Tutorial 6

CB2200 Business Statistics, YANG Yihang

October 20, 2019

Intended Learning Outcomes

After this tutorial, you may know how to

- distinguish sample mean with mean of sample mean.
- find the distribution of sample mean from statistics, such as mean and standard deviation.
- apply central limit theorem in solving specific problems.

Recap

- Sampling distribution
- The distribution of sample mean is generated by sampling from the population.
- The population mean of X and the population mean of $\bar{X}$ are identical
- Central limit theorem: As sample size gets large enough, sample mean distribution becomes Normal regardless of population distribution

Recap

- Mean of sample means
  - $\mu_{\bar{X}} = \mu$
  - Works for sampling with and without replacement if the samples are unbiased
- Standard deviation of sample means
  - $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
  - Also called standard error of the mean
  - Works for sampling with replacement, or sampling from large populations without replacement
  - As n increases, $\sigma_{\bar{X}}$ decreases

Question 1

Q1

Given a normal distribution with $\mu = 100$ and $\sigma = 12$, if you select a sample of $n = 36$, what is the probability that $\overline{X}$ is

a) Less than 95?

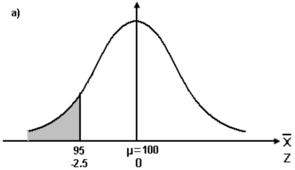
b) Between 95 and 97.5?

c) Above 102.2?

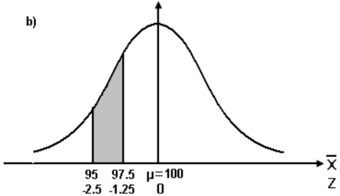
d) There is a 65% chance that $\overline{X}$ is above what value?

Q1

Since the population is normal [$X \sim N(100, 12^2)$], the sampling distribution of the mean is also normal [$\overline{X} \sim N(100, (\frac{12}{\sqrt{36}})^2)$]

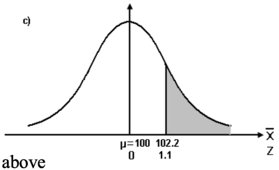


a) $$P(\overline{X} < 95) = P(Z < \frac{95-100}{12/\sqrt{36}}) = P(Z < -2.5) = 0.0062$$



b) $$P(95 < \overline{X} < 97.5) = P(\frac{95-100}{12/\sqrt{36}} < Z < \frac{97.5-100}{12/\sqrt{36}}) = P(-2.5 < Z < -1.25) = 0.1056 - 0.0062 = 0.0994$$

c) $P(\overline{X} > 102.2) = P(Z > \frac{102.2 - 100}{12/\sqrt{36}})$
$= P(Z > 1.1)$
$= 1 - 0.8643$
$= 0.1357$



d) Let Xo be the value that 65 % chance that $\overline{X}$ is above

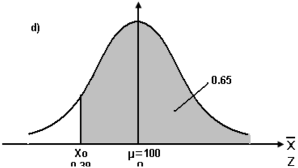
$$P(\overline{X} > x_0) = 0.65$$

$$P(Z > \frac{x_0 - 100}{12/\sqrt{36}}) = 0.65$$

Since $P(Z < -0.39) = 0.35$,

$$\frac{x_0 - 100}{12/\sqrt{36}} = -0.39$$

$x_0 = 99.22$



Q2

The diameter of a brand of Ping-Pong balls is normally distributed, with a mean of 1.30 inches and a standard deviation of 0.05 inch. If you select a random sample of 25 Ping-Pong balls,

a) What is the sampling distribution of the mean?
b) What is the probability that the sample mean is less than 1.28 inches?
c) What is the probability that the sample mean is between 1.31 and 1.33 inches?
d) The probability is 60% that the sample mean will be between what two values, symmetrically distributed around the population mean?

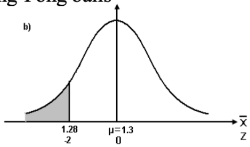
Q2

a) Since the population is normal [$X \sim N(1.30, 0.05^2)$], the sampling distribution of the mean is also normal [$\overline{X} \sim N(1.30, (\frac{0.05}{\sqrt{25}})^2)$]

b) Let $\overline{X}$ be the sample mean of the diameter of a brand of Ping-Pong balls

$$P(\overline{X} < 1.28) = P(Z < \frac{1.28 - 1.30}{0.05/\sqrt{25}})$$

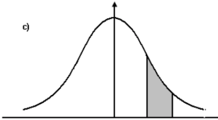
$$= P(Z < -2) = 0.0228$$



Question 2

c) $P(1.31 < \overline{X} < 1.33) = P(\frac{1.31 - 1.30}{0.05/\sqrt{25}} < Z < \frac{1.33 - 1.30}{0.05/\sqrt{25}})$

