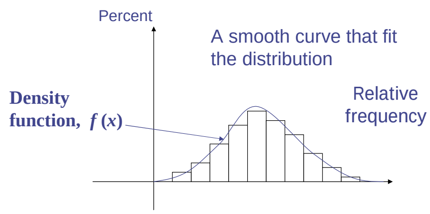
Modeling Continuous Distribution

Normal Distribution

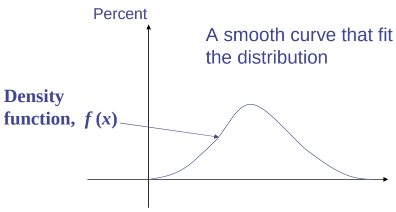
1

Density Curve



2

Density Curve

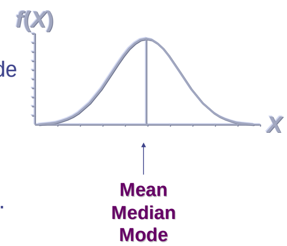


Use a mathematical model to describe the variable.

3

Normal Distribution

1. 'Bell-Shaped' & Symmetrical
2. Mean, Median, Mode Are Equal
3. Distribution Has Infinite Range ($-\infty < x < \infty$)
4. 99.7% within 3 s.d. of mean (Empirical Rule)



4

Normal Probability Density Function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{1}{2}\right)\left(\frac{x-\mu}{\sigma}\right)^2}$$

$f(x)$ = Density of Random Variable x
σ = Standard Deviation of the Distribution
π = 3.14159...; e = 2.71828...
x = Value of Random Variable ($-\infty < x < \infty$)
μ = Mean of the Distribution

Notation: $\mathcal{N}(\mu, \sigma) \Rightarrow$ A normal distribution with mean μ and standard deviation σ

5

Example

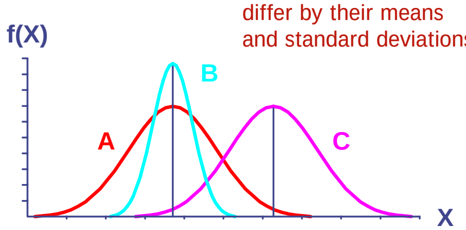
$\mathcal{N}(72, 5) \Rightarrow$ A normal distribution with mean 72 and standard deviation 5.
Possible situations: Test scores, pulse rates, …

$\mathcal{N}(130, 24) \Rightarrow$ A normal distribution with mean 130 and standard deviation 24.
Possible situations: Weight, Cholesterol levels, …

6

Effect of Varying Parameters (μ & σ)

Normal Distributions differ by their means and standard deviations.

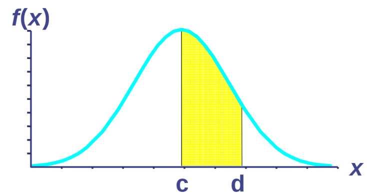


7

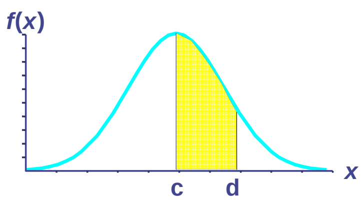
Normal Distribution Probability

Probability is area under curve!

$$P(c \le x \le d) = ?$$



8



$$P(c \le x \le d) = P(c < x < d)$$
$$= P(c \le x < d)$$
$$= P(c < x \le d)$$

9

Standard Normal Distribution

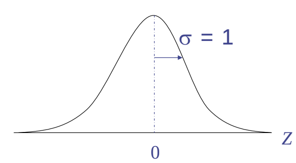
Standard Normal Distribution:

A normal distribution with *mean* = 0 and *standard deviation* = 1.

Notation:

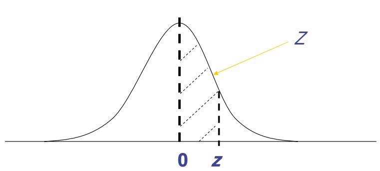
$Z \sim \mathcal{N}(\mu = 0, \sigma = 1)$

Cap letter **Z**



10

Area under Standard Normal Curve



How to find the proportion of the are under the standard normal curve below such as P (0 < Z < z) = ?

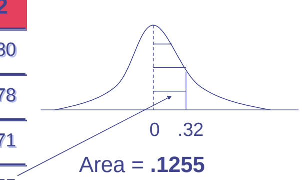
Use Standard Normal Table!!!

