Statistics II: TUTORIAL 4 - Sampling Distribution

August 27, 2019

1. If $X \sim N(200, 80)$ and a random sample of size 5 is taken from the distribution, find the probability that the sample mean

   (a) is greater than 207,
   Let X_1be the random sample of size 5 taken from the distribution with $X_1 \sim N\left(200, \frac{80}{5}\right)$

$$\begin{aligned}
P\left(X_1 > 207\right) &= P\left(Z > \frac{207-200}{\sqrt{\frac{80}{5}}}\right) \\
&= P\left(Z > \frac{207-200}{\sqrt{\frac{80}{5}}}\right) \\
&= P\left(Z > 1.75\right) \\
&= 0.040059
\end{aligned}$$

   (b) lies between 201 and 209.

$$\begin{aligned}
P\left(201 \leq X_1 \leq 209\right) &= P\left(\frac{201-200}{\sqrt{\frac{80}{5}}} \leq Z \leq \frac{209-200}{\sqrt{\frac{80}{5}}}\right) \\
&= P\left(0.25 \leq Z \leq 2.25\right) \\
&= 0.38907
\end{aligned}$$

2. The heights of a new variety of sunflower are normally distributed with mean 2m (2 meters) and standard deviation 40cm. 100 samples of 50 flowers each are measured. How many samples with the sample mean

   (a) between 195cm and 205cm?
   Let X be the height of a new variety of sunflower with $X \sim N\left(200, 40^2\right)$

      i. $n = 50$
      ii. $\mu_{\bar{X}} = \mu = 200$

iii. $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{50}} = 5.6569$

iv.

$$\begin{aligned} P\left(195 < \bar{X} < 205\right) &= P\left(\frac{195-200}{5.6569} < \frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} < \frac{205-200}{5.6569}\right) \\ &= P\left(-0.88 < Z < 0.88\right) \\ &= 1 - P\left(Z > 0.88\right) - P\left(Z < -0.88\right) \\ &= 1 - 2 \cdot P\left(Z > 0.88\right) \\ &= 1 - 2 \cdot (0.1894) \\ &= 0.6212 \end{aligned}$$

v. Samples $= 0.6212 * 100 = 62.12 \approx 62 samples$

(b) $P\left(\bar{X} < 197\right)$

$$\begin{aligned} P\left(Z < \frac{197-200}{5.6569}\right) &= P\left(Z < -0.53\right) \\ &= P\left(Z > 0.53\right) \\ &= 0.2981 \end{aligned}$$

i. Samples $= 0.2981 * 100 = 29.81 \approx 30 samples$

3. If large number of samples size n are taken from a population which follows a normal distribution with mean 74 and standard deviation 6, find n if the probability that the sample mean

   (a) exceeds 75 is 0.281,
   Let X be the sample mean taken from a population with size n where $\bar{X} \sim \left(74, \frac{6^2}{n}\right)$

$$\begin{aligned} P\left(\bar{X} > 75\right) &= 0.281 \\ P\left(Z > \frac{75-74}{\frac{6}{\sqrt{n}}}\right) &= 0.281 \\ \frac{\sqrt{n}}{6} &= 0.58 \\ \sqrt{n} &= 3.48 \\ n &= 12.11 \\ n &\approx 13 \end{aligned}$$

(b) is less than 70.4 is 0.00135.

$$
\begin{aligned}
P\left(\bar{X} < 70.4\right) &= 0.00135 \\
-P\left(\bar{X} > 70.4\right) &= 0.00135 \\
-P\left(Z > \frac{70.4 - 74}{\frac{6}{\sqrt{n}}}\right) &= 0.00135 \\
-\frac{70.4 - 74}{\frac{6}{\sqrt{n}}} &= 3 \\
\frac{3.6\sqrt{n}}{6} &= 3 \\
n &= \left(\frac{3 * 6}{3.6}\right)^2 \\
&= 25
\end{aligned}
$$

4.

