Department of Statistics
Faculty of Science
Yarmouk University

SATS 101

Introduction to Probability and Statistics

Yarmouk University

Second Semester

2009/2010

Done by: Osama Alkhoun
Mobile: 0796484613

Chapter 7
Sampling Distributions

Sampling plan: the way a sample is selected

- Simple random sample:
  - N: population size.
  - n : sample size

simple random sample of size n:
is a simple which each sample of size n has same chance or probability of being selected

example:
population N = 5 {1, 2, 3, 4, 5 }
select all possible simple random sample of size n = 2
sample =
{(1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5),(4, 5)}
P ((1, 2)) = P ((1, 3)= ………. = P ((4, 5)) = $\frac{1}{10}$

The number of sample size (n) from (N):

$\binom{N}{n}$ = The number of all possible sample of size (n) select from (N)

N = 5 elements
n = 2 elements

$$\binom{N}{n} = \binom{5}{2} = \frac{5!}{(5-2)!\,2!} = 10$$

Example:
Find The number of all possible sample of size (n = 3) select from

(N = 5)
N = 5 elements
n = 3 elements

$$\binom{N}{n} = \binom{5}{3} = \frac{5!}{(5-3)!\,3!} = 10$$

Sampling Plans

1. **Stratified random sample:** is obtain divide the population into group subpopulations or **strata** and select a simple random sample from each strata.
2. **Cluster sample:** is obtained divide the population into subgroups called **clusters;** select a simple random sample of clusters and take a census of every element in the cluster.
3. **1-in-k systematic sample**: is obtained by starting at random position Randomly select one of the first k elements in an ordered population, and then select every k-th element thereafter.

Example:
If we need to select every seventh students how enter the library?

Sampling distribution:
The sampling distribution of statistics is the probability distribution for all possible values of statistic that result when random sample of size n are repeatedly drown from the population.

Example:
1. Find The number of all possible sample of size (n = 3) select from (N = 5)
N = 5 elements
n = 3 elements

$$\binom{N}{n} = \binom{5}{3} = \frac{5!}{(5-3)!\,3!} = 10$$

2. find the sampling distribution of the $\overline{x} = \frac{\sum Xi}{n}$

{(1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5),(4, 5)}

(1, 2) = $\frac{1+2}{2}$ = 1.5

(1, 3) = $\frac{1+3}{2}$ = 2

$\overline{x}$	1.5	2	2.5	3	3.5	4	4.5
P(x)	0.1	0.1	0.2	0.2	0.1	0.1	0.1

3. Find E ($\overline{x}$) = $\sum \overline{x} P(\overline{x})$

$E(\overline{X}) = \sum \overline{X} P(\overline{X}) = 3$ $\quad$ $E(\overline{X}) = \frac{\sum \overline{X}}{n} = \frac{21}{7} = 3$

4. find the population mean $\mu = \frac{\sum Xi}{n} = \frac{1+2+3+4+5}{5} = 3$

$$E(\overline{X}) = \mu$$

Central Limit Theorem (CLT) for $\overline{x}$ in sampling distribution

If a random samples of n observations are drawn from a non normal population with finite mean μ and standard deviation of the sample mean ($\overline{x}$) is approximately normally distribution with

mean $\mu_{\overline{X}}$ = E ($\overline{x}$) = μ and **variance** $\sigma^2_{\overline{X}}$ = variance($\overline{x}$)= $\frac{\sigma^2}{n}$

Characteristics of Central Limit Theorem (CLT)

1. sampling is a large
2. normal distribution $\overline{x} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$
3. $\overline{x}$ distributed approximately normal distribution such that we have mean μ and variance $\frac{\sigma^2}{n}$

Example:

If we take a sample of size 40 from a population with mean 80 and variance 100, define $\overline{x}$ to be the sample mean of the sample.

1. find the mean of $\overline{x}$

   $\mu_{\overline{X}}$ = E ($\overline{x}$) = μ = 80

2. find the expectation of $\overline{x}$

   $\mu_{\overline{X}}$ = E ($\overline{x}$) = μ = 80

3. find the variance of $\overline{x}$

   $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{100}{40} = 2.5$

4. What is the sampling distribution? Why?

   the distribution is normal.

   $\overline{x} \approx N(80, 2.5)$, sample from a population sampling distribution of $\overline{x}$ if we have n observation from a population.

If random sample of size n is selected from a population with mean μ and variance σ^2 then

1. E ($\overline{x}$) = μ
2. $\sigma^2 = \dfrac{\sigma^2}{n}$

Case 1:
If sampling distribution of $\overline{x}$ is **normal** with mean μ and variance σ^2 then $\overline{x}$: N $\left(\mu, \dfrac{\sigma^2}{n}\right)$

Case 2:
If the sample is large then sampling distribution of $\overline{x}$ is **approximately normal** with mean μ and variance σ^2 then:

$\overline{x} \approx N\left(\mu, \dfrac{\sigma^2}{n}\right)$ then:

Standard error (S.E)

$\left(S.E_{\bar{x}}\right) = \sqrt{var(\bar{x})}$

$\left(S.E_{\bar{x}}\right) = \dfrac{\sigma}{\sqrt{n}}$

If X:N $\left(\mu, \sigma^2\right)$ → Z = $\dfrac{X-\mu}{\sigma}$

If X:N $\left(\mu, \dfrac{\sigma^2}{n}\right)$ → Z = $\dfrac{X-\mu}{\frac{\sigma}{\sqrt{n}}}$

- What is the distribution of $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ → $\overline{x}$: N (0, 1)
- Z = $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Example:
If we take a sample of size 30 from a normal with mean 8 and standard deviation 4 define $\overline{x}$ to be the sample mean of the sample.

