BS2280 - Econometrics 1

Lecture 1 - Part 2: Review of Fundamental Statistical Methods

Dr. Yichen Zhu

Outline

The Summation Operator I

- Additions are very common in econometrics and statistics
- Example: three Italian restaurants in Birmingham sell Pizza and report their revenues (R) as follows:

Company 1	Company 2	Company 3
100	150	200

- Calculating total revenues (TR) across these three firms: $TR = R_1 + R_2 + R_3 = 100 + 150 + 200 = 450$

The Summation Operator I

- Additions are very common in econometrics and statistics
- Example: three Italian restaurants in Birmingham sell Pizza and report their revenues (R) as follows:

Company 1	Company 2	Company 3
100	150	200

- Calculating total revenues (TR) across these three firms: $TR = R_1 + R_2 + R_3 = 100 + 150 + 200 = 450$

The Summation Operator I

- Additions are very common in econometrics and statistics
- Example: three Italian restaurants in Birmingham sell Pizza and report their revenues (R) as follows:

Company 1	Company 2	Company 3
100	150	200

- Calculating total revenues (TR) across these three firms: $TR = R_1 + R_2 + R_3 = 100 + 150 + 200 = 450$

The Summation Operator II

- Calculate total revenues for 9 Italian restaurants in Birmingham:

C1	C2	C3	C4	C5	C6	C7	C8	C9
100	150	200	50	25	75	100	20	80

- Calculating total revenues (TR) across these three firms:
$TR = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 =$
$= 100 + 150 + 200 + 50 + 25 + 75 + 100 + 20 + 80 = 800$
Problem
- Revenues of 20 or 500 or 1 million companies?
- Calculations become cumbersome to write down - need a compact way of writing out calculations:
- Summation operator $\sum$ can simplify analysis

The Summation Operator II

- Calculate total revenues for 9 Italian restaurants in Birmingham:

C1	C2	C3	C4	C5	C6	C7	C8	C9
100	150	200	50	25	75	100	20	80

- Calculating total revenues (TR) across these three firms:
  $TR = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 =$
  $= 100 + 150 + 200 + 50 + 25 + 75 + 100 + 20 + 80 = 800$
  Problem
- Revenues of 20 or 500 or 1 million companies?
- Calculations become cumbersome to write down - need a compact way of writing out calculations:
- Summation operator $\sum$ can simplify analysis

The Summation Operator II

- Calculate total revenues for 9 Italian restaurants in Birmingham:

C1	C2	C3	C4	C5	C6	C7	C8	C9
100	150	200	50	25	75	100	20	80

- Calculating total revenues (TR) across these three firms:
$TR = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 =$
$= 100 + 150 + 200 + 50 + 25 + 75 + 100 + 20 + 80 = 800$
Problem
- Revenues of 20 or 500 or 1 million companies?
- Calculations become cumbersome to write down - need a compact way of writing out calculations:
- Summation operator $\sum$ can simplify analysis

The Summation Operator II

- Calculate total revenues for 9 Italian restaurants in Birmingham:

C1	C2	C3	C4	C5	C6	C7	C8	C9
100	150	200	50	25	75	100	20	80

- Calculating total revenues (TR) across these three firms:
  $TR = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 =$
  $= 100 + 150 + 200 + 50 + 25 + 75 + 100 + 20 + 80 = 800$
  Problem
- Revenues of 20 or 500 or 1 million companies?
- Calculations become cumbersome to write down - need a compact way of writing out calculations:
- Summation operator $\sum$ can simplify analysis

The Summation Operator II

- Calculate total revenues for 9 Italian restaurants in Birmingham:

C1	C2	C3	C4	C5	C6	C7	C8	C9
100	150	200	50	25	75	100	20	80

- Calculating total revenues (TR) across these three firms:
  $TR = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 =$
  $= 100 + 150 + 200 + 50 + 25 + 75 + 100 + 20 + 80 = 800$
  Problem
- Revenues of 20 or 500 or 1 million companies?
- Calculations become cumbersome to write down - need a compact way of writing out calculations:
- Summation operator $\sum$ can simplify analysis

The Summation Operator III

- Use summation operator to represent calculations across many units

$$TR = \sum_{i=1}^{n} R_i$$

- "*Sum the values of revenues across all companies, from the first company (i = 1) to the last company (i = n)*"
- Using previous examples:

$$TR = \sum_{i=1}^{3} R_i = R_1 + R_2 + R_3 = 450$$

$$TR = \sum_{i=1}^{9} R_i = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 = 800$$

