EE325

Introductory Econometrics

Weerawat Phattarasukkumjorn

Semester 1/2021

Course details

› Schedule

Section 2: Tue, Thu 11.00 – 12.30
Moodle class code: **2589**

› Instructor

Weerawat Phattarasukkumjorn, Room no. 437
weerawat@econ.tu.ac.th
Please communicate **in class group.**

› Evaluation

Homework and assignment	30 points
Midterm exam	30 points
Final exam	40 points

› Exam date and time

Midterm: Wed, Sep29 from 15.00 – 17.00 (2 hours)
Final: Mon, Dec 13 from 09.00 – 11.30 (2.5 hours)

What does ‘Econometrics’ mean?

We, first and foremost, try to understand the topic, analyzing the morphemes.

› Econometrics = economics + metrics

› Economics: you all know what this means.

› Metrics: a system of measurement.

To conclude, *“econometrics is the application of statistical methods to economic data in order to give empirical content to economic relationships. More precisely, it is the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference“.*

How can econometrics be used?

First of all, let's introduce how academics work to come up with a conclusion and policy recommendation. You are also to deal with these processes for your seminar.

› Introduction and problem statement (research question)

What question do you want to answer? How and why is it important to answer the question? What is expected to be your contribution to academic world?

› Literature review (research gap)

Have there been any other people trying to answer the same question yet? What and how other people have tried to answer this question? What are their results? What are their methods used and how they are still flawed or lacking?

› Theoretical framework (rationalize your hypothesis)

For economics, what is the rationale behind your speculation or hypothesis? Has there been anyone discover or derive any theory before? What type of explanation will you be using to couple and elaborate your results and implications.

How can econometrics be used?

› Research method (how to find the answer)

How you are going to prove your hypothesis? We have several ways to do, but mostly we can use these methods

- Qualitative methods: descriptive statistics, reviews, deduction, etc.
- Quantitative methods: forecast, quantitative simulation, econometrics, etc.

› Report results and policy implication

What are your results and findings? Moreover, what do those findings suggest? Do those results imply that we should implement which kind of policy? What is the limitation of your work and what can be possibly improved in the future?

Example of econometric process

› Research question

› Economic theory or model

› Empirical data

› Econometric modeling

› Estimation and hypothesis testing

› Forecast or prediction

› Empirical conclusion

› Policy recommendation

Strengths of econometrics

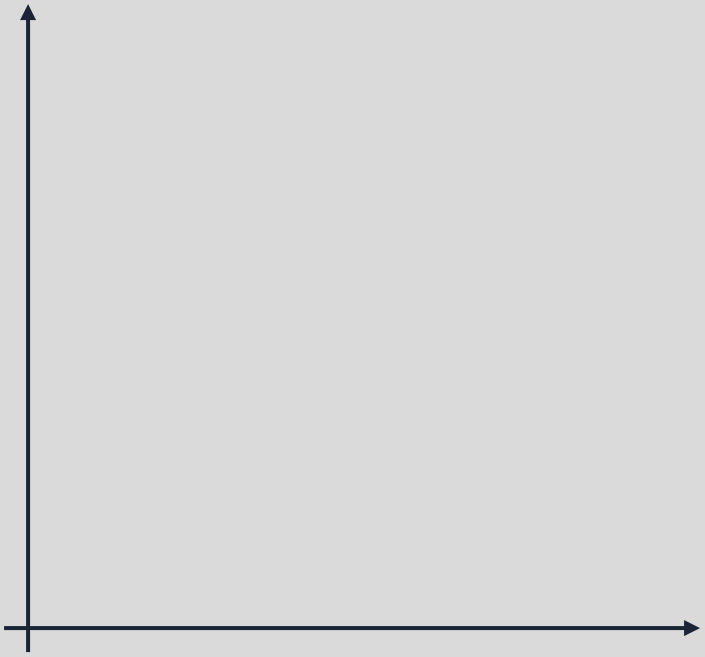
› Ability to quantify direction and magnitude of economic effects

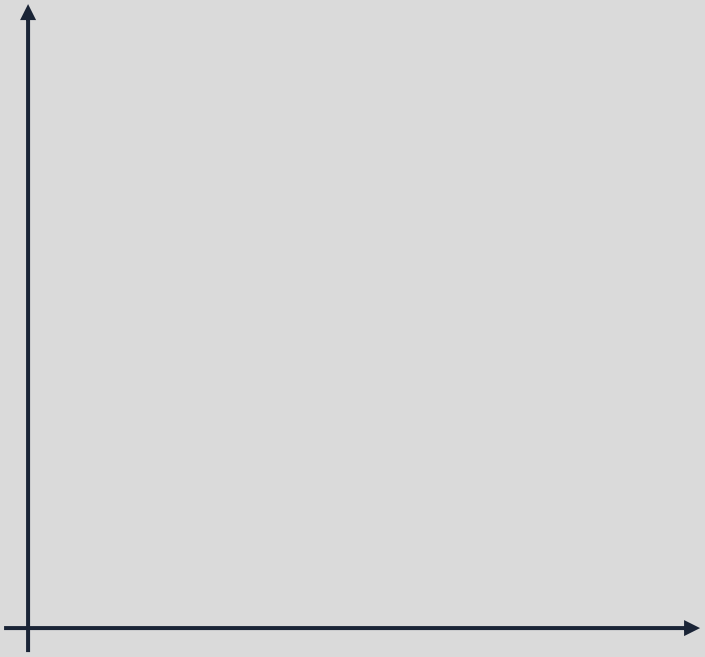
› Statistical testability

› Prediction or simulation (precise, though not always correct)

What are statistical relationships?

As a reminder, when we study mathematics, a function is usually determined. Meanwhile, studying statistics relationship is, most of the time, cannot be captured by a specific function.

