# *Lecture Three*

## *Section* 3.1 – Distribution of the Sample Mean / Proportion

Statistics such as $\bar{x}$ are random variables since their value varies from sample to sample. As such, they have probability distributions associated with them. We focus on the shape, center and spread of statistics such as $\bar{x}$.

### *Definitions*

The ***sampling distribution*** of a statistic is a probability distribution for all possible values of the statistic computed from a sample of size $n$.

The ***sampling distribution of the sample mean*** $\bar{x}$ is the probability distribution of all possible values of the random variable $\bar{x}$ computed from a sample of size $n$ from a population with mean $\mu$ and standard deviation $\sigma$.

### Properties

- Sample means target the value of the population mean. (That is, the mean of the sample means is the population mean. The expected value of the sample mean is equal to the population mean.)
- The distribution of the sample means tends to be a normal distribution.

### *Sampling Distributions*

***Step*** 1: Obtain a simple random sample of size $n$.
***Step*** 2: Compute the sample mean.
***Step*** 3: Assuming we are sampling from a finite population, repeat Steps 1 and 2 until all simple random samples of size $n$ have been obtained.

### *Definition*

The ***sampling distribution of the variance*** is the distribution of sample variances, with all samples having the same sample size $n$ taken from the same population. (The sampling distribution of the variance is typically represented as a probability distribution in the format of a table, probability histogram, or formula.)

### Biased Estimators

Sample *medians*, *ranges* and *standard deviations* are biased estimators. That is they do NOT target the population parameter.

***Note***: the bias with the standard deviation is relatively small in large samples so $s$ is often used to estimate.

## Unbiased Estimators

Sample means, variances and proportions are ***unbiased estimators***. That is they target the population parameter.
These statistics are better in estimating the population parameter

- *Mean* $\overline{x}$
- *Variance* $s^2$
- *Proportion* $\hat{p}$

## The Mean and Standard Deviation of the Sampling Distribution of $\overline{x}$

Suppose that a simple random sample of size $n$ is drawn from a large population with mean $\mu$ and standard deviation $\sigma$. The sampling distribution of will have mean $\mu_{\overline{x}} = \mu$ and standard deviation

$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$ The standard deviation of the sampling distribution of $\overline{x}$ is called the ***standard error of the mean*** and is denoted $\sigma_{\overline{x}}$ .

**Notation**

The mean of the sample means $\qquad \mu_{\overline{x}} = \mu$

The standard deviation of sample mean $\qquad \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$

## The Shape of the Sampling Distribution of $\overline{x}$ If *X* is Normal

If a random variable $X$ is normally distributed, the distribution of the sample mean $\overline{x}$ is normally distributed.

## Applying

***Individual value***: When working with individual value from a normally distributed population, use

$$z = \frac{x - \mu}{\sigma}$$

***Sample value***: When working with a mean for some sample (or group), use the value of $\frac{\sigma}{\sqrt{n}}$ for the standard deviation of the sample means. Use $z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

### *Example*

Assume the population of weights of men is normally distributed with a mean of 172 lb. and a standard deviation of 29 lb.

*a)* Find the probability that if an *individual* man is randomly selected, his weight is greater than 175 lb.

*b)* Find the probability that *20 randomly selected men* will have a mean weight that is greater than 175 lb. (so that their total weight exceeds the safe capacity of 3500 pounds).

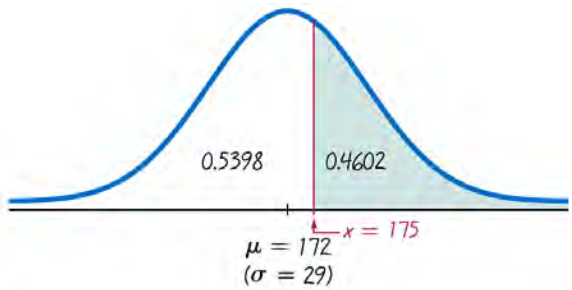
***Solution***

***a)*** We are dealing with an individual value from a normally distributed population.

$$z = \frac{x-\mu}{\sigma} = \frac{175-172}{29} = 0.10$$

Using the Table (*Normal Distribution*), $A_1 = 0.5398$

Therefore, the shaded region area is: $A = 1 - 0.5398 \ \underline{= 0.4602}$




The probability of a randomly selected man weighing more than 175 lb. is 0.4602.
(Calculator result is 0.4588)

