Statistics 6 Tutorial: Sampling

Additional Practice Questions

1. [N05/P2/26] In a country, 75% of the population have height exceeding 1.50m and 10% have height exceeding 1.90m. Assuming a normal distribution of heights, show that the height exceeded by 20% of the population is 1.81m, correct to 3 significant figures.
   A random sample of 80 people is taken from the population. Find the probability that the sample mean exceeds 1.69m.

   Solution
   Let X be the r.v. that denotes the height of a randomly chosen person from the population.

   $X \sim N\left(\mu, \sigma^2\right)$

   $P\left(X > 1.50\right) = 0.75 \quad \& \quad P(X > 1.9) = 0.1$

   $P(Z \le \frac{1.5-\mu}{\sigma}) = 0.25 \;\& \quad P(Z \le \frac{1.9-\mu}{\sigma}) = 0.9$

   $\frac{1.5-\mu}{\sigma} = -0.6745 \quad \& \quad \frac{1.9-\mu}{\sigma} = 1.282$

   $\mu = 1.6377 \quad \& \quad \sigma = 0.2045$

   Want to find a when $P\left(X > a\right) = 0.20$

   $P\left(X \le a\right) = 0.80$

   $a = 1.809811542 \approx 1.81$

   $\bar{X} \sim N\left(1.6377, \frac{0.2045^2}{80}\right)$

   $P\left(\bar{X} > 1.69\right) = 0.111$ (note: answer should be 0.0111)

2. [UCLES] The random variable X has a normal distribution with mean 2 and variance 3. The random variable S is the sum of 200 independent observations of X, and the random variable T is the sum of a further 300 independent observations of X. Find
   (i) $\mathrm{P}\left(S > 405\right)$,
   (ii) $\mathrm{P}\left(3S < 40 + 2T\right)$,
   (iii) The random variable N is the sum of n independent observations of X. Find the approximate values of $\mathrm{P}\left(N > 2.5n\right)$, $\mathrm{P}\left(N < 1.7n\right)$ and $\mathrm{P}\left(1.7n < N < 2.5n\right)$ as n becomes very large.

Solution

(i) $S = X_1 + ... + X_{200} \sim N(400, 600)$

$P(S > 405) = 0.419$

(ii)

$3S \sim N(1200, 5400) \quad 2T \sim N(1200, 3600)$

$3S - 2T \sim N(0, 9000)$

$P(3S - 2T < 40) = 0.663$

(iii)

$N \sim N(2n, 3n)$

$P(N > 2.5n)$

$= P(Z > \frac{0.5\sqrt{n}}{\sqrt{3}}) \to 0 \text{ as } n \to \infty$

$P(N < 1.7n)$

$= P(Z < \frac{-0.3\sqrt{n}}{\sqrt{3}}) \to 0 \text{ as } n \to \infty$

$P(1.7n < N < 2.5n)$

$= P(\frac{-0.3\sqrt{n}}{\sqrt{3}} < Z < \frac{0.5\sqrt{n}}{\sqrt{3}}) \to 1 \text{ as } n \to \infty$

3. [2011 RI/II/12(modified)] The weight, x kg, of each student in a random sample of 120 students from a secondary school is measured, and the results are summarized by

$$\sum (x - 50) = -100, \quad \sum (x - 50)^2 = 1158$$

(i) Find unbiased estimates of the population mean and variance.

(ii) Another random sample of n students ($n \geq 50$) is taken from the school. Given that the probability of the sample mean weight exceeding 49.9 kg is at most 0.01, find the least value of n.

