Chapter 6: The Normal Distribution

© FAROUQ ALAM, Ph.D. (2018)
Department of Statistics, KAU

**Textbook:**
The normal distribution is AKA a bell curve or a Gaussian distribution curve.

Many continuous variables (e.g. weights, heights) have distributions that are bell-shaped, and these are called approximately normally distributed variables.
If a random variable has a probability distribution whose graph is continuous, symmetric, and bell-shaped, it is called a normal distribution. The graph is called the normal distribution curve.

The position and shape of a normal distribution curve depend on the mean (μ) and the standard deviation (σ), respectively.
6 – 1: Normal Distributions (cont.)

(a) Same means but different standard deviations

(b) Different means but same standard deviations

(c) Different means and different standard deviations

\( \mu_1 = \mu_2 \)

\( \sigma_1 > \sigma_2 \)
Summary of the Properties of the Theoretical Normal Distribution (page 314)

- Mean = Median = Mode
- A normal distribution curve is unimodal.
- The curve never touches the x axis.
- The total area under a normal distribution curve is equal to 1.00.
Summary of the Properties of the Theoretical Normal Distribution (cont.)

- About 68% of the data falls within one standard deviation of the mean ($\mu \pm \sigma$).
- About 95% of the data falls within two standard deviations of the mean ($\mu \pm 2\sigma$).
- About 99.7% of the data falls within three standard deviations of the mean ($\mu \pm 3\sigma$).
The **standard** normal distribution is a normal distribution with **a mean of** $\mu = 0$ and **a standard deviation of** $\sigma = 1$.

**All** normally distributed variables $X$ can be transformed into the standard normally distributed variable $Z$ by using the formula for the standard score:

$$Z = \frac{X - \mu}{\sigma}$$
The Standard Normal Distribution (cont.)
Finding the Area Under the Standard Normal Distribution Curve

1. To the left of any \( z \) value:
   Look up the \( z \) value in the table and use the area given.

2. To the right of any \( z \) value:
   Look up the \( z \) value and subtract the area from 1.

\[
P(Z < -z) \quad \text{or} \quad P(Z < +z) \]

\[
P(Z > -z) \quad \text{or} \quad P(Z > +z)
\]

3. Between any two \( z \) values:
   Look up both \( z \) values and subtract the corresponding areas.

\[
P(-z < Z < +z) \quad \text{or} \quad P(z_1 < Z < z_2) \quad \text{or} \quad P(-z_1 < Z < -z_2)
\]
Find the area to the left of $z = 2.09 = 2.0 + 0.09$. 
Table E (continued)

