Topic 4: Sampling Distributions Solutions

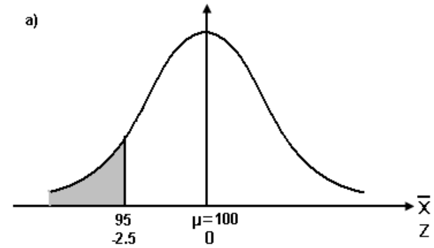
Q1

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

$= P(Z < -2.5)$

$= 0.0062$

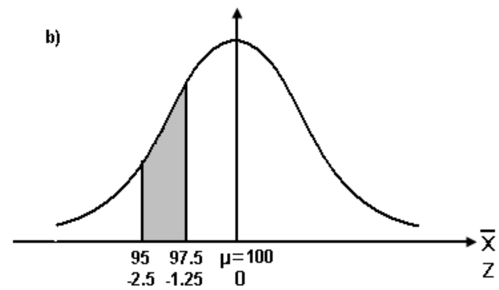


b) $P(95 \leq \overline{X} \leq 97.5) = P(\frac{95-100}{12/\sqrt{36}} \leq Z \leq \frac{97.5-100}{12/\sqrt{36}})$

$= P(-2.5 \leq Z \leq -1.25)$

$= 0.1056 - 0.0062$

$= 0.0994$

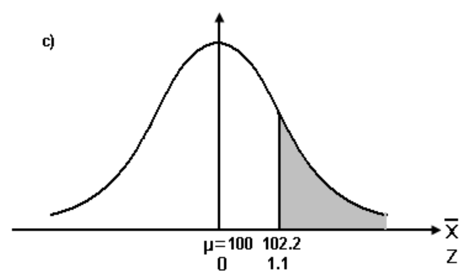


c) $P(\overline{X} > 102.2) = P(Z > \frac{102.2-100}{12/\sqrt{36}})$

$= P(Z > 1.1)$

$= 1 - 0.8643$

$= 0.1357$



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

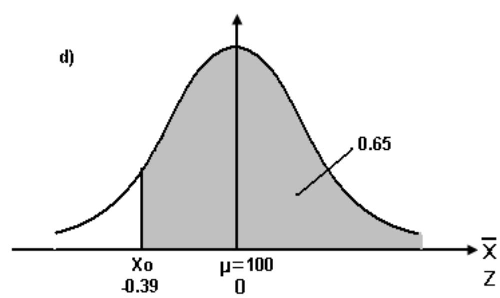
$P(\overline{X} > x_0) = 0.65$

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

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

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

$x_0 = 99.22$



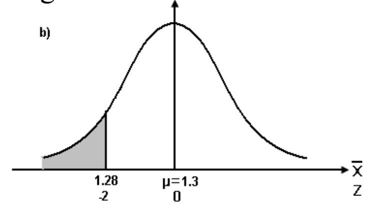
Q2

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

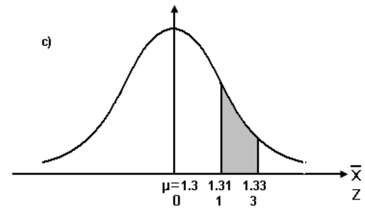
b) Let $\overline{X}$ be the sample mean of the diameter of a brand of Ping-Pong balls

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

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

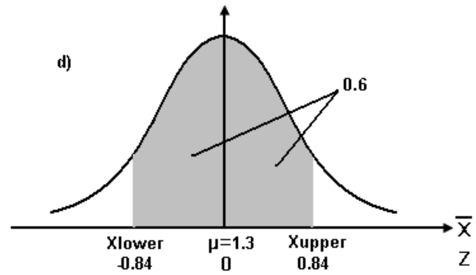


c) $P(1.31 \le \overline{X} \le 1.33) = P(\frac{1.31-1.30}{0.05/\sqrt{25}} \le Z \le \frac{1.33-1.30}{0.05/\sqrt{25}})$
$= P(1 \le Z \le 3) = 0.99865 - 0.8413 = 0.15735$



d) Let Xlower be the lower value and Xupper be the upper value
$P(x_{lower} \le \overline{X} \le x_{upper}) = 0.60$
$P(\frac{x_{lower} - 1.30}{0.05/\sqrt{25}} \le Z \le \frac{x_{upper} - 1.30}{0.05/\sqrt{25}}) = 0.60$
For symmetric distribution of probability (0.6) on both sides of μ,
Since $P(Z \le -0.84) = 0.2$ and $P(Z \le 0.84) = 0.8$,

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

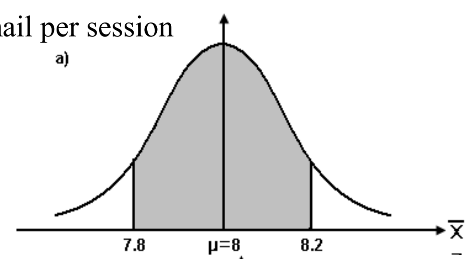


Hence, 60% of the sample means will be between 1.2916 inches and 1.3084 inches.