Solution

(i) Unbiased estimate of population mean

$$= \frac{\sum (x - 50)}{n} + 50 = \frac{-100}{120} + 50$$

$$= 49\frac{1}{6} \text{ or } 49.167 \text{ (5 s.f)} = 49.2 \text{ (3 s.f)}$$

Unbiased estimate of population variance

$$= \frac{1}{n-1}\left[\sum (x - 50)^2 - \frac{\left(\sum (x - 50)\right)^2}{n}\right]$$

$$= \frac{1}{119}(1158 - \frac{(-100)^2}{120}) \approx 9.0308 = 9.03 \text{ (3 s.f)}$$

(ii) Since n is large, by Central Limit Theorem,

$$\bar{X} \sim N(49.167, \frac{9.0308}{n}) \text{ approximately.}$$

$$P(\bar{X} > 49.9) \le 0.01$$

Method 1 : Using Standardization

$$P(\bar{X} < 49.9) \ge 0.99$$

$$P(Z < \frac{49.9 - 49.167}{\sqrt{\frac{9.0308}{n}}}) \ge 0.99$$

From GC, $\frac{49.9 - 49.167}{\sqrt{\frac{9.0308}{n}}} \ge 2.3263$

$n \ge 90.960$

$\therefore$ Least value of $n = 91$

Method 2 : Using table of values

Key in Y_1 = normalcdf(49.9, E99, 49.167, $\sqrt{9.0308/X}$)

n	$P(\bar{X} > 49.9)$
90	0.01033 > 0.01
91	0.00999 < 0.01
92	0.00965 < 0.01

$\therefore$ Least value of n = 91

4. A population has variance 4. Given that the probability that $\overline{X}$ differs from the population mean μ by less than 0.1 is at least 0.95, find the least value of the sample size required.

Solution

Let n denote the sample size.

$\overline{X} \sim \mathrm{N}\left(\mu, \frac{4}{n}\right)$, by CLT, assuming n is large.

$$\mathrm{P}\left(\left|\overline{X} - \mu\right| < 0.1\right) \geq 0.95$$
$$\mathrm{P}\left(-0.1 < \overline{X} - \mu < 0.1\right) \geq 0.95$$
$$\mathrm{P}\left(-0.1 + \mu < \overline{X} < 0.1 + \mu\right) \geq 0.95$$
$$\mathrm{P}\left(\frac{-0.1}{\sqrt{4/n}} < Z < \frac{0.1}{\sqrt{4/n}}\right) \geq 0.95$$
$$\mathrm{P}\left(Z < -0.05\sqrt{n}\right) \leq 0.025$$
$$-0.05\sqrt{n} \leq -1.959963986$$
$$\sqrt{n} \geq 39.19927972$$
$$n \geq 1536.583531$$

So, least value of n = 1537.

---

5. In the study of the sodium content, x, of a cereal, a random sample of 50 similar servings of Yummy Bites were taken. The following are the measurements of x (in milligrams):

$$\sum(x-190) = 1600 \quad \text{and} \quad \sum(x-190)^2 = 65000 \quad .$$

(i) Find the unbiased estimates of the mean and variance of X.

(ii) State the distribution of $\overline{X}$, the mean mass of 50 servings and any assumptions made. Hence find the least value of a such that $\mathrm{P}\left(\overline{X} > a\right) < 0.9$.

Solution

(i) $\mathit{Let}\ y = x - 190$

$$\overline{y} = \overline{x} - 190$$
$$\overline{x} = \overline{y} + 190$$
$$\overline{x} = \frac{1600}{50} + 190 = 222$$
$$s^2 = \frac{1}{49}[65000 - \frac{(1600)^2}{50}] = \frac{13800}{49} = 281.63$$

(ii) $\bar{X} \sim \mathrm{N}\left(222, \frac{281.63}{50}\right)$ approx by CLT since n=50 is large

$$\mathrm{P}\left(\bar{X} > a\right) < 0.9$$

$$\mathrm{P}\left(\bar{X} < a\right) > 0.1$$

$$\mathrm{P}\left(Z < \frac{a-222}{\sqrt{\frac{281.63}{50}}}\right) \geq 0.1$$

$$\frac{a-222}{\sqrt{\frac{281.63}{50}}} \geq -1.28156$$

$$a \geq 218.96$$

$$Least\, a = 219$$

---

6. **[**N03/P2/27] The random variable X has the distribution N(1, 20).

(i) Given that $\mathrm{P}\left(X < a\right) = 2\mathrm{P}\left(X > a\right)$, find a.

(ii) A random sample of n observations of X is taken. Given that the probability that the sample mean exceeds 1.5 is at most 0.01, find the set of possible values of n.