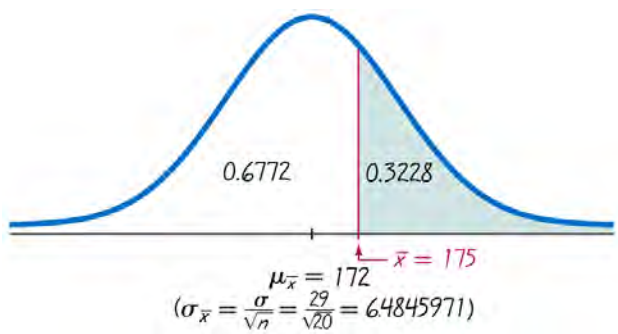
***b)*** Although the sample size is not greater than 30. We use a normal distribution because the original population of men has a normal distribution.

$$\mu_{\overline{x}} = \mu = 172$$

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{29}{\sqrt{20}} = 6.4845971$$

$$z = \frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{175-172}{6.4845971} = 0.46$$

$$A_2(<0.46) = 0.6772$$




Therefore, the shaded region area is: $A = 1 - 0.6772 \ \underline{= 0.3228}$

So, the probability that the 20 men have a mean weight greater than 175 lb. is 0.3228.
(Calculator result is 0.3218)

✓ There is 0.4602 probability that an individual man will weigh more than 175 lb., and there is a 0.3228 probability that 20 men will have a mean weight of more than 175 lb. Given that the safe capacity of the water taxi is 3500 pounds, there is a fairly good chance (with probability 0.3228) that it will be overloaded with 20 randomly selected men. The capacity of 20 passengers is just not safe enough.

### *Example*

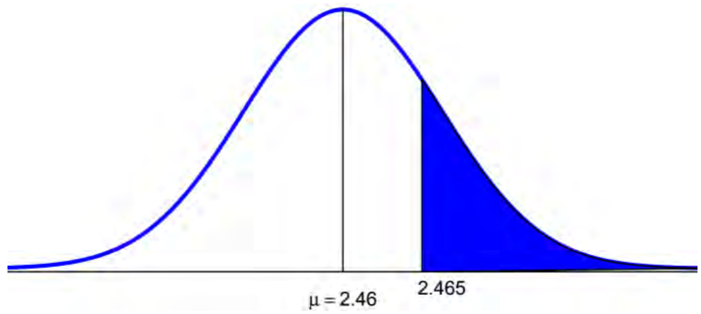
The weights of pennies minted after 1982 are approximately normally distributed with mean 2.46 grams and standard deviation 0.02 grams.
What is the probability that in a simple random sample of 10 pennies minted after 1982, we obtain a sample mean of at least 2.465 grams?

***Solution***

$$\mu_{\bar{x}} = 2.46, \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.02}{\sqrt{10}} = .0063$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2.465 - 2.46}{0.0063} = 0.79$$

$$P(z > 0.79) = 1 - 0.7852 \underline{= 0.2148}$$




### *Example*

The following table and histogram give the probability distribution for rolling a fair die.

| *Face on Die* | *Relative Frequency* |
|---|---|
| 1 | 0.1667 |
| 2 | 0.1667 |
| 3 | 0.1667 |
| 4 | 0.1667 |
| 5 | 0.1667 |
| 6 | 0.1667 |

***Solution***

$\underline{\mu = 3.5}$

$\underline{\sigma = 1.708}$

The population distribution is ***NOT*** normal.

## Correction for a Finite Population

When sampling without replacement and the sample size $n$ is greater than 5% of the finite population of size $N$ (that is, $n > 0.05N$), adjust the standard deviation of sample means by multiplying it by the *finite population correction factor*:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \underbrace{\sqrt{\frac{N-n}{N-1}}}_{\textit{finite population correction factor}}$$

## Example

Cans of regular Coke are labeled to indicate that they contain 12 *oz*. The corresponding sample statistics are $n = 36$ and $\overline{x} = 12.19\ oz$. If the Coke cans filled so that $\mu = 12\ oz$ and the population standard deviation is $\sigma = 0.11\ oz$ (based on the sample result), find the probability that a sample of 36 cans will have a mean of 12.19 *oz*. or greater. Do these results suggest that the Coke cans are filled with an amount greater than 12.00 *oz*.?