$= P(1 < Z < 3) = 0.99865 - 0.8413 = 0.15735$



d) Let Xlower be the lower value and Xupper be the upper value

$P(x_{lower} < \overline{X} < x_{upper}) = 0.60$

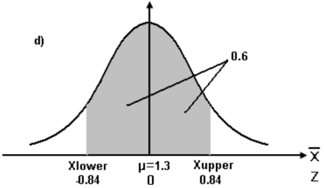
$P(\frac{x_{lower} - 1.30}{0.05/\sqrt{25}} < Z < \frac{x_{upper} - 1.30}{0.05/\sqrt{25}}) = 0.60$

For symmetric distribution of probability (0.6) on both sides of μ,

Since $P(Z < -0.84) = 0.2$ and $P(Z < 0.84) = 0.8$,

$$\begin{cases} \frac{x_{lower} - 1.30}{0.05/\sqrt{25}} = -0.84 \\ \frac{x_{upper} - 1.30}{0.05/\sqrt{25}} = 0.84 \end{cases}$$

$$\begin{cases} x_{lower} = -0.84(\frac{0.05}{\sqrt{25}}) + 1.30 = 1.2916 \\ x_{upper} = 0.84(\frac{0.05}{\sqrt{25}}) + 1.30 = 1.3084 \end{cases}$$



Hence, 60% of the sample means will be between 1.2916 inches and 1.3084 inches.

Question 3

Q3

Time spent using e-mail per session is normally distributed with $\mu = 8$ minutes and $\sigma = 2$ minutes. If you select a random sample of 16 sessions,

a) What is the probability that the sample mean is between 7.8 and 8.2 minutes?

b) What is the probability that the sample mean is between 7.5 and 8 minutes?

c) If you select a random sample of 100 sessions, what is the probability that the sample means is between 7.8 and 8.2 minutes?

d) Explain the difference in the results of (a) and (c).

Question 3

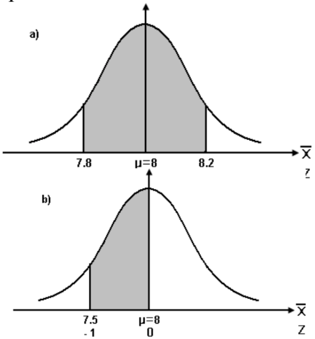
Q3

Since the population is normal [$X \sim N(8, 2^2)$], the sampling distribution of the mean is also normal [$\overline{X} \sim N(8, (\frac{2}{\sqrt{16}})^2)$]

a) Let $\overline{X}$ be the sample mean of the time spent using e-mail per session

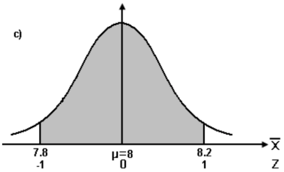
$$P(7.8 < \overline{X} < 8.2) = P(\frac{7.8-8}{2/\sqrt{16}} < Z < \frac{8.2-8}{2/\sqrt{16}})$$
$$= P(-0.4 < Z < 0.4)$$
$$= 0.6554 - 0.3446$$
$$= 0.3108$$

b) $$P(7.5 < \overline{X} < 8) = P(\frac{7.5-8}{2/\sqrt{16}} < Z < \frac{8-8}{2/\sqrt{16}})$$
$$= P(-1 < Z < 0)$$
$$= 0.5 - 0.1587$$
$$= 0.3413$$

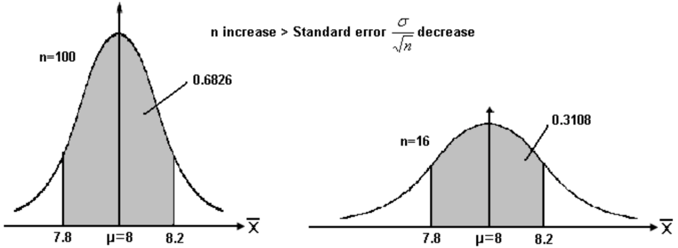


c) $n = 100$

$$P(7.8 < \overline{X} < 8.2) = P(\frac{7.8-8}{2/\sqrt{100}} < Z < \frac{8.2-8}{2/\sqrt{100}})$$
$$= P(-1 < Z < 1)$$
$$= 0.8413 - 0.1587$$
$$= 0.6826$$



d) With the sample size increasing from n = 16 to n = 100, more sample means will be closer to the population mean. The standard error of the sample mean of size 100 is much smaller than that of size 16, so the likelihood(probability) that the sample mean will fall within 0.2 minutes of the population mean is much higher for samples of size 100 (probability = 0.6826) than for samples of size 16 (probability = 0.3108).



Question 4

Q4
In a recent survey concerning the age (to the nearest year) and weight (to the nearest 10 lb) of first-year university students, the following probability distribution was obtained:

Age	Weight				
	100	110	120	130	140
19	0.02	0.09	0.09	0.01	0.02
20	0.06	0.15	α	0.05	0.03
21	0.02	0.06	0.11	0.04	0.05

A sample of 36 first-year students is taken. Find the approximate chance that their total weight is at most 4350 lb.