The Summation Operator III

- Use summation operator to represent calculations across many units

$$TR = \sum_{i=1}^{n} R_i$$

- "*Sum the values of revenues across all companies, from the first company (i = 1) to the last company (i = n)*"
- Using previous examples:

$$TR = \sum_{i=1}^{3} R_i = R_1 + R_2 + R_3 = 450$$

$$TR = \sum_{i=1}^{9} R_i = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 = 800$$

The Summation Operator III

- Use summation operator to represent calculations across many units

$$TR = \sum_{i=1}^{n} R_i$$

- "*Sum the values of revenues across all companies, from the first company (i = 1) to the last company (i = n)*"
- Using previous examples:

$$TR = \sum_{i=1}^{3} R_i = R_1 + R_2 + R_3 = 450$$

$$TR = \sum_{i=1}^{9} R_i = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 = 800$$

The Summation Operator III

- Use summation operator to represent calculations across many units

$$TR = \sum_{i=1}^{n} R_i$$

- "*Sum the values of revenues across all companies, from the first company (i = 1) to the last company (i = n)*"
- Using previous examples:

$$TR = \sum_{i=1}^{3} R_i = R_1 + R_2 + R_3 = 450$$

$$TR = \sum_{i=1}^{9} R_i = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 = 800$$

The Summation Operator Example

- Consider the following data for three Italian restaurants:

company (i)	price (p)	quantity (q)	chefs (c)	waiters (w)
1	3	150	2	5
2	3	200	3	5
3	3	250	2	5

- Calculate:
  - Total Number of waiters (TW)
  - Total Revenue (TR)
  - Total number of employees (TE)

The Summation Operator Example

- Consider the following data for three Italian restaurants:

company (i)	price (p)	quantity (q)	chefs (c)	waiters (w)
1	3	150	2	5
2	3	200	3	5
3	3	250	2	5

- Calculate:
  - Total Number of waiters (TW)
  - Total Revenue (TR)
  - Total number of employees (TE)

The Summation Operator Example

- Consider the following data for three Italian restaurants:

company (i)	price (p)	quantity (q)	chefs (c)	waiters (w)
1	3	150	2	5
2	3	200	3	5
3	3	250	2	5

- Calculate:
  - Total Number of waiters (TW)
  - Total Revenue (TR)
  - Total number of employees (TE)

The Summation Operator Example

- Consider the following data for three Italian restaurants:

company (i)	price (p)	quantity (q)	chefs (c)	waiters (w)
1	3	150	2	5
2	3	200	3	5
3	3	250	2	5

- Calculate:
  - Total Number of waiters (TW)
  - Total Revenue (TR)
  - Total number of employees (TE)

The Summation Operator Example

- Consider the following data for three Italian restaurants:

company (i)	price (p)	quantity (q)	chefs (c)	waiters (w)
1	3	150	2	5
2	3	200	3	5
3	3	250	2	5

- Calculate:
  - Total Number of waiters (TW)
  - Total Revenue (TR)
  - Total number of employees (TE)

The Summation Operator IV

- 3 main summation operator rules will make our life much easier during the module!

1. $\sum_{i=1}^{n} a = na$ where a is a constant
2. $\sum_{i=1}^{n} ax_i = a \sum_{i=1}^{n} x_i$ where a is a constant
3. $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

The Summation Operator IV

- 3 main summation operator rules will make our life much easier during the module!

1. $\sum_{i=1}^{n} a = na$ where a is a constant
2. $\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i$ where a is a constant
3. $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

The Summation Operator IV

- 3 main summation operator rules will make our life much easier during the module!

1. $\sum_{i=1}^{n} a = na$ where a is a constant
2. $\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i$ where a is a constant
3. $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

The Summation Operator IV

- 3 main summation operator rules will make our life much easier during the module!

1. $\sum_{i=1}^{n} a = na$ where a is a constant
2. $\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i$ where a is a constant
3. $\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

Rule 1

company (i)	price (p)	quantity (q)	chefs (c)	waiters (w)
1	3	150	2	5
2	3	200	3	5
3	3	250	2	5

- Calculate Total number of waiters (TW)

$$TW = 5 + 5 + 5 = 15$$

- Could have written above more compactly as:

$$TW = \sum_{i=1}^{n} w_i$$

- since w is a constant equal to 5:

$$TW = \sum_{i=1}^{n} w = nw = \sum_{i=1}^{3} 5 = 3 \times 5 = 15$$

Rule 2

company (i)	price (p)	quantity (q)	chefs (c)	waiters (w)
1	3	150	2	5
2	3	200	3	5
3	3	250	2	5

- Calculate Total Revenue ($p \times q$)

$$\begin{aligned} TR &= (p_1 \times q_1) + (p_2 \times q_2) + (p_3 \times q_3) = \\ &= (3 \times 150) + (3 \times 200) + (3 \times 250) = 1800 \end{aligned}$$

