Sampling Distribution

The notion of **sampling** is considered in many cases, like if $X =$ Length of life of a light bulb

Here, it is impossible to observe all values of X in order to provide information to customers the necessary information about the mean of light bulb length of life in the population. Therefore, we must depend on selecting representative samples of size n: $\boldsymbol{X_1, X_2, \ldots, X_n}$ from the population to make the desired inference concerning the entire population. The inference is usually based on some suitable **statistics** defined as measures on the samples, say:

The sample mean: $\overline{\boldsymbol{X}}$ or the sample variance $\boldsymbol{S^2}$ or the sample proportion $\hat{\boldsymbol{p}}$ or any other options.

Non representative sampling is called **biased**, and this defective is eliminated by selecting **random samples**, in the sense that observations are made independently and at random.

In general, sampling scheme aims at providing information about the population parameters, like;

mean $\boldsymbol{\mu}$**,** variance $\boldsymbol{\sigma^2}$ and proportion $\boldsymbol{P}$

Sampling Distribution of means $\overline{\boldsymbol{X}}$ is the population of all observed values $\bar{x}$ under same conditions. Moreover, each observation of $\overline{\boldsymbol{X}}$ has a ***pdf*** $f_{\overline{X}}(.)$, a mean $\mu_{\overline{X}}$ (or $E(\overline{\boldsymbol{X}})$**)** and a variance $\sigma^2_{\overline{X}}$(or $V(\overline{\boldsymbol{X}})$**)**.

Theorem: Let $f_X(.)$ is the population of a variable $\boldsymbol{X}$ with mean μ and variance σ^2, then the sampling distribution of $\overline{X}$, $f_{\overline{X}}(.)$, will have the following mean and variance:

$$\mu_{\overline{X}} = E(\overline{\boldsymbol{X}}) = \mu \qquad \sigma^2_{\overline{X}} = V(\overline{\boldsymbol{X}}) = \sigma^2/n$$

The square root $\sigma_{\overline{X}}$ is called the standard error (***s.e***) of $\overline{\boldsymbol{X}}$**.**

The population $f_{\overline{X}}(.)$ is covered by the next corner stone theorem in statistic.

Central Limit Theorem (*c.l.t*)

If $\boldsymbol{\overline{X}}$ is the mean of the random sample $\boldsymbol{X_1, X_2, \ldots, X_n}$ taken from any population $f_{\boldsymbol{X}}(.)$ with mean μ and finite variance σ^2 then, if the sample size $\boldsymbol{n}$ is large, the distribution of:

$$\boldsymbol{Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}}$$

is approximately standard normal $N(0,1)$.

In other words ***c.l.t*** says as $\boldsymbol{n}$ is large enough that it is **approximately** true:

$$\boldsymbol{Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \Leftrightarrow f_{\overline{X}}(.) = N(\mu, \sigma^2/n)}$$

And it will be accepted that $\boldsymbol{n}$ is large enough if $\boldsymbol{n \geq 30.}$

Remark: approximation becomes exact at any size $\boldsymbol{n}$ if $f_{\boldsymbol{X}}(.)$ is normal.

Q. The average life of a certain battery is 5 years, with a standard deviation of 1 year. Assume that the live of the battery approximately follows a normal distribution.

1) The sample mean $\overline{X}$ of a random sample of 5 batteries selected from this product has a mean $E(\overline{X}) = \mu_{\overline{X}}$ equal to:

(A) 0.2 (B) 5 (C) 3 (D) None of these

2) The variance $Var(\overline{X}) = \sigma^2_{\overline{X}}$ of the sample mean $\overline{X}$ of a random sample of 5 batteries selected from this product is equal to:

(A) 0.2 (B) 5 (C) 3 (D) None of these

3) The probability that the average life of a random sample of size 16 of such batteries will be between 4.5 and 5.4 years is:

(A) 0.1039 (B) 0.2135 (C) 0.7865 (D) 0.9224

4) The probability that the average life of a random sample of size 16 of such batteries will be less than 5.5 years is:

(A) 0.9772 (B) 0.0228 (C) 0.9223 (D) None of these

Q. The student heights in a secondary school has a mean of 174.5 ***c.m*** and a standard deviation of 6.9 ***c.m***. Suppose 200 random samples of size 36 are drawn from this population and the means recorded to the nearest tenth of a centimeter. Determine:

(a) The mean and standard deviation of the sampling distribution of $\bar{X}$;

(b) The expected number of sample means that fall between 172.5 and 175.8 centimeters inclusive;

(c) The expected number of sample means falling below 173.0 centimeters.

Solution:

a) $E(\bar{X}) = \mu_{\bar{X}} = \mu = 174.5, \quad V(\bar{X}) = \frac{6.9}{\sqrt{36}} = 1.15$

b)$P(172.5 < \bar{X} < 175.8) = P\left(\frac{172.5-174.5}{\frac{6.9}{\sqrt{36}}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{175.8-174.5}{\frac{6.9}{\sqrt{36}}}\right)$

$$= P(-1.74 < Z < 1.13) = 0.8708 - 0.0409 = 0.8299$$

The expected number of sample means that fall between 172.5 and 175.8 centimeters inclusive=200*0.8299=165.98≈ 166

c)$P(\bar{X} < 173.0) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{172-174.5}{6.9/\sqrt{25}}\right) = P(Z < -1.30) = 0.0968$

The expected number of sample means falling below 173.0 centimeters=200*0.0968=19.36≈ 19

Special cases of $f_{\overline{X}}(.)$:

If σ^2 is unknown and is estimated by the sample variance s^2 , then:

- If $n \geq 30$, ***c.l.t*** is not broken down and it is still **approximately** true:

$$Z = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim N(0,1) \Leftrightarrow f_{\overline{X}}(.) = N\left(\mu, \frac{s^2}{n}\right)$$

  And $s_{\overline{X}} = s/\sqrt{n}$ is called the estimate of ***s.e*** of $\overline{X}$**(i.e** $s_{\overline{X}}$ is estimate of $\sigma_{\overline{X}}$)

- If $n < 30$ but $f_X(.)$ is assumed normal, then the statistic: $t = \frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t_{(n-1)}$ the student t-distribution with degree of freedom $df = (n-1)$.

Q. A random sample of size *n* = 36 is taken from a population with a mean μ = 70. If the sample standard deviation equals to ***s*** = 4.3. Let $\bar{X}$ be the sample mean, then find P (70 < $\bar{X}$ < 71).

Solution:

$$E(\bar{X}) = \mu = 70,\ s^2{}_{\bar{X}} = \frac{s^2}{n} = \frac{18.49}{36} = 0.5136$$

$$P(70 < \bar{X} < 71) = P\left(\frac{70 - 70}{\sqrt{0.5136}} < \frac{\overline{X} - \mu}{s/\sqrt{n}} < \frac{71 - 70}{\sqrt{0.5136}}\right) = P(0 < Z < 1.4) = 0.4192$$

Sampling Distribution of $\hat{p}$ is no more than a special case of $f_{\overline{X}}(.)$ when $f_X(.)$ is a Bernoulli population, i.e., $X = 0, 1$ with mean $\mu = P$ the proportion of somewhat, say proportion of smokers, and with variance:$\sigma^2 = P(1-P)$. In this special case the sample mean is equal to:

$$\overline{X} = \hat{p}$$

and ***c.l.t*** says also as n is large enough that it is **approximately** true:

$$Z = \frac{\hat{p}-P}{\sqrt{P(1-P)/n}} \sim N(0,1) \Leftrightarrow f_{\hat{p}}(.) = N(P, \frac{P(1-P)}{n})$$

Moreover, the approximation is better here if both $\boldsymbol{n\,P} \geq \mathbf{5}$ and $\boldsymbol{n(1-P)} \geq \mathbf{5}.$

Q: Suppose that you take a random sample of size n=100 from a population with proportion of diabetic equal to $\boldsymbol{p}$ =0.25. Let $\hat{p}$ be the sample proportion of of diabetic.

