AUA CS108, Statistics, Fall 2020

Lecture 01

Michael Poghosyan

26 Aug 2020

Welcome to the AUA Statistics Course

Welcome to the AUA Statistics Course

And Happy New ~~Year~~ Semester ! ◡̈

Contents

- Syllabus highlights
- Intro to the Course
- Intro to the Descriptive Statistics

Syllabus Highlights

- Course name: **CS 108, Statistics**

Syllabus Highlights

- Course name: **CS 108, Statistics**
- No. of Credits: **3**

Syllabus Highlights

- Course name: **CS 108, Statistics**
- No. of Credits: **3**
- Instructor: **MP**

Syllabus Highlights

- Course name: **CS 108, Statistics**
- No. of Credits: **3**
- Instructor: **MP**
- Instructor's Office: **#336W, PAB**

Syllabus Highlights

- Course name: **CS 108, Statistics**
- No. of Credits: **3**
- Instructor: **MP**
- Instructor's Office: **#336W, PAB**
- Instructor's OH: By an Appointment

Syllabus Highlights

- Course name: **CS 108, Statistics**
- No. of Credits: **3**
- Instructor: **MP**
- Instructor's Office: **#336W, PAB**
- Instructor's OH: By an Appointment
- Teaching Associates: **Lilit Avetisyan, Arusyak Mikayelyan**

Syllabus Highlights

- Course name: **CS 108, Statistics**
- No. of Credits: **3**
- Instructor: **MP**
- Instructor's Office: **#336W, PAB**
- Instructor's OH: By an Appointment
- Teaching Associates: **Lilit Avetisyan, Arusyak Mikayelyan**
- PSS day/time: **TBD**

Syllabus Highlights

- Course name: **CS 108, Statistics**
- No. of Credits: **3**
- Instructor: **MP**
- Instructor's Office: **#336W, PAB**
- Instructor's OH: By an Appointment
- Teaching Associates: **Lilit Avetisyan, Arusyak Mikayelyan**
- PSS day/time: **TBD**
- TA's OH: **TBD**

Course Materials

- Moodle Page: Aha, we will use it ⌣̈

Course Materials

- Moodle Page: Aha, we will use it ⌣̈
- Moodle Enrollment Key: **Stat-F20**

Course Materials

- Moodle Page: Aha, we will use it ⌣̈
- Moodle Enrollment Key: **Stat-F20**
- Discussions and Info Platform: **Slack**

Course Materials

- Moodle Page: Aha, we will use it ⌣̈
- Moodle Enrollment Key: **Stat-F20**
- Discussions and Info Platform: **Slack**
- Syllabus: uploaded to our Moodle page

Course Materials

- Moodle Page: Aha, we will use it ⌣̈
- Moodle Enrollment Key: **Stat-F20**
- Discussions and Info Platform: **Slack**
- Syllabus: uploaded to our Moodle page
- Textbooks: uploaded to our Moodle page

Course Materials

- Moodle Page: Aha, we will use it ⌣̈
- Moodle Enrollment Key: **Stat-F20**
- Discussions and Info Platform: **Slack**
- Syllabus: uploaded to our Moodle page
- Textbooks: uploaded to our Moodle page
- Software: **R** and **R Studio** (freeware)

Course Materials

- Moodle Page: Aha, we will use it ⌣̈
- Moodle Enrollment Key: **Stat-F20**
- Discussions and Info Platform: **Slack**
- Syllabus: uploaded to our Moodle page
- Textbooks: uploaded to our Moodle page
- Software: **R** and **R Studio** (freeware)
- R Textbooks: uploaded to our Moodle Page

Course Materials

- Moodle Page: Aha, we will use it ⌣̈
- Moodle Enrollment Key: **Stat-F20**
- Discussions and Info Platform: **Slack**
- Syllabus: uploaded to our Moodle page
- Textbooks: uploaded to our Moodle page
- Software: **R** and **R Studio** (freeware)
- R Textbooks: uploaded to our Moodle Page
- Other: I have prepared some R Intro Slides

Course Materials

- Moodle Page: Aha, we will use it ⌣̈
- Moodle Enrollment Key: **Stat-F20**
- Discussions and Info Platform: **Slack**
- Syllabus: uploaded to our Moodle page
- Textbooks: uploaded to our Moodle page
- Software: **R** and **R Studio** (freeware)
- R Textbooks: uploaded to our Moodle Page
- Other: I have prepared some R Intro Slides
- Question: Do we need some supplementary R Labs? YESSS/No

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,
  - half Moodle-based (say, MCh Problems)

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,
  - half Moodle-based (say, MCh Problems)
  - half R problems $\ddot{\smile}$?

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,
  - half Moodle-based (say, MCh Problems)
  - half R problems $\ddot{\smile}$?
- Quizzes: Yeah, we will have them!

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,
  - half Moodle-based (say, MCh Problems)
  - half R problems $\ddot\smile$?
- Quizzes: Yeah, we will have them! Almost weekly.

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,
  - half Moodle-based (say, MCh Problems)
  - half R problems $\ddot{\smile}$?
- Quizzes: Yeah, we will have them! Almost weekly.Online.

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,
  - half Moodle-based (say, MCh Problems)
  - half R problems $\ddot{\smile}$?
- Quizzes: Yeah, we will have them! Almost weekly.Online.
- Final Grade Formula:

$$Total = 0.25 \cdot (MT + Q) + 0.2 \cdot HW + 0.3 \cdot F$$

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,
  - half Moodle-based (say, MCh Problems)
  - half R problems $\ddot{\smile}$?
- Quizzes: Yeah, we will have them! Almost weekly.Online.
- Final Grade Formula:

$$Total = 0.25 \cdot (MT + Q) + 0.2 \cdot HW + 0.3 \cdot F$$

- No Makeups for Quizzes! Sorry!

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,
  - half Moodle-based (say, MCh Problems)
  - half R problems $\ddot\smile$?
- Quizzes: Yeah, we will have them! Almost weekly.Online.
- Final Grade Formula:

$$Total = 0.25 \cdot (MT + Q) + 0.2 \cdot HW + 0.3 \cdot F$$

- No Makeups for Quizzes! Sorry! But one of the Quiz grades will be dropped.

Syllabus Highlights, Cont'd

- Exams: **1 Midterm** and a **Final Exam**
- Homework: **(almost) weekly**, due on Fridays (???)

