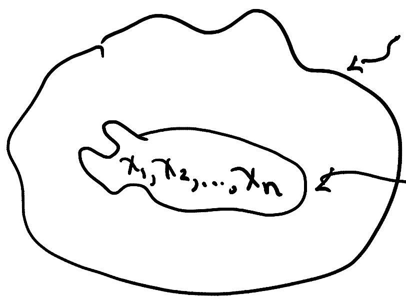
Chapter 6 (continued)

02/22/2023

Sampling Distributions and Estimators

One of the goals of statistical analysis is to make inferences about population characteristics based on information contained in a sample.



Population (μ, σ, p, ...) — proportion of "successes" in the population

Sample ($\bar{x}$, s, $\hat{p}$, ...)

NOTE: Values of population mean (μ), standard deviation (σ), and proportion (p) are consider fixed values but unknown.

Consider using the sample mean $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ to estimate the population mean μ.

$\bar{x}$ is often called a statistic. It is a quantity computed from the sample. μ is called a population parameter.

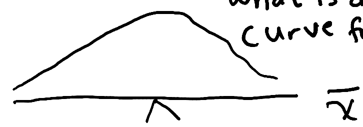
We use statististics to estimate parameters.

The sample mean, $\bar{x}$, varies from sample-to-sample and thus has a sampling distribution.

and thus has a sampling distribution.

<u>Def</u> : The <u>sampling distribution</u> of a statistic is a probability or density function that summarizes all of the possible values of the statistic.

What is density curve for $\bar{x}$?



<u>Ex</u> : x = blood platelet size (fl) of patient with non-cardiac chest pain.

NOTE: $1 m^3 = 1 \times 10^{18}$ fl

Suppose $X \sim N(\underset{\mu}{\underset{\shortparallel}{8.25 \text{ fl}}}, \underbrace{(0.75 \text{ fl})^2}_{\substack{\shortparallel \\ \sigma^2}})$

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

$n = 5$

From random sample 1, compute $\bar{x}$.
" " sample 2, compute $\bar{x}$.
⋮
From random sample 500, compute $\bar{x}$.

} Construct the histogram of the sampling distribution of $\bar{x}$.

Repeat above process with $n=10$, $n=20$, $n=30$.

Comment on the results. See Canvas Handout #3.

02/27/2023

<u>Ex</u> : x = blood tranfusion recipient stricken with viral hepatitis.

$$= \begin{cases} 1, & \text{if Yes} \\ 0, & \text{if No} \end{cases}$$

p = population proportion of blood recipients stricken with viral hepatitis.

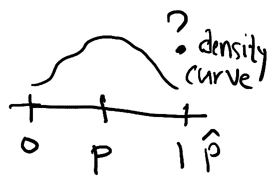
p = population proportion of blood recipients stricken with viral hepatitis.

$\hat{p}$ = sample proportion " " " " "

$= \dfrac{\text{number of Yes's in sample}}{n}$

$= \dfrac{\sum_{i=1}^{n} x_i}{n}$

What is the sampling distribution of $\hat{p}$?



Suppose $p = 0.07$.

$n = 10$

From random sample 1, compute $\hat{p}$

⋮

From random sample 500, compute $\hat{p}$.

Repeat process for $n = 25$, $n = 50$, $n = 100$.

Describe sampling distribution of $\hat{p}$.

See Handout #3 in Canvas.

General Properties of Sampling Distributions

1. Sampling distribution of a statistic is often centered at the value of the population parameter.
2. The variability of the sampling distribution of the statistic decreases as n increases.
3. As n increases, sampling distribution of many statistics become more and more bell-shaped.

The Sampling Distribution of $\bar{x}$ and the Central Limit Theorem (CLT)

<u>Result</u>: If $x_1, x_2, \ldots, x_n$ is a random sample from a population with mean μ and standard deviation σ, then

$$\mu_{\bar{x}} = \text{mean of the sampling distribution of } \bar{x}$$
$$= \mu$$

$$\sigma_{\bar{x}} = \text{standard deviation of the sampling distribution of } \bar{x}$$
$$= \frac{\sigma}{\sqrt{n}}$$
$$= \text{standard error of } \bar{x}\text{, where } \bar{x} \text{ is used to estimate } \mu.$$

<u>Result</u>: If the population from which the random sample is obtained is $N(\mu, \sigma^2)$, then

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Notation:
$\mu_{\bar{x}} = \mu$
$\sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \text{where } Z \sim N(0,1).$$

<u>Big Result</u>: Central Limit Theorem (CLT)

Suppose the shape of the population is unknown. When n is "large", then sampling distribution of $\bar{x}$ is approximately normally distributed. i.e.,

$\bar{x} \overset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$ when n is "large".

