Objective

Random

Definition

Probability Distributi of RV

Expected Value of a RV

Other Properties of RV

Sampling

Population vs Sampl

Estimation and

Daine Cationation

Evaluation of Estima

Hypothesis Testing

Thankyou

Random Variables, Sampling, Estimation and Inference

Parthasarathi Edupally¹

¹DataConscientious LLP, Mumbai

Elements of Econometrics, Russell Square International College, Mumbai

Outline

Objectives

Objective

Random

Definition

Probability Distribution of RV

Expected Value of a RV

Other Properties of R

Sampling

Population vs Sample

Estimation and

Point Estimation

Hypothesis Testing

-->1-----

Outline

- Objectives
- 2 Random Variables

Definition
Probability Distribution of RV
Expected Value of a RV
Other Properties of RVs

Objective

Randon Variable

Definition
Probability Distributi

Expected Value of a RV

Other Properties of I

Samplin

Random Sample

Estimation and

Point Estimation

H---th--i-T--ti--

Hypothesis Testing

Outline

Objectives

2 Random Variables

Definition Probability Distribution of RV Expected Value of a RV Other Properties of RVs

3 Sampling

Population vs Sample Random Sample

Objective

Random

Variable

Probability Distribut

Expected Value of a

Other Properties of I

Samping

Random Sample

Estimation and Inference

Point Estimation

Evaluation of Estima Hypothesis Testing

Thankyou

Outline

Objectives

Random Variables

Definition Probability Distribution of RV Expected Value of a RV Other Properties of RVs

3 Sampling Population vs Sample

Random Sample

Estimation and Inference

Point Estimation Evaluation of Estimator Hypothesis Testing

Outline

- Objectives
- 2 Random Variables

Definition
Probability Distribution of RV
Expected Value of a RV
Other Properties of RVs

3 Sampling

Population vs Sample Random Sample

4 Estimation and Inference

Point Estimation Evaluation of Estimator Hypothesis Testing

5 Thankyou

Sampl

Population vs Sam Random Sample

Estimation an Inference

Point Estimation Evaluation of Estimat

Hypothesis Testing

Objectives

Random

Definition

Probability Distributi of RV

Expected Value of a RV Other Properties of RVs

Sampling

Population vs Sam Random Sample

Estimation and

IIIICICIICC

Evaluation of Estima

--, ------

Objectives

 Review basic concepts of Random Variables, Sampling, Estimation and Inference

Objectives

Expected Value of a RV

Random Sample

• Review basic concepts of Random Variables, Sampling, Estimation and Inference

Objectives

Intiutively understand these concepts with examples

Objectives

Random Variable

Definition

Expected Value of a RV

Other Properties of RV

Sampling

Population vs Samp Random Sample

Estimation and

Inference
Point Estimation

Hypothesis Testing

--, [------

Objectives

- Review basic concepts of Random Variables, Sampling, Estimation and Inference
- Intiutively understand these concepts with examples
- Be able to use them in a practical problem setting

Objective:

Objectives

Definition

B 1 100 B 100 B

of RV

Expected Value of a RV

Other Properties of R

Sampling

Population vs Sample Random Sample

Estimation and

Point Estimation

Evaluation of Estima

riypotiiesis resting

• In a given probabilistic experiment, a random variable X is a

function from sample space to real line

Definition

Objective

Random

Definition

Probability Distribut

Expected Value of a

Other Properties of F

Sampling

Population vs Samp Random Sample

Estimation a

Inference

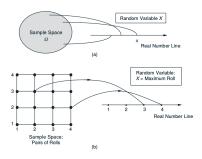
Evaluation of Estima

Hypothesis Testing

Thanky

Definition

• In a given probabilistic experiment, a random variable X is a function from sample space to real line



- Intuitively, we can think of rv as a bunch of values from real line
- It is often much easier to deal with summary variables than the actual outcomes

Objectives

Random

Definitio

Probability Distribution of RV

Expected Value of a RV

Sampling

Population vs Sample Random Sample

Estimation and

Point Estimation

Hypothesis Testing

--/ [------

Probability Distribution of RV

• It is description of all possible values, a RV can take along with their frequencies of occurance (probabilities)

Objectives

Objective

Definition

Probability Distribution of RV

Expected value of a r

Samplin

opulation vs Samp

Estimation an

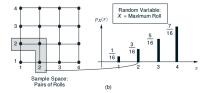
Point Estimation

Hypothesis Testing

...

