STAT 583 Lecture 1

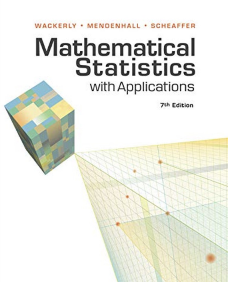
Linjun (Leon) Zhang
linjun.zhang@rutgers.edu

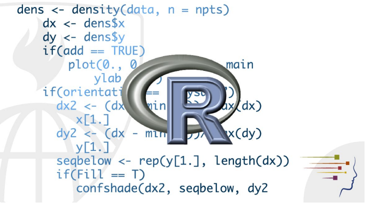
Department of Statistics
Rutgers University

Jan 21, 2020

Welcome to the class!

- Textbook: Math Stat with Applications, 7th Ed. Wackerley, et al.
- Software: R





What you should already know

- Probability at the level of STAT 582
  - Random variables
  - Probability distributions, probability density/mass function
  - Mean/variance/SD/quantile of a distribution
  - Transformation of a random variable
  - Jointly distributed random variables, conditional probability, independence
  - Covariance, correlation
  - Normal distribution, binomial distribution
  - Moment generating function
  - Law of large numbers, central limit theorem
  - ...
- Calculus will also be used

Syllabus overview

- Overview:
  - Basic concepts and probability review (today)
  - Point estimation and confidence intervals ($\sim$4 weeks)
  - Hypothesis testing ($\sim$4 weeks)
  - Linear regression ($\sim$3 weeks)
- Homework
- Midterm/Quiz and Final exams
- Grading: 40% homework, 20% midterm, and 40% final examination.
- Lecture slides will be posted on Canvas.

Survey

Probability and Statistics

After the course of probability theory, we are now ready to learn statistics.

What's Statistics?

BIG question: How do we learn from data?

We will cover

- Methodology: Examine a collection of important statistical concepts and methods
- Theory: Understand when, how, and why to apply these methods
- Computing: Apply these methods to real world data

What's Statistics?

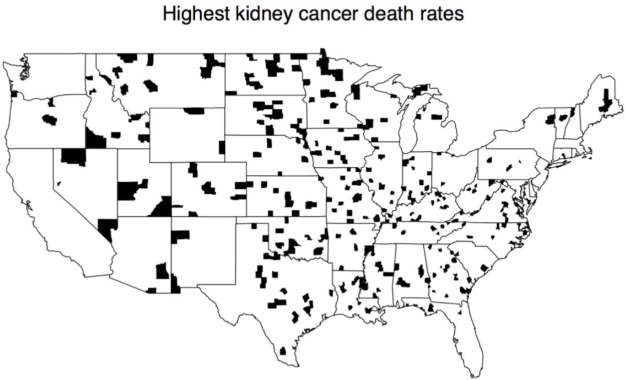
What's Data?

$$\text{Data} = \text{Signal} + \text{Random Noise}.$$

The goal of statistics, is to find Signal from the "noisy data".

Statistical Thinking: Example 1

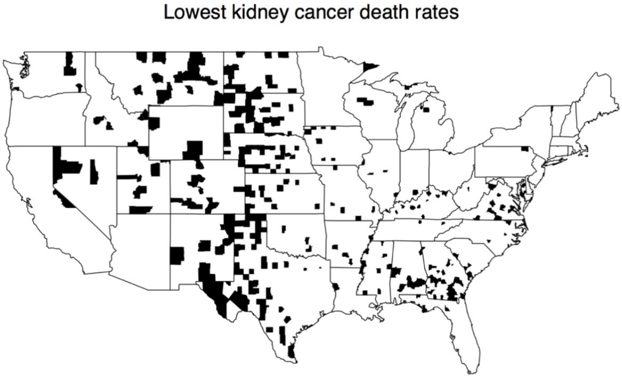
The counties with the highest kidney cancer death rates



Data observer: the shaded parts are mostly in rural areas.

Statistical Thinking: Example 1

The counties with the lowest kidney cancer death rates



Data observer: the shaded parts are mostly in rural areas, again.

Statistical Thinking: Example 1

- The counties in rural areas have less people

Statistical Thinking: Example 1

- The counties in rural areas have less people
- Consider a county with 100 people: 0 death v.s. 1 death

Statistical Thinking: Example 1

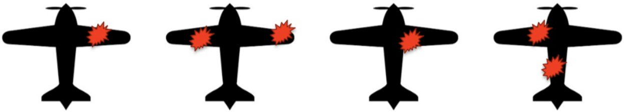
- The counties in rural areas have less people
- Consider a county with 100 people: 0 death v.s. 1 death
- The observed rates for smaller counties are much more variable

Statistical Thinking: Example 1

- The counties in rural areas have less people
- Consider a county with 100 people: 0 death v.s. 1 death
- The observed rates for smaller counties are much more variable
- Sample size of data matters!

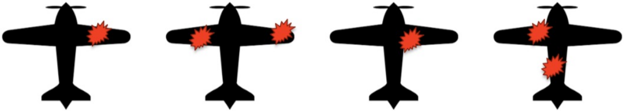
Statistical Thinking: Example 2

- Statistics played a hugely important role in WWII
- The statistician Abraham Wald studied losses of bombers
- Looked at where damage appeared on returning aircraft
- Where should aircraft be reinforced?



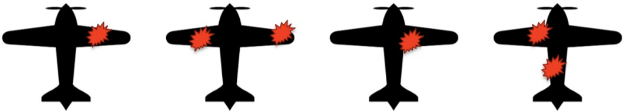
Statistical Thinking: Example 2

- Wald recommended reinforcing area of aircraft with NO damage



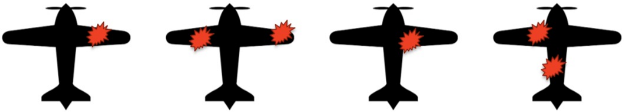
Statistical Thinking: Example 2

- Wald recommended reinforcing area of aircraft with NO damage
- Missing data: aircraft that do not return



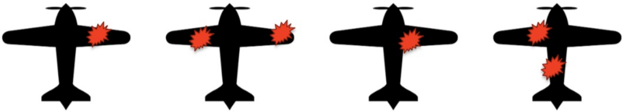
Statistical Thinking: Example 2

- Wald recommended reinforcing area of aircraft with NO damage
- Missing data: aircraft that do not return
- If the bomber is damaged



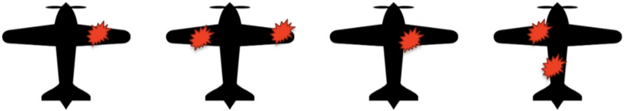
Statistical Thinking: Example 2

- Wald recommended reinforcing area of aircraft with NO damage
- Missing data: aircraft that do not return
- If the bomber is damaged
  - where you see damage in your data, the bomber can make it back



Statistical Thinking: Example 2

- Wald recommended reinforcing area of aircraft with NO damage
- Missing data: aircraft that do not return
- If the bomber is damaged
  - where you see damage in your data, the bomber can make it back
  - where you don't, the bomber crashes–it never makes it into the data set



Three Statistical Tasks

Given a question of interest...

