Basic Statistics 02

Probability Distributions

Probability Distributions

1. Probability Distributions
   - Random variable and probability distribution.
   - The mean, variance, and standard deviation of a (discrete) probability distribution.
2. Normal Distribution
   - Normal & Standard Normal Distribution
   - Calculating *z* value
   - Determining the probability

Random Variables & Probability Dist.

- A ***random variable*** is a numerical value determined by the outcome of an experiment.
- A ***probability distribution*** is the listing of all possible outcomes of an experiment and the corresponding probability.
  - The main features of a probability distribution are:
    - The sum of the probabilities of the various outcomes is 1.
    - The probability of a particular outcome is between 0 and 1.

The Mean of a Probability Dist.

- The ***mean***:
  - is the long-run average value of the random variable.
  - is also referred to as its expected value, $E(X)$, in a probability distribution.

$$\mu = \Sigma[xP(x)] = E[x]$$

- where μ is the mean and $P(x)$ is the probability of the various outcomes x.

The Variance of a Probability Dist.

- The variance measures the amount of spread (variation) of a distribution.
- The variance of a (discrete) distribution is denoted by the Greek letter σ^2 (sigma squared).

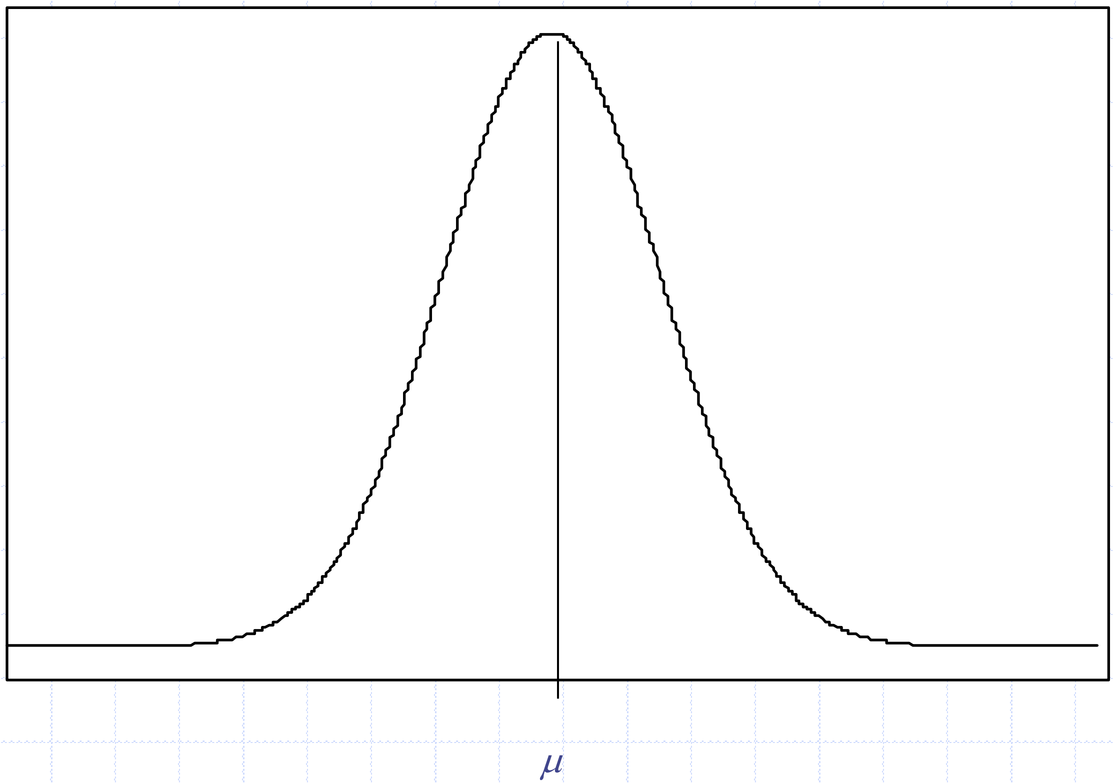
$$\sigma^2 = \Sigma[(x-\mu)^2 P(x)]$$

$$= E[(x-\mu)^2] = Var[x]$$

Characteristics of a Normal Dist.

- It is ***bell-shaped*** and has a single peak at the exact center of the distribution.
- It is ***symmetrical*** about its mean.
  - The arithmetic mean is located at the peak.
  - Half the area under the curve is above the mean.
- It is ***asymptotic***, that is the curve gets closer to the *X*-axis but never actually touches it.

Normal Distribution



Normal curve is symmetrical

Theoretically, curve extends to infinity

Mean, median, and mode are equal

The Standard Normal Dist.

