RAFFLES INSTITUTION
H2 Mathematics (9758)
2025 Year 6

Tutorial S4: Sampling

1 VJC Prelim 9740/2014/02/5 1st part

There were 500 spectators seated at the grandstand of a stadium. A surveyor took the seat plan of the grandstand and threw a die on the seat plan 20 times and interviewed the spectators at the seats on which the die landed. Give a reason why this might not result in a random sample of spectators. **[1]**

Solution

The surveyor is likely to avoid throwing at the edge of the map. Hence not all the 500 seats have an equal chance of being selected.

Alternative Solution

After a seat is selected, the surveyor is likely to avoid the nearby seats. Hence the selections are not independent of one another.

2 9758/2019/02/Q6

In a certain country there are 100 professional football clubs, arranged in 4 divisions. There are 22 clubs in Division One, 24 in Division Two, 26 in Division Three and 28 in Division Four.

(i) Alice wishes to find out about approaches to training by clubs in Division One, so she sends a questionnaire to the 22 clubs in Division One. Explain whether these 22 clubs form a sample or a population. **[1]**

(ii) Dilip wishes to investigate the facilities for supporters at the football clubs. but does not want to obtain the detailed information necessary from all 100 clubs. Explain how he should carry out his investigation, and why he should do the investigation in this way. **[2]**

(iii) Find the number of different possible samples of 20 clubs, with 5 clubs chosen from each division. **[3]**

Solution

(i) These 22 clubs form the population as they are **ALL** the clubs in Division One whose approaches to training she is interested to find out.

(ii) **How**

Assuming no special treatment with regard to facilities for supporters of different divisions, he could randomly pick a certain number of clubs out of the 100, say 10, to do a thorough investigation.

Why

This is to avoid bias and it would also be more cost effective and manageable.

(iii) Number of different possible samples = ${}^{22}C_5 \times {}^{24}C_5 \times {}^{26}C_5 \times {}^{28}C_5 = 7.24 \times 10^{18}$

3 In each of the following cases, find unbiased estimates of the population mean and population variance of X.

(a) sample size = 100, $\sum x = 160$, $\sum x^2 = 265$,

(b) sample size = 150, $\sum(x-154) = 150$, $\sum(x-154)^2 = 11000$,

(c) sample size = 50, $\sum(x-10) = 328$, $\sum x^2 = 16062$,

(d) sample size = 20, $\sum x = 1100$, sample variance = 10.

[(a) $\frac{8}{5}$, $\frac{1}{11}$ (b) 155, $\frac{10850}{149}$ (c) $\frac{414}{25}$, $\frac{8394}{175}$ (d) 55, $\frac{200}{19}$]

Solution

	An unbiased estimate of population mean	**An unbiased estimate of population variance**
(a)	$\bar{x} = \frac{\sum x}{100} = \frac{8}{5} = 1.6$	$s^2 = \frac{1}{99}\left(265 - \frac{(160)^2}{100}\right) = \frac{1}{11}$
(b)	Let $y = x - 154$ Then $\bar{y} = \frac{\sum y}{150} = \frac{150}{150} = 1$ $\bar{y} = \bar{x} - 154 \Rightarrow \bar{x} = \bar{y} + 154 = 155$ An unbiased estimate of population mean is 155	$s_y{}^2 = \frac{1}{n-1}\left(\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right)$ $= \frac{1}{149}\left(11000 - \frac{(150)^2}{150}\right)$ $= \frac{10850}{149} = 72.8$ (3 sf). $s_y{}^2 = s_x{}^2$ So an unbiased estimate of population variance is $\frac{10850}{149}$ or 72.8 (3 sf).
Recall: If $y = ax+b$, then $\bar{y} = a\bar{x} + b$ and $s_y^2 = a^2 s_x^2$		
(c)	Let $y = x - 10$ Then $\bar{y} = \frac{\sum y}{50} = \frac{328}{50} = \frac{164}{25}$ $\bar{y} = \bar{x} - 10 \Rightarrow \bar{x} = \bar{y} + 10 = \frac{414}{25}$ An unbiased estimate of population mean is $\frac{414}{25}$ or 16.56	$s_x{}^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right]$ $= \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right)$ $= \frac{1}{49}\left(16062 - 50(16.56)^2\right)$ $= \frac{8394}{175}$