Probability Distribution of RV

• It is description of all possible values, a RV can take along with their frequencies of occurance (probabilities)



• Other way of describing a distribution of RV is through its cumulative distribution function (cdf)

Objective

. .

Definition

Probability Distribution of RV

Expected value of t

Other Properties of I

Samplin

Random Sample

Estimation an

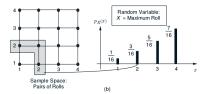
Inference

Evaluation of Estimat

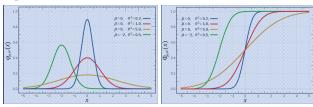
Thankwa

Probability Distribution of RV

• It is description of all possible values, a RV can take along with their frequencies of occurance (probabilities)



- Other way of describing a distribution of RV is through its cumulative distribution function (cdf)
- Example: Distribution of Normal RV



Objective

Random

Definition

Probability Distribution

Expected Value of a RV

Other Properties of P

Sampling

Population vs Sample Random Sample

Estimation and

Point Estimation

Evaluation of Estim

-->1-----

Expected Value of a RV

• It is merely average value of a RV, where average is taken after weighting values with their probabilities

Objective

Randon

Definition

Probability Distributi

Expected Value of a RV

Other Properties of R'

Samplin

Population vs San

mar a

Inference

Point Estimation

Evaluation of Estim

Expected Value of a RV

- It is merely average value of a RV, where average is taken after weighting values with their probabilities
- Can think of it as a typical or expected value of an observation of the random variable

$$E(X) = \begin{cases} \sum_{x \in X} x P(X = x), & \text{if X is discrete} \\ \int_{-\infty}^{\infty} x f(x) dx, & \text{if X is continuous} \end{cases}$$

Expected Value of a RV

Objective

Randon

Definition

Probability Distributi of RV

Expected Value of a RV Other Properties of RVs

Sampli

Population vs San Random Sample

Estimation :

Inference

Point Estimation

Hypothesis Testing

Thankyou

- It is merely average value of a RV, where average is taken after weighting values with their probabilities
- Can think of it as a typical or expected value of an observation of the random variable

$$E(X) = \begin{cases} \sum_{x \in X} x P(X = x), & \text{if X is discrete} \\ \int_{-\infty}^{\infty} x f(x) dx, & \text{if X is continuous} \end{cases}$$

- Expectation is a linear operation and hence following identities hold
 - $E(g(X)) = \sum_{x \in X} g(x)P(X = x)$
 - E(aX + bY + c) = aE(X) + bE(Y) + c

Objective

Random

Definitio

Probability Distributi of RV

Expected Value of a RV

Other Properties of RVs

Sampling

Population vs Sa

Random Sample

Estimation and

Interence

Evaluation of Estima

Hypothesis Testi

Thankyou

Properties of RVs

 Variance of a RV is measure of dispersion of its distribution and is given by

$$VarX = E[(X - \mu)^2],$$
 where μ is expected value of RV

Objective

,

Variable

Definition

of RV

Other Properties of RVs

Samplin

Population vs San Random Sample

Estimation ar

Inference

Point Estimation

Hypothesis Testin

Thankvou

Properties of RVs

• Variance of a RV is measure of dispersion of its distribution and is given by

$$VarX = E[(X - \mu)^2]$$
, where μ is expected value of RV

• If (X, Y) is a bivariate RV and have marginal pdfs f(x) and f(y), then they are independent if