11

Standard Normal Distribution

P(0 < Z < 0.32) = Area between 0 and .32 = .1255

Z	.00	.01	**.02**
0.0	.0000	.0040	.0080
0.1	.0398	.0438	.0478
0.2	.0793	.0832	.0871
0.3	.1179	.1217	**.1255**

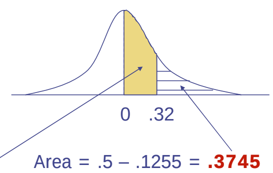


12

Standard Normal Distribution

P(Z > 0.32) = Area above .32 = **.3745**

Z	.00	.01	.02
0.0	.0000	.0040	.0080
0.1	.0398	.0438	.0478
0.2	.0793	.0832	.0871
0.3	.1179	.1217	.1255

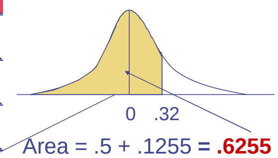


13

Standard Normal Distribution

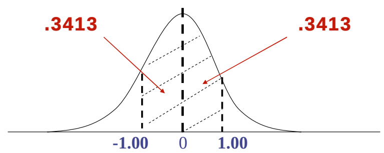
P(Z< 0.32) = Area below .32 = **.6255**

Z	.00	.01	.02
0.0	.0000	.0040	.0080
0.1	.0398	.0438	.0478
0.2	.0793	.0832	.0871
0.3	.1179	.1217	.1255



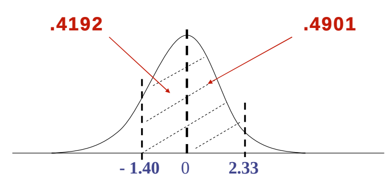
14

$P(-1.00 < Z < 1.00) =$ **.6826**



15

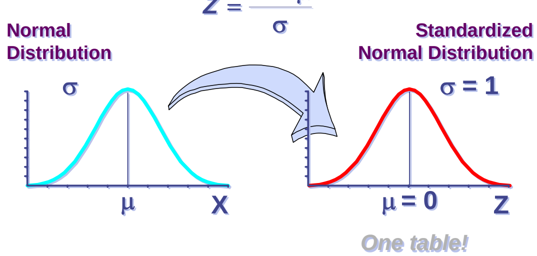
$P(-1.40 < Z < 2.33) =$ **.9093**



16

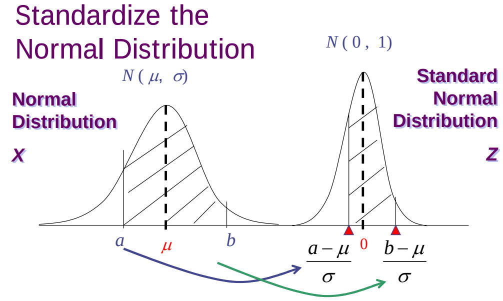
Standardize the Normal Distribution

$$Z = \frac{X - \mu}{\sigma}$$



17

Standardize the Normal Distribution



$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

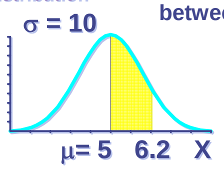
18

Standardizing Example

For a normal distribution that has a mean = 5 and s.d. = 10, what percentage of the distribution is between 5 and 6.2?

Normal Distribution

σ = 10



19

Standardizing Example

$$Z = \frac{X - \mu}{\sigma} = \frac{6.2 - 5}{10} = .12$$

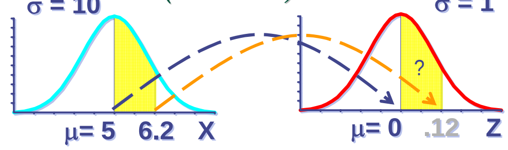
Normal Distribution

σ = 10

$P(5 \le X \le 6.2)$
$= P(0 \le Z \le .12)$

Standardized Normal Distribution

σ = 1



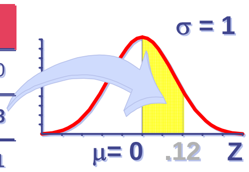
20

Obtaining the Probability

Standardized Normal Probability Table (Portion)