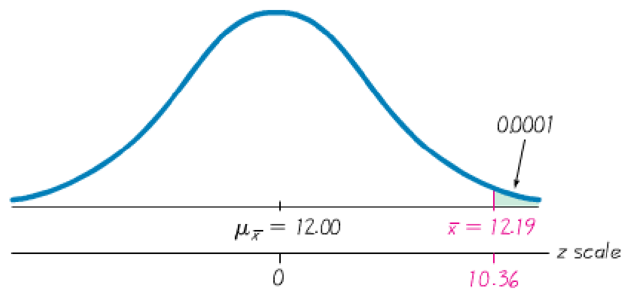
**Solution**

Because the sample size $n = 36$ exceeds 30, we apply the central limit theorem and conclude that the distribution of sample means is approximately a normal distribution with these parameters:

$$\mu_{\overline{x}} = \mu = 12$$

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.11}{\sqrt{36}} = 0.018333$$

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{12.19 - 12}{0.018333} = 10.36$$




Since the value of $z = 10.36$ is off the chart.
However, for values of *z* above 3.49, we use 0.9999 for the cumulative left area.
Therefore, we conclude that the shaded region is 0.0001.

✓ The result shows that there is an extremely small probability of getting a sample mean of 12.19 oz. or greater when 36 cans are randomly selected. It appears that the company has found a way to ensure that very few cans have less than 12 oz. while not wasting very much of their product.

## Example

It is not totally unreasonable to think that screws labeled as being 3/4 in. in length would have a mean length that is somewhat close to 3/4 in. The lengths of a sample of 50 such screws have a mean length of 0.7468 in. Assume that the population of all such screws has a standard deviation described by $\sigma = 0.0123\ in$.

*a)* Assuming that the screws have mean length of 0.75 in. as labeled, find the probability that a sample of 50 screws has a mean length of 0.7468 in. or less.

*b)* The probability of getting a sample mean that is "at least as extreme as the given sample mean" is twice the probability found in part (*a*). Find this probability. (Note that the sample mean of 0.7468 in. misses the labeled mean of 0.75 in. by 0.0032 in., so any other mean is at least as extreme as the sample mean if it is below 0.75 in. by 0.0032 in. or more, or if it is above 0.75 in. by 0.0032 in. or more.)

*c)* Based on the result in part (b), does it appear that the sample mean misses the labeled mean of 0.75 in. by a significant amount? Explain.

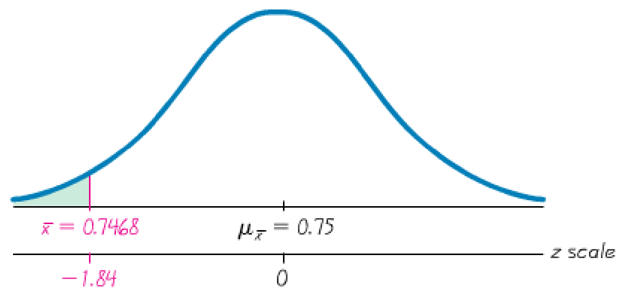
**Solution**

***a)*** Because the sample size $n = 50$ exceeds 30, we apply the central limit theorem and conclude that the distribution of sample means is a normal distribution with these parameters:

$$\mu_{\bar{x}} = \mu = 0.75$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.0123}{\sqrt{50}} = 0.001739$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.7468 - 0.75}{0.001739} = -1.84$$




$$A(< -1.84) \underline{= 0.0329}$$

The probability of getting a sample mean of 0.7468 in. or less is 0.0329.

**b)** $P(\textit{at least as extreme as the given sample mean}) = 2P(\textit{part } a)$

$$= 2 \times 0.0329$$

$$\underline{= 0.0658}$$

**c)** The result from part (b) shows that there is a 0.0658 probability of getting a sample mean that is at least as extreme as the given sample. Using a 0.05 cutoff probability of 0.0658 exceeds 0.05, so the sample mean is not unusual. We conclude that the given sample mean does not miss the labeled mean of 0.75 in. by a substantial amount. The labeling of 0.75 in. appears to be justified.

