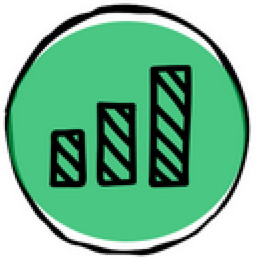
Statistical Techniques for CS/BIT

[Stat] Home Page

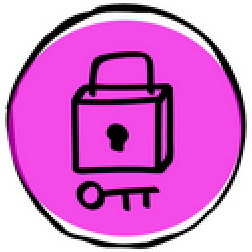
General Information & Planning

Course Materials

Assignments



SPSS



Sample Tests

Today:

Lectorial (meeting #6)

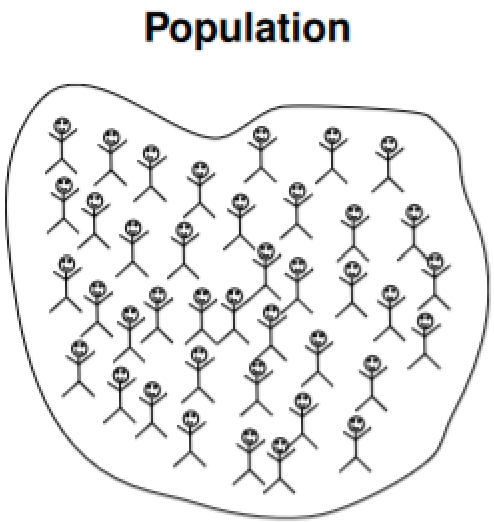
- Lecture CH 3 (3.4 & 3.3)

  08:45-09:30
- Break

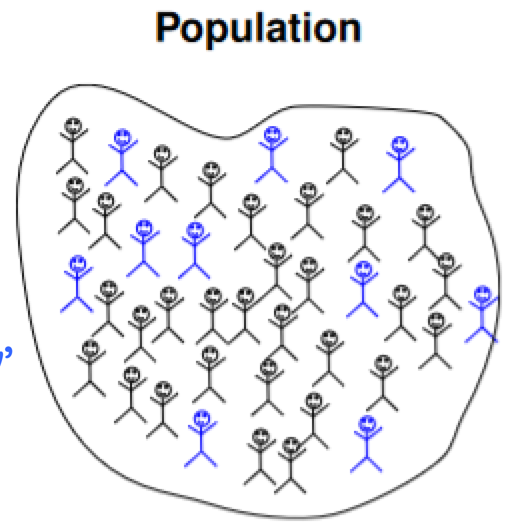
  09:30-09:45
- Tutorial

  09:45-10:30

Population proportion



P: 'success probability'
!



p: the parameter that describes a percentage value in the population

e.g.

p: defective items

p: having a smart watch

P: voting for a party

Dichotomous (1 or 0) (Yes or No)

Population proportion

Determine if you are dealing with a proportion problem!

Binomial distribution (No reference of a mean or average)

X is a binomial random variable

$X \sim B(n, p)$

n: the number of trials

p: the probability of a success

To form a proportion, take *X*, the random variable for the number of successes and divide it by *n*, the number of trials (or the sample size).

The random variable $\hat{p}$ is that proportion, $\hat{p} = \frac{X}{n}$

Population proportion

Determine if you are dealing with a proportion problem!

Binomial distribution (No reference of a mean or average)

X is a binomial random variable

$$X \sim B(n, p)$$

n: the number of trials

p: the probability of a success

To form a proportion, take X, the random variable for the number of successes and divide it by n, the number of trials (or the sample size).

The random variable $\hat{p}$ is that proportion, $\hat{p}=\frac{X}{n}$

Example

See Reader Example 3.4.1

In a survey with 1000 randomly selected voters, 258 voters found for the Labour party.

p: the proportion of the voters among all voters

$\hat{p}$: estimator for p,

$\hat{p}=\frac{258}{1000} = 0.258$, *p remains unknown!*

X: the number of voters of the Labour party in the sample of 1000 voters.

$$X \sim B(1000, p)$$

Distribution - Population proportion

$\widehat{\boldsymbol{p}}=\frac{X}{n}$ $\quad X \sim \mathcal{B}(n, p)$

The **sample proportion** $\widehat{\boldsymbol{p}}=\frac{X}{n}$ is an unbiased and consistent estimator of the **population proportion** $\boldsymbol{p}$.