1. Collect data
   - Which variables should be measured and how?
2. Summarize and explore data
   - Its hard to think about a long list of numbers
   - Better to use summary statistics / tables / graphical displays
3. Draw conclusions and make decisions based on data (inferential statistics)
   - Modeling: turning question of interest into numerical conjectures about parameters in the model
   - Inference on the model parameters: estimation / confidence intervals / hypothesis testing

Statistics

Overview of this lecture

- The relationship between statistics and probability theory
    - Data and random variables
    - Sample and population
    - Estimates and parameters
- Sampling distributions
    - Central Limit Theorem
    - chi-squared distribution and t-distribution

Statistics and Probability Theory

- Recall that

$$\text{Data} = \text{Signal} + \text{Random Noise}.$$

- Recall that

$$\text{Data} = \text{Signal} + \text{Random Noise}.$$

- We have learned how to describe the "random noise" in Probability Theory.

Statistics and Probability Theory

- Recall that

$$\text{Data} = \text{Signal} + \text{Random Noise}.$$

- We have learned how to describe the "random noise" in Probability Theory.
- Main idea: Use tools of probability theory to explore how to "estimate" signal from data.

Probability and Statistics

- In probability theory, we assume a model P first, then generate $X_1, \ldots, X_n$.

Probability and Statistics

- In probability theory, we assume a model P first, then generate $X_1, \ldots, X_n$.
- The whole statistical theory works when $X_1, \ldots, X_n$ are drawn from the true probability distribution P.

Probability and Statistics

- In probability theory, we assume a model P first, then generate $X_1, ..., X_n$.
- The whole statistical theory works when $X_1, ..., X_n$ are drawn from the true probability distribution P.
- P has some shape, has mean μ, has variance σ^2, but in real life we don't see or know these values. We do not see or know the true distribution. We just assume that it exists. All we see is data which are drawn from P.

Probability and Statistics

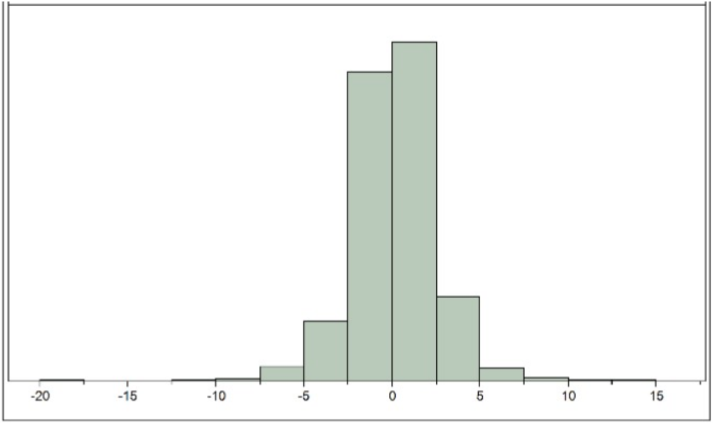
- In probability theory, we assume a model P first, then generate $X_1, ..., X_n$.
- The whole statistical theory works when $X_1, ..., X_n$ are drawn from the true probability distribution P.
- P has some shape, has mean μ, has variance σ^2, but in real life we don't see or know these values. We do not see or know the true distribution. We just assume that it exists. All we see is data which are drawn from P.
- No one knows P. Key idea of statistical analysis is that we can learn about P from the data we observe. We do know X_i's , don't we?

Probability and Statistics

- In probability theory, we assume a model P first, then generate $X_1, ..., X_n$.
- The whole statistical theory works when $X_1, ..., X_n$ are drawn from the true probability distribution P.
- P has some shape, has mean μ, has variance σ^2, but in real life we don't see or know these values. We do not see or know the true distribution. We just assume that it exists. All we see is data which are drawn from P.
- No one knows P. Key idea of statistical analysis is that we can learn about P from the data we observe. We do know X_i's , don't we?
- If we assume that data are drawn from the same distribution, then what model (distribution) can we assume behind that the real data? Look at the data! If we have lots of data and it looks normal, then by the Law of Large Numbers the true probability distribution is also normal.

Histogram

Explore data by using histogram



Histogram

A histogram is a graphical representation of the distribution of numerical data.

- Construction: create buckets (bins) for the data. For example, (0%, 2.5%), (2.5%, 5%) etc, and count the number of observations in each bucket.
- The buckets are on the horizontal axis and the height of the bars show the relative frequency counts of the number of observations in each bucket.
- Higher bars indicate more observations.
- On reviewing the histogram, we learn that the observations are centered about zero and that most of them lie within the buckets (-2.5%, 0%) and (0%, 2.5%) but that there are some extreme data points too. The shape of the distribution is very symmetric.

Law of Large Numbers

First Law of Large Numbers

The histogram of the dataset of n independent identically distributed (i.i.d.) random variables with pdf $p(x)$ gets closer and closer to the pdf $p(x)$.

Second Law of Large Numbers

If $\bar{X}$ is a sample average of n independent random variables, then the SD of $\bar{X}$ gets closer and closer to 0. Thus, $\bar{X}$ gets closer and closer to μ.

What is data; Data and random variables

Suppose $X_1, \ldots, X_n$ are drawn from the same true probability distribution P.

- The concept of data is in contrast to what we mean by the word random variable.

What is data; Data and random variables

Suppose $X_1, ..., X_n$ are drawn from the same true probability distribution P.

- The concept of data is in contrast to what we mean by the word random variable.
- By data we mean the observed value of a random variable once some experiment has been performed. It is crucial to distinguish between the before the experiment concept of a random variable and the after the experiment concept of data.

What is data; Data and random variables

Suppose $X_1, ..., X_n$ are drawn from the same true probability distribution P.

- The concept of data is in contrast to what we mean by the word random variable.
- By data we mean the observed value of a random variable once some experiment has been performed. It is crucial to distinguish between the before the experiment concept of a random variable and the after the experiment concept of data.
- Coin flipping example:

What is data; Data and random variables

Suppose $X_1, ..., X_n$ are drawn from the same true probability distribution P.