Z	.00	.01	.02
0.0	.0000	.0040	.0080
0.1	.0398	.0438	**.0478**
0.2	.0793	.0832	.0871
0.3	.1179	.1217	.1255

σ = 1



Area = **.0478**

P(5 < X < 6.2) = .0478

21

Example $P(3.8 \le X \le 5)$

$$Z = \frac{X - \mu}{\sigma} = \frac{3.8 - 5}{10} = -.12$$

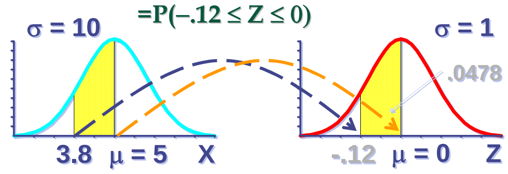
Normal Distribution

σ = 10

$P(3.8 \le X \le 5)$
$= P(-.12 \le Z \le 0)$

Standardized Normal Distribution

σ = 1



Area = **.0478**

22

Example $P(2.9 \le X \le 7.1)$

$$Z = \frac{X - \mu}{\sigma} = \frac{2.9 - 5}{10} = -.21$$

$$Z = \frac{X - \mu}{\sigma} = \frac{7.1 - 5}{10} = .21$$

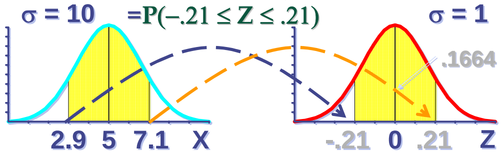
Normal Distribution

σ = 10

$P(2.9 \le X \le 7.1)$
$= P(-.21 \le Z \le .21)$

Standardized Normal Distribution

σ = 1



Area = .0832 + .0832 = **.1664**

Example $P(X > 8)$

$$Z = \frac{X - \mu}{\sigma} = \frac{8 - 5}{10} = .30$$

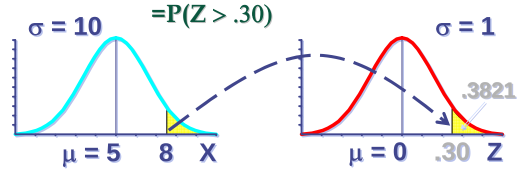
Normal Distribution

σ = 10

$P(X > 8)$
$= P(Z > .30)$

Standardized Normal Distribution

σ = 1



Area = **.3821**

24

More on Normal Distribution

The work hours per week for residents in Ohio has a **normal distribution** with μ **= 42** hours & σ **= 9** hours. Find the percentage of Ohio residents whose work hours are

A. between 42 & **60** hours.

$P(42 \le X \le 60) = ?$

B. less than 20 hours.

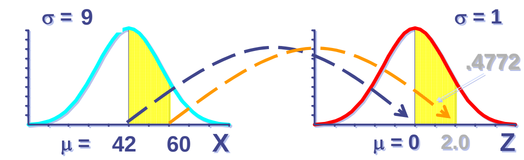
$P(X \le 20) = ?$

25

Solution*
P(42 ≤ X ≤ 60)

$$Z = \frac{X-\mu}{\sigma} = \frac{60-42}{9} = 2$$

Normal Distribution $P(0 \le Z \le 2) = .4772$ **Standardized Normal Distribution**

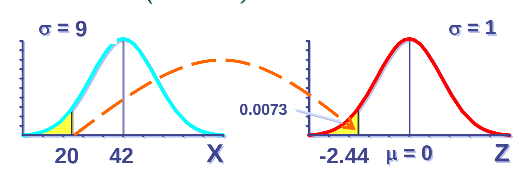


26

Solution*
P(X ≤ 20)

$$Z = \frac{X-\mu}{\sigma} = \frac{20-42}{9} = -2.44$$

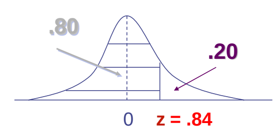
Normal Distribution $P(Z \le -2.44) = 0.0073$ **Standardized Normal Distribution**



27

Finding Z Values for Known Probabilities

What is z given P(Z < z) = .80 ?



Upper Tail Area =**.80 - .5**
= **.30**
z = .84

Standardized Normal Probability Table (Portion)

Z	…	.04	.05
⋮	.2357	.2389	.2422
0.7	.2673	.2704	.2734
0.8	.2967	.2995	.3023
0.9	.3238	.3264	.3289

28

Finding X Values for Known Probabilities

Example: The weight of new born infants is normally distributed with a mean 7 lb and standard deviation of 1.2 lb. Find the 80th percentile.