(a) $n = 20, p = 0.6$

i. $\mu_{\bar{x}} = np = 12$

ii. $\sigma_{\bar{x}} = \sqrt{\frac{20*0.4*0.6}{100}} = 0.2191$

iii. $P\left(\bar{X} > 12.4\right)$

$$
\begin{aligned}
P\left(\bar{X} > 12.4\right) &= P\left(Z > \frac{12.4 - 12}{0.2191}\right) \\
&= P\left(Z > 1.83\right) \\
&= 0.0336
\end{aligned}
$$

(b) $P\left(\bar{X} < 12.2\right)$

$$
\begin{aligned}
P\left(\bar{X} < 12.2\right) &= P\left(Z < \frac{12.2 - 12}{0.2191}\right) \\
&= P\left(Z < 0.91\right) \\
&= 1 - P\left(Z > 0.91\right) \\
&= 1 - 0.1814 \\
&= 0.8186
\end{aligned}
$$

5. If a large number of samples of size n is taken from Bin (20 , 0.2) and approximately 90% of the sample means are less than 4.354, estimate n.

(a) Since sample size is large ($n > 30$), we can use Central Limit Theorem where $\bar{X} = N(4, 3.2)$

$$\begin{aligned}
P\left(\bar{X} < 4.354\right) &= 0.9 \\
P\left(\bar{X} \geq 4.354\right) &= 0.1 \\
P\left(Z > \frac{4.354 - 4}{\sqrt{\frac{3.2}{n}}}\right) &= 0.1 \\
P\left(Z > \frac{0.354}{\sqrt{\frac{3.2}{n}}}\right) &= 0.1 \\
\frac{0.354}{\sqrt{\frac{3.2}{n}}} &= 1.28 \\
\sqrt{\frac{3.2}{n}} &= \frac{0.354}{1.28} \\
\frac{3.2}{n} &= 0.2765^2 \\
n &= \frac{3.2}{0.07649} \\
&= 41.83 \\
&\approx 42
\end{aligned}$$

6. $\mu = 2.9, \sigma = 2.9$

   (a) $\mu_{\bar{X}} = 2.9$

   (b) $\sigma_{\bar{X}} = \sqrt{\frac{2.9}{n}}$

$$\begin{aligned}
P\left(\bar{X} > 3.41\right) &= 0.01 \\
P\left(Z > \frac{3.41 - 2.9}{\sqrt{\frac{2.9}{n}}}\right) &= 0.01 \\
\frac{0.51}{\frac{1.7029}{\sqrt{n}}} &= 2.33 \\
0.2995\sqrt{n} &= 2.33 \\
\sqrt{n} &= 7.7796 \\
n &= 60.5227 \\
&\approx 61/
\end{aligned}$$

7. $\mu = 1.5, \sigma^2 = \frac{(-1.5)^2+(-0.5)^2+(0.5)^2+(1.5)^2}{4} = \frac{5}{4} = 1.25$

   (a) $n = 36$

(b) $\mu_{\bar{X}} = \mu = 1.5$

(c) $\sigma_{\bar{X}} = \frac{\sqrt{1.25}}{\sqrt{36}} = 0.1863$

(d) $P\left(1.4 \leq \bar{X} \leq 1.8\right)$

$$\begin{aligned} P\left(\frac{1.4-1.5}{0.1863} \leq \bar{X} \leq \frac{1.8-1.5}{0.1863}\right) &= P\left(-0.5368 \leq Z \leq 1.6103\right) \\ &= 1 - P\left(Z \geq 1.61\right) - P\left(Z \geq 0.54\right) \\ &= 1 - 0.0537 - 0.2946 \\ &= 0.6517 \end{aligned}$$

8. $n = 150$, $p = 0.5$

(a) Since n is big ($n > 30$) and $np > 5, nq > 5$. By Central Limit Theorem, we can use normal approximation. Let X be the proportion of heads occuring in 150 throws with $X \sim N\left(0.5, \frac{0.25}{150}\right)$.

$$\begin{aligned} P\left(\hat{P} < 0.4\right) &= P\left(Z < \frac{0.4-0.5}{\sqrt{\frac{0.25}{150}}}\right) \\ &= P\left(Z < -2.45\right) \\ &= P\left(Z > 2.45\right) \\ &= 0.00714 \end{aligned}$$

