Sampling Distribution

Sampling Distribution: Single Mean

$$* \ \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$* \ E(\bar{X}) = \bar{X} = \mu \qquad * \ Var(\bar{X}) = \frac{\sigma^2}{n}$$

Sampling Distribution: Two Means

$$* \ \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2 \ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$* \ E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 \qquad * \ Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Sampling Distribution: Single Proportion

$$* \ \hat{p} \sim N\left(p, \frac{pq}{n}\right)$$

$$* \ E(\hat{p}) = p \qquad * \ Var(\hat{p}) = \frac{pq}{n}$$

Sampling Distribution: Two Proportions

$$* \ \hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

$$* \ E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 \qquad * \ Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

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 $Var(\overline{X}) = \sigma^2_{\overline{X}}$ 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

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

Solution:

$$\mu = 5 \quad ; \quad \sigma = 1 \quad ; \quad n = 5$$

(1): $E(\bar{X}) = \mu = 5$

(2): $Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{1}{5} = 0.2$

(3): $n = 16 \rightarrow \frac{\sigma}{\sqrt{n}} = \frac{1}{4}$

$$P(4.5 < \bar{X} < 5.4) = P\left(\frac{4.5-\mu}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{5.4-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(\frac{4.5-5}{\frac{1}{4}} < Z < \frac{5.4-5}{\frac{1}{4}}\right)$$

$$= P(-2 < Z < 1.6)$$

$$= P(Z < 1.6) - P(Z < -2)$$

$$= 0.9452 - 0.0228 = 0.9224$$

(4): $P(\bar{X} < 5.5) = P\left(Z < \frac{5.5-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z < \frac{5.5-5}{1/4}\right) = P(Z < 2) = 0.9772$

(5): $P(\bar{X} > 4.75) = P\left(Z > \frac{4.75-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z > \frac{4.75-5}{\frac{1}{4}}\right) = P(Z > -1)$

$$= 1 - P(Z < -1) = 1 - 0.1587 = 0.841$$

(6): $P(\bar{X} > a) = 0.1492 \;\; ; \;\; n = 9$

$$P\left(Z > \frac{a-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = 0.1492$$

$$\Rightarrow 1 - P\left(Z < \frac{a-5}{\frac{1}{3}}\right) = 0.1492$$

$$\Rightarrow P\left(Z < \frac{a-5}{\frac{1}{3}}\right) = 0.8508$$

$$\frac{a-5}{\frac{1}{3}} = 1.04$$

$$a = 5 + \frac{1.04}{3} = 5.347$$

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 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean.

(a) What is the approximated sampling distribution of $\bar{X}$?
(b) What is the mean of $\bar{X}$?
(c) What is the standard error (standard deviation) of $\bar{X}$?

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

Solution

$$\mu = 55 \;\; ; \;\; \sigma = 10 \;\; ; \;\; n = 64$$

(a) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = \bar{X} \sim N\left(55, \frac{100}{64}\right)$

(b) $E(\bar{X}) = \mu = 55$

(c) $S.D(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{64}} = \frac{10}{8}$

(d) $P(\bar{X} > 52) = P\left(Z > \frac{52-55}{\frac{10}{8}}\right)$

$$= P(Z > -2.4)$$

$$= 1 - P(Z < -2.4)$$

$$= 1 - 0.0082 = 0.9918$$

Old exams:

Question: Suppose that the hemoglobin levels (in g/dl) of healthy Saudi females are approximately normally distributed with mean of 13.5 and a standard deviation of 0.7. If 15 healthy adult Saudi female is randomly chosen, then:

(1) The mean of $\bar{x}$ ($E(\bar{x})$ or $\mu_{\bar{x}}$) is:

(A) 0.7 (B) 13.5 (C) 15 (D) 3.48

(2) The standard error of $\bar{x}$ ($\sigma_{\bar{x}}$) is:

