Distribution of Sample Means

Today we are learning...

The distribution of a sample mean.

I will konw if I have been successful if...

I understand what iid means.

I know the distribution of the sample mean and its parameters.

I can calculate associated probabilites and check my answer using another method.

Distribution of Sample Means

Natural Variation - The variation that occurs between different members of the population.

Sampling Variation - The variation in estimating a population parameter by taking different samples of the population.

Distribution of Sample Means

iid (Independent and identically distributed) -

The random variables X_1, X_2, ..., X_n are said to be iid if they independent of one another and from the same distribution.

Distribution of the Sample Mean

Suppose that $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ then....

$E(\overline{X}) = \mu$ and $Var(\overline{X}) = \frac{\sigma^2}{n}$

Standard Error

The **standard error** is a numerical indicator of the precision of an estimate.

If the standard error is small there is high precision (very accurate).

If the standard error is large there is a low precision (not very accurate).

Standard Error = $\sqrt{V(\overline{X})} = \frac{\sigma}{\sqrt{n}}$

Distribution of the Sample Mean

If the random variable X_i follows a Normal distribution then so does the random variable $\overline{X}$.

If $X_i \sim N(\mu, \sigma^2)$ then....

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Distribution of Sample Means

Example on Page 81

EX3.3

(Use your graphical calculator to check your answer)

Central Limit Theorem

Today we are learning...

What the CLT is and how to apply it.

I will know if I have been successful if...

I understand the conditions required to use the CLT.

Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables, each with expected value μ and variance σ^2. Then for sufficiently large n...

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

(Sufficiently large means n > 20, but the larger n is the better the approximation will be)

Continuity Correction

If we are approximating a discrete random variable with a continuous random variable (the Normal distribution), we must still apply a continuity correct. These were:

$P(X < x) = P(Y < x - 0.5)$

$P(X \leq x) = P(Y \leq x + 0.5)$

$P(X > x) = P(Y > x - 0.5)$

$P(X \geq x) = P(Y \geq x + 0.5)$

Example on Page 83

Complete EX3.4 over Half Term

Distribution of the Sample Proportion

Today we are learning...

What the distribution of the sample proportion is.

I will know if I have been successful if...

I can state the distribution of the sample proportion using the CLT to justify its use.

I know the parameters of the distribution.

I can state the standard error.

Proportion

Occasionally we are interested in proportion rather than the mean. Examples are:

1) Polls - Estimating the proportion of voters who voted for each candidate.

2) Genetics - Estimating the proportion of people who carry a particular gene.

3) Television Rating - Estimating the proportion of households who watch a particular TV show.

Proportion

p - The Population proportion.

$\hat{p}$ - The Sample Proportion

The Distribution of the Sample Proportion

When n > 20, then by the CLT and provided np > 5 and nq > 5 the distribution of the sample proportion is approximately normal.

$$\hat{p} \sim N\left(p, \frac{pq}{n}\right)$$

$$\sqrt{\frac{pq}{n}}$$

is referred to as the standard error of the sample proportion.

Where does this come from?

You are not required to prove this but I shall do it now.

Example: A sample of 300 bolts were taken from a production line and 15 were found to be defective.

The sample proportion is therefore ....

$$\hat{p} = \frac{X}{n} = \frac{15}{300}$$

Where does this come from?

Example: A sample of 300 bolts were taken from a production line and 15 were found to be defective.

The sample proportion is therefore ....

$$\hat{p} = \frac{X}{n} = \frac{15}{300}$$

X in this case is the random variable that counts the number of defective bolts and follows a Binomial distribution.

Where does this come from?

$$\hat{p} = \frac{X}{n}$$

If X ~ B (np, npq) then

$E(\hat{p}) =$

$Var(\hat{p}) =$

And so if n > 20 the CLT applies and np > 5 and nq > 5 we can say

$$p \sim N\left(p, \frac{pq}{n}\right)$$

Example

A certain company's customers is made up of 43% women and 57% men. In a sample of 50 customers what is the probability that at least 46% are women?

Confidence Interval for the Population Mean

Today we are learning...

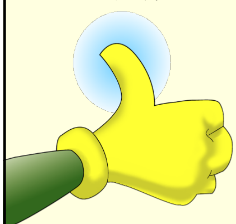
How to construct a confidence interval for a population mean.

I will know if I have been successful if...

I understand how to standardise.

I can find the correct k value in the tables.

I know the general form of the confidence interval.



Previously....

The CLT says that for large enough n, when $X_1, ..., X_n$ are independent and identically distributed random variables each with expected value μ and variance σ^2 then...

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Previously....

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

To calculate $P(\bar{X} < x)$ we first need to standardise to use the tables.

Then $P\left(\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} < x\right)$ or simply $P(Z < x)$

where $\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = Z$

Let's say we've done some calculations and our question has boiled down to this....

$P(-1.96 < Z < 1.96) = ?$

Use your tables to work out the ?

Let's say we've done some calculations and our question has boiled down to this....

$P(-1.96 < Z < 1.96) = 0.95$

Mr Welford will now do some rearranging to obtain a truly remarkable result...

A Confidence Interval for the Population Mean

A 95% CI for the Population Mean is given by

$$\left(\overline{X} - 1.96 \frac{\sigma}{\sqrt{n}} \;,\; \overline{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

For 90% use k = 1.64

For 95% use k = 1.96

For 99% use k = 2.58