- The ***standard normal distribution*** has a mean of 0 and a standard deviation of 1.
- A ***z-value*** is the distance between a selected value, designated X, and the population mean μ, divided by the population standard deviation, σ. The formula is:

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

EXAMPLE 1

The monthly starting salaries of recent MBA graduates follows the normal distribution with a mean of $2,000 and a standard deviation of $200. What is the ***z-value*** for a salary of $2,200?

$$z = \frac{X - \mu}{\sigma} = \frac{\$2,200 - \$2,000}{\$200} = 1.00$$

What is the z-value corresponding to $1,600?

$$z = \frac{X - \mu}{\sigma} = \frac{\$1,600 - \$2,000}{\$200} = -2.0$$

Areas Under the Normal Curve

About 68 percent of the area under the normal curve is within one standard deviation of the mean.

$$\mu \pm \sigma$$

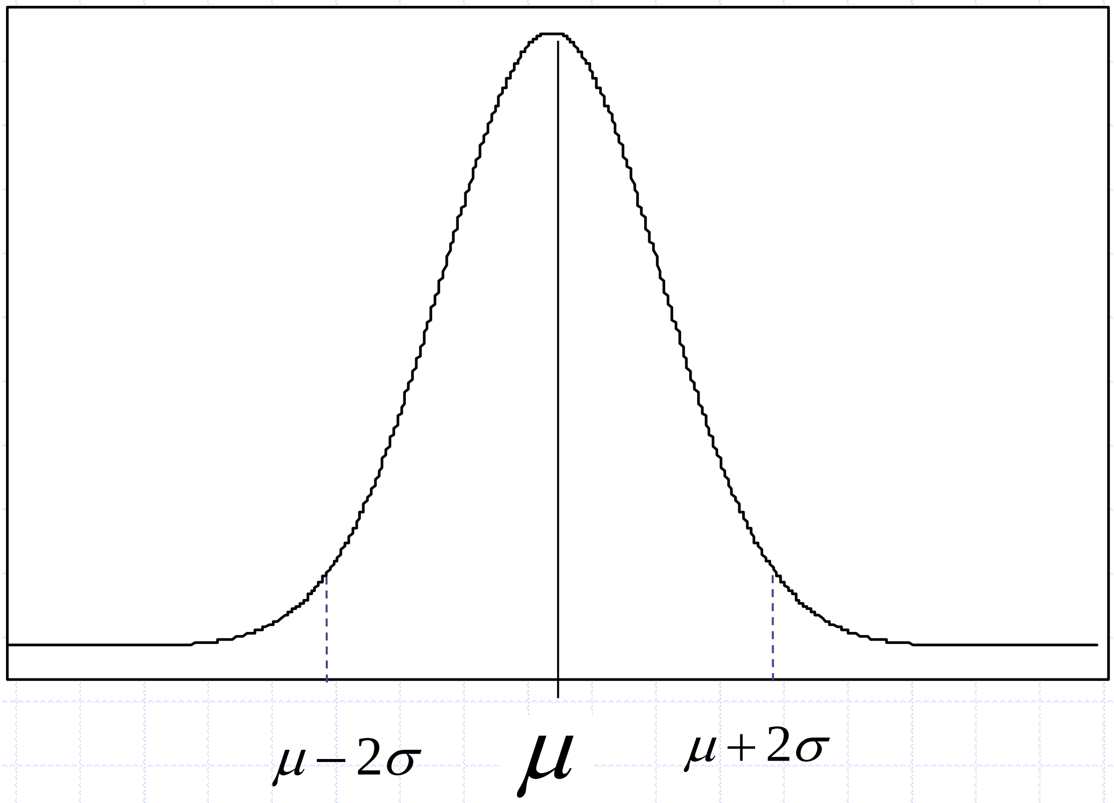
About 95 percent is within two standard deviations of the mean.

$$\mu \pm 2\sigma$$

Practically all is within three standard deviations of the mean.

$$\mu \pm 3\sigma$$

Areas Under the Normal Curve



Between:

$\pm 1\sigma$ - 68.26%

$\pm 2\sigma$ - 95.44%

$\pm 3\sigma$ - 99.74%

EXAMPLE 2

The daily water usage per person in a city follows a normal distribution with a mean of 20 gallons and a standard deviation of 5 gallons. About 68 percent of those living in a city will use how many gallons of water?

About 68% of the daily water usage will lie between 15 and 25 gallons since 20 ± 5.

EXAMPLE 3

What is the probability that a person from this city selected at random will use between 20 and 24 gallons per day?