- The concept of data is in contrast to what we mean by the word random variable.
- By data we mean the observed value of a random variable once some experiment has been performed. It is crucial to distinguish between the before the experiment concept of a random variable and the after the experiment concept of data.
- Coin flipping example:
  - We are going to flip a coin and let $X = 1$ if the coin turns up head: random variable

What is data; Data and random variables

Suppose $X_1, ..., X_n$ are drawn from the same true probability distribution P.

- The concept of data is in contrast to what we mean by the word random variable.
- By data we mean the observed value of a random variable once some experiment has been performed. It is crucial to distinguish between the before the experiment concept of a random variable and the after the experiment concept of data.
- Coin flipping example:
  - We are going to flip a coin and let $X = 1$ if the coin turns up head: random variable
  - We flipped the coin, and found that $X = 1$, which implies that the coin turned up head: data

What is a sample?

- A sample is any collection of random variables/data.
  - $X_1, X_2, ..., X_n$ a sequence of $i.i.d.$ random variables with common distribution P.
  - $X_1, X_2, ..., X_n$ a sequence of draws from a population.

What is a sample?

- A sample is any collection of random variables/data.
  - $X_1, X_2, ..., X_n$ a sequence of $i.i.d.$ random variables with common distribution P.
  - $X_1, X_2, ..., X_n$ a sequence of draws from a population.
- n is called the sample size

What is a sample?

- A sample is any collection of random variables/data.
  - $X_1, X_2, ..., X_n$ a sequence of $i.i.d.$ random variables with common distribution P.
  - $X_1, X_2, ..., X_n$ a sequence of draws from a population.
- n is called the sample size
- Example: Roll a die 25 times and record the outcomes.

What is a sample?

- A sample is any collection of random variables/data.
  - $X_1, X_2, ..., X_n$ a sequence of $i.i.d.$ random variables with common distribution P.
  - $X_1, X_2, ..., X_n$ a sequence of draws from a population.
- n is called the sample size
- Example: Roll a die 25 times and record the outcomes.
- Example: Write 100 numbers on tickets, place in a box, draw out 50 tickets at random without replacing the tickets.

What is a sample?

- A sample is any collection of random variables/data.
  - $X_1, X_2, ..., X_n$ a sequence of $i.i.d.$ random variables with common distribution P.
  - $X_1, X_2, ..., X_n$ a sequence of draws from a population.
- n is called the sample size
- Example: Roll a die 25 times and record the outcomes.
- Example: Write 100 numbers on tickets, place in a box, draw out 50 tickets at random without replacing the tickets.
- Example: Choose 10 digit phone numbers at random, dial, then ask the responders a question. Record the responses.

The population-sample paradigm

We can view population as a box containing an infinite numbers of X_i's drawn from a distribution P.

- Population: the entire collection of X_i's.
- Sample: the part (subset) of the population chosen.

Why sample?

- The sample was drawn in order to learn a particular characteristic of the population.

Why sample?

- The sample was drawn in order to learn a particular characteristic of the population.
- Terminology:

Why sample?

- The sample was drawn in order to learn a particular characteristic of the population.
- Terminology:
  - A population characteristic is called a parameter (usually an unknown constant).

Why sample?

- The sample was drawn in order to learn a particular characteristic of the population.
- Terminology:
  - A population characteristic is called a parameter (usually an unknown constant).
  - The process of using samples to learn about an unknown population parameter is called statistical inference.

Why sample?

- The sample was drawn in order to learn a particular characteristic of the population.
- Terminology:
  - A population characteristic is called a parameter (usually an unknown constant).
  - The process of using samples to learn about an unknown population parameter is called statistical inference.
- So, parameters of the population are estimated by statistics (estimates) computed from a sample.

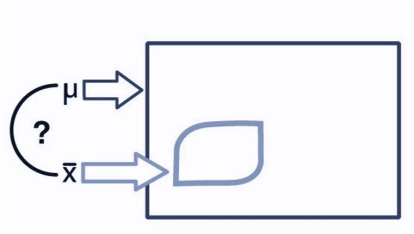
Why sample?

- The sample was drawn in order to learn a particular characteristic of the population.
- Terminology:
  - A population characteristic is called a parameter (usually an unknown constant).
  - The process of using samples to learn about an unknown population parameter is called statistical inference.
- So, parameters of the population are estimated by statistics (estimates) computed from a sample.
- We want to know parameters, but can only calculate statistics.

The population-sample paradigm

The following schematic present the population as the dark blue rectangle and the sample as the light blue embedded subset. The unknown population mean μ is the parameter of interest and the sample mean $\bar{x}$ is used to estimate it.

Figure 1: An illustration of the population/sample paradigm



Parameters, r.v., data

- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X.

Parameters, r.v., data

- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X.
    - What are the parameter, random variable here?

Parameters, r.v., data

- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X.
  - What are the parameter, random variable here?
  - p is the parameter, X is a random variable

Parameters, r.v., data

- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X.
  - What are the parameter, random variable here?
  - p is the parameter, X is a random variable
  - I flipped the coin 100 times, and get $X = x$.

Parameters, r.v., data

- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X.
  - What are the parameter, random variable here?
  - p is the parameter, X is a random variable
  - I flipped the coin 100 times, and get $X = x$.
  - What are the data, and sample size?

Parameters, r.v., data

- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X.
    - What are the parameter, random variable here?
    - p is the parameter, X is a random variable
    - I flipped the coin 100 times, and get $X = x$.
    - What are the data, and sample size?
    - x is the data, the sample size is 1.

Parameters, r.v., data

- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X.
  - What are the parameter, random variable here?
  - p is the parameter, X is a random variable
  - I flipped the coin 100 times, and get $X = x$.
  - What are the data, and sample size?
  - x is the data, the sample size is 1.
- We are estimating the expected return of the stock Amazon, and we model the data by a normal distribution $N(\mu, \sigma^2)$. We recorded the stock prices in the past 100 days as $X_1, \ldots, X_{100}$.

Parameters, r.v., data

- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X.
  - What are the parameter, random variable here?
  - p is the parameter, X is a random variable
  - I flipped the coin 100 times, and get $X = x$.
  - What are the data, and sample size?
  - x is the data, the sample size is 1.
- We are estimating the expected return of the stock Amazon, and we model the data by a normal distribution $N(\mu, \sigma^2)$. We recorded the stock prices in the past 100 days as $X_1, ..., X_{100}$.
  - What are the parameter, random variable here?