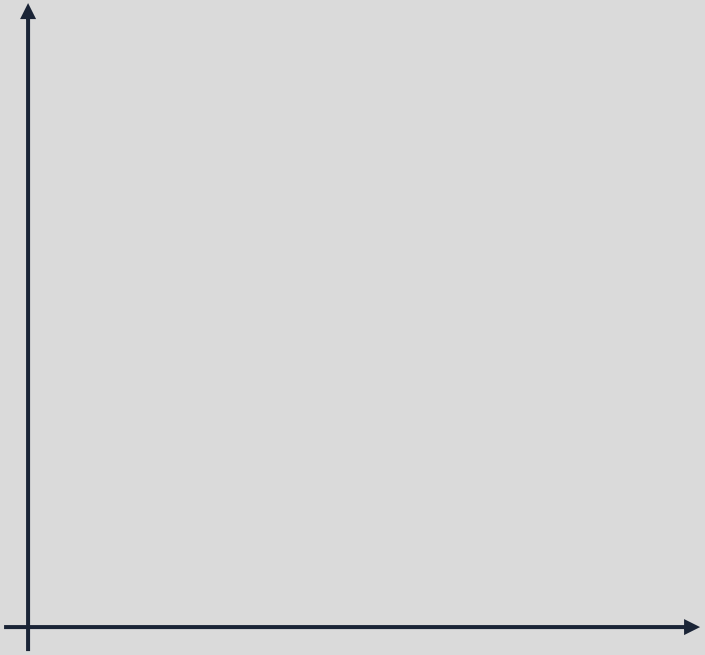


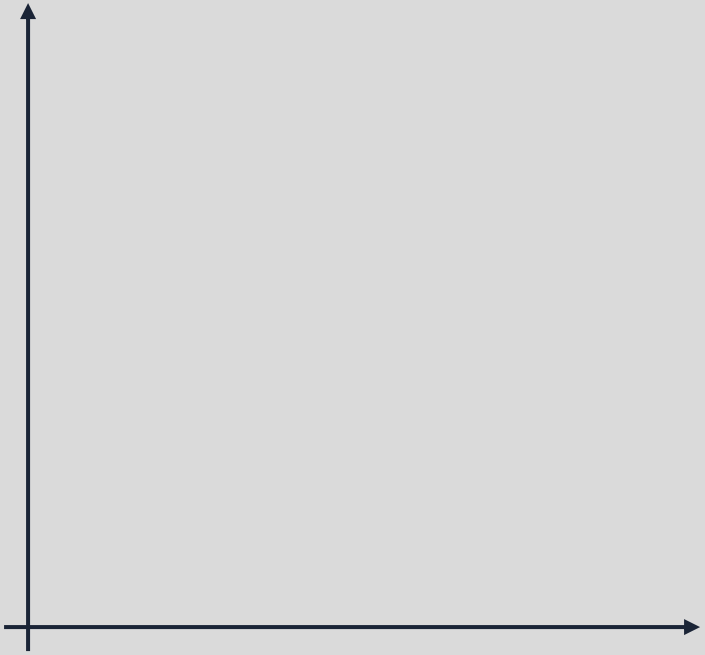


What are statistical relationships?

(1) Correlation relationship is any statistical association or the degree of association between two or more variables (pairwise, linear).

(2) Regression relationship is a relationship derived from a set of statistical processes for estimating the relationships between a dependent and one or more independent variables (linear or nonlinear).





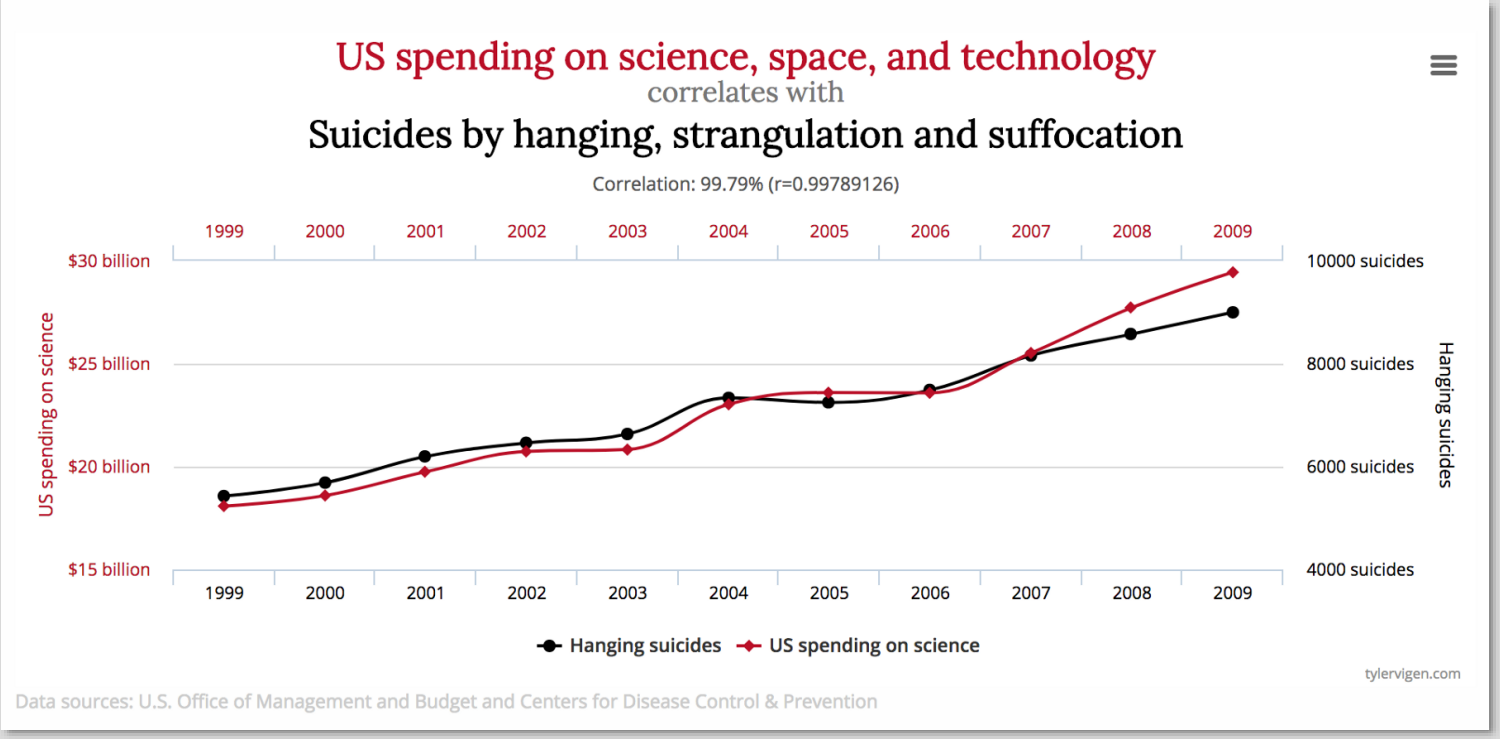
What are statistical relationships?

(3) Causal relationship is influence by which one event, process or state (a cause) contributes to the production of another event, process or state (an effect) where the cause is partly responsible for the effect, and the effect is partly dependent on the cause.

It is very difficult to draw a conclusion that two variables are causal, especially for social science. **Neither** correlation **nor** regression imply causality at all.

Most causality can be base upon economic theory and explanation. Aligning empirical results can also support for theory. Couple both can provide strong evidence, though does not reveal universal truth.

