AP Statistics
Mr. Murphy

HW Pg. Pg. 431 #8.3, 8.4, 8.7(a), 8.8, Pg. 439 #8.10, 8.11, 8.14, 8.17, 8.19

8.1 The Sampling Distribution of a Sample Mean

***Note*: Sampling Distribution = Shape (i.e. Normal, Uniform, Skewed Left/Right) OR a Chart with two rows, x values on the top row, probabilities on the second row.**

Goals: 1. Explain the concepts of sampling variability and sampling distribution.
2. Determine the sampling distribution for $\bar{x}$.
3. Calculate probabilities based on the distribution of $\bar{x}$.
4. Explain the Central Limit Theorem.

Ex1 Consider a small population consisting of the 20 students enrolled in an upper division class. The students are numbered 1 to 20, and the amount of money (in dollars) each student spent on textbooks for the current semester is shown in the following table:

Student	Amount Spent on Books	Student	Amount Spent on Books
1	267	11	319
2	258	12	263
3	342	13	265
4	261	14	262
5	275	15	333
6	295	16	184
7	222	17	231
8	270	18	159
9	278	19	230
10	168	20	323

Calculate μ and σ. Describe the SOCS.

Now take a random sample of 5 students and calculate $\bar{x}$.

Did we all get the same $\bar{x}$?

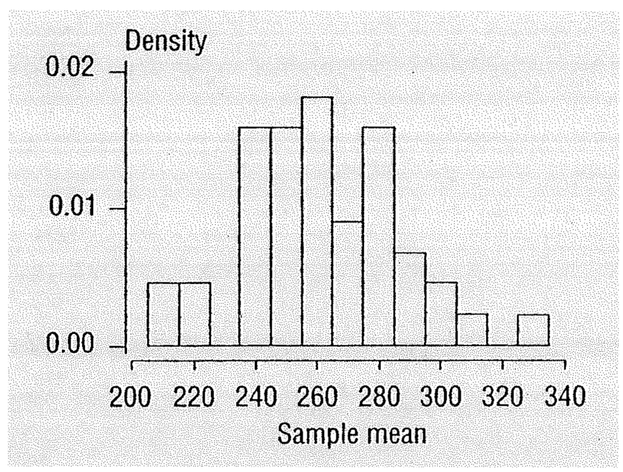
- The observed value of a statistic depends on the particular sample selected from the population; typically, it varies from sample to sample. This variability is called **sampling variability**.

- The distribution of the values of a statistic is called its **sampling distribution.**

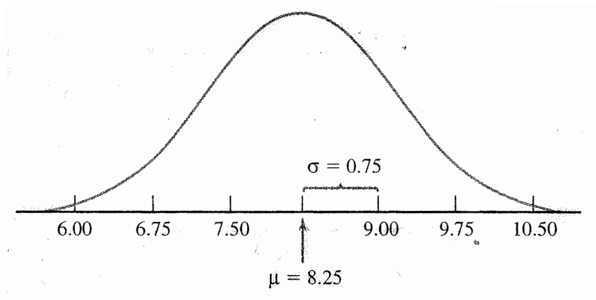
I took 45 samples, each of size 5, and calculated each $\bar{x}$. Here are the sample means:

Sample #	$\bar{x}$	Sample #	$\bar{x}$	Sample #	$\bar{x}$
1	274.6	16	255.0	31	253.4
2	256.6	17	307.2	32	279.6
3	263.8	18	280.0	33	252.6
4	254.8	19	277.4	34	242.2
5	238.2	20	241.2	35	262.6
6	275.6	21	216.0	36	215.4
7	279.2	22	270.0	37	266.2
8	241.6	23	301.0	38	261.4
9	255.4	24	247.0	39	275.0
10	263.8	25	273.8	40	301.4
11	239.6	26	282.8	41	237.0
12	248.2	27	220.4	42	287.4
13	330.8	28	213.6	43	249.2
14	288.8	29	287.6	44	236.8
15	252.8	30	214.8	45	264.6

