ECON 214
Elements of Statistics for Economists

Session 8 – Sampling Distributions

Lecturer: Dr. Bernardin Senadza, Dept. of Economics
Contact Information: bsenadza@ug.edu.gh

UNIVERSITY OF GHANA
College of Education
School of Continuing and Distance Education
2014/2015 – 2016/2017

Session Overview

- We begin our discussion of statistical estimation with sampling distributions in this Session.
- A (sample) statistic is a random variable and as such has a probability distribution.
- The probability distribution of a statistic is called its sampling distribution.
- We examine the sampling distribution of the sample mean and sample proportion in this Session.

Session Overview

- At the end of the session, the student will
  - Be able to explain why a sample is the only feasible way to learn about a population
  - Be able to explain methods for selecting a sample
  - Be able to define and construct a sampling distribution of the sample mean
  - Be able to define and construct a sampling distribution of the sample proportion
  - Be able to explain the Central Limit Theorem

Session Outline

The key topics to be covered in the session are as follows:

- Probability Sampling
- Sampling Distribution of the Sample Mean
- Sampling Distribution of the Sample Proportion

Reading List

- Michael Barrow, "Statistics for Economics, Accounting and Business Studies", 4th Edition, Pearson
- R.D. Mason , D.A. Lind, and W.G. Marchal, "Statistical Techniques in Business and Economics", 10th Edition, McGraw-Hill

Topic One

PROBABILITY SAMPLING

Why sample the population?

- The physical impossibility of checking all items in the population.
- The cost of studying all the items in a population.
- The sample results are usually adequate.
- Contacting the whole population would often be time-consuming.
- The destructive nature of certain tests.

Probability Sampling

- A probability sample is a sample selected in such a way that each item or person in the population being studied has a known likelihood of being included in the sample.

Methods of Probability Sampling

- **Simple Random Sample**: A sample formulated so that each item or person in the population has the same chance of being included.
- **Systematic Random Sampling**: The items or individuals of the population are arranged in some order.
- A random starting point is selected and then every *k*th member of the population is selected for the sample.

Methods of Probability Sampling

- **Stratified Random Sampling**: A population is first divided into subgroups, called strata, and a sample is selected from each stratum.
- **Cluster Sampling**: A population is first divided into subgroups (strata), and a sample of the strata is selected. The sample is then taken from these selected strata.
- A **sampling error** is the difference between a sample statistic and its corresponding parameter.

Topic Two

SAMPLING DISTRIBUTION OF THE SAMPLE MEAN

Sampling Distribution of the Sample Mean

- Generally there are a large number of possible samples that can be selected from a population.
- The value of any statistic (such as the sample mean) computed from a sample will vary from sample to sample.
- A statistic is therefore a random variable and as such has a probability distribution.
- The probability distribution of a statistic is called its sampling distribution.

Sampling Distribution of the Sample Mean

- The sampling distribution of the (sample) mean is the probability distribution for the possible values of the sample mean.
- For any given sample of size ***n*** taken from a population with mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$, the value of the sample mean would vary from sample to sample.

Sampling Distribution of the Sample Mean

- **Illustration**: A PhD class in economics has five students. Each of the students reported the number of hours they studied during the first week of the 2nd semester.

No.	Student	Hours
1	Kofi	22
2	Ama	26
3	Esi	30
4	Eric	26
5	Mensah	22

- If two students are selected randomly, how many different samples are possible?

Sampling Distribution of the Sample Mean

Sample	Total Hrs	Mean
1, 2	48	24
1, 3	52	26
1, 4	48	24
1, 5	44	22
2, 3	56	28
2, 4	52	26
2, 5	48	24
3, 4	56	28
3, 5	52	26
4, 5	48	24

- This is the combination of 5 objects taken 2 at a time. That is,

$$^{5}C_{2} = \frac{5!}{2!3!} = 10$$

Sampling Distribution of the Sample Mean

- Organize the sample means into a sampling distribution.

Sample mean	Frequency	Rel. Freq. (Probability)
22	1	$^{1}/_{10}$
24	4	$^{4}/_{10}$
26	3	$^{3}/_{10}$
28	2	$^{2}/_{10}$

- Note that the sum of probabilities equals 1

Sampling Distribution of the Sample Mean

- Compute the mean of the sample means and compare it with the population mean:
  - The mean of the sample means = [(22)(1) + (24)(4) + (26)(3) + (28)(2)]/10=25.2
  - The population mean = (22+26+30+26+22)/5 = 25.2
  - *Observe that the mean of the sample means is equal to the population mean.*

Sampling Distribution of the Sample Mean

- The sampling distribution of the (sample) mean is described by determining its **expected value** (mean) and standard deviation (or **standard error**).

Sampling Distribution of the Sample Mean

- The expected value (mean) of the sample mean is equal to the population mean

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

- The standard error is the population standard deviation divided by the square root of the sample size.