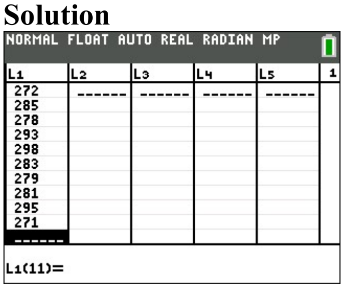
(d)	$\bar{x} = \dfrac{\sum x}{20} = 55$	$s^2 = \dfrac{n}{n-1}$(sample variance) $= \dfrac{20}{19}(10)$ $= \dfrac{200}{19}$ Note : Sample variance = $\dfrac{1}{n}\left(\sum x^2 - \dfrac{1}{n}\left(\sum x\right)^2\right)$ and $s^2 = \dfrac{1}{n-1}\left(\sum x^2 - \dfrac{1}{n}\left(\sum x\right)^2\right)$ so an unbiased estimate of population variance is $\dfrac{n}{n-1}$(Sample variance)

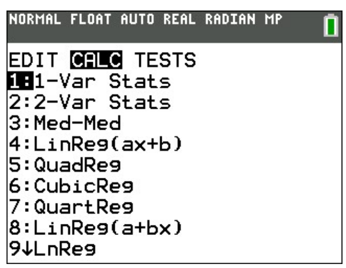
4 The amount, x mg, of Vitamin B2 in a packet of cereals was measured. 10 packets of cereals were taken and the following data were obtained:

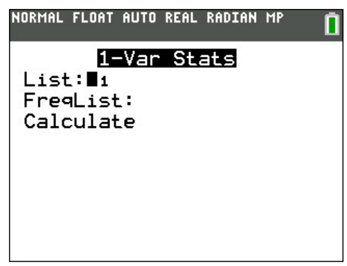
272, 285, 278, 293, 298, 283, 279, 281, 295, 271

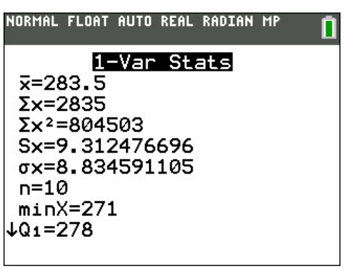
Find unbiased estimates of the population mean and population variance of Vitamin B2 in packets of the cereal. [283.5, 86.7]

Solution









An unbiased estimate of the population mean of Vitamin B2 in a packet of cereal is <u>283.5mg</u>.

A unbiased estimate of the population variance of Vitamin B2 in a packet of cereal is $s^2 \approx (9.3125)^2 \text{mg}^2$

$= \underline{86.7 \text{ mg}^2}$ (3 s.f.)

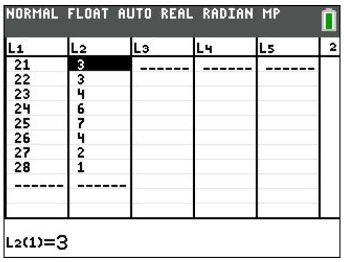
5 A sample of 30 salesmen was surveyed and the ages of each of the salesman are given as follows:

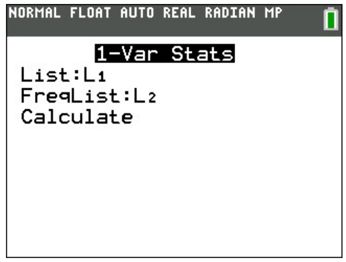
Age	21	22	23	24	25	26	27	28
Frequency	3	3	4	6	7	4	2	1

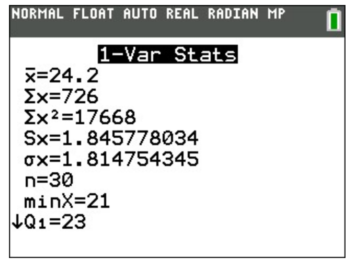
Calculate unbiased estimates for the population mean μ and population variance σ^2.