- Could have written above more compactly as:

$$TR = \sum_{i=1}^{n} p_i q_i = \sum_{i=1}^{3} p_i q_i$$

- since p is a constant equal to 3:

$$TR = \sum_{i=1}^{3} 3q_i = 3\sum_{i=1}^{3} q_i = 3(150 + 200 + 250) = 1800$$

Rule 3

company (i)	price (p)	quantity (q)	chefs (c)	waiters (w)
1	3	150	2	5
2	3	200	3	5
3	3	250	2	5

- Calculate Total Number of Employees (w + c)

$$TE = (c_1 + w_1) + (c_2 + w_2) + (c_3 + w_3) =$$
$$= (2 + 5) + (3 + 5) + (2 + 5) = 22$$

- Could have written above more compactly as:

$$TE = \sum_{i=1}^{n} (c_i + w_i) = \sum_{i=1}^{3} (c_i + w_i)$$

- this is identical to

$$TE = \sum_{i=1}^{3} c_i + \sum_{i=1}^{3} w_i = 7 + 15 = 22$$

Scientific notation

Extremely large or small numerical values

- GDP of UK (2018): £2,855,000,000,000 (2.9 trillion £)
- Apples A14 chip: 0.000000005m (5 nanometre)

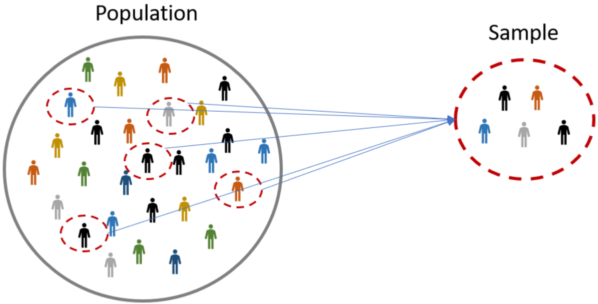
Scientific notation

- GDP of UK (2018): £2.9×10^{12}
- Apples new A14 chip: 5×10^{-9}m

Scientific notation in R

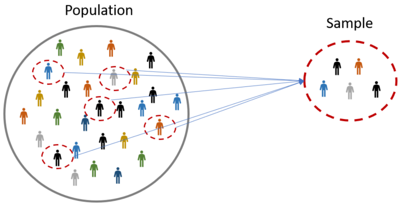
- GDP of UK (2018): `2.9e12`
- Apples new A14 chip: `5e-09`

Review: Population vs Sample



- Population: refers to *all* cases or situations the 'statistician' wants his inferences or guesses or estimates to apply to.
- Sample: a relatively small selection from the population. We use samples to make inferences for population.

Review: Population vs Sample



- Examples:

Question	Population	Sample
Starting salary of graduate	All graduates	Aston graduates
Voting forecasts	All UK voters	A random 10% selection of UK voters
Salt content of meal	The whole pot of the meal	A spoon of the meal

Review: Measures of location and dispersion

Measure of central tendency, E.g. mean

- Is a single value that attempts to describe a set of data by identifying the central position within that set of data.
- Example: What is the average income of economics graduates in the UK?

Measure of dispersion, E.g. variance, standard deviation

- Measure how spread out a set of data is.
- Example: How dispersed is the income of economics graduates in the UK?

Review: Measures of location and dispersion

Measure of central tendency, E.g. mean

- Is a single value that attempts to describe a set of data by identifying the central position within that set of data.
- Example: What is the average income of economics graduates in the UK?

Measure of dispersion, E.g. variance, standard deviation

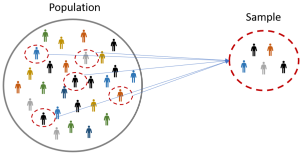
- Measure how spread out a set of data is.
- Example: How dispersed is the income of economics graduates in the UK?

Review: Measures of location and dispersion

- Arithmetic Mean (μ or $\bar{x}$): measure of a mean, or average
- Variance (σ^2 or s^2): Average of the squared discrepancies from the mean.
- Standard deviation (σ or s): Positive square root of variance.

Measures		Population	Sample
Measure of central tendency	Arithmetic Mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$ $N \ldots$ population size	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ $n \ldots$ sample size
Measure of dispersion	Variance (σ^2 or s^2)	$\sigma^2 = \frac{\sum_{i=1}^{N}(Y_i - \mu)^2}{N}$	$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$
	Standard deviation (σ or s)	$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \mu)^2}{N}}$	$\hat{\sigma} = s = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$

Review: Measures of location and dispersion



Measures		Population	Sample
Measure of central tendency	Arithmetic Mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$ $N \ldots$ population size	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ $n \ldots$ sample size
Measure of dispersion	Variance (σ^2 or s^2)	$\sigma^2 = \frac{\sum_{i=1}^{N} (Y_i - \mu)^2}{N}$	$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}$
	Standard deviation (σ or s)	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \mu)^2}{N}}$	$\hat{\sigma} = s = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}}$

Review: Variance and Standard Deviation

Student Task:

- (i) What is the average revenue of the three Italian restaurants? ($q_1 = 150$, $q_2 = 200$ and $q_3 = 250$)?
- (ii) What is the Standard Deviation of revenue?

