

5A Estimators, Bias and Standard Error

Random sample of size n:

Statistic:

1. A sample, $X_1, X_2, \dots, X_n$, is taken from a population with unknown population parameters μ and σ . State whether or not each of the following are statistics.

a $\frac{X_1 + X_3 + X_5}{3}$ b $\max(X_1, X_2, \dots, X_n)$ c $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2$

Sampling distribution of a statistic T:

2. The weights, in grams, of a consignment of apples are normally distributed with a mean μ and standard deviation 4. A random sample of size 25 is taken and the statistics R and T are calculated as follows:

$$R = X_{25} - X_1 \quad \text{and} \quad T = X_1 + X_2 + \dots + X_{25}$$

Find the distributions of R and T.

3. A large bag contains counters. 60% of the counters have the number 0 on them and 40% have the number 1.
- a) Find the population mean μ and population variance σ^2 of the values shown on the counters.

x	0	1
P(X = x)		

A simple random sample of size 3 is taken from this population.

- b) List all the possible observations from this sample.

- c) Find the sampling distribution for the mean

$$\bar{X} = \frac{X_1 + X_2 + X_3}{3}$$

where X_1 , X_2 and X_3 are the values shown on the three counters in the sample.

$\bar{x}$	0	$\frac{1}{3}$	$\frac{2}{3}$	1
P($\bar{X} = \bar{x}$)				

- d) Hence find $E(\bar{X})$ and $\text{Var}(\bar{X})$.

- e) Find the sampling distribution for the mode, M.

m	0	1
P(M = m)		

- f) Hence find $E(M)$ and $\text{Var}(M)$.

Estimator and estimate:

4. A random sample $X_1, X_2, \dots, X_n$ is taken from a population with $X \sim N(\mu, \sigma^2)$. Show that $E(\bar{X}) = \mu$.

Unbiased estimator:

Bias:

Unbiased estimator for σ^2 (the sample variance S^2) and ways to calculate s^2 :

5. A random sample $X_1, X_2, \dots, X_n$ is taken from a population with $X \sim N(\mu, \sigma^2)$.

$$\text{Show that } \text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Standard error:

Standard error of $\bar{X}$:

6. The table below summarises the number of breakdowns, X , on a town's bypass on 30 randomly chosen days.

Number of breakdowns	2	3	4	5	6	7	8	9
Number of days	3	5	4	3	5	4	4	2

- a) Calculate unbiased estimates of the mean and variance of the number of breakdowns.

20 more days were randomly sampled and this sample had $\bar{x} = 6.0$ and $s^2 = 5.0$.

- b) Treating the 50 results as a single sample, obtain further unbiased estimates of the population mean and variance.

- c) Find the standard error of this new estimate of the mean.

- d) Estimate the size of sample required to achieve a standard error of less than 0.25.

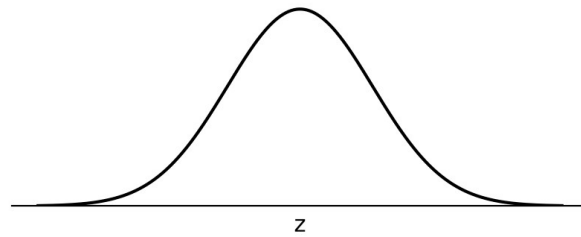
5B Confidence Intervals

Confidence interval:

Distribution of $\bar{X}$ (central limit theorem):

1. Show that a 95% confidence interval for μ , based on a sample of size n , is given by

$$\left(\bar{x} - 1.96 \times \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \times \frac{\sigma}{\sqrt{n}} \right)$$



Confidence limits:

Complete the table of percentage points of $N(0, 1^2)$ used for confidence intervals.

Confidence level	90%	95%	98%	99%
z				

Interpreting a 95% confidence interval:

2. The breaking strains of reels of string produced at a certain factory have a standard deviation of 1.5 kg. A sample of 100 reels from a certain batch were tested and their mean breaking strain was 5.30 kg.
- a) Find a 95% confidence interval for the mean breaking strain of string in this batch.

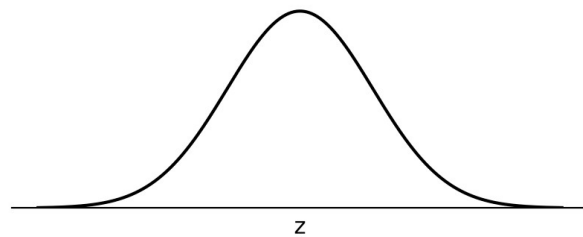
The manufacturer becomes concerned if the lower 95% confidence limit falls below 5 kg. A sample of 80 reels from another batch gave a mean breaking strain of 5.31 kg.

- b) Will the manufacturer be concerned?

Width of a confidence interval:

Factors that affect the width:

3. A random sample of size 25 is taken from a normal population with standard deviation 2.5. The mean of the sample is 17.8.
- a) Find a 99% C.I. for the population mean μ .
- b) What size sample is required to obtain a 99% C.I. of width of at most 1.5?
- c) What confidence level would be associated with the interval based on the above sample of 25 but of width 1.5, i.e. (17.05, 18.55)?



5C Hypothesis Testing for the Difference Between Means

Distribution of $\bar{X} - \bar{Y}$, where $\bar{X}$ and $\bar{Y}$ are the means of samples from independent normal populations X and Y:

Test statistic for the difference between two means:

When the distributions of X and Y are not normal:

1. The weights of boys and girls in a certain school are known to be normally distributed with standard deviations of 5 kg and 8 kg respectively. A random sample of 25 boys had a mean weight of 48 kg and a random sample of 30 girls had a mean weight of 45 kg.
Stating your hypotheses clearly, test, at the 5% level of significance, whether there is evidence that the mean weight of the boys in the school is greater than the mean weight of the girls.

2. The weights of boys and girls in a certain school are known to be normally distributed with standard deviations of 5 kg and 8 kg respectively. A random sample of 25 boys had a mean weight of 48 kg and a random sample of 30 girls had a mean weight of 45 kg.
Stating your hypotheses clearly, test, at the 5% level of significance, whether there is evidence that the mean weight of the boys in the school is more than 2 kg greater than the mean weight of the girls.
3. A manufacturer of personal stereos can use batteries made by two different manufacturers. The standard deviation of lifetimes for Never Die batteries is 3.1 hours and for Everlasting batteries is 2.9 hours. A random sample of 80 Never Die batteries and a random sample of 90 Everlasting batteries were tested and their mean lifetimes were 7.9 hours and 8.2 hours respectively.
Stating your hypotheses clearly, test, at the 5% level of significance, whether there is evidence of a difference between the mean lifetimes of the two makes of batteries.

5D Use of Large Sample Results for an Unknown Population

If σ is unknown and the population is normal (or can be assumed to be), for large samples:

If the population is not normal, for large samples:

1. As part of a study of the health of young schoolchildren, a random sample of 220 children from area A and a second, independent random sample of 180 children from area B were weighed. The results are given in the table below.

	n	$\bar{x}$	s
Area A	220	37.8	3.6
Area B	180	38.6	4.1

- a) Test, at the 5% level of significance, whether there is evidence of a difference in the mean weight of children in the two areas. State your hypotheses clearly.
- b) State an assumption you have made in carrying out this test.
- c) Explain the significance of the central limit theorem to this test.