Spurious correlation



Data by collection

(1) Primary data

Collected by a researcher from first-hand sources, using methods like surveys, interviews, or experiments.

(2) Secondary data

Gathered from studies, surveys, or experiments that have been run by other people or for other researches.

Broad categorization

observations

student	GPA
student 1	GPA 1
student 2	GPA 2
student 3	GPA 3
student 4	GPA 4
student 5	GPA 5
student 6	GPA 6

(1) Cross-sectional data

A type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at the one point or a period of time. The analysis has no regard to differences in time.

observations

time	GPA
t=1	GPA 1 (t=1)
t=2	GPA 1 (t=2)
t=3	GPA 1 (t=3)
t=4	GPA 1 (t=4)
t=5	GPA 1 (t=5)
t=6	GPA 1 (t=6)

(2) Time series

A series of data points indexed in time order.

Broad categorization

(3) Panel data

A set of data collected over time and over the same individuals.

student	GPA	time
student 1	GPA 1	t=1
student 2	GPA 2	t=1
student 3	GPA 3	t=1
student 1	GPA 1	t=2
student 2	GPA 2	t=2
student 3	GPA 3	t=2
student 1	GPA 1	t=3
student 2	GPA 2	t=3
student 3	GPA 3	t=3

time	student 1	student 2	student 3	student 4
t=1				
t=1				
t=3				

Broad categorization

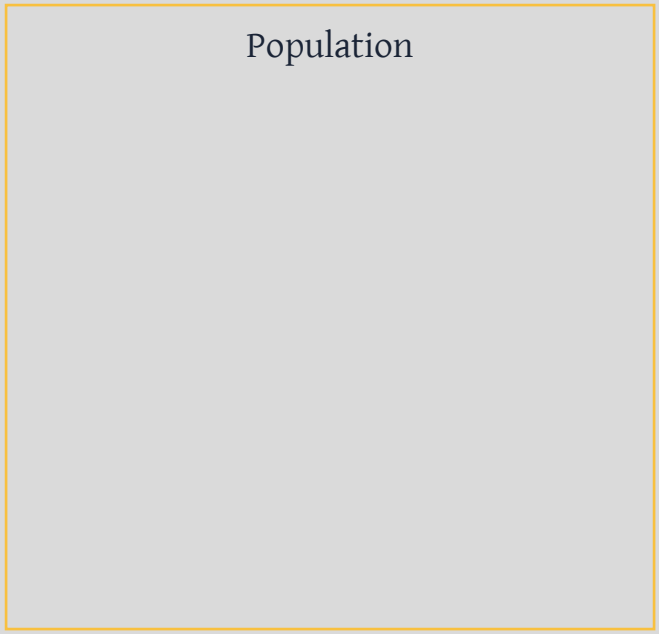
(4) Pooled cross-sectional data

A multiple cross-sectional data pooled without observing the same subject.

student	GPA	time
student 1	GPA 1	t=1
student 2	GPA 2	t=1
student 3	GPA 3	t=1
student 2	GPA 2	t=2
student 3	GPA 3	t=2
student 4	GPA 4	t=2
student 1	GPA 1	t=3
student 3	GPA 3	t=3
student 4	GPA 4	t=3

time	student 1	student 2	student 3	student 4
t=1				
t=1				
t=3				

Inferencing from sample



(1) Population

Refers to a group of observations of interest. If the data cover all the observations of interest, the set of data is called a **census**.

(2) Sample

Refers to a subset of population of interest. Most of the time they are statistically random samples which is called **survey**.

Secondary data in Thailand

(1) Cross-sectional data

Not free most of the time. University students can request from the National Statistics Office (NSO) for their project. Project proposal must be submitted.

› Household Socio-Economic Survey (SES) – income, expenditure, debt, asset (recently added), etc. Unit of analysis is household.

› Labor Force Survey (LFS) – wage, working hour, occupation, job search. Unit of analysis is individual.

› Health and Welfare Survey (HWS) – health insurance, health benefit, partial utilization, etc. Unit of analysis is individual.

› Office of the National Economic and Social Development Council (NESDC) – the NESDC takes NSO data to calculate important statistics. https://www.nesdc.go.th/more_news.php?cid=74

Secondary data in Thailand

(2) Time series data

Widely available because they are not identity specified.

› Bank of Thailand (BOT) – GDP, financial and monetary statistics, currency exchange, etc.
https://www.bot.or.th/Thai/Statistics/Pages/default.aspx

› Ministry of Finance (MOF) and Fiscal Policy Office – tax revenue and expenditure, tax disbursement, national debt, etc.
http://www.fpo.go.th/main/Statistic-Database.aspx

› Others include Ministry of Commerce (MOC), Stock Exchange of Thailand (SET), international organization such as International Monetary Fund (IMF), World Bank, United Nations (UN), OECD, etc.

Secondary data in Thailand

(3) Panel data

Very limited in Thailand. The ones that I know of are

› Panel SES (by the NSO) from 2005, 2006, 2007, 2010, 2012 – not free.

› Townsend Thai data (cooperated with UTCC and TRF) from 1997 to 2017 – request needed.
http://riped.utcc.ac.th/panel/data/townsend-thai-data

Content

Chapter number

Sub-chapter

How to use this handout?

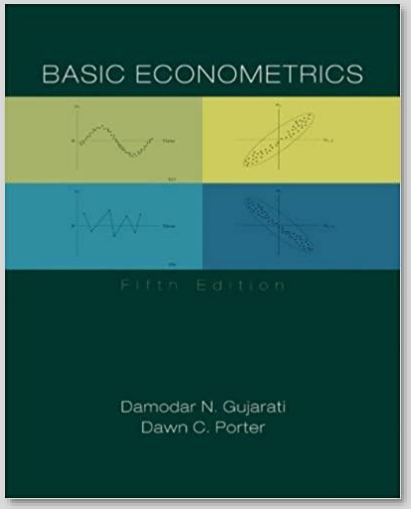
Main topic

Illustration or content

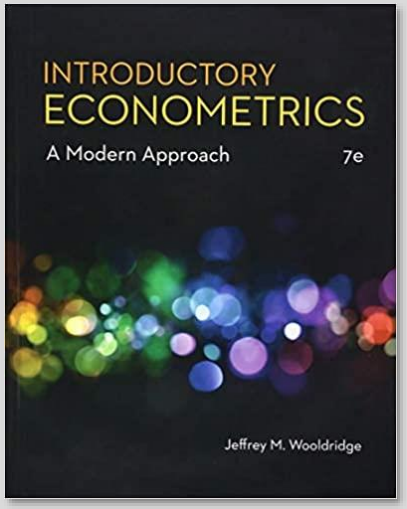
Illustration or content

Content

Main textbook



› Gujarati, D.N., and D.C. Porter, **Basic Econometrics**. 5th ed., N.Y., McGraw-Hill, 2009.