Solution I (using R)

(i) What is the average revenue of the three Italian restaurants? ($q_1 = 150$, $q_2 = 200$ and $q_3 = 250$)?

- First, we assign our values to an object. Note we have to combine the values to a vector using the `c` command:

```
> revenue <- c(150, 200, 250)
> revenue
[1] 150 200 250
```

- Use `mean()` command to calculate mean:

```
> mean.revenue <- mean(revenue)
> mean.revenue
[1] 200
```

Solution II (using R)

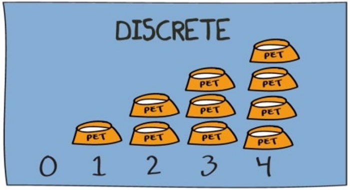
(ii) What is the Standard Deviation of revenue?

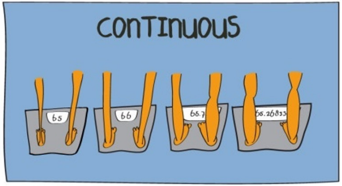
- Use `sd()` command to calculate standard deviation:

```
> sd.revenue <- sd(revenue)
> sd.revenue
[1] 50
```

Discrete Random Variables and Expectations I

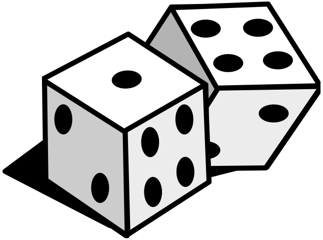
- A random variable (RV) is any variable whose value is not known in advance and cannot be predicted exactly
- Two types of RVs exist: discrete RVs and continuous RVs





Discrete Random Variables and Expectations I

- Population of an RV is the set of all possible values of an RV



Example of an RV

Total score when you through a dice twice - call it X.
We do not know which score we get but we know the population of X will be numbers from 2 to 12.

Discrete Random Variables and Expectations II

First dice

Second dice	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- Each of the numbers from 2 to 12 is an experimental outcome
- 36 possible combinations
- X can only take 11 possible values from 2 to 12
- Calculate frequencies & probability distributions on these numbers

Discrete Random Variables and Expectations II

First dice

Second dice	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- Each of the numbers from 2 to 12 is an experimental outcome
- 36 possible combinations
- X can only take 11 possible values from 2 to 12
- Calculate frequencies & probability distributions on these numbers

Discrete Random Variables and Expectations II

First dice

Second dice	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- Each of the numbers from 2 to 12 is an experimental outcome
- 36 possible combinations
- X can only take 11 possible values from 2 to 12
- Calculate frequencies & probability distributions on these numbers

Discrete Random Variables and Expectations II

First dice

Second dice	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- Each of the numbers from 2 to 12 is an experimental outcome
- 36 possible combinations
- X can only take 11 possible values from 2 to 12
- Calculate frequencies & probability distributions on these numbers

Discrete Random Variables and Expectations II

First dice

Second dice	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

- Each of the numbers from 2 to 12 is an experimental outcome
- 36 possible combinations
- X can only take 11 possible values from 2 to 12
- Calculate frequencies & probability distributions on these numbers

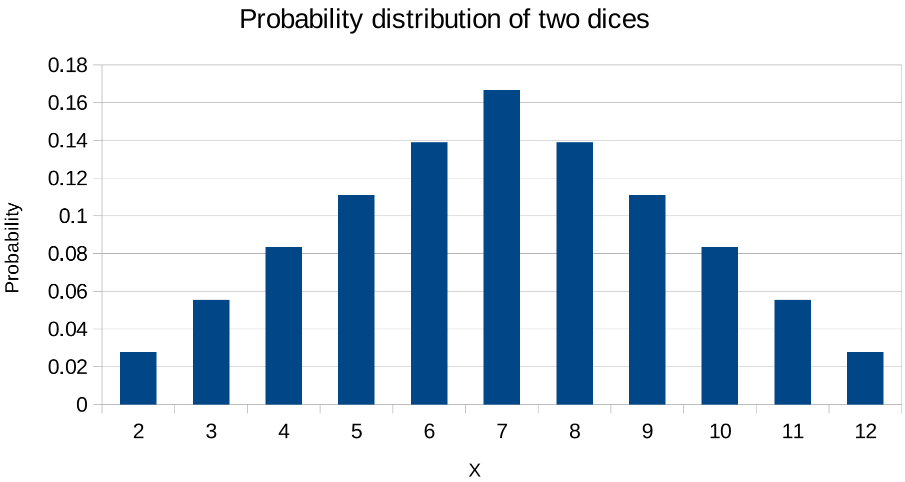
Frequency and probability distribution

Total score on rolling two dice

Value of X	2	3	4	5	6	7	8	9	10	11	12
Frequency	1	2	3	4	5	6	5	4	3	2	1
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

