innovate achieve lead

BITS Pilani
Pilani Campus

Course No: MATH F113

Probability and Statistics

Sec 5.3: Statistics and Their Distributions

Sumanta Pasari

sumanta.pasari@pilani.bits-pilani.ac.in

BITS Pilani
Pilani Campus

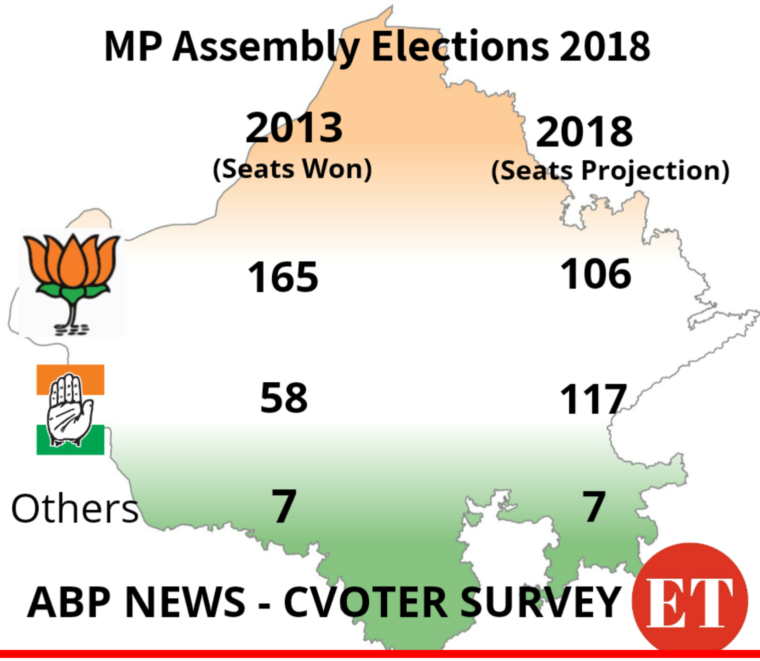
Learning Objectives

- Concept of a statistic(s)
- Difference between statistic(s) and Statistics
- Concept of a random sample
- Distribution of sample statistic
- Concept of deriving sampling distribution
- In particular, distribution of the sample mean
- Concept of CLT (Central Limit Theorem)

What is Statistics?

- A discipline of science that pertains to the **collection**, **analysis**, **interpretation**, and **presentation** of data
- Science of ***learning from data***
- Mathematical Statistics: application of Mathematics to Statistics
- In 18th century, “Statistics” designated to a systematic collection of demographic and economic data by states.
- Main concern is to collect, analyze, and present data in the context of **uncertainty** and **decision making**

Example 1: Election Forecast



Parties and coalitions	Popular vote			Seats	
	Votes	%	±pp	Won	+/−
INC +	15,595,153	40.9%	▲4.59%	114	▲56
BJP	15,642,980	41%	▼3.88%	109	▼56

Example 2: Student Information

	Age	Gender	Height (cm)	Weight (kg)	Nose length (mm)
1	20	M	165	62	40
2	21	F	152	56	44
3	20	F	158	62	42
4	21	F	**191**	54	40
5	19	M	167	65	41
6	20	F	159	60	43
7	18	M	190	101	47
8	19	M	182	95	46
9	20	M	170	81	44
10	21	M	172	74	41
11	22	F	170	55	42
12	19	M	178	75	45
13	20	F	157	55	40
14	21	M	169	70	48
15	20	M	164	63	45
16	19	M	174	67	44
17	18	F	154	56	**72**
18	21	F	156	58	47
19	19	M	171	59	42
20	19	F	151	**102**	45

Observe and Answer:

1. ***Any outliers present in this raw data?***
2. Any relation between height and weight of students?
3. Do the female students have more nose-length?
4. What is the average height of male students?
5. *Do the male students come from a rich family?*
6. *Are the female students more intelligent than male students?*

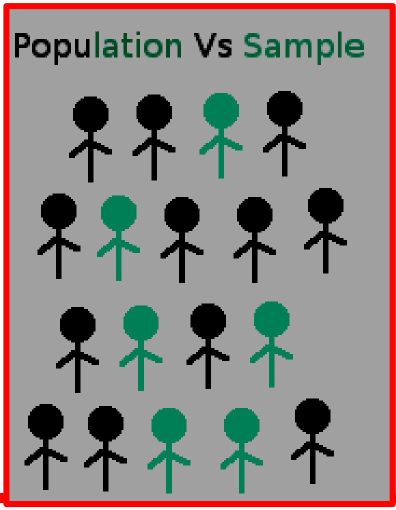
Descriptive and Inferential Statistics

1. **Descriptive Statistics** consists of the collection, organization, summarization, and presentation of data.
   - process of describing data and trying to reach a conclusion
   - data and charts observed in general newspapers or articles

2. **Inferential Statistics** provides a scientific procedure to make inferences about a population based on sample.
   - generalizing from samples to populations, performing estimations and hypothesis testing, determining relationships among variables, and making predictions
   - ages of students of a class are 25, 24, 28, 29, 30, 25, 26, 25, 28, 25, and 25 years. Do the students (overall) follow a normal distribution?

Population and Sample

- **Population** is a collection of all distinct individuals or objects or items under study; Size of population: ***N***
- **Sample** is a portion (representative portion) of a population; Size of a sample: ***n***
- **Sampling**: How to choose a sample? What are the desired criteria?
- To represent the population well, a sample should be **randomly** collected and adequately **large**.