(b)

$$\begin{aligned} P\left(0.44 < \hat{P} < 0.56\right) &= P\left(\frac{0.44-0.5}{\sqrt{\frac{0.25}{150}}} < \hat{P} < \frac{0.56-0.5}{\sqrt{\frac{0.25}{150}}}\right) \\ &= P\left(-1.47 < Z < 1.47\right) \\ &= 1 - P\left(Z > 1.47\right) - P\left(Z > 1.47\right) \\ &= 1 - 2\left(P\left(Z > 1.47\right)\right) \\ &= 1 - 2\left(0.0708\right) \\ &= 0.8584 \end{aligned}$$

9. Let p be the probability that the tools produced by a certain machine are defective, where $p = 0.05$. $n = 400$. Since $n > 30$,$np > 5, nq > 5$. We can approximate the binomial distribution of $B\left(400, 0.05\right)$ with $N\left(\mu_{\bar{x}} = p = 0.02, \sigma_{\bar{x}}^2 = \frac{pq}{n} = \frac{0.02*0.98}{400}\right)$ using central limit theorem

(a) $P\left(\hat{P} > 0.03\right)$

$$\begin{aligned} P\left(\hat{P} > 0.03\right) &= P\left(Z > \frac{0.03 - 0.02}{\sqrt{\frac{0.02*0.98}{400}}}\right) \\ &= P\left(Z > 1.43\right) \\ &= 0.0764 \end{aligned}$$

(b) $P\left(\hat{P} < 0.02\right)$

$$\begin{aligned} P\left(\hat{P} < 0.02\right) &= \frac{0.02 - 0.02}{\sqrt{\frac{0.02*0.98}{400}}} \\ &= P\left(Z < 0\right) \\ &= P\left(Z > 0\right) \\ &= 0.5 \end{aligned}$$

10. A certain candidate standing for election is known to have the support of proportion 0.46 of the electors. Find the probability that the candidate will have a majority in a random sample of
Let $\hat{P}$be the proportion of electors who support this candidate.

   (a) 200 electors,

      i. Let p= probability of support, where $p = 0.46$. $n = 200$. Since $n > 30$, $np > 5, nq > 5$, we can approximate the binomial distribution of $B\left(200, 0.46\right)$ with $\hat{P} \sim N\left(\mu_{\bar{x}} = 0.46, \sigma_{\bar{x}}^2 = \frac{0.46*0.54}{200} = 0.001242\right)$. Majority is considered where $\hat{P} > 0.5$

$$\begin{aligned} P\left(\hat{P} > 0.5\right) &= P\left(Z > \frac{0.5 - 0.46}{\sqrt{0.001242}}\right) \\ &= P\left(Z > 1.14\right) \\ &= 0.1271 \end{aligned}$$

   (b) 1000 electors.

      i. Let p follow part (a) and $n = 1000$. Since n is bigger, we can still use Central Limit Theorem with $\hat{P} \sim N\left(\mu_{\bar{x}} = 0.46, \sigma_{\bar{x}}^2 = \frac{0.46*0.54}{1000}\right)$

$$\begin{aligned} P\left(\hat{P} > 0.5\right) &= P\left(Z > \frac{0.5 - 0.46}{\sqrt{\frac{0.46*0.54}{1000}}}\right) \\ &= P\left(Z > 2.54\right) \\ &= 0.00554 \end{aligned}$$

11.

(a) Since they are normally distribution, $\bar{X}_1 \sim N\left(57, \frac{12}{36}^2\right)$, $\bar{X}_2 \sim N\left(25, \frac{6}{36}^2\right)$

   i. $\mu_{\bar{X}_1 - \bar{X}_2} = 57 - 25 = 32$

(b) $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{12^2}{36} + \frac{6^2}{36}} = 2.2361$

(c) $\bar{X}_1 - \bar{X}_2 \sim N(32, 5)$, it is a normal distribution.

