Exercise Point estimates and confidence intervals for μ (using normal distribution)

1. The concentrations, in milligrams per litre, of a trace element in 7 randomly chosen samples of water from a spring were

   240.8, 237.3, 236.7, 236.6, 234.2, 233.9, 232.5.

   Determine the unbiased estimates of the mean and the variance of the concentration of the trace element per litre of water from the spring. (*L*)

$\mu = \bar{x} = 236$

$(2.75318)^2$

$\sigma^2 = 7.58$

2. Find the best unbiased estimates of the mean μ and variance σ^2 of the population from which each of the following samples is drawn. It is a good idea to do parts (a) to (c) both with and without a calculator.

   (a) 46, 48, 51, 50, 45, 53, 50, 48

   (b) 1.684, 1.691, 1.687, 1.688, 1.689, 1.688, 1.690, 1.693, 1.685

   (c)

x	20	21	22	23	24	25
f	4	14	17	26	20	9

   (d) $\Sigma x = 120$, $\Sigma x^2 = 2\,102$, $n = 8$

   (e) $\Sigma x = 100$, $\Sigma x^2 = 1\,028$, $n = 10$

   (f) $n = 34$, $\Sigma x = 330$, $\Sigma x^2 = 23\,700$

(a) $\mu = 49$

$\sigma^2 = (2.6424)^2 = 6.98.$

(c) $\mu = 22.8$

$\sigma^2 = S_{n-1}^2 = (1.3449)^2 = 1.81.$

(d) $\mu = \bar{x} = \frac{\Sigma x}{n} = 15$

$\sigma^2 = S_{n-1}^2$

$= \frac{n}{n-1} S_n^2$

$= \frac{n}{n-1} \times 37.75$

$= 43.1$

3. A measuring rule was used to measure the length of a rod of stated length 1 m. On 8 successive occasions the following results, in millimetres, were obtained.

   1000, 999, 999, 1002, 1001, 1000, 1002, 1001.

   Calculate unbiased estimates of the mean and, to two significant figures, the variance of the errors occurring when the rule is used for measuring a 1 m length. (*L*)

4. Cartons of orange are filled by a machine. A sample of 10 cartons selected at random from the production contained the following quantities (in millilitres)

   201.2 205.0 209.1 202.3 204.6
   206.4 210.1 201.9 203.7 207.3

   Calculate unbiased estimates of the mean and variance of the population from which the sample was taken. (*L*)

5. A certain type of tennis ball is known to have a height of bounce which is normally distributed with standard deviation 2 cm. A sample of 60 tennis balls is tested and the mean height of bounce of the sample is 140 cm.

   (a) Find a 95% confidence interval for the mean height of bounce of this type of tennis ball.

   (b) State any assumptions made in calculating your interval.

6. A random sample of 6 items taken from a normal population with mean μ and variance 4.5 cm^2 gave the following data:
   Sample values: 12.9 cm, 13.2 cm, 14.6 cm, 12.6 cm, 11.3 cm, 10.1 cm.

   (a) Find the 95% confidence interval for μ.

   (b) What is the width of this confidence interval?

7. A factory produces cans of meat whose masses are normally distributed with standard deviation 18 g. A random sample of 25 cans is found to have a mean mass of 458 g.

   (a) Obtain the 99% confidence interval for the population mean mass of a can of meat produced at the factory.

   (b) Explain what the interval means.

   (c) Would the interval be wider if a 90% confidence interval was calculated? Explain your reasoning.

8. A random sample of 100 observations from a normal population with mean μ gave the following data: $\Sigma x = 8200$, $\Sigma x^2 = 686\,800$.

   (a) Find a 98% confidence interval for μ.

   (b) Find a 99% confidence interval for μ.

   (c) Would your answers have been different if the population was not normal? Explain your answer.

9. Eighty employees at an insurance company were asked to measure their pulse rates when they woke up in the morning. The researcher then calculated the mean and the standard deviation of the sample and found these to be 69 beats and 4 beats respectively. Calculate a 97% confidence interval for the mean pulse rate of all the employees at the company, stating any assumptions that you have made.

10. One hundred and fifty bags of flour are taken from a production line and found to have a mean mass of 748 g and standard deviation of 3.6 g.

    (a) Calculate an unbiased estimate of the standard deviation of a bag of flour produced on this production line.