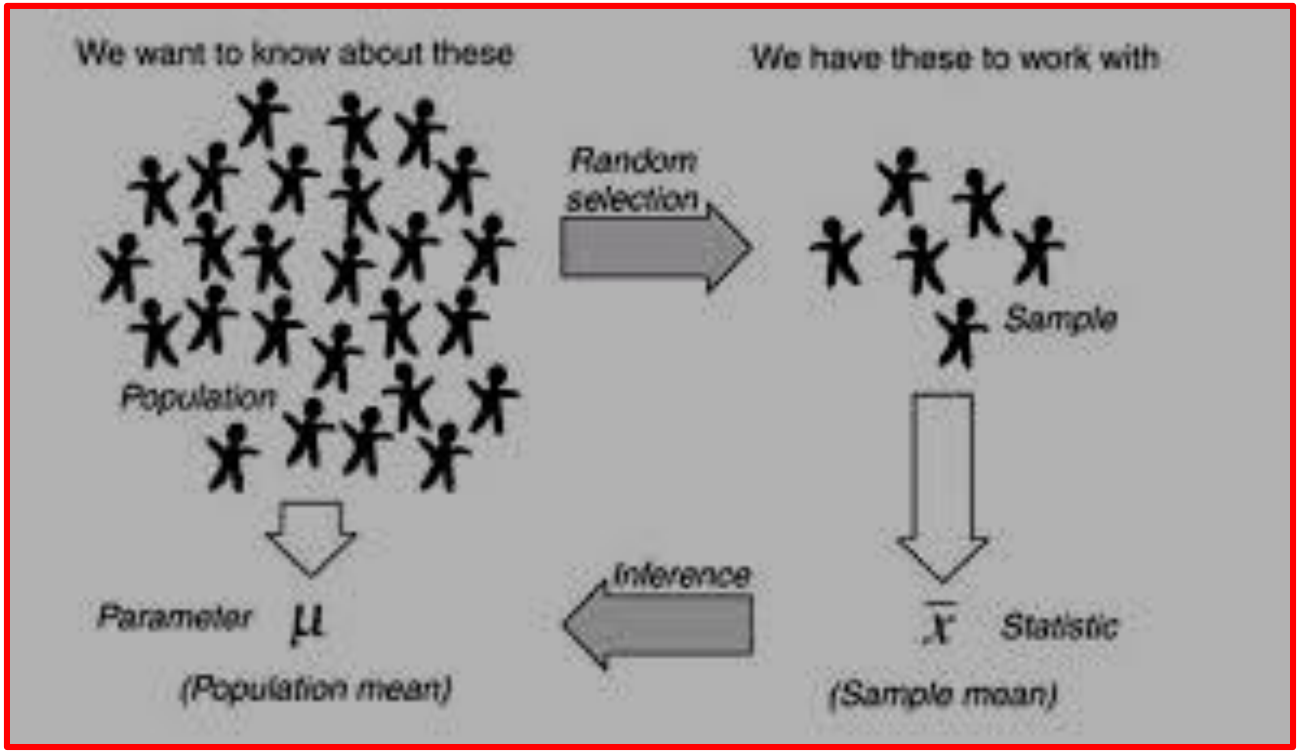


Ex: Election forecast, lifetime of tube light,

efficiency of a drug, campus placements

Parameter and Statistic

- **Parameter** is a descriptive measure of some characteristics of the population. Parameter is usually unknown.
- **Statistic** is a descriptive measure obtained from a sample; It is a function of observations in a random sample.



Random Sample: Example

- Suppose there are 2500 managers in a company. We would like to develop a profile of managers, with (a) their mean annual salary, (b) proportion of managers who completed company's management training program. …..suppose the population mean salary is known as ***μ = $51,800*** and ***population s.d. σ = $4000,*** and 1500 managers have already completed the training program, i.e., ***population proportion p =0.6.***

Random Number Table

10097	32533	76520	13586	34673	54876	80959	09117	39292	74945
37542	04805	64894	74296	24805	24037	20636	10402	00822	91665
08422	68953	19645	09303	23209	02560	15953	34764	35080	33606
99019	02529	09376	70715	38311	31165	88676	74397	04436	27659
12807	99970	80157	36147	64032	36653	98951	16877	12171	76833
66065	74717	34072	76850	36697	36170	65813	39885	11199	29170
31060	10805	45571	82406	35303	42614	86799	07439	23403	09732
85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
73796	45753	03529	64778	35808	34282	60935	20344	35273	88435
98520	17767	14905	68607	22109	40558	60970	93433	50500	73998
11805	05431	39808	27732	50725	68248	29405	24201	52775	67851
83452	99634	06288	98083	13746	70078	18475	40610	68711	77817
88685	40200	86507	58401	36766	67951	90364	76493	29609	11062
99594	67348	87517	64969	91826	08928	93785	61368	23478	34113
65481	17674	17468	50950	58047	76974	73039	57186	40218	16544
80124	35635	17727	08015	45318	22374	21115	78253	14385	53763
74350	99817	77402	77214	43236	00210	45521	64237	96286	02655
69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
09893	20505	14225	68514	46427	56788	96297	78822	54382	14598

Random sample

- Recall previous example: once 30 managers are randomly selected, we calculate **sample statistic (?)**.

Salary	Training?	Salary	Training?	Salary	Training?
49094.3	Yes	45922.6	Yes	45120.9	Yes
53263.9	Yes	57268.4	No	51753.0	Yes
49643.5	Yes	55688.8	Yes	54391.8	No
49894.9	Yes	51564.7	No	50164.2	No
47621.6	No	56188.2	No	52973.6	No
55924.0	Yes	51766.0	Yes	50241.3	No
49092.3	Yes	52541.3	No	52793.6	No
51404.4	Yes	44980.0	Yes	50979.4	Yes
50957.7	Yes	51932.6	Yes	55860.9	Yes
55109.7	Yes	52973.0	Yes	57309.1	No

Sample Statistic

innovate achieve lead

- Recall previous example: once 30 managers are randomly selected, we calculate **sample statistic (?)**.

Salary	Training?	Salary	Training?	Salary	Training?
49094.3	Yes	45922.6	Yes	45120.9	Yes
53263.9	Yes	57268.4	No	51753.0	Yes
49643.5	Yes	55688.8	Yes	54391.8	No
49894.9	Yes	51564.7	No	50164.2	No
47621.6	No	56188.2	No	52973.6	No
55924.0	Yes	51766.0	Yes	50241.3	No
49092.3	Yes	52541.3	No	52793.6	No
51404.4	Yes	44980.0	Yes	50979.4	Yes
50957.7	Yes	51932.6	Yes	55860.9	Yes
55109.7	Yes	52973.0	Yes	57309.1	No

From sample,

$$\bar{x} = \$51,814$$

$$s = \$3,348$$

$$\bar{p} = \frac{19}{30} = 0.63$$

Actually,

$$\mu = \$51800$$

$$\sigma = \$4000$$

$$p = 0.60$$

Sampling Distributions

- Suppose we collect another random sample of size 30, and we found sample mean = \$ 52670, and sample proportion = 0.70
- Let us repeat this "**random experiment**" 500 times – random variables come into picture.

Sample Number	Sample mean	Sample proportion
001	51814	0.63
002	52670	0.70
003	51780	0.67
004	51588	0.53
…	…	…
…	…	…
500	51752	0.50

Each $X_1, X_2, \cdots, X_n$ is a random variable (and, X_i are i.i.d), with mean μ and standard deviation σ.

So, how does $\bar{X}$ behave?

What is the distribution of $\bar{X}$?