(a) What is the mean and the standard error of $\hat{\boldsymbol{p}}$?

(b) What is the approximated sampling distribution of $\hat{p}$?

(c) Find the probability that the sample proportion $\hat{p}$ is less than 0.2.

Solution:

a) The mean of $\hat{p} = 0.25$ & $\sqrt{Var(\hat{p})} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.25*0.75}{100}} = 0.0433$

b) $\hat{p} \sim N(0.25, 0.0433^2)$

c) $P(\hat{P} < 0.2) = P\left(Z < \frac{0.2-0.25}{0.0433}\right) = P(Z < -1.15) = 0.1251$

Sampling Distribution of the difference between tow means $\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2$:

Indeed, the population $f_{\overline{X}}(.)$ can be similarly generated to the population $f_{\overline{X}_1-\overline{X}_2}(.)$ of all observed values $\bar{x}_1 - \bar{x}_2$ under the same conditions, where $\bar{x}_1$and $\bar{x}_2$are observations of $\overline{\boldsymbol{X}}_1$and $\overline{\boldsymbol{X}}_2$ the sample means of two independent random samples of sizes $\boldsymbol{n_1}$and $\boldsymbol{n_2}$ from two populations f_1 and f_2 each of which has mean $\mu_1$and μ_2 and variance $\sigma_1^2$and σ_2^2, respectively.

The population $f_{\overline{X}_1-\overline{X}_2}(.)$ will have the following mean and variance:

$$\boldsymbol{\mu_{\overline{X}_1-\overline{X}_2} = E(\overline{X}_1 - \overline{X}_2) = \mu_1 - \mu_2}$$

$$\boldsymbol{\sigma^2_{\overline{X}_1-\overline{X}_2} = V(\overline{X}_1 - \overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

The square root $\sigma_{\overline{X}_1-\overline{X}_2}$ is called the standard error (***s.e***) of $\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2$**.**

Central Limit Theorem (*c.l.t*) concerning $\overline{X}_1 - \overline{X}_2$

Is also valid for any two populations f_1 and f_2 each of which has mean μ_1 and μ_2 and finite variance σ_1^2 and σ_2^2 respectively, if the sample sizes $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ are large, then the distribution of:

$$\boldsymbol{Z} = \frac{(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) - (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})}{\sqrt{\frac{\boldsymbol{\sigma_1^2}}{\boldsymbol{n_1}} + \frac{\boldsymbol{\sigma_2^2}}{\boldsymbol{n_2}}}}$$

is approximately standard normal $N(0,1)$.

In other words ***c.l.t*** says as $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ are large enough that it is **approximately** true:

$$\boldsymbol{Z} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \boldsymbol{N(0,1)} \Leftrightarrow \; f_{\overline{X}_1 - \overline{X}_2}(.) = \boldsymbol{N}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}, \frac{\boldsymbol{\sigma_1^2}}{\boldsymbol{n_1}} + \frac{\boldsymbol{\sigma_2^2}}{\boldsymbol{n_2}})$$

And it will be accepted that $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ are large enough if $\boldsymbol{n_1}$ and $\boldsymbol{n_2} \geq \boldsymbol{30.}$

Actually, approximation becomes exact at any sizes $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ if f_1 and f_2 are normal.

Q. A random sample of size $n_1 = 36$ is taken from a normal population with a mean $\mu_1 = 70$ and a standard deviation $\sigma_1 = 4$. A second independent random sample of size $n_2 = 49$ is taken from a normal population with a mean $\mu_2 = 85$ and a standard deviation $\sigma_2 = 5$. Let $\overline{\boldsymbol{X}}_1$ and $\overline{\boldsymbol{X}}_2$ be the averages of the first and second samples, respectively.