    (b) Calculate a 98% confidence interval for the mean mass of a bag of flour produced on this production line.

    (c) State any assumptions you have made.

1. 236, 7.58
2. (a) 48.875, 6.98 (b) 1.69, 8×10^{-6} (1 s.f.)
   (c) 22.79, 1.81
   (d) 15, 43.14 (e) 10, 3.11 (f) 9.71, 621.12
3. 0.5, 1.428
4. 205.16, 9.223
5. (a) (139.16, 140.5) (b) random sample
6. (a) (10.75, 14.15) (b) 3.4
7. (a) (448.7, 467.3)
   (b) The probability that this interval includes μ is 0.99.
   (c) No, z value less
8. (a) (79.19, 84.81) (b) (78.89, 85.11)
   (c) No, the central limit theorem can be used, since n is large.
9. (68.0, 70.0), random sample, central limit theorem can be applied.
10. (a) 3.612 (b) (747.3 g, 748.7 g)
    (c) random sample, central limit theorem can be applied.

Estimators

1a. *[3 marks]*

A random variable X has a population mean μ.

Explain briefly the meaning of

(i) an estimator of μ; A ~~parame~~ statistics to estimate the value of unknown μ.

(ii) an unbiased estimator of μ. $E(x) = \mu$.

1b. *[12 marks]*

A random sample $X_1,\ X_2,\ X_3$ of three independent observations is taken from the distribution of X.

An unbiased estimator of $\mu,\ \mu \neq 0$, is given by $U = \alpha X_1 + \beta X_2 + (\alpha - \beta)X_3$,

$\alpha + \beta + \alpha - \beta = 1$

where $\alpha,\ \beta \in \mathbb{R}$.

✓ (i) Find the value of α. $\frac{1}{2}$

✓ (ii) Show that $\mathrm{Var}(U) = \sigma^2\left(2\beta^2 - \beta + \frac{1}{2}\right)$ where $\sigma^2 = \mathrm{Var}(X)$.

$\mathrm{Var}(U) = \alpha^2 \mathrm{Var}(x_1) + \beta^2 \mathrm{Var}(x_2) + (\alpha^2 - 2\alpha\beta + \beta^2)\mathrm{Var}(x_3)$

$= 2\alpha^2 \mathrm{Var}(x) - 2\alpha\beta \mathrm{Var}(x) + 2\beta^2 \mathrm{Var}(x)$

$= \sigma^2\left(\frac{1}{2} - \beta + 2\beta^2\right)$ ✓

(iii) Find the value of β which gives the most efficient estimator of μ of this form.

~~$2\beta^2 - \beta + \frac{1}{2} = 1$ $2\beta^2 - \beta - \frac{1}{2} = 0$~~

(iv) Write down an expression for this estimator and determine its variance.

~~$(2\beta - 1)^2$ $4\beta^2 - 2\beta - 1 = 0$~~

~~$\frac{2 \pm \sqrt{2+16}}{8} = \frac{2 \pm 3\sqrt{2}}{8}$~~

(v) Write down a more efficient estimator of μ than the one found in (iv), justifying your answer.

μ

2a. *[3 marks]*

The continuous random variable X takes values in the interval $[0,\ \theta]$ and

$\mathrm{E}(X) = \frac{\theta}{2}$ and $\mathrm{Var}(X) = \frac{\theta^2}{24}$.

To estimate the unknown parameter θ, a random sample of size n is obtained from the distribution of X. The sample mean is denoted by $\overline{X}$ and $U = k\overline{X}$ is an unbiased estimator for θ.

Find the value of k.

2b. *[4 marks]*

(i) Calculate an unbiased estimate for θ, using the random sample, 8.3, 4.2, 6.5, 10.3, 2.7, 1.2, 3.3, 4.3.

(ii) Explain briefly why this is not a good estimate for θ.

2c. *[8 marks]*

(i) Show that $\mathrm{Var}(U) = \frac{\theta^2}{6n}$.

(ii) Show that U^2 is not an unbiased estimator for θ^2.

(iii) Find an unbiased estimator for θ^2 in terms of U and n.

3a. *[4 marks]*

A biased cubical die has its faces labelled $1, 2, 3, 4, 5$ and 6. The probability of rolling a 6 is p, with equal probabilities for the other scores.