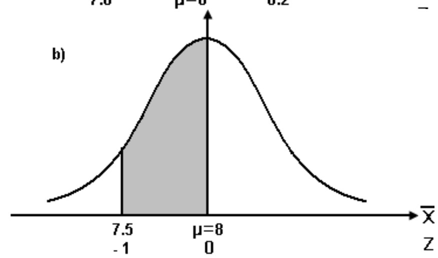
Q3

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

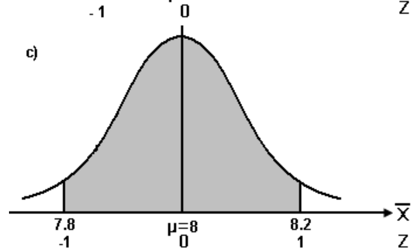
a) Let $\overline{X}$ be the sample mean of the time spent using e-mail per session
$P(7.8 \le \overline{X} \le 8.2) = P(\frac{7.8-8}{2/\sqrt{16}} \le Z \le \frac{8.2-8}{2/\sqrt{16}})$
$= P(-0.4 \le Z \le 0.4)$
$= 0.6554 - 0.3446$
$= 0.3108$



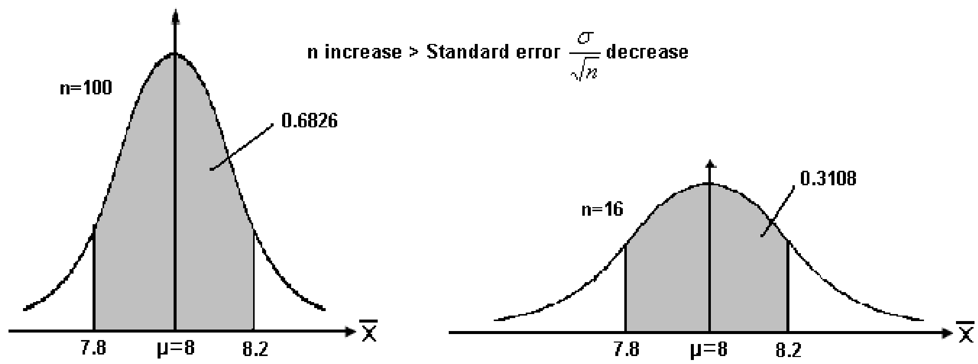
b) $P(7.5 \le \overline{X} \le 8) = P(\frac{7.5-8}{2/\sqrt{16}} \le Z \le \frac{8-8}{2/\sqrt{16}})$
$= P(-1 \le Z \le 0)$
$= 0.5 - 0.1587$
$= 0.3413$



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



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).



Q4

Let X be the weight of the student

$\because n = 36 > 30$

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

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

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

$= P(Z \leq 1.58) = 0.9429$

Q5

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

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

P(Total time of 5 computers ≥ 15)

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

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

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

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 standard error $6/\sqrt{344} = 0.3235$ (as estimated by the sample mean and sample standard deviation)

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.

Q8

a) Sample proportion, $\hat{p} = \frac{\text{no. of students who own shares of stock}}{\text{sample size}(n)}$

$= 8/40 = 0.2$

b) Standard error of the proportion, $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.2(1-0.2)}{40}} = 0.0632$

Q9

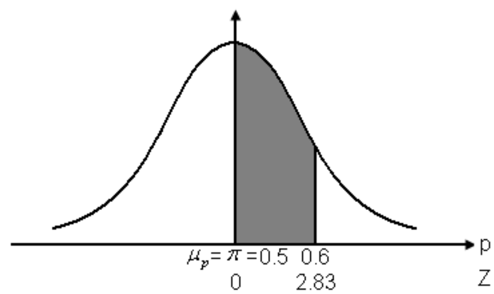
a) $\because$ An individual has no ability to distinguish between the two brands.

