PRACTICE 1

1. A recent study reported that high school students spend an average of 94 minutes per day texting. Jenna claims that the average for the students at her large high school is greater than 94 minutes. She will conduct a study to investigate this claim.

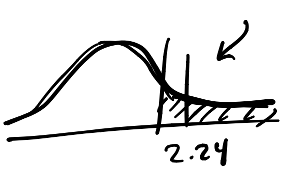
   (a) To collect data Jenna will select a sample of size 32 from the population. State Jenna's population of interest.

   • The population of interest is all students at Jenna's high school. ✓

   (b) Name and describe a sampling method Jenna could use that will satisfy the conditions needed for inference.

   ✓ One possible method is to use a simple random sample. Obtain a numbered list of all the names of students at the school. Then use a random number generator to generate 32 different numbers from the list. The students whose names are associated with selected numbers are used as the sample.

   (c) Based on a sample of 32 students, Jenna calculated a sample mean of 96.5 minutes and a sample standard deviation of 6.3 minutes. Assume all conditions of inference are met. At the significance level of $\alpha = 0.05$, do the data provide convincing statistical evidence to support Jenna's claim? Complete an appropriate inference procedure to support your answer.



$\bar{x} = 96.5$

$s = 6.3$

✓ 1. We will conduct a one-sample t-test for μ.

$H_0: \mu = 94$

$H_a: \mu > 94$

✓ 2. We are told that conditions are met for inference

3. Test Statistic.

$$t = \frac{96.5 - 94}{\frac{6.3}{\sqrt{32}}} \approx 2.24.$$

using a t-distribution with $n-1 = 31$ degrees of freedom we get $P(t_{31} > 2.24) = 0.016$, so p-value = 0.016.

4. The p-value of 0.016 is less than $\alpha = 0.05$, so the null hypothesis is rejected. There is convincing statistical evidence that the mean number of minutes that all students at Jenna's highschool spend per day on

at Jenna's highschool sp...
texting is greater than 94 minutes.

2. A large university offers STEM internship to women in STEM majors at the university, A woman must be 20 years or older to meet the age requirement for the internships. The table shows the probability distribution of the ages of the woman in STEM majors at the university:

Age (years)	17	18	19	20	21	22	23 or older
Probability	0.005	0.107	0.111	0.252	0.249	0.213	0.063

(a) Suppose one woman is selected at random from the women in STEM majors at the university. What is the probability that the woman selected will not meet the age requirement for the internship?

- $P(17, 18 \text{ or } 19) = 0.005 + 0.107 + 0.111 = 0.223$.

There is a 0.223 probability that a student is 17, 18, or 19 and will not qualify for the internship.

(b) The university will select a sample of 100 women in STEM majors to participate in a focus group about the internships. Suppose a simple random sampling process is used to select the sample of 100 women. What is the probability that at least 30 percent of the women in the sample will not meet the age requirement for the internships?

Let X represent the number of woman in the sample who do not meet the age requirement.

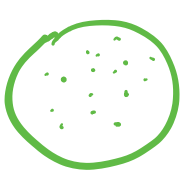
$X \sim \text{binomial}(n=100,\ p=0.223)$

$P(x \geq 30) = 1 - P(x \leq 29) = 1 - 0.9547 = \boxed{0.0453}$

(c) Suppose a stratified random sampling design is used to select a sample of 30 women who do not meet the age requirement and a sample of 70 women who do meet the age requirement. Based on the probability distribution, is a woman who does not meet the age requirement more likely, less likely, or equally likely to be selected with a stratified random sample than with a simple random sample? Justify your answer.

SRS

- on average expect. 22.3% of people to not qualify



Stratified sample.

- $\frac{30}{100}$ or 30% of girls sampled will not meet age requirement.

Therefore, a woman who does not meet the age requirement is more likely to make it into the stratified sample than the SRS.

3. An automobile manufacturer sold 30,000 new cars, one to each of 30,000 customers, in a certain year. The manufacturer was interested in investigating the proportion of the new cars that experienced a mechanical problem within the first 5,000 miles driven.

   (a) A list of the names and addresses of all customers who bought the new cars is available. Describe a sampling plan that could be used to obtain a simple random sample of 1,000 customers from the list.

   1. Number the customers from 1 to 30 000,

   2. use a random number generator to generate 1000 random numbers between 1 and 30000 without replacement. If the random numbers are non-unique, the repeated numbers are thrown out, this is repeated until 1000 unique numbers are obtained.

   3. The customers whos names correspond to each randomly generated number are selected for sample.

   (b) Each customer from a simple random sample of 1,000 customers who bought one of the new cars was asked whether they experienced any mechanical problems within the first 5,000 miles driven. Forty customers from the sample reported a problem. Of the 40 customers who reported a problem, 13 customers, or 32.5%, reported a problem specifically with the power door locks.

   Explain why 0.325 should not be used to estimate the population proportion of the 30,000 new cars sold that experienced a problem with the power door locks within the first 5000 miles driven.

   32.5% should not be used as an estimate of the entire population with power lock problems because it represents the percent of cars that had door lock problems given that the car had some sort of problem. But only 40/1000 or 4%, of the cars in the sample had any sort of problem.

   (c) Based on the results of the sample give a point estimate of the number of new cars sold that experience a problem with the power door locks within the first 5000 miles driven.

   $\frac{13}{1000} = 0.013$ or 1.3%