Brock Franson
A02052161

7.2

1a) $\hat{\mu}_1 = \frac{\mu}{2} + \frac{\mu}{2}$

$= \mu$

bias $= 0$

unbiased

$\hat{\mu}_2 = \frac{\mu}{4} + \frac{3\mu}{4}$

$= \mu$

bias $= 0$

unbiased

$\hat{\mu}_3 = \frac{\mu}{6} + \frac{3\mu}{8} + 9$

$= \frac{4\mu}{6} + 9$

biased!

1b) $Var(\hat{\mu}_1) = (\frac{1}{2})^2 \cdot 10 + (\frac{1}{2})^2 \cdot 15$

$= 6.25$

$Var(\hat{\mu}_2) = (\frac{1}{4})^2 \cdot 10 + (\frac{3}{4})^2 \cdot 15$

$= 9.0625$

$Var(\hat{\mu}_3) = (\frac{1}{6})^2 \cdot 10 + (\frac{1}{3})^2 \cdot 15$

$= 1.9444$

2a) $\hat{\mu}_1 = \frac{\mu}{3} + \frac{\mu}{3} + \frac{\mu}{3}$

$= \mu$

unbiased.

$\hat{\mu}_2 = \frac{\mu}{4} + \frac{\mu}{3} + \frac{\mu}{5}$

$= \frac{47}{60}\mu$

biased

$\hat{\mu}_3 = \frac{\mu}{6} + \frac{\mu}{3} + \frac{\mu}{4} + 2$

$= \frac{11}{4}\mu$

biased

Homework Sampling Distribution for Means

1. Textbook 7.2.1ab ✓
2. Textbook 7.2.2a ✓
3. Read sections 7.1 and 7.2. (Pay attention to the vocabulary, don't try to understand the proofs.)
4. A trucking company knows that the population mean billing time is $\mu = 19.5$ days. The population standard deviation is $\sigma = 4.2$ days. They decide to take a sample of $n = 65$ days and look at the sample mean.

   (a) What is the mean of the sampling distribution? $\mu_{\overline{X}}$

   $\mu_{\overline{X}} = \mu = 19.5$

   (b) What is the standard deviation of the sampling distribution? $\sigma_{\overline{X}}$

   $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{4.2}{\sqrt{65}} = .52$

   (c) Can we use the normal curve as an approximation of the sampling distribution? (Hint: is the population normal or the sample size at least 30?)

   yes $n = 65$, $n \geq 30$ ✓

   (d) What is the probability that they get a sample mean that is less than 18 days?

   $P(X \leq 18) = P(Z \leq -2.88) = .002$

   $Z = \frac{18 - 19.5}{.52} = -2.88$

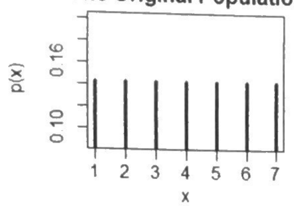
   (e) What is the probability that they get a sample mean that is greater than 21 days?

   $P(X \geq 21) = 1 - P(Z \geq 2.88) = 1 - .9980 = .002$

   $Z = \frac{21 - 19.5}{.52} = 2.88$

5. **What is a sampling distribution of the sample mean?**
Let's start with a very small population with the values 1, 2, 3, 4, 5, 6, 7.
The mean of the population is $\mu = 4$. The standard deviation of the population is $\sigma = 2$.
Here is the graph for the probability mass function, or pmf.

The Original Population



(a) Let's look at all the possible samples of size three that we could get from this population. I found most of the sample means for you. You can fill in the missing ones.

sample	sample mean $\bar{x}$	sample	sample mean $\bar{x}$	sample	sample mean $\bar{x}$
1,2,3	2	1,5,6	4	2,6,7	5
1,2,4	2.33	1,5,7	4.33	3,4,5	4
1,2,5	2.67	1,6,7	4.67	3,4,6	4.33
1,2,6	3	2,3,4	3	3,4,7	4.67
1,2,7	3.33	2,3,5	3.33	3,5,6	4.67
1,3,4	2.67	2,3,6	3.67	3,5,7	5
1,3,5	3	2,3,7	4	3,6,7	5.33
1,3,6	3.33	2,4,5	3.67	4,5,6	5
1,3,7	3.67	2,4,6	4	4,5,7	5.33
1,4,5	3.33	2,4,7	4.33	4,6,7	5.67
1,4,6	3.67	2,5,6	4.33	5,6,7	6
1,4,7	4	2,5,7	4.67		

(b) So for just the population of 1,2,3,4,5,6,7, there were 35 possible samples of size 3 that we could have drawn. Those possible samples each give us a sample mean.