$\therefore$ p = proportion of students which can distinguish the brand = 0.5

n = 200 > 30, n p=200(0.5)= 100 > 5, n(1-p)=200(1-0.5)=100 > 5

So sampling distribution of $\hat{p}$ is approximately normal

$$\hat{p} \sim N\left(p, \frac{p(1-p)1}{n}\right) \Rightarrow \hat{p} \sim N\left(0.5, \frac{0.5(1-0.5)}{200}\right)$$



$$P(0.5 \le \hat{p} \le 0.6) = P\left(\frac{0.5 - p}{\sqrt{\frac{\pi(1-p)}{n}}} \le Z \le \frac{0.6 - p}{\sqrt{\frac{\pi(1-p)}{n}}}\right)$$

$$= P\left(\frac{0.5 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}} \le Z \le \frac{0.6 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}}\right)$$

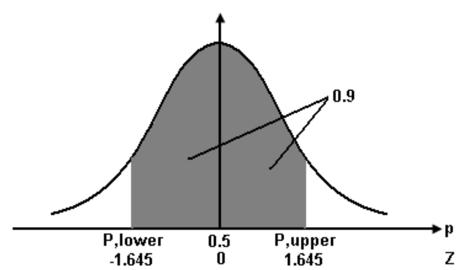
$= P(0 \le Z \le 2.83)$

$= 0.9977 - 0.5$

$= 0.4977$

b) $P(\hat{p}_{lower} \le \hat{p} \le \hat{p}_{upper}) = 0.90$

$$P\left(\frac{\hat{p}_{lower} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}} \le Z \le \frac{\hat{p}_{upper} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}}\right) = 0.90$$



For symmetric distribution of probability (0.9) on both sides,

$$P\left(Z \le \frac{\hat{p}_{lower} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}}\right) = 0.05 \text{ and } P\left(Z \le \frac{\hat{p}_{upper} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}}\right) = 0.95$$

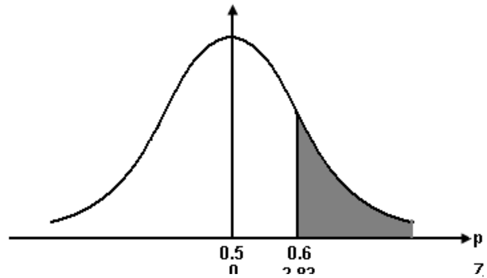
Since $P(Z \le -1.645) = 0.05$ and $P(Z \le 1.645) = 0.95$,

$$\begin{cases} \dfrac{\hat{p}_{lower} - 0.5}{\sqrt{\dfrac{0.5(1-0.5)}{200}}} = -1.645 \\ \dfrac{\hat{p}_{upper} - 0.5}{\sqrt{\dfrac{0.5(1-0.5)}{200}}} = 1.645 \end{cases} \qquad \begin{cases} \hat{p}_{lower} = -1.645\left(\sqrt{\dfrac{0.5(1-0.5)}{200}}\right) + 0.5 = 0.4418 \\ \hat{p}_{upper} = 1.645\left(\sqrt{\dfrac{0.5(1-0.5)}{200}}\right) + 0.5 = 0.5582 \end{cases}$$

Hence, 90% of the sample proportion will be between 0.4418 and 0.5582.

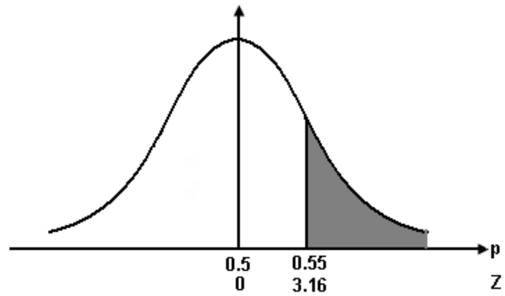
c) $P(\hat{p} > 0.65) = P(Z > \frac{0.65 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}}) = P(Z > 4.24) \approx 0$

d) se 1: $P(\hat{p} > 0.6) = P(Z > \frac{0.6 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}})$



$= P(Z > 2.83)$
$= 1 - 0.9977$
$= 0.0023$

Case 2: $P(\hat{p} > 0.55) = P(Z > \frac{0.55 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{1000}}})$



$= P(Z > 3.16)$
$= 1 - 0.99921$
$= 0.00079$

More than 60% correct identification in a sample of 200 is more likely to occur than more than 55% correct in a sample of 1000.