Introduction to Statistics for the Life Sciences
Chapter 4: The Normal Distribution
Recall: Discrete Random Variables

- We can describe a discrete random variable with a table. For example:

<table>
<thead>
<tr>
<th>$k$</th>
<th>0</th>
<th>1</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = k)$</td>
<td>0.1</td>
<td>0.5</td>
<td>0.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Continuous Random Variables

- Numeric
- Not discrete
- Any two different values have a valid value between
Continuous Random Variables

- **Numeric**
- **Not discrete**
- Any two different values have a valid value between
- Let $Y$ be a continuous random variable equal to the exact height of a Red Kangaroo.
Continuous Random Variables

- Numeric
- Not discrete
- Any two different values have a valid value between
- Let $Y$ be a continuous random variable equal to the exact height of a Red Kangaroo.
- $Y = 2.1\, m$ and $Y = 2.0\, m$ are valid heights. $Y = 2.05\, m$ is also valid. Etc.
Let $Y$ be the continuous random variable that gives the temperature tomorrow in degrees $F$.

Can we make a table of probabilities? Why or why not?

What is the probability that the temperature is 70 tomorrow? $P(Y = 70)$?
Continuous Random Variables

Let $Y$ be the continuous random variable that gives the temperature tomorrow in degrees $F$.

Can we make a table of probabilities? Why or why not?

What is the probability that the temperature is 70 tomorrow? $P(Y = 70)$?

$P(Y = 70.8888118100416)$?
Continuous Random Variables

Let \( Y \) be the continuous random variable that gives the temperature tomorrow in degrees \( F \).

Can we make a table of probabilities? Why or why not?

What is the probability that the temperature is 70 tomorrow?

\[ P(Y = 70) \]

\[ P(Y = 70.8888118100416) \]

The probabilities are 0.
Since we don’t have a table that gives probabilities of certain values of $Y$ in the case of a continuous random variable, what do we do?

### Probability Density Function

Let $Y$ be a continuous random variable. Then associated with $Y$ is a function $f(y)$, called the probability density function (pdf), which is used to answer questions like $P(Y < c)$, where $c$ is some number. The pdf has several properties:

- $f(y) \geq 0$ for all $y$
- $\int_{-\infty}^{\infty} f(y)dy = 1$***
- Equivalently, the area under the curve of $f(y)$ is equal to 1.
Example

Suppose that the probability density function of $Y$ is given by $f(y) = 2y$ for $0 \leq y \leq 1$. Calculate the probability that $Y$ is less than $1/2$, or $P(Y < 1/2)$. This corresponds to the area under the curve given by $f(y) = 2y$ from 0 to $1/2$. 
The blue area is equal to \((1/2)(1/2)(1) = 1/4 = P(Y < 1/2)\)
Same example, slightly different

Suppose that the probability density function of $Y$ is given by $f(y) = 2y$ for $0 \leq y \leq 1$. Calculate the probability that $Y$ is less than or equal to $1/2$, or $P(Y \leq 1/2)$.

How does the solution change?

$$P(Y \leq 1/2) = P(Y < 1/2) + P(Y = 1/2)$$

$$= P(Y < 1/2) + 0$$

$$= 1/4$$
Continuous Random Variables

Normal Distribution

Note

If $Y$ is a continuous random variable, for any fixed value $c$, $P(Y = c) = 0$

So only probabilities of intervals like $P(a < Y \leq b)$ may have probabilities greater than 0.
In nature, the symmetric unimodal “bell curve” keeps appearing.
Continuous Random Variables
Normal Distribution

IQ

Frequency

0.1% 2.2% 13.6% 34.1% 34.1% 13.6% 2.2% 0.1%

55 70 85 100 115 130 145

x
For Fun

- **Uninvolved**: Checked out, very disinterested; not involved
- **Good Soldier**: Reads the materials and attends all meetings
- **Perfect Director**: Very engaged, works collaboratively with both the management team and other directors
- **High Maintenance**: Regularly works with management on a variety of specific topics and issues
- **Meddler**: Constantly in management’s face about day-to-day tactics
Definition of the Normal Distribution

Normal Distribution

$Y$ is a normal random variable with mean $\mu$ and standard deviation $\sigma$, denoted by $Y \sim N(\mu, \sigma)$, if

$$f(y) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2\sigma^2} (y-\mu)^2}$$

for $-\infty < y < \infty$. 
Suppose that $Y \sim N(\mu, \sigma)$

- The pdf $f(y)$ is a symmetric bell-shaped curve.
- $f(y)$ is centered at $y = \mu$.
- The points of inflection (where the curve changes concavity) are located at $\mu \pm \sigma$
- The probability of $Y$ occurring in certain intervals, say $P(Y < b)$, is equal to the area under the curve of $f(y)$.
- Values can go to infinity in either direction
Definition

