Stat 201: Statistics I
Week 5

June 19, 2017

Chapter 6
Normal Probability Distributions

Section 6.2
The Standard Normal Distribution

Continuous probability distributions

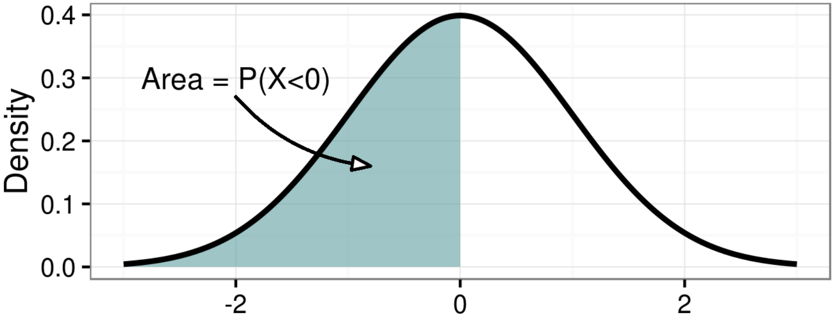
A **continuous probability distribution** is a description of the probabilities of all possible values of a continuous random variable.

- Unlike discrete probability distributions, cannot be displayed in a table. There are infinite possible values.
- Distributions are precisely defined by a probability density function (PDF) or a cumulative density function (CDF).
- Probabilities of single values are technically always zero, $P(X = x) = 0$.
- Only probabilities of ranges of values have meaning.

Density curves

A continuous probability distribution is visualized by a **density curve**, a graph of the probability density function.

- The height of the curve (the y-value) is always between 0 and 1.
- The total area under the graph is always 1.
- Probabilities are defined as the area under the curve for the range of values of the random variable.

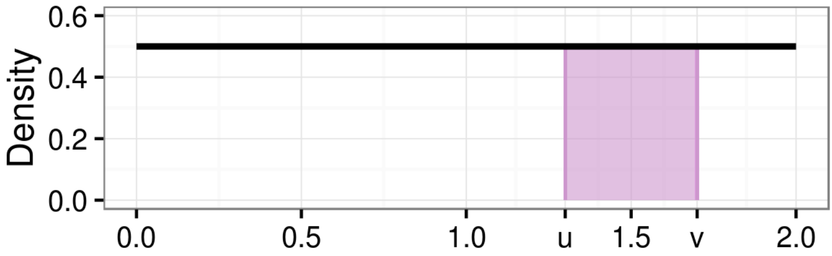


Uniform probability distribution

A random variable has an **uniform distribution** if all possible values are equally likely.

- $X \sim U(a, b)$, a is the minimum and b is the maximum
- PDF: $f(x) = c$, where c is a constant $(c = 1/(b-a))$
- $P(u < x < v)$ is the area of the rectangle $(v-u) \times c$, or

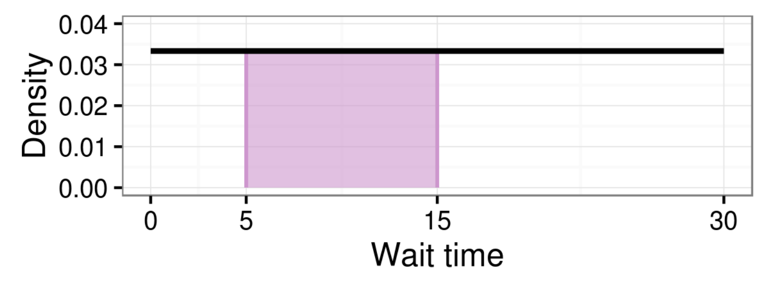
$$P(u < x < v) = \frac{v-u}{b-a}$$



Uniform probability distribution, example

Example

Bernice is waiting at the bus stop in a torrential downpour. She knows the bus will arrive any time between 0 and 30 minutes with equal likelihood. What is the probability she will have to wait between 5 and 15 minutes?



- $X \sim U(0, 30)$
- $P(5 < X < 15) = \dfrac{15 - 5}{30} = \dfrac{1}{3} = 0.33$

Normal distributions

Recall, normal distributions were defined as having a particular shape.

