$$E(X) = \mu_X = 1(0.4) + 2(0.1) + 3(0.3) + 4(0.2) = 0.4 + 0.2 + 0.9 + 0.8 = 2.3$$

$$\text{Var}(X) = (1 - 2.3)^2 (0.4) + (2 - 2.3)^2 (0.1) + (3 - 2.3)^2 (0.3) + (4 - 2.3)^2 (0.2) = 1.41$$

$$\sigma_X = \sqrt{1.41} \approx 1.187$$

b) You randomly select a sample of size $n = 50$. What is the expected value for the sample mean? What is the standard error for the sampling distribution?

By the Central Limit Theorem, for $n = 50$,

$$\text{expected value } E(\bar{X}) = \mu_X = 2.3 \text{ and standard error } \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{\sqrt{1.41}}{\sqrt{50}} \approx 0.16793$$

c) You randomly select a sample of size $n = 50$. What is the probability that the sample mean is less than 2.2?

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow \frac{2.2 - 2.3}{\sqrt{\frac{1.41}{50}}} \approx -0.60, P(\bar{X} \leq 2.2) = P(Z \leq -0.60) = 0.2743$$

d) You randomly select a sample of size $n = 50$. What is the probability that the sample mean is between 2.2 and 2.5?

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow \frac{2.5 - 2.3}{\sqrt{\frac{1.41}{50}}} \approx 1.19$$

$$P(2.2 < \bar{X} < 2.5) = P(Z < 1.19) - P(Z < -0.60) = 0.8830 - 0.2743 = 0.6087$$

3. You flip a coin 100 times. Define a random variable $X = 0$ for tails and $X = 1$ for heads.

a) What are the expected value and standard deviation of X for a single toss?

$$E(X) = \mu_X = 0(0.5) + 1(0.5) = 0.5 = 50\%$$

$$\text{Var}(X) = (0 - 0.5)^2(0.5) + (1 - 0.5)^2(0.5) = 0.25 \quad \sigma_X = \sqrt{0.25} = 0.5$$

b) What are the expected value and standard error for the sampling distribution?

By the Central Limit Theorem, for $n = 100$,

$$\text{expected value } E(\bar{X}) = \mu_X = 0.5 \text{ and standard error } = \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{0.5}{\sqrt{100}} = 0.05$$

c) What is the probability that the average (mean) value for X is between 0.45 and 0.60?

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow \frac{0.45 - 0.5}{0.5 / \sqrt{100}} = -1; \quad Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow \frac{0.60 - 0.5}{0.5 / \sqrt{100}} = 2$$

$$P(0.45 < \bar{X} < 0.60) = P(Z < 2) - P(Z < -1) = 0.9772 - 0.1587 = 0.8185$$

d) What is the probability that the average (mean) value for X is greater than 0.60?

$$P(\bar{X} > 0.60) = P(Z > 2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$$

4. You pay \$1 to play a game in which you roll one standard six-sided die. You lose your dollar if the die is 1, 2, 3 or 4. You get your dollar back if the die is a 5, and if the die is a 6 you get your dollar back plus \$2 more (total of \$3).

a) Calculate expected value and standard deviation for a single toss. (Be sure to include the dollar you pay to play the game.)

The game has the following probability distribution.

X	-1	0	2
$P(X = x)$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X) = \mu_X = -1\left(\frac{2}{3}\right) + 0\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) = -\frac{1}{3}$$

$$\text{Var}(X) = \left(-1 - \left[-\frac{1}{3}\right]\right)^2 \left(\frac{2}{3}\right) + \left(0 - \left[-\frac{1}{3}\right]\right)^2 \left(\frac{1}{6}\right) + \left(2 - \left[-\frac{1}{3}\right]\right)^2 \left(\frac{1}{6}\right) = \frac{11}{9} \approx 1.2222 \quad \sigma_X = \sqrt{\frac{11}{9}} = \frac{\sqrt{11}}{3} \approx 1.1055$$

b) If you play the game 100 times, what are the expected value and standard error for the sampling distribution?