**Standard Normal Distribution**

Let $Z$ be a normal random variable with $\mu = 0$ and $\sigma = 1$, or $Z \sim N(0, 1)$. We then say $Z$ has the **standard normal distribution**.
Recall: Empirical Rule

- The 68–95–99.7 Rule.
- For every normal curve:
  1. $\approx 68\%$ of the area is within one SD of the mean
  2. $\approx 95\%$ of the area is within two SDs of the mean
  3. $\approx 99.7\%$ of the area is within three SDs of the mean
Compute areas

- There is no formula to calculate general areas under the normal curve.
- We will learn to use normal tables for this.
- LEARN the normal table.
We have tables for $\mathcal{N}(0, 1)$

- The standard normal table lists the area to the left of $z$ under the standard normal curve for each value from $-3.49$ to $3.49$ by $0.01$ increments.
- The normal table is located: Table 3 in your book
- Numbers in the margins represent $z$.
- Numbers in the middle of the table are areas to the left of $z$. 
Warning!

- Learn the normal table today.
- Make it part of your being.
Standard Normal Table: Computing Probabilities
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area = 1 (total probability rule)

- $P(Z \leq 0) = \frac{1}{2}$ by symmetry
- $P(Z = 0) = 0$
- $P(Z < 1) = 0.8413$ (table, front cover)
- $P(0 \leq Z \leq 1) =
- P(-1 \leq Z \leq 1) =
- P(Z > 1.5) =

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area $= 1$ (total probability rule)

$P(Z \leq 0) = \text{by symmetry}$

$P(Z = 0) = 0$

$P(Z < 1) = 0.8413$ (table, front cover)

$P(0 \leq Z \leq 1) =$

$P(-1 \leq Z \leq 1) =$

$P(Z > 1.5) =$

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area = 1 (total probability rule)

$P(Z \leq 0) = \text{by symmetry}$
$P(Z = 0) = 0$
$P(Z < 1) = .8413 \text{ (table, front cover)}$
$P(0 \leq Z \leq 1) =$
$P(-1 \leq Z \leq 1) =$
$P(Z > 1.5) =$

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal $Z \sim \mathcal{N}(0, 1)$

- Whole area $= 1$ (total probability rule)
- $P(Z \leq 0) = .5$ by symmetry
- $P(Z = 0) = 0$
- $P(Z < 1) = .8413$ (table, front cover)
- $P(0 \leq Z \leq 1) =$
- $P(-1 \leq Z \leq 1) =$
- $P(Z > 1.5) =$

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area = 1 (total probability rule)

$P(Z \leq 0) = .5$ by symmetry

$P(Z = 0) = 0$

$P(Z < 1) = .8413$ (table, front cover)

$P(0 \leq Z \leq 1) =$

$P(-1 \leq Z \leq 1) =$

$P(Z > 1.5) =$

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area = 1 (total probability rule)

$P(Z \leq 0) = .5$ by symmetry

$P(Z = 0) = 0$

$P(Z < 1) = .8413$ (table, front cover)

$P(0 \leq Z \leq 1) =

P(-1 \leq Z \leq 1) =

P(Z > 1.5) =$

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area = 1 (total probability rule)

$P(Z \leq 0) = .5$ by symmetry

$P(Z = 0) = 0$

$P(Z < 1) = .8413$ (table, front cover)

$P(0 \leq Z \leq 1) = 0.8413 - 0.5 = .3413$

$P(-1 \leq Z \leq 1) =$

$P(Z > 1.5) =$

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal \( Z \sim \mathcal{N}(0, 1) \)

Whole area = 1 (total probability rule)

- \( P(Z \leq 0) = 0.5 \) by symmetry
- \( P(Z = 0) = 0 \)
- \( P(Z < 1) = 0.8413 \) (table, front cover)
- \( P(0 \leq Z \leq 1) = 0.8413 - 0.5 = 0.3413 \)
- \( P(-1 \leq Z \leq 1) = \)
- \( P(Z > 1.5) = \)

Draw a picture and make use of symmetry.

\[ P(-0.5 \leq Z \leq 0.3) = \]
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area = 1 (total probability rule)