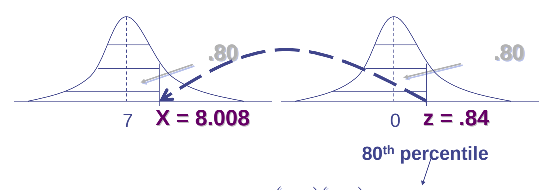
Area to the left of 80th percentile in 0.200. In the table there is a area value 0.200 corresponding to a **z**-score of **.84**.

80th percentile = 7 + .84 x 1.2 = **8.008** lb

29

Finding X Values for Known Probabilities

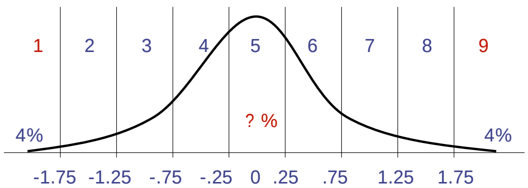
Normal Distribution **Standardized Normal Distribution**



$$X = \mu + Z \cdot \sigma = 7 + (.84)(1.2) = 8.008$$

30

Stanine Score



31

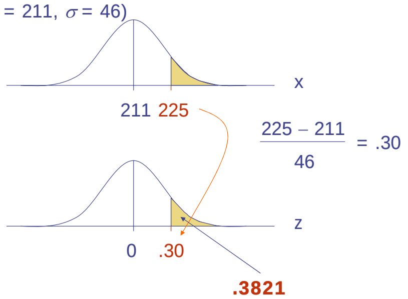
A Normal Distribution

Example: Consider the distribution of serum cholesterol levels for 40- to 70-year-old males living in community A has a **mean** of **211** mg/100 ml, and the **standard deviation** of **46** mg/100 ml. If an individual is selected from this population, what is the probability that his/her serum cholesterol level is **higher than 225**?

32

P(X > 225) = ?

$X \sim N(\mu = 211, \sigma = 46)$



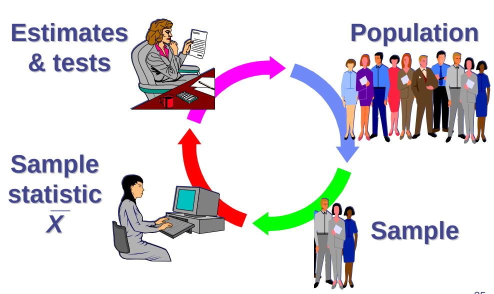
33

Statistical Inference

- Estimation
- Hypothesis Testing

34

Inference Process



35

Estimators

1. Statistics Used for Estimating Population Parameters
   - Statistic: descriptive measure of a sample. (sample mean, sample s.d., sample median)
   - Parameter: descriptive measure of a population. (population mean, population s.d., ...)
2. Theoretical Basis Is Sampling Distribution

36

Population Parameters & Their Estimators

Estimate Population Parameter		with Sample Statistic
Mean	μ	$\bar{x}$
Proportion	p	$\hat{p}$
Variance	σ^2	s^2
Differences	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$

37

Sampling Distribution

Theoretical Probability Distribution of the **Sample Statistic.**

What is the Shape of this distribution?

What are the values of the parameters such as mean and standard deviation?

38

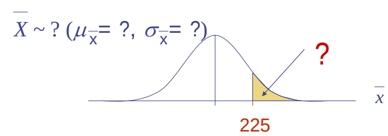
Probability Related to Mean

Example: Consider the distribution of serum cholesterol levels for 40- to 70-year-old males living in community A has a mean of **211** mg/100 ml, and the standard deviation of **46** mg/100 ml. If a **random sample of 100** individuals is taken from this population, what is the probability that the average serum cholesterol level of these 100 individuals is **higher than 225**?

39

$P(\bar{X} > 225) = ?$

What is probability that mean of the sample is greater than 225?



What is the distribution of sample mean?

40

Parameters of the Sampling Distribution

If a random sample is taken from a population that has a mean μ and a standard deviation σ, the sampling distribution of the sample mean, $\bar{x}$, will have a mean that is the same as the population mean, and will have a standard deviation that is equal to the standard deviation of the population divided by the square root of the sample size.