What are the mean and s.d. of $\bar{X}$?

Random Sample

- The random variables $\boldsymbol{X_i}$ constitute **a random sample of size n** if and only if,

    1) Random variables $\boldsymbol{X_i}$ are independent, and
    2) Random variables $\boldsymbol{X_i}$ are identically distributed, that is, each $\boldsymbol{X_i}$ has same distribution having pdf $\boldsymbol{f(x)}$, mean $\boldsymbol{\mu}$, and variance $\boldsymbol{\sigma^2}$. **We say that X_i are i.i.d.**

- **Examples?**

Sample Statistic

- A **statistic** of a random sample *(X_1, ..., X_n)* from population *X* is a function *H (X_1, ..., X_n)* of the n-dim r.v. *(X_1, ..., X_n)*. For example,

(i) Sample mean, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

(ii) Sample variance,

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n\left(n-1\right)}$$

Other Statistics

(iii) Sample minimum $X^{(1)} = \underset{i}{Min}\{X_i\}$

(iv) Sample maximum $X^{(n)} = \underset{i}{Max}\{X_i\}$

(v) Sample range $= X^{(n)} - X^{(1)} = \underset{i}{Max}\{X_i\} - \underset{i}{Min}\{X_i\}$

(vi) Sample median $= \begin{cases} X^{\left(\frac{n+1}{2}\right)}; \; n \text{ odd} \\ \dfrac{X^{\left(\frac{n}{2}\right)} + X^{\left(\frac{n+1}{2}\right)}}{2}; \; n \text{ even} \end{cases}$

Order Statistic: $X^{(1)}, X^{(2)}, \cdots, X^{(n)}$ is called an ordered statistic.

Random Sample: Finite Population

- Suppose there are 2500 managers in a company. We would like to develop a profile of managers, with (a) their mean annual salary, (b) proportion of managers who completed company's management training program.
- Suppose there are 2000 oak trees in a managed small forest. We need to estimate tree diameter at breast height.
- How to choose *n* samples in each case? **Random number table or any other computer programs?**

Random Sample: Infinite Population

- Consider the population of 8 million school students in a certain state. Suppose, we are interested to determine μ, the unknown population mean distance from students' schools to their hometowns. Suppose, the sample mean from 100 students show a distance of 1.2 km.

- Suppose a tyre manufacturer is interested to test whether the new design provides increasing mileage. To estimate the mean useful life, the manufacturer choose a sample of 120 tires. Test result shows a sample mean of 36000 miles.

Other Examples

- Suppose you would like to select a random sample of 50 customers in a restaurant to complete a feedback survey – how to go ahead?
- In order to improve the facilities of an amusement park (e.g., Nicco Park, Kolkata), the manager wants to collect a sample of 50 persons.
- A quality control manager is concerned about the proper machine-filling of breakfast boxes, filled with 100 gm of cereals.
- **Discuss, in each of the above examples, how to avoid selection bias.**
- Infinite populations: customers entering a retail store, repeated experimental trials in a laboratory, number of trees in a forest, telephone calls arriving at a technical support centre
- **For practical purposes, if $n/N \leq 0.05$ (at most 5% of the population is sampled), one may consider infinite population.**

Remarks

(1) Here the sample of size *n* can be thought as ordered and with repetition allowed, i.e., an *n*-tuple $(x_1,\ldots,x_n)$ represents an observed sample.

(2) Let r.v. *X* with density $f_X(x)$ denote the population. The ***random sample of size n from (infinite) population X*** can be thought as the *n*-dimensional random variable $(X_1, \ldots, X_n)$ whose joint density is

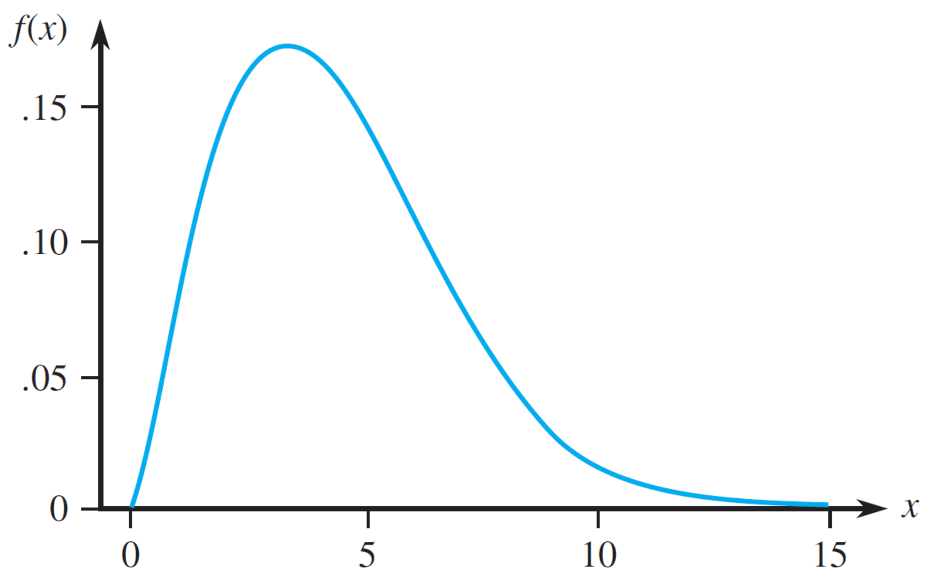
$$f_{X_1\ldots X_n}(x_1,\ldots,x_n) = f_X(x_1)\ldots f_X(x_n)$$

Remarks

(3) The random sample and its characteristics will be denoted by capital letters and a particular sample and its characteristics by corresponding small letters. This is in tune with the practice of denoting random variables by capital letters and its values by corresponding small letters. Thus a particular sample is a value $(x_1, \ldots, x_n)$ of the n-dimensional r.v. $(X_1, \ldots, X_n)$, or equivalently, a point in n-dimensional space. The values $x_1, \ldots, x_n$ are also called observed values of X.

Another Example (from textbook)

- Suppose that material strength for a randomly selected specimen of a particular type has a Weibull distribution with parameter values $\alpha = 2$ (shape) and $\beta = 5$ (scale). The corresponding density curve is shown in below figure.



The Weibull density curve

Another Example (contd...)

cont'd

Using the formulas,

- $E(X) = e^{\mu+\sigma^2/2}$ and $V(X) = e^{2\mu+\sigma^2} \cdot (e^{\sigma^2} - 1)$
- where $\mu = \beta\Gamma\left(1 + \frac{1}{\alpha}\right)$ and $\sigma^2 = \beta^2\left\{\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2\right.$

$\mu = E(x) = 4.4311$ and $\sigma^2 = V(X) = 5.365$, $\sigma = 2.316$

- We used statistical software to generate six different samples, each with n = 10, from this distribution (material strengths for six different groups of ten specimens each).