For **large** $\boldsymbol{n}$ ($n > 25, np > 5$ and $n(1-p) > 5$)

<u>When n is large and p is not close to zero or one</u>

$$\widehat{\boldsymbol{p}} \overset{\text{CLT}}{\sim} \boldsymbol{N}\left(\boldsymbol{p}, \frac{\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})}{\boldsymbol{n}}\right) \text{ and } \frac{\widehat{\boldsymbol{p}} - \boldsymbol{p}}{\sqrt{\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})/\boldsymbol{n}}} \overset{\text{CLT}}{\sim} \boldsymbol{N}(\mathbf{0}, \mathbf{1})$$

Distribution - Population proportion

$\hat{p}=\frac{X}{n}$ $X \sim B(n, p)$

The **sample proportion** $\hat{p} = \frac{X}{n}$ is an unbiased and consistent estimator of the **population proportion** p.

For **large** n ($n > 25, np > 5$ and $n(1-p) > 5$)

When n is large and p is not close to zero or one

$$\hat{p} \overset{\text{CLT}}{\sim} N\left(p, \frac{p(1-p)}{n}\right) \text{ and } \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \overset{\text{CLT}}{\sim} N(0,1)$$

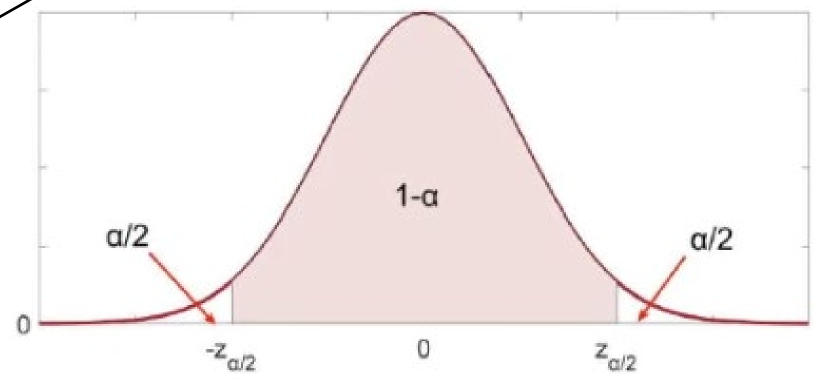
Example (cont.)

n=1000, large n,

X *is approximately* $N(np, np(1-p))$ distributed

$\hat{p}=\frac{X}{n}$ *is also approximately* $N(p, \frac{p(1-p)}{n})$ distributed

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \overset{\text{CLT}}{\sim} N(0,1)$$



Confidence Interval - Population proportion

$$\hat{\boldsymbol{p}} \overset{\text{CLT}}{\sim} \boldsymbol{N}\left(\boldsymbol{p}, \frac{\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})}{\boldsymbol{n}}\right) \text{ and } \frac{\hat{\boldsymbol{p}} - \boldsymbol{p}}{\sqrt{\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})/\boldsymbol{n}}} \overset{\text{CLT}}{\sim} \boldsymbol{N}(\mathbf{0},\mathbf{1})$$

An approximate confidence interval for p can be found by solving

$$\boldsymbol{P}\left(-\boldsymbol{c} < \frac{\hat{\boldsymbol{p}} - \boldsymbol{p}}{\sqrt{\boldsymbol{p}(\mathbf{1}-\boldsymbol{p})/\boldsymbol{n}}} < \boldsymbol{c}\right) = \mathbf{1} - \boldsymbol{\alpha},$$

where $\boldsymbol{\Phi}(\boldsymbol{c}) = \mathbf{1} - \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\alpha}$

