Chapter 4: Sampling Distribution

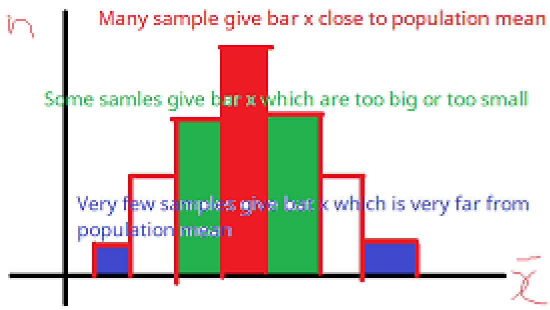
July 2, 2019

Population - set of all the objects of interest in a particular study. Sample - consists of some elements of the population selected for making observations (measurement, testing, etc). The observed values obtained in a sample are called data. eg. If we wish to study the average life of light bulbs produced by a factory, we may randomly pick 100 bulbs for testing. The set of all bulbs produced by the factory form the "population". Those 100 bulbs selected for testing form a sample.

Sample size means the number of data in the sample. In the example above, the sample size is 100. In inferential statistics, we attempt to make conclusions about the population based on the sample data. For unbiased conclusions, we have to use random sample. A simple random sample is a sample selected in such a way that all the items in the population have equal chance to be included. (This is equivalent to saying that all the samples of that size have equal chance to be selected.) For reliable conclusion, we need big samples. Suppose X is a measurable quantity of a population, such as mass, time, length, etc. The mean value of X in the population is μ, and the variance is σ^2

. These values that describe the population are called population parameters. The corresponding variables whose values are determined from sample data are called sample statistics. Population parameters are constants. Sample statistics are random variables. We may use Greek alphabet for population parameters, and Roman alphabet for sample statistics.

Pop. parameters	Samp. parameters
Mean $\mu = \frac{1}{n}\sum_{i=1}^{n} X_i$, where n= population size	Sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$, n=sample size
Variance $\sigma^2 = \frac{1}{N}\sum_{i=1}^{n}(X_i - \mu)^2$	Variance $= s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Proportion (many books use p) π=$\frac{X}{N}$,X is the number of items of a certain kind	Sample proportion $\hat{p} = \frac{X}{N}$, X is the number of items of that kind



The distribution appears to be a normal distribution with centre at μ.

1 Example

1.

(a) Let X be the number of soft drink filled into cans by a machine where $X \sim N(325, 2)$

i. $P(X > 326)$

$$\begin{aligned} P(X > 325.5) &= P\left(Z > \frac{0.5}{2}\right) \\ &= P\left(Z > \frac{1}{4}\right) \\ &= 0.30854 \end{aligned}$$

ii. Conclusion: The probability that the content of one randmoly selected can exceeds 326ml is 0.30854

$$\begin{aligned} P(\bar{x} > 326) &= P\left(\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{326 - 325}{\frac{2}{\sqrt{16}}}\right) \\ &= P(Z > 2) \\ &= 0.02275 \end{aligned}$$

(b)

$$\begin{aligned} P\left[\sum_{i=1}^{24} X_i > 7810\right] &= P\left(\frac{1}{24}\sum_{1}^{24} X_i > \frac{7810}{24}\right) \\ &= P\left(\bar{X} > 325.4167\right) \\ &= P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{325.4167 - 325}{\frac{2}{\sqrt{24}}}\right) \\ &= P(Z > 1.02) \\ &= 0.1537 \end{aligned}$$

2. $\mu = 8, \sigma = 2$

$$P\left(7.8 \leq \sum_{i=1}^{25} X_i \leq 8.2\right) = P\left(\frac{7.8-8}{\frac{2}{\sqrt{25}}} < Z < \frac{8.2-8}{\frac{2}{\sqrt{25}}}\right)$$
$$= P\left(-0.5 < Z < 0.5\right)$$
$$= 0.38292$$

3.

$$P\left(\bar{X} > 72\right) = 0.854$$
$$P\left(Z > -\frac{72-74}{\frac{6}{\sqrt{n}}}\right) = 0.854$$
$$P\left(Z > -\frac{\sqrt{n}}{3}\right) = 0.146$$
$$-\frac{\sqrt{n}}{3} = 1.0539$$
$$n \approx 10$$

(a) **Linear interpolation:**

$$\frac{Z - 1.05}{1.06 - 1.05} = \frac{0.146 - 0.1469}{0.1446 - 0.1409}$$
$$Z = 1.05 + ... * 0.01$$
$$= 1.0539$$

4.