Cumulative Standard Normal Distribution

<table>
<thead>
<tr>
<th>z</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>.5000</td>
<td>.5040</td>
<td>.5080</td>
<td>.5120</td>
<td>.5160</td>
<td>.5199</td>
<td>.5239</td>
<td>.5279</td>
<td>.5319</td>
<td>.5359</td>
</tr>
<tr>
<td>0.1</td>
<td>.5398</td>
<td>.5438</td>
<td>.5478</td>
<td>.5517</td>
<td>.5557</td>
<td>.5596</td>
<td>.5636</td>
<td>.5675</td>
<td>.5714</td>
<td>.5753</td>
</tr>
<tr>
<td>0.2</td>
<td>.5793</td>
<td>.5832</td>
<td>.5871</td>
<td>.5910</td>
<td>.5948</td>
<td>.5987</td>
<td>.6026</td>
<td>.6064</td>
<td>.6103</td>
<td>.6141</td>
</tr>
<tr>
<td>0.3</td>
<td>.6179</td>
<td>.6217</td>
<td>.6255</td>
<td>.6293</td>
<td>.6331</td>
<td>.6368</td>
<td>.6406</td>
<td>.6443</td>
<td>.6480</td>
<td>.6517</td>
</tr>
<tr>
<td>0.4</td>
<td>.6554</td>
<td>.6591</td>
<td>.6628</td>
<td>.6664</td>
<td>.6700</td>
<td>.6736</td>
<td>.6772</td>
<td>.6808</td>
<td>.6844</td>
<td>.6879</td>
</tr>
<tr>
<td>0.5</td>
<td>.6915</td>
<td>.6950</td>
<td>.6985</td>
<td>.7019</td>
<td>.7054</td>
<td>.7088</td>
<td>.7123</td>
<td>.7157</td>
<td>.7190</td>
<td>.7224</td>
</tr>
<tr>
<td>0.6</td>
<td>.7257</td>
<td>.7291</td>
<td>.7324</td>
<td>.7357</td>
<td>.7389</td>
<td>.7422</td>
<td>.7454</td>
<td>.7486</td>
<td>.7517</td>
<td>.7549</td>
</tr>
<tr>
<td>0.7</td>
<td>.7580</td>
<td>.7611</td>
<td>.7642</td>
<td>.7673</td>
<td>.7704</td>
<td>.7734</td>
<td>.7764</td>
<td>.7794</td>
<td>.7823</td>
<td>.7852</td>
</tr>
<tr>
<td>0.8</td>
<td>.7881</td>
<td>.7910</td>
<td>.7939</td>
<td>.7967</td>
<td>.7995</td>
<td>.8023</td>
<td>.8051</td>
<td>.8078</td>
<td>.8106</td>
<td>.8133</td>
</tr>
<tr>
<td>0.9</td>
<td>.8159</td>
<td>.8186</td>
<td>.8212</td>
<td>.8238</td>
<td>.8264</td>
<td>.8289</td>
<td>.8315</td>
<td>.8340</td>
<td>.8365</td>
<td>.8390</td>
</tr>
<tr>
<td>1.0</td>
<td>.8413</td>
<td>.8438</td>
<td>.8461</td>
<td>.8485</td>
<td>.8508</td>
<td>.8531</td>
<td>.8554</td>
<td>.8577</td>
<td>.8599</td>
<td>.8621</td>
</tr>
<tr>
<td>1.1</td>
<td>.8643</td>
<td>.8665</td>
<td>.8686</td>
<td>.8708</td>
<td>.8729</td>
<td>.8749</td>
<td>.8770</td>
<td>.8790</td>
<td>.8810</td>
<td>.8830</td>
</tr>
<tr>
<td>1.2</td>
<td>.8849</td>
<td>.8869</td>
<td>.8888</td>
<td>.8907</td>
<td>.8925</td>
<td>.8944</td>
<td>.8962</td>
<td>.8980</td>
<td>.8997</td>
<td>.9015</td>
</tr>
<tr>
<td>1.3</td>
<td>.9032</td>
<td>.9049</td>
<td>.9066</td>
<td>.9082</td>
<td>.9099</td>
<td>.9115</td>
<td>.9131</td>
<td>.9147</td>
<td>.9162</td>
<td>.9177</td>
</tr>
<tr>
<td>1.4</td>
<td>.9192</td>
<td>.9207</td>
<td>.9222</td>
<td>.9236</td>
<td>.9251</td>
<td>.9265</td>
<td>.9279</td>
<td>.9292</td>
<td>.9306</td>
<td>.9319</td>
</tr>
<tr>
<td>1.5</td>
<td>.9332</td>
<td>.9345</td>
<td>.9357</td>
<td>.9370</td>
<td>.9382</td>
<td>.9394</td>
<td>.9406</td>
<td>.9418</td>
<td>.9429</td>
<td>.9441</td>
</tr>
<tr>
<td>1.6</td>
<td>.9452</td>
<td>.9463</td>
<td>.9474</td>
<td>.9484</td>
<td>.9495</td>
<td>.9505</td>
<td>.9515</td>
<td>.9525</td>
<td>.9535</td>
<td>.9545</td>
</tr>
<tr>
<td>1.7</td>
<td>.9554</td>
<td>.9564</td>
<td>.9573</td>
<td>.9582</td>
<td>.9591</td>
<td>.9599</td>
<td>.9608</td>
<td>.9616</td>
<td>.9625</td>
<td>.9633</td>
</tr>
<tr>
<td>1.8</td>
<td>.9641</td>
<td>.9649</td>
<td>.9656</td>
<td>.9664</td>
<td>.9671</td>
<td>.9678</td>
<td>.9686</td>
<td>.9693</td>
<td>.9699</td>
<td>.9706</td>
</tr>
<tr>
<td>1.9</td>
<td>.9713</td>
<td>.9719</td>
<td>.9726</td>
<td>.9732</td>
<td>.9738</td>
<td>.9744</td>
<td>.9750</td>
<td>.9756</td>
<td>.9761</td>
<td>.9767</td>
</tr>
<tr>
<td>2.0</td>
<td>.9772</td>
<td>.9778</td>
<td>.9783</td>
<td>.9788</td>
<td>.9793</td>
<td>.9798</td>
<td>.9803</td>
<td>.9808</td>
<td>.9812</td>
<td>.9817</td>
</tr>
<tr>
<td>2.1</td>
<td>.9821</td>
<td>.9826</td>
<td>.9830</td>
<td>.9834</td>
<td>.9838</td>
<td>.9842</td>
<td>.9846</td>
<td>.9850</td>
<td>.9854</td>
<td>.9857</td>
</tr>
</tbody>
</table>
Example 6 – 1 (cont.)

- Find the area to the left of $z = 2.09 = 2.0 + 0.09$.

$$P(\text{the area lower than } z = 2.09) = P(z < 2.09) = 0.9817$$
MegaStat

MegaStat®

Probability
- Descriptive Statistics...
- Frequency Distributions
- Normal Distribution
  - Counting Rules
  - Discrete Probability Distributions
  - Continuous Probability Distributions

Normal Distribution

- z = 2.09
- Calculate P given z
- Calculate z given P

- Mean = 0
- Standard Deviation = 1

- Show axis points
- Show center line
- Show shading
  - lower
  - upper
  - two-tail

- Color
  - Solid color
  - Transparent
  - Patterned
  - No shading

OK
Normal distribution

\[ P(\text{lower}) = 0.9817 \]
\[ P(\text{upper}) = 0.0183 \]
\[ z = 2.09 \]
Other examples

- Examples 6 – 2 to 6 – 4.
Problem #1: Finding Probabilities Given Specific Data Values \((X \rightarrow P)\).