Confidence Interval - Population proportion

$$\hat{p} \overset{\text{CLT}}{\sim} N\left(p, \frac{p(1-p)}{n}\right) \text{ and } \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \overset{\text{CLT}}{\sim} N(0,1)$$

An approximate confidence interval for p can be found by solving

$$P\left(-c < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < c\right) = 1 - \alpha,$$

where $\Phi(c) = 1 - \frac{1}{2}\alpha$

$$\longrightarrow \quad P\left(-c < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < c\right) = 1 - \alpha,$$

Confidence Interval - Population proportion

$$\hat{p} \overset{\text{CLT}}{\sim} N\left(p, \frac{p(1-p)}{n}\right) \text{ and } \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \overset{\text{CLT}}{\sim} N(0,1)$$

An approximate confidence interval for p can be found by solving

$$P\left(-c < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < c\right) = 1 - \alpha,$$

where $\Phi(c) = 1 - \frac{1}{2}\alpha$

→

$$P\left(-c < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < c\right) = 1 - \alpha,$$

↓

?

Confidence Interval - Population proportion

$$\hat{p} \overset{\text{CLT}}{\sim} N\left(p, \frac{p(1-p)}{n}\right) \text{ and } \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \overset{\text{CLT}}{\sim} N(0,1)$$

An approximate confidence interval for p can be found by solving

$$P\left(-c < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < c\right) = 1 - \alpha,$$

where $\Phi(c) = 1 - \frac{1}{2}\alpha$

$$\longrightarrow \quad P\left(-c < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < c\right) = 1 - \alpha,$$

a further approximation:

the standard deviation $\sqrt{\frac{p(1-p)}{n}}$ is estimated by $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, the standard error of $\hat{p}$.

Confidence Interval - Population proportion

$$-c < \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} < c \Leftrightarrow -c < \frac{p - \hat{p}}{\sqrt{\hat{p}(1-\hat{p})/n}} < c$$

$$\Leftrightarrow \hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$(1-\alpha)$ 100% - $\mathrm{CI}(p)$: $\left(\hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

Confidence Interval - Population proportion

$$-c < \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} < c \iff -c < \frac{p - \hat{p}}{\sqrt{\hat{p}(1-\hat{p})/n}} < c$$

$$\iff \hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

(1– α) 100% - CI(p) : $\left(\hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

Example (cont)

α=0.10, C=1.645, $\hat{p} = 0.258$

90 % - CI(p) :

$$\left(0.258 - 1.645\sqrt{\frac{0.258(1-0.258)}{1000}}, 0.258 + 1.645\sqrt{\frac{0.258(1-0.258)}{1000}}\right)$$

≈ (0.235, 0.281)

- Approximate confidence interval for the population proportion of the Labour party voters at a 90% level of confidence.
- "We are 90% confident that the Labour party will have between 23.5 and 28.1 % voters"

Sample size n

What if we want to estimate the proportion of the Labour party voters with precision 0.1 and confidence level 95%?..

$X \sim B(n, p)$

$$95\% \text{ - } \mathrm{CI}(p): \quad \left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Sample size n

What if we want to estimate the proportion of the Labour party voters with precision 0.1 and confidence level 95%?..

$X \sim \mathcal{B}(n, p)$

95% - $\mathrm{CI}(p)$: $\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

$\frac{1.96}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} = 0.1$ = W, or $\mathrm{n} \geq (2x1.96)^2 \frac{\hat{p}(1-\hat{p})}{0.1^2}$

Sample size n

What if we want to estimate the proportion of the Labour party voters with precision 0.1 and confidence level 95%?..

$X \sim \mathcal{B}(n, p)$

$$95\% \text{ - CI}(p): \left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$\frac{1.96}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} = 0.1$ = W, or $n \geq (2x1.96)^2 \frac{\hat{p}(1-\hat{p})}{0.1^2}$

$\hat{p}$: unknown!

What to do ?

Sample size n

What if we want to estimate the proportion of the Labour party voters with precision 0.1 and confidence level 95%?..