Find $E(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)$ & $Var(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)$ **&** $P(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 > \boldsymbol{-16})$

Solution:

$$E(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) = \mu_1 - \mu_2 = 70 - 85 = -15 \quad \& \quad Var(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) = \frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2} = \frac{16}{36} + \frac{25}{49} = 0.9546$$

$$P(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 > \boldsymbol{-16}) = P\left(Z > \frac{-16 + 15}{\sqrt{0.9546}}\right) = 1 - P(Z < 1.02) = 0.1539$$

Q: The distribution of heights of a certain breed of terrier has a mean of 72 centimeters and a standard deviation of 10 centimeters, whereas the distribution of heights of a certain breed of poodle has a mean of 28 centimeters with a standard deviation of 5 centimeters. Assuming that the sample means can be measured to any degree of accuracy, find the probability that the sample mean for a random sample of heights of 64 terriers exceeds the sample mean for a random sample of heights of 100 poodles by at most 44.2 centimeter

Solution: $n_1 = 64, \mu_1 = 72, \sigma_1 = 10$ & $n_2 = 100, \mu_2 = 28, \sigma_2 = 5$

$$P(\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}} < 44.2) = P\left(Z < \frac{44.2 - (72 - 28)}{\sqrt{\frac{100}{64} + \frac{25}{100}}}\right) = P(Z < 0.15) = 0.5596$$

Special cases of $\boldsymbol{f}_{\overline{X}_1 - \overline{X}_2}(.)$

if σ_1^2 and σ_2^2 are unknown and are estimated by the sample variances s_1^2 and s_2^2, then:

- If $\boldsymbol{n_1}$ and $\boldsymbol{n_2} \geq \boldsymbol{30}$, ***c.l.t*** is not broken down and it is **approximately** true::

$$\boldsymbol{Z} = \frac{(\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}) - (\boldsymbol{\mu_1} - \boldsymbol{\mu_2})}{\sqrt{\frac{s_1^2}{\boldsymbol{n_1}} + \frac{s_2^2}{\boldsymbol{n_2}}}} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{1}) \Leftrightarrow f_{\overline{X}_1 - \overline{X}_2}(.) = \boldsymbol{N}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}, \frac{s_1^2}{\boldsymbol{n_1}} + \frac{s_2^2}{\boldsymbol{n_2}})$$

Where $\sqrt{\frac{s_1^2}{\boldsymbol{n_1}} + \frac{s_2^2}{\boldsymbol{n_2}}}$ is an estimation of ***s.e*** of $\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}$

- If $\boldsymbol{n_1}$ or $\boldsymbol{n_2} < \boldsymbol{30}$ but f_1 and f_2 are normal, then it is distinguished between the two cases:

    1- $\sigma_1^2 = \sigma_2^2$

    2- $\sigma_1^2 \neq \sigma_2^2$

If the first case is only considered, and the sample variances s_1^2 and s_2^2 should be pooled together as following $s_p^2 = \frac{(\boldsymbol{n_1} - 1)s_1^2 + (\boldsymbol{n_2} - 1)s_2^2}{(\boldsymbol{n_1} + \boldsymbol{n_2} - \boldsymbol{2})}$ and $\sqrt{\frac{s_p^2}{\boldsymbol{n_1}} + \frac{s_p^2}{\boldsymbol{n_2}}} = \boldsymbol{s_p}\sqrt{\frac{1}{\boldsymbol{n_1}} + \frac{1}{\boldsymbol{n_2}}}$ is an estimate of ***s.e*** of $\overline{\boldsymbol{X}}_{\boldsymbol{1}} - \overline{\boldsymbol{X}}_{\boldsymbol{2}}$.