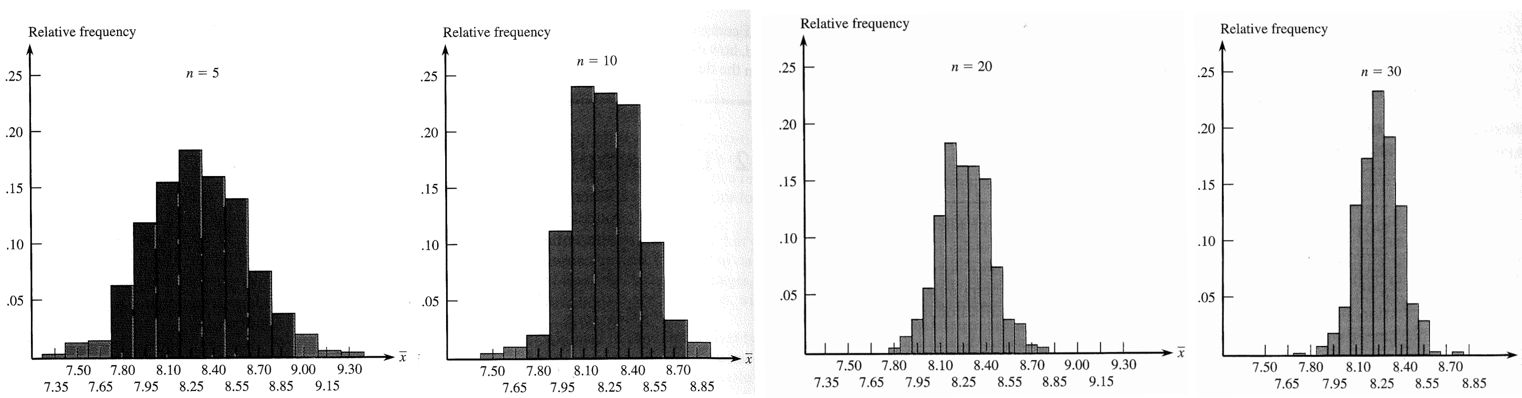
Here is a histogram of all the $\bar{x}$'s. What shape does the **sampling distribution** have? What can be said about the spread of this distribution as compared to the spread of the original data?



<u>Ex2</u> Data suggested that the distribution of platelet size for patients with non-cardiac chest pain is approximately normal with mean $\mu = 8.25$ and standard deviation $\sigma = 0.75$. The figure below shows the corresponding normal curve.

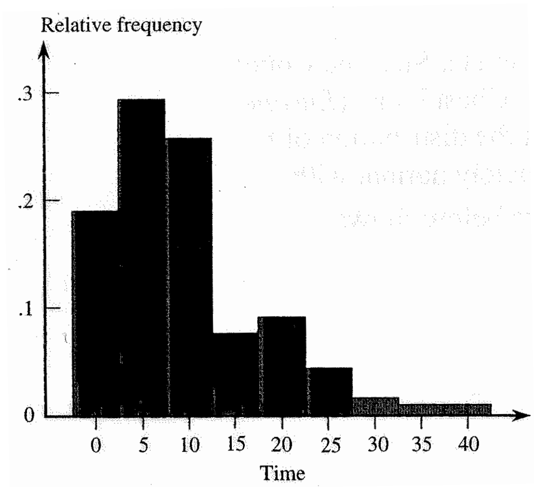


MINITAB was used to select 500 random samples from this normal distribution, with each sample consisting of n = 5 observations. The process was repeated for sample of size n = 10, n = 20, and n = 30. The resulting 500 $\bar{x}$ values appear in the following histograms.



What can be said about the histograms? In terms of SOCS?

Ex3 Now consider the properties of the $\bar{x}$ distribution when the population is quite skewed (and thus very unlike a normal distribution). The Winter 1995 issue of *Chance* magazine gave data on the length of overtime period for all 251 National Hockey League play-off games between 1970 and 1993 that went into overtime. In hockey, the overtime period ends as soon as one of the teams scores a goal. This histogram of the data is below.

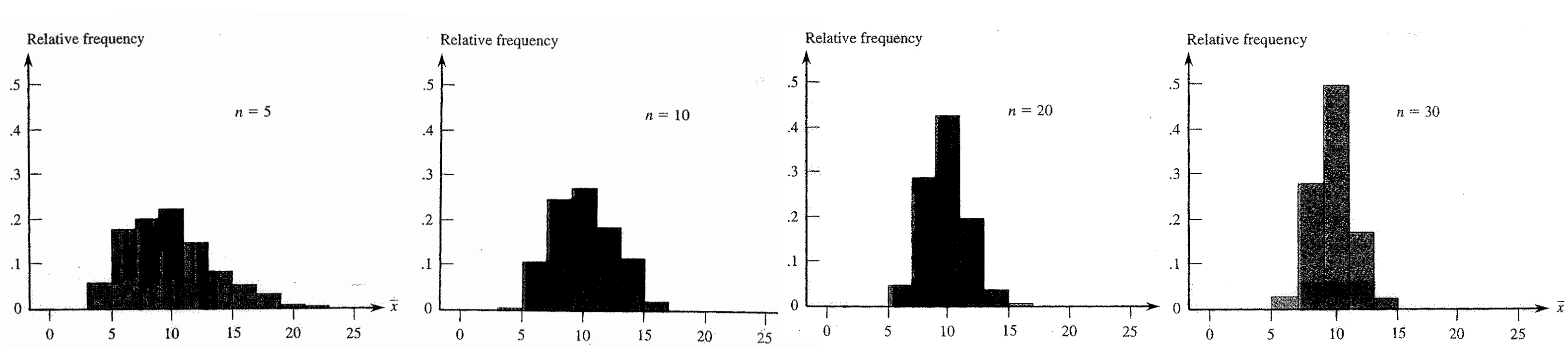


Describe this histogram. (SOCS)

Let's say we found that $\mu = 9.841$, so that is the balance point for the population histogram. The median is 8. What does that tell us about our histogram?

For each of the sample sizes n = 5, 10, 20, and 30, we selected 500 random samples of size n. We then constructed the following histograms of the 500 $\bar{x}$ values.