  Will it be OK to have our HW:

  - half paper-based,
  - half Moodle-based (say, MCh Problems)
  - half R problems $\ddot\smile$?
- Quizzes: Yeah, we will have them! Almost weekly.Online.
- Final Grade Formula:

$$Total = 0.25 \cdot (MT + Q) + 0.2 \cdot HW + 0.3 \cdot F$$

- No Makeups for Quizzes! Sorry! But one of the Quiz grades will be dropped.
- No late HWs (except some veeery special cases)

Syllabus Highlights, Cont'd

- No Grades Curving. ☹

Syllabus Highlights, Cont'd

- No Grades Curving. ☹
- Advice: Always ask your questions during lectures, ask your questions during OHs, solve HWs by yourself!

Syllabus Highlights, Cont'd

- No Grades Curving. ☹
- Advice: Always ask your questions during lectures, ask your questions during OHs, solve HWs by yourself!
- Advice: Run over the Probability Topics, especially, about RVs and Distributions

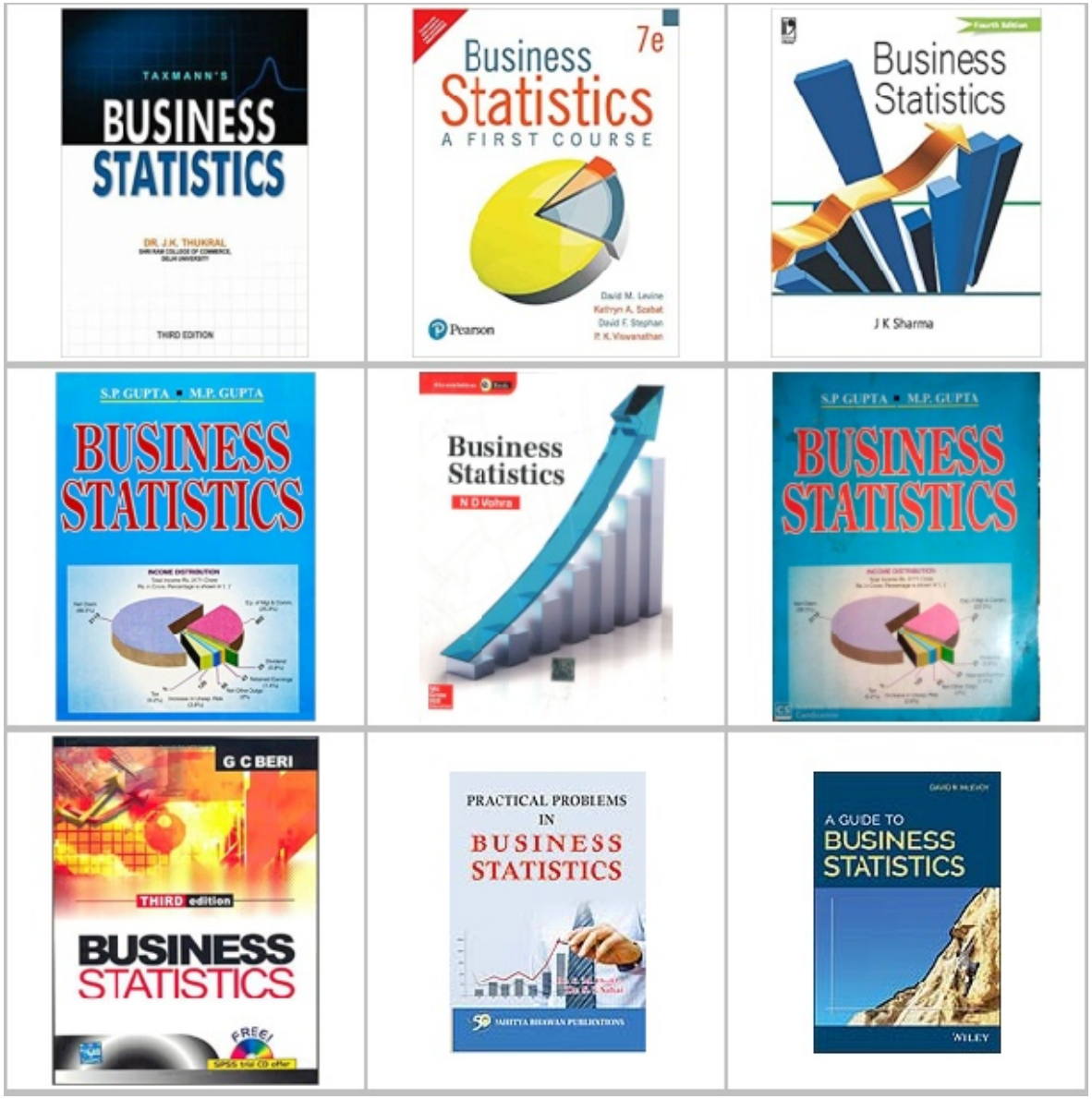
QA Session

Questions?