1. find the mean of $\overline{x}$

$$\mu_{\overline{X}} = E(\overline{x}) = \mu = 8$$

2. find the expectation of $\overline{x}$

$$\mu_{\overline{X}} = E(\overline{x}) = \mu = 8$$

3. find the variance of $\overline{x}$

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{16}{30} = 0.53$$

4. find the Standard error (S.E) of $\overline{x}$

$$\left(S.E_{\bar{x}}\right) = \sqrt{var(\bar{x})} = \sqrt{\frac{16}{30}}$$

5. What is the sampling distribution of $\overline{x}$?

$$\overline{x} : N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow \overline{x} : N\left(8, \frac{16}{30}\right) [\overline{x} \text{ has a normal }]$$

6. What is the sampling distribution of $\frac{\sqrt{30}\,(\bar{x} - 8)}{4}$?

$\overline{x}$: N (0, 1) standard deviation

7. find the P ($\overline{x}$ < 7) $\overline{x}$: $N\left(8, \frac{16}{30}\right)$

$$P\left(Z < \frac{7-8}{\frac{4}{\sqrt{30}}}\right) = P(Z < -1.37) = 0.0853$$

8. find the P (x < 8)

$$P\left(\frac{X-8}{4} < \frac{7-8}{4}\right) = P(Z < 0) = 0.5$$

9. find the P (7 < $\overline{x}$ < 9)

$$P\left(\frac{7-8}{\frac{4}{\sqrt{30}}} < Z < \frac{9-8}{\frac{4}{\sqrt{30}}}\right)$$

P (-1.37 < Z < 1.37) = P (Z< 1.37) – P (Z <-1.37)
= 0.9147 - 0.0853 = 0.8294

Example:

The distribution of grades of STAT 101 is normal with mean 70 and variance 16 if we take the sample of size 36.

1. find the probability that mean (average) of graded of 36 students is greater than 72.

$$P(\overline{x} > 72) = P\left(\frac{\overline{X}-70}{\frac{4}{6}} < \frac{72-70}{\frac{4}{6}}\right)$$

$$= P(Z > 3) = 1 - P(Z < 3)$$

$$= 1 - 0.9987$$

$$= 0.0013$$

1. find the probability that the total grade of those students will be less than 2556.

$$P\left(\frac{\overline{\sum Xi}}{n} < \frac{2556}{n}\right) = P\left(\overline{x} < \frac{2556}{36}\right)$$

$$= P(\overline{x} < 71)$$

$$= P\left(Z > \frac{71-70}{\frac{4}{6}}\right) = P(Z < 1.5)$$

$$= 0.9332$$

3. What is the sampling distribution of $\overline{x}$?

$$\overline{x} : N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow \overline{x} : N\left(70, \frac{16}{36}\right) \text{ [} \overline{x} \text{ has a normal]}$$

4. if we select one student, find the probability that has grade with be less 70.

$$P(x < 70) = P\left(Z > \frac{70-70}{\frac{4}{6}}\right) = P(Z < 0)$$

$$= 0.5$$

A sampling distribution is created by, as the name suggests, **sampling**.

The method we will employ on the **rules of probability** and the **laws of expected value and variance** to derive the sampling distribution.

For example, consider the roll of one and two dice…

Sampling Distribution of the Mean…
A fair **die** is thrown infinitely many times,
with the random variable X = # of spots on any throw.

The probability distribution of X is:

x	1	2	3	4	5	6
P(x)	1/6	1/6	1/6	1/6	1/6	1/6

…and the mean and variance are calculated as well:

$$\mu = \sum xP(x) = 1(\tfrac{1}{6}) + 2(\tfrac{1}{6}) + \ldots + 6(\tfrac{1}{6}) = 3.5$$

$$\sigma^2 = \sum (x-\mu)^2 P(x) = (1-3.5)^2(\tfrac{1}{6}) + \ldots + (6-3.5)^2(\tfrac{1}{6}) = 2.92$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.92} = 1.71$$

