The Normal Distribution

**Definition**  A continuous random variable has a **normal distribution** if its probability density function can be written as for \(-\infty < y < \infty\) as

\[
f(y) = \frac{1}{\sigma y^{1/2}} \pi e^{-(y-\mu y)^2 / 2 \sigma^2}
\]

Notation: \(Y \sim N(\mu Y, \sigma Y)\) where \(\mu Y\) is the population mean and \(\sigma Y\) is the population standard deviation

Examples of random variables in the biological context that have a normal distribution

- Example 4.1.1  \(Y =\) serum cholesterol (mg/dLi)
- Example 4.1.2  \(Y =\) eggshell thickness (mm)
- Example 4.1.3  \(Y =\) nerve cell interspike times (ms)

The normal density curve is

(i)  Continuous over \(-\infty < y < \infty\)
(ii) Symmetric about \(y = \mu\)
(iii) Unimodal and hence, “bell shaped”

For infinitely many combinations of \(\mu\) and \(\sigma\), we have many bell-shaped curves...

**Figure 4.2.2**

![Normal Distribution Curves](image)

We’ll learn the following four topics about the normal distribution:

1. For a value in the distribution, find probabilities associated with it
2. For a probability, find the value in the distribution
3. Standardize a value from a normal distribution to find its Z-score
4. Given a set of data, try to say whether it came from a normal distribution
**Finding Probabilities Associated with a Normal Distribution**

Given a normal random variable, we can find probabilities associated with a value from the distribution by finding area under the curve (integrating).

R accomplishes this using the function `pnorm(x,mean,sd)`. This function gives the area under the curve to the left of “x” by default. Typing `?pnorm` will give the remainder of the options available for the “norm” family of functions.

**Example 4.3.1**  Suppose $Y = \text{length of herring (mm)}$ and also that $Y \sim N(\mu=54, \sigma=4.5)$

Find $P\{Y < 60\}$

Find the percent of herring in this distribution that are between 50 and 60 mm long.
**Percentiles of the Normal Distribution**

**Definition** The point of a distribution below which p\% lies is called the \( p^{\text{th}} \) percentile of the distribution.

R computes normal percentiles using the function `qnorm(p,mean,sd)`

Example 4.3.2 Find the 70\textsuperscript{th} percentile of the herring length distribution.

Find the 25\textsuperscript{th} percentile

Find the 50\textsuperscript{th} percentile
**Z-scores**

**Definition** The Standardization Formula for $Y \sim N(\mu, \sigma)$ is $Z = \frac{Y - \mu}{\sigma}$ and is called the Z-score.

If $Y \sim N(\mu, \sigma)$, then $Z \sim N(0, 1)$. We call this the standard normal distribution.

A Z-score can be interpreted as ________________________________________________________________

______________________________________________________________

Example  Find the Z-score for a herring with length 60mm.

Now, Find $P\{-1 < Z < 1\}$

Find $P\{-2 < Z < 2\}$

Find $P\{-3 < Z < 3\}$
**Definition**  The upper $\alpha$ critical point from $Z \sim N(0, 1)$ is the value $Z_\alpha$, such that $P\{Z > Z_\alpha\} = \alpha$

Notation: $Z_\alpha$

Draw a picture...

Find $Z_{0.05}$
Assessing Normality

Many statistical procedures are valid only if the data appear to have come from a normal distribution. How do we determine that our data appears to have come from a normal distribution? Will a well constructed histogram work?

A graphical display called a QQplot or normal probability plot is often used to determine whether the data appear to have come from a normal distribution. A QQplot plots the values of the normal score vs. the data values. Normal scores are the expected values of ordered observations in a sample and different computer packages compute these expected values differently.

Now, if our data were from a normal distribution, we’d expect to see the points form a linear pattern on a QQplot.

Example  The heights in inches of 11 women are listed below.
61  62.5  63  64  64.5  65  66.5  67  68  68.5  70.5
Using the QQplot, check the assumption that the data is distributed normally.

Based on the QQplot, would you say these data come from a normal distribution? Why or why not?

![QQplot for brain weight](image)

If there is a violation of the normality assumption, we can try a ________________________.

Some common ones are________________________________________________________

The following is the resulting QQplot after taking the logarithm of the brain weight data. Assess normality.

![QQplot after log transformation](image)
Common shapes that indicate a violation of the normality assumption in a QQplot: