

Stat 201: Statistics I

Week 6



Week 6

Binomial and Normal Distributions

Section 6.1

Binomial Probability Distributions

Binomial distribution

A **binomial distribution** is a probability distribution representing the number of “successes” in a fixed number of trials with two possible outcomes.

- The term “success” is traditional, but can refer to any outcome of interest.

Example

- The number of students passing the midterm in class
 - Passing the midterm is a “success”, failing it is a “failure”
- The number of heads in three coin flips
 - Getting a head is a “success”, getting a tail is a “failure”
- The number of of car crashes that result in fatalities
 - A fatality is a “success”, no fatalities is a “failure”

Requirements for binomial distributions

There are four requirements to be considered a binomial distribution:

- The distribution must represent a fixed number of trials
- Each trial must have exactly two possible outcomes (success / failure)
- Each trial must be independent of the other trials
- Each trial must have the same probability for success

The last two requirements are often summarized as “independent and identically distributed” and abbreviated as “iid”.

Requirements for binomial distributions, example

Example

Recall the Metro State Statistics Club, with 6 male members and 4 females members. If the three officer positions are selected randomly, does the number of women selected follow a binomial distribution?

- No. The events are not independent.

Example

An instructor is trying a new method of testing students. A test is given once a day for five days, with opportunities to practice and ask questions in between each test. Does the number of times a student passes the test follow a binomial distribution?

- No. The probability of success changes (hopefully) with each test.

Requirements for binomial distributions, example

Example

Recall the Youth Risk Behavior Survey (YRBS) which found the probability of a teenaged driver had texted or emailed while driving was 0.404. Suppose this is the probability for the population of teenaged drivers. Suppose 30 teenaged drivers are selected at random. Does the number of those drivers that had texted or emailed while driving follow a binomial distribution?

- Yes.
 - There is a fixed number of trials (30).
 - Each trial has only two possible outcomes (had or had not texted).
 - Each trial is independent.
 - Each trial has the same probability of “success.”

Notation for binomial distributions

$$X \sim \text{Binom}(n, p)$$

- The random variable X follows a binomial distribution with n trials and p probability of a success.
- $q = 1 - p$ = the probability of failure
- $P(X = x)$ is the probability of getting exactly x successes. x must be between 0 and n .

Example

The probability a teenaged driver had texted or emailed while driving is 0.404. If the random variable Y is the number of teenaged drivers who had texted or emailed while driving out of a sample of 30,

$$Y \sim \text{Binom}(30, 0.404)$$

Probabilities for binomial distributions

The formula for binomial probabilities is

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

- $\binom{n}{x}$ is read as “ n choose x ”. It is the number of ways to get x successes in n trials. For example, we previously determined that there were three ways to get two heads in three flips, { HHT, HTH, THH }. Thus, $\binom{3}{2}$ is 3.
- $p^x q^{n-x}$ is the probability of getting x successes and $n - x$ failures in one particular order.

Practice: cancer screening

Recall the cancer screening example. Out of 1000 people screened, the probability of a positive test was 0.1. Suppose a sample 40 of those screened were randomly selected.

The random variable X is the number who tested positive in the sample. Does X follow a binomial distribution?

- Yes. All the requirements are met.

What is the notation for X ?

- $X \sim \text{Binom}(40, 0.1)$

What is the probability that exactly 5 people in the sample tested positive?

- $P(X = 5) = 0.165$

Parameters for Binomial Distributions

For $X \sim \text{Binom}(n, p)$,

- Mean:

$$\mathbb{E}(X) = \mu = np$$

- Variance:

$$\sigma^2 = npq$$

- Standard deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{npq}$$

Parameters, example

Example

The random variable Y is the number of teenaged drivers who had texted or emailed while driving out of a sample of 30,

$$Y \sim \text{Binom}(30, 0.404)$$

Parameters of Y :

- $\mathbb{E}(Y) = \mu = np = (30)(0.404) = 12.12$
- $\sigma^2 = npq = (30)(0.404)(1 - 0.404) = 7.22$
- $\sigma = \sqrt{\sigma^2} = \sqrt{7.22} = 2.69$

Unusual/significant values

For a given binomial distribution, the boundaries for unusual values can be found. From the range rule of thumb, the lower boundary is $\mu - 2\sigma$ and the upper boundary is $\mu + 2\sigma$.