[24.2, 3.41]

Solution







Unbiased estimates of mean μ and variance σ^2 are 24.2 and $(1.8458)^2 \approx 3.41$

6 Given that the random variables X and Y have distributions $\mathrm{N}(30,18)$ and $\mathrm{N}(20,16)$ respectively, and that $\bar{X}$ and $\bar{Y}$ denote the means of 15 independent observations of X and 8 independent observations of Y respectively.
State the sampling distributions of

(i) $\bar{X}-\bar{Y}$, **(ii)** $5\bar{X}+3\bar{Y}$, **(iii)** $4\bar{X}-2\bar{Y}$, **(iv)** $\dfrac{X_1+X_2+\ldots+X_{15}+Y_1+Y_2+\ldots+Y_8}{23}$.

Solution

Given $X \sim \mathrm{N}(30,18)$ and $Y \sim \mathrm{N}(20,16)$ with $n_x = 15$ and $n_y = 8$

Then, $\bar{X} \sim \mathrm{N}\left(30,\ \dfrac{18}{15}\right)$ and $\bar{Y} \sim \mathrm{N}\left(20,\ \dfrac{16}{8}\right)$

i.e., $\bar{X} \sim \mathrm{N}(30,\ 1.2)$ and $\bar{Y} \sim \mathrm{N}(20,\ 2)$

(i)

$\mathrm{E}\left(\bar{X}-\bar{Y}\right) = \mathrm{E}\left(\bar{X}\right) - \mathrm{E}\left(\bar{Y}\right) = 30-20 = 10$

$\mathrm{Var}\left(\bar{X}-\bar{Y}\right) = \mathrm{Var}\left(\bar{X}\right) + \mathrm{Var}\left(\bar{Y}\right) = 1.2+2 = 3.2$

$\bar{X}-\bar{Y} \sim \mathrm{N}(10,\ 3.2)$

(ii)

$\mathrm{E}\left(5\bar{X}+3\bar{Y}\right) = 5\mathrm{E}\left(\bar{X}\right) + 3\mathrm{E}\left(\bar{Y}\right) = 5(30)+3(20) = 210$

$\mathrm{Var}\left(5\bar{X}+3\bar{Y}\right) = 5^2\mathrm{Var}\left(\bar{X}\right) + 3^2\mathrm{Var}\left(\bar{Y}\right) = 5^2(1.2)+3^2(2) = 48$

$5\bar{X}+3\bar{Y} \sim \mathrm{N}(210,\ 48)$

(iii) $\mathrm{E}\left(4\overline{X}-2\overline{Y}\right)=4\mathrm{E}\left(\overline{X}\right)-2\mathrm{E}\left(\overline{Y}\right)=4(30)-2(20)=80$

$\mathrm{Var}\left(4\overline{X}-2\overline{Y}\right)=4^2\mathrm{Var}\left(\overline{X}\right)+2^2\mathrm{Var}\left(\overline{Y}\right)=2^2(1.2)+2^2(2)=27.2$

$4\overline{X}-2\overline{Y}\sim \mathrm{N}\left(80,\ 27.2\right)$

(iv) Let $T=\dfrac{X_1+X_2+\ldots+X_{15}+Y_1+Y_2+\ldots+Y_8}{23}$

$\mathrm{E}(T)=\dfrac{1}{23}\left[15\mathrm{E}(X)+8\mathrm{E}(Y)\right]=\dfrac{610}{23}$

$\mathrm{Var}(T)=\left(\dfrac{1}{23}\right)^2\left[15\mathrm{Var}(X)+8\mathrm{Var}(Y)\right]=\dfrac{398}{529}$

$T\sim \mathrm{N}\left(\dfrac{610}{23},\dfrac{398}{529}\right)$

7 A large number of samples of size n are taken from a normal population with mean 74 and standard deviation 6. It was found that 28.2% of the samples have sample mean that exceeds 75. Estimate the value of n. [12]

Solution

Let X be a random variable taken from the normal population with mean 74 and standard deviation 6.