These probabilities should add up to 1

Distribution of Scores (discrete variable)



- **Expected value** of a random variable X is the weighted average of all its values, with probability of each outcome used as weight
- If X can take n particular values, $x_1, x_2, x_3, \ldots, x_n$ and probability of each x_i is p_i, the expected value of X, E(X) is:

$$E(X) = x_1 p_1 + x_2 p_2 + x_3 p_3 + \cdots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

- From previous dice rolling experiment, the values of $x_1, x_2, x_3, \ldots, x_n$ range from 2 to 12, with their corresponding probabilities, $p_1, p_2, p_3, \ldots, p_n$

$$E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + \cdots + 12 \times \frac{1}{36} = 7$$

- **Expected value** of a random variable X is the weighted average of all its values, with probability of each outcome used as weight
- If X can take n particular values, $x_1, x_2, x_3, \ldots, x_n$ and probability of each x_i is p_i, the expected value of X, E(X) is:

$$E(X) = x_1 p_1 + x_2 p_2 + x_3 p_3 + \cdots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

- From previous dice rolling experiment, the values of $x_1, x_2, x_3, \ldots, x_n$ range from 2 to 12, with their corresponding probabilities, $p_1, p_2, p_3, \ldots, p_n$

$$E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + \cdots + 12 \times \frac{1}{36} = 7$$

- **Expected value** of a random variable X is the weighted average of all its values, with probability of each outcome used as weight
- If X can take n particular values, $x_1, x_2, x_3, \ldots, x_n$ and probability of each x_i is p_i, the expected value of X, E(X) is:

$$E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \cdots + x_np_n = \sum_{i=1}^{n} x_ip_i$$

- From previous dice rolling experiment, the values of $x_1, x_2, x_3, \ldots, x_n$ range from 2 to 12, with their corresponding probabilities, $p_1, p_2, p_3, \ldots, p_n$

$$E(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + \cdots + 12 \times \frac{1}{36} = 7$$

Student Task: Expected Value

- Expected values are important for decision making! (but not only...)
- Assume you have two investment options (A, B) whereby their return will depend on the economic performance:

Econ. Perf.	Prob.	Portfolio A	Portfolio B
Boom	0.5	20	50
Slow down	0.3	10	0
Recession	0.2	-10	-50

- What is the expected value of investment A and B? Which investment would you select?

Student Task: Expected Value

- Expected values are important for decision making! (but not only...)
- Assume you have two investment options (A, B) whereby their return will depend on the economic performance:

Econ. Perf.	Prob.	Portfolio A	Portfolio B
Boom	0.5	20	50
Slow down	0.3	10	0
Recession	0.2	-10	-50

- What is the expected value of investment A and B? Which investment would you select?

Student Task: Expected Value

- Expected values are important for decision making! (but not only...)
- Assume you have two investment options (A, B) whereby their return will depend on the economic performance:

Econ. Perf.	Prob.	Portfolio A	Portfolio B
Boom	0.5	20	50
Slow down	0.3	10	0
Recession	0.2	-10	-50

- What is the expected value of investment A and B? Which investment would you select?

More rules for expected values...

- Rules below valid for both discrete and continuous random variables
  1. $E(X+Y+Z) = E(X) + E(Y) + E(Z)$
  2. $E(bX) = bE(X)$ where b is a constant
  3. $E(b) = b$ where b is a constant
- Example: Calculate E(Y) for the following (b_1 , b_2 are constants)

$$Y = b_1 + b_2X$$

More rules for expected values...

- Rules below valid for both discrete and continuous random variables

1. $E(X + Y + Z) = E(X) + E(Y) + E(Z)$
2. $E(bX) = bE(X)$ where b is a constant
3. $E(b) = b$ where b is a constant

- Example: Calculate E(Y) for the following (b_1 , b_2 are constants)

$$Y = b_1 + b_2 X$$

- Applying above rules:

$$E(Y) = E(b_1 + b_2 X) = E(b_1) + E(b_2 X) = b_1 + b_2 E(X)$$

Continuous Random Variable Distributions

- Many variables are continuous, e.g. temperature, time waiting in the morning for a Cross City Line train.
- Example: Assume a MCQ test had a mean value of 53.4 and a s.d. of 15.1
- Most students have a mark around the average, some of the further away from the mean, the more unlikely the mark gets.
- MCQ marks could be approx. normal
- Normal distribution will be the key distribution for econometrics!!!
- What does it look like?