Another Example (contd...)

cont'd

- The results appear in the table below.

Sample	1	2	3	4	5	6
1	6.1171	5.07611	3.46710	1.55601	3.12372	8.93795
2	4.1600	6.79279	2.71938	4.56941	6.09685	3.92487
3	3.1950	4.43259	5.88129	4.79870	3.41181	8.76202
4	0.6694	8.55752	5.14915	2.49759	1.65409	7.05569
5	1.8552	6.82487	4.99635	2.33267	2.29512	2.30932
6	5.2316	7.39958	5.86887	4.01295	2.12583	5.94195
7	2.7609	2.14755	6.05918	9.08845	3.20938	6.74166
8	10.2185	8.50628	1.80119	3.25728	3.23209	1.75468
9	5.2438	5.49510	4.21994	3.70132	6.84426	4.91827
10	4.5590	4.04525	2.12934	5.50134	4.20694	7.26081
$\bar{x}$	4.401	5.928	4.229	4.132	3.620	5.761
$\tilde{x}$	4.360	6.144	4.608	3.857	3.221	6.342
s	2.642	2.062	1.611	2.124	1.678	2.496

Samples from the Weibull Distribution

Another Example (contd…)

cont'd

- Followed by the values of the sample mean, sample median, and sample standard deviation for each sample. Notice first that the ten observations in any particular sample are all different from those in any other sample.

- Second, the six values of the sample mean are all different from one another, as are the six values of the sample median and the six values of the sample standard deviation.

Another Example (contd...)

cont'd

- Furthermore, the value of the sample mean from any particular sample can be regarded as a ***point estimate*** ("point" because it is a single number, corresponding to a single point on the number line) of the population mean μ, whose value is known to be 4.4311.

- None of the estimates from these six samples is identical to what is being estimated.

- The estimates from the second and sixth samples are much too large, whereas the fifth sample gives a substantial underestimate.

Another Example (contd…)

cont'd

- Similarly, the sample standard deviation gives a point estimate of the population standard deviation. All six of the resulting estimates are in error by at least a small amount.

- In summary, the values of the individual sample observations vary from sample to sample, so will in general the value of any quantity computed from sample data, and the value of a sample characteristic used as an estimate of the corresponding population characteristic will virtually never coincide with what is being estimated.

What is an outlier?

- What is an outlier?
- Why does it occur?
- Is it really an outlier?
- How to detect?
- What to do then?

(Out of present syllabus)

Let’s Watch…

https://www.youtube.com/watch?v=jihVhB_TXeU

Some Properties

If $X_1, X_2, \cdots, X_n$ is a random sample (that is, X_i are i.i.d), with mean μ and standard deviation σ, then

(i) $E\left(X_1 + X_2 + \cdots + X_n\right) = \sum_{i=1}^{n} E\left(X_i\right) = n\mu$ (true, without independent)

(ii) $\mathrm{Var}\left(X_1 + X_2 + \cdots + X_n\right) = \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) = n\sigma^2$ $\left(\text{as, } \mathrm{cov}\left(X_i, X_j\right) = 0, i \neq j\right)$

(iii) $E\left(X_1 X_2 \cdots X_n\right) = \left[E\left(X_i\right)\right]^n = \mu^n$

(iv) Joint pdf $f_{X_1 X_2 \cdots X_n}\left(x_1, x_2, \cdots x_n\right) = f_{X_1}\left(x_1\right) f_{X_2}\left(x_2\right) \cdots f_{X_n}\left(x_n\right)$

(v) Joint cdf $F_{X_1 X_2 \cdots X_n}\left(x_1, x_2, \cdots x_n\right) = F_{X_1}\left(x_1\right) F_{X_2}\left(x_2\right) \cdots F_{X_n}\left(x_n\right)$

So, how does $\bar{X}$ behave? What is the distribution of $\bar{X}$?

What are the mean and standard deviation of $\bar{X}$?

Mean and Variance of Sample Mean

Ex.1. If $X_1, X_2, \cdots, X_n$ is a random sample, each X_i having mean μ and standard deviation σ, then

$$\text{(i) } \mu_{\bar{X}} = E\left(\bar{X}\right) = \mu \qquad \text{(ii) } \sigma_{\bar{X}}^2 = \text{Var}\left(\bar{X}\right) = \frac{\sigma^2}{n}$$

Proof. (i) $E\left(\bar{X}\right) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \frac{n\mu}{n} = \mu$

$$\text{(ii) Var}\left(\bar{X}\right) = \text{Var}\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) \quad \left(\text{as, Var}\left(aX\right) = a^2\text{Var}\left(X\right)\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}\left(X_i\right) \quad \left(\text{as, } X_1, X_2, \cdots, X_n \text{ are independent}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Comments

- From this theorem, it follows that larger the sample size, sample mean can be expected to lie closer to the population mean.
- Thus choosing large sample makes estimation more reliable.
- But how large is large? What should be desired sample size to carry out inferential statistics?

Sample Variance

Recall sample variance,

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1} = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

Then,

$$E\left(S^2\right) = \sigma_X^2$$

Proof: to be discussed in class!!

Problem Solving

Ex.2. Let $X_1, X_2, \cdots, X_{25}$ be a random sample from the distribution of X having mean 10 and variance 50. Find the mean and standard deviation of (i) $a\bar{X}+5$ (ii) $7\bar{X}+5a$, a is a scalar.

Sol.

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i=\frac{1}{25}\sum_{i=1}^{25}X_i,$$

$$E\left(\bar{X}\right)=\mu=10$$

$$\operatorname{Var}\left(\bar{X}\right)=\frac{\sigma^2}{n}=\frac{50}{25}=2$$

(i) $\mu_{a\bar{X}+5}=E\left(a\bar{X}+5\right)=aE\left(\bar{X}\right)+5=10a+5$

$$\sigma^2_{a\bar{X}+5}=\operatorname{Var}\left(a\bar{X}+5\right)=a^2\operatorname{Var}\left(\bar{X}\right)=2a^2\Rightarrow\sigma_{a\bar{X}+5}=|a|\sqrt{2}$$

Problem Solving

HW.1. Let the mean and variance of a sample mean are 5 and 2, respectively. If the random sample $X_1, X_2, \cdots, X_n$ comes from a distribution of X having variance 10, find the mean of X_i and the sample size *n*.