What do we notice about the histograms? In terms of SOCS.



What do we notice about the histograms in terms of SOCS?

• **General Properties of the Sampling Distribution of $\bar{x}$**
Let $\bar{x}$ denote the mean of the observations in a random sample of size n from a population having mean μ and standard deviation σ. Denote the mean value of the $\bar{x}$ distribution by $\mu_{\bar{x}}$ and the standard deviation of the $\bar{x}$ distribution by $\sigma_{\bar{x}}$. Then the following rules hold:

Rule #1: $\mu_{\bar{x}} = \mu$

Rule #2: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

Assumption #1: When the population distribution is normal, the sampling distribution of $\bar{x}$ is also normal for any sample size n.

Assumption #2 (Central Limit Theorem): When n is sufficiently large, the sampling distribution of $\bar{x}$ is well approximated by a normal curve, even when the population distribution is not itself normal.

One of the assumptions must be met in order to continue.

• If n is large or if the population distribution is normal, then the standardized variable

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

has (at least approximately) a standard normal (z) distribution.

• <u>The Central Limit Theorem can be safely applied if n exceeds 30.</u>

<u>Ex4</u> What are the mean and the standard deviation of a sampling distribution consisting of samples of size 16? These samples were drawn from a population whose mean is 25 and whose standard deviation is 5.

(a) 25, 1.25
(b) 5, 5
(c) 25, 5
(d) 5, 1.25
(e) 25, $\sqrt{5}$

Can we consider this distribution normal?

Ex5 The mean TOEFL score of international students at a certain university is normally distributed with a mean of 490 and a standard deviation of 80. Suppose that groups of 30 students are studied. The mean and the standard deviation for the distribution of sample means, respectively, will be

(a) 490 and 8/3
(b) 16.33 and 80
(c) 490 and 14.61
(d) 490 and 213.33

Can we consider this distribution normal?

Ex6 A certain brand of lightbulb has a mean lifetime of 1500 hours with a standard deviation of 100 hours. If the bulbs are sold in boxes of 25, the parameters of the distribution of sample means are

(a) 1500 and 100
(b) 1500 and 4
(c) 1500 and 2
(d) 1500 and 20

Can we consider this distribution normal?

Ex7 A soft-drink bottler claims that, on average, cans contain 12 oz. of soda. Let x denote the actual volume of soda in a randomly selected can. Suppose that x is normally distributed with $\sigma = 0.16$ oz. Sixteen cans are to be selected, and the soda volume will be determined for each one. Let $\bar{x}$ denote the resulting sample mean soda volume.

What is the distribution of x ?

Calculate $P(11.96 \le x \le 12.08)$.

Calculate $P(x \le 12.08)$

What is the distribution of $\bar{x}$?

Calculate $P(11.96 \le \bar{x} \le 12.08)$.

Calculate $P(\bar{x} \le 12.08)$.

Ex8 Samples of size 49 are drawn from a distribution that's highly skewed to the right with a mean of 70 and a standard deviation of 14. What's the probability of getting a sample mean between 71 and 73?

(a) 0.0563
(b) 0.00023
(c) 0.2417
(d) 0
(e) We can't answer this question
because the distribution is highly skewed.

Checkpoint:
Multiple Choice
1. As the sample size increases,
(a) the population mean decrease.
(b) the population standard deviation decreases.
(c) the standard deviation for the distribution of the sample mean increases.
(d) the standard deviation for the distribution of the sample mean decreases.

2. Samples of size 49 are drawn from a population with a mean of 36 and a standard deviation of 15. Then $P(\bar{x} < 33)$ is

(a) 0.5808
(b) 0.4192
(c) 0.1608
(d) 0.0808

3. A tire manufacturer claims that its tires will last an average of 40,000 miles with a standard deviation of 3,000 miles. Forty-nine tires were placed on test, and the average failure was recorded. The probability that the average value of failure miles is less than 39,500 is

(a) 0.3790
(b) 0.8790
(c) 0.1217
(d) 0.6210

4. Lloyd's Cereal Company packages cereal in 1-pound boxes (1 pound = 16 ounces). It is assumed that the amount of cereal per box varies according to a normal distribution with a standard deviation of 0.05 pound. One box is selected at random from the production line every hour, and if the weight is less than 15 ounces, the machine is adjusted to increase the amount of cereal dispensed. The probability that the amount dispensed per box will have to be increased during a 1-hour period is

(a) 0.3944
(b) 0.8944
(c) 0.1056
(d) 0.6056

Free Response

Suppose that $T = 2X + 3Y$, where X is your score on the multiple-choice part of a test, Y is your score on the written part of a test, and T is your total score. In this case, your total score is calculated by doubling your multiple-choice score and tripling your written score. Suppose $X \sim N(30,7)$ and $Y \sim N(20,13)$.

(a) What is the probability that T will be greater than 130?
(b) Suppose that we find three students' T scores independently of one another. What is the probability that at least one of these measurements will be greater than 130?
(c) What is the probability that the mean of the three independent scores will be greater than 130?