# Intro to Hypothesis Testing on the Population Mean

Nicky Grant

University of St Andrews, 31st July 2019

# Motivation

**Student A** claims the average economics grade in the UK population of UGs is no more than 65%

**Student B** disagrees believing this average grade is larger than $65\%$

**How could we form a test of the validity of Student A's claim against B?**

# Motivation

**Student A** claims the average economics grade in the UK population of UGs is no more than $65\%$

**Student B** disagrees believing this average grade is larger than $65\%$

**How could we form a test of the validity of Student A's claim against B?**

*First formally state the claims to be tested on the unknown population mean* $\mu$

# Motivation

**Student A** claims the average economics grade in the UK population of UGs is no more than 65%

**Student B** disagrees believing this average grade is larger than $65\%$

**How could we form a test of the validity of Student A's claim against B?**

*First formally state the claims to be tested on the unknown population mean* $\mu$

**Null Hypothesis:** $H_0 : \mu <= 65$

**Alternative Hypothesis:** $H_A : \mu > 65$

# Motivation

Generally we cannot observe the whole population of data

# Motivation

Generally we cannot observe the whole population of data

Instead we observe a sample drawn from this population

# Motivation

Generally we cannot observe the whole population of data

Instead we observe a sample drawn from this population

*Suppose we sample 25 UK students at random and find the their econ average grade is* **65%**

# Motivation

Generally we cannot observe the whole population of data

Instead we observe a sample drawn from this population

*Suppose we sample 25 UK students at random and find the their econ average grade is* **65%**

**Q: Does this imply Student A's claim (the Null Hypothesis) is correct?**

# Motivation

Generally we cannot observe the whole population of data

Instead we observe a sample drawn from this population

*Suppose we sample 25 UK students at random and find the their econ average grade is* **65%**

**Q: Does this imply Student A's claim (the Null Hypothesis) is correct?**

**Q: If the mean in our sample was 67% does this imply Student A is incorrect?**

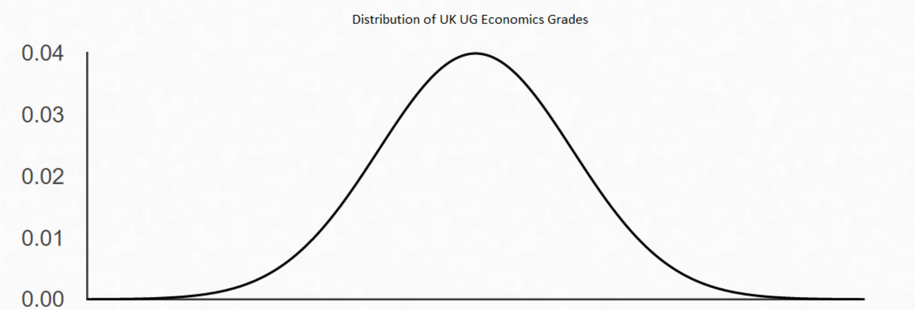
# Sampling Distribution of the Mean

Sample mean of N=25 students economics grades ($\hat{\mu}$) is a random variable

# Sampling Distribution of the Mean

Sample mean of N=25 students economics grades ($\hat{\mu}$) is a random variable

Suppose student grades (x) are normally distributed with mean $\mu$ and variance $\sigma^2 = 100$