(A) 0.181 (B) .0327 (C) 0.7 (D) 13.5

(3) $P(\bar{X} < 14) =$

(A) 0.99720 (B) 0.99440 (C) 0.76115 (D) 0.9971

(4) $P(\bar{X} > 13.5) =$

(A) 0.99 (B) 0.50 (C) 0.761 (D) 0.622

(5) $P(13 < \bar{X} < 14) =$

(A) 0.9972 (B) 0.9942 (C) 0.7615 (D) 0.5231

Question: If the uric acid value in normal adult males is approximately normally distributed with a mean and standard derivation of 5.7 and 1 mg percent, respectively, find the probability that a sample of size 9 will yield a mean:

(1) Greater than 6 is

A)	0.2109	(B)	0.1841	(C)	0.8001	(D)	0.8159

(2) At most 5.2 is

(A)	0.6915	(B)	0.9331	(C)	0.8251	(D)	0.0668

(3) Between 5 and 6 is

(A)	0.1662	(B)	0.7981	(C)	0.8791	(D)	0.9812

Sampling Distribution: Two Means:

$$* \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2 \ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$* E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 \qquad * Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Q1. 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 $\overline{X}_1$ and $\overline{X}_2$ be the averages of the first and second samples, respectively.

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

d) Find $P(\overline{X}_1 - \overline{X}_2 > -16)$.

Solution

$$n_1 = 36 \, , \mu_1 = 70 \, , \sigma_1 = 4$$

$$n_2 = 49 \, , \mu_2 = 85 \, , \sigma_2 = 5$$

(b): $E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 = 70 - 85 = -15$

$$Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{16}{36} + \frac{25}{49} = 0.955$$

(d): $P(\bar{X}_1 - \bar{X}_2 > -16) = P\left(Z > \frac{-16-(-15)}{\sqrt{0.955}}\right) = 1 - P\left(Z < \frac{-16-(-15)}{\sqrt{0.955}}\right)$

$$= 1 - P(Z < -1.02) = 0.8461$$

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 sample 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) 0.2480 (C) 0.8499 (D) 0.9499

Solution

$$n_1 = 25\,, \mu_1 = 100\,, \sigma_1 = 6$$

$$n_2 = 36\,, \mu_2 = 97\,, \sigma_2 = 5$$

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 = 100 - 97 = 3$$

$$Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{36}{25} + \frac{25}{36} = 2.35$$

(1) $P(\bar{X}_1 > \bar{X}_2 + 6) = P(\bar{X}_1 - \bar{X}_2 > 6)$

$$= P\left(Z > \frac{6-(3)}{\sqrt{2.35}}\right) = P(Z > 2.05)$$

$$= 1 - P(Z < 2.05)$$

$$= 1 - 0.9798 = 0.0202$$

(2) $P(\bar{X}_1 - \bar{X}_2 < 2) = P\left(Z < \frac{2-(3)}{\sqrt{2.35}}\right)$

$$= P(Z < -0.68) = 0.2483$$

Question: Given two normally distributed populations with equal means and variances of

$\sigma_1^2 = 100,\ \sigma_2^2 = 350$. *Two random samples of sizes* $n_1 = 40,\ n_2 = 35$ *are drawn and the sample means* $\bar{X}_1,\ \bar{X}_2$ *are calculated, respectively, then*

(1) $P(\bar{X}_1 - \bar{X}_2 > 12)$ *is*

(A)	*0.1499*	*(B)*	*0.8501*	*(C)*	*0.9997*	*(D)*	*0.0003*

(2) $P(5 < \bar{X}_1 - \bar{X}_2 < 12)$ *is*

(A)	*0.0783*	*(B)*	*0.9217*	*(C)*	*0.8002*	*(D)*	*None of these*

Sampling Distribution: Single Proportion

$$* \hat{p} \sim N\left(p, \frac{pq}{n}\right)$$

$$* E(\hat{p}) = p \qquad * Var(\hat{p}) = \frac{pq}{n}$$

Q1. Suppose that 20% of the students in a certain university smoke cigarettes. A random sample of 5 students is taken from this university. Let $\hat{p}$ be the proportion of smokers in the sample.