$$f(x,y) = f(x)f(y)$$

Objective

,

Variable

Definition

of RV

Other Properties of RVs

Samplin

Population vs San Random Sample

Estimation ar

Inference

Point Estimation

Hypothesis Testin

Thankvou

Properties of RVs

• Variance of a RV is measure of dispersion of its distribution and is given by

$$VarX = E[(X - \mu)^2]$$
, where μ is expected value of RV

• If (X, Y) is a bivariate RV and have marginal pdfs f(x) and f(y), then they are independent if

$$f(x,y) = f(x)f(y)$$

Objective

Randon

Variab

Definition

Probability Distril

Expected Value of a

Other Properties of RVs

Samplin

Population vs Samp

Estimation a

Inference

Evaluation of Estimator Hypothesis Testing

Thankvou

Properties of RVs

 Variance of a RV is measure of dispersion of its distribution and is given by

$$VarX = E[(X - \mu)^2]$$
, where μ is expected value of RV

• If (X, Y) is a bivariate RV and have marginal pdfs f(x) and f(y), then they are independent if

$$f(x, y) = f(x)f(y)$$

- To quantify extent of linear relation between two RVs X and Y, we define
 - Covariance of X and Y as

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

Correlation of X and Y as

$$\rho_{xy} = Cov(X, Y) / \sqrt{Var(X)} \sqrt{Var(Y)}$$

Objectives

Variables
Definition

Probability Distribution of RV

Other Properties of R'

Samplin

Population vs Sample

Random Sample

Estimation and

Point Estimation

Hypothesis Testin

Population, Sample and Realization

- Data collected in an experiment consists of several observations of variables of interest
- For example, to understand the effect of education on future employment of individuals in a particular state
 - data consists of age, education, current employment and other factors of n individuals from the population of interest

Objectives

Definition
Probability Distribution

Expected Value of a R

Other Properties of R

Samplin

Population vs Sample Random Sample

Estimation as

Point Estimation

Evaluation of Estimate Hypothesis Testing

Thankyou

Population, Sample and Realization

- Data collected in an experiment consists of several observations of variables of interest
- For example, to understand the effect of education on future employment of individuals in a particular state
 - data consists of age, education, current employment and other factors of n individuals from the population of interest
- From statistical point of view, we have two sets of RVs in this experiment

Objectives

Variables
Definition
Probability Distribution RV

Expected Value of a Other Properties of F

Sampli

Population vs Sample Random Sample

Estimation a Inference

Point Estimation Evaluation of Estimato

Hypothesis Testing

Thankyou

Population, Sample and Realization

- Data collected in an experiment consists of several observations of variables of interest
- For example, to understand the effect of education on future employment of individuals in a particular state
 - data consists of age, education, current employment and other factors of n individuals from the population of interest
- From statistical point of view, we have two sets of RVs in this experiment
 - one corresponding to the whole population called population RVs
 - other corresponds to the sample, sample RVs, because we could have chosen any set of n individuals from population (so sample is itself random)
- After the observation is made, it is called 'realization' of the corresponding RV

Objective

Random

Definition

Probability Distributi

Expected Value of a

Other Properties of I

Samping

Population vs Sar

Random Sample

Estimation and

Inference

Evaluation of Estima

Hypothesis Testi

Thankvou

Random Sample

- Consider observing age of individuals in a population:
 - these sample observations $X_1, X_2...X_n$ of age are from the same population RV X
 - also each observation is independent of the other

Random Sample

Objectiv

Random Variable

Definition

Probability Distribu of RV

Other Properties of I

Samplin

Population vs Sam Random Sample

Estimation a

Inference

Foint Estimation

Evaluation of Estimat

Hypothesis Testing

Tl. . . . l

- Consider observing age of individuals in a population:
 - these sample observations X₁, X₂...X_n of age are from the same population RV X
 - also each observation is independent of the other
- This is an example of Random sample
- Note: X₁, X₂...X_n are RVs (sample RVs) because we could have easily realized different set of n individuals if experiment was conducted again!
- Note: They are all having same distribution as X (population RV)

