WIIT 7752

Lesson A Guided Notes

The Normal Distribution

Probability Density Curve

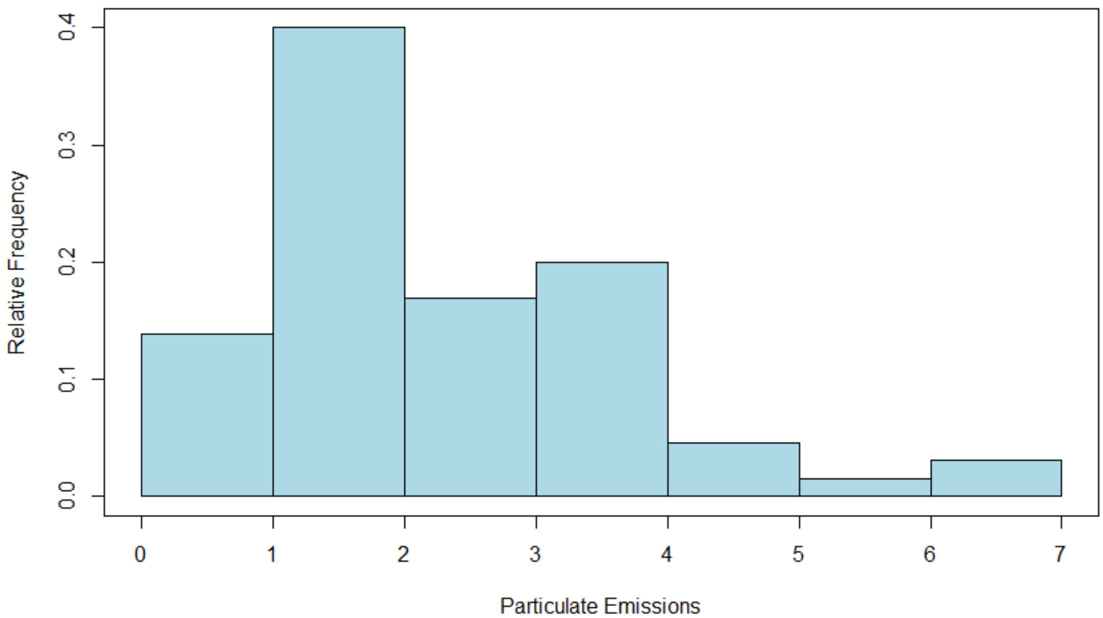
The emissions for 65 vehicles, in units of grams of particles per gallon of fuel, are given.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1.5 | 0.87 | 1.12 | 1.25 | 3.46 | 1.11 | 1.12 | 0.88 | 1.29 | 0.94 | 0.64 | 1.31 | 2.49 |
| 1.48 | 1.06 | 1.11 | 2.15 | 0.86 | 1.81 | 1.47 | 1.24 | 1.63 | 2.14 | 6.64 | 4.04 | 2.48 |
| 1.4 | 1.37 | 1.81 | 1.14 | 1.63 | 3.67 | 0.55 | 2.67 | 2.63 | 3.03 | 1.23 | 1.04 | 1.63 |
| 3.12 | 2.37 | 2.12 | 2.68 | 1.17 | 3.34 | 3.79 | 1.28 | 2.1 | 6.55 | 1.18 | 3.06 | 0.48 |
| 0.25 | 0.53 | 3.36 | 3.47 | 2.74 | 1.88 | 5.94 | 4.24 | 3.52 | 3.59 | 3.1 | 3.33 | 4.58 |