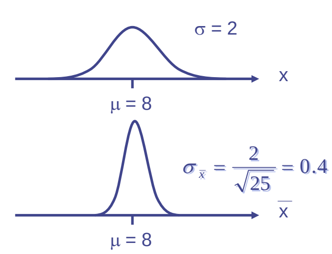
$$\mu_{\bar{x}} = \mu \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

41

Sampling Distribution

Population Distribution

Sampling Distribution of Mean (sample size 25)



42

Standard Error of Mean

1. Formula $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$
2. Standard Deviation of the sampling distribution of the Sample Means, $\bar{X}$
3. Less Than Pop. Standard Deviation

$$\frac{\sigma}{\sqrt{n}} < \sigma$$

43

What is the shape of the distribution?

$$\bar{X} \sim ?\ (\mu_{\bar{x}} = \mu,\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}})$$

44

Sampled from Normal Population

If a **random sample** is taken from a normal population (or normal distribution) that has a mean μ and a standard deviation σ, the distribution of the sample means is **normal** with

$$\bar{X} \sim N \qquad \mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

45

$P(\bar{X} > 225) = ?$

Cholesterol Level has a mean 211, s.d. 46.

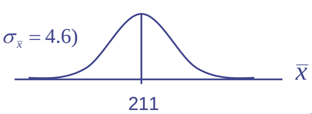
Parameters of the sampling distribution of the mean: $\mu_{\bar{x}} = \mu = 211$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{46}{\sqrt{100}} = 4.6$$

If the population is normally distributed, the sampling distribution of the mean is normally distributed.

$\bar{X} \sim N(\mu_{\bar{x}} = 211, \sigma_{\bar{x}} = 4.6)$

$n = 100$



46

Central Limit Theorem

What if the population sampled is not normally distributed?

47

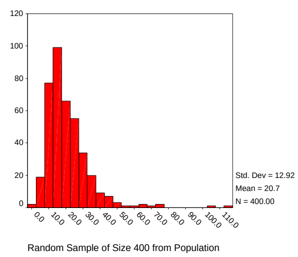
Central Limit Theorem

If a **relative large random sample** is taken from a population that has a mean μ and a standard deviation σ, **regardless of the distribution of the population**, the distribution of the sample means is **approximately normal** with

$$\bar{X} \sim N \qquad \mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

48

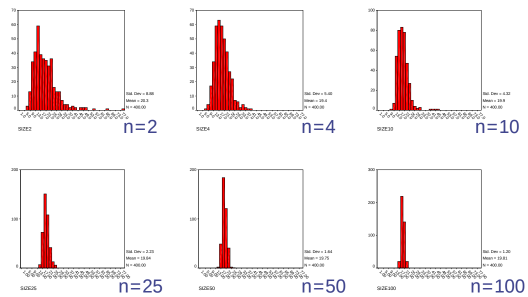
A Random Sample from Population



Population mean = 19.9, standard deviation = 12.6

49

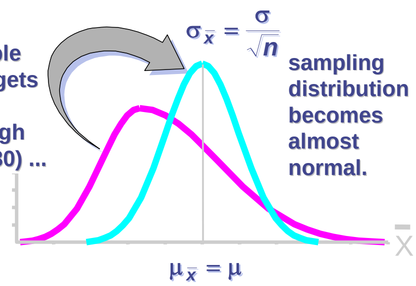
Simulated Sampling Distribution of Means



50

Central Limit Theorem

As sample size gets large enough ($n \geq 30$) ...



sampling distribution becomes almost normal.

51

Probability Related to Mean

Example: Consider the distribution of serum cholesterol levels for 40- to 70-year-old males living in community A has a mean of **211** mg/100 ml, and the standard deviation of **46** mg/100 ml. If a **random sample of 100** individuals is taken from this population, <u>what is the probability that the average serum cholesterol level of these 100 individuals is **higher than 225**?</u>

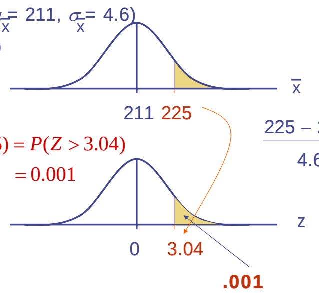
52

$P(\bar{X} > 225) = ?$

Cholesterol Level has a mean 211, s.d. 46.

$\bar{X} \sim N(\mu_{\bar{x}} = 211,\ \sigma_{\bar{x}} = 4.6)$

$n = 100$



$\frac{225 - 211}{4.6} = 3.04$

$P(\bar{X} > 225) = P(Z > 3.04)$
$= 0.001$

53