

ILRST/STSCI 2100 Discussion 7: Ch 8 Sampling Distributions

Sampling Distribution:

When given a population, you can draw samples from it. The mean of the samples can be calculated and a probability distribution can be generated from these sample means, referred to as **sampling distribution of the sample means**.

Ex. Given the population distribution, find the sampling distribution of the sample means, assuming that sample size is 2. Also, find the relationship between sampling mean and population mean, and between population variance and the variance of the sampling distribution.

x	f(x)
3	.3
4	.2
5	.5

Note that the sum of $f(x)$ must equal to 1.

To find the sampling distribution, we must **(1) determine all the possible samples**.

Sample	Sample mean (\bar{x})
3,3	3
3,4	3.5
3,5	4
4,3	3.5
4,4	4
4,5	4.5
5,3	4
5,4	4.5
5,5	5

Note that there should be $3 * 3$ number of possible samples, since there are 3 total objects in the population and we are selected 2 to form a sample. After calculating the sample mean, also **(2) calculate the probability of obtaining the sample**. The probability can be found by multiplying the individual probabilities. Note that the sum of the "Probability" column is also equal to 1

Sample	Sample mean \bar{x}	Probability
3,3	3	.09
3,4	3.5	.06
3,5	4	.15
4,3	3.5	.06
4,4	4	.04
4,5	4.5	.10
5,3	4	.15
5,4	4.5	.10
5,5	5	.25

The "Sampling Distribution of the Sample Mean" can now be constructed by listing the **(3) different sample means and its probability**.

Sample mean	Probability $f(\bar{x})$
3	.09
3.5	.12
4	.34
4.5	.20
5	.25

Note that the sum of the probability column add up to 1.

(4) We can now calculate the sample mean by using the equation $\bar{x} * f(\bar{x})$ and summing the column

\bar{x}	$f(\bar{x})$	$\bar{x} * f(\bar{x})$
3	.09	.27
3.5	.12	.42
4	.34	1.36
4.5	.20	.9
5	.25	1.25

(5) The sum is $.27 + .42 + 1.36 + .9 + 1.25 = 4.2$. So the mean of the sampling distribution is 4.2.

(6) Compare to the population mean

The population mean is found by summing $x * f(x) = 3 * .3 + 4 * .2 + 5 * .5 = 4.2$. The mean of the population must equal the mean of the sampling distribution.

(7) Next, calculate the variance of the sampling distribution

\bar{x}	$f(\bar{x})$	$(\bar{x} - \mu)^2 f(\bar{x})$
3	.09	.13
3.5	.12	.06
4	.34	.014
4.5	.20	.018
5	.25	.16

The variance is then found by **(8) the sum of the $(\bar{x} - \mu)^2 f(\bar{x})$ column.** So sampling distribution variance is equal to $.13 + .06 + .014 + .018 + .16 = 0.38$. It is important that the sampling mean or the population mean is first calculated because the variance is dependent on this number.

9. Find population variance.

x	f(x)	$(x - \mu)^2 f(x)$
3	.3	0.432
4	.2	0.008
5	.5	0.32

The population variance is found by taking the sum of $(x - \mu)^2 f(x) = 0.432 + 0.008 + 0.32 = 0.76$

Note that the population variance $\sigma^2 = 0.76$ and that $s^2 = .38$ and that $\frac{\sigma^2}{n} = s^2$

$$\frac{0.76}{2} = 0.38$$

Central limit theorem: The CLT states for any distribution, the sample means will be normally distributed when the sample size is large enough.