Solution

(i) $P(X < a) = 2P(X > a)$ $3\mathrm{P}(X \leq a) = 2$ $\mathrm{P}(X \leq a) = \frac{2}{3}$ $a = 2.93$	(ii) $\bar{X} \sim \mathrm{N}\left(1, \frac{20}{n}\right)$ $\mathrm{P}(\bar{X} > 1.5) \leq 0.01$ $\mathrm{P}(\bar{X} \leq 1.5) \geq 0.99$ $\mathrm{P}\left(Z \leq \frac{1.5-1}{\sqrt{\frac{20}{n}}}\right) \geq 0.99$ $\frac{1.5-1}{\sqrt{\frac{20}{n}}} \geq 2.326$ $0.5\sqrt{n} \geq \sqrt{20} \times 2.326$ $n \geq 433$ The set of values of n is $\{n$: n is an integer, $n \geq 433\}$

7. The random variable X has mean μ and variance 9. A random sample of size n ($n > 50$) is taken from the population and the sample mean is denoted by $\overline{X}$. Find the least value of n such that $\mathrm{P}(|\overline{X} - \mu| < 0.5) > 0.96$.

Solution

Assumptions: n is large, $n > 50$ or X follows a normal distribution

By CLT, $\overline{X} \sim \mathrm{N}(\mu, \frac{9}{n})$ approximately

$$\mathrm{P}(|\overline{X} - \mu| < 0.5) > 0.95 \quad \Rightarrow \mathrm{P}(-0.5 < \overline{X} - \mu < 0.5) > 0.95$$

$$\Rightarrow \mathrm{P}\left(\frac{-0.5\sqrt{n}}{3} < Z < \frac{0.5\sqrt{n}}{3}\right) > 0.95$$

$$\Rightarrow \frac{0.5\sqrt{n}}{3} > 1.96$$

$$\Rightarrow n > 138.3$$

$$\Rightarrow \text{Least value of } n = 139$$

8. **2021 Prelim/ACJC/H2/P2/Q9 [Modified]**

The Particle Filtration Efficiency (PFE) of a mask is a measure of how well a mask filters airborne particles such as pollen or dust. A mask with higher PFE is deemed to be of better efficiency as it filters more particles. A mask with a PFE of 95% would have met the requirement for surgical masks.

A factory manufactures Brand BEY surgical masks that is known to have expected PFE of 95.8%. During a routine check of the manufacturing process, the quality control manager suspects that the efficiency of the Brand BEY surgical masks produced is compromised such that the mean PFE is reduced. The PFE, x%, of a random sample of 50 masks is taken and the summarised results are as follows.

$$\sum(x-90) = 289 \qquad \sum(x-90)^2 = 1670.56$$

(i) State what it means for a sample to be random in this context. [1]

The manager carries out a test on the PFE of the Brand BEY masks, which requires the distribution of the sample mean.

(ii) Calculate the unbiased estimates of the population mean and variance for the PFE of Brand BEY masks. [2]

(iii) State the distribution of the sample mean and explain whether there is a need for the manager to make any assumption about the population distribution of the PFE of the masks. [1]

Solution:

8(i)	A sample is random if <u>every Brand BEY surgical mask manufactured has an equal chance of being selected to be in the sample of the 50 taken</u> and that the selections are independent of each other. Or A sample is random if <u>every subset of *n* of Brand BEY surgical masks has an equal chance of being in the sample of the 50 taken</u>.
(ii)	Let X be random variable for the PFE of Brand BEY masks. Unbiased estimate of population mean, $\bar{x} = \frac{289}{50} + 90 = 95.78$ Unbiased estimate of population variance, $s^2 = \frac{1}{49}\left[1670.56 - \frac{(289)^2}{50}\right] = 0.0028571.$
(iii)	As $n = 50$ is large, by Central Limit Theorem, $\bar{X} \sim N(95.78, \frac{0.0028571}{50})$ approx. i.e. $\bar{X} \sim N(95.78, 0.000051742)$ There is no need for any assumptions to be made about the population distribution of PFE of masks since n = 50 is large, by central limit theorem, the sample mean PFE will follow a normal distribution approximately.

