6.1 Introducing Normally Dist’ed Variables

- It is a continuous random variable (Defined on an interval and every value is a possible outcome).

- One of the most useful and frequently encountered continuous RV's.

- Completely determined by its

- Symmetrical

- Its spread is determined by the value of its

**Note:** The normal dist’n with a mean of \( \mu \) and a SD of \( \sigma \) is denoted by \( N(\mu, \sigma) \).
Empirical Rule - 

A variable having approximately a bell shaped distribution should have almost all the observations (99.74%) within three standard deviations from the mean.
How to calculate probabilities?

The areas under the normal curve correspond to probabilities for $x$.

The area $A$ is the probability that $x$ assumes a value between $a$ and $b$.

**Note:** When $X$ is a continuous RV,

\[ P( a \leq X \leq b ) = P( a < X < b ). \]
To calculate normal probabilities we need calculus or tables. But tables are only available for RV’s that are normally dist’ed with mean 0 and SD of 1 (Table II, pgs A6 - A7).

To calculate normal probabilities use the fact that if

\[ X \sim N(\mu, \sigma) \]

then

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

the standardized normal RV. (Read ‘Z is normal 0, 1’).
6.2 Areas Under the Standard Normal N(0,1) Curve

Basic Properties:

1. The total area under the N(0,1)
2. The N(0,1) is symmetric
3. Most of the area under N(0,1) curve lies between
Using the standard normal table:

Ex: Determine the area under the N(0,1) curve that

a. lies to the left of 2.11

\[ P(Z \leq 2.11) = \]
b. lies to the right of -1.25

\[ P (Z \leq -1.25) = \]

\[ = 0.8944 \]
c. lies between -0.5 and 2.47 inclusive.

\[ P(-0.5 \leq Z \leq 2.47) = 0.6847 \]
Finding the z-score for a specified area:

Ex. Determine the z-score having an area of

a. 0.05 to its left

There is

\[
0.0495 = 0.0505 =
\]

Hence (using interpolation)

The z-score is
b. 0.025 to its right

Same as : Area to its left is

From table:

\[ 0.975 = \]

Hence, z-score is
Notation: \( z \) denotes the z-score having area (\( \alpha \)) to its right under \( N(0,1) \) curve.

From above: \( z_{0.025} = 1.96 \)

What is \( z_{0.05} \)?

\[
z_{0.05} = \]

\[
-4.0 -3.0 -2.0 -1.0 0.0 1.0 2.0 3.0 4.0
\]
6.3 Working with Normally Distributed Variables

To determine a percentage or probability for a normally dist'ed variable:

Steps:

1. Sketch the normal curve.

2. Shade the region of interest and mark delimiting x-values.

3. Compute the z-scores for the delimiting x-values found in (2).

4. Use table II to obtain the area under the N(0,1) curve.
Ex. Each year, thousands of college seniors take the Graduate Record Examination (GRE). The scores are transformed so that they have a mean of 500 and a SD of 100. Furthermore, the scores are known to be normally dist’ed. Determine the percentage of students that score:

a. between 350 and 600 inclusive.

Step 1 & 2:
Step 3:

\[ P(350 \leq X \leq 600) = ? \] (Shaded area).

To use Normal tables, first transform the normal RV \( X \) into the standard normal RV \( Z \) (find z-scores):

\[ \leq \quad \leq \]

or

\[ \leq Z \leq \]

Step 4:

\[ P(350 \leq X \leq 600) = \]

\[ = \]

\[ = \]

\[ = 0.7745 \]

Represents the area under \( N(0,1) \) over the interval from -1.5 to 1.
b. 375 or greater.

\[ P(X \geq 375) = P\left( Z \geq \right) \]

\[ = P\left( Z \geq \right) \]

\[ = \]

\[ = 0.8944 \]
c. between 300 and 450.

\[ P(300 \leq X \leq 450) = \]

\[ = \]

\[ = \]

\[ = 0.2857 \]

d. exactly equal to 680.

\[ P(X = 680) = \]
e. What score is exceeded by exactly 5%? (95 percentile).

\[ P(Z < a) = \]

\[ \text{Excel: norminv}(0.95, 0, 1) \]

\[ = 664.5 \]
6.4 **Assessing Normality; Normal Probability Plots**

If variable is normally dist’ed, to assess normality:

- **Large sample** → look at histogram. Bell shaped?
- **Small/large sample** → look at normal probability plots. Fairly linear?

**Normal probability plot:**

→ If plot is roughly linear, then accept as reasonable that the variable is approx. normal.

*Note:* You do not need to know the mechanics of how to construct a normal probability plot.