Then $X\sim \mathrm{N}(74,\ 6^2)$

$\overline{X}\sim \mathrm{N}(74,\dfrac{36}{n})$

$\mathrm{P}(\overline{X}>75)=0.282$

$\Rightarrow \mathrm{P}\left(Z>\dfrac{75-74}{6/\sqrt{n}}\right)=0.282$

$\Rightarrow \mathrm{P}\left(Z>\dfrac{\sqrt{n}}{6}\right)=0.282$

From GC, $\mathrm{P}(Z>0.57691)=0.282$

$\Rightarrow \dfrac{\sqrt{n}}{6}\approx 0.57691$

$\Rightarrow n\approx 11.982$

$\therefore n\approx 12$

OR:

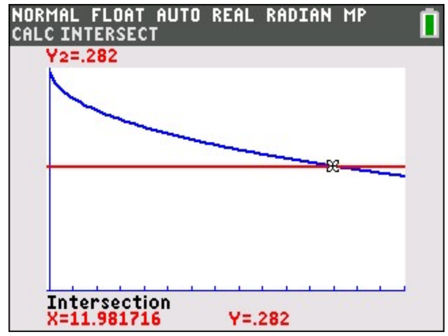
$\overline{X} \sim \mathrm{N}(74, \frac{36}{n})$

$\mathrm{P}(\overline{X} > 75) = 0.282$

By sketching y=$\mathrm{P}(\overline{X} > 75)$ and $y = 0.282$

where $Y_1 = normalcdf(75, 2^{99}, 74, \frac{6}{\sqrt{X}})$

$Y_2 = 0.282$



From the graphs, $n \approx 11.9817 \approx 12$

OR : Sketch $Y_1 = normalcdf(75, 2^{99}, 74, \frac{6}{\sqrt{X}}) - 0.282$ and find intersection with x-axis

8 TJC Prelim 9758/2021/02/Q9(i) – (iii)

A random variable X has probability distribution given by

$$\mathrm{P}(X = x) = \begin{cases} kx(6-x), & x = 1,2,3,4,5 \\ 0, & \text{otherwise.} \end{cases}$$

(i) Show that $k = \frac{1}{35}$. **[1]**

(ii) Find the mean and variance of X. **[3]**

40 independent observations of X are taken.

(iii) Find the probability that the sum of the 40 observations of X exceeds 110. **[3]**

[**(ii)** $\mathrm{E}(X) = 3$, $\mathrm{Var}(X)$=1.6, **(iii)** 0.894]

Solution

(i)

$\sum \mathrm{P}(X = x) = 1$

$5k + 8k + 9k + 8k + 5k = 1$

$35k = 1$

$k = \frac{1}{35}$

(ii)

x	1	2	3	4	5
P(X=x)	$5k$	$8k$	$9k$	$8k$	$5k$

$\mathrm{E}(X) = 3$ (by symmetry)

$\mathrm{Var}(X)$=$\mathrm{E}(X^2) - [\mathrm{E}(X)]^2$

$=1^2(5k) + 2^2(8k) + 3^2(9k) + 4^2(8k) + 5^2(5k) - 3^2$ $=k(371) - 3^2$

$=\frac{1}{35}(371) - 3^2$ $=1.6$

(iii)	Let $S = X_1 + X_2 + \ldots + X_{40}$. Since sample size, $n = 40$ is large, by Central Limit Theorem, $S \sim \mathrm{N}\left(40\times 3, 40\times 1.6\right)$ approximately. i.e. $S \sim \mathrm{N}\left(120, 64\right)$ approximately. $\mathrm{P}\left(S > 110\right) = 0.894$ (3 s.f.)

9 HCI Prelim 9758/2021/02/Q7 ((v) modified)

A toy factory manufactures gel beads which are polymer beads that increase in size when soaked in water. On average, 8% of the gel beads are defective. The gel beads are packed in bags of 500. A significant number of customers recently gave feedback that many of the gel beads they bought could not expand in water or cracked while expanding. The quality control department decides to take a random sample of 20 gel beads from each bag to test. If more than 4 gel beads are found to be defective in the sample of 20, the bag is rejected. Otherwise the bag is accepted.