Below is the frequency distribution

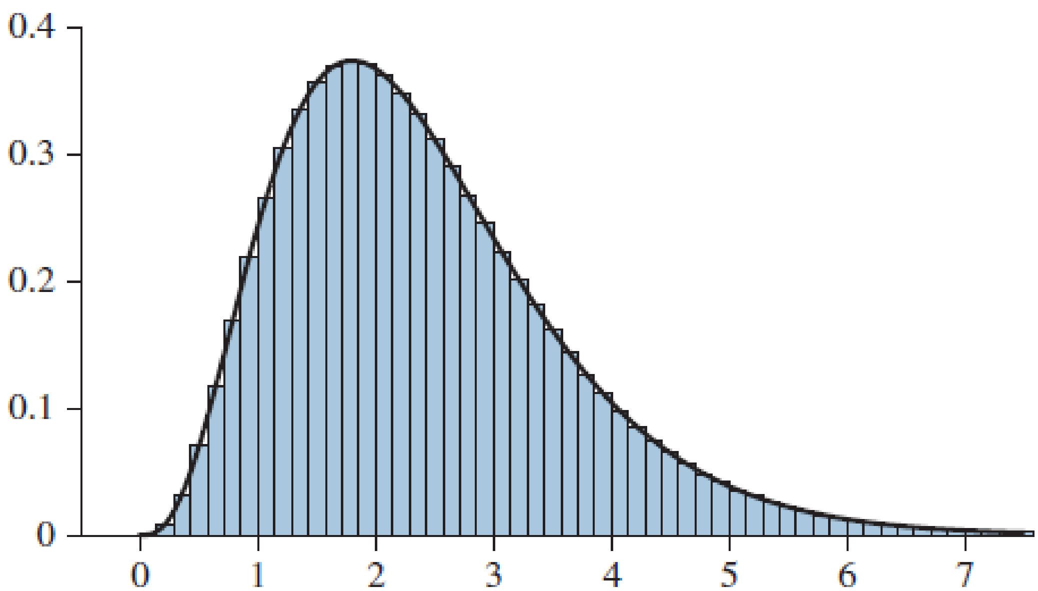
| Class | Frequency | Relative Frequency |
|---|---|---|
| 0 .00 – 0.99 | 9 | 0.14 |
| 1.00 – 1.99 | 26 | 0.40 |
| 2.00 – 2.99 | 11 | 0.17 |
| 3.00 – 3.99 | 13 | 0.20 |
| 4.00 – 4.99 | 3 | 0.05 |
| 5.00 – 5.99 | 1 | 0.02 |
| 6.00 – 6.99 | 2 | 0.03 |

The following figure presents a relative frequency histogram for the particulate emissions of a sample of 65 vehicles.



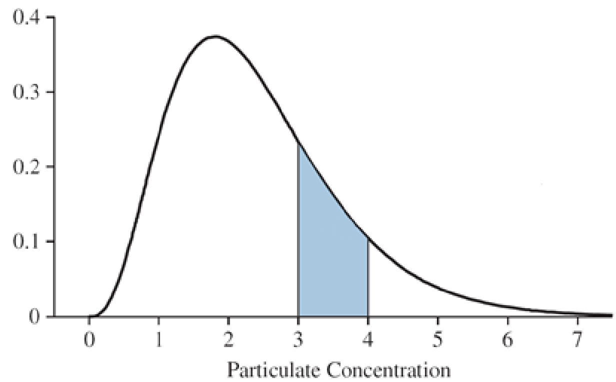

If we had information on the entire population, containing millions of vehicles, we could make the rectangles extremely narrow.

The histogram would then look smooth and could be approximated by a curve.  Since the amount of emissions is a continuous variable, the curve used to describe the distribution of this variable is called the **probability density curve** of the random variable.



Area under Probability Density Curve

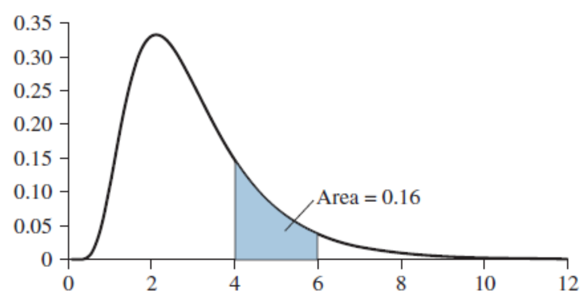
The **area under a probability density curve** between any two values *a* and *b* has two interpretations:




1. It represents the proportion of the population whose values are between *a* and *b*.

2. It represents the probability that a randomly selected value from the population will be between *a* and *b*.

Example 1 The diagram is a probability density




curve for a population.

a) What proportion of the population is between 4 and 6?

b) If a value is chosen at random from this population, what is the probability that it is not between 4 and 6?