Random Sample

Objective

Variable

Definition

Probability Distribu of RV

Other Properties of I

Sampli

Population vs Sam Random Sample

Estimation a

Inference

Point Estimation

Hypothesis Testing

Thankyo

- Consider observing age of individuals in a population:
 - these sample observations X₁, X₂...X_n of age are from the same population RV X
 - also each observation is independent of the other
- This is an example of Random sample
- Note: X₁, X₂...X_n are RVs (sample RVs) because we could have easily realized different set of n individuals if experiment was conducted again!
- Note: They are all having same distribution as X (population RV)
- Alternatively, $X_1, X_2...X_n$ are called independent and identically distributed RVs

Objectives

Random

Definitio

Probability Distribution of RV

Expected Value of a RV

Other Properties of RV

Sampling

Population vs Sample Random Sample

Estimation and

Point Estimation

Evaluation of Estimate

Point Estimation

• Why do we need to estimate population parameter?

Objectiv

Random

Definition

Probability Distribut of RV

Other Properties of RV

Sampling

Population vs Samp

Estimation and

Point Estimation

Evaluation of Estimate

Thankyou

Point Estimation

- Why do we need to estimate population parameter?
 - The population of interest is completely defined by the probability density function $f(x|\theta)$
 - Thus parameter θ yields knowledge of the entire population
 - Also there might be important physical interpretation for θ (for example population mean)

Objective

Random

Definition

Probability Distribu of RV

Expected Value of a R

Other Properties of RV

Commilian

Population vs Samr

Random Sample

Estimation an Inference

Point Estimation

Hypothesis Testing

Point Estimation

- Why do we need to estimate population parameter?
 - The population of interest is completely defined by the probability density function $f(x|\theta)$
 - Thus parameter θ yields knowledge of the entire population
 - Also there might be important physical interpretation for θ (for example population mean)
- Estimator vs Estimate
 - Estimator of a population parameter is some function of the sample,

$$\theta = W(X_1, X_2...X_n)$$
, where $X_1, X_2...X_n$ are sample RVs

Objectiv

Random

D. G. isi.

Probability Distribution of RV

Expected Value of a R

Other Properties of R

Sampli

Population vs Samp

Estimation ar Inference

Point Estimation

Evaluation of Estimal

Hypothesis Testing

Thankyo

Point Estimation

- Why do we need to estimate population parameter?
 - The population of interest is completely defined by the probability density function $f(x|\theta)$
 - Thus parameter θ yields knowledge of the entire population
 - Also there might be important physical interpretation for θ (for example population mean)
- Estimator vs Estimate
 - Estimator of a population parameter is some function of the sample,

$$\theta = W(X_1, X_2...X_n)$$
, where $X_1, X_2...X_n$ are sample RVs

• So θ is a Random Variable in itself

Objective

Dandom

v air 100

Probability Distri

Expected Value of a R

Sampling

Population vs Samp

Estimation ar Inference

Point Estimation

Hypothesis Testing

Thankyo

Point Estimation

- Why do we need to estimate population parameter?
 - The population of interest is completely defined by the probability density function $f(x|\theta)$
 - Thus parameter θ yields knowledge of the entire population
 - Also there might be important physical interpretation for θ (for example population mean)
- Estimator vs Estimate
 - Estimator of a population parameter is some function of the sample,

$$\theta = W(X_1, X_2...X_n)$$
, where $X_1, X_2...X_n$ are sample RVs

- So θ is a Random Variable in itself
- Estimate is the realized value of an estimator,

$$\theta = W(x_1, x_2...x_n)$$
, where $x_1, x_2...x_n$ are realized values

Objective

Random

Definition

Probability Distribut

Expected Value of a RV

Sampling

Population vs Samn

Random Sample

Estimation an

Inference Point Estimation

Evaluation of Estimat

Hypothesis Testii

Point Estimation

• Sample mean can be used as an esimator of population mean,

$$\overline{X} = (X_1 + X_2 + X_3 ... + X_n)/n$$

• Sample variance is a good estimator of population variance,

$$S^{2} = \sum_{i=1}^{n} (X_{i} - \overline{X})^{2} / (n-1)$$

Objective

Random

Definition

Probability Distribution RV

Expected Value of a I Other Properties of R

Samplin

Population vs Samp

Estimation .