Eventually, it is used the statistic: $= \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim \boldsymbol{t_{(n_1+n_2-2)}}$ **,** the student t-distribution with degree of freedom $df = (\boldsymbol{n_1+n_2-2})$.

Q The following data represent the length of time, in days, to recovery for patients randomly treated with one of two medications to clear up severe bladder infections:

Medication1	**Medication2**
n_1= 14	**n_2=16**
$\boldsymbol{\bar{X}_1 = 17}$	$\boldsymbol{\bar{X}_2 = 19}$
$\boldsymbol{S_1^2 = 1.5}$	$\boldsymbol{S_2^2 = 1.8}$

Estimate $Var(\boldsymbol{\bar{X}_1 - \bar{X}_2})$, assuming normal populations **with equal variances**.

Solution: $S_p{}^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{(13)1.5+(15)1.8}{28} = 1.6607$

The estimation of $Var(\boldsymbol{\bar{X}_1 - \bar{X}_2}) = 1.6607\left(\frac{1}{14}+\frac{1}{16}\right) = 0.2224$

Sampling Distribution of $\boldsymbol{\hat{p}_1 - \hat{p}_2}$ is also no more than a special case of $f_{\bar{X}_1-\bar{X}_2}(.)$ if $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$ are Bernoulli populations with means: $\mu_1 = P_1$ and $\mu_2 = P_2$ the proportions of somewhat, say proportions of smokers, and variances:

$$\boldsymbol{\sigma_1^2} = P_1(1-P_1) \text{ And } \boldsymbol{\sigma_2^2} = P_2(1-P_2)$$

In this special case the sample means are equal to: $\boldsymbol{\bar{X}_1 = \hat{p}_1}$ and $\boldsymbol{\bar{X}_2 = \hat{p}_2}$ and ***c.l.t*** says as $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ are large enough that it is **approximately** true:

$$\boldsymbol{Z} = \frac{(\boldsymbol{\hat{p}_1-\hat{p}_2})-(P_1-P_2)}{\sqrt{\frac{P_1(1-P_1)}{\boldsymbol{n_1}}+\frac{P_2(1-P_2)}{\boldsymbol{n_2}}}} \sim \boldsymbol{N(0,1)} \Leftrightarrow \boldsymbol{f}_{(\boldsymbol{\hat{p}_1-\hat{p}_2})}(.) = \boldsymbol{N}\left(P_1-P_2, \frac{P_1(1-P_1)}{\boldsymbol{n_1}}+\frac{P_2(1-P_2)}{\boldsymbol{n_2}}\right)$$

Q. Suppose that 25% of the male students and 20% of the female students in a certain university smoke cigarettes. A random *sample* of 50 male students is taken. Another random sample of 100 female students is independently taken from this university. Let $\hat{p}_1$ and $\hat{p}_2$ be the proportions of smokers in the two samples, respectively.

(1) Find $E(\hat{p}_1 - \hat{p}_2) = \mu_{\hat{p}_1 - \hat{p}_2}$, the mean of $\hat{p}_1 - \hat{p}_2$.

(2) Find $Var(\hat{p}_1 - \hat{p}_2) = \sigma^2_{\hat{p}_1 - \hat{p}_2}$, the variance of $\hat{p}_1 - \hat{p}_2$.

(3) Find an approximate distribution of $\hat{p}_1 - \hat{p}_2$.

(4) What is the probability that $\hat{p}_1$ is greater than $\hat{p}_2$ by at least 0.06

Solution:

1) $E(\hat{p}_1 - \hat{p}_2) = 0.25 - 0.20 = 0.05$

2) $Var(\hat{p}_1 - \hat{p}_2) = \frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2} = \frac{0.25*0.75}{50} + \frac{0.20*0.80}{100} = 0.00535$

3) $\hat{p}_1 - \hat{p}_2 \approx N(0.05, 0.00535)$

4) $P(\hat{p}_1 - \hat{p}_2 > 0.06) = P(Z > \frac{0.06 - 0.05}{\sqrt{0.00535}}) = P(Z > 0.14) = 0.4447$