Parameters, r.v., data

- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X.
  - What are the parameter, random variable here?
  - p is the parameter, X is a random variable
  - I flipped the coin 100 times, and get $X = x$.
  - What are the data, and sample size?
  - x is the data, the sample size is 1.
- We are estimating the expected return of the stock Amazon, and we model the data by a normal distribution $N(\mu, \sigma^2)$. We recorded the stock prices in the past 100 days as $X_1, ..., X_{100}$.
  - What are the parameter, random variable here?
  - μ, σ^2 are parameters. $X_1, ..., X_{100}$ are data.

Statistic

A statistic is a functions of data/random variable. When we use a statistic to estimate an interested parameter, we call this statistic an estimate.

- Note: an estimator is a rule for calculating an estimate of a given parameter based on observed data

Statistic

A statistic is a functions of data/random variable. When we use a statistic to estimate an interested parameter, we call this statistic an estimate.

- Note: an estimator is a rule for calculating an estimate of a given parameter based on observed data
- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X. I decide to use $\hat{p} = \frac{X}{n}$ to estimate p.

Statistic

A statistic is a functions of data/random variable. When we use a statistic to estimate an interested parameter, we call this statistic an estimate.

- Note: an estimator is a rule for calculating an estimate of a given parameter based on observed data
- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X. I decide to use $\hat{p} = \frac{X}{n}$ to estimate p.
- p is the parameter, $\hat{p}$ is a statistic or an estimate.

Statistic

A statistic is a functions of data/random variable. When we use a statistic to estimate an interested parameter, we call this statistic an estimate.

- Note: an estimator is a rule for calculating an estimate of a given parameter based on observed data
- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X. I decide to use $\hat{p} = \frac{X}{n}$ to estimate p.
- p is the parameter, $\hat{p}$ is a statistic or an estimate.
- I flipped the coin 100 times, and get $X = x$, then $\hat{p} = \frac{x}{n}$

Statistic

A statistic is a functions of data/random variable. When we use a statistic to estimate an interested parameter, we call this statistic an estimate.

- Note: an estimator is a rule for calculating an estimate of a given parameter based on observed data
- A coin is unfair with $\mathbb{P}(H) = p$. To find p, I am going to flip the coin 100 times, and denote the number of heads I will get as X. I decide to use $\hat{p} = \frac{X}{n}$ to estimate p.
- p is the parameter, $\hat{p}$ is a statistic or an estimate.
- I flipped the coin 100 times, and get $X = x$, then $\hat{p} = \frac{x}{n}$
- We still call $\hat{p}$ as a statistic, or an estimate.

Statistic

- We are estimating the expected return rate and volatility of the stock Amazon, and we model the data by a normal distribution $N(\mu, \sigma^2)$. We recorded the stock prices in the past 100 days as $X_1, ..., X_{100}$, and estimate μ by $\hat{\mu} = \bar{X} = \frac{X_1 + ... + X_{100}}{100}$.

Statistic

- We are estimating the expected return rate and volatility of the stock Amazon, and we model the data by a normal distribution $N(\mu, \sigma^2)$. We recorded the stock prices in the past 100 days as $X_1, \ldots, X_{100}$, and estimate μ by $\hat{\mu} = \bar{X} = \frac{X_1 + \ldots + X_{100}}{100}$.
- μ, σ^2 are parameters. $\hat{\mu}$ is a statistic or an estimate.

Statistic

- We are estimating the expected return rate and volatility of the stock Amazon, and we model the data by a normal distribution $N(\mu, \sigma^2)$. We recorded the stock prices in the past 100 days as $X_1, ..., X_{100}$, and estimate μ by $\hat{\mu} = \bar{X} = \frac{X_1 + ... + X_{100}}{100}$.
- μ, σ^2 are parameters. $\hat{\mu}$ is a statistic or an estimate.
- We also estimate σ^2 by

$$s^2 = \frac{(X_1 - \hat{\mu})^2 + ... + (X_{100} - \hat{\mu})^2}{100}.$$

Statistic

- We are estimating the expected return rate and volatility of the stock Amazon, and we model the data by a normal distribution $N(\mu, \sigma^2)$. We recorded the stock prices in the past 100 days as $X_1, ..., X_{100}$, and estimate μ by $\hat{\mu} = \bar{X} = \frac{X_1 + ... + X_{100}}{100}$.
- μ, σ^2 are parameters. $\hat{\mu}$ is a statistic or an estimate.
- We also estimate σ^2 by

$$s^2 = \frac{(X_1 - \hat{\mu})^2 + ... + (X_{100} - \hat{\mu})^2}{100}.$$

- μ, σ^2 are parameters. The goral of Statistics is to learn about these parameters. $\hat{\mu}, s^2$ are statistics or estimates.

Intuition

- Recall that $\sigma^2 = \mathbb{E}((X-\mu)^2)$.
- Let $Y_i = (X_i - \mu)^2$, we then have Y_i's are $i.i.d.$ with mean $\mathbb{E}(Y_i) = \mathbb{E}((X_i-\mu)^2) = \sigma^2$.
- We estimate σ^2 by $\bar{Y}$, and approximate μ by $\hat{\mu}$, we have

$$s^2 = \frac{(X_1-\hat{\mu})^2 + \ldots + (X_n - \hat{\mu})^2}{n}.$$

Sampling Distributions

Sampling Distributions

- Recall that a sample is any sequence of observations drawn from a common distribution P: random variables $X_1, ..., X_n \sim P$.

Sampling Distributions

- Recall that a sample is any sequence of observations drawn from a common distribution P: random variables $X_1, ..., X_n \sim P$.
- The sample is $i.i.d.$ if the random variables are also independent.

Sampling Distributions

- Recall that a sample is any sequence of observations drawn from a common distribution P: random variables $X_1, ..., X_n \sim P$.
- The sample is *i.i.d.* if the random variables are also independent.
- In statistics, we construct samples to learn about a population: in such cases the sample is
  - a sequence of draws with or without replacement, from a deck of cards with numbers on them.
  - each card represents an individual in the population.

Sampling Distributions

- Recall that the sample mean, $\bar{X}$ is defined as follows:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- MAJOR POINT: the sample mean is a random variable. It has an expected value, a standard deviation and a probability distribution.

The R experiment

From the R experiment, we can see the following points:

- Law of Large Numbers: $\bar{X} \to 3.5$, as n increases.

The R experiment

From the R experiment, we can see the following points:

- Law of Large Numbers: $\bar{X} \to 3.5$, as n increases.
- If we run the simulation "roll a fair die 1000 times" once, we get one single number $\bar{X} = \frac{X_1 + \ldots + X_{1000}}{1000} = 3.457$.