HW.2. Let the mean and standard deviation of a sample mean are 10 and 2, respectively. If the random sample $X_1, X_2, \cdots, X_n$ comes from a distribution of X having variance 60, find the sample size *n*. (**Sol.** $n = 15$?)

Problem Solving

HW3. Let $X_1, X_2, \cdots, X_n$ be a random sample from the distribution of X having mean 10 and variance 50. Find the minimum number of sample size n, such that variance of $3\bar{X}$ becomes less than 5.

HW4. Let $X_1, X_2, \cdots, X_n$ be a random sample from the distribution of X having mean 10 and variance 50. Find the minimum number of sample size n, such that the standard deviation of $8\bar{X}$ becomes less than 10.

Problem Solving

innovate achieve lead

HW5. A random sample of size 5 provides the following observations on X (height of students, in cm) and Y (weight of students, in kg).

X_{obs}: 185 175 165 170 156 Y_{obs}: 69 67 61 64 56

(a) Calculate sample means and sample standard deviations of height and weight. Can we compare the standard deviations of height and weight?

(b) Is there a (linear) relationship between observed height and weight data? (wait, sample covariance/correlation will be discussed later!)

HW6. Analyze the datset in Example 2. Find observed values of sample mean, sample variance, sample maximum, sample minimum, sample median, sample range for height, weight, and nose lengths.

Problem Solving

HW7. A random sample of size 9 yields the following observations on the random variable X, the coal consumption in millions of tons by electric utilities for a given year:

406 395 400 450 390 410 415 401 408

Find the observed value of the sample mean, sample median, and sample standard deviation.

Ex. 37 from textbook (page 222)

A particular brand of dishwasher soap is sold in three sizes: 25 oz, 40 oz, and 65 oz. Twenty percent of all purchasers select a 25-oz box, 50% select a 40-oz box, and the remaining 30% choose a 65-oz box. Let X_1 and X_2 denote the package sizes selected by two independently selected purchasers.

a. Determine the sampling distribution of $\bar{X}$, calculate $E(\bar{X})$, and compare to μ.

b. Calculate $P(\bar{X} \leq 42)$.

c. Determine the sampling distribution of the sample variance S^2, calculate $E(S^2)$, and compare to σ^2.

d. Obtain the distribution of sample range R.

Solution of Ex. 37

The joint pmf of X_1 and X_2 is presented below. Each joint probability is calculated using the independence of X_1 and X_2; e.g., $p(25, 25) = P(X_1 = 25) \cdot P(X_2 = 25) = (.2)(.2) = .04$.

	$p(x_1, x_2)$	x_1 = 25	40	65	
x_2	25	.04	.10	.06	.2
	40	.10	.25	.15	.5
	65	.06	.15	.09	.3
		.2	.5	.3	

a. For each coordinate in the table above, calculate $\bar{x}$. The six possible resulting $\bar{x}$ values and their corresponding probabilities appear in the accompanying pmf table.

$\bar{x}$	25	32.5	40	45	52.5	65
$p(\bar{x})$	.04	.20	.25	.12	.30	.09

From the table, $E(\bar{X}) = (25)(.04) + 32.5(.20) + \ldots + 65(.09) = 44.5$. From the original pmf, $\mu = 25(.2) + 40(.5) + 65(.3) = 44.5$. So, $E(\bar{X}) = \mu$.

b. For each coordinate in the joint pmf table above, calculate $s^2 = \frac{1}{2-1}\sum_{i=1}^{2}(x_i - \bar{x})^2$. The four possible resulting s^2 values and their corresponding probabilities appear in the accompanying pmf table.

s^2	0	112.5	312.5	800
$p(s^2)$	.38	.20	.30	.12

From the table, $E(S^2) = 0(.38) + \ldots + 800(.12) = 212.25$. From the original pmf, $\sigma^2 = (25 - 44.5)^2(.2) + (40 - 44.5)^2(.5) + (65 - 44.5)^2(.3) = 212.25$. So, $E(S^2) = \sigma^2$.

How to Derive a Sampling Distribution?

- Probability rules can be used to obtain the distribution of a statistic provided that it is a "simple (nice?)" function of the X_i's and either there are relatively few different X values in the population or else the population distribution has a "nice" form.

- Let us consider an example.

Example 3

- A certain brand of MP3 player comes in three configurations: a model with 2 GB of memory, costing \$80, a 4 GB model priced at \$100, and an 8 GB version with a price tag of \$120.

- If 20% of all purchasers choose the 2 GB model, 30% choose the 4 GB model, and 50% choose the 8 GB model, then the probability distribution of the cost *X* of a single randomly selected MP3 player purchase is given by

-

x	80	100	120
$p(x)$	.2	.3	.5

with $\mu = 106$, $\sigma^2 = 244$ (....5.2)

Example 3

cont'd

- Suppose on a particular day only two MP3 players are sold. Let X_1 = the revenue from the first sale and X_2 the revenue from the second.

- Suppose that X_1 and X_2 are independent, each with the probability distribution shown in (5.2) [so that X_1 and X_2 constitute a random sample from the distribution (mentioned above)].

Example 3

cont'd

- Table 5.2 lists possible (x_1, x_2) pairs, the probability of each [computed using (5.2) and the assumption of independence], and the resulting $\bar{x}$ and s^2 values. [Note that if $n = 2$, $s^2 = (x_1 - \bar{x})^2 (x_2 - \bar{x})^2$.]

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
80	80	.04	80	0
80	100	.06	90	200
80	120	.10	100	800
100	80	.06	90	200
100	100	.09	100	0
100	120	.15	110	200
120	80	.10	100	800
120	100	.15	110	200
120	120	.25	120	0

Outcomes, Probabilities, and Values of $\bar{x}$ and s^2 for Example 20

Table 5.2

Example 3

cont'd

- Now to obtain the probability distribution of $\overline{X}$, the sample average revenue per sale, we must consider each possible value $\overline{x}$ and compute its probability. For example, $\overline{x}$ = 100 occurs three times in the table with probabilities .10, .09, and .10, so

- $P_{\overline{X}}(100) = P(\overline{X} = 100) = .10 + .09 + .10 = .29$

- Similarly,

- $p_{S^2}(800) = P(S^2 = 800) = P[(X_1 = 80, X_2 = 120)$ or $(X_1 = 120, X_2 = 80)]$

- $= .10 + .10 = .20$

Example 3

cont'd

- The complete sampling distributions of $\overline{X}$ and S^2 appear in (5.3) and (5.4).