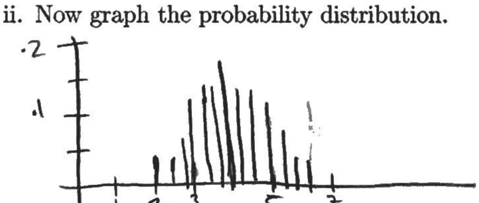
So those 35 **possible sample means** are the **sampling distribution of the sample mean.**

(c) It's hard to get an idea of what values of the sample mean are possible from the table above.

i. Let's summarize it by counting how many times we could get each sample mean. I

sample mean $\bar{x}$	Frequency	Probability	sample mean $\bar{x}$	Frequency	Probability
2	1	1/35	4.33	4	4/35
2.33	1	1/35	4.67	4	4/35
2.67	2	2/35	5	3	3/35
3	3	3/35	5.33	2	2/35
3.33	4	4/35	5.67	1	1/35
3.67	4	4/35	6	1	1/35
4	5	5/35			

ii. Now graph the probability distribution.



iii. How does the graph of the sampling distribution compare to the original population?

It looks more normal

(d) So we have a list of all the possible sample means: 2 2.33 2.67 2.67 3 3 3 3.33 3.33 3.33 3.33 3.67 3.67 3.67 3.67 4 4 4 4 4 4.33 4.33 4.33 4.33 4.67 4.67 4.67 4.67 5 5 5 5.33 5.33 5.67 6

We can find the mean and standard deviation of those values just like any other set of numbers.

i. The average, or mean, is $\mu_{\overline{X}} = 4$.

A. How does that compare to the original population mean?

original $\mu = 4$

Same

ii. The standard deviation is $\sigma_{\overline{X}} = 1.1547$.

A. How does that compare to the original population standard deviation?

original $\sigma = 2$

Smaller.

B. What do you get if you calculate $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$?

$\sigma_{\overline{X}} = \frac{2}{\sqrt{3}} = 1.1547$

(e) So even with a small population of 7 values and a very small sample size of 3, it still took me a long time to find all the possible samples and all the possible sample means. In practice we won't want to do that very often. So instead we learn the rules of what will happen every time.

If you look at the **sampling distribution of the sample mean** (all the possible sample means):

- Shape: it will always look at least a little more normal than the original population.
  - By the time you get to a sample size of at least 30, the sampling distribution will look very normal, no matter what your original population looked like.
  - So if the original population was normal, or the sample size is at least 30, we can use the normal distribution to find probabilities of getting a certain sample mean.
- Mean: the mean of all the possible sample means is the same as the population mean.

$$\mu_{\overline{X}} = \mu$$

- Standard Deviation: the standard deviation of all the possible sample means is the original population standard deviation divided by the square root of the sample size.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

6. Why does the sample size have to be at least 30?

because that is usually big enough for any distribution to normalize

7. Suppose the time X between text messages arriving on your cell phone has a normal distribution with a population mean $\mu = 25$ minutes and standard deviation $\sigma = 15$ minutes. You record the times between your **next 10** messages and you are interested in the sample mean.

   (a) What is the mean of the sampling distribution? $\mu_{\overline{X}}$

   $\mu_{\overline{X}} = 25$

   (b) What is the standard deviation of the sampling distribution? $\sigma_{\overline{X}}$

   $\sigma_{\overline{X}} = \frac{15}{\sqrt{10}} = 4.74$

   (c) Can we use the normal curve as an approximation of the sampling distribution? (Hint: is the population normal or the sample size at least 30?) yes, original was normal.

   (d) What is the probability that the sample mean is less than 28 minutes?

   $P(\overline{X} \leq 28) = P(Z \leq .63) = .7357$

   (e) What is the probability that the sample mean is between 20 and 22 minutes?

   $P(20 \leq \overline{X} \leq 22) = P(-1.05 \leq Z \leq -.63) = .1174$

   (f) What $\overline{x}$ value has 10% of the area to the right?

