

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

Math 1710 Class 19

CLT, Association, Correlation
Dr. Back

Oct. 9, 2009

Proportion Case of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

p = true proportion

\hat{p} = observed proportion

Proportion Case of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

p = true proportion

\hat{p} = observed proportion

$N(p, \sqrt{\frac{pq}{n}})$ is the sampling distribution of \hat{p} .

Proportion Case of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

p = true proportion

\hat{p} = observed proportion

$N(p, \sqrt{\frac{pq}{n}})$ is the sampling distribution of \hat{p} .

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}}.$$

Proportion Case of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

p = true proportion

\hat{p} = observed proportion

$N(p, \sqrt{\frac{pq}{n}})$ is the sampling distribution of \hat{p} .

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}}.$$

$SD(\hat{p})$ stands for **S**tandard **D**eviation of the **S**ampling **D**istribution of \hat{p} .

Proportion Case of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

p = true proportion

\hat{p} = observed proportion

$N(p, \sqrt{\frac{pq}{n}})$ is the sampling distribution of \hat{p} .

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}}.$$

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

Proportion Case of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

p = true proportion

\hat{p} = observed proportion

$N(p, \sqrt{\frac{pq}{n}})$ is the sampling distribution of \hat{p} .

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}}$$

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$SE(\hat{p})$ stands for the Standard Error of \hat{p} .

Proportion Case of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

p = true proportion

\hat{p} = observed proportion

$N(p, \sqrt{\frac{pq}{n}})$ is the sampling distribution of \hat{p} .

$$SD(\hat{p}) = \sqrt{\frac{pq}{n}}.$$

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

SE is just an approximate version of SD.

Quantitative Case of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

- Suppose X is a random variable with mean μ and standard deviation σ .
- Let X_1, X_2, \dots, X_n be n independent copies of X .

- Set

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

- Then

- 1 The mean of \bar{X} is μ .
- 2 The standard deviation of \bar{X} is $\frac{\sigma}{\sqrt{n}}$.
- 3 \bar{X} is approximately normal as n gets large.

Prop. Case a Special Case of Quant. Case

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

Suppose we poll n people, each with prob. p of answering "yes."

Prop. Case a Special Case of Quant. Case

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

Suppose we poll n people, each with prob. p of a answering
"yes."

Let X_i be an RV describing the number of yeses from person i .

Prop. Case a Special Case of Quant. Case

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

Suppose we poll n people, each with prob. p of a answering "yes."

Let X_i be an RV describing the number of yeses from person i .
 X_i is Bernoulli(p), with $\mu = p$ and $\sigma = \sqrt{pq}$.

Prop. Case a Special Case of Quant. Case

Math 1710
Class 19

V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

Suppose we poll n people, each with prob. p of a answering "yes."

Let X_i be an RV describing the number of yeses from person i . X_i is Bernoulli(p), with $\mu = p$ and $\sigma = \sqrt{pq}$.

Then

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

is just the RV describing the observed proportion \hat{p} .

Prop. Case a Special Case of Quant. Case

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

Suppose we poll n people, each with prob. p of answering "yes."

Let X_i be an RV describing the number of yeses from person i . X_i is Bernoulli(p), with $\mu = p$ and $\sigma = \sqrt{pq}$.

Then

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

is just the RV describing the observed proportion \hat{p} .

Thus for n large, the RV describing the sampling distribution of \hat{p} is approximately

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

in agreement with our earlier proportion result.

CLT Proportion Conditions

Math 1710
Class 19

V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

- random sample
- (plausible independence) Anything silly about data collection?
- success/failure
- 10% condition if pop involved

CLT Quantitative Conditions

Math 1710
Class 19

V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

- plausible independence
- random sampling
- 10% condition
- n large enough

$n < 15$ unimodal and symmetric

$15 \leq n \leq 40$ avoid strong skewness and outliers

$n > 40$ pretty much ok

Conditions for the RV Case of CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

- independence
- n large enough

What Happens if Not Satisfied?