The R experiment

From the R experiment, we can see the following points:

- Law of Large Numbers: $\bar{X} \to 3.5$, as n increases.
- If we run the simulation "roll a fair die 1000 times" once, we get one single number $\bar{X} = \frac{X_1 + \ldots + X_{1000}}{1000} = 3.457$.
- If we run the simulation "roll a fair die 1000 times" many times, we get different values of $\bar{X}$.

Example

- Suppose I roll a die 1000 times. Let $X_1, X_2, ..., X_{1000}$ be the random variables representing the rolls. It is obvious that $\bar{X}$, the average of the rolls before the experiment, is a random variable.

Example

- Suppose I roll a die 1000 times. Let $X_1, X_2, ..., X_{1000}$ be the random variables representing the rolls. It is obvious that $\bar{X}$, the average of the rolls before the experiment, is a random variable.
- Now let's reveal the rolls. The sample mean is $\bar{x} = 3.457$.

Example

- Suppose I roll a die 1000 times. Let $X_1, X_2, ..., X_{1000}$ be the random variables representing the rolls. It is obvious that $\bar{X}$, the average of the rolls before the experiment, is a random variable.
- Now let's reveal the rolls. The sample mean is $\bar{x} = 3.457$.
- Point of confusion: $\bar{x}$ is a number (3.457 in fact). It sure doesn't look random! Why not?

Example

- Suppose I roll a die 1000 times. Let $X_1, X_2, ..., X_{1000}$ be the random variables representing the rolls. It is obvious that $\bar{X}$, the average of the rolls before the experiment, is a random variable.
- Now let's reveal the rolls. The sample mean is $\bar{x} = 3.457$.
- Point of confusion: $\bar{x}$ is a number (3.457 in fact). It sure doesn't look random! Why not?
- After the sample is drawn (in this case, the dice are rolled) the sample mean becomes data (a single observation of a random variable).

Example

- Suppose I roll a die 1000 times. Let $X_1, X_2, ..., X_{1000}$ be the random variables representing the rolls. It is obvious that $\bar{X}$, the average of the rolls before the experiment, is a random variable.
- Now let's reveal the rolls. The sample mean is $\bar{x} = 3.457$.
- Point of confusion: $\bar{x}$ is a number (3.457 in fact). It sure doesn't look random! Why not?
- After the sample is drawn (in this case, the dice are rolled) the sample mean becomes data (a single observation of a random variable).
- It is easy to forget, when confronted with a single number, that that number may be an outcome of a random experiment, which if repeated, may result in a different number.

Key Take-Away

- The sample mean is random. A different sample will result in a different sample mean.
- When we look at data, there is only one sample in front of us.
- The sample mean, before the experiment, is a random variable. After the experiment, it's a single number.

Normal example

Suppose that $X_1, X_2, ..., X_5$ are *i.i.d.* samples from $N(16, 5^2)$, what's the distribution of $\bar{X} = \frac{X_1+X_2+...+X_5}{5}$?

http://onlinestatbook.com/stat_sim/sampling_dist/index.html

Sampling distribution

- Sampling distribution – probability distribution that describes how a statistic such as the mean varies from sample to sample.
- Since the distribution is too hard to find, we appeal to the variance and the standard deviation of a random variable.

The square root law

- Let's consider an important statistic, sample mean: $\bar{X} = \frac{X_1 + \ldots + X_n}{n}$, where $X_1, \ldots, X_n$ are $i.i.d.$ with mean μ and variance σ^2.

The square root law

- Let's consider an important statistic, sample mean: $\bar{X} = \frac{X_1 + \ldots + X_n}{n}$, where $X_1, ..., X_n$ are $i.i.d.$ with mean μ and variance σ^2.
- If you have an $i.i.d.$ sample, then the expected value of $\bar{X}$ is μ and the variance is given by the formula:

$$Var(\bar{X}) = \frac{Var(X_1)}{n} = \frac{\sigma^2}{n}.$$

Thus

$$sd(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

The square root law

- Let's consider an important statistic, sample mean: $\bar{X} = \frac{X_1+...+X_n}{n}$, where $X_1, ..., X_n$ are $i.i.d.$ with mean μ and variance σ^2.
- If you have an $i.i.d.$ sample, then the expected value of $\bar{X}$ is μ and the variance is given by the formula:

$$Var(\bar{X}) = \frac{Var(X_1)}{n} = \frac{\sigma^2}{n}.$$

Thus

$$sd(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

- This is called The Square Root Law.

The accuracy of a sample mean

- Consider a special example: $X_1, ..., X_n \ i.i.d. \sim N(\mu, \sigma^2)$. For the sample mean $\bar{X}$, we have

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}).$$

- In this case, we know the sampling distribution exactly.
- The larger the sample size n, the less spread out the sampling distribution, and it's more accurate to estimate μ by the sample mean $\bar{X}$.

The Sampling Distribution of $\bar{X}$

- In general, the sampling distribution is difficult to find exactly. If X_i's are not normal, we only know that $\mathbb{E}(\bar{X}) = \mu$, and $Var(\bar{X}) = \frac{\sigma^2}{n}$.

The Sampling Distribution of $\bar{X}$

- In general, the sampling distribution is difficult to find exactly. If X_i's are not normal, we only know that $\mathbb{E}(\bar{X}) = \mu$, and $Var(\bar{X}) = \frac{\sigma^2}{n}$.
- However, it turns out that the sampling distribution for $\bar{X}$ is close to normal for n large enough.

The Sampling Distribution of $\bar{X}$

- In general, the sampling distribution is difficult to find exactly. If X_i's are not normal, we only know that $\mathbb{E}(\bar{X}) = \mu$, and $Var(\bar{X}) = \frac{\sigma^2}{n}$.
- However, it turns out that the sampling distribution for $\bar{X}$ is close to normal for n large enough.
- Play with this applet to convince yourself that the distribution of $\bar{X}$ is approximately normal no matter what the distribution of X_i's is: `http://onlinestatbook.com/stat_sim/sampling_dist/index.html`

Central Limit Theorem

- Astonishing Fact: For $X_1, ..., X_n$ $i.i.d.$ from any population with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$
  1. has mean $\mathbb{E}(\bar{X}) = \mu$
  2. has standard deviation $sd(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
  3. is exactly normal when $X_1, ...X_n$ $i.i.d. \sim N(\mu, \sigma^2)$
  4. is approximately normal when n is large

  Remark: the last bullet is known as the Central Limit Theorem (CLT). For practical purposes, normality can be assumed when $n > 30$.