Example

For $Y \sim \text{Binom}(30, 0.404)$, $\mu = 12.12$ and $\sigma = 2.69$. What are the unusual values?

- The lower bound for unusual values is $\mu - 2\sigma = 6.74$
- The upper bound for unusual values is $\mu + 2\sigma = 17.5$
- In a random sample of 30 teenaged drivers, it would be unusual to get 6 or fewer, or 18 or more, drivers who had texted or emailed while driving.

Section 6.2

Normal Distributions

Continuous probability distributions

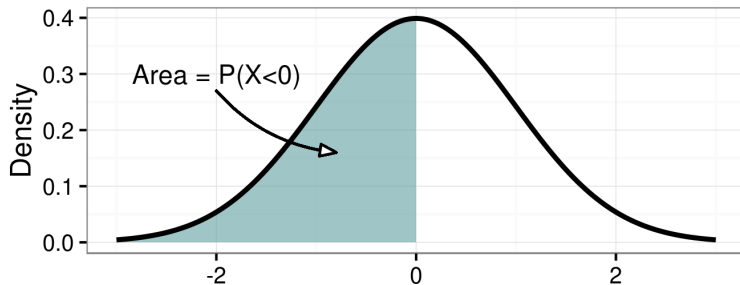
A **continuous probability distribution** is a description of the probabilities of all possible values of a continuous random variable.

- 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 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.



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”

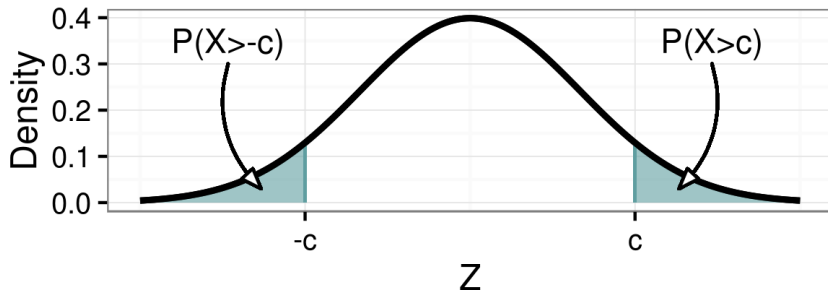
Formally, random variable X has a normal distribution with mean μ and standard deviation σ ...