- Start with low values, rise to a maximum value, and end with low values.
- Distribution is symmetric (mirror image) around maximum.
- "Bell curve"

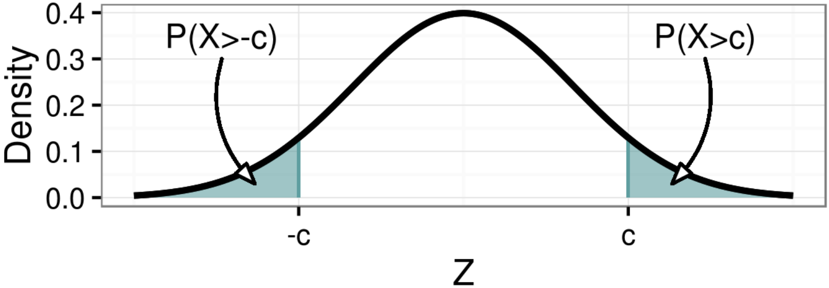
Formally, random variable X has a normal distribution with mean μ and standard deviation σ...

- $X \sim N(\mu, \sigma)$
- PDF: $f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma}}$
- $P(a < x < b \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma}} \, dx$

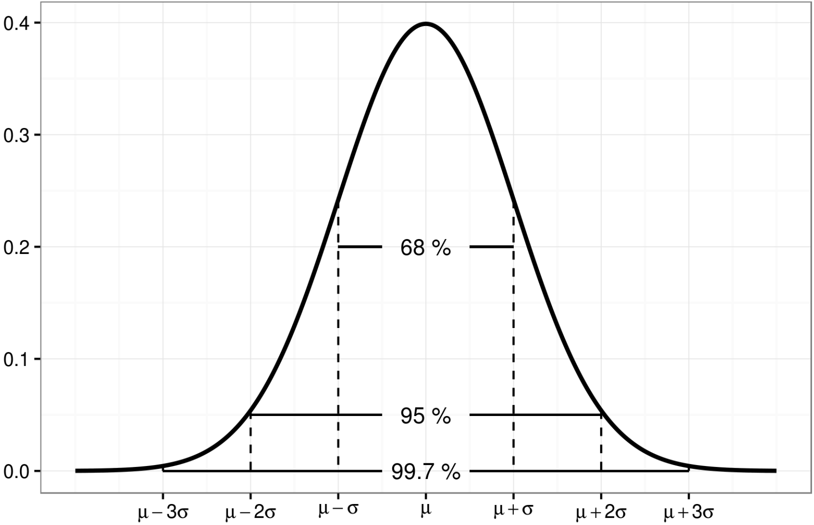
Symmetry of normal distribution

Normal distributions are perfectly symmetrical, mathematically speaking. That means, the probability a value is greater than some number is equal to the probability of being below the negative of that number.

- $P(X > c) = P(X < -c)$



Distribution of normal distributions

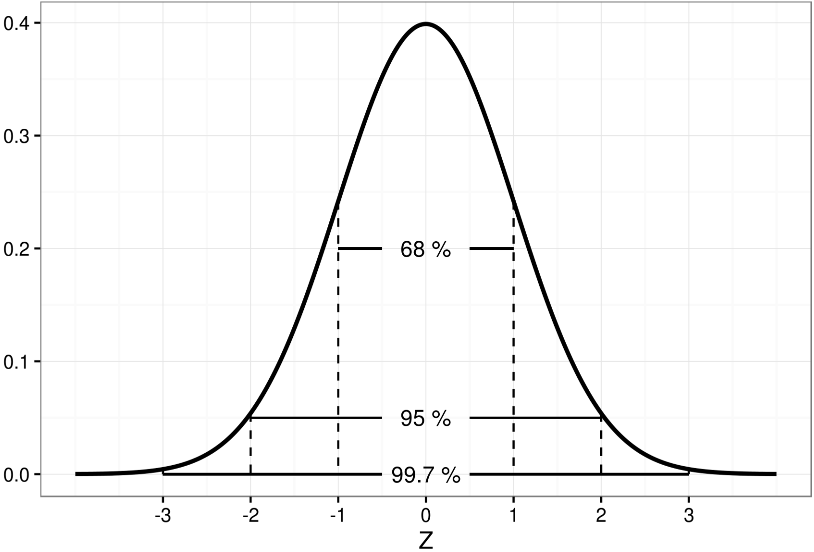


Standard normal distribution

A **standard normal distribution** is a normal distribution with a mean $\mu = 0$ and a standard deviation $\sigma = 1$.