The die is rolled once, and the score X_1 is noted.

(i) Find $\mathrm{E}(X_1)$.

(ii) Hence obtain an unbiased estimator for p.

3b. *[7 marks]*

The die is rolled a second time, and the score X_2 is noted.

(i) Show that $k(X_1 - 3) + \left(\frac{1}{3} - k\right)(X_2 - 3)$ is also an unbiased estimator for p for all values of $k \in \mathbb{R}$.

(ii) Find the value for k, which maximizes the efficiency of this estimator.

Exercise - The distribution of the sample mean, $\bar{X}$, for

- samples of any size from a normal population
- large samples from a non-normal population

1. The volumes of wine in bottles are normally distributed with a mean of 758 ml and a standard deviation of 12 ml. A random sample of 10 bottles is taken and the mean volume found.
Calculate the probability that the sample mean is less than 750 ml.

2. The heights of a new variety of sunflower can be modelled by a normal distribution with mean 2 m and standard deviation of 40 cm.

(a) A random sample containing 50 sunflowers is taken and the mean height calculated. What is the probability that the sample mean lies between 195 cm and 205 cm?
(b) A hundred such samples, each of 50 observations, are taken. In how many of these would you expect the sample mean to be greater than 210 cm?

3. In an examination taken by a large number of students the mean mark was 64.5 and the variance was 64. The mean mark in a random sample of 100 scripts is denoted by $\bar{X}$. Find
(a) $P(\bar{X} > 65.5)$
(b) $P(63.8 < \bar{X} < 64.5)$

4. The mean of 50 observations of X, where $X \sim B(12, 0.4)$, is $\bar{X}$.
(a) State the approximate distribution of $\bar{X}$.
(b) Hence find $P(\bar{X} < 5)$

5. A normal variable X has standard deviation σ. The mean of 20 independent observations of X is $\bar{X}$.
(a) Given that $\text{Var}(\bar{X}) = 3.2$, find the value of σ.
(b) Would your answer be different if the variable was not normal?

6. Independent observations are taken from a normal distribution with mean 30 and variance 5.
(a) Find the probability that the average of 10 observations exceeds 30.5.
(b) Find the probability that the average of 40 observations exceeds 30.5.
(c) Find the probability that the average of 100 observations exceeds 30.5.
(d) Find the least value of n such that the probability that the average of n observations exceeds 30.5 is less than 1%.

7. The standard deviation of the masses of articles in a large population is 4.55 kg. Random samples of size 100 are drawn from the population. Find the probability that a sample mean will differ from the population mean by less than 0.8 kg.

8. The variable X is such that $X \sim N(\mu, 4)$. A random sample of size n is taken from the population. Find the least n such that $P(|\bar{X} - \mu| < 0.5) > 0.95$.

9. (a) A large number of random samples of size n are taken from $B(20, 0.2)$. Approximately 90% of the sample means are less than 4.354. Estimate n.
(b) A large number of random samples of size n are taken from Po(2.9). Approximately 1% of the sample means are greater than 3.41. Estimate n.

10. The random variable X has standard deviation σ. The mean of 40 observations of X is $\bar{X}$. Given that $\text{Var}(\bar{X}) = 0.625$, find the value of σ.

11. The mean of a sample of 100 observations of the random variable X is denoted by $\bar{X}$. The mean of $\bar{X}$ is 20 and the standard deviation of $\bar{X}$ is 0.3. Find the mean and the standard deviation of X.

12. A sample of n independent observations is taken from a normal population with mean 74 and standard deviation 6. The sample mean is denoted by $\bar{X}$.
(a) Find n if $P(\bar{X} > 75) = 0.282$.
(b) Find n if $P(\bar{X} < 70.4) = 0.0037$.

1. 0.0176
2. (a) 0.6234 (b) Approx. 4
3. (a) 0.1056 (b) 0.3092
4. (a) $\bar{X} \sim N\left(4.8, \frac{2.88}{50}\right)$ (b) 0.7975
5. (a) 8 (b) no
6. (a) 0.2399 (b) 0.0787 (c) 0.0127 (d) $n = 109$
7. 0.9212
8. 62
9. (a) 42 (b) 60
10. 5
11. 20, 3
12. (a) 12 (b) 20