› Wooldridge, J. M. **Introductory Econometrics: A Modern Approach**. 7e ed. Thompson: South-Western, 2019.

Chapter 2

Statistics revision

Flow of study in this chapter

› Probability, random variable and density

Revision the basic concept of probability and its distribution, graphing random variable and its probability.

› Bivariate probability density

When two random variables, or two events, coexist, what aspects of the distribution can be studied.

› Central tendency and dispersion

How to measure, and why, central tendency and dispersion of a distribution for a random variable.

› Common distribution functions

The distributions we are relying on for statistical in this semester.

Further reading can be found in Gujarati and Porter (2009), Appendix A, page 801-837

(1) Event, Sample Space and Probability

Let A be an event of interest, occurring within a given sample space S and $P(A)$ be the probability that A will occur, $P(A)$ is defined as

› $P(A) = \frac{number\ of\ times\ event\ A\ will\ occur}{number\ of\ all\ possible\ outcome\ in\ sample\ space\ S}$

Example tossing 2 fair coins, the sample space is

› $S = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$

If the event of interest is having at least a coin turning head (H) is

› $A = \{\text{HH}, \text{HT}, \text{TH}\}$.

The probability of this event is then

› $P(A) =$

(1) Event, Sample Space and Probability

Probability Axioms

(1) $0 \leq P(A) \leq 1$

(2) $P(S) = 1$

(3) If and A are B mutually exclusive, then $P(A \cup B) = P(A) + P(B)$

(2) Random variable

Definition 2.1

Let X be a **random variable**, the results of an experiment in the form of value, which value is given by one of the results.

Example Tossing 2 fair coins again, let X be 0 if the result shows **at least** a coin turned up head, be 1 otherwise. The sample space was defined as

› $S = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$

Transforming theses events into random variable, we get

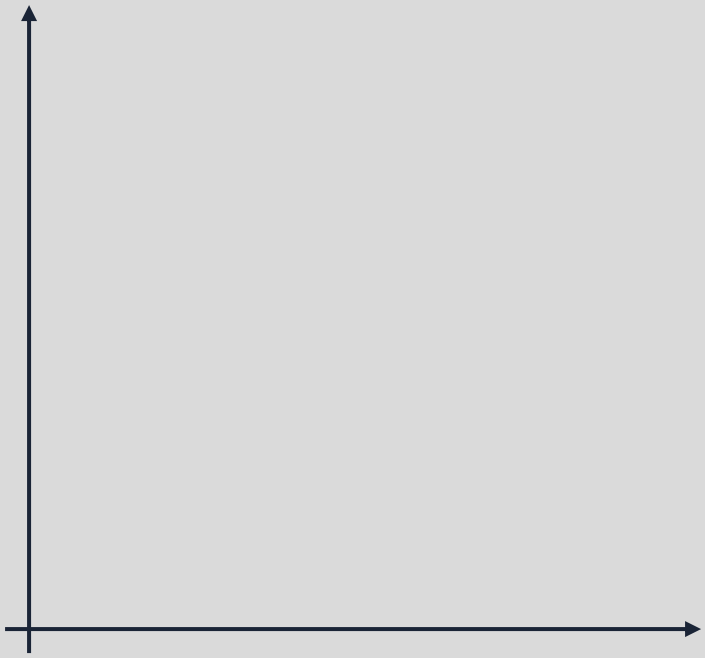
› $S_X = \{0,1\}$

Therefore, if we put probability function with specific value of random variable X, we have

› $P(X = 0) =$

› $P(X = 1) =$

(2) Random variable



Graphing the random variable and its probability here on the left.

› **Discreate random variable** is a random variable that can take specific values of event.

› **Continuous random variable** is a random variable that can takes infinite amounts of value of event.

(3) Probability Density Function (PDF)

Definition 2.2

A function whose value at any given sample (or point) in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample.

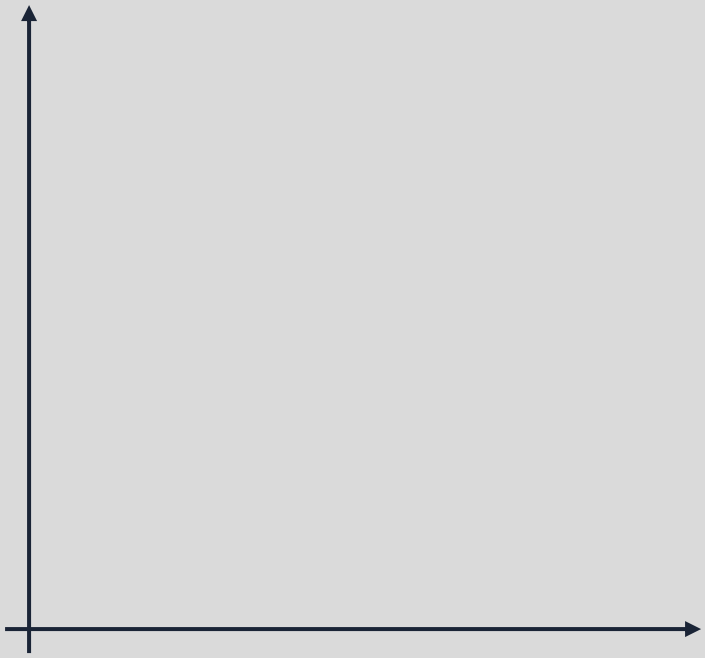
› $f(x_i) = P(X = x_i)$ for $x_i \in S_X$

› $f(x_i) = 0$ for $x_i \notin S_X$

Example Let X be a random variable of total points from rolling 2 fair dices, the sample space would be

› $S_X =$

(3) Probability Density Function (PDF)



Graphing the random variable and its probability here on the left.

Now figure out these probabilities.