$X \sim \mathcal{B}(n, p)$

$$95\% \text{ - CI}(p): \left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$\frac{1.96}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} = 0.1$ = W, or $n \geq (2x1.96)^2 \frac{\hat{p}(1-\hat{p})}{0.1^2}$

$\hat{p}$: unknown!

i) Replace $\hat{p}\ (1-\hat{p})$ by its maximal 'reasonable 'value

ii) Choose $\hat{p}$ to maximize $\hat{p}\ (1-\hat{p})$ ($\hat{p}$ =1/2)

Sample size n

What if we want to estimate the proportion of the Labour party voters with precision 0.1 and confidence level 95%?..

$X \sim B(n, p)$

95% - $\mathrm{CI}(p)$: $\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

$\frac{1.96}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} = 0.1$ = W, or $n \geq (2x1.96)^2 \frac{\hat{p}(1-\hat{p})}{0.1^2}$ → Solve for n!

$\hat{p}$: unknown!

i) Replace $\hat{p}(1-\hat{p})$ by its maximal 'reasonable 'value

ii) Choose $\hat{p}$ to maximize $\hat{p}(1-\hat{p})$ ($\hat{p}$ =1/2) → Solve for n!

$$n \geq (2x1.96)^2 \frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{0.1^2} = 384.16$$

$$n \geq 385$$

Sample size n

What if we want to estimate the proportion of the Labour party voters with precision 0.1 and confidence level 95%?..

$X \sim B(n, p)$

95% - $\text{CI}(p)$: $\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

$\frac{1.96}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} = 0.1 = W$, or $n \geq (2x1.96)^2 \frac{\hat{p}(1-\hat{p})}{0.1^2}$

Solve for n!

$\hat{p}$: unknown!

i) Replace $\hat{p}\,(1-\hat{p})$ by its maximal 'reasonable 'value

ii) Choose $\hat{p}$ to maximize $\hat{p}\,(1-\hat{p})$ ($\hat{p}$ =1/2)

$$n \geq (2x1.96)^2 \frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{0.1^2} = 384.16$$

$$n \geq 385$$

What if we want to estimate the proportion of the Labour party voters with precision 0.01 and confidence level 95%?..

How will the sample size change?

Sample size n

What if we want to estimate the proportion of the Labour party voters with precision 0.1 and confidence level 95%?..

$X \sim B(n, p)$

$$95\%\text{ - CI}(p):\ \left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

$\frac{1.96}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} = 0.1$ = W, or $n \geq 1.96^2\,\frac{\hat{p}(1-\hat{p})}{0.1}$ → Solve for n!

$\hat{p}$: unknown!

i) Replace $\hat{p}\,(1-\hat{p})$ by its maximal 'reasonable 'value

ii) Choose $\hat{p}$ to maximize $\hat{p}\,(1-\hat{p})$ ($\hat{p}$ =1/2)

$$n \geq 1.96^2\,\frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{0.1^2} = 96.04$$

$$n \geq 97$$

What if we want to estimate the proportion of the Labour party voters with precision 0.01 and confidence level 95%?..

$$n \geq (2x1.96)^2\frac{(\widehat{p}(1-\hat{p}))^2}{0.01^2}$$

Sample size n

What if we want to estimate the proportion of the Labour party voters with precision 0.1 and confidence level 95%?..

$X \sim B(n, p)$

95% - $\mathrm{CI}(p)$: $\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

$\frac{1.96}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} = 0.1$ = W, or $n \geq 1.96^2\frac{\hat{p}(1-\hat{p})}{0.1}$

Solve for n!