Transform \(X\) to \(Z\) using

\[
Z = \frac{X - \mu}{\sigma}
\]

then find the area under the normal curve \((P)\).
Problem #2: Finding Data Values Given Specific Probabilities \((P \rightarrow X)\).

Find \(Z\) that corresponds to an area under the normal curve \((P)\) and transform to \(Z\) using

\[ X = \mu + \sigma Z \]
Problem #3:
Number of units (individual/items) satisfying a condition =
Total number of units given in the problem ×
area calculated based on condition

Examples 6 – 6 to 6 – 10.
Example 6 – 6 (Problem #1)

- Each month, an American household generates an average of 28 pounds ($= \mu$) of newspaper for garbage or recycling. Assume the standard deviation is 2 pounds ($= \sigma$). If a household is selected at random, find the probability of its generating
  a. Between 27 and 31 pounds per month
  b. More than 30.2 pounds per month
  c. Less than 30.2 pounds per month

- Assume the variable is approximately normally distributed.
Example 6 – 6 (cont.)

X: amount of newspaper for garbage or recycling

\[ \mu = 28, \quad \sigma = 2 \]

a. Between 27 and 31 pounds per month

\[ P(27 < X < 31) = 0.6247 \]

b. More than 30.2 pounds per month

\[ P(X > 30.2) = 0.1357 \]

c. Less than 30.2 pounds per month

\[ P(X < 30.2) = 0.8643 \]
To qualify for a police academy, candidates must score in the top 10% on a general abilities test. The test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify. Assume the test scores are normally distributed.

\[ P = 0.1, \quad \mu = 200, \quad \sigma = 20 \quad \Rightarrow \quad X = 226 \]
Example 6 – 8 (Problem #3)

- Americans consume an average of 1.64 cups of coffee per day. Assume the variable is approximately normally distributed with a standard deviation of 0.24 cup. If 500 individuals are selected, approximately how many will drink less than 1 cup of coffee per day?

- \( P(X < 1) = 0.0038 \Rightarrow \) Number of people who drink less than 1 cup = \( 0.0038 \times 500 \approx 2 \)
A **sampling distribution of sample means** is a distribution using the means computed from all possible random samples of a specific size taken from a population.

**Sampling error** is the difference between the sample measure and the corresponding population measure because the sample is not a perfect representation of the population.
Properties of the Distribution of Sample Means

- The **mean of the sample means** will be the **same as** the **population mean**.

- The **standard deviation of the sample means** will be **smaller** than the **standard deviation of the population**, and it will be equal to the population standard deviation divided by the square root of the sample size.
The Central Limit Theorem

- As the **sample size** \( n \) increases without limit, the shape of the distribution of the sample means \( \bar{X} \) taken **with replacement** from a population with **mean** \( \mu \) and standard deviation \( \sigma \) will approach a normal distribution.

\[
Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}
\]

This is called the **standard error of the sample mean**! It is usually denoted by \( \sigma_{\bar{X}} \).
Individual data \((X)\): mean \(\mu\) and standard deviation \(\sigma\) are given. Hence,
\[
Z = \frac{X - \mu}{\sigma}
\]

Central limit theorem \((\bar{X})\): sample size \(n\), mean \(\mu\) and standard deviation \(\sigma\) are given. Hence,
\[
Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}
\]
Examples

- Examples 6 – 13 and 6 – 14.
Example 6 – 15

- The average number of pounds of meat that a person consumes per year is 218.4 pounds. Assume that the standard deviation is 25 pounds and the distribution is approximately normal.

a. Find the probability that a person selected at random consumes less than 224 pounds per year.

b. If a sample of 40 individuals is selected, find the probability that the mean of the sample will be less than 224 pounds per year.
Example 6 – 15 (cont.)

- The **average** number of pounds of meat that a person consumes per year is **218.4** pounds. Assume that the **standard deviation is 25** pounds and the distribution is approximately normal.

  a. Find the probability that **a person selected at random** consumes less than **224** pounds per year.

  b. **If a sample of 40 (\(= n\)) individuals is selected**, find the probability that **the mean of the sample** will be less than **224** pounds per year.
Example 6 – 15 (cont.)

a. Here $\mu = 218.4$ and $\sigma = 25$, so
$P(X < 224) = 0.5871$

b. Here $\mu = 218.4$ and $\sigma_{\bar{X}} = \frac{25}{\sqrt{40}}$, so
$P(\bar{X} < 224) = 0.9222$