Continuous Random Variable Distributions

- Many variables are continuous, e.g. temperature, time waiting in the morning for a Cross City Line train.
- Example: Assume a MCQ test had a mean value of 53.4 and a s.d. of 15.1
- Most students have a mark around the average, some of the further away from the mean, the more unlikely the mark gets.
- MCQ marks could be approx. normal
- Normal distribution will be the key distribution for econometrics!!!
- What does it look like?

Continuous Random Variable Distributions

- Many variables are continuous, e.g. temperature, time waiting in the morning for a Cross City Line train.
- Example: Assume a MCQ test had a mean value of 53.4 and a s.d. of 15.1
- Most students have a mark around the average, some of the further away from the mean, the more unlikely the mark gets.
- MCQ marks could be approx. normal
- Normal distribution will be the key distribution for econometrics!!!
- What does it look like?

Continuous Random Variable Distributions

- Many variables are continuous, e.g. temperature, time waiting in the morning for a Cross City Line train.
- Example: Assume a MCQ test had a mean value of 53.4 and a s.d. of 15.1
- Most students have a mark around the average, some of the further away from the mean, the more unlikely the mark gets.
- MCQ marks could be approx. normal
- Normal distribution will be the key distribution for econometrics!!!
- What does it look like?

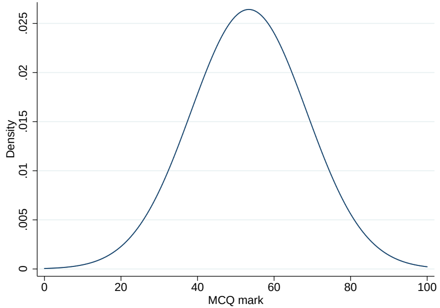
Continuous Random Variable Distributions

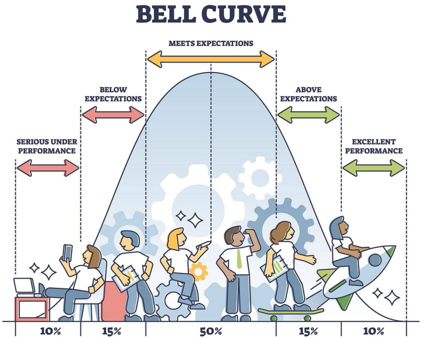
- Many variables are continuous, e.g. temperature, time waiting in the morning for a Cross City Line train.
- Example: Assume a MCQ test had a mean value of 53.4 and a s.d. of 15.1
- Most students have a mark around the average, some of the further away from the mean, the more unlikely the mark gets.
- MCQ marks could be approx. normal
- Normal distribution will be the key distribution for econometrics!!!
- What does it look like?

Continuous Random Variable Distributions

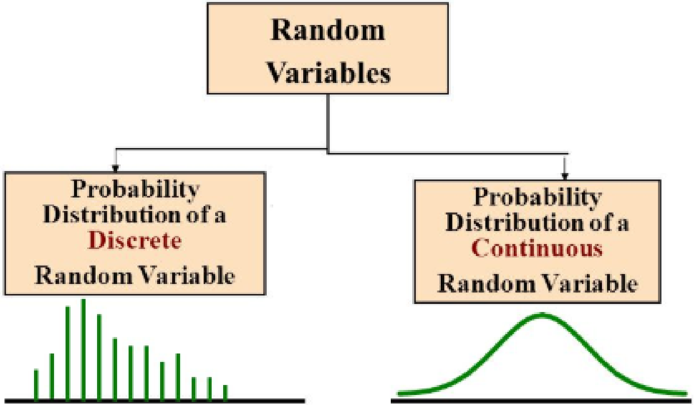
- Many variables are continuous, e.g. temperature, time waiting in the morning for a Cross City Line train.
- Example: Assume a MCQ test had a mean value of 53.4 and a s.d. of 15.1
- Most students have a mark around the average, some of the further away from the mean, the more unlikely the mark gets.
- MCQ marks could be approx. normal
- Normal distribution will be the key distribution for econometrics!!!
- What does it look like?

