Marmara University, **2021**

Probability and Statistics

Subject 6

The Normal Probability Distribution

Mujdat Soyturk, Ph.D.
Associate Professor

Contents

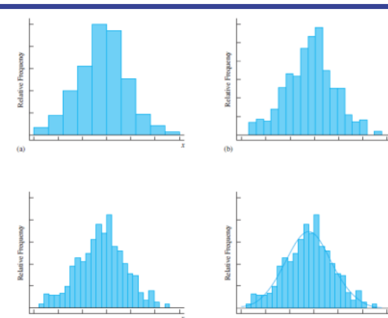
- Probability Distributions for Continuous Random Variables
- The Normal Probability Distribution
- Tabulated Areas of the Normal Probability Distribution
- The Normal Approximation to the Binomial Probability Distribution

Most parts of the slides are derived from the textbook: "Mendenhall, Beaver, Beaver, *Introduction to Probability and Statistics*, 14th Ed., Brooks/Cole, Cengage Learning, 2013"

Continuous Random Variables

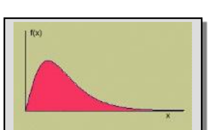
- Continuous random variables can assume the infinitely many values corresponding to points on a line interval.
- Examples:
  - heights, weights
  - length of life of a particular product
  - experimental laboratory error

Continuous Random Variables



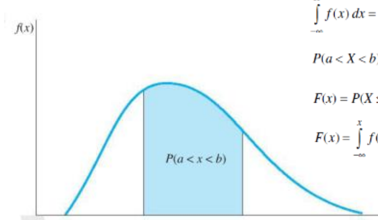
Continuous Random Variables

A **smooth curve** describes the probability distribution of a continuous random variable.



The depth or density of the probability, which varies with ***x***, may be described by a mathematical formula ***f(x)***, called the **probability distribution** or **probability density function** (***pdf***) for the random variable *x*.

Continuous Random Variables



$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

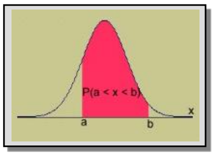
$$P(a < X < b) = \int_{b}^{a} f(x)\,dx$$

$$F(x) = P(X \le x)$$

$$F(x) = \int_{-\infty}^{x} f(x)\,dx$$

Properties of Continuous Probability Distributions

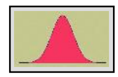
- The area under the curve is equal to **1.**
- P($a \le x \le b$) = **area under the curve** between *a* and *b*.

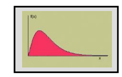


There is no probability attached to any single value of *x*. That is, ***P(x = a) = 0.***

Continuous Probability Distributions

- There are many different types of continuous random variables
- We try to pick a model that
  - Fits the data well
  - Allows us to make the best possible inferences using the data.
- One important continuous random variable is the **normal random variable**.





The Normal Distribution

The formula that generates the normal probability distribution is:

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

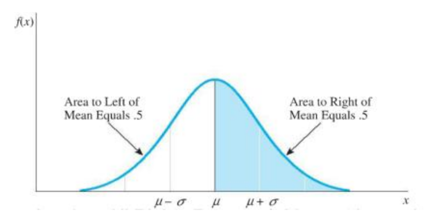
$$e = 2.7183 \qquad \pi = 3.1416$$

μ and σ are the population mean and standard deviation.

The ***shape*** of the normal curve changes as the ***standard deviation*** changes.

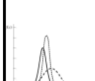
The Normal Distribution

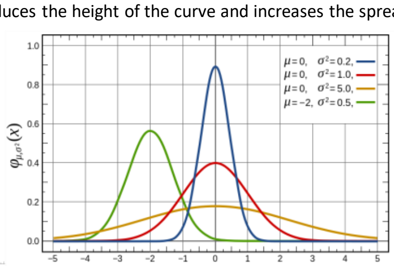
The mean μ locates the ***center*** of the distribution, and the distribution is ***symmetric*** about its mean μ.