9. **2020 Prelim/NYJC/H2/P2/Q8(Part of)**

A manufacturer produces teacups, saucers and plates. Past records indicate that on average 5% of the teacups, 2% of the saucers and 1% of the plates produced are flawed. The quality of the teacups, saucers and plates are independent of one another. A box contains a teacup, a saucer and a plate. A box is considered imperfect if any of the three items are flawed. A consignment consists of 250 boxes. The number of boxes in a consignment that are imperfect is denoted by X. You may assume that X can be modelled by a binomial distribution.

(i) Show that the probability that a randomly selected box is imperfect is 0.07831. [1]

(iv) Using a suitable approximation, find the probability that the total number of boxes that are imperfect in a sample of 100 consignments is at most 2000. State an assumption that you have made in your calculations. [4]

[Solution]:

(i)	$\mathrm{P}\left(\text{a box is imperfect}\right) = 1 - \mathrm{P}(\text{a box is perfect})$ $= 1-(0.95)(0.98)(0.99)$ $= 0.07831$
(iv)	Let X be the random variable denoting number of boxes out of 250 (in a consignment) that are imperfect. $X \sim \mathrm{B}(250, 0.07831)$ $\mathrm{E}(X) = np = 250\times 0.07831 = 19.5775$ (exact value, accept 19.578) $\mathrm{Var}(X) = np\left(1-p\right) = 250\times 0.07831\times\left(1-0.07831\right) = 18.044$ (5.s.f.) Let $T = X_1 + ... + X_{100}$. Since $n = 100$ is large, by Central Limit Theorem, $T \sim \mathrm{N}(19.5775\times 100, 18.044\times 100)$ approximately i.e. $T \sim \mathrm{N}(1957.75, 1804.4)$ Required probability = $\mathrm{P}(T \le 2000) = 0.84004 = 0.840$ An assumption is that whether a consignment has imperfect boxes is independent of other consignments. Or: The 100 consignments are randomly chosen.

10. **2021 Prelim/SAJC/H2/P2/Q11(Modified)**

Anne, a Bubble Tea (BBT) seller, intends to increase the sales of BBT using a drink vending machine which delivers BBT into a cup when cash payment is made into the machine. The volume of BBT dispensed is normally distributed with mean 210 ml and standard deviation 5 ml. The capacity of a cup is 220 ml and the nominal amount of BBT in a cup is stated as 212 ml.

(a) A customer bought *n* cups of BBT. Find the approximate probability that the total volume of BBT dispensed exceeds 211*n* ml as *n* becomes very large. [2]

(b) Anne wishes to gather feedback about her BBT. She decided to interview 50 customers who bought BBT from the vending machine during lunch time.
Give a reason as to whether Anne would obtain a random sample of customers. [1]

(c) On another occasion, Anne deployed a staff to operate a BBT counter at a wedding reception. The average volume of BBT per serving is 200 ml and the standard deviation of the volume of BBT is given to be 10 ml. Find the probability that the mean volume of 60 servings of BBT prepared by the staff is less than 198 ml. [3]

[Solution]:

(a)	$X_1 + X_2 + ... + X_n \sim N(210n, 25n)$ $P(X_1 + X_2 + ... + X_n > 211n)$ $= P(Z > \frac{211n - 210n}{5\sqrt{n}})$ $= P(Z > \frac{1}{5}\sqrt{n}) --- (*)$ Since *n* becomes very large , $P(Z > \frac{1}{5}\sqrt{n}) \to 0.$
(b)	The BBT seller would not obtain a random sample as she only interviewed customers who bought her BBT during lunch time. She would not be able to get feedback from customers who made the purchase during other hours, hence **not all the customers have an equal chance of being selected.** The sample is not randomly selected as result.
(c)	Let *T* be the random variable denoting the volume of bubble tea served in ml. $\overline{T} \sim N(200, \frac{100}{60})$ approximately by Central Limit Theorem since the sample size 60 > 30. Using GC, $P(\overline{T} < 198) = 0.0607$ (3s.f)