$n \geq 1.96^2\frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{0.1^2} = 96.04$

$n \geq 97$

$\hat{p}$: unknown!

i) Replace $\hat{p}\,(1-\hat{p})$ by its maximal 'reasonable 'value

ii) Choose $\hat{p}$ to maximize $\hat{p}\,(1-\hat{p})$ ($\hat{p}$ =1/2)

What if we want to estimate the proportion of the Labour party voters with precision 0.01 and confidence level 95%?..

$$n \geq (2x1.96)^2\frac{(\widehat{p}(1-\hat{p}))^2}{0.01^2}$$

Solve for n!

$$n \geq (2x1.96)^2\frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{0.01^2} = 38416$$

Sample size n

What if we want to estimate the proportion of the Labour party voters with precision 0.1 and confidence level 95%?..

$X \sim B(n, p)$

$$95\% - \mathrm{CI}(p): \left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

$\frac{1.96}{\sqrt{n}}\sqrt{\hat{p}(1-\hat{p})} = 0.1 = W$, or $n \geq 1.96^2 \frac{\hat{p}(1-\hat{p})}{0.1}$

Solve for n!

$$n \geq 1.96^2 \frac{\frac{1}{2}(1-\frac{1}{2})}{0.1^2} = 96.04$$

$$n \geq 97$$

$\hat{p}$: unknown!

i) Replace $\hat{p}\ (1-\hat{p})$ by its maximal 'reasonable 'value

ii) Choose $\hat{p}$ to maximize $\hat{p}\ (1-\hat{p})$ ($\hat{p}$ =1/2)

What if we want to estimate the proportion of the Labour party voters with precision 0.01 and confidence level 95%?..

$$n \geq (2x1.96)^2 \frac{(\widehat{p}(1-\hat{p}))^2}{0.01^2}$$

Solve for n!

$$n \geq (2x1.96)^2 \frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{0.01^2} = 38416$$

Sample size n (for given interval width (W) and 1 - α)

1. The width of a $(1-\alpha)100\%$-CI(μ) for known $\boldsymbol{\sigma^2}$ is at most $\boldsymbol{W}$: solve $\boldsymbol{n}$ from $\boldsymbol{2 \cdot c \frac{\sigma}{\sqrt{n}} \leq W}$

2. The width of a $(1-\alpha)100\%$-CI(μ) for unknown $\boldsymbol{\sigma^2}$ is at most $\boldsymbol{W}$: can we solve $\boldsymbol{n}$ from $\boldsymbol{2 \cdot c \frac{s}{\sqrt{n}} \leq W}$?
   2 problems:
   - **$\boldsymbol{\sigma}$ and $\boldsymbol{s}$ are unknown** → use an estimate or maximum value
   - **$\boldsymbol{df = n - 1}$ unknown** → use c from the $N(0,1)$-table (approximately correct if n is large)

3. The width of a $(1-\alpha)100\%$-CI($\boldsymbol{p}$) is at most $\boldsymbol{W}$: Solve $\boldsymbol{n}$ from $\boldsymbol{2 \cdot c \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \leq W}$

   Problem: $\boldsymbol{\hat{p}}$ is unknown.
   - use an estimate of p (if available, e.g. from a initial small sample)
   - or otherwise use the property: $\boldsymbol{\hat{p}(1-\hat{p}) \leq 0.25}$

Questions ?

Confidence Interval for the variance σ²

Confidence interval for the population mean μ

$(1-\alpha)$-CI for μ when σ is unknown

$$\text{CI}(\mu, S, \alpha) = \left(\bar{X} - t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right),$$

Example: Average height of Dutch men

Random Sample
176.2
180.4
190.1
194.8
188.4
190.7
189.4
185.3
170.6
173.2
$n = 10$
$\bar{x} = 183.91$
$s = 8.292$

- Determine a 95%-confidence interval for the average height of Dutch men!
- $\text{CI} = \left(\bar{X} - t_{n-1,1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + t_{n-1,1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$.
- $t_{n-1,1-\frac{\alpha}{2}} = t_{9,0.975} = 2.262$
- $n = 10$, $\bar{x} = 183.91$ and $s = 7.9980$.
- The 95%-confidence interval is given by

$$(177.978\,,\ 189.842).$$

Confidence Interval for the variance σ^2

It is also important that this μ does not have too much variability!