- Also known as the Z distribution.
- $Z \sim N(0, 1)$
- Values of the standard normal are known as z-scores.
- A z-score of 1 ($z = 1$) is one standard deviation above the mean, $z = -2$ is two standard deviations below the mean, etc.

Z distribution



Probabilities of standard normal variables

Before technological resources, such as computers or statistical calculators, were widely available, tables (like Table A-2 in the book) were used to determine probabilities of events of normal random variables.

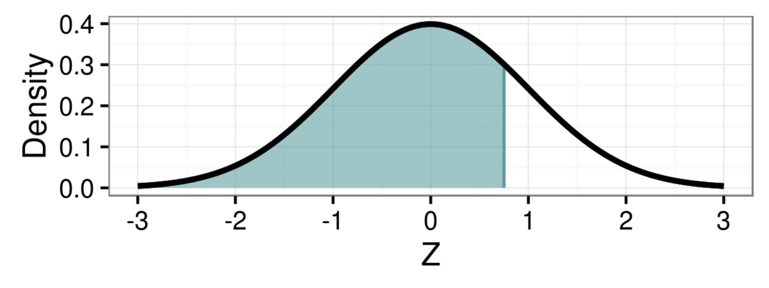
- The table gives the probability for values less than a specified z-score, $P(Z < z)$
- To find probabilities for values greater than a specified z-score, subtract table probability from one, $P(Z > z) = 1 - P(Z < z)$
- To find probabilities of ranges of values, subtract lower probability from higher, $P(z_1 < Z < z_2) = P(Z < z_2) - P(Z < z_1)$

However, using technology is usually quicker and more accurate.

Probabilities, example

Example

Using the standard normal distribution, find the probability a value is less than 0.75 standard deviations above the mean, $P(Z < 0.75)$

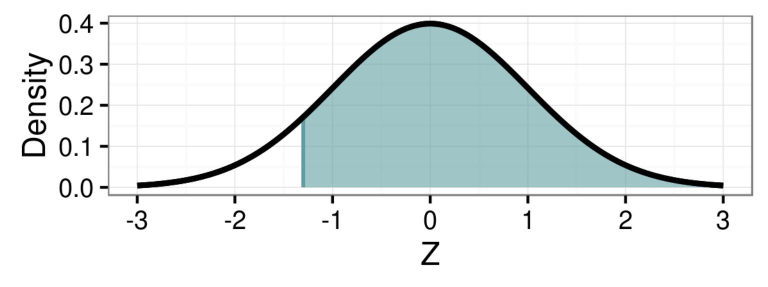


- $P(Z < .75) = \mathbf{0.773}$

Probabilities, example

Example

Using the standard normal distribution, find the probability a value is greater than 1.3 standard deviations below the mean, $P(Z > -1.3)$

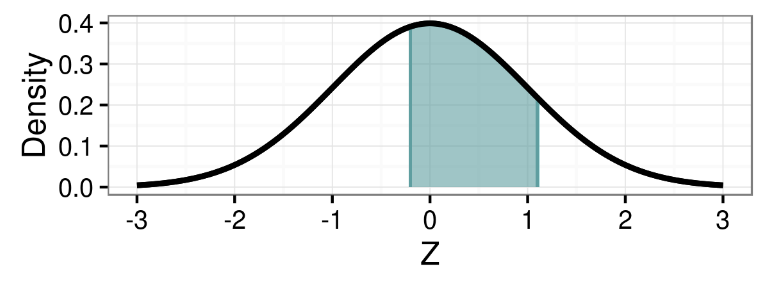


- $P(Z > -1.3) = \mathbf{0.903}$

Probabilities, example

Example

Using the standard normal distribution, find the probability a value is between -0.2 and 1.9 standard deviations, $P(-0.2 < Z < 1.1)$