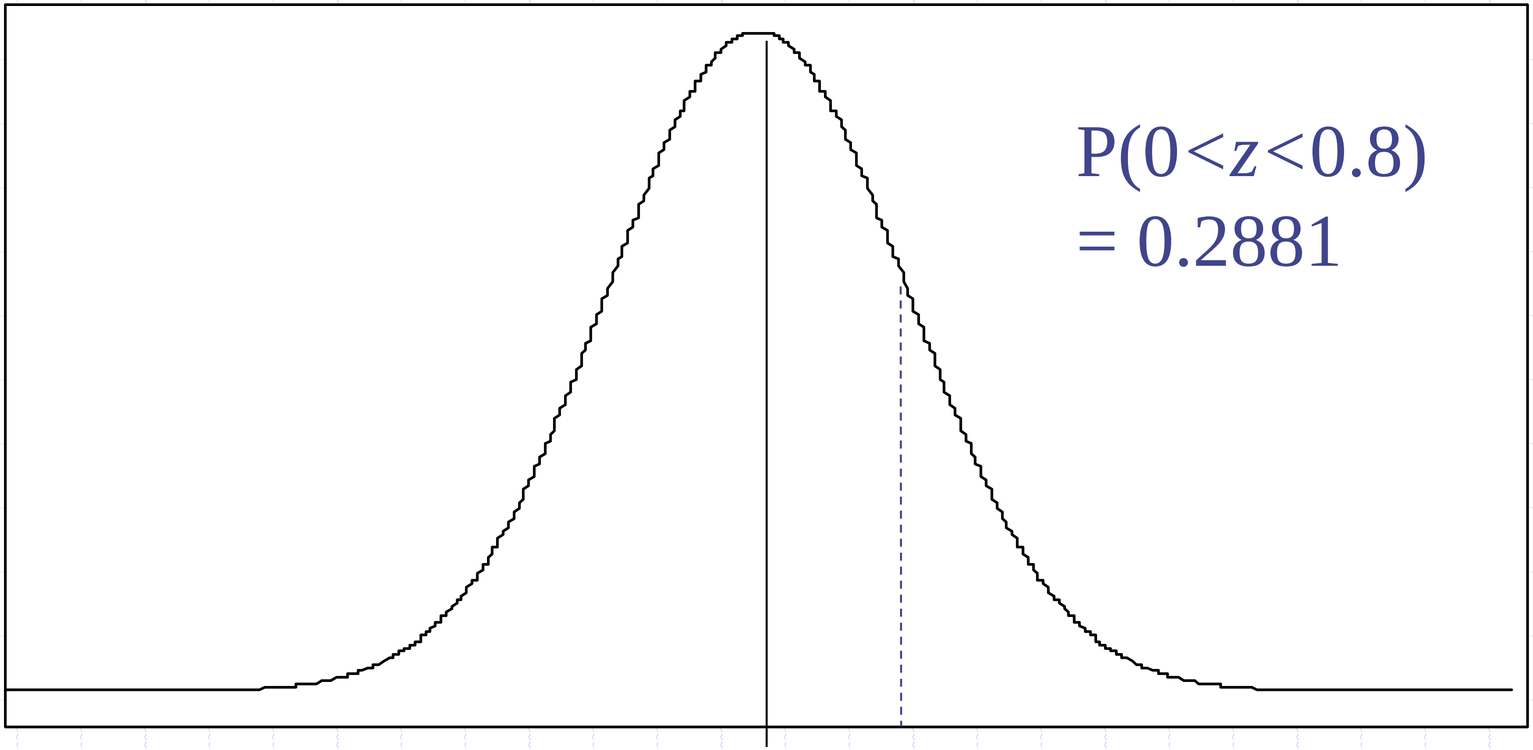
$$z = \frac{X - \mu}{\sigma} = \frac{20 - 20}{5} = 0.00$$

$$z = \frac{X - \mu}{\sigma} = \frac{24 - 20}{5} = 0.80$$

Example 3 *continued*

- The area under a normal curve between a *z*-value of 0 and a *z*-value of 0.80 is 0.2881 (=0.7881-0.5, see p.351)
- We conclude that 28.81 percent of the residents use between 20 and 24 gallons of water per day.
- See the following diagram.

EXAMPLE 3 *continued*



$0<x<0.8$

EXAMPLE 3 *continued*

- What percent of the population use between 18 and 26 gallons per day?

$$z = \frac{X - \mu}{\sigma} = \frac{18 - 20}{5} = -0.40$$

$$z = \frac{X - \mu}{\sigma} = \frac{26 - 20}{5} = 1.20$$

Example 3 *continued*

- The area associated with a *z*-value of –0.40 is 0.1554.
- The area associated with a *z*-value of 1.20 is 0.3849.
- Adding these areas, the result is 0.5403.
- We conclude that 54.03 percent of the residents use between 18 and 26 gallons of water per day.