The question: What is the variability ?

(1- α)%-CI(σ^2)

Confidence Interval for the variance σ^2

It is also important that this μ does not have too much variability!

The question: What is the variability ?

$(1-\alpha)\%$-CI(σ^2)

To construct this CI, we need a quantity that:

(i) involves σ^2 and its estimator s^2 Why?

Confidence Interval for the variance σ^2

It is also important that this μ does not have too much variability!

The question: What is the variability ?

(1- α)%-CI(σ^2)

To construct this CI, we need a quantity that:

(i) involves σ^2 and its estimator s^2 (so that we can construct interval for the σ^2 based on s^2)

(ii) has a known distribution Why?

Confidence Interval for the variance σ^2

It is also important that this μ does not have too much variability!

The question: What is the variability ?

(1- α)%-CI(σ^2)

To construct this CI, we need a quantity that:

(i) involves σ^2 and its estimator s^2 (so that we can construct interval for the σ^2 based on s^2)

(ii) has a known distribution (To construct the boundaries)

a. If $X_1, \ldots, X_n$ is a random sample taken from a $N(\mu, \sigma^2)$-distribution, then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

b. If $Y \sim \chi^2_n$, then we have: $E(Y) = n$ and $var(Y) = 2n$

Confidence Interval for the variance σ^2

It is also important that this μ does not have too much variability!

The question: What is the variability ?

(1- α)%-CI(σ^2)

To construct this CI, we need a quantity that:

(i) involves σ^2 and its estimator s^2 (so that we can construct interval for the σ^2 based on s^2)

(ii) has a known distribution (To construct the boundaries)

a. If $X_1, \ldots, X_n$ is a random sample taken from a $N(\mu, \sigma^2)$-distribution, then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