Where we use Statistics

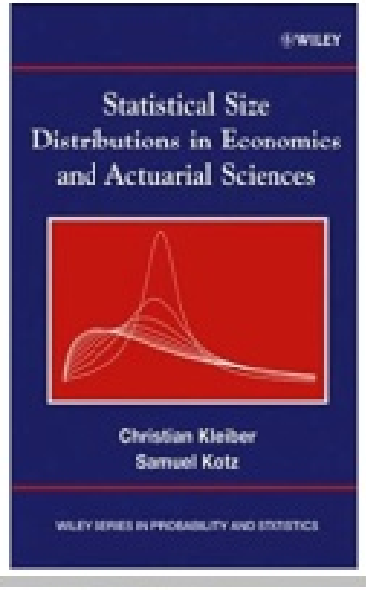
Where we use Statistics

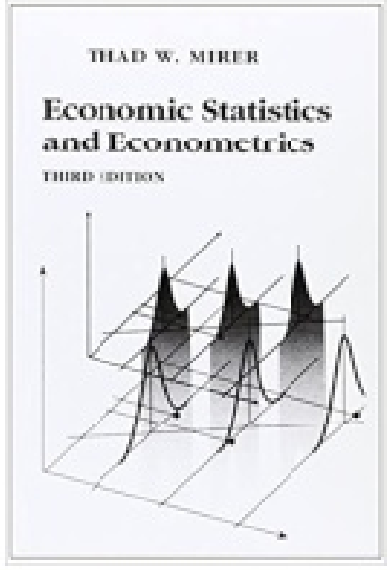
Business

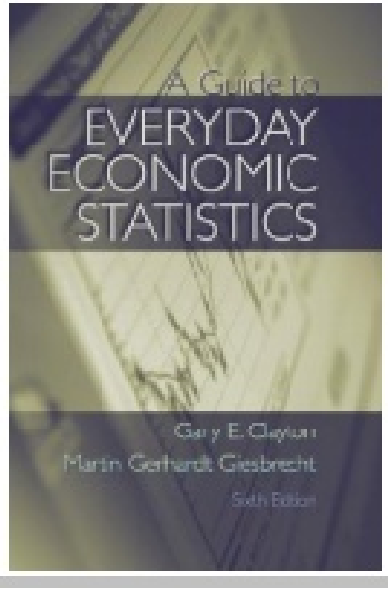


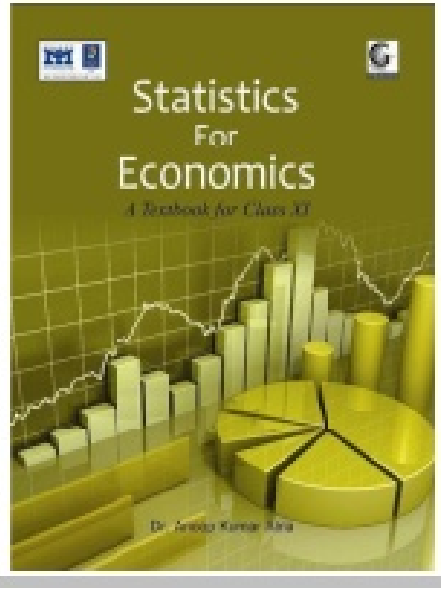
Where we use Statistics

Economics









Where we use Statistics

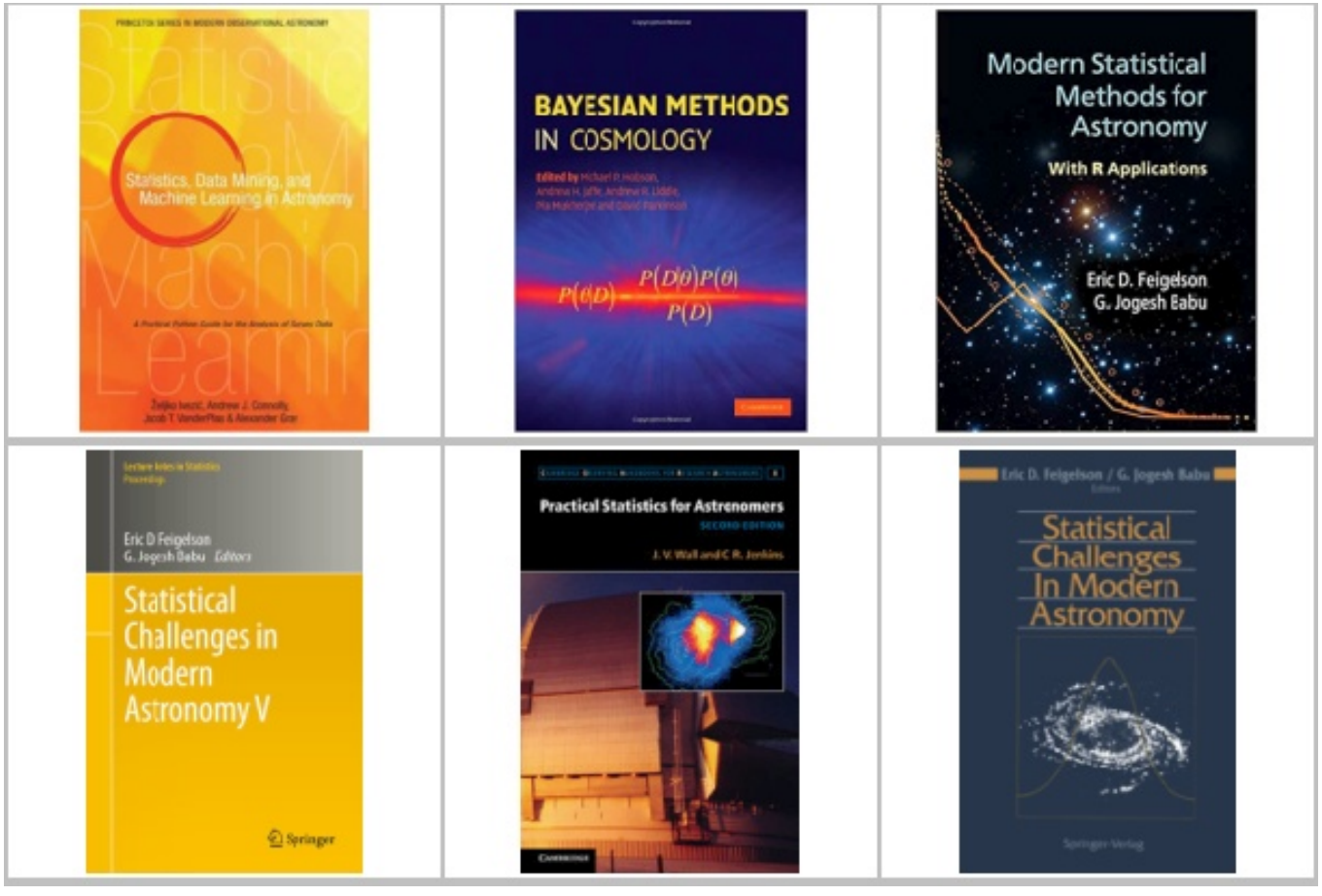
Agriculture

Where we use Statistics

Finance

Where we use Statistics

Astronomy



Where we use Statistics

Biology

Where we use Statistics

BioStatistics

Where we use Statistics

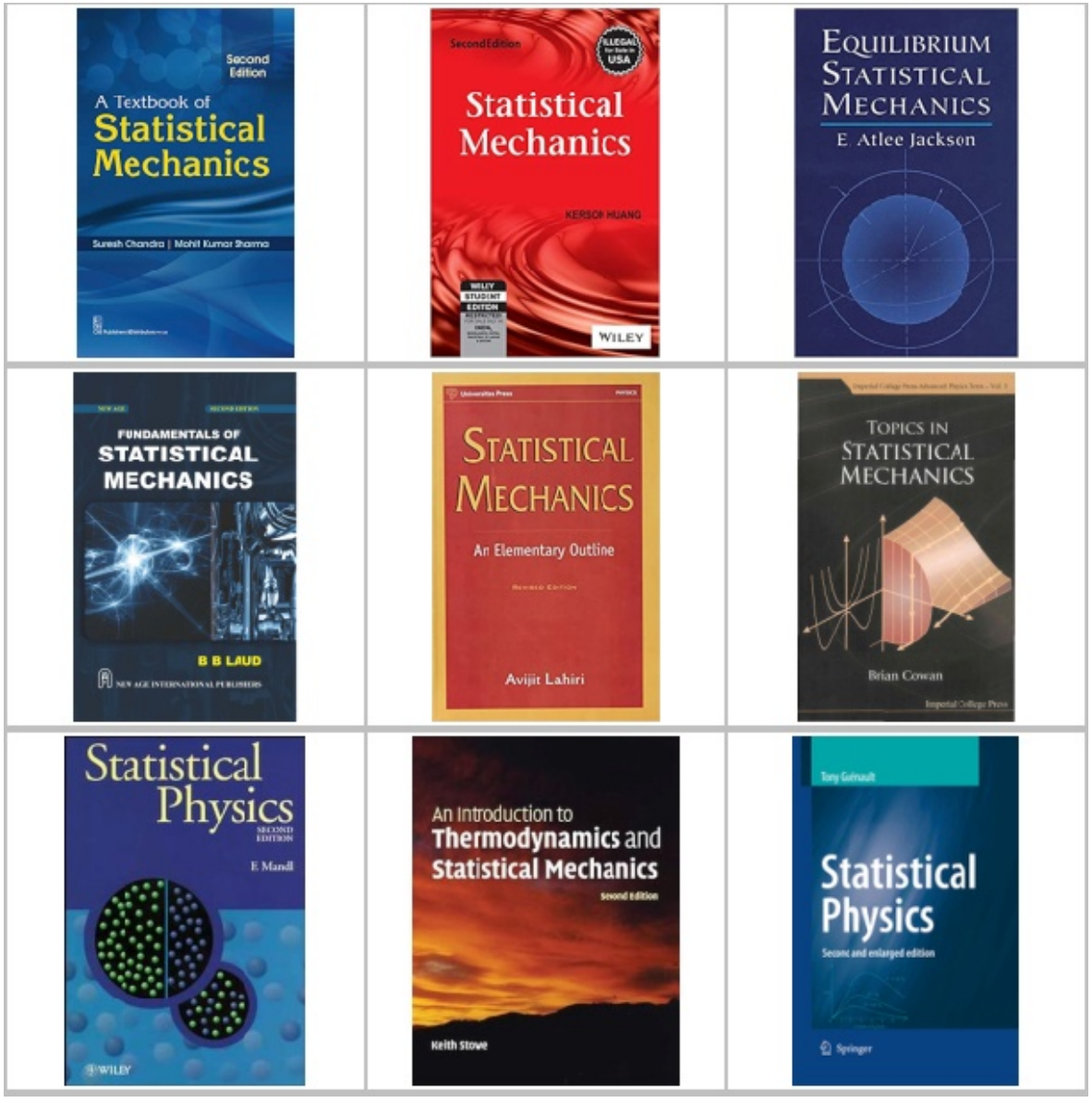
Psychology

Where we use Statistics

Medicine

Where we use Statistics

Physics and Mechanics



Where we use Statistics

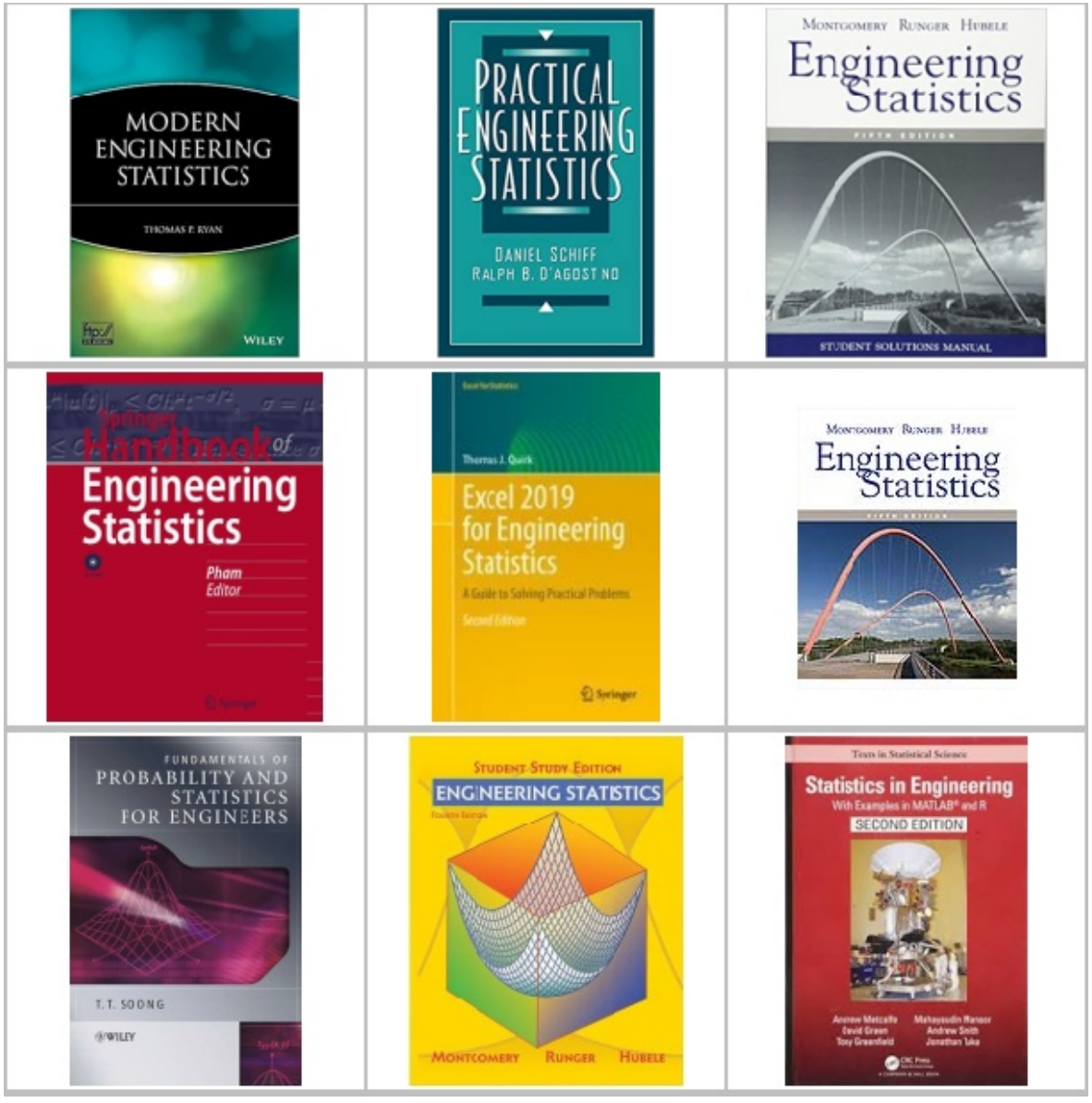
Marketing

Where we use Statistics

Language Study

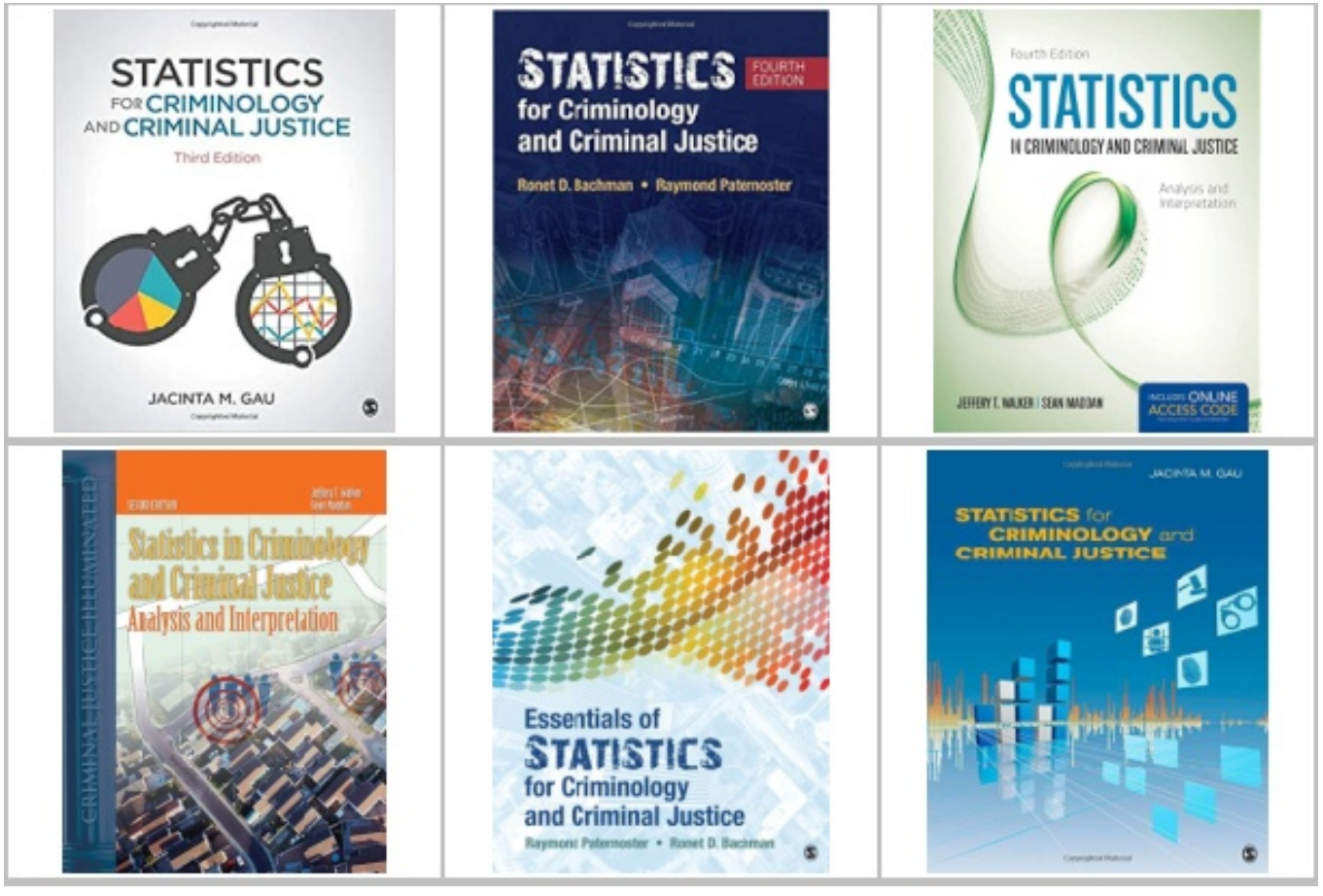
Where we use Statistics

Engineering



Where we use Statistics

Criminology

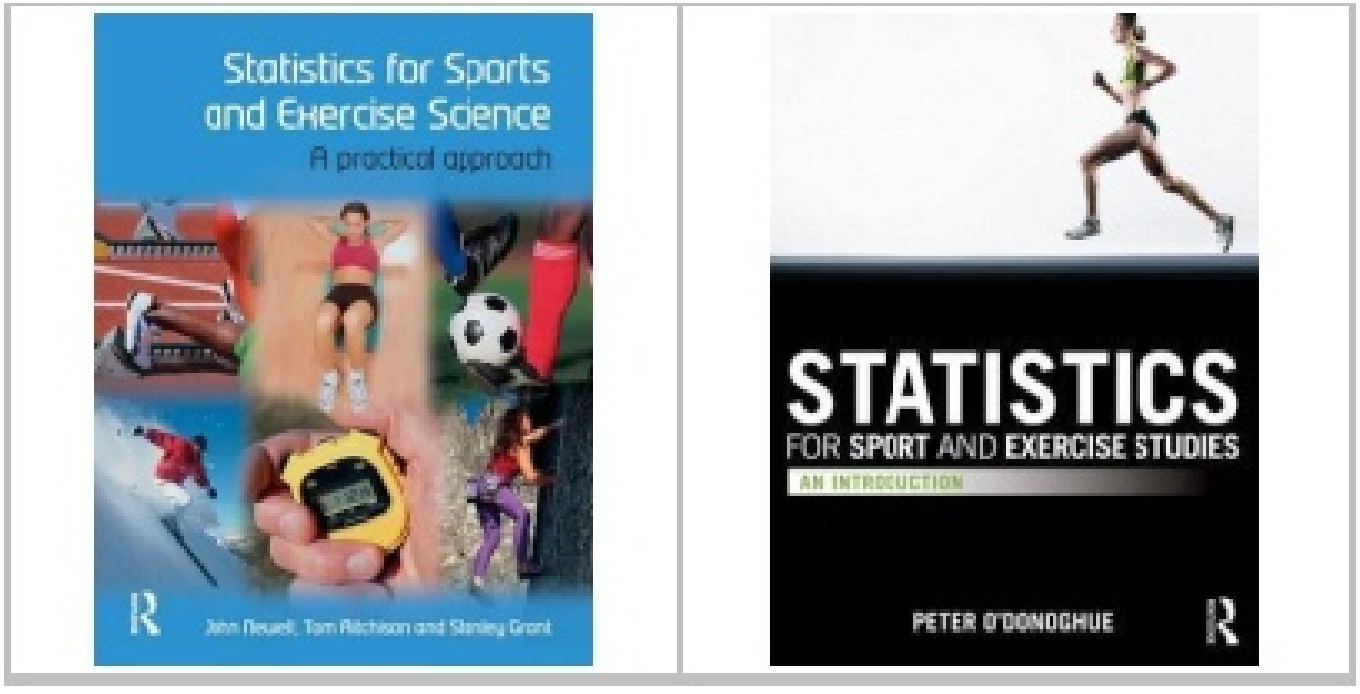


Where we use Statistics

Quality Control

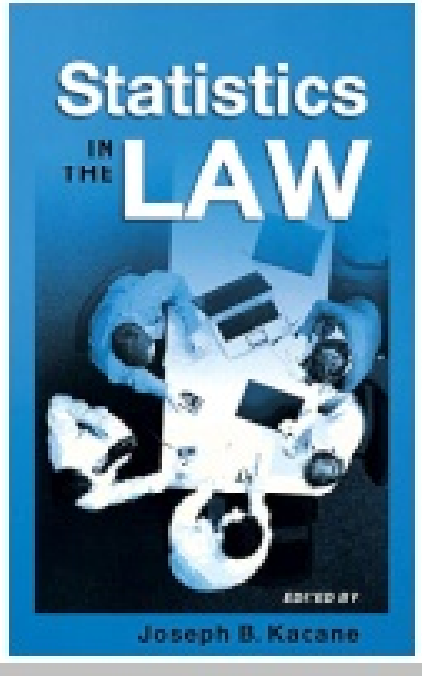
Where we use Statistics

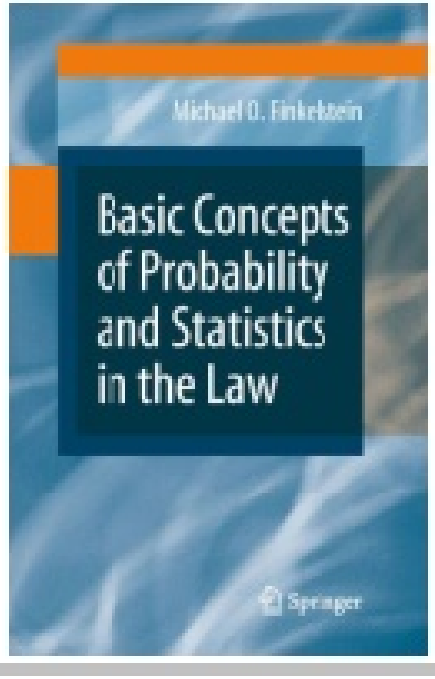
Sport



Where we use Statistics

Law





Where we use Statistics

Food

Where we use Statistics

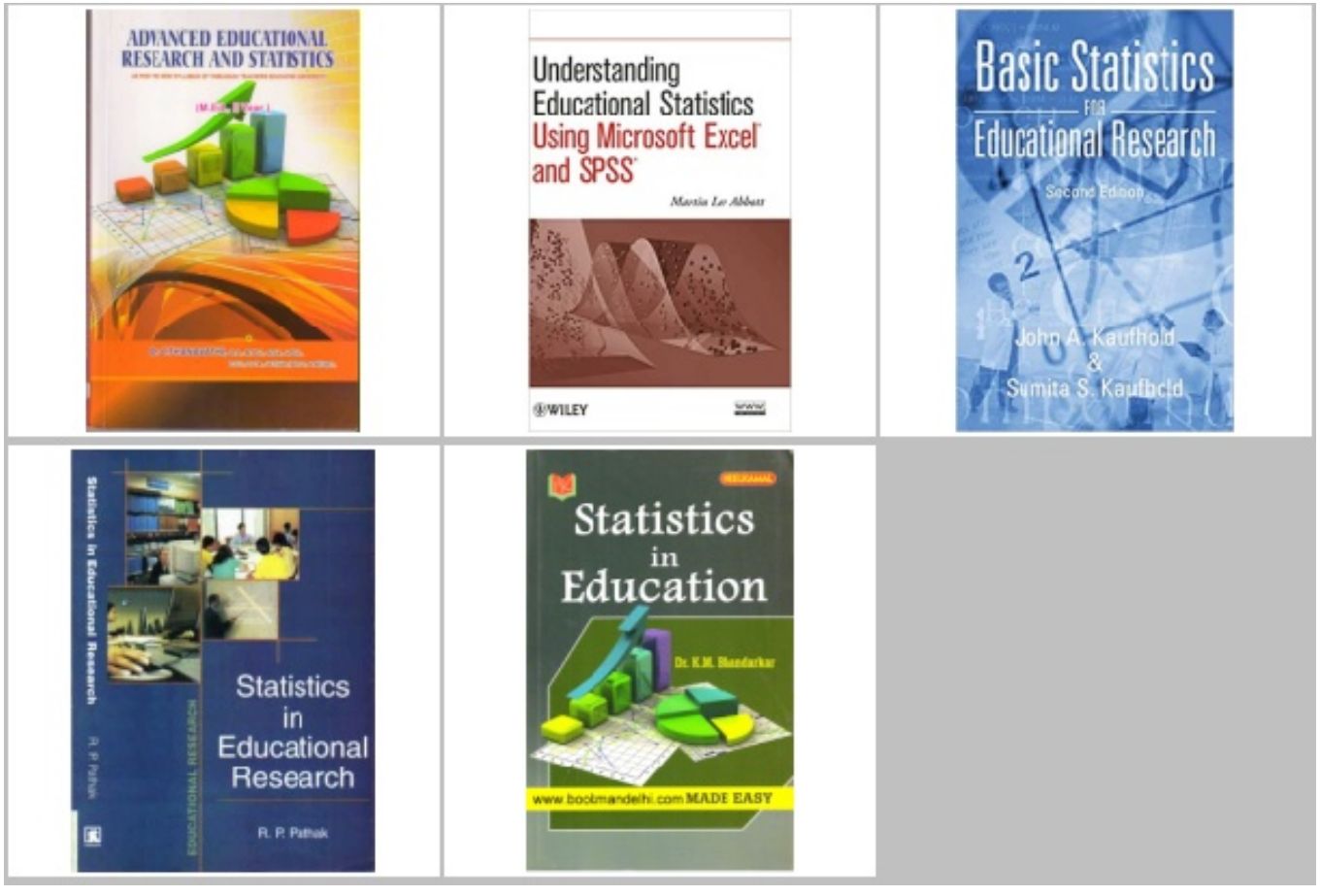
Genetics

Where we use Statistics

Chemistry

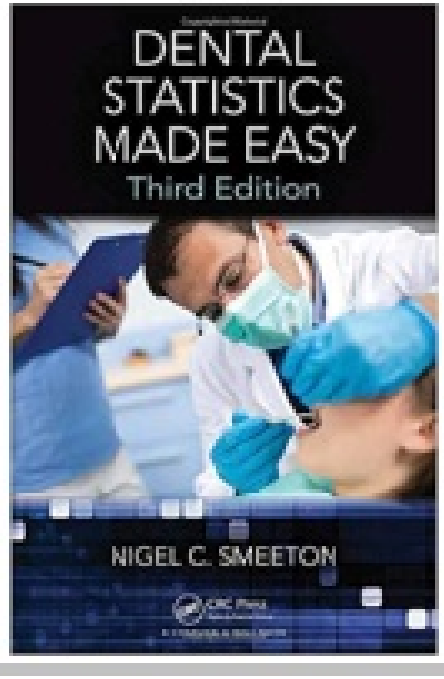
Where we use Statistics

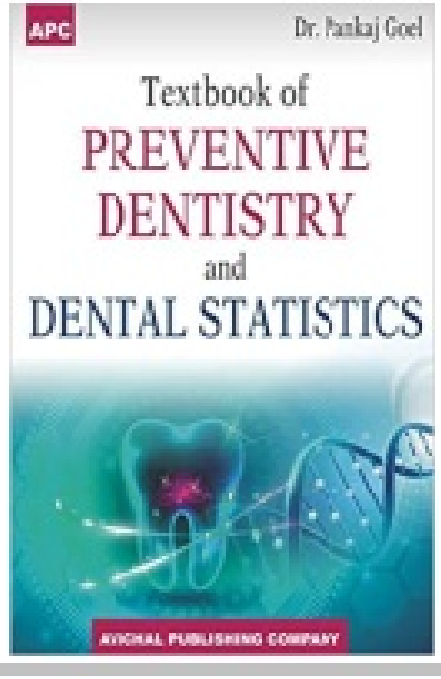
Education