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

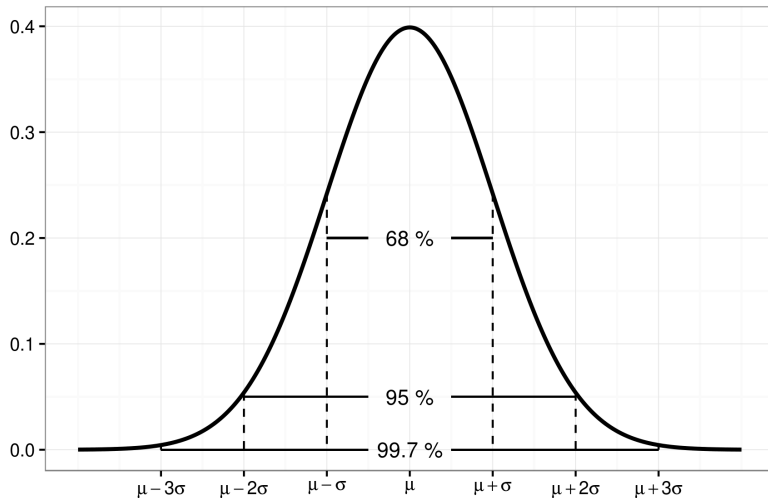
Symmetry of normal distribution

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

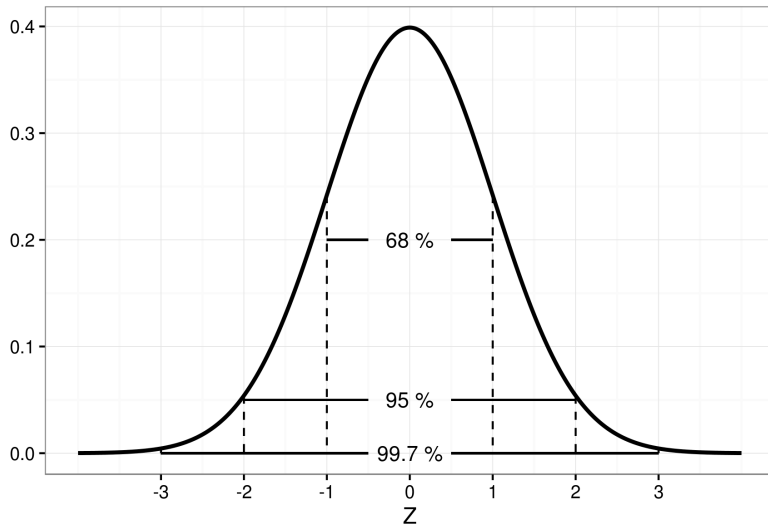


Standard normal distribution

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 were used to determine probabilities of events of normal random variables.

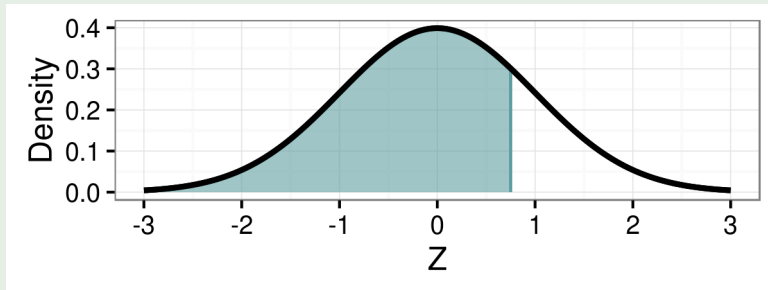
- The table gives the probability for values less than a specified z -score, $P(Z < z)$
- To find probabilities for values greater than a specified z -score, subtract table probability from one, $P(Z > z) = 1 - P(Z < z)$
- To find probabilities of ranges of values, subtract lower probability from higher,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