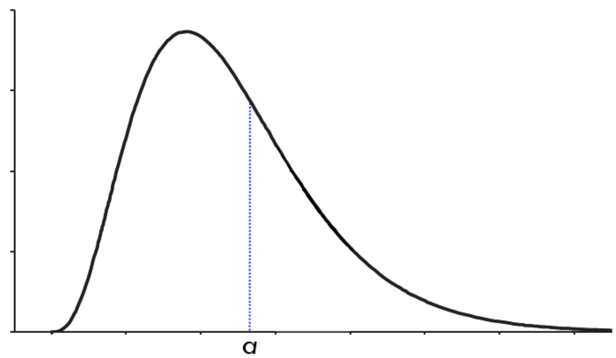
Facts About the Probability Desnity Curve

The region above a single point has zero width. Therefore, if *X* is a continuous random variable, ***P*(*X* = *a*) = 0** for any number *a*.

If *X* is a continuous random variable, then

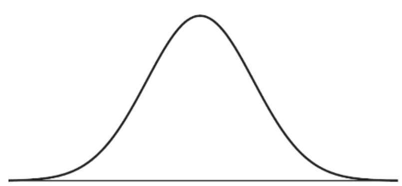
***P*(*a* < *X* < *b*) = *P*(*a* ≤ *X* ≤ *b*)** for any number *a* and *b*.

For any probability density curve, the **area under the entire curve is 1**, because this area represents the entire population.




Normal Curves

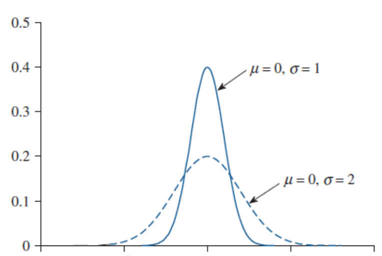
Probability density curves comes in many varieties, depending on the characteristics of the populations they represent. Many important statistical procedures can be carried out using only one type of probability density curve, called a **normal curve**.

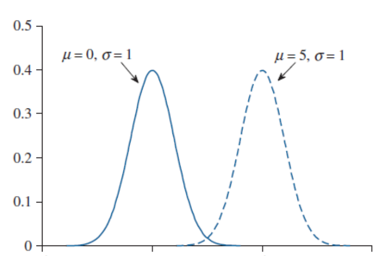


A population that is represented by a normal curve is said to be **normally distributed**, or to have a **normal distribution**.

The **population mean** (or mode) determines the **location of the peak**. The **population standard deviation** measures the **spread** of the population.

Therefore, the normal curve is wide and flat when the population standard deviation is large, and tall and narrow when the population standard deviation is small.



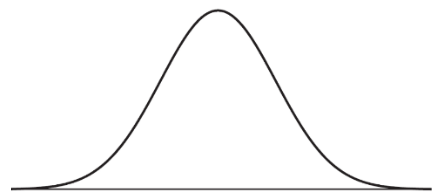



Properties of Normal Distributions

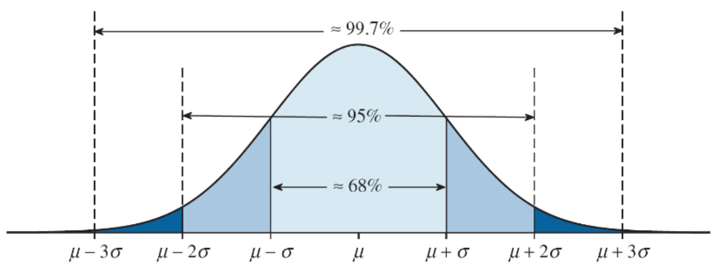
Normal distributions have **one mode**.

Normal distributions are **symmetric** around the mode.

The **mean** and **median** of a normal distribution are both **equal to the mode**.



The normal distribution follows the empirical rule:




Standard Normal Curves

A normal distribution can have any mean and any positive standard deviation.

The distribution with a **mean of 0** and **standard deviation of 1** is known as the **standard normal distribution**.

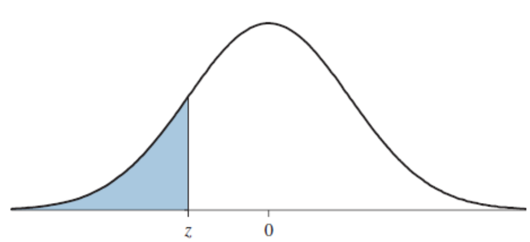
The probability density function for the standard normal distribution is called the **standard normal curve.**

Z-scores

When finding an area under the standard normal curve, we use the letter *z* to indicate a value on the horizontal axis beneath the curve. We refer to such a value as a ***z*-score**.