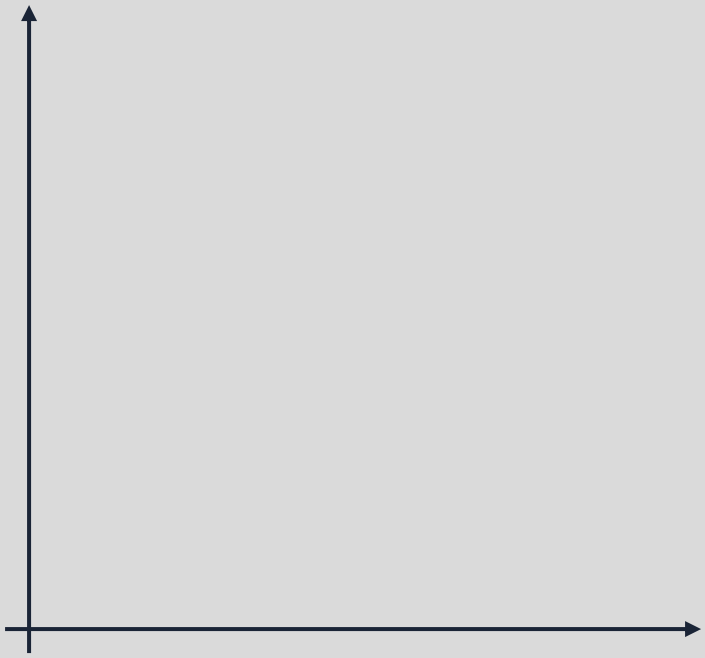
› $P(X = 4) =$

› $P(X = 7) =$

› $P(X < 3) =$

› $P(X \leq 4) + P(X > 9) =$

(3) Probability Density Function (PDF)



PDF can be both discrete and continuous.

Discreate PDF

› $0 \leq f(x) \leq 1$

› $\sum_{-\infty}^{\infty} f(x) = 1$

› $\sum_{a}^{b} f(x) = P(a \leq X \leq b)$

Continuous PDF

› $0 \leq f(x) \leq 1$

› $\int_{-\infty}^{\infty} f(x)dx = 1$

› $\int_{a}^{b} f(x)dx = P(a \leq X \leq b)$

(1) Conditional probability

There are a few basic concepts of bivariate distribution worth revising

› Joint probability density function

› Marginal probability

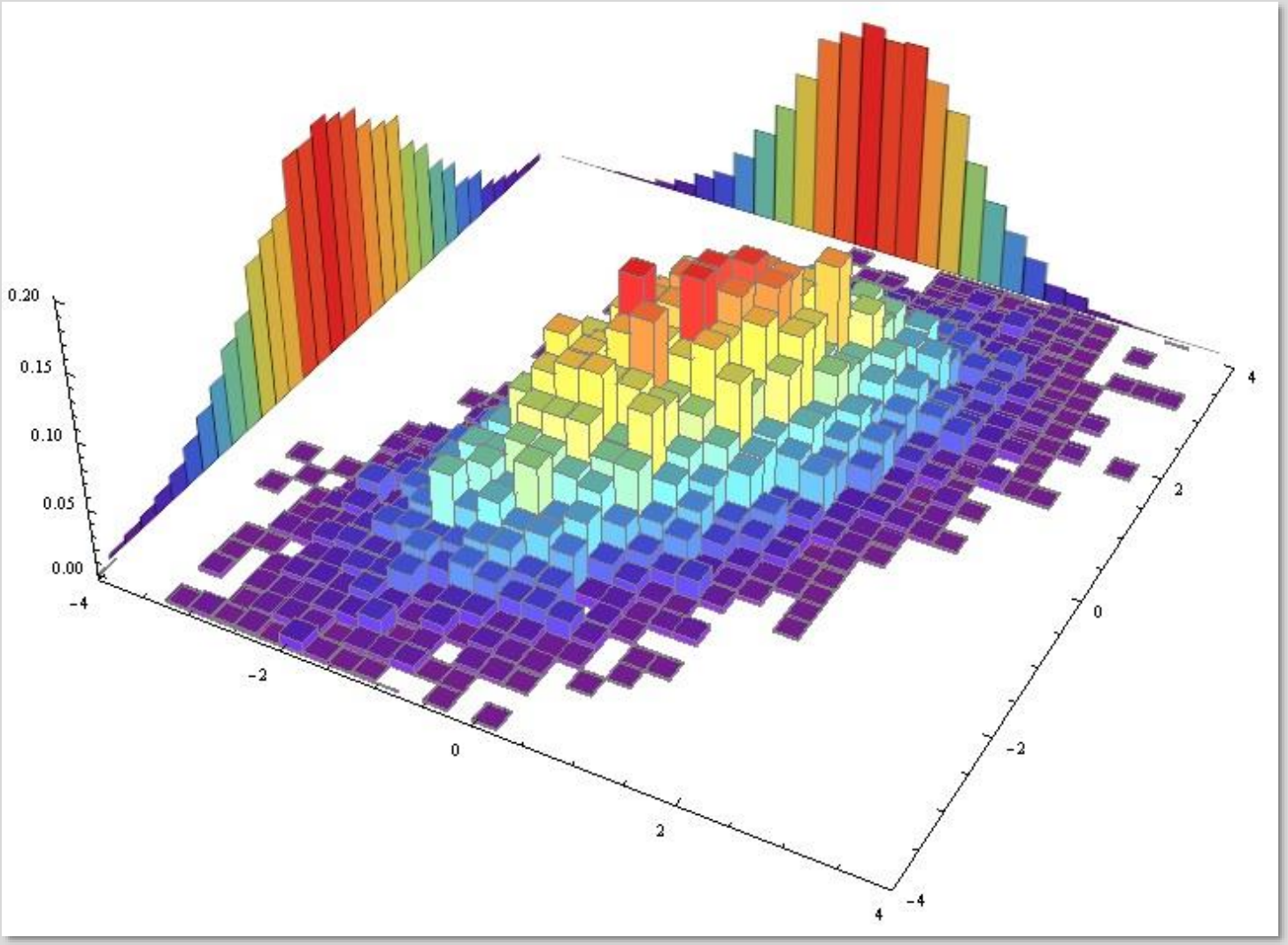
This class we will focus on

Definition 2.3

Conditional probability is a measure of the probability of an event occurring given that another event has (by assumption, presumption, assertion or evidence) occurred. Conditional probability is defined as

› $f(X|Y) = P(X = x|Y = y) = \frac{f(x,y)}{f(y)}$

(1) Conditional probability



(1) Conditional probability

Example Two archers are competing shooting a target for 3 times each.

Let X and Y be number of times archer 1 and archer and archer 2 hit the target respectively.

Thousand of rounds has been competed and the probability is computed as a result in this table.

		X		
		1	2	3
Y	1	0.35	0.2	0.07
	2	0.1	0.05	0.09
	3	0.08	0.04	0.02

› Find $f(X = 1|Y = 2) =$

› Find $f(Y = 2|X = 3) =$

(2) Statistical independence

Definition 2.4

Two random variables are considered **independent** if and only if the condition below is satisfied.

› $f(x, y) = f(x) \cdot f(y)$

Example Using the same archer example, prove that archers' performance is independent.