Where we use Statistics

Dentistry





Where we use Statistics

Other. . .

Where we use Statistics

ML, Statistical Learning

Where we use Statistics

. . . and so on . . .

About Statistics

- What is Statistics?

About Statistics

- What is Statistics?

Statistics is an Art and Science of Learning from Data.

About Statistics

- What is Statistics?

Statistics is an Art and Science of Learning from Data. In a little bit detailed form,

Statistics is the Art and Science of ***Collecting, Describing and Analyzing Data****, getting insight from Data.*

About Statistics

- What is the difference between Statistics and Probability?

About Statistics

- What is the difference between Statistics and Probability?

In some sense, Statistics and Probability are inverse to each other:

About Statistics

- What is the difference between Statistics and Probability?

In some sense, Statistics and Probability are inverse to each other:

- in Probability Theory, we assume the Reality, the Truth, the Generating Process is given, and we get information about possible outcomes, possible observations.

About Statistics

- What is the difference between Statistics and Probability?

In some sense, Statistics and Probability are inverse to each other:

- in Probability Theory, we assume the Reality, the Truth, the Generating Process is given, and we get information about possible outcomes, possible observations. Say,

We assume that the daily number of customers who are ordering Pepperoni Pizza at a particular pizzeria is following the $Pois(15.2)$ Distribution (*the Reality, the Generating Process*), and we are interested in the Probability that the daily number will exceed 17 (*Probability of a possible outcome*)

About Statistics

- in Statistics, we have Observations, we have Data, Outcomes, we want to learn about the Reality, about the Truth, about the Generating Process (from which that Data is obtained).

About Statistics

- in Statistics, we have Observations, we have Data, Outcomes, we want to learn about the Reality, about the Truth, about the Generating Process (from which that Data is obtained). Say,

We have Data about daily number of customers ordering Pepperoni Pizza at that pizzeria for some days:

$$10,\ 15,\ 14,\ 20,\ 7,\ 12,\ 10,$$

and we want to find out the Distribution of the r.v.

$X =$ the daily number of customers ordering PP at that pizzeria.

About Statistics

It is easy to see that this problem is more challenging than its Probabilistic counterpart, its “inverse” one -

About Statistics

It is easy to see that this problem is more challenging than its Probabilistic counterpart, its “inverse” one - because we want to learn about the Reality, the Truth from a finite number of observations.

About Statistics