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

Sampling Distribution of the Sample Mean

- **Example**: Suppose the mean of a large population of measurements is μ =100 and the population standard deviation is σ =15. For samples of size n = 36, the expected value and standard error of the mean are:

$$E(\bar{X}) = \mu = 100$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{36}} = 2.5$$

Sampling Distribution of the Sample Mean

- In general, if samples of size *n* are randomly drawn from a Normally distributed population of mean μ and variance σ^2, the sample mean is distributed as

$$\overline{X} \sim N\left(\mu, \sigma^2/n\right)$$

- **Example**: If samples of 50 women are chosen, and the mean height is computed. The sample mean is distributed $\overline{X} \sim N(166, 40.32/50)$

Finite correction factor

- When the sample is taken from a population that is finite, a finite correction factor is required in the formula of the standard error
- That is when $n \geq .05N$ (sample is $\geq 5\%$ of population)
- Then the formula for the standard error becomes

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

- When the standard deviation of the population is unknown, we use the sample standard deviation.

Calculating probabilities for the sampling distributions

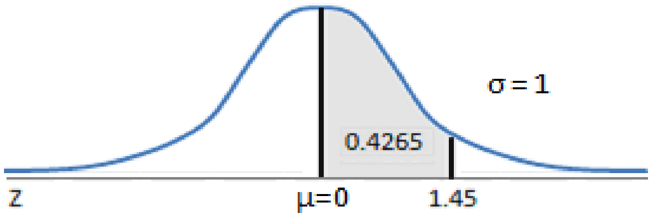
- We can calculate the probability that the sample mean will assume a given range of values.
- **Example**: What is the probability of drawing a sample of 50 women whose *average* height is greater than 168 cm?

$$Z = \frac{\overline{X} - \mu}{\sqrt{\sigma^2 / n}} = \frac{168 - 166}{\sqrt{40.32/50}} = 2.23$$

- We read from the standard normal table the area to the right of *Z* = 2.23 which gives .0129 [P(Z>2.23)=.0129]

Areas Under the One-Tailed Standard Normal Curve

This table provides the area between the mean and some Z score.
For example, when Z score = 1.45 the area = 0.4265.



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916

The Central Limit Theorem

- If samples are taken from a population that is normally distributed, the distribution of the sample mean is also normal.
- However, for a non-normal population, as the sample size gets large, $n > 25$, the distribution of the sample mean is (approximately) normal.
- This result is called the **Central Limit Theorem**.
- The approximation gets better, the larger the sample size.
- We can then proceed to calculate probabilities as before.

The Central Limit Theorem

- **Example**: An Auditor takes a sample of 50 accounts receivable to audit. If the mean value in the accounts is 200 and the standard deviation is 45, what is the probability that the sample mean will be less than 190.
- Since *n* is large, the distribution of the sample mean is approximately normal
- Thus the Z value is: $$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{190 - 200}{45 / \sqrt{50}} = -1.57$$
- So P(Z<-1.57) = .0582 (read from the Z table)

Topic Three

SAMPLING DISTRIBUTION OF THE SAMPLE PROPORTION

Sampling distribution of the proportion

- Many statistical investigations are carried out with the aim of estimating the proportion of the population having a specified characteristic.
- For example, we may be interested in the proportion of drug stores that sell a particular drug.
- In such cases the sample proportion is generally used as an estimate of the population proportion.
- Assume we have a population of 1000 drug stores.
- A sample of 40 is taken of which 32 sell the particular drug.

Sampling distribution of the proportion

- The sample proportion is

$$p = \frac{X}{n} = \frac{32}{40} = .8$$

- The population proportion is denoted

$$\pi = \frac{X}{N}$$

Sampling distribution of the proportion

- If we take many samples of size *n*, the value of the sample proportion will vary from sample to sample.
- For the sampling distribution of the proportion, we have the expected value as

$$E(p) = \pi$$

- And the standard error
- $$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \quad \text{or} \quad s_p = \sqrt{\frac{p(1-p)}{n}}$$

Calculating probabilities for the sampling distributions

- Just as with the sample mean, we can calculate the probability that the sample proportion will assume a given range of values.
- **Example**: It is known that only 70 percent of drug stores sell a particular drug in Ghana. A sample of 40 drug stores is selected. Find the probability that more than 32 sell this drug.
- **Solution**: We want to find the probability that the sample mean, ***p***, is greater than 0.8 (or 32/40).

Calculating probabilities for the sampling distributions

- The population proportion is 0.7, the sample proportion, p = 0.8, and sample size, n = 40.
- Using the standard normal $Z = \dfrac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}}$

$$P(p > 0.8) = P\left(Z > \frac{0.8 - 0.7}{\sqrt{\dfrac{0.7(1-0.7)}{40}}} \right) = P(Z > 1.38) = .0838$$

References

- Michael Barrow, "Statistics for Economics, Accounting and Business Studies", 4th Edition, Pearson
- R.D. Mason , D.A. Lind, and W.G. Marchal, "Statistical Techniques in Business and Economics", 10th Edition, McGraw-Hill