The Normal Distribution

- The shape of the distribution is determined by σ. Large values of σ reduces the height of the curve and increases the spread.





The Standard Normal Distribution

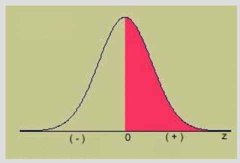
- To find **P(a < *x* < b)**, we need to find the area under the appropriate normal curve.
- There are infinitely large number of normal distributions, one for each different μ and σ.

  → the use of separate table of areas for each curve is impractical.
- To simplify the tabulation of these areas, we **standardize** each value of ***x*** by expressing it as a ***z-score***, the number of standard deviations σ it lies from the mean μ.

$$z = \frac{x-\mu}{\sigma} \quad \text{or} \quad x = \mu + z\sigma$$

The Standardized Normal (z) Distribution

IMPORTANT:
Standardization procedure allows to use the same table for all normal distributions.

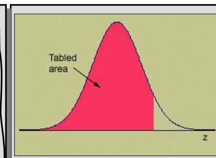


- Mean = 0; Standard deviation = 1
- When $x = \mu$, $z = 0$
- Symmetric about $z = 0$
- Values of z to the left of center are negative
- Values of z to the right of center are positive
- Total area under the curve is 1.

Using Standardized Normal Distribution Table

The four digit probability in a particular row and column of Table 3 gives the area under the z curve to the left that particular value of z.

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8328	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278



Area for z = 1.36

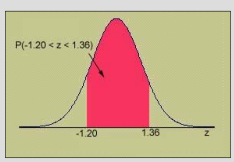
Example

Use Table 3 to calculate these probabilities:

$P(z \leq 1.36) = .9131$

$P(z > 1.36)$
$= 1 - .9131 = .0869$

$P(-1.20 \leq z \leq 1.36) =$
$.9131 - .1151 = .7980$

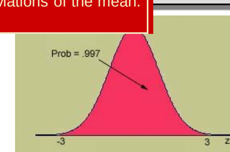


Using Table 3

- To find an area to the left of a z-value, find the area directly from the table.
- To find an area to the right of a z-value, find the area in Table 3 anc
- To find the ar , find the two areas in Table he other.

Remember the Empirical Rule: Approximately 95% of the measurements lie within 2 standard deviations of the mean.

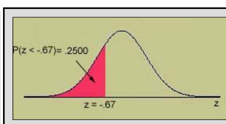
$P(-3 \leq z \leq 3)$
$= .9987 -$
$.0013 = .9974$



Working Backwards

Find the value of z that has area 0.25 to its left.

1. Look for the four digit area closest to .2500 in Table 3.
2. What row and column does this value correspond to?
3. $z = -.67$



4. What percentile does this value represent?

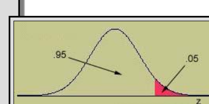
25th percentile, or 1st quartile (Q_1)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08
		0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635
		0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894
		0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177
		[illegible]	.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483
			3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810

Working Backwards

Find the value of z that has area 0.05 to its right.

1. The area to its left will be 1 - .05 = .95
2. Look for the four digit area closest to .9500 in Table 3.
3. Since the value .9500 is halfway between .9495 and .9505, we choose z halfway between 1.64 and 1.65.
4. $z = 1.645$



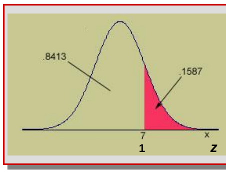
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Finding Probabilities for the General Normal Random Variable

- To find an area for a normal random variable x with mean μ and standard deviation σ, *standardize or rescale* the interval in terms of z.
- Find the appropriate area using Table 3.

Example: x has a normal distribution with $\mu = 5$ and $\sigma = 2$. Find $P(x > 7)$.