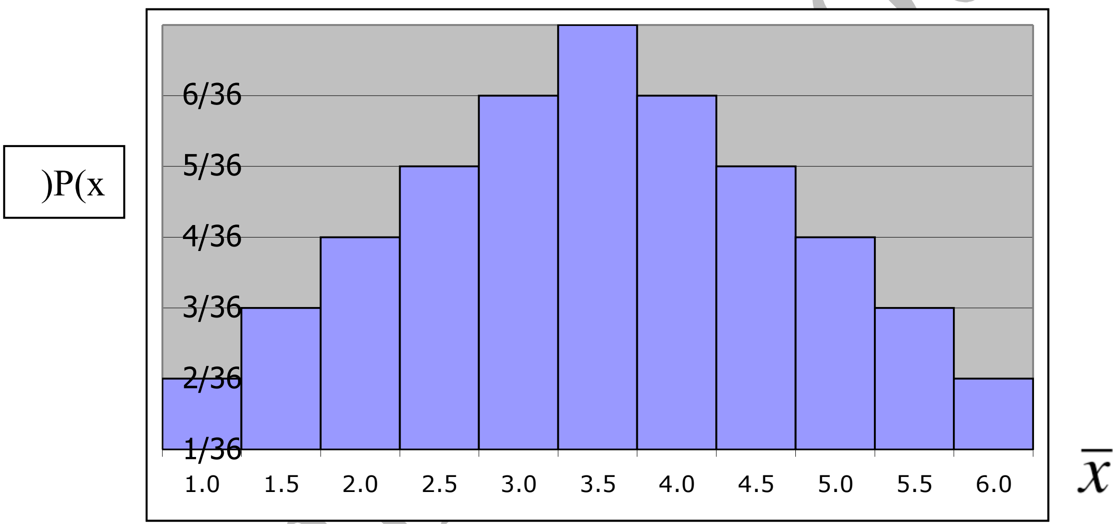
Sampling Distribution of Two Dice
A sampling distribution is created by looking at
all samples of size n=2 (i.e. two dice) and their means…

Sample		Sample		Sample	
1, 1	1.0	3, 1	2.0	5, 1	3.0
1, 2	1.5	3, 2	2.5	5, 2	3.5
1, 3	2.0	3, 3	3.0	5, 3	4.0
1, 4	2.5	3, 4	3.5	5, 4	4.5
1, 5	3.0	3, 5	4.0	5, 5	5.0
1, 6	3.5	3, 6	4.5	5, 6	5.5
2, 1	1.5	4, 1	2.5	6, 1	3.5
2, 2	2.0	4, 2	3.0	6, 2	4.0
2, 3	2.5	4, 3	3.5	6, 3	4.5
2, 4	3.0	4, 4	4.0	6, 4	5.0
2, 5	3.5	4, 5	4.5	6, 5	5.5
2, 6	4.0	4, 6	5.0	6, 6	6.0

While there are 36 possible samples of size 2, there are only 11 values for, and some occur more frequently than others.

The ***sampling distribution*** of $\overline{x}$ is shown below:

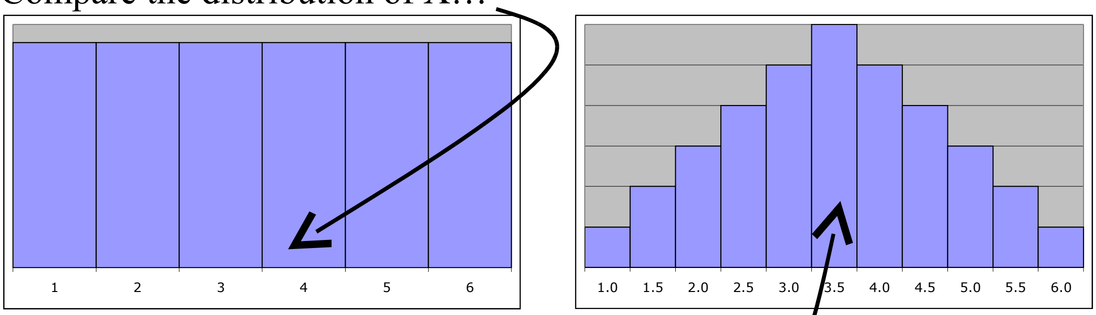
1.0	1/36
1.5	2/36
2.0	3/36
2.5	4/36
3.0	5/36
3.5	6/36
4.0	5/36
4.5	4/36
5.0	3/36
5.5	2/36
6.0	1/36



$$\sigma_{\bar{x}}^2 = \sum (\bar{x} - \mu_{\bar{x}})^2 P(\bar{x}) = (1.0 - 3.5)^2(\tfrac{1}{36}) + \ldots + (6.0 - 3.5)^2(\tfrac{1}{36}) = 1.46$$

$$\mu_{\bar{x}} = \sum \bar{x} P(\bar{x}) = 1.0(\tfrac{1}{36}) + 1.5(\tfrac{2}{36}) + \ldots + 6.0(\tfrac{1}{36}) = 3.5$$

Compare…
Compare the distribution of X…



…with the sampling distribution of $\bar{x}$

As well, note that: $\mu_{\bar{x}} = \mu$

$$\sigma_{\bar{x}}^2 = \sigma^2 / 2$$

Generalize…
We can generalize the mean and variance of the sampling of two dice:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}}^2 = \sigma^2 / 2$$

…to **n**-dice:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

The standard deviation of the sampling distribution is called the ***standard error***:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$