Demonstrate the CLT using java applet.

Rule of Thumb for CLT involving $\bar{x}$

CLT for $\bar{x}$ holds when $n > 30$, where the less symmetric the population is, the larger n will have to be to ensure normality for the sampling distribution of $\bar{x}$.

If n is sufficiently large, then

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

where $Z \overset{\text{approx}}{\sim} N(0,1)$.

Ex: X = weight of a newborn manatee.

suppose $\mu = 60$ lbs
$\sigma = 5$ lbs

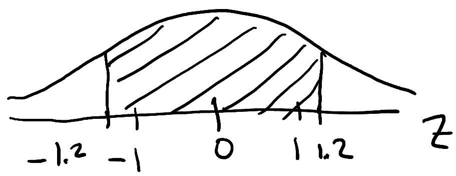
i.) Suppose a random sample of 36 newborn manatees is obtained. Find probability the sample mean weight is between

59 lbs and 61 lbs.

$n = 36$

$$P(59 \le \bar{X} \le 61) = P\left(\frac{59-60}{5/\sqrt{36}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{61-60}{5/\sqrt{36}}\right)$$

$$= P(-1.2 \le Z \le 1.2)$$



$$= P(Z \le 1.2) - P(Z < -1.2)$$

$$= 0.8849 - 0.1151$$

$$= 0.7698$$

ii) $n = 64$

$$P(59 \le \bar{X} \le 61) = P\left(\frac{59-60}{5/\sqrt{64}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{61-60}{5/\sqrt{64}}\right)$$

$$= P(-1.6 \le Z \le 1.6)$$

$$= 0.9452 - 0.0548$$

$$= 0.8904$$

Remark : If individual X-values are normally distributed, then above probabilities are exact.

If population of weights is not normally distributed, above probabilities are approximations of the true values.

iii) $n = 64$

using CLT

$$P(\bar{X} \ge 62) \doteq P\left(Z \ge \frac{62-60}{5/\sqrt{64}}\right) = P(Z \ge 3.2)$$

$= P(Z$

$P(x \geq 62) = P(Z \leq \quad /\sqrt{64}) = P(Z \leq -3.2 \quad)$

$= P(Z \leq -3.2)$

$= 0.0007$

<u>Remarks</u>: 1. $Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

2. $\bar{x} = \mu + z \cdot \dfrac{\sigma}{\sqrt{n}}$

The Sampling Distribution of the Sample Proportion ($\hat{p}$)

<u>Ex</u>: Smoking status (Yes, No)

p = Population proportion of smokers

$\hat{p}$ = sample proportion of smokers

n = sample size

x_i = response of i^{th} person in sample $= \begin{cases} 1, & \text{if "Yes"} \\ 0, & \text{if "No"} \end{cases}$

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

<u>Remark</u>: $0 \leq \hat{p} \leq 1$

$0 \leq p \leq 1$

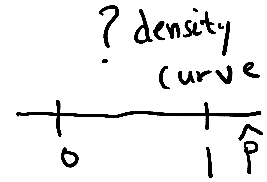
<u>Result</u>: If $x_1, x_2, \ldots, x_n$ is a random sample from a population where p = population proportion of "successes", then

$$\mu_{\hat{p}} = p$$

? density

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$



<u>Big Result</u> : CLT for $\hat{p}$

If $np \geq 5$ <u>and</u> $n(1-p) \geq 5$, then the CLT for $\hat{p}$ results in

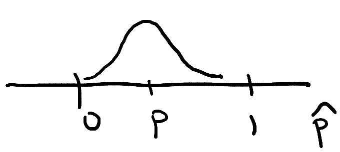
$$\hat{p} \overset{\text{approx}}{\sim} N\left(p, \frac{p(1-p)}{n}\right).$$

NOTE: $\mu_{\hat{p}} = p$, $\sigma^2_{\hat{p}} = \frac{p(1-p)}{n}$

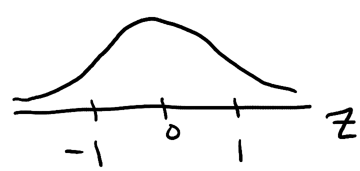
$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$$

where $Z \overset{\text{approx}}{\sim} N(0,1)$.

↑ Exam 2



$\Longleftrightarrow$



using the CLT for $\hat{p}$