- $P(-0.2 < Z < 1.1) = \mathbf{0.444}$

Finding percentiles

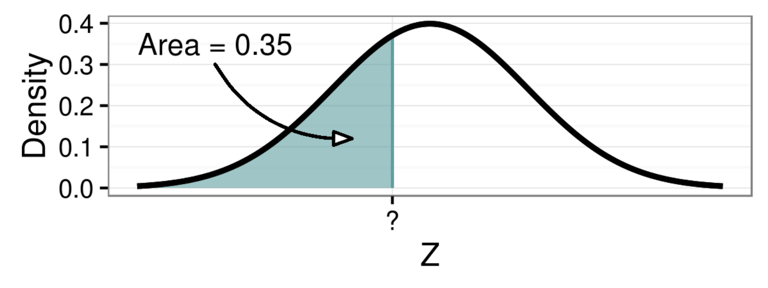
Often it is desirable to find a z-score that is greater than a specified probability, in other words, a percentile. This can be accomplished with the table by locating the desired probability and finding the corresponding z-score.

Again, technology provides an easier and more accurate method.

Finding percentiles, example

Example

What is the z-score greater than 35% of values? What is P_{35}? For what z-score is there a 0.35 probability of being less than $P(Z < z) = 0.35$?

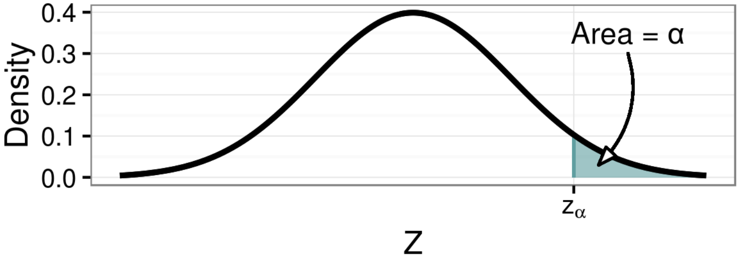


- $P(Z < \mathbf{-0.385}) = 0.35$

Critical values

In a standard normal distribution, the z-score separating usual outcomes from unusual outcomes is known as a **critical value**.

- The probability denoting unusual events is designated with α (alpha).
- Then z_α is the critical value such that $P(Z > z_\alpha) = \alpha$



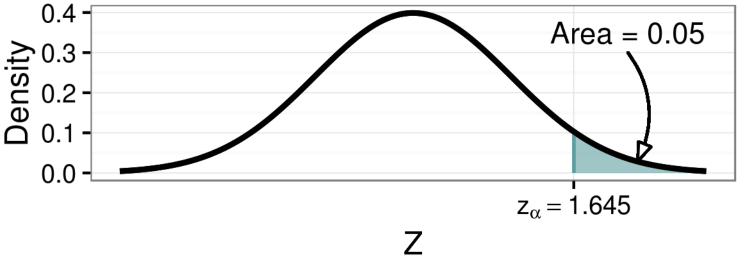
Critical values, example

Example

Let $\alpha = 0.05$.

Find the critical value for α. That is, find z_α or $z_{0.05}$.

- $z_\alpha = 1.645$
- $P(Z > z_\alpha) = \alpha$ or $P(Z < -z_\alpha) = \alpha$



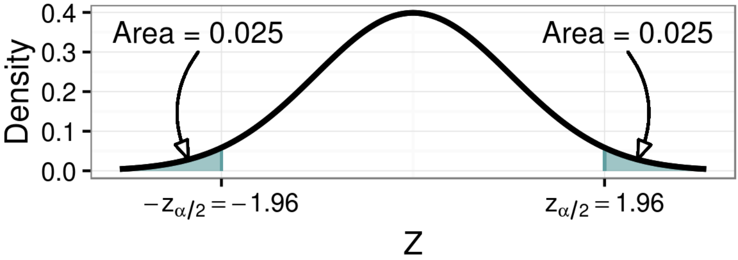
Critical values, example

Example

Let $\alpha = 0.05$.