## Central Limit Theorem

Regardless of the shape of the underlying population, the sampling distribution of $\bar{x}$ becomes approximately normal as the sample size, $n$, increases.

The mean of the sampling distribution is equal to the mean of the parent population and the standard deviation of the sampling distribution of the sample mean is $\frac{\sigma}{\sqrt{n}}$ regardless of the sample size.

### *Example*

Suppose that the mean time for an oil change at a "10-minute oil change joint" is 11.4 minutes with a standard deviation of 3.2 minutes.

*a)* If a random sample of $n = 35$ oil changes is selected, describe the sampling distribution of the sample mean.

*b)* If a random sample of $n = 35$ oil changes is selected, what is the probability the mean oil change time is less than 11 minutes?

***Solution***

**a)** $\mu_{\bar{x}} = 11.4$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.2}{\sqrt{35}} = 0.5409$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.7468 - 0.75}{0.001739} \underline{= -1.84}$$

## Point Estimate of a Population Proportion

Suppose that a random sample of size $n$ is obtained from a population in which each individual either does or does not have a certain characteristic. The sample proportion, denoted $\hat{p}$ (read "*p*-hat") is given by

$$\hat{p} = \frac{x}{n}$$

where $x$ is the number of individuals in the sample with the specified characteristic. The sample proportion $\hat{p}$ is a statistic that estimates the population proportion, $p$.

### *Example*

In a Quinnipiac University Poll conducted in May of 2008, 1745 registered voters nationwide were asked whether they approved of the way Bush is handling the economy. 349 responded "yes". Obtain a point estimate for the proportion of registered voters who approve of the way Bush is handling the economy.

***Solution***

$$\hat{p} = \frac{x}{n} = \frac{349}{1745} = 0.2$$

### Sampling Distribution of $\hat{p}$

For a simple random sample of size $n$ with population proportion $p$:

- ✓ The shape of the sampling distribution of $\hat{p}$ is approximately normal provided $np(1-p) \geq 10$.
- ✓ The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$
- ✓ The standard deviation of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
- ✓ The model on the previous slide requires that the sampled values are independent. When sampling from finite populations, this assumption is verified by checking that the sample size $n$ is no more than 5% of the population size $N$ ($n \leq 0.05N$).
- ✓ Regardless of whether $np(1-p) \geq 10$ or not, the mean of the sampling distribution of $\hat{p}$ is $p$, and the standard deviation is $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

### *Example*

According to a *Time* poll conducted in June of 2008, 42% of registered voters believed that gay and lesbian couples should be allowed to marry. Suppose that we obtain a simple random sample of 50 voters and determine which voters believe that gay and lesbian couples should be allowed to marry. Describe the sampling distribution of the sample proportion for registered voters who believe that gay and lesbian couples should be allowed to marry.

***Solution***

The sample of $n = 50$ is smaller than 5% of the population size (all registered voters in the U.S.).

$$npq = 50(0.42)(0.58) \underline{= 12.8 \geq 10} \quad \& \quad \underline{\mu \approx 0.42}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.42(1-0.42)}{50}} \underline{= 0.0698}$$

### Example

According to the Centers for Disease Control and Prevention, 18.8% of school-aged children, aged 6-11 years, were overweight in 2004.

*a)* In a random sample of 90 school-aged children, aged 6-11 years, what is the probability that at least 19% are overweight?

*b)* Suppose a random sample of 90 school-aged children, aged 6-11 years, results in 24 overweight children. What might you conclude?

***Solution***

$n = 90$ is less than 5% of the population size

$$npq = 90(0.188)(1-0.188) \underline{\approx 13.7 \geq 10} \qquad \& \qquad \underline{\mu = 0.188}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.188(1-0.188)}{90}} \underline{= 0.0412}$$

**a)** $z = \dfrac{0.19 - 0.188}{0.0412} \underline{= 0.0485}$

$$P(z > 0.05) = 1 - 0.5199 \underline{= 0.4801}$$

**b)** $\hat{p} = \dfrac{24}{90} = 0.2667$

$$z = \frac{0.2667 - 0.188}{0.0412} \underline{= 1.91}$$

$$P(z > 1.91) = 1 - 0.9719 \underline{= 0.028}$$

We would only expect to see about 3 samples in 100 resulting in a sample proportion of 0.2667 or more. This is an unusual sample if the true population proportion is 0.188.