$P(Z \leq 0) = .5$ by symmetry

$P(Z = 0) = 0$

$P(Z < 1) = .8413$ (table, front cover)

$P(0 \leq Z \leq 1) = 0.8413 - 0.5 = .3413$

$P(-1 \leq Z \leq 1) = 2 \times 0.3413 = .6826$

$P(Z > 1.5) = $

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area = 1 (total probability rule)

$P(Z \leq 0) = .5$ by symmetry

$P(Z = 0) = 0$

$P(Z < 1) = .8413$ (table, front cover)

$P(0 \leq Z \leq 1) = 0.8413 - 0.5 = .3413$

$P(-1 \leq Z \leq 1) = 2 \times 0.3413 = .6826$

$P(Z > 1.5) =$

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area = 1 (total probability rule)

- $P(Z \leq 0) = 0.5$ by symmetry
- $P(Z = 0) = 0$
- $P(Z < 1) = 0.8413$ (table, front cover)
- $P(0 \leq Z \leq 1) = 0.8413 - 0.5 = 0.3413$
- $P(-1 \leq Z \leq 1) = 2 \times 0.3413 = 0.6826$
- $P(Z > 1.5) = 1 - P(Z \leq -1.5) = 1 - 0.9332 = 0.0668$

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) =$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area = 1 (total probability rule)

- $P(Z \leq 0) = 0.5$ by symmetry
- $P(Z = 0) = 0$
- $P(Z < 1) = 0.8413$ (table, front cover)
- $P(0 \leq Z \leq 1) = 0.8413 - 0.5 = 0.3413$
- $P(-1 \leq Z \leq 1) = 2 \times 0.3413 = 0.6826$
- $P(Z > 1.5) = 1 - P(Z \leq -1.5)$
  $$= 1 - 0.9332 = 0.0668$$

Draw a picture and make use of symmetry.

$$P(-0.5 \leq Z \leq 0.3) =$$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area $= 1$ (total probability rule)

- $P(Z \leq 0) = .5$ by symmetry
- $P(Z = 0) = 0$
- $P(Z < 1) = .8413$ (table, front cover)
- $P(0 \leq Z \leq 1) = 0.8413 - 0.5 = .3413$
- $P(-1 \leq Z \leq 1) = 2 \times 0.3413 = .6826$
- $P(Z > 1.5) = 1 - P(Z \leq -1.5) = 1 - 0.9332 = .0668$

Draw a picture and make use of symmetry.

$$P(-0.5 \leq Z \leq 0.3) = P(Z \leq 0.3) - P(Z \leq -0.5)$$
The standard normal $Z \sim \mathcal{N}(0, 1)$

Whole area $= 1$ (total probability rule)

$P(Z \leq 0) = .5$ by symmetry

$P(Z = 0) = 0$

$P(Z < 1) = .8413$ (table, front cover)

$P(0 \leq Z \leq 1) = 0.8413 - 0.5 = .3413$

$P(-1 \leq Z \leq 1) = 2 \times 0.3413 = .6826$

$P(Z > 1.5) = 1 - P(Z \leq -1.5)$

$= 1 - 0.9332 = .0668$

Draw a picture and make use of symmetry.

$P(-0.5 \leq Z \leq 0.3) = P(Z \leq 0.3) - P(Z \leq -0.5)$

$= .6179 - .3085 = .3094$
Standard Normal Table: Finding Quantiles
Basic Idea

- This is the inverse problem.

  *Before.* Question: here is the interval; Answer: Probability of the interval

  *Now.* Question: Here is the probability. Answer: What is the interval that matches with the probability.

- Problem: many intervals have the same area.

- We ask for a certain type of interval.
The Reverse Question: Finding quantiles

\[ P(Z \leq ?) = 0.975 \]

the value \(?\) is a **quantile**.

Use Table 3 (or front cover) backward

\[ P(Z \leq ?) = 0.20 \]
\[ P(Z \geq ?) = 0.80 \]
The Reverse Question: Finding quantiles

\[ P(Z \leq ?) = 0.975 \quad ? = 1.96 \]

the value \( ? \) is a **quantile**.