(i) State, in context, two assumptions needed for the number of defective gel beads in the sample to be well modelled by a binomial distribution. **[2]**

Assume now that the number of defective gel beads in a sample of 20 is modelled by a binomial distribution.

(ii) Find the probability that a randomly chosen bag of gel beads is rejected. **[1]**

(iii) An officer from the quality control department is in charge of inspecting 10 randomly chosen bags of gel beads. Find the probability that the last bag inspected is the second bag that is being rejected. **[2]**

(iv) A random sample of 20 gel beads is taken from a particular bag. Given that the bag is rejected, find the probability that there are more than 13 gel beads with no defects in the random sample of 20 gel beads. **[3]**

(v) The quality control department now decides to investigate 50 randomly chosen bags of gel beads. For each bag, a random sample of 20 gel beads were tested. Estimate the probability that the average number of defective gel beads of the 50 samples will not exceed 1.5. **[2]**

[**(ii)** 0.0183 **(iii)** 0.00261 **(iv)** 0.965 **(v)** 0.280]

Solution	
(i)	Assumptions • The event that a gel bead is defective is independent of other gel beads. • The probability that a gel bead is defective is a constant at 0.08.
(ii)	Let X be the number of gel beads, out of 20, that are defective. $X \sim \mathrm{B}\left(20, 0.08\right)$

	P(randomly chosen bag is rejected) $=\mathrm{P}(X>4)$ $=1-\mathrm{P}(X\le 4)$ $=0.018344$ $=0.0183$ (3 s.f.)
(iii)	<u>Method 1</u> Required probability $= {}^{9}C_{1}\times 0.018344\times(1-0.018344)^{8}\times 0.018344$ $=0.00261$ (3 s.f.) <u>Method 2</u> Let Y be the number of bags that is being rejected, out of 9. Then $Y\sim \mathrm{B}(9, 0.018344)$ Required probability $=\mathrm{P}(Y=1)\times\mathrm{P}(\text{10th bag is rejected})$ $=0.14236\times 0.018344$ $=0.00261$ (3 s.f.)
(iv)	Let X' be the number of non-defective gel beads in a random sample of 20. Then $X'=20-X$. $\mathrm{P}(X'>13 \mid X\ge 5)$ $=\dfrac{\mathrm{P}(X'\ge 14 \text{ and } X\ge 5)}{\mathrm{P}(X\ge 5)}$ $=\dfrac{\mathrm{P}(X\le 6 \text{ and } X\ge 5)}{\mathrm{P}(X\ge 5)}$ $=\dfrac{\mathrm{P}(5\le X\le 6)}{\mathrm{P}(X\ge 5)}$ $=0.965$ (3 s.f.)
(v)	$\mathrm{E}(X)=20\times 0.08=1.6$ $\mathrm{Var}(X)=20\times 0.08\times 0.92=1.472$ Since $n=50$ is large, by Central Limit Theorem, $\overline{X}\sim \mathrm{N}\left(1.6, \dfrac{1.472}{50}\right)$ approximately. $\mathrm{P}(\overline{X}\le 1.5)=0.280$ (3 s.f.)

10 VJC Prelim 9758/2021/02/Q10 (modified)

In this question you should state clearly all the distributions that you use, together with the values of the appropriate parameters.

A certain bakery bakes two types of cookies; butter cookies and chocolate cookies. The masses of butter cookies have the distribution $\mathrm{N}\left(15, 0.4^2\right)$ and the masses of chocolate cookies have the distribution $\mathrm{N}\left(20, 1.2^2\right)$. The units for mass are grams.

(i) Find the probability that the mass of a randomly chosen butter cookie is more than 15.5 grams. **[1]**

(ii) 10 butter cookies are randomly chosen. Find the probability that at least 4 of them each has mass more than 15.5 grams. **[3]**

The cookies are sold by weight. Butter cookies cost \$6 per 100 grams and chocolate cookies cost \$7.50 per 100 grams.

(iii) Miss Lee bought 12 butter cookies and 12 chocolate cookies for her family. Find the probability that she paid less than \$29. **[4]**

(iv) State an assumption needed for your calculations in part **(iii)**. **[1]**

The waiting time, T minutes, before a customer is served at the bakery has a mean of 16 minutes and a standard deviation of 9 minutes.