Q4

Let X be the weight of the student

$\because n = 36 > 30$

By Central Limit Theorem, the Sampling Distribution of Mean is normal.

$$\overline{X} \sim N(118, \left(\frac{10.77}{\sqrt{36}}\right)^2)$$

$$\Pr(\text{Total weight of 36 students} \le 4350) = \Pr(\overline{X} \le \frac{4350}{36}) = \Pr\left(Z \le \frac{120.83 - 118}{10.77 / \sqrt{36}}\right)$$

$$= \Pr(Z \le 1.58) = 0.9429$$

Q5
At the CityU Computer Service Centre, the loading time for e-Portal page on Internet Explorer is normally distributed with mean 3 seconds.

A random sample of 5 computers is drawn. What is the chance that their total loading time is at least 15 seconds?

Question 5

Q5

Let $\bar{X}$ be the average loading time

Since the population is normally distribution, sampling distribution of mean is approximately normal, $\bar{X} \sim N(3, \left(\frac{\sigma}{\sqrt{5}}\right)^2)$

P(Total time of 5 computers $\geq$ 15)

$$= P(\bar{X} \geq \frac{15}{5}) = P(\bar{X} \geq 3) = P(Z \geq \frac{3-3}{\frac{\sigma}{\sqrt{5}}}) = P(Z \geq 0) = 0.5$$

Q6

Suppose there is a population with population size N = 3. The variable of interest is the salary (X) of individuals. The values of X are 18, 20 and 22 (in thousand dollars).

a) Find the mean (μ) and standard deviation (σ) for the population distribution.

In the process of developing sampling distribution, all possible samples (taken with replacement) of size n = 2 are obtained. The sample mean ($\bar{X}$) is considered as the sample statistic.

b) What are the possible values of this sample mean random variable? Develop the probability distribution of the sample mean.

c) Show that the sample statistic $\bar{X}$ is an unbiased estimator of μ.

d) Denote $\sigma_{\bar{X}}$ the standard deviation of $\bar{X}$, verify the following relationship: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

e) Does the sampling distribution of $\bar{X}$ follows a Normal Distribution? Explain.

Q6

a) $\mu = 20, \ \sigma = 1.63$

b) Possible values and probability distribution of $\overline{X}$

$\overline{X}$	P($\overline{X}$)
18	1/9
19	2/9
20	3/9
21	2/9
22	1/9

c) $\mu = 20, \ \mu_{\overline{X}} = 20$, so $\mu_{\overline{X}} = \mu$

d) $\sigma_{\overline{X}} = 1.15 \ , \ \dfrac{\sigma}{\sqrt{n}} = 1.15$, so $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

e) The sampling distribution of $\overline{X}$ does not follow a Normal Distribution.
Reasons: - population distribution of X is not normal
- n < 30, cannot apply Central Limit Theorem

Question 7

Q7

To investigate the length of time working for an employer, researchers at the CityU sampled 344 business students and asked them a question: Over the course of your lifetime, what is the maximum number of years you expect to work for any one employer? The resulting sample had sample mean $\overline{X}$=19.1 years and sample standard deviation s=6 years. Assume the sample of students was randomly selected from the 5800 undergraduate students in CityU.

a) What are reasonable estimators of population mean and population standard deviation?
b) What is the sampling distribution of $\overline{X}$? Why?
c) If the population mean was 18.5 years, what is P($\overline{X} \geq$19.1 years)?
d) If the population mean was 19.5, what is P($\overline{X}$=19.1 years)?
e) If P($\overline{X} \geq$19.1 years) = 0.5, what is the population mean?
f) If P($\overline{X} \geq$19.1 years) = 0.2, without calculation, can you tell that the population mean is greater or less than 19.1 years? Explain.

Q7

a) Sample mean and sample standard deviation are reasonable estimators of population mean and population standard deviation.

b) According to Central Limit Theorem, with large sample size (n = 344), sample mean $\bar{x}$ follows a normal distribution approximately, with mean 19.1 and s.d. $6/\sqrt{344} = 0.3235$

c) $P(\bar{x} \geq 19.1 \text{ years}) = 1 - P(\bar{x} < 19.1 \text{ years})$

$$= 1 - P(z < \frac{19.1 - 18.5}{6/\sqrt{344}}) = 1 - P(z < 1.8547) = 1 - 0.9678 = 0.0322$$

d) $P(\bar{x} = 19.1 \text{ years}) = 0$

e) $P(\bar{x} \geq 19.1 \text{ years}) = 0.5 \Rightarrow P(\bar{x} < 19.1 \text{ years}) = 0.5$

$$\Rightarrow \frac{19.1 - \mu}{6/\sqrt{344}} = 0 \Rightarrow \mu = 19.1 \text{ years}$$

f) Yes. The population mean will be less than 19.1 years because for a normal distribution mean is equal to median. However, now we observe $P(\bar{x} \geq 19.1 \text{ years}) = 0.2 < 0.5$. Therefore, the population mean should be less than 19.1 years.