Math 1710
Class 19

V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

random sampling -

- only relevant to population case
- could be critical
- might be ok if "representative"
- representative hard/impossible to define

What Happens if Not Satisfied?

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

plausible independence -

- could be critical
- sometimes just a working hypothesis

What Happens if Not Satisfied?

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

10% condition -

- only relevant to population case
- results in overestimation of samp. dist. st dev
gradual breakdown in formulas, not method
(finite population correction)

What Happens if Not Satisfied?

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

success/failure -

- progressive reduction of accuracy
- accuracy varies regardless for smaller values of n

What Happens if Not Satisfied?

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

n large enough -

- no guarantee
- progressive reduction of accuracy

RV vs. Population View of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

The CLT talks about averages of independent RV's.

RV vs. Population View of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

The CLT talks about averages of independent RV's.

Since independence is assumed, the 10% condition is irrelevant to this mathematical version.

RV vs. Population View of the CLT

Math 1710
Class 19

V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

The CLT talks about averages of independent RV's.

Since independence is assumed, the 10% condition is irrelevant to this mathematical version.

However, a population gives rise to an RV.

RV vs. Population View of the CLT

Math 1710
Class 19

V2u

Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

The CLT talks about averages of independent RV's.

Since independence is assumed, the 10% condition is irrelevant to this mathematical version.

However, a population gives rise to an RV.

For example, the weights of the 6 billion citizens of planet Earth could be viewed as giving a probability model for the weights of people.

RV vs. Population View of the CLT

Math 1710
Class 19

V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

The CLT talks about averages of independent RV's.

Since independence is assumed, the 10% condition is irrelevant to this mathematical version.

However, a population gives rise to an RV.

For example, the weights of the 6 billion citizens of planet Earth could be viewed as giving a probability model for the weights of people.

An RV.

RV vs. Population View of the CLT

Math 1710
Class 19

V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

The CLT talks about averages of independent RV's.

Since independence is assumed, the 10% condition is irrelevant to this mathematical version.

However, a population gives rise to an RV.

For example, the weights of the 6 billion citizens of planet Earth could be viewed as giving a probability model for the weights of people.

An RV.

If this is the kind of RV we are using, then the 10% condition addresses whether assuming independence with a bunch of copies might be a reasonable approximation.

Son's Heights from Their Fathers

Math 1710
Class 19

V2u

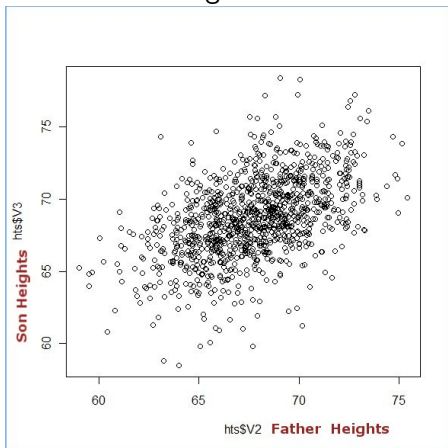
Last Time

A Special Case

CLT
"Conditions"

Graphs and
Association

Galton's Original 1886 Data



Son's Heights from Their Fathers

Math 1710
Class 19

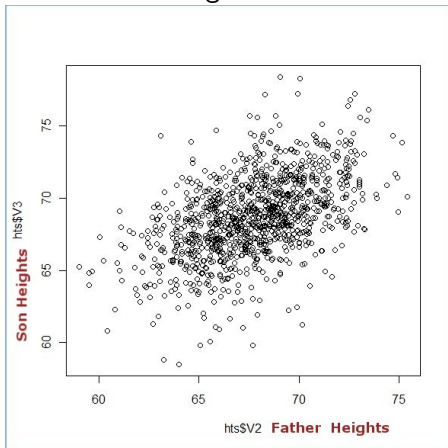
V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

Galton's Original 1886 Data



If you know a father's height, what can you say about his son's?