   $P(\overline{X} > x) = .1 \therefore P(\overline{X} < x) = .9$, $Z = 1.28$, $1.28 = \frac{\overline{X} - 25}{15/\sqrt{10}}$ $\overline{X} = 31.07$

8. **REQUIRED! Watch this Video:**

   (a) The handout for the video is at the very end of this document.

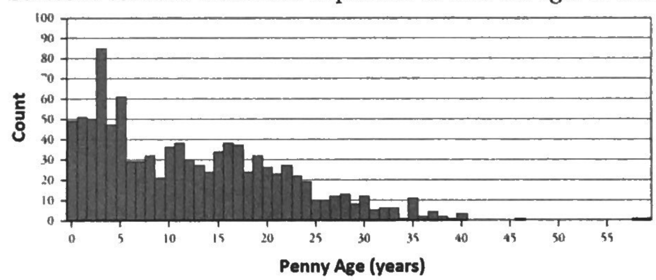
   (b) https://youtu.be/dySXByKeTk4

9. If the sample size is at least 30, which of the following is true?

   - **Wrong answer:** the population itself will be approximately normally distributed
   - **Another wrong answer:** an actual sample will be approximately normally distributed
   - **Correct answer:** the distribution of all the possible sample means will be approximately normally distributed, and then we can use the normal distribution to find probabilities associated with the sample mean

10. **Note:** As the sample size gets bigger, it does NOT affect the standard deviation of your one sample you are taking. But the distribution of ALL THE POSSIBLE SAMPLE MEANS has a standard deviation that gets smaller as the sample size gets bigger.

11. Someone counted thousands of pennies to find the ages of each pennies. This is the population.



The population mean is $\mu = 12.3$ years and the standard deviation is $\sigma = 9.6$. You decide to take a sample of 50 pennies and look at the average age.

   (a) Will the sampling distribution of the mean be normally distributed? (Hint: is the population normal or the sample size at least 30?) yes, because $50 \geq 30$ ✓

   (b) What is the average of all the possible sample means?

   $\mu_{\overline{X}} = 12.3$

(c) What is the standard deviation of all the possible sample means?

$\sigma_{\bar{x}} = \frac{2.6}{\sqrt{50}} = .36$

(d) What is the probability that the sample mean is greater than 15 years?

$P(\bar{x} > 15) = 1 - P(z < 1.99) = 1 - .9767 = .0233$

(e) What is the probability that the sample mean is between 10 and 15 years?

$P(10 < \bar{x} < 15) = P(-1.69 < z < 1.99) = .9312$

(f) What is the probability that the sample mean is less than 9 years or greater than 15 years?

$P(\bar{x} < 9 \text{ or } \bar{x} > 15) = P(z < -2.43 \text{ or } z > 1.99) = .0308$

12. **Video Summary of Sampling Distributions:** https://youtu.be/VVpAYgG6NBI

13. Here is a population. You want to take a sample of size $n = 5$. The population mean is $\mu = 10$ and the population standard deviation is $\sigma = 8$.

(a) Will the sampling distribution of the mean be normally distributed?

no.

(b) What is the mean of all the possible sample means?

$\mu_{\bar{x}} = 10$

(c) What is the standard deviation of all the possible sample means?

$\sigma_{\bar{x}} = \frac{8}{\sqrt{5}} = 3.58$

(d) What is the probability that the sample mean is greater than 11?

not normal, can't do this

(e) You want to take another sample so that you can have $\sigma_{\bar{x}}$ be 1.5 or less. How big of a sample do you need?

$1.5 = \frac{8}{\sqrt{n}}$, $n = 28.444$ you need 29+

14. Suppose that of all the adults in the U.S. the mean spending for Halloween is \$50 and the standard deviation is \$35. You decide to take a sample of 100 U.S. adults and look at the mean spending.

(a) What is the shape of the distribution of all the possible sample means?

normal, $100 \geq 30$ ✓

(b) What is the mean of all the possible sample means? $\mu_{\overline{X}}$

$\mu_{\bar{x}} = 50$

(c) What is the standard deviation of all the possible sample means? $\sigma_{\overline{X}}$

$\sigma_{\bar{x}} = \frac{35}{\sqrt{100}} = 3.5$

(d) Find the probability that the **total spending** for the 100 people is greater than \$5700.