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

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

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

(4) Find P($\hat{p}$ >0.25).

Solution

$$p = 0.2 \quad ; \quad n = 5 \quad ; \quad q = 1 - p = 0.8$$

(1): $E(\hat{p}) = p = 0.2$

(2): $Var(\hat{p}) = \frac{pq}{n} = \frac{0.2 \times 0.8}{5} = 0.032$

(3): $\hat{p} \sim N(0.2, 0.032)$

(4): $P(\hat{p} > 0.25) = P\left(Z > \frac{0.25 - 0.2}{\sqrt{0.032}}\right) = P(Z > 0.28)$

$$= 1 - P(Z < 0.28) = 1 - 0.6103 = 0.3897$$

Question: A random sample of 35 students in a certain university resulted in the sample proportion of smokers $\hat{p}$ *= 0.15. Then:*

1. *The point estimate of* p *is:*

(A) 0.35	*(B) 0.85*	*(C)* ***0.15***	*(D) 0.80*

2. *The standard deviation of* $\hat{p}$ *is:*

(A) 0.3214	*(B) .0036*	*(C) 0.1275*	*(D)* ***0.0604***

Question: In a study, it was found that 31% of the adult population in a certain city has a diabetic disease. 100 people are randomly sampled from the population. Then

(6) The mean for the sample proportion $(\mu_{\hat{p}}$ ***or*** $E(\hat{p}))$ *is:*

(A) 0.4 *(B) 0.31* *(C) 0.69* *(D) 0.1*

(7) $P(\hat{p} > 0.4) =$

(A) 0.02619 *(B) 0.02442* *(C) 0.0256* *(D) 0.7054*

<u>Sampling Distribution: Two Proportions:</u>

$$* \hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}\right)$$

$$* E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 \qquad * Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

Q1. Suppose that 25% of the male students and 20% of the female students in a certain university smoke cigarettes. A random sample of 5 male students is taken. Another random sample of 10 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) Find $P(0.10 < \hat{p}_1 - \hat{p}_2 < 0.20)$.

<u>Solution</u>

$$p_1 = 0.25 \;\; ; \quad n_1 = 5$$

$$p_2 = 0.2 \;\; ; \quad n_2 = 10$$

(1): $E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2 = 0.25 - 0.2 = 0.05$

(2): $Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = \frac{0.25 \times 0.75}{5} + \frac{0.2 \times 0.8}{10} = 0.054$

(3): $\hat{p}_1 - \hat{p}_2 \sim N(0.05, 0.054)$

(4): $P(0.1 < \hat{p}_1 - \hat{p}_2 < 0.2) = \left(\frac{0.1 - 0.05}{\sqrt{0.054}} < Z < \frac{0.2 - 0.05}{\sqrt{0.054}}\right)$

$$= (0.22 < Z < 0.65)$$

$$P(Z < 0.65) - P(Z < 0.22)$$

$$= 0.7422 - 0.5871 = 0.1551$$

Question: Suppose that 7 % of the pieces from a production process A are defective while that proportion of defective for another production process B is 5 %. A random sample of size 400 pieces is taken from the production process A while the sample size taken from the production process B is 300 pieces. If $\hat{P}_1$ and $\hat{P}_2$ be the proportions of defective pieces in the two samples, respectively, then:

3. *The sampling distribution of $\hat{P}_1 - \hat{P}_2$ is:*

(A) N(0, 1)	*(B)* ***Normal***	*(C) T*	*(D) unknown*

4. *The value of the standard error of the difference $(\hat{P}_1 - \hat{P}_2)$ is:*

(A) ***0.02***	*(B) 0.10*	*(C) 0*	*(D) 0.22*