Use Table 3 (or front cover) backward

\[ P(Z \leq ?) = 0.20 \]
\[ P(Z \geq ?) = 0.80 \]
The Reverse Question: Finding quantiles

\[ P(Z \leq ?) = 0.975 \quad ? = 1.96 \]
the value \( ? \) is a quantile.
Use Table 3 (or front cover) backward

\[ P(Z \leq ?) = 0.20 \quad ? = -0.84 \]
\[ P(Z \geq ?) = 0.80 \]
The Reverse Question: Finding quantiles

\[ P(Z \leq ?) = 0.20 \quad ? = -0.84 \]
\[ P(Z \geq ?) = 0.80 \quad ? = -0.84 \]
Instead of \( ? \) we introduce a fancy notation.

Let \( z_\alpha \) be the number such that

\[
P(Z < z_\alpha) = 1 - \alpha
\]

See Figures 4.3.11, 4.3.12 (p.129) (\( \alpha \) to the right; \( 1 - \alpha \) to the left)

Example. \( z_{0.025} = 1.96 \)
General Normal Distribution
We are good if $Y \sim \mathcal{N}(0, 1)$

What if $Y \sim \mathcal{N}(10, 5)$ or $\mathcal{N}(11, 3)$, or ...
General form $\mathcal{N}(\mu, \sigma)$

- All normal distributions have the same shape.
- The question: How many standard deviations away from the mean?

Transformation

If

$$Y \sim \mathcal{N}(\mu, \sigma)$$

then

$$Z = \frac{Y - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$
Standardization

- All normal curves have the same shape, and are simply rescaled versions of the standard normal density.
- Consequently, every area under a general normal curve corresponds to an area under the standard normal curve.
- The key standardization formula is

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

- Solving for \( x \) yields

\[ x = \mu + z\sigma \]

which says algebraically that \( x \) is \( z \) standard deviations above the mean.
Use $z = \frac{x - \mu}{\sigma}$.

If $X \sim N(100, 2)$, find $P(X > 97.5)$. 
Use \( z = \frac{x - \mu}{\sigma} \):

- If \( X \sim \mathcal{N}(100, 2) \), find \( P(X > 97.5) \).
- Solution:

\[
P(X > 97.5) = P\left( \frac{X - 100}{2} > \frac{97.5 - 100}{2} \right)
= P(Z > -1.25)
= 1 - P(Z < -1.25)
= 1 - .1056 = .8944
\]
Quantile Example

Use \( x = \mu + z\sigma \)

If \( X \sim N(100, 2) \), find the cutoff values for the middle 70% of the distribution.
Quantile Example: Solution

Use $x = \mu + z\sigma$

- If $X \sim N(100, 2)$, find the cutoff values for the middle 70% of the distribution.
- Solution: The cutoff points will be the 0.15 and 0.85 quantiles.
- From the table, $1.03 < z < 1.04$ (1.04 is closer)

$$P(Z < 1.03) = .8485, \quad P(Z < 1.04) = .8508$$

Thus, the cutoff points are the mean plus or minus 1.04 standard deviations.

$$100 - 1.04(2) = 97.92, \quad 100 + 1.04(2) = 102.08$$
A Case Study
Exam Hint

- This how exam questions will often look.
- I give you some question in the context of a scientific area; you figure out the probability calculations to use.
- (word problems)
Case Study

Example

Body temperature varies within individuals over time (it can be higher when one is ill with a fever, or during or after physical exertion). However, if we measure the body temperature of a single healthy person when at rest, these measurements vary little from day to day, and we can associate with each person an individual resting body temperature. There is, however, variation among individuals of resting body temperature.
The Question

Example

In the population, suppose that:

- the mean resting body temperature is 98.25 degrees Fahrenheit;
- the standard deviation is 0.73 degrees Fahrenheit;
- resting body temperatures are normally distributed.

Let $X$ be the resting body temperature of a randomly chosen individual. Find:

1. $P(X < 98)$, the proportion of individuals with temperature less than 98.
2. $P(98 < X < 100)$, the proportion of individuals with temperature between 98 and 100.
$P(X < 98) = P(X - 98.25 < 98 - 98.25)$

$= P \left( \frac{X - 98.25}{.73} < \frac{98 - 98.25}{.73} \right)$

$= P(Z < -0.34)$

$= .367$
\[ P(98 < X < 100) = P(X < 100) - P(X < 98) \]

So find \( P(X < 100) \) and plug in previous answer.

\[ P(X < 100) = P(X - 98.25 < 100 - 98.25) \]
\[ = P \left( \frac{X - 98.25}{.73} < \frac{100 - 98.25}{.73} \right) \]
\[ = P \left( Z < 2.39726 \right) \]
\[ .9917 \]

So that \( P(98 < X < 100) = .9917 - .367 = .6247 \)