Find the critical value for $\alpha/2$. That is, find $z_{\alpha/2}$ or $z_{0.025}$.

- $z_{\alpha/2} = 1.96$
- $P(Z < -z_{\alpha/2}) + P(Z > z_{\alpha/2}) = \alpha$



Section 6.3
Applications of Normal Distribution

Non-standard normal distributions

Any value from a normal distribution, given the mean μ and standard deviation σ, can be converted into a standard normal z-score with the transformation,

$$z = \frac{x - \mu}{\sigma}$$

- When using tables, this is the only way to find probabilities for non-standard normal random variables.
- With technology, this is no longer necessary.
- However, it is still useful to use z-scores for comparing values from different distributions.

Z-score comparison, example

Example

In the United States, adult women have a mean height of 63.7 in with a standard deviation of 5.96 in. Adult men have a mean height of 69.2 in with a standard deviation of 5.79 in.

Jane is 71 inches tall and Rafael is 74 inches tall. Who is taller, relative to their genders? Are either of them unusually tall?

- Jane: $z = \dfrac{x-\mu}{\sigma} = \dfrac{71-63.7}{5.96} = 1.22$
- Rafael: $z = \dfrac{74-69.2}{5.79} = 0.83$
- Jane is taller for a woman, than Rafael is for a man.
- Neither z-score is greater than 2 (or 1.96). Neither is unusual.

Values in non-standard normal distributions

Similarly, z-scores can be converted into values in any other normal distribution with the following transformation,

$$x = \mu + z\sigma$$

Example

A cruel statistics professor (not me) fails any student whose score on the final is less than 1.5 standard deviations below the mean. For one class, the scores on the final had a mean of 84.3 with a standard deviation of 7.8. What portion of the class does he fail? What is the boundary for failing scores for this class?

- $P(Z < -1.5) = 0.067$
- $x = \mu + z\sigma = 84.3 + (-1.5)(7.8) = 72.6$

Probabilities, percentiles and critical values

Probabilities, percentiles and critical values can all be found for non-standard normal distributions.

- Using tables, values are converted to z-scores, the relevant table look-up performed, and then converted back into original distribution.
- Again, technology makes the process easier.

Non-standard normal distributions, example

Example

An amusement park has made safety their highest priority. They design all their rides to have zero chance of causing serious head trauma, as long as the rider is under 78 inches tall. What proportion of men are in danger at this park?

- $X_m \sim N(69.2, 5.79)$
- $P(X_m > 78) = \mathbf{?}$
- $P(X_m > 78) = \mathbf{0.064}$

Non-standard normal distributions, example

Example

The amusement park is designing a new ride. It wants make sure that 85% of adult women can ride it safely. What is the maximum height the ride should be designed for?

- $X_f \sim N(63.7, 5.96)$
- $P(X_f < \mathbf{?}) = 0.85$
- $P(X_f < \mathbf{69.88}) = 0.85$

Non-standard normal distributions, example

Example

The amusement park is growing weary of accommodating the very tall and the very short. It has decided to exclude the most extreme heights among adult men. But it doesn't want to lose to much business, so it will only exclude 5% of the adult male population. What are the critical values for the tallest and shortest men, for a total of 5%?

- $X_m \sim N(69.2, 5.79), \qquad \alpha = 0.05$
- $z_{\alpha/2} = z_{0.025} = 1.96$
- $P(X_m < \mathbf{57.85}) = 0.025, \qquad P(X_m < \mathbf{80.55}) = 0.025$

Section 6.4
Sampling Distributions and Estimators

Samples, statistics and sampling distributions

Recall, a **sample** is a subset of a population. A **statistic** is a value calculated from the data of a sample.

A **sampling distribution** is a probability distribution of a statistic from all possible samples of a certain size from a population.

A sampling distribution is a mathematical construction. Understanding how statistics from samples are distributed, allows judgements to be made about the predictive value of individual samples that might be collected in real life.