Exercises Single Mean:

Q1. A machine is producing metal pieces that are cylindrical in shape. A random sample of size 5 is taken and the diameters are 1.70, 2.11, 2.20, 2.31 and 2.28 centimeters. Then,

1) The sample mean is:

(A) 2.12 (B) 2.32 (C) 2.90 (D) 2.20 (E) 2.22

2) The sample variance is:

(A) 0.59757 (B) 0.28555 (C) 0.35633 (D) 0.06115 (E) 0.53400

Q2. The average life of a certain battery is 5 years, with a standard deviation of 1 year. Assume that the live of the battery approximately follows a normal distribution.

1) The sample mean $\overline{X}$ of a random sample of 5 batteries selected from this product has a mean; $E(\overline{X}) = \mu_{\overline{X}}$ equal to:

(A) 0.2 (B) 5 (C) 3 (D) None of these

2) The variance $Var(\overline{X}) = \sigma_{\bar{x}}^2$ of the sample mean $\overline{X}$ of a random sample of 5 batteries selected from this product is equal to:

(A) 0.2 (B) 5 (C) 3 (D) None of these

3) The probability that the average life of a random sample of size 16 of such batteries will be between 4.5 and 5.4 years is:

(A) 0.1039 (B) 0.2135 (C) 0.7865 (D) 0.9224

4) The probability that the average life of a random sample of size 16 of such batteries will be less than 5.5 years is:

(A) 0.9772 (B) 0.0228 (C) 0.9223 (D) None of these

5) The probability that the average life of a random sample of size 16 of such batteries will be more than 4.75 years is:

(A) 0.8413 (B) 0.1587 (C) 0.9452 (D) None of these

6) If $P(\ \overline{X} > a\ ) = 0.1492$ where $\overline{X}$ represents the sample mean for a random sample of size 9 of such batteries, then the numerical value of a is:

(A) 4.653 (B) 6.5 (C) 5.347 (D) None of these

Q3. The random variable X, representing the lifespan of a certain light bulb, is distributed normally with a mean of 400 hours and a standard deviation of 10 hours.

1. What is the probability that a particular light bulb will last for more than 380 hours?
2. Light bulbs with lifespan less than 380 hours are rejected. Find the percentage of light bulbs that will be rejected.
3. If 9 light bulbs are selected randomly, find the probability that their average lifespan will be less than 405.

Q4. Suppose that you take a random sample of size n=64 from a distribution with mean μ=55 and standard deviation σ=10. Let $\overline{X}$ be the sample mean.

(a) What is the approximated sampling distribution of $\overline{X}$?

(b) What is the mean of $\overline{X}$?

(c) What is the standard error (standard deviation) of $\overline{X}$?

(d) Find the probability that the sample mean $\bar{X}$ exceeds 52.

Q5. The amount of time that customers using ATM (Automatic Teller Machine) is a random variable with the mean 3.0 minutes and the standard deviation of 1.4 minutes. If a random sample of 49 customers is observed, then

(1) the probability that their mean time will be at least 2.8 minutes is

(A) 1.0 (B) 0.8413 (C) 0.3274 (D) 0.4468

(2) the probability that their mean time will be between 2.7 and 3.2 minutes

(A) 0.7745 (B) 0.2784 (C) 0.9973 (D) 0.0236

Q6. The average (mean) life of a certain type of batteries is 5 years, with a standard deviation of ne year. Assuming the live of the battery follows approximately a normal distribution.