$\overline{x}$	80	90	100	110	120
$p_{\overline{X}}(\overline{x})$	.04	.12	.29	.30	.25

(5.3)

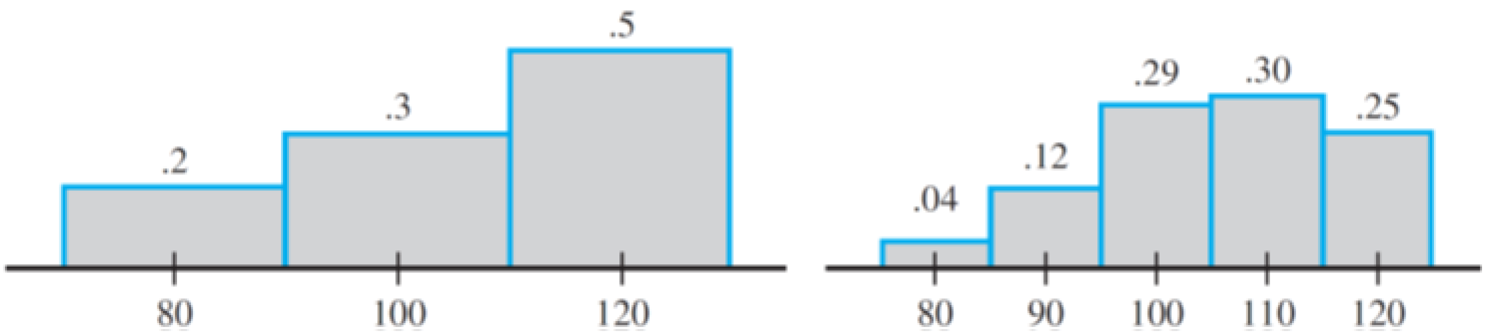
s^2	0	200	800
$pS^2(s^2)$	.38	.42	.20

(5.4)

Example 3

cont'd

Figure 5.7 pictures a probability histogram for both the original distribution (5.2) and the $\overline{X}$ distribution (5.3). The figure suggests first that the mean (expected value) of the distribution $\overline{X}$ is equal to the mean 106 of the original distribution, since both histograms appear to be centered at the same place.



Probability histograms for the underlying distribution and $\bar{x}$ distribution in Example 20

Figure 5.7

Example 3

cont'd

From (5.3),

$$\mu_{\overline{X}} = E(\overline{X}) = \sum \bar{x} p_{\overline{X}}(\bar{x})$$

$$= (80)(.04) + \ldots + (120)(.25) = 106 = \mu$$

Second, it appears that the $\overline{X}$ distribution has smaller spread (variability) than the original distribution, since probability mass has moved in toward the mean. Again from (5.3),

$$\sigma^2_{\overline{X}} = V(\overline{X}) = \sum \bar{x}^2 \cdot p_{\overline{X}}(\bar{x}) - \mu^2_{\overline{X}}$$

$$= (80^2)(.04) + \cdots + (120^2)(.25) - (106)^2$$

$$= 122 = \frac{244}{2} = \frac{\sigma^2}{2}$$

Example 3

cont'd

The variance of $\overline{X}$ is precisely half that of the original variance (because n = 2). Using (5.4), the mean value of S^2 is

$$\mu S^2 = E(S^2) = \Sigma\, S^2 \cdot p_{S^2}(s^2)$$

$$= (0)(.38) + (200)(.42) + (800)(.20) + 244 = \sigma^2$$

That is, the $\overline{X}$ sampling distribution is centered at the population mean μ, and the S^2 sampling distribution is centered at the population variance σ^2.

Class Notes:
Functions of Random Variables

Functions of RV

innovate achieve lead

Theorem 1 : Let X and Y be random variables with moment generating functions $m_X(t)$ and $m_Y(t)$, respectively. If $m_X(t) = m_Y(t)$ for all t in some open interval about 0, then X and Y have the same distribution. (Proof is out of syllabus)

Theorem 2 : Let X_1 and X_2 be independent random variables with moment generating functions $m_{X_1}(t)$ and $m_{X_2}(t)$, respectively. Let $Y = X_1 + X_2$. The moment generating function for Y is given by:

$$m_Y(t) = m_{X_1}(t).\, m_{X_2}(t) \qquad \text{(proof: trivial)}$$

Functions of RV

Ex 4 : (Distribution of the sum of independent normally distributed random variables)

Let $X_1, X_2, X_3, \ldots, X_n$ be independent normal random variables with means $\mu_1, \mu_2, \mu_3, \ldots, \mu_n$ and variances $\sigma^2{}_1, \sigma^2{}_2, \sigma^2{}_3, \ldots, \sigma^2{}_n$ respectively.

Let $Y = X_1+X_2+X_3+\ldots+X_n$. Note that the moment generating function for X_i is given by:

$$m_{X_i}(t) = e^{(\mu_i t + (\sigma^2{}_i t^2/2))} \quad i = 1,2,3,\ldots,n$$

Functions of RV

and the moment generating function for Y is (why?)

$$m_Y(t) = \prod_{i=1}^{n} m_{X_i}(t) = \exp\left[\left(\sum_{i=1}^{n}\mu_i\right)t + \left(\sum_{i=1}^{n}\sigma_i^2\right)\frac{t^2}{2}\right]$$

The function on the right is nothing but the moment generating function for a normal random variable Y with mean $\mu = \sum_{i=1}^{n}\mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n}\sigma_i^2$

Functions of RV

innovate achieve lead

Theorem 3:

Let X be a random variable with moment generating function $m_x(t)$. Let $Y = \alpha + \beta X$. The moment generating function for Y is

$$m_Y(t) = e^{\alpha t}\, m_x(\beta t)$$

Functions of RV

Theorem 4: (Distribution of $\overline{X}$ -normal population)

Let X_1, X_2, X_3,…, X_n be a random sample of size ‘n’ from a normal distribution with mean μ and variance σ^2.

Then $\overline{X}$ is normally distributed with mean μ and variance σ^2/n.