Central Limit Theorem

- Astonishing Fact: For $X_1, ..., X_n$ $i.i.d.$ from any population with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$

  1. has mean $\mathbb{E}(\bar{X}) = \mu$
  2. has standard deviation $sd(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
  3. is exactly normal when $X_1, ...X_n$ $i.i.d. \sim N(\mu, \sigma^2)$
  4. is approximately normal when n is large

  Remark: the last bullet is known as the Central Limit Theorem (CLT). For practical purposes, normality can be assumed when $n > 30$.

- The Astonishing Fact says that the sampling distribution of the sample mean $\bar{X}$ is always approximately $N(\mu, \frac{\sigma^2}{n})$.

Empirical rule

Empirical rule: if $X \sim N(\mu, \sigma^2)$, then

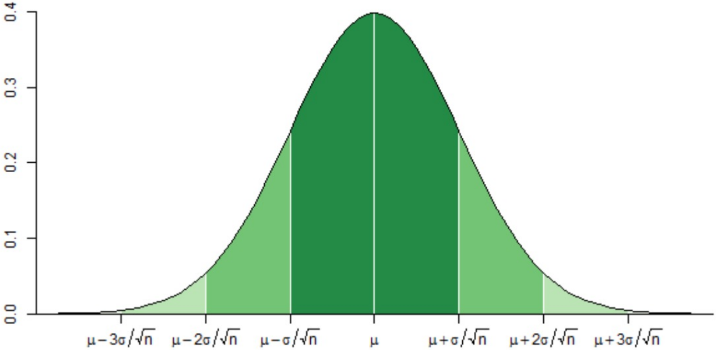
1. 68% of the data lies within 1 σ of the mean
2. 95% of the data lies within 2 σ of the mean
3. Essentially all (99.7%) of the data lies within 3 σ of the mean

The Standard Deviation of $\bar{X}$

Since $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$, the astonishing fact enables us to make general statements like

1. $\mathbb{P}(\mu - \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + \frac{\sigma}{\sqrt{n}}) \approx 0.68$
2. $\mathbb{P}(\mu - 2\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95$
3. $\mathbb{P}(\mu - 3\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 3\frac{\sigma}{\sqrt{n}}) \approx 0.997$

The Sampling distribution of $\bar{X}$



Central Limit Theorem

- The information in *i.i.d.* samples allows us to

Central Limit Theorem

- The information in $i.i.d.$ samples allows us to
  - estimate population parameters like μ by the corresponding statistics based on the sample, $\bar{X}$.

Central Limit Theorem

- The information in $i.i.d.$ samples allows us to
  - estimate population parameters like μ by the corresponding statistics based on the sample, $\bar{X}$.
  - estimate how close the sample statistics are likely to come to the corresponding population parameters. The key ingredient is the standard deviation of the statistic.

Central Limit Theorem

- The information in $i.i.d.$ samples allows us to
  - estimate population parameters like μ by the corresponding statistics based on the sample, $\bar{X}$.
  - estimate how close the sample statistics are likely to come to the corresponding population parameters. The key ingredient is the standard deviation of the statistic.
- For the sample mean, the standard deviation is $\frac{\sigma}{\sqrt{n}}$. If we further estimate σ from the data, then we will know how accurate is the estimate $\bar{X}$.

Central Limit Theorem

- The information in $i.i.d.$ samples allows us to
  - estimate population parameters like μ by the corresponding statistics based on the sample, $\bar{X}$.
  - estimate how close the sample statistics are likely to come to the corresponding population parameters. The key ingredient is the standard deviation of the statistic.
- For the sample mean, the standard deviation is $\frac{\sigma}{\sqrt{n}}$. If we further estimate σ from the data, then we will know how accurate is the estimate $\bar{X}$.
- The sampling distribution of a sample mean is close to Normal, a remarkable fact.

Example

A small freight ship has space for 50 containers and a weight limit of 55 tons. The container weight is $i.i.d.$ normally distributed with a mean weight of 1 ton and a standard deviation of 0.4 ton. Approximately, what is the probability of exceeding the weight limit?

Answer

- $X_1, ..., X_{50}$ $i.i.d. \sim N(1, 0.4^2)$.

Example

A small freight ship has space for 50 containers and a weight limit of 55 tons. The container weight is $i.i.d.$ normally distributed with a mean weight of 1 ton and a standard deviation of 0.4 ton. Approximately, what is the probability of exceeding the weight limit?

Answer

- $X_1, ..., X_{50}$ $i.i.d. \sim N(1, 0.4^2)$.
- By CLT, $\bar{X} \sim N(1, \frac{0.4^2}{50})$.

Example

A small freight ship has space for 50 containers and a weight limit of 55 tons. The container weight is $i.i.d.$ normally distributed with a mean weight of 1 ton and a standard deviation of 0.4 ton. Approximately, what is the probability of exceeding the weight limit?

Answer

- $X_1, ..., X_{50}$ $i.i.d. \sim N(1, 0.4^2)$.
- By CLT, $\bar{X} \sim N(1, \frac{0.4^2}{50})$.
- $\mathbb{P}(X_1 + ... + X_{50} > 55) = \mathbb{P}(\frac{X_1 + ... + X_{50}}{50} > \frac{55}{50}) = \mathbb{P}(\bar{X} > \frac{55}{50})$

Example

A small freight ship has space for 50 containers and a weight limit of 55 tons. The container weight is *i.i.d.* normally distributed with a mean weight of 1 ton and a standard deviation of 0.4 ton. Approximately, what is the probability of exceeding the weight limit?

Answer

- $X_1, ..., X_{50}$ *i.i.d.* $\sim N(1, 0.4^2)$.
- By CLT, $\bar{X} \sim N(1, \frac{0.4^2}{50})$.
- $\mathbb{P}(X_1 + ... + X_{50} > 55) = \mathbb{P}(\frac{X_1 + ... + X_{50}}{50} > \frac{55}{50}) = \mathbb{P}(\bar{X} > \frac{55}{50})$
- $\mathbb{P}(\bar{X} > \frac{55}{50}) = \mathbb{P}(Z > \frac{\frac{55}{50} - 1}{\sqrt{\frac{0.4^2}{50}}}) = \mathbb{P}(Z > 1.77) = \mathbb{P}(Z < -1.77) \approx \mathbb{P}(Z \leq -1.8) = 3.5\%$

Example

In a lottery game in a Casino one can win 0, 1 or 2 dollars with equal probabilities. A ticket to participate in the game costs 1.1 dollars. In one evening, 100 people participate in the game. What is the probability that the Casino will lose? Make sure to carefully define the appropriate random variables.

Answer

- Define X as the winning of one person. We then have $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \frac{1}{3}$.

