Chapter 3 – Normal Distribution

**Density curve:** A density curve is an idealized histogram, a mathematical model; the curve tells you what values the quantity can take and how likely they are.

**Example:** Height of students in a statistics class

A probability density curve satisfies several rules:
- It never goes below the horizontal axis, i.e. it’s never negative.
- The total area under the curve is 1.
- The chance of the quantity falling between \(a\) and \(b\) is the area under the curve between the point \(a\) and \(b\).
Example: Time required to wait for a DART train in Mockingbird Station.

Q: What is the value of k?

Q: What is the chance that you will wait longer than 6 minutes?

Q: What is the chance that you will wait in between 1 and 3 minutes?

[This is called the Uniform distribution.]

Normal Curves

1. Normal distributions are symmetric, unimodal, bell-shaped density curve.

2. A lot of data follows the bell-shaped curve.

Example: Scores on intelligence tests are likely to be normally distributed. The mean is about 100 and a typical person is likely to score within about 15 points of the mean, that is, between 85 and 115.
3. The exact density curve for a particular normal distribution is described by giving its mean $\mu$ and its standard deviation $\sigma$.

4. How do we use it?
The 68-95-99.7 Rule

- This rule applies generally to a variable $X$ having normal distribution with mean $\mu$ and standard deviation $\sigma$.
- However, this rule does not apply to distributions that are not normal.

Approximately 68% of the observations fall within 1 standard deviation of the mean

That is, $0.68 = 68\%$ of the observations fall within one standard deviation of the mean, or, $68\%$ of the observations are between $(\mu - \sigma)$ and $(\mu + \sigma)$.

Approximately 95% of the observations fall within 2 standard deviations of the mean
Approximately 99.7% of the observations fall within 3 standard deviations of the mean.

Only a small fraction of observations (0.3% = 1 in 333) lie outside this range.

Another way of looking at it:

Example: Scores on intelligence tests (cont’d)
- If the psychologist knows the mean and the typical deviation from the mean (i.e. the standard deviation), the researcher can determine what proportion of scores is likely to fall in any given range.
- For instance, in the range between one standard deviation below the mean (about 85 for IQ scores) and one deviation above the mean (about 115 for IQ scores), one expects to find the scores of about 68% of all test takers.
- Further, only about 2.5% of test takers will score higher than two standard deviations above the mean (about 130).
Example: Fish in a certain lake have a distribution that is normal with mean 10 inches and standard deviation 2 inches.

Q: Approximate what proportion of fish is:
   a) Greater than 12 inches
   b) Greater than 8 and less than 12 inches
   c) Less than 6 inches
   d) Greater than 16 inches
What about values not at ± 1, 2, 3 standard deviation(s)?

**Standardizing and z-scores:**
If $x$ is an observation from a distribution with mean $\mu$ and standard deviation $\sigma$, the **standardized value** of $x$ is

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

A standardized value is often called a **z-score**.

**Standard Normal Distribution:** Normal distribution with mean 0 and standard deviation 1

Normal distribution with mean $\mu$ and s.d. $\sigma$

$\Rightarrow$ **Standard Normal distribution**, then use standard normal table.
Example: Fish example (cont’d)

Q: Approximate what proportion of fish is:
   e) Greater than 13 inches

   f) Greater than 8 less than 11 inches

Q: Catch-and-release policy is applied in this lake, only the longest 10% fish can be removed from the lake.
   g) What is the minimum length of fish can be removed from the lake?
**Supplementary Note on the use of Table A**

Table A: Standard Normal Probabilities (p. 690 & 691)

Table entry of $z$ is the area under the standard normal curve to the left of $z$.

Example 1: $z = 1.72$, the area under the standard normal curve to the left of $z$ can be found from Table A:

<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>...</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>.500</td>
<td>.5040</td>
<td>.5080</td>
<td>.5120</td>
<td>.5160</td>
<td>.5199</td>
<td>...</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>...</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>...</td>
<td>.6141</td>
</tr>
<tr>
<td>...</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>...</td>
<td>:</td>
</tr>
<tr>
<td>...</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>...</td>
<td>:</td>
</tr>
<tr>
<td>...</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>...</td>
<td>:</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>...</td>
<td>.9545</td>
</tr>
<tr>
<td>1.7</td>
<td>.9554</td>
<td>.9564</td>
<td><strong>.9573</strong></td>
<td>.9582</td>
<td>.9591</td>
<td>.9599</td>
<td>...</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>...</td>
<td>.9706</td>
</tr>
</tbody>
</table>

i.e.

Example 2: What is the value of $z$ that the area under the standard normal curve to the left of $z$ is 95% = 0.95?
Using Table A to find Normal Proportion:

1. State the problem in terms of the observed variable $x$ (with mean $\mu$ and standard deviation $\sigma$).

2. Draw a picture that shows the proportion you want in terms of $x$.

3. Standardized $x$ to restate the problem in terms of a standard normal variable $z$:
   $$z = (x - \mu)/\sigma$$

4. Draw another picture that shows the proportion you want in terms of $z$.

5. Use Table A and the fact that the total area under the curve is 1 to find the required area under the standard normal curve.

Using Table A to find Normal Percentage Point:

1. Let the percentage point you want to find as $L$, then state the problem in terms of the observed variable $x$ (with mean $\mu$ and standard deviation $\sigma$).

2. Draw a picture that shows the value you want in terms of $x$.

3. Standardized $x$ to restate the problem in terms of a standard normal variable $z$:
   $$z = (x - \mu)/\sigma$$

4. Draw another picture that shows the proportion you want in terms of $z$.

5. Use Table A to find the percentage point for standard normal curve, say $W$, then solve for $L$ from
   $$W = (L - \mu)/\sigma$$
   \[i.e.\quad L = W\sigma + \mu\]