Average spending = 57 per person $P(\bar{x} > 57) = 1 - P(z < 2) = .0228$

(e) What is the cutoff to be in the lowest 20% of all the possible sample means?

$P(\bar{x} < x) = .2$, $z = -.84$, $-.84 = \frac{x - 50}{35/\sqrt{100}}$ 47.06

15. Suppose that of all the adults in the U.S. the mean spending for Halloween is \$50 and the standard deviation is \$35. You decide to take a different sample of U.S. adults and look at the mean spending. However, you want to be able to control the standard deviation of the sampling distribution of the means.

(a) You want $\sigma_{\bar{x}}$ to be \$5 or less. How big of a sample do you need?

$5 = \frac{35}{\sqrt{n}}$, $n = 49$

(b) You want $\sigma_{\bar{x}}$ to be less than \$2. How many adults do you need to include in your sample?

$2 = \frac{35}{\sqrt{n}}$, $n = 306.25$,

307 adults

Optional Problems from Textbook

Section 7.3 Sampling Distributions 9, 13, 14, 20, 22, 24

Instead of using software to analyze the data sets, you can use the following summary statistics:

#13 ($\overline{x} = 69.35$, $s = 17.59$ and $n = 200$)

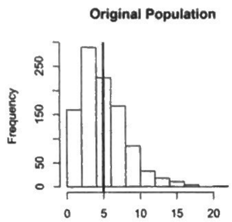
#14 ($\overline{x} = 3.291$, $s = 3.794$, $n = 55$)

#20, it is just asking $P(173 < \overline{X} < 175)$

#22, notice that they are asking for the probability that the point estimate of μ will be in the interval. So they are asking for the probability $P(109.9 < \overline{X} < 110.1)$

REQUIRED Extra Packet for Sampling Distributions for Sample Mean

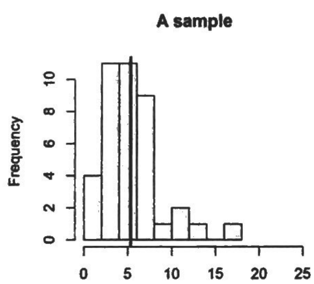
- Here is an actual population:



right skewed

  - mean $\mu = 4.853$
  - standard deviation $\sigma = 3.09$
  - this population has 1000 data values

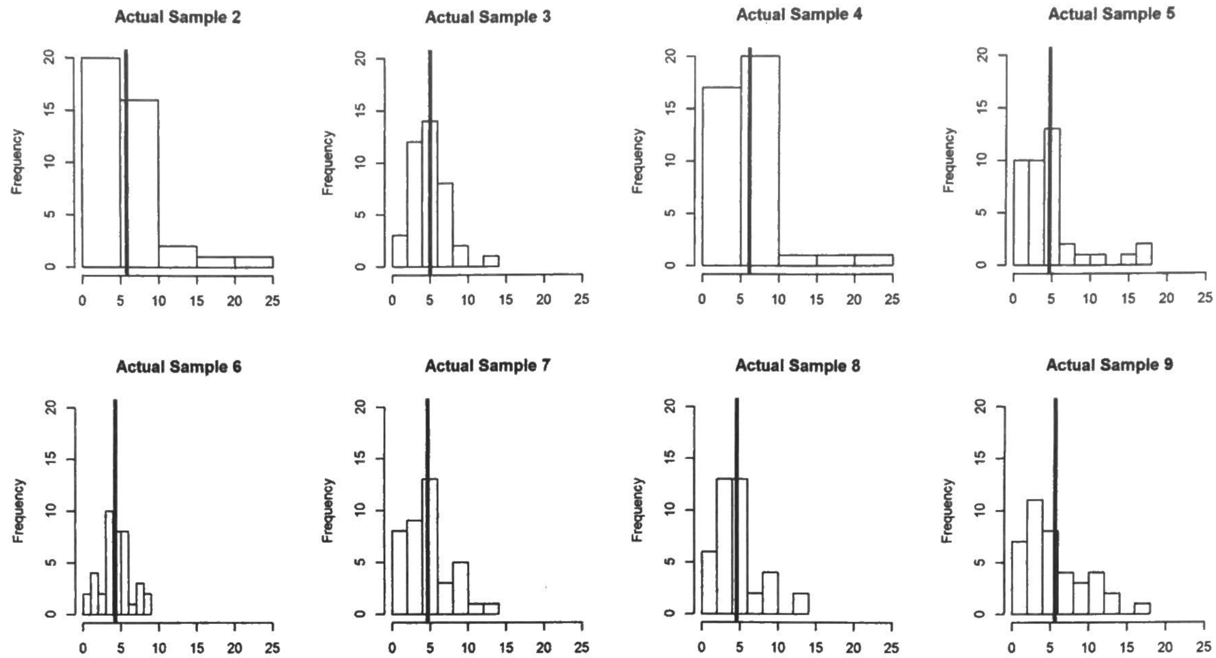
- Now we could take one sample: $n = 40$
  - actual data: 7.50, 3.61, 3.24, 4.02, 2.80, 4.44, 6.00, 0.89, 16.27, 1.87, 1.50, 8.74, 2.68, 1.06, 10.97, 5.11, 7.03, 4.50, 5.75, 7.26, 4.60, 6.71, 6.85, 2.20, 3.38, 5.67, 3.33, 2.62, 3.11, 2.20, 12.77, 2.56, 6.64, 5.33, 6.44, 7.27, 5.41, 4.04, 10.71, 5.38