(d) $P\left(29 \leq \bar{X}_1 - \bar{X}_2 \leq 31\right)$

   i. Answer

$$
\begin{aligned}
P\left(\frac{29-32}{2.2361} \leq Z \leq \frac{31-32}{2.2361}\right) &= P(-1.34 \leq Z \leq -0.45) \\
&= P(Z \geq 0.45) - P(Z \geq 1.34) \\
&= 0.3264 - 0.0901 \\
&= 0.2363
\end{aligned}
$$

(e) Let $\hat{P}$be the proportion of the the mean of the sample differ

   i. $\hat{P} \sim N\left(\mu = 32, \sigma^2 = 2.2361^2\right)$

$$
\begin{aligned}
P\left(\hat{P} > 34\right) &= P\left(Z > \frac{34-32}{2.2361}\right) \\
&= P(Z > 0.89) \\
&= 0.1867
\end{aligned}
$$

12. Since $n_1 > 30$, $n_2 > 30$, we can use CLT

$$\mu_{X_1} = 0.12, \sigma_{X_1} = 0.02, n = 49$$

$$\mu_{X_2} = 0.13, \sigma_{X_2} = 0.03, n = 64$$

(a) $E\left(\bar{X}_1 - \bar{X}_2\right) = 0.12 - 0.13 = -0.01$

(b) $\sigma_{X_1} - \sigma_{X_2} = \sqrt{\frac{0.02^2}{49} + \frac{0.03^2}{64}} = 0.004714420994 \approx 0.004714$

(c) $\bar{X}_1 - \bar{X}_2 \sim N\left(-0.01, 0.004714^2\right)$ .It is a normal distribution

(d) $P\left(\bar{X}_2 > \bar{X}_1\right)$

$$
\begin{aligned}
P\left(\bar{X}_1 - \bar{X}_2 < 0\right) &= P\left(Z < \frac{0-(-0.01)}{0.004714}\right) \\
&= P(Z < 2.12) \\
&= 1 - P(Z > 2.12) \\
&= 1 - 0.0170 \\
&= 0.983
\end{aligned}
$$

(e) $P\left(\bar{X}_1 - \bar{X}_2 > -0.005\right)$

$$\begin{aligned} P\left(\bar{X}_1 - \bar{X}_2 < -0.005\right) &= P\left(Z < \frac{-0.005-(-0.01)}{(0.004714)}\right) \\ &= P\left(Z < 1.06\right) \\ &= 1 - P\left(Z > 1.06\right) \\ &= 1 - 0.1446 \\ &= 0.8554 \end{aligned}$$

13. (DO THESE)$p_1 = 0.80, n_1 = 0.20, n_1 = 80$
$p_2 = 0.72, n_2 = 0.28, n_2 = 60$. Since both $n_1 \& n_2 > 30$

(a)

$$\begin{aligned} E\left(\hat{P}_1 - \hat{P}_2\right) &= P_1 - P_2 \\ &= 0.8 - 0.72 \\ &= 0.08 \end{aligned}$$

(b) Standard Deviation

$$\begin{aligned} \sigma_{\hat{P}_1 - \hat{P}_2} &= \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \\ &= \sqrt{\frac{67}{12500}} \\ &= 0.07321 \end{aligned}$$

(c)

$$\begin{aligned} P\left(\hat{P}_1 - \hat{P}_2 \geq 0.1\right) &= P\left(Z > \frac{0.1 - 0.08}{\sqrt{\frac{67}{12500}}}\right) \\ &= P\left(Z > 0.27\right) \\ &= 0.3936 \end{aligned}$$

14. (DO THESE) $p_1 = 0.3, n_1 = 80$
$p_2 = 0.18, n_2 = 70$

(a)

$$\begin{aligned} E\left(\hat{P}_1 - \hat{P}_2\right) &= p_1 - p_2 \\ &= 0.3 - 0.18 \\ &= 012 \end{aligned}$$

(b)

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{(0.3)\,(0.7)}{80}+\frac{(0.18)\,(0.82)}{70}}$$
$$= 0.0688$$

(c)

$$\hat{P}_1-\hat{P}_2 \sim N\left(0.12, 0.0688^2\right)$$

$$\begin{aligned} P\left(0.1<\hat{P}_1-\hat{P}_2<0.2\right) &= P\left(\frac{0.1-0.12}{0.0688}<Z<\frac{0.2-0.12}{0.0688}\right) \\ &= P\left(-0.29<Z<1.16\right) \\ &= 1-P\left(Z>0.29\right)-P\left(Z>1.16\right) \\ &= 0.4911 \end{aligned}$$