Sampling distributions, example

Example

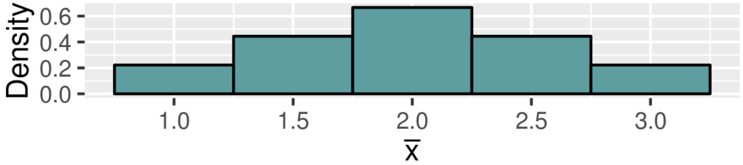
Suppose we have a population of 3 values: { 1, 2, 3 }.

There are 9 possible samples (with replacement) of size 2:
{ (1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3) }

The sample means, $\bar{x}$, are: { 1, 1.5, 2, 1.5, 2, 2.5, 2, 2.5, 3 }

The sampling distribution of the sample means is:

$\bar{x}$	1	1.5	2	2.5	3
Prob	1/9	2/9	3/9	2/9	1/9



Estimators

An **estimator** is statistic from a sample used to estimate a population parameter.

- For example, the sample mean $\bar{x}$ can be used to estimate the population mean μ.
- Any statistic can be used as an estimator. The population mean could be estimated by the constant value 4, but this is almost always a poor estimate (unless the population mean is, in fact, 4).

Unbiased estimators

If the expected value of an estimator, as calculated from the sampling distribution, is equal to the parameter it is estimating, the estimator is said to be **unbiased**.

- Sample mean, $\bar{x}$, is an unbiased estimator of population mean. μ.
- Sample standard deviation, s, is an unbiased estimator if population standard deviation, σ.
- Sample proportion, p, is an unbiased estimator for population proportion, π.
  - The book uses the notion of p for population proportion and $\hat{p}$ for sample proportion.

Recall, sample standard deviation is calculated as

$$s = \sqrt{\frac{1}{n-1} \sum (x - \bar{x})^2}$$

Section 6.5
Central Limit Theorem

Central Limit Theorem

The **Central Limit Theorem** (CLT) says that, given

- X is a random variable for a population with a mean μ and standard deviation σ
- A sampling distribution $S_{\bar{x}}$ of sample means $\bar{x}$ for samples of size n

Then,

- As n increases, $S_{\bar{x}}$ approaches a normal distribution
- The mean of $S_{\bar{x}}$, denoted $\mu_{\bar{x}}$, is μ
- The standard deviation of $S_{\bar{x}}$, denoted $\sigma_{\bar{x}}$ and known as the **standard error**, is $\dfrac{\sigma}{\sqrt{n}}$

Central Limit Theorem, example

Example

In the United States, adult women have a mean height of 63.7 in with a standard deviation of 5.96 in.

What is the probability that a sample of 10 adult women will have a mean height of less than 69.88 inches,

- The mean of the sample means is 63.7
- The standard deviation of the sample means, or standard error, is $\dfrac{5.96}{\sqrt{30}} = 1.88$
- $P(S_{\bar{x}} < 69.88) = 0.9995$

Thoughts on CLT

- If the population X has a normal distribution, the sampling distribution $S_{\bar{x}}$ will have a normal distribution regardless of sample size n.
- If X is not normally distributed, how normal $S_{\bar{x}}$ is, or how quickly it becomes normal as n increases, depends on how not normal X is.
- The rule of thumb generally used is, if sample size is $n = 30$ or greater, $S_{\bar{x}}$ can be considered normal.

Elevator example

Example

Suppose an elevator has a maximum capacity of 16 passengers with a total weight of 2500 lb. Assume male weights follow a normal distribution with a mean of 182.9 lb and a standard deviation of 40.8 lb.

Find the probability that 1 randomly selected male has a weight greater than 156.25 lb (2500 lbs./16).

- $X \sim N(182.9, 40.8)$
- $P(X > 156.25) = 0.743$

Find the probability that a sample of 16 males have a mean weight greater than 156.25 lb.

- $\mu_{\bar{x}} = 182.9, \qquad \sigma_{\bar{x}} = \frac{40.8}{\sqrt{16}} = 10.2$
- $S_{\bar{x}} \sim N(182.9, 10.2)$
- $P(S_{\bar{x}} > 156.25) = 0.996$