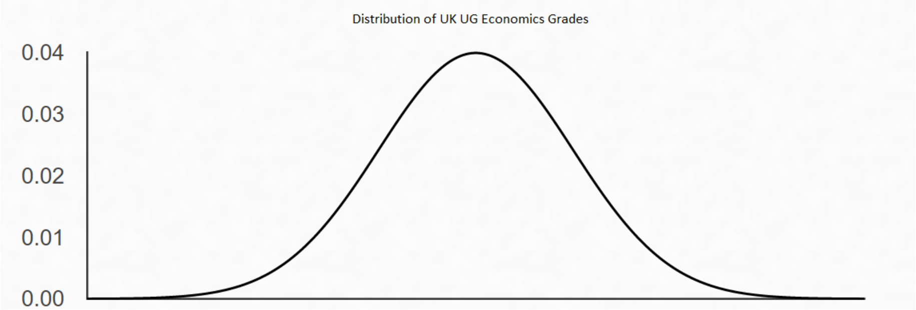



Sample Realisation of Nonstationary AR(1) Processes

# Sampling Distribution of the Mean

Sample mean of N=25 students economics grades ($\hat{\mu}$) is a random variable

Suppose student grades (x) are normally distributed with mean $\mu$ and variance $\sigma^2 = 100$




Sample Realisation of Nonstationary AR(1) Processes

**Sample Mean Distribution**: $\hat{\mu}$ is normally distributed with mean $\mu$ and variance $\frac{\sigma^2}{25} = 4$

# Importance of Sampling Distribution

Suppose again our sample mean of 25 students is $67\%$

**Q: How to work out the probability of observing a sample mean $\geq 67$ if Stud. A is correct?**

## Importance of Sampling Distribution

Suppose again our sample mean of 25 students is $67\%$

**Q: How to work out the probability of observing a sample mean $\geq 67$ if Stud. A is correct?**

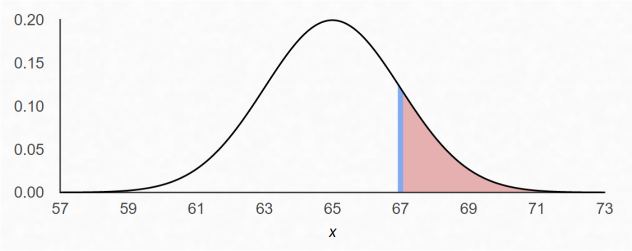
If Student A is correct (Null is true) and $\mu = 65 \Rrightarrow \hat{\mu}$ is normally dist. with $\mu = 65$ and variance 4

# Importance of Sampling Distribution

Suppose again our sample mean of 25 students is $67\%$

**Q: How to work out the probability of observing a sample mean** $\geq 67$ **if Stud. A is correct?**

If Student A is correct (Null is true) and $\mu = 65 \Rrightarrow \hat{\mu}$ is normally dist. with $\mu = 65$ and variance 4




Sample Realisation of Nonstationary AR(1)

Processes

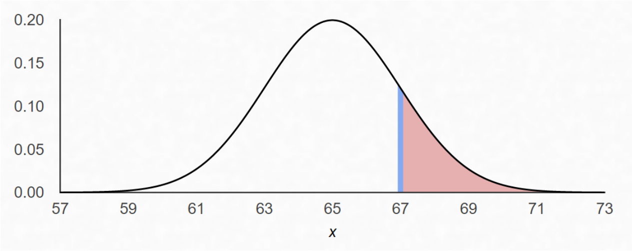
**Q: Does this imply Student A's claim is correct?** $\Pr\{\hat{\mu} > 67|\mu = 65\} = 0.159$

## Importance of Sampling Distribution

Suppose again our sample mean of 25 students is $67\%$

**Q: How to work out the probability of observing a sample mean $\geq 67$ if Stud. A is correct?**

If Student A is correct (Null is true) and $\mu = 65 \Rrightarrow \hat{\mu}$ is normally dist. with $\mu = 65$ and variance 4




Sample Realisation of Nonstationary AR(1)

Processes

**Q: Does this imply Student A's claim is correct?** $\Pr\{\hat{\mu} > 67 | \mu = 65\} = 0.159$

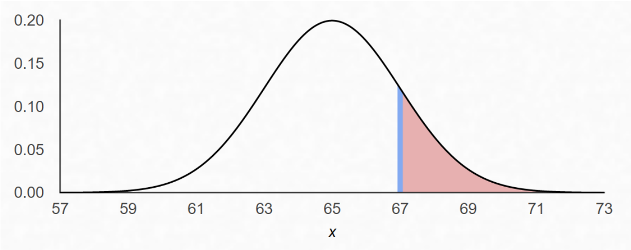
**Q: What if we observe sample mean of 70 where** $\Pr\{\hat{\mu} > 70 | \mu = 65\} = 0.0062$