(a) If a random sample of 5 batteries (to be selected from this production) has a mean of 3 years with a standard deviation of one year, then: the random variable $\bar{X}$ (the mean of all possible samples of size 5 batteries) has a mean $\mu_{\bar{X}}$ equal to:

(A) 0.2 (B) 5 (C) 3 (D) None of these

(b) The variance of the random variable $\bar{X}$, $\sigma^2_{\bar{X}}$, is equal to:

(A) 0.2 (B) 5 (C) 3 (D) None of these

(c) The probability that the mean life of a random sample of size 16 of such batteries will be between 4.5 and 5.4 years is:

(A) 0.1039 (B) 0.2135 (C) 0.7865 (D) None of these

(d) The probability that the mean life of a random sample of size 16 of such batteries will be less than 5.5 years is:

(A) 0.9772 (B) 0.0228 (C) 0.9223 (D) None of these

(e) If $P(\ \overline{X} > a\ ) = \ 0.1492$ where $\overline{x}$ represents the sample mean for a random sample of size 9 to be selected from the production lot, then the numerical value of a is:

(A) 4.653 (B) 6.5 (C) 5.347 (D) None of these

<u>Q7.</u> The average life of an industrial machine is 6 years, with a standard deviation of 1 year. Assume the life of such machines follows approximately a normal distribution. A random sample of 4 of such machines is selected. The sample mean life of the machines in the sample is $\overline{x}$.

(1) The sample mean has a mean $\mu_{\bar{x}} = E(\overline{X})$ equals to:

(A) 5 (B) 6 (C) 7 (D) 8

(2) The sample mean has a variance $\sigma^2_{\bar{x}} = Var(\overline{X})$ equals to:

(A) 1 (B) 0.5 (C) 0.25 (D) 0.75

(3) $P(\overline{X} < 5.5) =$

(A) 0.4602 (B) 0.8413 (C) 0.1587 (D) 0.5398

(4) If $P(\ \overline{X} > a\ ) = \ 0.1492$, then the numerical value of a is:

(A) 0.8508 (B) 1.04 (C) 6.52 (D) 0.2

Exercises two means:

Q1. A random sample of size n_1 = 36 is taken from a first population with a mean μ_1 = 70 and a standard deviation σ_1 = 4. A second independent random sample of size n_2 = 49 is taken from a second population with a mean μ_2 = 85 and a standard deviation σ_2 = 5. Let $\overline{X}_1$ and $\overline{X}_2$ be the averages of the first and second samples, respectively.

a) Find E($\overline{X}_1$), Var($\overline{X}_1$) and P(70 < $\overline{X}_1$ < 71)

b) Find E($\overline{X}_1 - \overline{X}_2$) , Var($\overline{X}_1 - \overline{X}_2$) and P($\overline{X}_1 - \overline{X}_2 > -16$).

Q2. A random sample of size 25 is taken from a normal population (first population) having a mean of 100 and a standard deviation of 6. A second random sample of size 36 is taken from a different normal population (second population) having a mean of 97 and a standard deviation of 5. Assume that these two samples are independent.

(1) the probability that the sample mean of the first sample will exceed the sample mean of the second population by at least 6 is

(A) 0.0013 (B) 0.9147 (C) 0.0202 (D) 0.9832

(2) the probability that the difference between the two sample means will be less than 2 is

(A) 0.099 (B) 0.2483 (C) 0.8499 (D) 0.9499

Exercises one and two propotions:

Q1. Suppose that 20% of the students in a certain university smoke cigarettes. A random sample of 50 students is taken from this university. Let $\hat{p}$ be the proportion of smokers in the sample. Find $\mu_{\hat{p}}$, $\sigma^2_{\hat{p}}$ and P($\hat{p}$ >0.25).

Q2. Suppose that 25% of the male students and 20% of the female students in a certain university smoke cigarettes. A random sample of 50 male students is taken. Another random sample of 100 female students is independently taken from this university. Let $\hat{p}_1$ and $\hat{p}_2$ be the proportions of smokers in the two samples, respectively.

Find $\mu_{\hat{p}_1-\hat{p}_2}$, $\sigma^2_{\hat{p}_1-\hat{p}_2}$ and P(0.10< $\hat{p}_1-\hat{p}_2$ <0.20).