HW 8 (Distribution of a sum of independent random variables)

Let $X_1, X_2, X_3, \ldots, X_n$ be a collection of independent random variables with moment generating functions $m_{X_i}(t)$ (i=1,2,3,.....,n, respectively). Let a_0, $a_1, a_2, \ldots, a_n$ be real numbers, and let

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_n X_n.$$

Show that the moment generating function for Y is given by

$$m_Y(t) = e^{a_0 t} \prod_{i=1}^{n} m_{X_i}(a_i t)$$

HW 9 (Distribution of a linear combination of independent normally distributed random variables)

Let $X_1, X_2, X_3, \ldots, X_n$ be independent normal random variables with means μ_i and σ_i^2 (i=1,2,3,…,n, respectively). Let a_0, $a_1, a_2, \ldots,$ a_n be real numbers, and let

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

Show that Y is normal with mean

$$\mu = a_0 + \sum_{i=1}^{n} a_i \mu_i \text{, and variance}$$

$$\sigma^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 .$$

HW 10 (Distribution of a sum of independent chi-squared random variables)

Let $X_1, X_2, X_3, \ldots, X_n$ be independent chi-squared random variables with $\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n$ degrees of freedom, respectively.

Let $Y = X_1 + X_2 \ldots + X_n$.

Show that Y is a chi-squared random variable with degrees of freedom where $\gamma = \sum \gamma_i$

Homework

HW 11. Let $X_1, X_2, X_3, \ldots, X_{100}$ be a random sample of size 100 from gamma distribution with $\alpha=5$ and $\beta=3$.

(a) Find the mgf of $Y = \sum_{i=1}^{100} X_i$

(b) What is the distribution of *Y*?

(c) Find the mgf of $\overline{X} = Y / n$

(d) What is the distribution of $\overline{X}$?

Central Limit Theorem (CLT)

- Regardless of the population distribution model, as the sample size increases, the sample mean tends to be normally distributed around the population mean, and its standard deviation shrinks as *n* increases.
- Two conditions must be satisfied to apply CLT – (a) samples must be i.i.d. (b) sample size must be large enough (usually, n ≥ 30, but depends on problem!)

Let $X_1, X_2, \cdots, X_n$ be a random sample (that is, X_i are i.i.d) of size n from a distribution with mean μ and variance σ^2. Then for large n, sample mean $\bar{X}$ is approximately normal with mean μ and variance σ^2/n; $\bar{X} \to N\left(\mu, \sigma/\sqrt{n}\right)$

Furthermore, for large n, the random variable $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$.

CLT: Examples

- A certain brand of tires has a mean life of 25,000 km with a s.d. of 1600 km. What is the probability that the mean life of 64 tires is less than 24,600 km?

- A hawker sells dolls at various prices with a mean of Rs.700 and s.d. of Rs. 250. Selling prices of a random sample of 60 dolls are observed. What is the probability that he earns Rs. 45000 or more by selling those 60 dolls?

- Can we find the distribution of sample mean (annual salary) of the managers?

$$\text{Knowing, } \mu_{\bar{X}} = 51800 \quad \text{and} \quad \sigma_{\bar{X}} = 730.3,$$

$$\text{(i) } P\left(51{,}300 < \bar{X} < 52{,}300\right) = ?$$

$$\text{(ii) } P\left(51{,}000 < \bar{X} < 52{,}000\right) = ?$$

Central Limit Theorem (CLT)

- **Randomization** – we assume that **samples constitute a random sample (i.i.d.)** from the population.
- **Large enough sample size – how large is large?**
  - If the population is **normal**, then the sampling distribution $\bar{X}$ will also be normal, no matter what is the sample size.
  - If the population is approximately **symmetric**, the distribution becomes approximately normal for relatively small values of n.
  - When the population is **skewed**, the sample size must be **at least 30** before the sampling distribution of $\bar{X}$ becomes approximately normal.

CLT: Step by Step

Step 1: Identify parts of the problem. Your question should state:

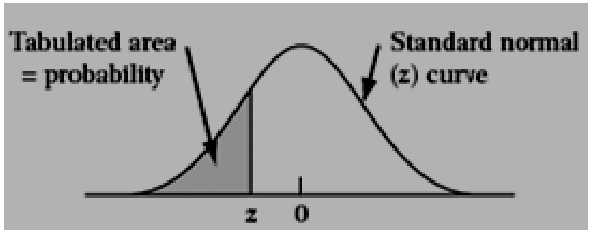
- The mean (average or μ)
- The standard deviation (σ)
- The sample size (n)

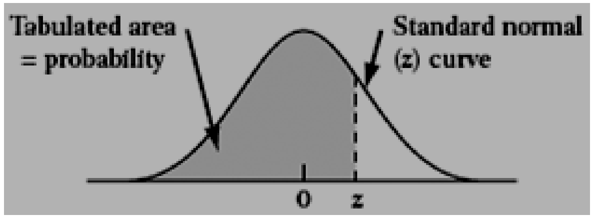
Step 2: Find $\bar{X}$ and express the problem in terms of "greater than" or "less than" the sample mean $\bar{X}$.

Step 3: Use CLT to find the distribution of $\bar{X}$ and $\bar{X} \to N\left(\mu, \sigma/\sqrt{n}\right)$

Step 4: Convert the normal variate $\bar{X}$ to a standard normal variate $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

Now you may draw a graph, centre with the 0 (mean of ***Z***) and shade the appropriate area to find the required probability.





Problem Solving

HW12. A certain brand of tyres has a mean life of 25,000 km with a s.d. of 1600 km. What is the probability that the mean life of 64 tyres is less than 24,600 km?

Sol.

Step 1: Here $X_1, X_2, \cdots, X_{64}$ constitute a random sample, and it is given that $E(X_i) = 25,000$ and $\sigma_{X_i} = 1600$.

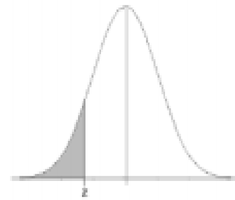
Step 2: $\bar{X} = \frac{1}{64}\sum_{i=1}^{64} X_i$ and our interest is to find out $P(\bar{X} < 24600)$.

Step 3: As we have a random sample of size 64 (sufficiently large n), we can use CLT to find the distribution of $\bar{X}$, that is, $\bar{X} \sim N\left(25000, \frac{1600}{\sqrt{64}}\right) \Rightarrow \bar{X} \sim N(25000, 200)$

Step 4: $P(\bar{X} < 24600) = P\left(\frac{\bar{X} - 25000}{200} < \frac{24600 - 25000}{200}\right)$

$$= P(Z < -2) = 0.0228 \text{ (using standard normal cdf table)}$$

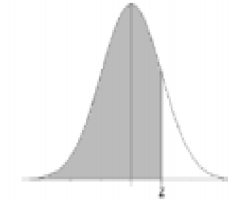
Standard Normal Cumulative Probability Table



Cumulative probabilities for NEGATIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Standard Normal Cumulative Probability Table



Cumulative probabilities for POSITIVE z-values are shown in the following table:

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Problem Solving

HW13. A hawker sells dolls at varying prices with a mean of Rs.700 and s.d. of Rs. 250. Selling prices of a random sample of 60 dolls are observed. What is the probability that he earns Rs. 45000 or more by selling those 60 dolls?