Example

In a lottery game in a Casino one can win 0, 1 or 2 dollars with equal probabilities. A ticket to participate in the game costs 1.1 dollars. In one evening, 100 people participate in the game. What is the probability that the Casino will lose? Make sure to carefully define the appropriate random variables.

Answer

- Define X as the winning of one person. We then have $\mathbb{P}(X=0)=\mathbb{P}(X=1)=\mathbb{P}(X=2)=\frac{1}{3}$.
- We have $\mathbb{E}(X)=1$, $Var(X)=1^2\times\frac{1}{3}+2^2\times\frac{1}{3}-1^2=\frac{2}{3}$.

Example

In a lottery game in a Casino one can win 0, 1 or 2 dollars with equal probabilities. A ticket to participate in the game costs 1.1 dollars. In one evening, 100 people participate in the game. What is the probability that the Casino will lose? Make sure to carefully define the appropriate random variables.

Answer

- Define X as the winning of one person. We then have $\mathbb{P}(X=0)=\mathbb{P}(X=1)=\mathbb{P}(X=2)=\frac{1}{3}$.
- We have $\mathbb{E}(X)=1$, $Var(X)=1^2\times\frac{1}{3}+2^2\times\frac{1}{3}-1^2=\frac{2}{3}$.
- By CLT, $\bar{X}=\frac{X_1+\ldots+X_{100}}{100}\sim N(1,\frac{\frac{2}{3}}{100})$.

Example

In a lottery game in a Casino one can win 0, 1 or 2 dollars with equal probabilities. A ticket to participate in the game costs 1.1 dollars. In one evening, 100 people participate in the game. What is the probability that the Casino will lose? Make sure to carefully define the appropriate random variables.

Answer

- Define X as the winning of one person. We then have $\mathbb{P}(X=0)=\mathbb{P}(X=1)=\mathbb{P}(X=2)=\frac{1}{3}$.
- We have $\mathbb{E}(X)=1$, $Var(X)=1^2\times\frac{1}{3}+2^2\times\frac{1}{3}-1^2=\frac{2}{3}$.
- By CLT, $\bar{X}=\frac{X_1+\ldots+X_{100}}{100}\sim N(1,\frac{\frac{2}{3}}{100})$.
- $\mathbb{P}(X_1+\ldots+X_{100}>1.1\times 100)=\mathbb{P}(\frac{X_1+\ldots+X_{100}}{100}>1.1)=\mathbb{P}(\bar{X}>1.1)$

Example

In a lottery game in a Casino one can win 0, 1 or 2 dollars with equal probabilities. A ticket to participate in the game costs 1.1 dollars. In one evening, 100 people participate in the game. What is the probability that the Casino will lose? Make sure to carefully define the appropriate random variables.

Answer

- Define X as the winning of one person. We then have $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \frac{1}{3}$.
- We have $\mathbb{E}(X) = 1$, $Var(X) = 1^2 \times \frac{1}{3} + 2^2 \times \frac{1}{3} - 1^2 = \frac{2}{3}$.
- By CLT, $\bar{X} = \frac{X_1 + \ldots + X_{100}}{100} \sim N(1, \frac{\frac{2}{3}}{100})$.
- $\mathbb{P}(X_1 + \ldots + X_{100} > 1.1 \times 100) = \mathbb{P}(\frac{X_1 + \ldots + X_{100}}{100} > 1.1) = \mathbb{P}(\bar{X} > 1.1)$
- $\mathbb{P}(\bar{X} > 1.1) = \mathbb{P}(Z > \frac{1.1-1}{\sqrt{\frac{2}{300}}}) = \mathbb{P}(Z > 1.22) \approx \mathbb{P}(Z \leq -1.2) = 0.1151$.

Sampling distribution of the sample variance

- Recall that if $y_1, ..., y_n$ constitute a random sample from a population with variance σ^2, then the estimate of σ^2 is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Sampling distribution of the sample variance

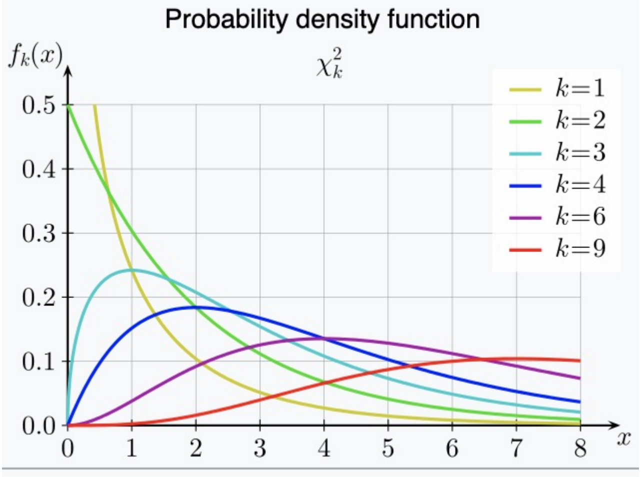
Theorem

If $y_1, ..., y_n$ are i.i.d. samples from $N(\mu, \sigma^2)$, then

1. $\bar{y}$ *and* s^2 *are independent*
2. $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

Note: If $Z_1, ..., Z_m \overset{i.i.d.}{\sim} N(0,1)$, then $Z_1^2 + ... + Z_m^2 \sim \chi^2_m$

Chi squared distribution



Example

Example

At a factory, the standard deviation of a part's thickness is supposed to be 0.02m. A quality control inspector takes a random samples of size 20 to make sure that the process maintains low variability. If the thickness is normally distributed, what is the probability the inspector would observe a sample standard deviation greater than 0.025m, if the process is in control?

Answer

- Denote the sample variance by s^2, we know

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1},$$

where $n = 10$, and $\sigma = 0.02$

Answer

- Denote the sample variance by s^2, we know

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1},$$

  where $n = 10$, and $\sigma = 0.02$

- The problem asks for $\mathbb{P}(s > 0.025)$

Answer

- Denote the sample variance by s^2, we know

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1},$$

  where $n = 10$, and $\sigma = 0.02$

- The problem asks for $\mathbb{P}(s > 0.025)$
- $\mathbb{P}(s > 0.025) = \mathbb{P}(s^2 > 0.025^2) = \mathbb{P}(\frac{9 \times s^2}{0.02^2} > \frac{9 \times 0.025^2}{0.02^2}) = \mathbb{P}(\chi^2_9 > 14.06) = 0.1202$