		X		
		1	2	3
Y	1	1/9	1/9	1/9
	2	1/9	1/9	1/9
	3	1/9	1/9	1/9

›

(1) Expected value

Definition 2.5

Expected value of a random variable is a generalization of the weighted average and intuitively is the arithmetic mean of independent realizations of that variable. Expected value is defined as

› $E(X) = \sum_{i=1}^{n} x_i \cdot f(x_i)$ - discrete

› $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$ - continuous

Grade	Prob
A	0.3
B	0.4
C	0.15
D	0.1
F	0.05

Example A student is trying hard for econometrics class. The probability getting grades are listed in the table.

›

(1) Expected value

Properties of expected value

(1) $E(\mathrm{a}) = \mathrm{a}$ for any constant a

(2) $E(\mathrm{a}X) = \mathrm{a}E(X)$

(3) $E(\mathrm{a}X + \mathrm{b}) = \mathrm{a}E(X) + \mathrm{b}$

(4) $E(X \pm Y) = E(X) \pm E(Y)$

(5) $E(XY) = E(X) \cdot E(Y)$
if and only if X and Y are independent.

(1) Expected value

Example Find the expected value of this distribution

$f(X) = \frac{1}{9}x^2$ for $0 \leq x \leq 3$

›

(2) Conditional expectation

Definition 2.6

Let $f(X,Y)$ be a joint probability density function, the expectation of X conditional on some value of Y is

› $E(X|Y) = \sum_X x_i \cdot f(X|Y=y)$ - discrete

› $E(X|Y) = \int_{-\infty}^{\infty} x \cdot f(X|Y=y)dx$ - continuous

		X			
		-2	0	2	3
Y	3	0.27	0.08	0.16	0
	6	0	0.04	0.1	0.35

Example Find $E(X|Y=3)$ from the PDF given in the table.

›

(3) Variance

Definition 2.7

Variance is a measure of data dispersion from the expected value. Given that μ is the expected value of X, then

› $var(X) = \sum_{i=1}^{n}(x_i - \mu)^2 \cdot f(x_i)$- discrete

› $var(X) = \int_{-\infty}^{\infty}(x - \mu)^2 \cdot f(x)dx$- continuous

Another formula for variance is

› $var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$

(3) Variance

Properties of variance

(1) $\text{var}(a) = 0$ for any constant a

(2) $\text{var}(aX + b) = a^2\text{var}(X)$

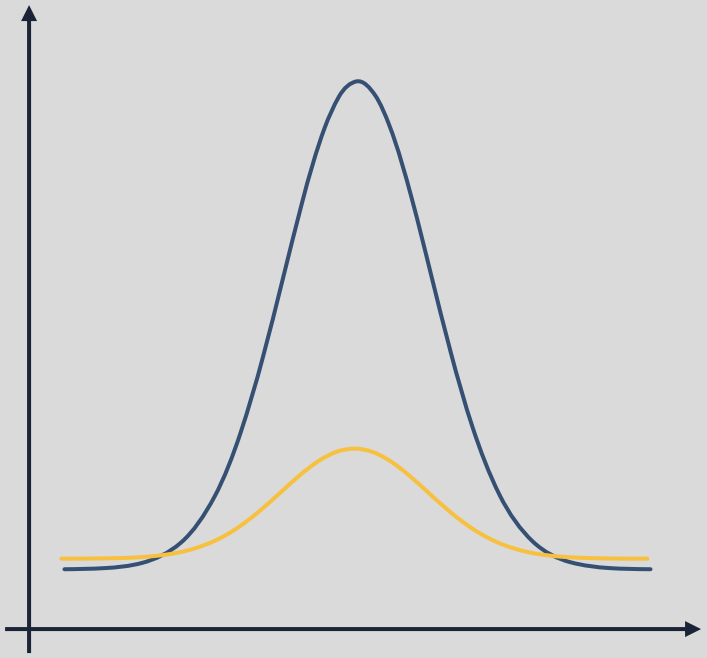
(3) $\text{var}(X \pm Y) = \text{var}(X) \pm \text{var}(Y)$
if and only if X and Y are independent.

X	-2	1	2
$f(X)$	5/8	1/8	2/8

Example Find variance from the PDF given in the table.

›

(3) Variance



Example Find variance from given distribution

› $f(X) = \frac{1}{9}x^2$ for $0 \leq x \leq 3$

(4) Conditional variance

Definition 2.8

Conditional variance is a measure variance, but coupled with a condition on another variable, defined as

› $var(X|Y) = \sum_X [x_i - E(X|Y = y)]^2 \cdot f(X|Y = y)$ - discrete

› $var(X|Y) = \int_{-\infty}^{\infty} [x - E(X|Y = y)]^2 \cdot f(X|Y = y)dx$ - continuous

(5) Covariance

Definition 2.8

Let X and Y be two random variables with expected value of μ_X and μ_Y respectively, the **covariance** is a measure of the joint variability of two random variables.

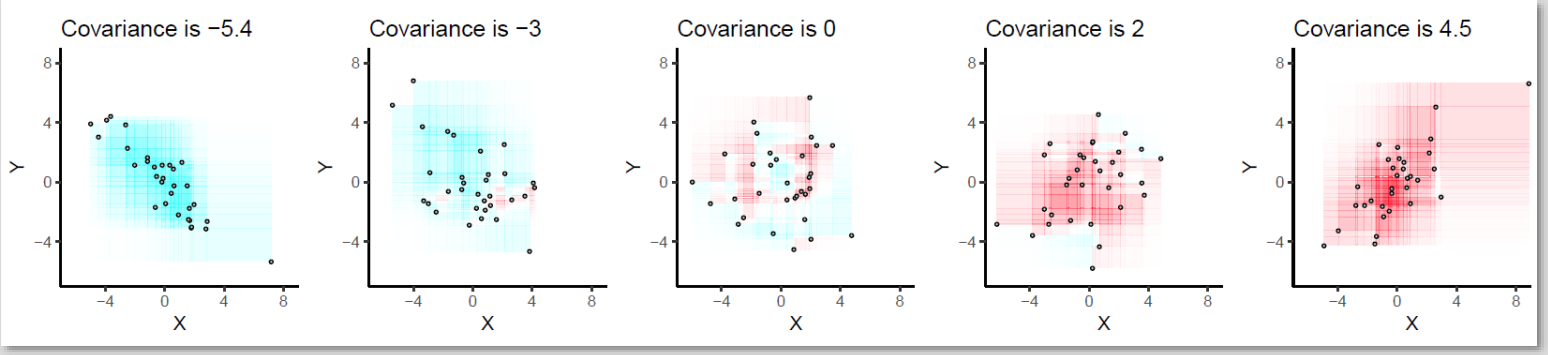
If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values. Defined as

› $cov(X,Y) = E\{(X-\mu_X)(Y-\mu_Y)\} = E(XY) - \mu_X\mu_Y$ - discrete

› $cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X-\mu_X)(Y-\mu_X)\cdot f(x,y)dxdy - \mu_X\mu_Y$

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XYf(x,y)\,dxdy - \mu_X\mu_Y$ - continuous

(5) Covariance



Further properties of variance

If X and Y are ***not*** independent, then

› $var(X \pm Y) = var(X) + var(Y) \pm 2cov(X, Y)$

Problems with interpretation

"A large covariance can mean a strong relationship between variables. However, you can't compare variances over data sets with different scales (like pounds and inches). A weak covariance in one data set may be a strong one in a different data set with different scales.

