5.4 Linear transformation & Standardization

\[ Y = aX + b \]
\[ S_Y = 1a \cdot S_X \]
\[ \{ \bar{Y} = a\bar{X} + b \]
\[ S_Y = 1a \cdot S_X \]

Exercise 5.12: Temperature transformation

\[ X = \text{Temp (Celsius)} \]
M T W Th F S S
40 41 39 41 41 40 38
\[ \bar{X} = 40 \degree C \]
\[ S_X = 0.15 \degree C \]

\[ Y = \text{Temp (Fahrenheit)} \]
\[ = \frac{9}{5} \bar{X} + 32 \] \[ \Rightarrow a = \frac{9}{5}, b = 32 \]
\[ \bar{Y} = a\bar{X} + b = \frac{9}{5} \times 40 + 32 = 104 \degree F \]
\[ S_Y = 1a \cdot S_X = \frac{9}{5} \times 0.15 = 2.07 \degree F \]
Standardization:

A variable \( X \) is said to be standardized if the variable has a mean of zero and a standard deviation of 1.

\[
Y = \frac{X - \bar{X}}{S_x}
\]

\[
\bar{Y} = \frac{1}{S_x} \bar{X} - \frac{\bar{X}}{S_x} = 0
\]

\[
S_Y = \frac{1}{S_x} |S_x| = 1
\]

Standardization of \( X \)
A density function is a (nonnegative) function or curve that describes the overall shape of a distribution. The total area under the entire curve is equal to 1, and proportions are measured as areas under the density function.

6.3.1 Normal distributions

General notation: 
\[ X \sim N(\mu, \sigma) \]

\( X \) is normally distributed with mean \( \mu \) and standard deviation \( \sigma \).
How to calculate areas under a normal distribution.

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

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

Standardized normal variable.

\[ Z \sim N(0, 1) \]

The z-score or standard score for an observed value is

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

Finding proportions for the standard Normal distribution

Table II

\[ z = 2.13 \]

Area = 0.9834.
Find the area under the standard normal distribution.

\[ z = 1.22 \quad \text{to the left of } z = 1.22 \quad = 0.8888 \]

\[ z = -1.22 \quad = 0.1112 \]

Area to the left of 1.22 = area to the right of 1.22 = 1 - area to the left of 1.22 = 1 - 0.8888 = 0.1112

Find the proportions for general normal \( \mathcal{N}(\mu, \sigma) \).

1. Draw a picture.

2. Convert the interval to an interval for the standard score.

3. Draw a picture of \( \mathcal{N}(0, 1) \).

4. Find the proportion for the standard score interval. (Table II)
Ex: Suppose height in adult males is normally distributed with \( \mu = 70 \) inches and \( \sigma = 4 \) inches.

\[ X \sim N(70, 4) \]

(i). the proportion of adult males between 70 and 82 inches

\[ X_1 = 70 \quad Z_1 = \frac{X_1 - \mu}{\sigma} = \frac{70 - 70}{4} = 0 \]

\[ X_2 = 82 \quad Z_2 = \frac{X_2 - \mu}{\sigma} = \frac{82 - 70}{4} = 3 \]

area between 0 & 3

\[ = \text{area to the left of } 3 - \text{area to the left of } 0 \]

\[ = 0.9987 - 0.5 = 0.4987 \]

Table II

Proportion between 70 & 82 = 0.4987.

(ii) the proportion of males who are taller than 75 inches

\[ \frac{75 - 70}{4} = 1.25 \]

area to the right of 1.25

\[ = 1 - \text{area to the left of } 1.25 = 0.1056 \]

\[ = \text{area to the left of } -1.25 = 0.1056 \]
Find the pth percentile for a variable with $N(\mu, \sigma)$

1) Find the pth percentile for the standard normal variable $Z$.
Table III p. 408

$90$th percentile $= 1.282$

2) pth percentile for $N(\mu, \sigma)$ is

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

pth percentile of $N(0, 1)$

Ex 6.5 $N(100, 16)$

What IQ score must a 2-year-old have to place in the top 1% of the distribution?

99th percentile of $N(100, 16)$

1) 99th percentile for $N(10, 1)$

$$2.326$$

2) $Z = \frac{X - 10}{1}$

$$= 10 + 2.326 \times 16 = 187.216$$
Definition:
The **68-95-99.7 rule** for any normal distribution \( N(\mu, \sigma) \):

- 68% of the observations falls within one standard deviation of the mean \([\mu - \sigma, \mu + \sigma]\).
- 95% \([\mu - 2\sigma, \mu + 2\sigma]\).
- 99.7% \([\mu - 3\sigma, \mu + 3\sigma]\).

Assessing Normality:
1. Histogram or stem-and-leaf plot. Check for unusual features such as gaps, outliers, or skewness.
2. Compare our observations to the **68-95-99.7 rule**.

\[ \overline{x}, s \]

- Check 68% \([\overline{x} - s, \overline{x} + s]\).
- 95% \([\overline{x} - 2s, \overline{x} + 2s]\).
- 99.7% \([\overline{x} - 3s, \overline{x} + 3s]\) \(\Rightarrow N(\overline{x}, s)\).
6.3.2 Uniform distribution

**Notation:** \( X \sim U(a, b) \)

**Density:**

\[
\frac{1}{b-a}
\]

\( a \leq x \leq b \)

1 = base \times height

\[
\frac{1}{(b-a)} \times \frac{b-a}{b-a} = \frac{1}{(b-a)}
\]

**Mean = Median = \( \frac{a+b}{2} \).**

**Example:** \( X \sim U(40, 70) \)

**1.**

\[
\frac{1}{30}
\]

\( 40 \leq x \leq 70 \)

**2.** What proportion will be late for class?

\( 10 \times \frac{1}{30} = \frac{1}{3} \approx 0.33 \)
Ex: \( X \sim U(49, 51) \)

Find IQR for the distribution.

\[
\text{IQR} = 51 - 49
\]

\[
\frac{49 + 51}{2} = 0.25 \Rightarrow 49 + 0.25 = 49.25
\]

\[
\frac{51 + 49}{2} = 0.75 \Rightarrow 51 - 0.75 = 50.25
\]

\[
\text{IQR} = 1
\]