Since the mean of the standard normal distribution is 0:

- The **mean** has a ***z*-score of 0**.
- Points on the horizontal axis to the **left** of the mean have **negative *z*-scores**.
- Points to the **right** of the mean have **positive *z*-scores**.



Z-Score

Who is taller, a man 73 inches tall or a woman 68 inches tall? The obvious answer is that the man is taller. However, men are taller than women on the average. Suppose the question is asked this way: Who is taller relative to their gender, a man 73 inches tall or a woman 68 inches tall?

One way to answer this question is with a ***z*-score**.

The z-score of an individual data value tells how many standard deviations that value is from its population mean.

For example, a value one standard deviation above the mean has a *z*-score of 1. A value two standard deviations below the mean has a *z*-score of –2.

**Definition:**

Let $x$ be a value from a population with mean $\mu$ and standard deviation $\sigma$ . The ***z*-score** for $x$ is

Example 2 A National Center for Health Statistics study states that the mean height for adult men in the U.S. is $\mu$ = 69.4 inches, with a standard deviation of $\sigma$ = 3.1 inches. The mean height for adult women is $\mu$ = 63.8 inches, with a standard deviation of $\sigma$ = 2.8 inches. Who is taller relative to their gender, a man 73 inches tall, or a woman 68 inches tall?

Calculating Probabilities for Normal Distribution

A probability density function (pdf) of a continuous random variable *X* is a function $f(x)$ such that for any two numbers $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

That is, the probability that *X* takes on a value in the interval $[a, b]$ is the area under the graph of the density function between the values of $x = a$ and $x = b$.

The cumulative distribution function (cdf) $F(x)$ for a continuous random variable *X* is defined for every number *x* by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

Computing Probabilities Using the cdf

Let *X* be a continuous random variable with pdf $f(x)$ and cdf $F(x)$. Then for any number *a*,

$$P(X > a) = 1 - F(a)$$

And for any two numbers *a* and *b* with *a* < *b*,

$$P(a \leq X \leq b) = F(b) - F(a)$$

# The Normal Distribution

## Description

Density, distribution function, quantile function and random generation for the normal distribution with mean equal to `mean` and standard deviation equal to `sd`.

## Usage

```
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean = 0, sd = 1)
```

## Arguments

| | |
|---|---|
| `x, q` | vector of quantiles. |
| `p` | vector of probabilities. |
| `n` | number of observations. If `length(n) > 1`, the length is taken to be the number required. |
| `mean` | vector of means. |
| `sd` | vector of standard deviations. |
| `log, log.p` | logical; if TRUE, probabilities p are given as log(p). |
| `lower.tail` | logical; if TRUE (default), probabilities are $P[X \le x]$ otherwise, $P[X > x]$. |

## Details

If `mean` or `sd` are not specified they assume the default values of `0` and `1`, respectively.

The normal distribution has density

$$f(x) = 1/(\sqrt{(2 \pi)}\ \sigma)\ e\text{^}-((x - \mu)\text{^}2/(2\ \sigma\text{^}2))$$

where $\mu$ is the mean of the distribution and $\sigma$ the standard deviation.

## Value

`dnorm` gives the density, `pnorm` gives the distribution function, `qnorm` gives the quantile function, and `rnorm` generates random deviates.

Example 3 Find each probability indicated.

(a) $P(Z < 1.26)$

(b) $P(X < 250)$ where $X \sim N(266,16)$

(c) $P(X > 35)$ where $X \sim N(28,3)$

(d) $P(-1 < Z < 2)$

(e) $P(Z > 1.5 \ or \ Z < -2)$

Quantiles For a continuous probability distribution for a random variable $X$, the qth quantile, denoted by $x_q$, is the value such that the proportion of the area under the pdf for $X$ to the left of $x_q$ is q.

Notation: $z_\alpha$ denotes that value for which $\alpha$ of the area under the standard normal density curve lies to the left of $z_\alpha$.

Example 4: Find each percentile:

a) Find $z_{0.95}$

b) Find $z_{0.50}$

c) Find $x_{0.84}$ of a normal distribution with mean 45 and variance 36.

d) Find the value of $x$ from a normal distribution with mean -7 and standard deviation 9 that has 0.23 of the area under the pdf to the right.