## Distribution of Sample Means

| ***Population (with mean $\mu$ and standard deviation $\sigma$)*** | ***Distribution of Sample Means*** | ***Mean of the Sample Means*** | ***Standard Deviation of the Sample Means*** |
|---|---|---|---|
| Normal | Normal (for any sample size $n$) | $\mu_{\bar{x}} = \mu$ | $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ |
| Not normal with $n > 30$ | Normal (approximately) | $\mu_{\bar{x}} = \mu$ | $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ |
| Not normal with $n \leq 30$ | Not normal | $\mu_{\bar{x}} = \mu$ | $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ |

# *Exercise* *Section* 3.1 – Distribution of the Sample Mean / Proportion

**1.** Assume that SAT scores are normally distributed with mean $\mu = 1518$ and standard deviation $\sigma = 325$.

*a)* If 1 SAT score is randomly selected, find the probability that it is less than 1500.
*b)* If 100 SAT scores are randomly selected, find the probability that they have a mean less than 1500.
*c)* If 1 SAT score is randomly selected, find the probability that it is greater than 1600.
*d)* If 64 SAT scores are randomly selected, find the probability that they have a mean greater than 1600.
*e)* If 1 SAT score is randomly selected, find the probability that it is between 1550 and 1575.
*f)* If 25 SAT scores are randomly selected, find the probability that they have a mean between 1550 and 1575.

**2.** Assume that weights of mean are normally distributed with a mean of 172 lb. and a standard deviation of 29 lb.

*a)* Find the probability that if an individual man is randomly selected, his weight will be greater than 180 lb.
*b)* Find the probability that 20 randomly selected men will have a mean weight that is greater than 180 lb.
*c)* If 20 mean have a mean weight greater than 180 lb., the total weight exceeds the 3500 lb. safe capacity of a particular water taxi. Based on the proceeding results, is this safety concern? Why or why not?

**3.** Membership requires an IQ score above 131.5. Nine candidates take IQ tests, and their summary results indicated that their mean IQ score is 133, (IQ scores are normally distributed with mean of 100 and a standard deviation of 15.)

*a)* If 1 person is randomly selected from the general population, find the probability of getting someone with an IQ score of at least 133.
*b)* If 9 people are randomly selected, find the probability that their mean IQ score is at least 133.
*c)* Although the summary results are available, the individual IQ test scores have been lost. Can it be concluded that all 9 candidates have IQ scores above 131.5 so that they are all eligible for membership?

**4.** For women aged 18-24, systolic blood pressures (in mm Hg) are normally distributed with a mean of 114.8 and a standard deviation of 13.1. Hypertension is commonly defined as a systolic blood pressure above 140.

*a)* If a woman between the ages of 18 and 24 is randomly selected, find the probability that her systolic blood pressure is greater than 140.
*b)* If 4 women in that age bracket are randomly selected, find the probability that their mean systolic blood pressure is greater than 140.
*c)* Given that part (*b*) involves a sample size that is not larger than 30, why can the central limit theorem be used?

*d)* If a physician is given a report stating that 4 women have a mean systolic, blood pressure below 140, can she conclude that none of the women have hypertension (with a blood pressure greater than 140)?

**5.** Engineers must consider the breadths of male heads when designing motorcycle helmets. Men have head breaths that are normally distributed with a mean of 6.0 in. and a standard deviation of 1.0 in.

*a)* If one male is randomly selected, find the probability that his head breadth is less than 6.2 in.

*b)* The Safeguard Helmet Company plans an initial production run of 100 helmets. Find the probability that 100 randomly selected men have a mean head breadth less than 6.2 in.

*c)* The production manager sees the result from part (b) and reasons that all helmets should be made for men with head breadths less than 6.2 in., because they would fit all but a few men. What is wrong with that reasoning?

**6.** Currently, quarters have weighs that are normally distributed with a mean 5,670 *g* and a standard deviation of 0.062 *g*. A vending machine is configured to accept only those quarters with weights between 5.550 *g* and 5.790 *g*.