By the Central Limit Theorem, for $n = 100$,

$$\text{expected value } E(\bar{X}) = \mu_X = -\frac{1}{3} \text{ and standard error } = \frac{\sigma_X}{\sqrt{n}} = \frac{\sqrt{11/9}}{\sqrt{100}} = \frac{\sqrt{11}}{30} \approx 0.1106$$

c) If you play the game 100 times, what is the probability that your average outcome will be positive? (That is, you walk away with more money than what you had before the game.)

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} = \frac{0 - \left[-\frac{1}{3}\right]}{\sqrt{11}/30} \approx 3.02, \quad P(\bar{X} > 0) = P(Z > 3.02) = 1 - P(Z < 3.02) = 1 - 0.9987 = 0.0013$$

d) If you play the game 100 times, your average winnings have a 90% probability of being below what value?

From the normal distribution table, the closest we can get to 0.9000 is $0.8997 = P(Z < 1.28)$.

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow 1.28 = \frac{\bar{x} - \left[-\frac{1}{3}\right]}{\sqrt{11}/30} \Rightarrow \bar{x} = 1.28 \left(\frac{\sqrt{11}}{30}\right) - \frac{1}{3} \approx -0.19$$

5.* A random variable X has probability density function $f(x) = \frac{1}{x}$, $1 \leq x \leq e$.

a) Find the expected value and standard deviation for X .

$$E(X) = \mu_X = \int_1^e x * \frac{1}{x} dx = \int_1^e 1 dx = [x]_1^e = e - 1 \approx 1.72$$

$$\text{Var}(X) = \int_1^e x^2 * \frac{1}{x} dx - (e-1)^2 = \left[\frac{x^2}{2} \right]_1^e - (e-1)^2 = \frac{e^2}{2} - \frac{1}{2} - e^2 + 2e - 1 = -\frac{e^2}{2} + 2e - \frac{3}{2} \approx 0.2420$$

$$\sigma_X = \sqrt{-\frac{e^2}{2} + 2e - \frac{3}{2}} \approx 0.4920$$

b) What is the probability that a single randomly-chosen value of X is less than 1.72?

$$P(X < 1.72) = \int_1^{1.72} \frac{1}{x} dx = [\ln x]_1^{1.72} = \ln(1.72) - \ln(1) = \ln(1.72) \approx 0.5423$$

side note: This mean is not at the “halfway point” for probability, i.e. $P(X < \mu) \neq 0.5$. The “halfway point” value is a different statistical measure called the median, i.e. $P(X < \text{median}) = 0.5$.

c) Given a sample size of 900, calculate the expected value and standard error for the sampling distribution.

By the Central Limit Theorem, for $n = 900$, expected value $E(\bar{X}) = \mu_X = e - 1$

$$\text{and standard error} = \sigma(\bar{X}) = \frac{\sigma_X}{\sqrt{n}} = \frac{\sqrt{-\frac{e^2}{2} + 2e - \frac{3}{2}}}{\sqrt{900}} = \frac{\sqrt{-\frac{e^2}{2} + 2e - \frac{3}{2}}}{30} \approx 0.0164$$

d) Given a sample size of 900, what is the probability that the sample mean is less than 1.72?

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow \frac{1.72 - (e-1)}{\frac{\sqrt{-\frac{e^2}{2} + 2e - \frac{3}{2}}}{30}} \approx 0.10 \quad P(\bar{X} < 1.72) = P(Z < 0.10) = 0.5398$$

Important note: If we had used the approximate value $\mu = 1.72$ rather than the exact value $\mu = e - 1$, we would have gotten a different and incorrect answer: $P(\bar{X} < 1.72) = P(Z < 0) = 0.5$!

e) Given a sample size of 900, there is a 5% probability that the sample mean is above what value?

$$P(\bar{X} > \text{what value?}) = 0.05 \text{ is the same as } P(\bar{X} < \text{what value?}) = 0.95.$$

From the normal distribution table, we get $0.9495 = P(Z < 1.64)$ and $0.9505 = P(Z < 1.65)$. The convention in statistics is to “split the difference” and use $z = 1.645$.