It is easy to see that this problem is more challenging than its Probabilistic counterpart, its "inverse" one - because we want to learn about the Reality, the Truth from a finite number of observations. And, moreover, if the Probabilistic counterpart has a unique correct answer (we just need to calculate $\mathbb{P}(X > 17)$, where X is the daily number of PP-hungry customers, $X \sim Pois(15.2)$),

About Statistics

It is easy to see that this problem is more challenging than its Probabilistic counterpart, its "inverse" one - because we want to learn about the Reality, the Truth from a finite number of observations. And, moreover, if the Probabilistic counterpart has a unique correct answer (we just need to calculate $\mathbb{P}(X > 17)$, where X is the daily number of PP-hungry customers, $X \sim Pois(15.2)$), its Statistical counterpart doesn't have a unique correct answer!

About Statistics

It is easy to see that this problem is more challenging than its Probabilistic counterpart, its "inverse" one - because we want to learn about the Reality, the Truth from a finite number of observations. And, moreover, if the Probabilistic counterpart has a unique correct answer (we just need to calculate $\mathbb{P}(X > 17)$, where X is the daily number of PP-hungry customers, $X \sim Pois(15.2)$), its Statistical counterpart doesn't have a unique correct answer! We cannot find out the Exact Distribution, we can just guess it, we can only **Estimate**.

About Statistics

- What is the difference between Statistical and Mathematical thinking?

About Statistics

- What is the difference between Statistical and Mathematical thinking?

In Mathematics, we prove facts (using other, already proven facts, and using some small number of Axioms).

About Statistics

- What is the difference between Statistical and Mathematical thinking?

In Mathematics, we prove facts (using other, already proven facts, and using some small number of Axioms). Say, our lovely MVT is correct in all cases!

About Statistics

- What is the difference between Statistical and Mathematical thinking?

In Mathematics, we prove facts (using other, already proven facts, and using some small number of Axioms). Say, our lovely MVT is correct in all cases! The real roots of $x^2 = 4$ are

About Statistics

- What is the difference between Statistical and Mathematical thinking?

In Mathematics, we prove facts (using other, already proven facts, and using some small number of Axioms). Say, our lovely MVT is correct in all cases! The real roots of $x^2 = 4$ are $x = \pm 2$.

About Statistics

- What is the difference between Statistical and Mathematical thinking?

In Mathematics, we prove facts (using other, already proven facts, and using some small number of Axioms). Say, our lovely MVT is correct in all cases! The real roots of $x^2 = 4$ are $x = \pm 2$.