Answer

- Denote the sample variance by s^2, we know

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1},$$

  where $n = 10$, and $\sigma = 0.02$

- The problem asks for $\mathbb{P}(s > 0.025)$
- $\mathbb{P}(s > 0.025) = \mathbb{P}(s^2 > 0.025^2) = \mathbb{P}(\frac{9 \times s^2}{0.02^2} > \frac{9 \times 0.025^2}{0.02^2}) = \mathbb{P}(\chi^2_9 > 14.06) = 0.1202$
- R command: 1-pchisq(14.06,9)

Z-transform

- Now let us learn t-distribution

Z-transform

- Now let us learn *t-distribution*
- We know that if $y_1, \ldots, y_n$ constitute a random sample from a $N(\mu, \sigma^2)$ population, then the Z-transform gives

$$\frac{\bar{y} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

Z-transform

- Now let us learn t-distribution
- We know that if $y_1, ..., y_n$ constitute a random sample from a $N(\mu, \sigma^2)$ population, then the Z-transform gives

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

- In practice, we usually don't know σ, so we use s instead

Z-transform

- Now let us learn *t-distribution*
- We know that if $y_1, ..., y_n$ constitute a random sample from a $N(\mu, \sigma^2)$ population, then the Z-transform gives

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

- In practice, we usually don't know σ, so we use s instead
- $\frac{\bar{y}-\mu}{s/\sqrt{n}}$ is no longer $N(0, 1)$

Z-transform

- Now let us learn *t-distribution*
- We know that if $y_1, ..., y_n$ constitute a random sample from a $N(\mu, \sigma^2)$ population, then the Z-transform gives

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

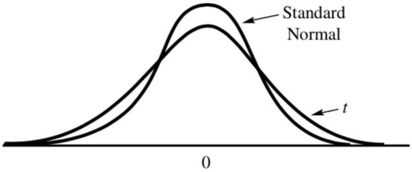
- In practice, we usually don't know σ, so we use s instead
- $\frac{\bar{y}-\mu}{s/\sqrt{n}}$ is no longer $N(0, 1)$
- $$\frac{\bar{y} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

t-distribution

Definition

If Z and W are independent, where $Z \sim N(0,1)$ and $W \sim \chi^2_m$, then $T = \frac{Z}{\sqrt{W/m}}$ has a t-distribution with m degrees of freedom, denoted by t_m.

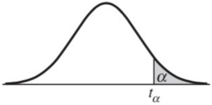
FIGURE 7.3
A comparison of the standard normal and t density functions.



t-table

$$\mathbb{P}(T > t_\alpha) = \alpha$$

Percentage Points of the t Distributions



$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
3.078	6.314	12.706	31.821	63.657	1
1.886	2.920	4.303	6.965	9.925	2
1.638	2.353	3.182	4.541	5.841	3
1.533	2.132	2.776	3.747	4.604	4
1.476	2.015	2.571	3.365	4.032	5
1.440	1.943	2.447	3.143	3.707	6
1.415	1.895	2.365	2.998	3.499	7
1.397	1.860	2.306	2.896	3.355	8
1.383	1.833	2.262	2.821	3.250	9

t-table

R codes:

- pt(y0,m) gives $\mathbb{P}(Y \leq y0)$, for $Y \sim t_m$
- qt(p,m) yields the pth quantile, the value of ϕ_p such that $\mathbb{P}(Y \leq \phi_p) = p$

Example

Example

The tensile strength for a type of wire is normally distributed with unknown mean μ and unknown variance σ^2. Six pieces of wire were randomly selected from a large roll; Y_i , the tensile strength for portion i, is measured for $i = 1, 2, ..., 6$. The population mean μ and variance σ^2 can be estimated by $\bar{Y}$ and s^2, respectively. Because $\sigma^2_{\bar{Y}} = \sigma^2/n$, it follows that σ^2 can be estimated by s^2/n. Find the approximate probability that $\bar{Y}$ will be within $2s/\sqrt{n}$ of the true population mean μ.

Example

Answer

- $\mathbb{P}(\mu - 2s/\sqrt{n} \leq \bar{Y} \leq \mu + 2s/\sqrt{n})$

Example

Answer

- $\mathbb{P}(\mu - 2s/\sqrt{n} \leq \bar{Y} \leq \mu + 2s/\sqrt{n})$
- $\mathbb{P}(-2 \leq \frac{\bar{Y}-\mu}{s/\sqrt{n}} \leq 2)$

Example

Answer

- $\mathbb{P}(\mu - 2s/\sqrt{n} \leq \bar{Y} \leq \mu + 2s/\sqrt{n})$
- $\mathbb{P}(-2 \leq \frac{\bar{Y}-\mu}{s/\sqrt{n}} \leq 2)$
- $\frac{\bar{Y}-\mu}{s/\sqrt{n}} \sim t_{6-1}$

Example

Answer

- $\mathbb{P}(\mu - 2s/\sqrt{n} \leq \bar{Y} \leq \mu + 2s/\sqrt{n})$
- $\mathbb{P}(-2 \leq \frac{\bar{Y}-\mu}{s/\sqrt{n}} \leq 2)$
- $\frac{\bar{Y}-\mu}{s/\sqrt{n}} \sim t_{6-1}$
- $\mathbb{P}(-2 \leq t_5 \leq 2) = 0.898$

Example

Answer

- $\mathbb{P}(\mu - 2s/\sqrt{n} \leq \bar{Y} \leq \mu + 2s/\sqrt{n})$
- $\mathbb{P}(-2 \leq \frac{\bar{Y}-\mu}{s/\sqrt{n}} \leq 2)$
- $\frac{\bar{Y}-\mu}{s/\sqrt{n}} \sim t_{6-1}$
- $\mathbb{P}(-2 \leq t_5 \leq 2) = 0.898$
- R command: 2*pt(2,5)-1

Summary

- The relationship between statistics and probability theory
  - Data and random variables
  - Sample and population
  - Estimates and parameters

Summary (very important)

Sampling distributions

- Central Limit Theorem (CLT): $\bar{X}$
  - $X_1, ..., X_n$ are *i.i.d.* random variables with mean μ, variance σ^2.
  - $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$.
- Assuming normality: $X_1, ..., X_n$ are *i.i.d.* normal $N(\mu, \sigma^2)$
  - $\bar{X}$: $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
  - s^2: $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$
  - $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

Homework

Install R, and run the Experiment 1 in "lecture_1_code.R":

- record the number.
- Change the code to run the simulation for tossing a 20-sided 2000 times. Repeat this experiment for 5 times, record the sample average and standard deviation. (Codes should also be submitted)
- Use CLT to derive the distribution of the sample average

Homework

Install R, and try functions 1-pchisq(10,3), 2*pt(5,4)-1; record the values, and explain their meanings.

7.9 (a,b,c), 7.19, 7.37, 7.42