Inference

Point Estimation

Hypothesis Testing

701 1

Thankyo

Point Estimation

• Sample mean can be used as an esimator of population mean,

$$\overline{X} = (X_1 + X_2 + X_3 ... + X_n)/n$$

• Sample variance is a good estimator of population variance,

$$S^{2} = \sum_{i=1}^{n} (X_{i} - \overline{X})^{2} / (n-1)$$

- When we are using a qualifier 'good', we have to be more precise,
 - this leads us to three properties of estimators unbiasedness, efficiency and consistency

Objectiv

Randon

variabi

Probability Distribu

Expected Value of a

Other Properties of I

Samplii

Population vs Sam Random Sample

Estimation Inference

_

Evaluation of Estimator

Hypothesis Testing

Thankyou

Unbiasedness and Efficiency

- Mean Square Error (MSE) to evaluate an estimator:
 - MSE of an estimator W of a parameter θ is a function of θ defined as $E_{\theta}(W \theta)^2$,

$$E_{\theta}(W - \theta)^2 = \underbrace{Var_{\theta}W}_{\text{Variance}} + \underbrace{(E_{\theta}W - \theta)^2}_{\text{Bias}}$$

- MSE incorporates to components, one measuring the variability of estimator (precision) and the other its bias (accuracy)
- An estimator is 'Unbiased' if its bias term is zero
- For example, population mean estimator \overline{X} is unbiased since $E\overline{X} = \mu_x$
- An estimator is said to be 'Efficient' if its variance component is least among all the estimators with same bias

Objective

Randon

Definition

Probability Distributi of RV

Expected Value of a RV

Other Properties of RV

Sampling

Population vs Sample

Estimation and

Point Estimation

Evaluation of Estimator

Thankvou

Unbiasedness and Efficiency

- Evaluating an estimator based on MSE is a trade-off between Variance and bias components
- An estimator which is biased could be a better estimator due to less variance

Objective

Randon

Definition

Probability Distribut

Expected Value of a RV

Samplin

Population vs Sam

Estimation a

Interence

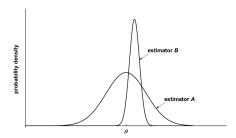
Evaluation of Estimator

Hypothesis Testing

Thankyou

Unbiasedness and Efficiency

- Evaluating an estimator based on MSE is a trade-off between Variance and bias components
- An estimator which is biased could be a better estimator due to less variance



Unbiasedness and efficiency are finite sample properties of estimators

Objective

Random

P. C. W.

Definitio

of RV

Expected Value of a RV

G 1'

Population vs Samn

Random Sample

Estimation and Inference

Point Estimation

Evaluation of Estimator Hypothesis Testing

Thankyou

Consistency of an Estimator

 Asymptotic properties like consistency are used to evaluate an estimator as the sample size tends to infinity

Objective

Random

variable

Probability Dietribu

Expected Value of a

Other Properties of

Sampling

Population vs San

Estimation a

Inference

Evaluation of Estimator

Hypothesis Testing

Thankvou

Consistency of an Estimator

- Asymptotic properties like consistency are used to evaluate an estimator as the sample size tends to infinity
- Probability limits (Convergence in probability)
 - A sequence of random variables, $X_1, X_2...$, converges in probability to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \ge \epsilon) = 0$$

• Intuitively, limit implies that for large n 'bulk 'of the distribution of X_n is concentrated near X

Objective

Variable

D.C.