# Importance of Sampling Distribution

Suppose again our sample mean of 25 students is $67\%$

**Q: How to work out the probability of observing a sample mean $\geq 67$ if Stud. A is correct?**

If Student A is correct (Null is true) and $\mu = 65 \Rrightarrow \hat{\mu}$ is normally dist. with $\mu = 65$ and variance 4




Sample Realisation of Nonstationary AR(1)

Processes

**Q: Does this imply Student A's claim is correct?** $\Pr\{\hat{\mu} > 67 | \mu = 65\} = 0.159$

**Q: What if we observe sample mean of 70 where** $\Pr\{\hat{\mu} > 70 | \mu = 65\} = 0.0062$

Both probabilities are known as **p-values** you may have seen reported on statistical software

# Formalising a Statistical Test

**Form a test where there is $5\%$ chance of rejecting the Null Hypothesis when it is true (Type I Error)**

# Formalising a Statistical Test

**Form a test where there is $5\%$ chance of rejecting the Null Hypothesis when it is true (Type I Error)**

$5\%$ is the **significance level** of our test - the probability with which we are willing to make a Type I Error (other common choices are 10% and 1%)

# Formalising a Statistical Test

**Form a test where there is $5\%$ chance of rejecting the Null Hypothesis when it is true (Type I Error)**

$5\%$ is the **significance level** of our test - the probability with which we are willing to make a Type I Error (other common choices are 10% and 1%)

Need to find '**critical value**' $c$ where $\Pr\{\hat{\mu} > c | \mu = 65\} = 0.05$

# Formalising a Statistical Test

**Form a test where there is $5\%$ chance of rejecting the Null Hypothesis when it is true (Type I Error)**

$5\%$ is the **significance level** of our test - the probability with which we are willing to make a Type I Error (other common choices are 10% and 1%)

Need to find '**critical value**' $c$ where $\Pr\{\hat{\mu} > c|\mu = 65\} = 0.05$

**Test Procedure: Accept** $H_0$ if $\hat{\mu} <= 68.29$ and **Reject** if $\hat{\mu} > 68.29$

# Standardised Test

Often a hypothesis test of the mean **standardise** the test statistic to form acc./rej. region

# Standardised Test

Often a hypothesis test of the mean **standardise** the test statistic to form acc./rej. region

$\hat{\mu} \leq 68.29$ **equivalent to** $\hat{\mu} - 65 \leq 3.29$ **equivalent to** $\frac{\hat{\mu}-65}{2} \leq 1.645$

# Standardised Test

Often a hypothesis test of the mean **standardise** the test statistic to form acc./rej. region

$\hat{\mu} \leq 68.29$ **equivalent to** $\hat{\mu} - 65 \leq 3.29$ **equivalent to** $\frac{\hat{\mu}-65}{2} \leq 1.645$

$\frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}}$ is the **standardised sample mean (z-stat)**

# Standardised Test

Often a hypothesis test of the mean **standardise** the test statistic to form acc./rej. region

$\hat{\mu} \leq 68.29$ **equivalent to** $\hat{\mu} - 65 \leq 3.29$ **equivalent to** $\frac{\hat{\mu}-65}{2} \leq 1.645$

$\frac{\hat{\mu}-\mu}{\sigma/\sqrt{N}}$ is the **standardised sample mean** (**z-stat**)

z-stat is distributed standard normal (mean 0 variance 1)

Perform test of Null Hypothesis by comparing z-stat with critical value from standard normal distribution tables

## Standardised Test

$\mu + 10$ $\mu + 20$ $\mu + 30$ $\mu - 10$ $\mu - 20$ $\mu - 30$