*a)* If 280 different quarters are inserted into the vending machine, what is the expected number of rejected quarter?

*b)* If 280 different quarters are inserted into the vending machine, what is the probability that the mean falls between the limits of 5.550 *g* and 5.790 *g*?

*c)* If you own the vending machine, which result would concern you more? The result from part (*a*) or the result from part (*b*)? Why?

**7.** The annual precipitation amounts in a certain mountain range are normally distributed with a mean of 100 inches, and a standard deviation of 12 inches. What is the probability that the mean annual precipitation during 36 randomly picked years will be less than 112.8 inches?

**8.** The annual precipitation amounts in a certain mountain range are normally distributed with a mean of 72 inches, and a standard deviation of 14 inches. What is the probability that the mean annual precipitation during 49 randomly picked years will be less than 74.8 inches?

**9.** The weights of the fish in a certain lake are normally distributed with a mean of 13 *lb*. and a standard deviation of 6. If 4 fish are randomly selected, what is the probability that the mean weight will be between 10.6 and 16.6 *lb*.?

**10.** For women aged 18-24, systolic blood pressures (in mm Hg) are normally distributed with a mean of 114.8 and a standard deviation of 13.1. If 23 women aged 18-24 are randomly selected, find the probability that their mean systolic blood pressure is between 119 and 122.

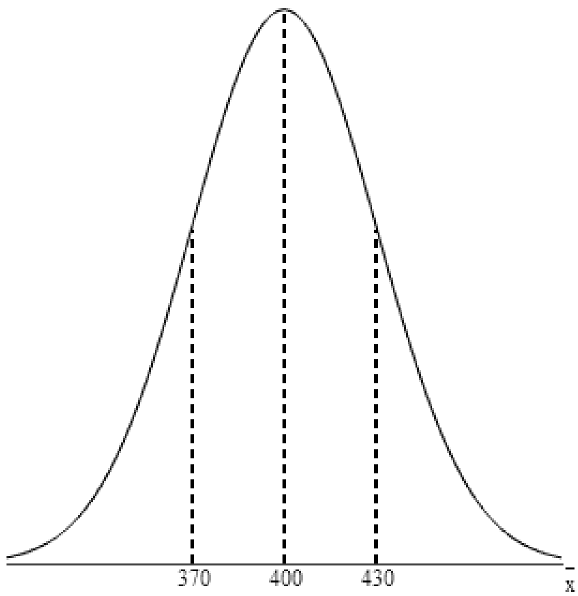
**11.** A study of the amount of time it takes a mechanic to rebuild the transmission for 2005 Chevy shows that the mean is 8.4 hours and the standard deviation is 1.8 hours. If 40 mechanics are randomly selected, find the probability that their mean rebuild time

*a)* Exceeds 8.7 hours.

*b)* Exceeds 8.1 hours.

**12.** A final exam in Math 160 has a mean of 73 with standard deviation 7.8. If 24 students are randomly selected, find the probability that the mean of their test scores is greater than 71.

**13.** The sampling distribution of the sample mean shown in the graph.




*a)* What is the value of $\mu_{\bar{x}}$ ?

*b)* What is the value of $\sigma_{\bar{x}}$ ?

*c)* If the sample size is $n = 9$, what is likely true about the shape of the population?

*d)* If the sample size is $n = 9$, what is the standard deviation of the population for which the sample was drawn?

**14.** A sample random of size $n = 81$ is obtained from a population with $\mu = 77$ *and* $\sigma = 18$.

*a)* Describe the sampling distribution of $\bar{x}$

*b)* What is $P(\bar{x} > 79.6)$?

*c)* What is $P(\bar{x} \le 72.5)$?

*d)* What is $P(75.1 < \bar{x} < 80.9)$?

**15.** The reading speed of second grade students is approximately normal, with a mean of 90 words per minute (*wpm*) and a standard deviation of 10 *wpm*.

*a)* What is the probability a randomly selected student will read more than 95 wpm?

*b)* What is the probability that a random sample of 11 second grade students results in a mean reading rate of more than 95 wpm?

*c)* What is the probability that a random sample of 22 second grade students results in a mean reading rate of more than 95 wpm?

*d)* What effect does increasing the sample size have on the probability?