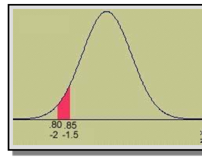
$$P(x>7) = P(z > \frac{7-5}{2})$$
$$= P(z>1) = 1 - .8413 = .1587$$



Example

The weights of packages of ground beef are normally distributed with mean *1 kg* and standard deviation *0.10*. What is the probability that a randomly selected package weighs between 0.80 and 0.85 kg?

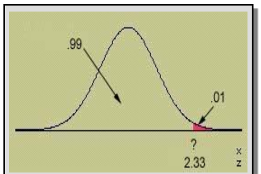
$$P(.80 < x < .85) =$$
$$P(-2 < z < -1.5) =$$
$$.0668 - .0228 = .0440$$



Example

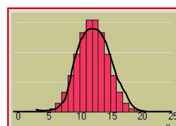
What is the weight of a package such that only 1% of all packages exceed this weight?

$$P(x > ?) = .01$$
$$P(z > \frac{?-1}{.1}) = .01$$
$$\text{From Table 3, } \frac{?-1}{.1} = 2.33$$
$$? = 2.33(.1) + 1 = 1.233$$



The Normal Approximation to the Binomial

- We can calculate binomial probabilities using
  - The binomial formula
  - The cumulative binomial tables
  - Java applets
- When n is large, and p is not too close to zero or one, areas under the normal curve with mean np and variance npq can be used to approximate binomial probabilities.



Approximating the Binomial

- Make sure to include the entire rectangle for the values of x in the interval of interest. This is called the **continuity correction.**
- Standardize the values of x using

$$z = \frac{x - np}{\sqrt{npq}}$$

- Make sure that ***np* and *nq* are both greater than 5** to avoid inaccurate approximations!

Example

Suppose x is a binomial random variable with $n = 30$ and $p = 0.4$. Using the normal approximation to find $P(x \le 10)$.

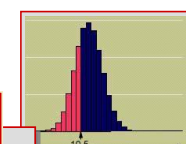
***n* = 30 *p* = .4 *q* = .6**

***np* = 12 *nq* = 18**

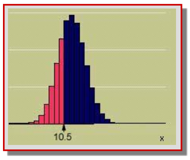
The normal approximation is ok!

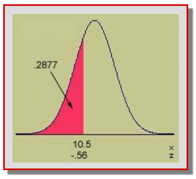
Calculate

$$\mu = np = 30(.4) = 12$$
$$\sigma = \sqrt{npq} = \sqrt{30(.4)(.6)} = 2.683$$



Example





$$P(x \le 10) \approx P(z \le \frac{10.5-12}{2.683})$$
$$= P(z \le -.56) = .2877$$

Example

A production line produces AA batteries with a reliability rate of 95%. A sample of n = 200 batteries is selected. Find the probability that at least 195 of the batteries work.

Success = working battery n = 200

***p* = .95 *np* = 190 *nq* = 10**

The normal approximation is ok!

$$P(x \ge 195) \approx P(z \ge \frac{194.5-190}{\sqrt{200(.95)(.05)}})$$
$$= P(z \ge 1.46) = 1 - .9279 = .0721$$

Key Concepts

I. Continuous Probability Distributions

1. Continuous random variables
2. Probability distributions or probability density functions
   a. Curves are smooth.
   b. The area under the curve between a and b represents the probability that x falls between a and b.
   c. $P(x = a) = 0$ for continuous random variables.

II. The Normal Probability Distribution

1. Symmetric about its mean μ.
2. Shape determined by its standard deviation σ.

Key Concepts

III. The Standard Normal Distribution

1. The normal random variable z has mean 0 and standard deviation 1.
2. Any normal random variable x can be transformed to a standard normal random variable using

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

3. Convert necessary values of x to z.
4. Use Table 3 in Appendix I to compute standard normal probabilities.
5. Several important z-values have tail areas as follows:

Tail Area:	.005	.01	.025	.05	.10
z-Value:	2.58	2.33	1.96	1.645	1.28