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, $P(Z < 0.75)$

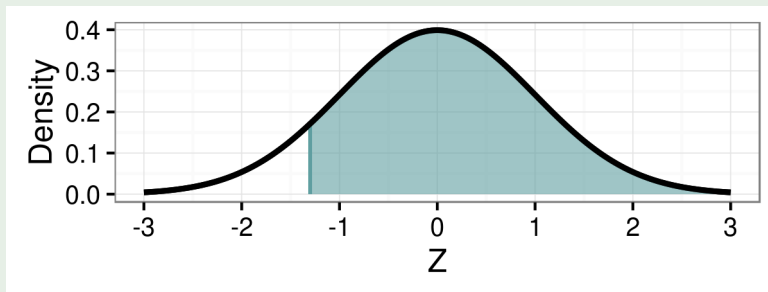


- $P(Z < .75) = 0.773$

Probabilities, example

Example

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

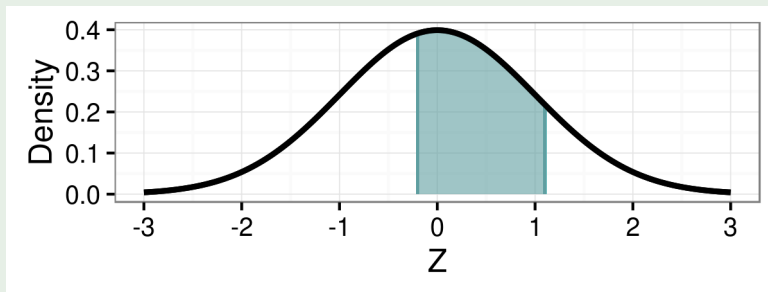


- $P(Z > -1.3) = 0.903$

Probabilities, example

Example

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



- $P(-0.2 < Z < 1.1) = 0.444$

Finding percentiles

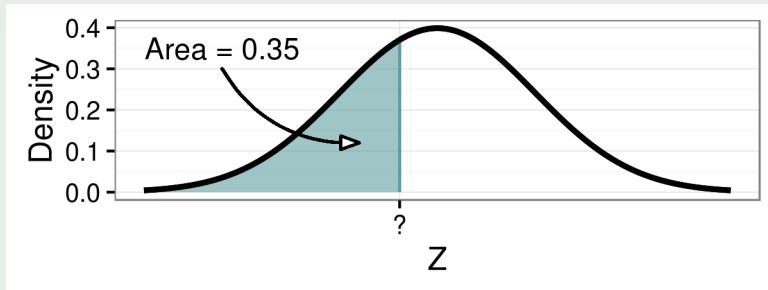
Often it is desirable to find a z -score that is greater than a specified probability, in other words, a percentile. This can be accomplished with the table by locating the desired probability and finding the corresponding z -score.

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 $P(Z < z) = 0.35$?

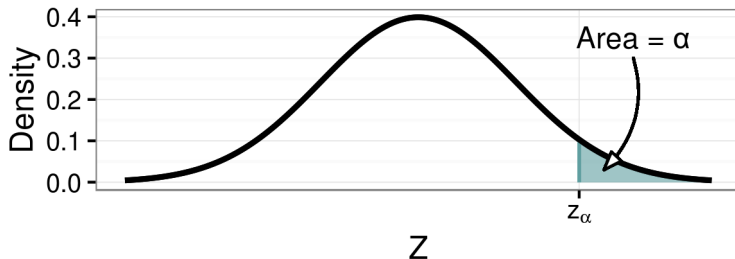


- $P(Z < -0.385) = 0.35$

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 $P(Z > z_\alpha) = \alpha$



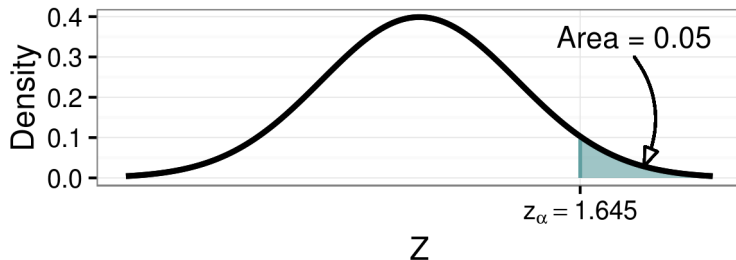
Critical values, example

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 $P(Z < -z_\alpha) = \alpha$



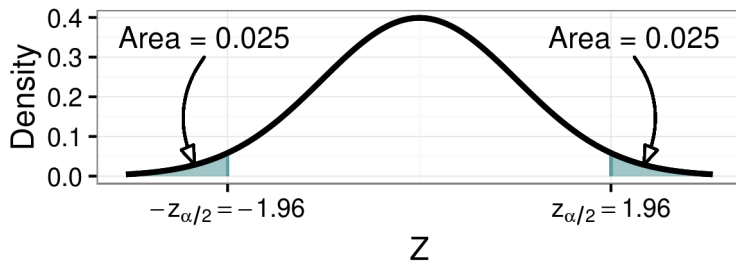
Critical values, example

Example

Let $\alpha = 0.05$.

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

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



Non-standard normal distributions

Non-standard normal distributions, that is normal distributions with a mean other than 0 and/or standard deviation other than 1, are as easy to work with as the standard normal distribution with the help of technology.

- However, it is still useful to use z -scores for comparing values from different distributions.
- Recall, to convert a value x from a non-standard normal distribution to a z -score, and vice versa,

$$z = \frac{x - \mu}{\sigma} \quad \text{and} \quad x = \mu + z\sigma$$

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?

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

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?

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

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 too 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%?

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