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow 1.645 = \frac{\bar{x} - (e-1)}{\frac{\sqrt{-\frac{e^2}{2} + 2e - \frac{3}{2}}}{30}} \Rightarrow \bar{x} = 1.645 \left(\frac{\sqrt{-\frac{e^2}{2} + 2e - \frac{3}{2}}}{30} \right) + e - 1 \approx 1.745$$

6.* A certain drug is to be rated either effective or ineffective. Suppose lab results indicate that 75% of the time the drug increases the lifespan of a patient by 5 years (effective) and 25% of the time the drug causes a complication which decreases the lifespan of a patient by 1 year (ineffective). As part of a study you administer the drug to 10000 patients.

a) Find $E(X)$ and σ_X for a single patient.

$$E(X) = \mu_X = 5(0.75) + (-1)(0.25) = 3.75 - 0.25 = 3.5 \text{ years}$$

$$\text{Var}(X) = (5 - 3.5)^2 (0.75) + (-1 - 3.5)^2 (0.25) = 6.75 \quad \sigma_X = \sqrt{6.75} \approx 2.5981$$

b) Find the expected value and standard error for the sampling distribution.

By the Central Limit Theorem, for $n = 10000$, expected value $E(\bar{X}) = \mu_X = 3.5$

$$\text{and standard error} = \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{\sqrt{6.75}}{\sqrt{10000}} = \frac{\sqrt{6.75}}{100} = 0.01\sqrt{6.75} \approx 0.0260$$

c) What is the probability that the lifespans of those in the study will be increased by an average of 3.55 years or more?

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow \frac{3.55 - 3.5}{\sqrt{6.75} / \sqrt{100}} \approx 1.92$$

$$P(\bar{X} \geq 3.55) = P(Z \geq 1.92) = 1 - P(Z \leq 1.92) = 1 - 0.9726 = 0.0274$$

7.* A honeybee drone has an expected lifespan which is exponentially distributed with mean 45 days. You collect one thousand honeybee drones.

a) What is the standard deviation for the lifespan of a honeybee drone?

An exponential probability distribution has probability density function $f(x) = \lambda e^{-\lambda x}$, $[0, \infty)$ with mean = standard deviation = $\frac{1}{\lambda}$. So, the standard deviation for the lifespan of a honeybee drone = 45.

b) For an individual bee what is the probability that its lifespan will be above 47 days?

$$\begin{aligned} P(X > 47) &= \int_{47}^{\infty} \frac{1}{45} e^{-\frac{1}{45}x} dx = 1 - \int_0^{47} \frac{1}{45} e^{-\frac{1}{45}x} dx = 1 - \left[-e^{-\frac{1}{45}x} \right]_0^{47} = 1 + e^{-\frac{1}{45}(47)} - e^{-\frac{1}{45}(0)} \\ &= 1 + e^{-\frac{47}{45}} - 1 = e^{-\frac{47}{45}} \approx 0.3519 \end{aligned}$$

c) What is the probability that the average lifespan for all 1000 honeybee drones will be above 47 days?

By the Central Limit Theorem, for $n = 1000$,

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow \frac{47 - 45}{45 / \sqrt{1000}} \approx 1.41$$

$$P(\bar{X} \geq 47) = P(Z \geq 1.41) = 1 - P(Z \leq 1.41) = 1 - 0.9207 = 0.0793$$

d) There is a 10% probability that the average lifespan for all 1000 honeybee drones will be less than what age?

From the normal distribution table, the closest we can get to 0.1000 is 0.1003 = $P(Z < -1.28)$.

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n}} \Rightarrow -1.28 = \frac{\bar{x} - 45}{45 / \sqrt{1000}} \Rightarrow \bar{x} = -1.28 \left(\frac{45}{\sqrt{1000}} \right) + 45 \approx 43.18 \text{ days}$$