Probability Distributi

Expected Value of a

Sampling

Population vs Sam

Estimation a

IIIICICIICC

Evaluation of Estimator

Hypothesis Testing

Thankyo

Consistency of an Estimator

- Asymptotic properties like consistency are used to evaluate an estimator as the sample size tends to infinity
- Probability limits (Convergence in probability)
 - A sequence of random variables, $X_1, X_2...$, converges in probability to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \ge \epsilon) = 0$$

- Intuitively, limit implies that for large n 'bulk 'of the distribution of X_n is concentrated near X
- An estimator is said to be consistent if it satisfies these two properties:
 - the estimator possesses probability limit
 - and the limit is the true value of the population parameter

Objectiv

Randon

Definitio

Probability Distribut

Expected Value of a RV

Other Properties of R'

Sampling

Population vs Sample

Estimation and

Point Estimation

Evaluation of Estimator

Hypothesis Testing

-->1------

Consistency of an Estimator

 For example, Sample mean is a consistent estimator of the population mean

$$\lim_{n\to\infty} P(|\overline{X} - \mu_x| \ge \epsilon) = 0$$

Objective:

Objective

Definitio

Probability Distribut of RV

Expected Value of a RV
Other Properties of RVs

Samplin

Population vs Sampl

Estimation an Inference

Interence

Evaluation of Estimator

Thankyou

Consistency of an Estimator

 For example, Sample mean is a consistent estimator of the population mean

$$\lim_{n\to\infty} P(|\overline{X}-\mu_x|\geq \epsilon)=0$$

- Unbiasedness is sufficient but not necessary condition for consistency
- An estimator may become unbiased as the sample size increases:

Objective

. .

Definition

Probability Distribu

Expected Value of a F

Other Properties of R'

Samplin

opulation vs Samp

Estimation an

Inference

Evaluation of Estimator

Hypothesis Testing

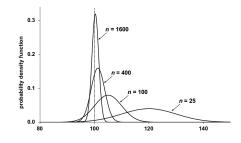
Thankvo

Consistency of an Estimator

 For example, Sample mean is a consistent estimator of the population mean

$$\lim_{n\to\infty} P(|\overline{X}-\mu_x|\geq \epsilon)=0$$

- Unbiasedness is sufficient but not necessary condition for consistency
- An estimator may become unbiased as the sample size increases:



Hypothesis Testing

It is yet another method of inference like point estimation

A hypothesis is a statement about the population parameter

- Goal of a hypothesis test is to decide, based on sample from the population, which of two complementary hypothesis (Null and Alternative) is true
- Hypothesis testing procedure is a rule that species for which values of sample do we accept one hypothesis over the other
- Some of the popular Hypothesis testing procedures include LRT, Bayesian tests etc

Hypothesis Testing

Hypothesis Testing - Example

Definition

Probability Distribution of RV Expected Value of a RV

Other Properties of R

Sampling

Random Sample

Estimation and Inference

Point Estimation

Hypothesis Testing

.,,,.....

Dougherty Slides from Study Guide

Evaluating Hypothesis tests

Dougherty Slides from Study Guide

Objectives

Variable

Definition

of RV

Expected Value of a RV

Other Properties of P

Other Properties of F

Samping

Random Sample

Estimation and Inference

Point Estimation

Evaluation of Estimato

Hypothesis Testing

I hankyou

Hypothesis tests - Confidence Interval approach

Objective

Kandon Variable

Definition

of RV Expected Value of a RV

Other Properties of R

Samplin

Population vs Samp

Random Sample

Estimation and

Inference

Evaluation of Estimat

Hypothesis Testing

Dougherty Slides from Study Guide

Objective

Random

Definition

of RV

Expected Value of a RV

Other Properties of R

Sampling

Population vs Sample Random Sample

Estimation and

Inference

Evaluation of Estima

Thankyou

"In God we trust, all others bring data."

William Edwards Deming (1900 - 1993).