Son's Heights from Their Fathers

If you know a father's height, what can you say about his son's?

>	hts	V2	V3
1		65.04851	59.77827
2		63.25094	63.21404
3		64.95532	63.34242
4		65.75250	62.79238
5		61.13723	64.28113
6		63.02254	64.24221
7		65.37053	64.08231
8		64.72398	63.99574
9		66.06509	64.61338
10		66.96738	63.97944
...	

Son's Heights from Their Fathers

Math 1710
Class 19

V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

If you know a father's height, what can you say about his son's?

```
>summary(hts$V3)  (Sons)
```

<i>Min.</i>	<i>1stQu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rdQu.</i>	<i>Max.</i>
58.51	66.93	68.62	68.68	70.47	78.36

```
>summary(hts$V2)  (Fathers)
```

<i>Min.</i>	<i>1stQu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rdQu.</i>	<i>Max.</i>
59.01	65.79	67.77	67.69	69.60	75.43

Son's Heights from Their Fathers

If you know a father's height, what can you say about his son's?

```
>summary(hts$V3) (Sons)
```

<i>Min.</i>	<i>1stQu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rdQu.</i>	<i>Max.</i>
58.51	66.93	68.62	68.68	70.47	78.36

```
>summary(hts$V2) (Fathers)
```

<i>Min.</i>	<i>1stQu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rdQu.</i>	<i>Max.</i>
59.01	65.79	67.77	67.69	69.60	75.43

```
> sd(hts$V3)
```

```
[1] 2.814702
```

```
> sd(hts$V2)
```

```
[1] 2.744868
```

Son's Heights from Their Fathers

If you know a father's height, what can you say about his son's?

```
>summary(hts$V3) (Sons)
```

<i>Min.</i>	<i>1stQu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rdQu.</i>	<i>Max.</i>
58.51	66.93	68.62	68.68	70.47	78.36

```
>summary(hts$V2) (Fathers)
```

<i>Min.</i>	<i>1stQu.</i>	<i>Median</i>	<i>Mean</i>	<i>3rdQu.</i>	<i>Max.</i>
59.01	65.79	67.77	67.69	69.60	75.43

```
> cor(hts$V3 , hts$V2)  
[1] 0.5013383
```

Son's Heights from Their Fathers

Math 1710
Class 19

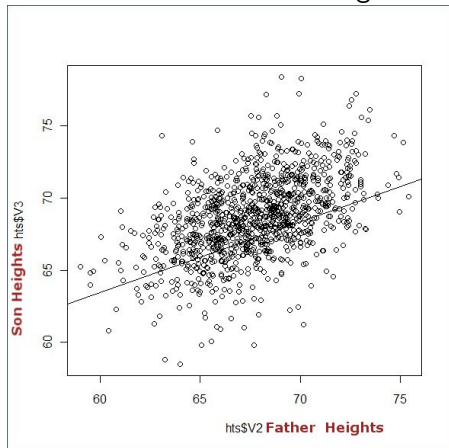
V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

Galton Data with Line of Regression



Son's Heights from Their Fathers

Math 1710
Class 19

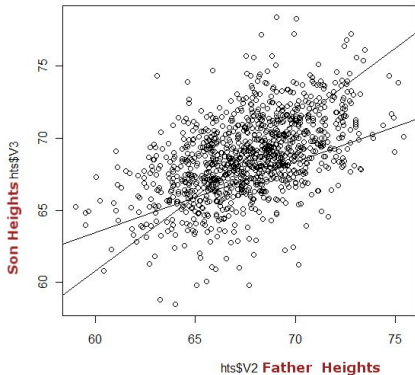
V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

Galton Data with line of slope $\frac{s_{\text{son}}}{s_{\text{father}}}$ added



Son's Heights from Their Fathers

Math 1710
Class 19

V2u

Last Time
A Special Case

CLT
"Conditions"

Graphs and
Association

Galton Data with Other Line of Regression
(predict father height from son)