Case Study: Student MCQ Marks III

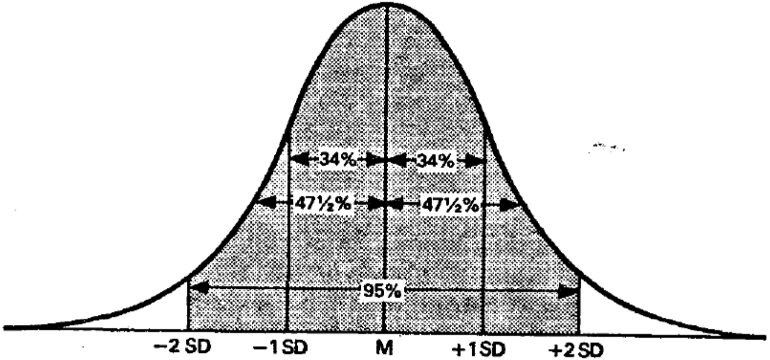




Continuous Random Variable Distributions



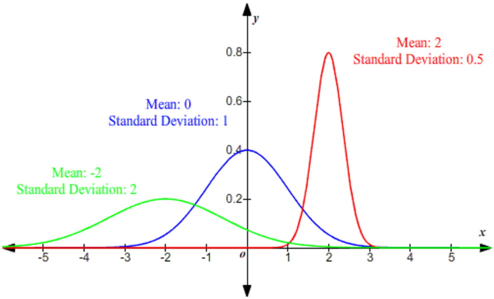
Continuous Random Variable Distributions



The shaded area represents 95% of the area under the curve. It is bounded on the left by a vertical line drawn up from whatever value of the variable falls at the −2 SD position. On the right it is bounded by the +2 SD point.

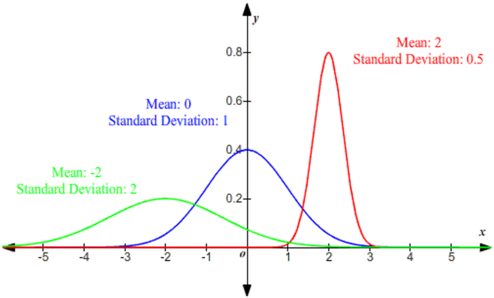
Continuous Random Variable Distributions

- Area below distribution curve shows probability
- Shape depends on mean value and standard deviation
- How can I compare different normal distribution? How can I undertake standardised hypothesis test?
- Use standard normal distribution, N(0,1)



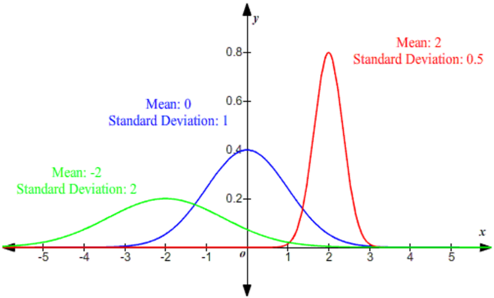
Continuous Random Variable Distributions

- Area below distribution curve shows probability
- Shape depends on mean value and standard deviation
- How can I compare different normal distribution? How can I undertake standardised hypothesis test?
- Use standard normal distribution, N(0,1)



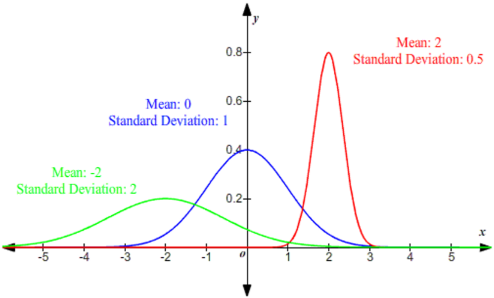
Continuous Random Variable Distributions

- Area below distribution curve shows probability
- Shape depends on mean value and standard deviation
- How can I compare different normal distribution? How can I undertake standardised hypothesis test?
- Use standard normal distribution, N(0,1)



Continuous Random Variable Distributions

- Area below distribution curve shows probability
- Shape depends on mean value and standard deviation
- How can I compare different normal distribution? How can I undertake standardised hypothesis test?
- Use standard normal distribution, N(0,1)



Distribution of Sample Means

- So far we analysed the distribution of one variable.
- Assume you take several samples from the whole population.
- How is the mean of all the samples distributed?
- We use *sample statistics* to estimate *population parameters*.
- This is the foundation of any estimation and hypotheses testing.

Distribution of Sample Means

- So far we analysed the distribution of one variable.
- Assume you take several samples from the whole population.
- How is the mean of all the samples distributed?
- We use *sample statistics* to estimate *population parameters*.
- This is the foundation of any estimation and hypotheses testing.

Distribution of Sample Means

- So far we analysed the distribution of one variable.
- Assume you take several samples from the whole population.
- How is the mean of all the samples distributed?
- We use *sample statistics* to estimate *population parameters*.
- This is the foundation of any estimation and hypotheses testing.