(6) Correlation coefficient

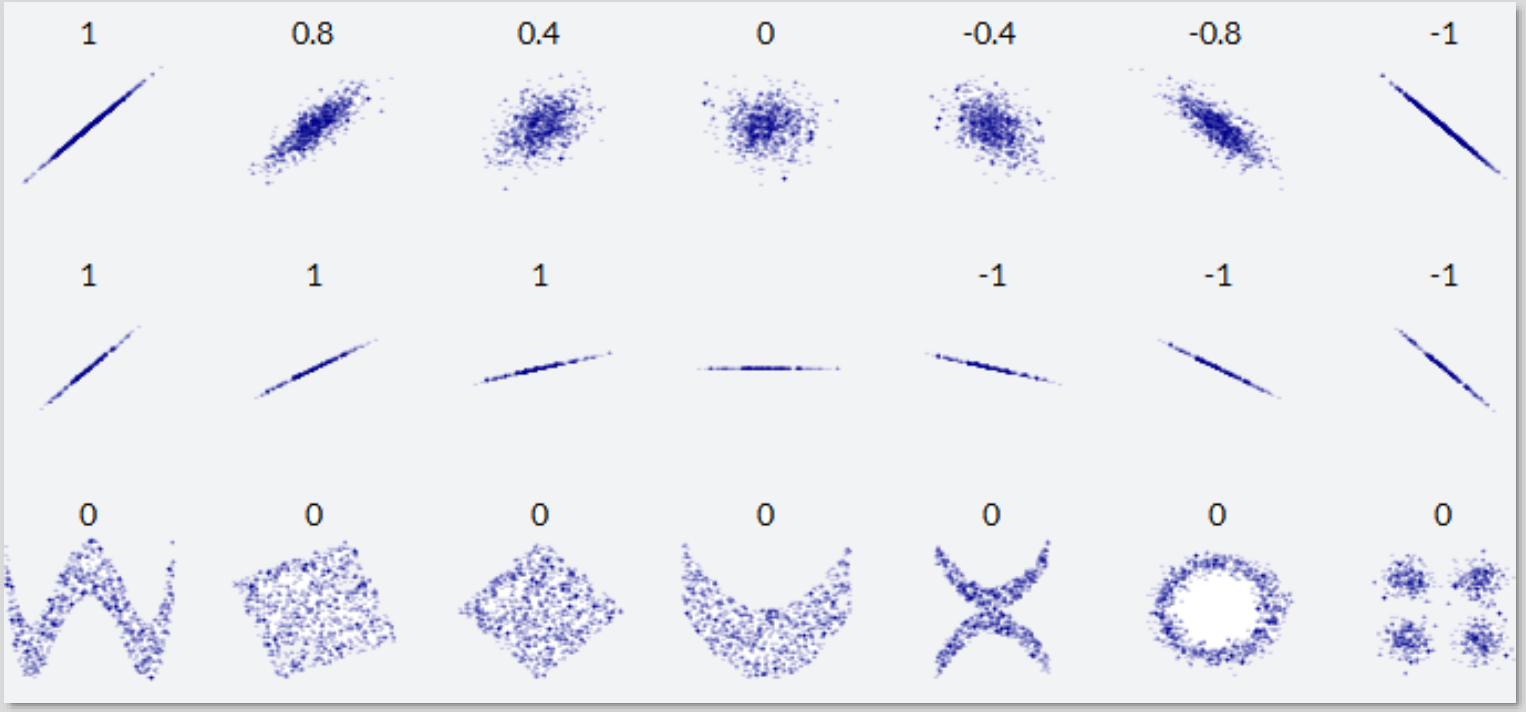
Definition 2.9

Correlation coefficient is a numerical measure of some type of correlation, meaning a statistical relationship between two variables, denoted by ρ_{XY}, r_{XY}, $\text{corr}(X, Y)$.

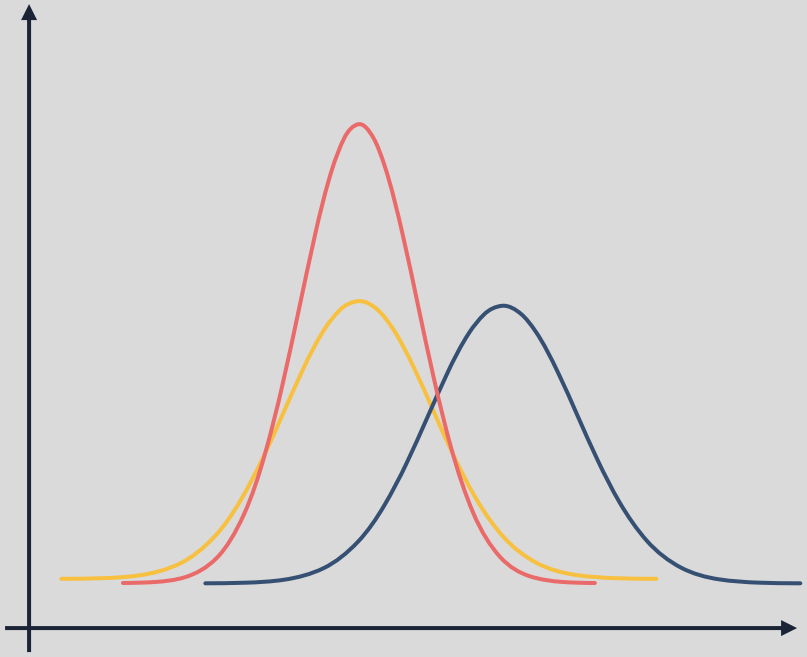
› $\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \epsilon [-1,1]$

where σ_X is standard deviation or $\sigma_X = \sqrt{\text{var}(X)}$

(6) Correlation coefficient



(1) Normal distribution



A continuous random variable X is normally distributed with mean μ and variance σ^2, denoted as $X \sim N(\mu, \sigma^2)$, if the PDF is