(a) $X \sim P_o\left(4.5\right)$

$$n \geq 30$$
$$\mu \& \sigma^2 = 4.5$$

i. By the central limit theorem, $\bar{X} \sim N\left(4.5, \frac{4.5}{30}\right)$

$$P\left(\bar{X} > 5\right) = P\left(Z \geq \frac{5-4.5}{\sqrt{\frac{4.5}{30}}}\right)$$
$$= P\left(Z > 1.29\right)$$
$$= 0.0985$$

(b) $X \sim B\left(9, 0.5\right)$

$$n \geq 30$$
$$\mu = 9\left(0.5\right) = 4.5$$
$$\sigma^2 = 9 * 0.5 * 0.5 = 2.25$$

i. By the central limit theorem,

$$P\left(\bar{X} > 5\right) = \left(Z > \frac{5-4.5}{\sqrt{\frac{2.25}{30}}}\right)$$
$$= (Z > 1.8257)$$
$$= 0.03395\,(Answer : 0.0336)$$

5. $Z \sim P_o\,(2.5)$

(a) Mean, $\mu = \lambda = 2.5$
Variance $= 2.5$

(b) Central limit theorem

$$\bar{X} \sim N\left(2.5, \frac{2.5}{n}\right)$$

(c) Given $P\left(\bar{X} < 2.025\right) = 0.05$

$$P\left(Z < \frac{2.025-2.5}{\sqrt{\frac{2.5}{n}}}\right) = P\left(Z < -\frac{0.475}{\sqrt{\frac{2.5}{n}}}\right) = P\left(Z > \frac{0.475}{\sqrt{\frac{2.5}{n}}}\right) = 0.05$$

(d)

$$\frac{0.475}{\sqrt{\frac{2.5}{n}}} = 1.6449$$

(e)

$$n = \left(1.6449 * \frac{\sqrt{2.5}}{0.475}\right)^2$$
$$\approx 30$$

6.

(a) More than 5% will be broken

i. Let $\hat{p} \equiv proportion$ of pies in the sample which are broken

A. $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ approximately where $n = 500$ and $p = 0.04$

B. $\hat{p} \sim N\left(0.04, \frac{6}{78125}\right)$

ii.

$$P\left(\hat{p} \geq 0.05\right) = P\left(\left(\frac{\hat{p}-P}{\sqrt{\frac{p(1-p)}{n}}}\right)/Z > \frac{0.05-0.04}{\sqrt{\frac{0.04*0.96}{500}}}\right)$$
$$= P\left(Z > 1.14\right)$$
$$= 0.1271$$

(b) $P(\hat{p} \leq 0.03)$

$$P\left(Z < \frac{0.03 - 0.04}{\sqrt{\frac{0.04*0.96}{500}}}\right) = P\left(Z < \frac{0.03 - 0.04}{\sqrt{\frac{6}{78125}}}\right)$$
$$= P(Z < -1.14)$$
$$= P(Z > 1.14)$$
$$= 0.1271$$

7. $\mu_1 = 150,\ \sigma_1 = 20,\ \mu_2 = 125, \sigma_2 = 25, n = 5$

(a)

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2 = 150 - 125 = 25$$

(b)

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{20^2}{5} + \frac{25^2}{5} = 205$$

(c)

$$P\left(\bar{X}_1 \leq \bar{X}_2\right) = P\left(\bar{X}_1 - \bar{X}_2 \leq 0\right)$$
$$= P\left(Z \leq \frac{0 - 25}{\sqrt{205}}\right)$$
$$= P(Z \leq -1.75)$$
$$= P(Z > 1.75)$$
$$= 0.0401$$

8. $\mu_1 = 6.5,\ \sigma_1 = 0.9,\ \mu_2 = 6, \sigma_2 = 0.8, n_1 = 36, n_2 = 49$

(a)

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2 = 0.5$$

(b)

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{0.9^2}{36} + \frac{0.8^2}{49} = 0.03556$$

(c)

$$P\left(\bar{X}_1 - \bar{X}_2 \geq 1\right) = P\left(\bar{X}_1 - \bar{X}_2 \geq 1\right)$$
$$= P\left(Z \geq \frac{1 - 0.5}{\sqrt{0.03556}}\right)$$
$$= P(Z \geq 2.6515)$$
$$= 0.00402$$

9.