(v) Give a reason why a normal distribution, with this mean and standard deviation, would not give a good approximation to the distribution of T. **[1]**

(vi) The waiting times of n randomly chosen customers at the bakery are taken, where $n > 30$. Given that the probability that the average waiting time of these n customers is between 16 minutes and 18 minutes is more than 0.48, estimate the least value of n. **[5]**

[(i) 0.106 (ii) 0.0155 (iii) 0.732 (vi) 86]

(i)	Let X and Y be the masses of a butter cookie and a chocolate cookie respectively. $X \sim \mathrm{N}\left(15, 0.4^2\right)$ and $Y \sim \mathrm{N}\left(20, 1.2^2\right)$ $\mathrm{P}\left(X > 15.5\right) \approx 0.10565 = 0.106 \quad (3 \text{ s.f.})$
(ii)	Let W be the number of butter cookies, out of 10, with mass more than 15.5g. $W \sim \mathrm{B}\left(10, 0.10565\right)$ $\mathrm{P}\left(W \geq 4\right) = 1 - \mathrm{P}\left(W \leq 3\right) = 0.0155 \;\; (3 \text{ s.f.})$

(iii)

Let $S = \frac{6}{100}\left(X_1 + \ldots + X_{12}\right) + \frac{7.5}{100}\left(Y_1 + \ldots + Y_{12}\right)$

$$\mathrm{E}(S) = \frac{6}{100}(12)\mathrm{E}(X) + \frac{7.5}{100}(12)\mathrm{E}(Y) = \frac{6}{100}(12)(15) + \frac{7.5}{100}(12)(20) = 28.8$$

$$\mathrm{Var}(S) = \left(\frac{6}{100}\right)^2(12)\mathrm{Var}(X) + \left(\frac{7.5}{100}\right)^2(12)\mathrm{Var}(Y) = \left(\frac{6}{100}\right)^2(12)\left(0.4^2\right) + \left(\frac{7.5}{100}\right)^2(12)\left(1.2^2\right)$$

$= 0.104112$ (exact value)

$S \sim \mathrm{N}\left(28.8, 0.104112\right)$

$\mathrm{P}\left(S < 29\right) = 0.732$ (3 s.f.)

(iv)

We assume that the masses of cookies are independent of each other.

Note: It is **insufficient** to say that "mass of a butter cookie is independent of the mass of a chocolate cookie". This is because for the 12 butter cookies bought, we also need the condition that the mass of a butter cookie is independent of the mass of any other butter cookie. The same goes for the 12 chocolate cookies bought.

(v)

If $T \sim \mathrm{N}\left(16, 9^2\right)$, then about 99.7% of the values would lie within $16 \pm 3(9)$, which contains a significant range of negative values.
In fact, $\mathrm{P}\left(T < 0\right) = 0.0377$ (3 s.f.)

Hence a normal distribution, with this mean and standard deviation, would not give a good approximation to the distribution of T.

(vi)

Since $n > 30$, by Central Limit Theorem,

$\overline{T} \sim \mathrm{N}\left(16, \frac{9^2}{n}\right)$ approximately.

$$\mathrm{P}\left(16 < \overline{T} < 18\right) > 0.48$$

$$\mathrm{P}\left(\overline{T} > 18\right) < 0.5 - 0.48 = 0.02$$

$$\mathrm{P}\left(Z > \frac{18-16}{9/\sqrt{n}}\right) < 0.02$$

$$\mathrm{P}\left(Z > \frac{2\sqrt{n}}{9}\right) < 0.02$$

From the GC, $\mathrm{P}\left(Z > 2.0537\right) = 0.02$

From the diagram,

$$\frac{2\sqrt{n}}{9} > 2.0537$$

$$n > 85.4$$

$\therefore$ Least $n = 86$

Alternative:

n	$\mathrm{P}\left(16 < \overline{T} < 18\right)$
85	$0.4798 < 0.48$
86	$0.4803 > 0.48$

$\therefore n \geq 86$