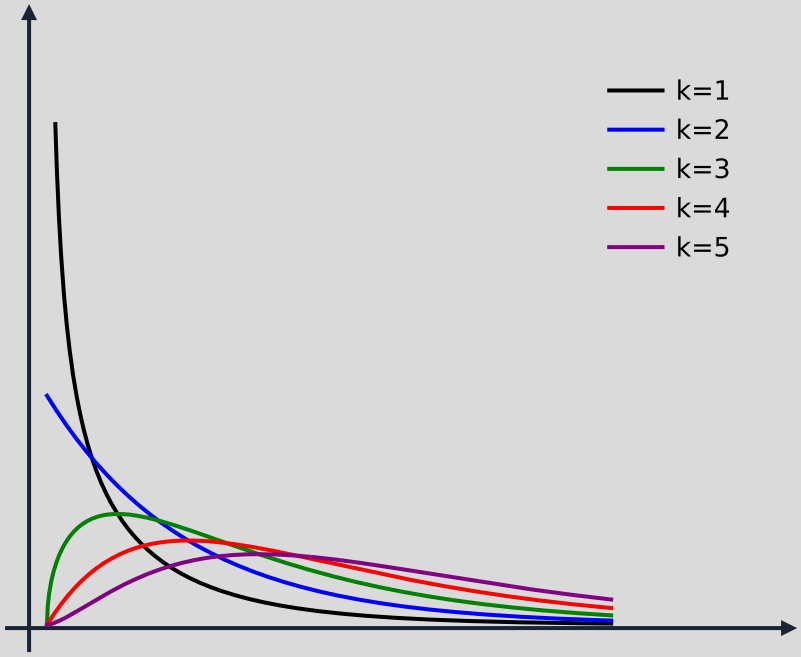
› $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

If the mean or variance changes, the position and shape of the distribution also shift.

We can convert any X that is normally distributed into a **standard normal distribution**, defined as Z, by weighting as follows.

› $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

(2) Chi-square distribution



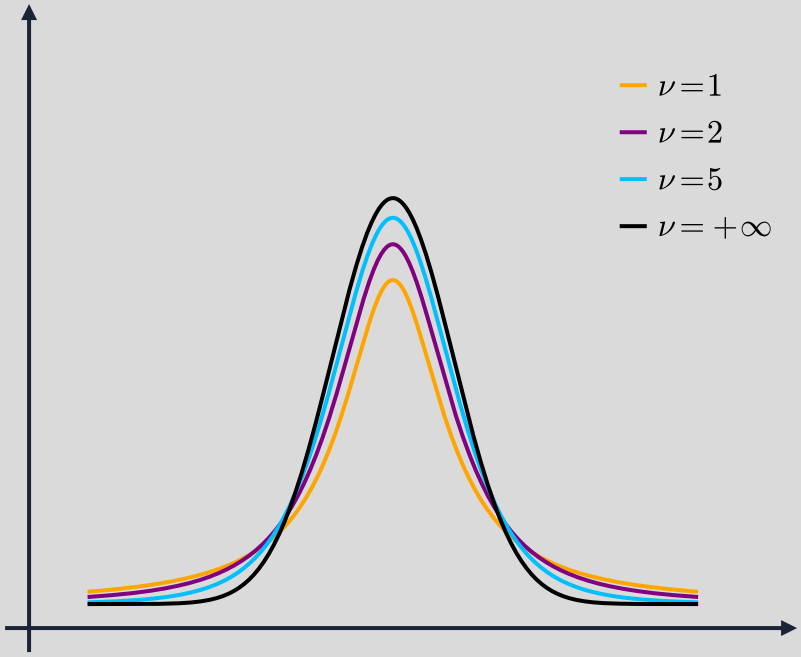
Chi-squared with k degrees of freedom (d.f.) is the distribution of a sum of the squares of k independent standard normal random variables.

› $\chi_k^2 = \sum_{i=1}^{k} Z_i^2$

Propertie of χ_k^2

› Chi-square is skewed depending on d.f. As the d.f. increases it becomes more and more symmetrical.

(3) Student's t-distribution



Let Z and χ_k^2 be random variables distributed as standard normal variable and chi-square respectively and they are independent, the t-distribution with k degrees of freedom can be represented as

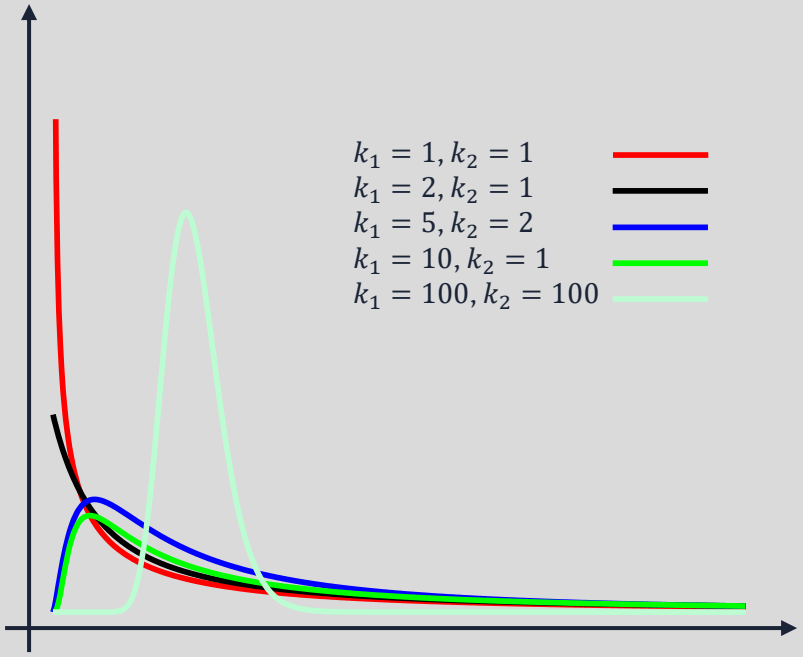
› $t = \frac{Z\sqrt{k}}{\chi_k^2} \sim t_v$

Properties of *t*

› The t-distribution is symmetric but flatter compared to the normal distribution.

› As the d.f. increases, t-distribution is converted to the normal distribution.

(4) F-distribution



Let χ_1^2 and χ_2^2 be random variables distributed as chi-square and they are independent with the d.f. of k_1 and k_2, the F-distribution can be represented as

› $F = \frac{\chi_1^2/k_1}{\chi_2^2/k_2} \sim F(k_1, k_2)$

Properties of F

› The F-distribution is skewed to the right but if k_1 and k_2 becomes larger, the F-distribution becomes normal distribution.

› The square of t-distributed random variable with k degrees of freedom is $t_v^2 = F_{1,v}$