In Statistics (Inferential Statistics), we use data to get an insight about the unknown process behind the generation of the data. Usually, our data is finite, and this is not giving a chance to get a complete and 100 percent information about that unknown process.

About Statistics

- What is the difference between Statistical and Mathematical thinking?

In Mathematics, we prove facts (using other, already proven facts, and using some small number of Axioms). Say, our lovely MVT is correct in all cases! The real roots of $x^2 = 4$ are $x = \pm 2$.

In Statistics (Inferential Statistics), we use data to get an insight about the unknown process behind the generation of the data. Usually, our data is finite, and this is not giving a chance to get a complete and 100 percent information about that unknown process. So, in Statsitsics, we are doing some *assumptions*, and from that "inferring", "guessing", ***estimating*** the unknown process, its parameters, and sometimes give the level of our confidence.

About Statistics

- What is the difference between Statistical and Mathematical thinking?

In Mathematics, we prove facts (using other, already proven facts, and using some small number of Axioms). Say, our lovely MVT is correct in all cases! The real roots of $x^2 = 4$ are $x = \pm 2$.

In Statistics (Inferential Statistics), we use data to get an insight about the unknown process behind the generation of the data. Usually, our data is finite, and this is not giving a chance to get a complete and 100 percent information about that unknown process. So, in Statsitsics, we are doing some *assumptions*, and from that "inferring", "guessing", ***estimating*** the unknown process, its parameters, and sometimes give the level of our confidence. Say, if I am tossing a coin 10 times and get 8 Hs and 2 Ts, is it a sign that the coin is not fair?

About Statistics

- What is the difference between Data Scienece and Statistics?

About Statistics

- What is the difference between Data Scienece and Statistics?

Well, maybe we will talk a little bit about this at the end of the course.

- Why I need to learn Statistics?

About Statistics

- What is the difference between Data Scienece and Statistics?

Well, maybe we will talk a little bit about this at the end of the course.

- Why I need to learn Statistics?

Simple - to pass this course ¨‿

About Statistics

- What is the difference between Data Scienece and Statistics?

Well, maybe we will talk a little bit about this at the end of the course.

- Why I need to learn Statistics?

Simple - to pass this course ⌣̈ Or, it is very important if you want to become a Statistician (one of the fastest growing and Best Jobs in USA), learn DS/ML, become a Biostatistician, learn Econometrics, make Investments and play in Financial Markets, and many-many more, see, for example, the Wiki page.

About Statistics

- What is the difference between Data Scienece and Statistics?

Well, maybe we will talk a little bit about this at the end of the course.

- Why I need to learn Statistics?

Simple - to pass this course ⌣̈ Or, it is very important if you want to become a Statistician (one of the fastest growing and Best Jobs in USA), learn DS/ML, become a Biostatistician, learn Econometrics, make Investments and play in Financial Markets, and many-many more, see, for example, the Wiki page.

And finally, to understand the everyday usage of Statistical language, graphs and estimates, say, about polls and salaries ⌣̈

Some Problems that Statistics can consider

- Estimate the proportion of adults in Armenia that are computer-literate;

Some Problems that Statistics can consider

- Estimate the proportion of adults in Armenia that are computer-literate;
- Estimate the number/proportion of defective production in a manufacturing plant;

Some Problems that Statistics can consider

- Estimate the proportion of adults in Armenia that are computer-literate;
- Estimate the number/proportion of defective production in a manufacturing plant;
- Show that there is a strong correlation between education and income;

Some Problems that Statistics can consider

- Estimate the proportion of adults in Armenia that are computer-literate;
- Estimate the number/proportion of defective production in a manufacturing plant;
- Show that there is a strong correlation between education and income;
- Determine whether it is true that, in average, women earn less than men;

Some Problems that Statistics can consider

- Find the relationship between the High School Math average grade and the average of the University Calc 1 and Calc 2 courses;

Some Problems that Statistics can consider

- Find the relationship between the High School Math average grade and the average of the University Calc 1 and Calc 2 courses;

Note: Clearly, there is **no** exact formula relating the average of Calc1 and Calc2 grades to HS Math average grade.

Some Problems that Statistics can consider

- Find the relationship between the High School Math average grade and the average of the University Calc 1 and Calc 2 courses;

Note: Clearly, there is **no** exact formula relating the average of Calc1 and Calc2 grades to HS Math average grade. In statistics, we do not *find* exactly, but try to give a reasonable *Estimate*, even we assess how good is our Estimate;

Some Problems that Statistics can consider

- Find the relationship between the High School Math average grade and the average of the University Calc 1 and Calc 2 courses;

Note: Clearly, there is **no** exact formula relating the average of Calc1 and Calc2 grades to HS Math average grade. In statistics, we do not *find* exactly, but try to give a reasonable *Estimate*, even we assess how good is our Estimate;

- Calculate the percentage of body fat based on only easily calculated body measurements (say, biceps and abdomen measurements);

Some Problems that Statistics can consider

- Find the relationship between the High School Math average grade and the average of the University Calc 1 and Calc 2 courses;

Note: Clearly, there is **no** exact formula relating the average of Calc1 and Calc2 grades to HS Math average grade. In statistics, we do not *find* exactly, but try to give a reasonable *Estimate*, even we assess how good is our Estimate;

- Calculate the percentage of body fat based on only easily calculated body measurements (say, biceps and abdomen measurements);
- Predict apartment prices based on various (given) characteristics of that apartment;

Some Problems that Statistics can consider

- Find the relationship between the High School Math average grade and the average of the University Calc 1 and Calc 2 courses;

Note: Clearly, there is **no** exact formula relating the average of Calc1 and Calc2 grades to HS Math average grade. In statistics, we do not *find* exactly, but try to give a reasonable *Estimate*, even we assess how good is our Estimate;

- Calculate the percentage of body fat based on only easily calculated body measurements (say, biceps and abdomen measurements);
- Predict apartment prices based on various (given) characteristics of that apartment;
- Check if the newly developed drug is more effective (in curing some disease) than the old one;

Some Problems that Statistics can consider

- Find the relationship between the High School Math average grade and the average of the University Calc 1 and Calc 2 courses;

Note: Clearly, there is **no** exact formula relating the average of Calc1 and Calc2 grades to HS Math average grade. In statistics, we do not *find* exactly, but try to give a reasonable *Estimate*, even we assess how good is our Estimate;

- Calculate the percentage of body fat based on only easily calculated body measurements (say, biceps and abdomen measurements);
- Predict apartment prices based on various (given) characteristics of that apartment;
- Check if the newly developed drug is more effective (in curing some disease) than the old one;
- Check if the drug is effective against Corona