(a) Lectures - Sampling Distribution of Difference Between Two Proportions

i. Sample probability of successes

A. $\hat{P}_1 = \frac{x_1}{n_1}$

B. $\hat{P}_2 = \frac{x_2}{n_2}$

ii. Mean

A. $\hat{P}_1 - \hat{P}_2 = \mu_{\hat{p}_1 - \hat{p}_2} = E\left[\hat{P}_1 - \hat{P}_2\right] = p_1 - p_2$

B. Variance $= \hat{P}_1 - \hat{P}_2 = \sigma^2_{\hat{P}_1 - \hat{P}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$

(b) Question:

i. $P_1 = 0.08, n_1 = 150$

ii. $P_2 = 0.05, n_2 = 300$

(c) Answer

i.

$$\mu_{\hat{P}_1 - \hat{P}_2} = 0.08 - 0.05 = 0.03$$

ii.

$$\begin{aligned} \sigma^2_{\hat{P}_1 - \hat{P}_2} &= \frac{0.08 * 0.92}{150} + \frac{0.05 * 0.95}{300} \\ &= 6.49 \times 10 - 4 \end{aligned}$$

iii.

$$\begin{aligned} P\left(\left|\hat{P}_1 - \hat{P}_2\right| \leq 0.01\right) &= P\left(-0.01 \leq \hat{P}_1 - \hat{P}_2 \leq 0.01\right) \\ &= P\left(\frac{-0.01 - 0.03}{\sqrt{6.49 \times 10^{-4}}} \leq \frac{\left(\hat{P}_1 - \hat{P}_2\right) - p_2 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \leq \frac{0.01 - 0.03}{\sqrt{6.49 \times 10^{-4}}}\right) \\ &= P\left(\frac{-0.01 - 0.03}{\sqrt{6.49 \times 10^{-4}}} \leq \frac{\left(\hat{P}_1 - \hat{P}_2\right) - p_2 - p_2}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \leq \frac{0.01 - 0.03}{\sqrt{6.49 \times 10^{-4}}}\right) \\ &= P\left(-1.57 \leq Z \leq -0.785\right) \\ &= 0.2162 - 0.0582 \\ &= 0.158 \end{aligned}$$

10.

(a) Summary

Summary

Sampling distributions	Z value
$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$	$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$
$\hat{P} \sim N\left(p, \frac{p(1-p)}{n}\right)$	$Z = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$
$\overline{X}_1 - \overline{X}_2 \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$	$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
$\hat{P}_1 - \hat{P}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$	$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$

i. First is sample mean

ii. Second is population

iii. Third is proportion of sample mean

iv. Fourth is proportion of population

A. Normal distribution if the sample size is big

(b) **Exam formulas**

i. No need to memorize

(c) **Example**

i. $\mu = 300, \sigma = 2, n = 100$

ii. Since it is at exactly center $a_2 = a_1$ and $a_1 = -a_1$

iii.

$$P(a_1 \leq X \leq a_2) = 0.9$$
$$P(-a \leq X \leq a) = 0.9$$
$$P\left(\frac{-a - 300}{\frac{2}{\sqrt{100}}} < Z < \frac{a - 300}{\frac{2}{\sqrt{100}}}\right) = 0.9$$
$$Fromtable, P(-1.645 < Z < 1.645) = 0.09$$

$$a_1, a_2 = 300 \pm 1.645 * \frac{2}{\sqrt{100}} = 300 \pm 0.329$$

iv. Conclusion: The limits are $300 + 0.329$ and $300 - 0.329$