Step 1: Here $X_1, X_2, \cdots, X_{60}$ constitute a random sample, and it is given that $E(X_i) = 700$ and $\sigma_{X_i} = 250$.

Step 2: $\bar{X} = \frac{1}{60}\sum_{i=1}^{60} X_i$ and our aim is to find out $P\left(\sum_{i=1}^{60} X_i > 45000\right)$ i.e., $P(\bar{X} > 750)$.

Step 3: As we have a random sample of size 60 (sufficiently large n), we can use CLT to find the distribution of $\bar{X}$, that is, $\bar{X} \sim N\left(700, \frac{250}{\sqrt{60}}\right) \Rightarrow \bar{X} \sim N(700, 32.27)$

Step 4: $P(\bar{X} > 750) = P\left(\frac{\bar{X} - 700}{32.27} > \frac{750 - 700}{32.27}\right) = P(Z > 1.55)$

$$= P(Z < -1.55) = 0.0606 \text{ (using standard normal cdf table)}$$

Problem Solving

HW14. A population of 30 year – old males has a mean salary of Rs. 75000 with a standard deviation of Rs. 10000. If a sample of 100 men is taken, what is the probability that their mean salary will be less than Rs. 77500?

HW15. A certain group of welfare recipients receives pension benefits of Rs. 45000 per month with a standard deviation of Rs. 7500. If a random sample of 25 people is taken, what is the probability that their mean pension benefit will be greater than Rs. 47000 or less than Rs. 43000 per month?

HW16. A certain population of dogs weigh an average of 5 kg, with a standard deviation of 2 kg. If 40 dogs are chosen at random, what is the probability they have an average weight of greater than 6.0 kg or less than 4.5 kg?

Problem Solving

innovate achieve lead

HW 17. Suppose the weights of a certain type of FedEx freight boxes follow a distribution with mean 206 pounds and standard deviation 21 pounds. If 49 such FedEx freight boxes are chosen at random, what is the probability that their combined weight lies between 9800 pounds and 10241 pounds? **(Ans: 0.8185?)**

HW 18. Phone bills for residents of a city have a mean of Rs. 850 and a standard deviation of Rs. 144. If a random sample of 36 phone bills is drawn from this population, what is the probability that the mean bill will lie between Rs. 800 and Rs. 875? **(Ans: 0.8320?)**

Problem Solving

HW19. Human pregnancies follow a normal distribution with mean of 268 days and s.d. of 11 days. We study the mean pregnancy length of 70 women (call this random variable as sample mean $\bar{X}$). What is the expected value and s.d. of this statistic?

Sol. Step 1: $X_1, X_2, \cdots, X_{70}$ constitute a random sample of size 70, and $X_i \sim N(268, 11)$

Step 2: Here $\bar{X} = \dfrac{1}{70}\sum_{i=1}^{70} X_i$ and our interest is to find out $E(\bar{X})$ and s.d. of $\bar{X}$.

Step 3: Each sample follows normal distribution, so is their mean. Using CLT,

$\bar{X} \sim N\left(268, \dfrac{11}{\sqrt{70}}\right)$. So $E(\bar{X}) = 268$ and $\sigma_{\bar{X}} = \dfrac{11}{\sqrt{70}} = 1.314.$

Note. If $X_1, X_2, \cdots, X_{70}$ is just a random sample with $E(X_i) = 268$ and $\sigma_{X_i} = 11,$ we can use CLT to find mean and s.d. of $\bar{X}$.

Textbook Problems

Q 37: Already discussed

Q 46:

The inside diameter of a randomly selected piston ring is a random variable with mean value 12 cm and standard deviation 0.04 cm.

a. If $\overline{X}$ is the sample mean diameter for a random sample of $n=16$ rings, where is the sampling distribution of $\overline{X}$ centered, and what is the standard deviation of the $\overline{X}$ distribution?

b. Answer the questions posed in part (a) for a sample size of $n = 64$ rings.

Textbook Problems

Q 47:

Refer to Exercise 46. Suppose the distribution of diameter is normal.

a. Calculate $P(11.99 \le \bar{X} \le 12.01)$ when $n = 16$.

b. How likely is it that the sample mean diameter exceeds 12.01 when $n = 25$?

Textbook Problems

Q 59. Let X_1, X_2, and X_3 represent the times necessary to perform three successive repair tasks at a certain service facility. Suppose they are independent, normal rv's with expected values μ_1, μ_2, and μ_3 and variances σ_1^2 , σ_2^2 ,and σ_3^2 respectively.

a. If $\mu_1 = \mu_2 = \mu_3 = 60$ and $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 15$, calculate $P(T_0 \leq 200)$ and $P(150 \leq T_0 \leq 200)$.

b. Using the μ_i's and σ_i's given in part (a), calculate both $P(55 \leq \bar{X})$ and $P(58 \leq \bar{X} \leq 62)$.

c. Using the μ_i's and σ_i's given in part (a), calculate $P(-10 \leq X_1 - 0.5X_2 - 0.5X_3 \leq 5)$.

Textbook Problems

Q 59.

d. If $\mu_1 = 40$, $\mu_2 = 50$, $\mu_3 = 60$, $\sigma_1^2 = 10$, $\sigma_2^2 = 12$ and $\sigma_3^2 = 14$, calculate $P(X_1 + X_2 + X_3 \leq 160)$

d. If $\mu_1 = 40$, $\mu_2 = 50$, $\mu_3 = 60$, $\sigma_1^2 = 10$, $\sigma_2^2 = 12$ and $\sigma_3^2 = 14$, calculate $P(X_1 + X_2 \geq 2X_3)$.

Textbook Problems

Q 75. (a) Let X_1 and X_2 be two chi-square independent random variables with parameters v_1 and v_2, respectively. Let Y = X1 + X2. Find the distribution of Y.

(b) If Z is a standard normal rv, then Z^2 has a chi-squared distribution with $\nu = 1$. Let $Z_1, Z_2, \ldots, Z_n$ be n independent standard normal rv's. What is the distribution of $Z_1^2 + Z_2^2 + \ldots + Z_n^2$? Justify your answer.

(c) Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. What is the distribution of the sum

$$Y = \sum_{i=1}^{n} [(X_i - \mu)/\sigma]^2?$$