  - sample size $n = 40$
  - mean $\overline{x} = 5.31$
  - standard deviation $s = 3.23$
  - shape: right skew, sim to pop.

For a sample, shape, mean, st. dev. close to pop.

- But if I took different samples I would get different results:



	sample 2	sample 3	sample 4	sample 5	sample 6	sample 7	sample 8	sample 9
sample mean $\overline{x}$	5.81	5.01	6.14	4.71	4.25	4.68	4.58	5.61
sample standard deviation s	4.36	2.55	4.07	4.02	2.08	3.01	2.92	3.99

– What do you notice about the _______ of each individual sample?

* shape
  usually right skewed
* mean
  close to pop. μ
* standard deviation
  close to pop σ

- With our specific population of 1000 values, and a sample size of $n = 40$, then there are

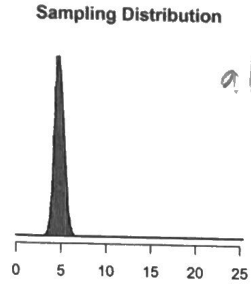
Combination $$\binom{1000}{40} = 5.5597 \times 10^{71}$$

different samples we could possible choose. Each one would look different and have a different mean and standard deviation.

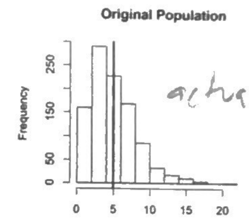
- We need some way to understand and predict what behaviors the actual samples will follow. So we started looking at what would happen if we took thousands or millions of different samples, each with $n = 40$.

- If we looked at every possible sample mean from every possible sample, we call it the **sampling distribution of the sample mean.**

- So if we take for now 1,000,000 different samples each with $n = 40$, here is what we get:



all possible sample means



actual

- This distribution of the different sample means is its own data set:
  - shape normal
  - mean $\mu_{\overline{X}} = 4.853$ equal
  - standard deviation $\sigma_{\overline{X}} = .4783$ much less.
    - $\sigma_{\overline{X}}$ is often called the standard error
    - We use the sample mean $\overline{x}$ from a sample to estimate the population mean μ.
    - $\sigma_{\overline{X}}$ tells us how accurate that estimate is likely to be.
    - $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ $\quad \frac{3.09}{\sqrt{40}} = .4885$ Close but not exact cause only 1 mil.

Things We Know

- Population:
  - can be any shape
  - mean: μ
  - standard deviation: σ
- An Individual Sample with n data values:
  - shape: usually close to original population
  - mean $\overline{x}$: usually close to original population mean μ
  - standard deviation s: usually close to original population standard deviation σ
  - **Notice that there are no guarantees for the individual sample. All of the values are usually close to the original population value, but sometimes you get one of those few weird samples that aren't.
- The Sampling Distribution of ALL the Possible Sample Means:
  - shape: normal if $n \geq 30$ or original population is normal
  - mean $\mu_{\overline{X}}$: same as original population mean μ

$$\mu_{\overline{X}} = \mu$$

  - standard deviation $\sigma_{\overline{X}}$: smaller than original population mean σ

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

- **Notice that the sample size being big ($n \geq 30$) only affects the sampling distribution. It does not affect the population or individual sample.