Distribution of Sample Means

- So far we analysed the distribution of one variable.
- Assume you take several samples from the whole population.
- How is the mean of all the samples distributed?
- We use *sample statistics* to estimate *population parameters*.
- This is the foundation of any estimation and hypotheses testing.

Distribution of Sample Means

- So far we analysed the distribution of one variable.
- Assume you take several samples from the whole population.
- How is the mean of all the samples distributed?
- We use *sample statistics* to estimate *population parameters*.
- This is the foundation of any estimation and hypotheses testing.

Application III

- Do we always take several samples of the population to derive the mean and standard deviation of sample means?
- Usually, we only have one sample, but that is sufficient!
- We know that $E(\bar{x}) = \mu$ and the variance of sample means will be $\frac{\sigma^2}{n}$.
- The standard deviation of the sampling distribution of the mean is called **Standard Error** of the Mean.
- For a mathematical derivation see Barrow page 134.

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \tag{1}$$

Application III

- Do we always take several samples of the population to derive the mean and standard deviation of sample means?
- Usually, we only have one sample, but that is sufficient!
- We know that $E(\bar{x}) = \mu$ and the variance of sample means will be $\dfrac{\sigma^2}{n}$.
- The standard deviation of the sampling distribution of the mean is called **Standard Error** of the Mean.
- For a mathematical derivation see Barrow page 134.

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \tag{1}$$

Application III

- Do we always take several samples of the population to derive the mean and standard deviation of sample means?
- Usually, we only have one sample, but that is sufficient!
- We know that $E(\bar{x}) = \mu$ and the variance of sample means will be $\frac{\sigma^2}{n}$.
- The standard deviation of the sampling distribution of the mean is called **Standard Error** of the Mean.
- For a mathematical derivation see Barrow page 134.

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \tag{1}$$

Application III

- Do we always take several samples of the population to derive the mean and standard deviation of sample means?
- Usually, we only have one sample, but that is sufficient!
- We know that $E(\bar{x}) = \mu$ and the variance of sample means will be $\frac{\sigma^2}{n}$.
- The standard deviation of the sampling distribution of the mean is called **Standard Error** of the Mean.
- For a mathematical derivation see Barrow page 134.

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (1)$$

Application III

- Do we always take several samples of the population to derive the mean and standard deviation of sample means?
- Usually, we only have one sample, but that is sufficient!
- We know that $E(\bar{x}) = \mu$ and the variance of sample means will be $\frac{\sigma^2}{n}$.
- The standard deviation of the sampling distribution of the mean is called **Standard Error** of the Mean.
- For a mathematical derivation see Barrow page 134.

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \tag{1}$$

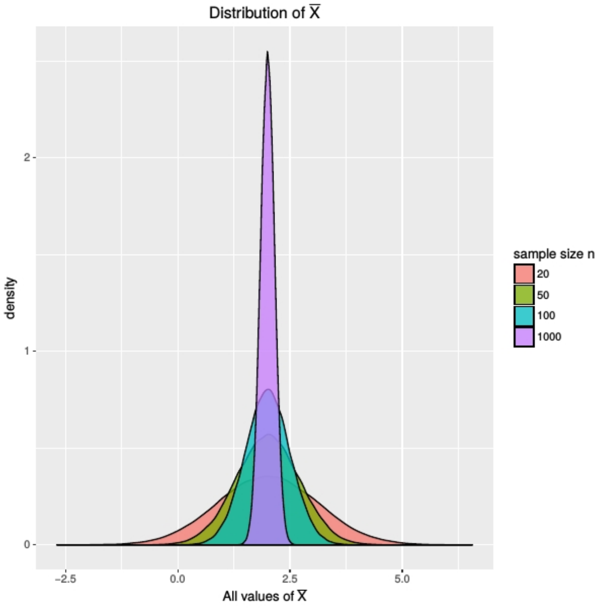
Application IV

Theorem

The mean, $\bar{x}$, of a random sample drawn from a population which has a normal distribution with a mean μ and a variance σ^2, has a sampling distribution which is Normal, with mean μ and variance $\frac{\sigma^2}{n}$, where n is the sample size.

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \leftarrow \text{Sample Mean Distribution}$$

$$x \sim N\left(\mu, \sigma^2\right) \quad \leftarrow \text{Population Distribution}$$


Distribution of $\overline{X}$
density
2
1
0
-2.5
0.0
2.5
5.0
All values of $\overline{X}$
sample size n
20
50
100
1000

Another Important Theorem

Central Limit Theorem

- If the sample size is large ($n > 30$) the population does not have to be Normally distributed, the sample mean is (approximately) Normal whatever the shape of the population distribution.
- The approximation gets better, the larger the sample size. 30 is a safe minimum to use.