b. If $Y \sim \chi^2_n$, then we have: $E(Y) = n$ and $var(Y) = 2n$

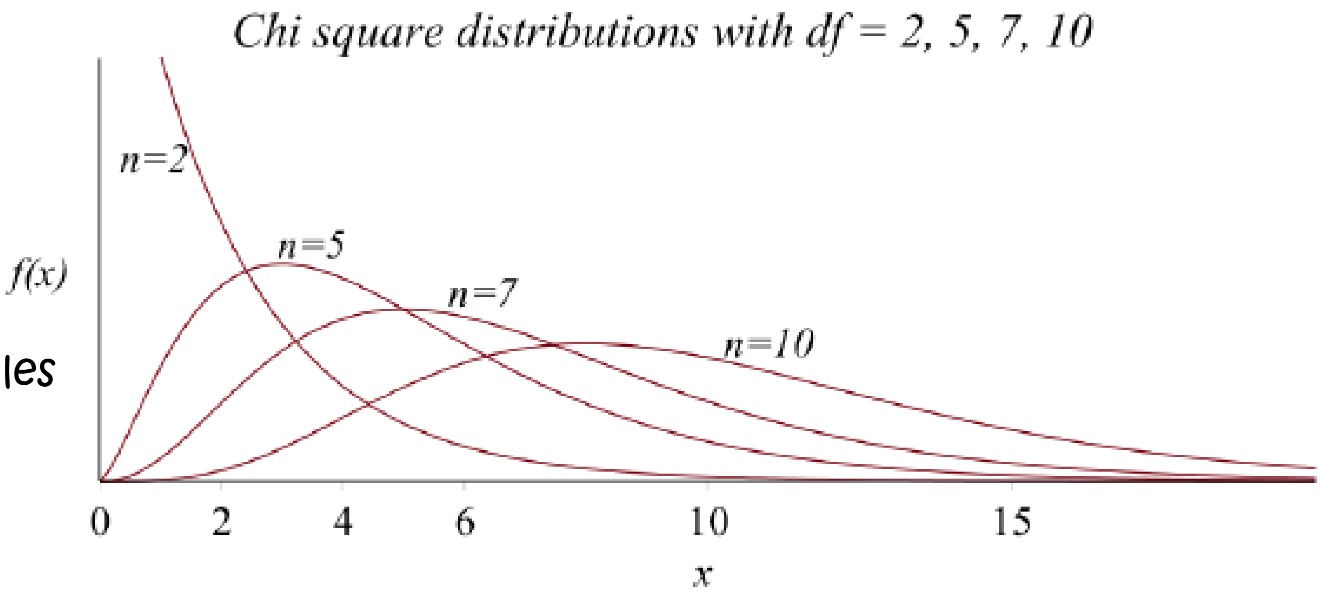
Chi-square distribution

The distribution of the chi-square statistic is called the chi-square distribution

`$X_1, \ldots, X_n$ is a random sample taken from a $N(\mu, \sigma^2)$-distribution,

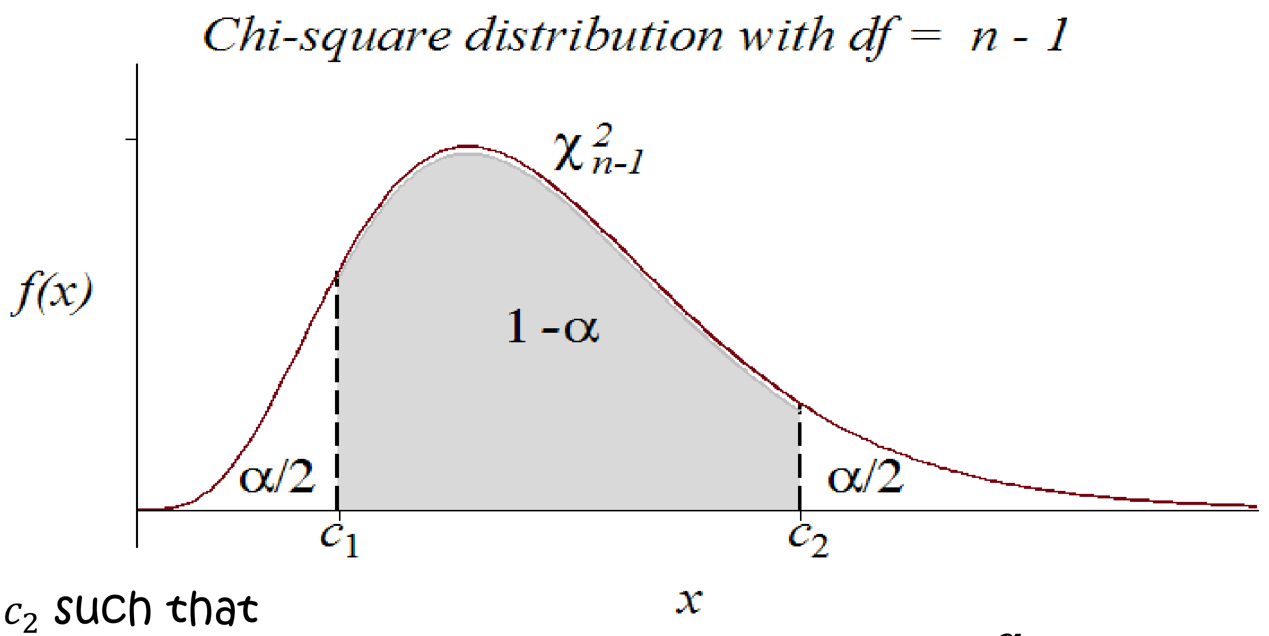
$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

The **chi-square distribution** (χ^2**-distribution**)
with *k* degrees of freedom is
the distribution of a sum of the squares
of *k* independent *standard normal* random variables



Does not depend on μ
χ^2_{n-1} is not normalized wrt n

Construction of confidence interval for σ²



Determine c_1 and c_2 such that

$$P(\chi^2_{n-1} \le c_1) = \boldsymbol{P}(\chi^2_{n-1} \ge c_1) = \frac{\alpha}{2}$$

$$P\left(c_1 < \frac{(n-1)S^2}{\sigma^2} < c_2\right) = 1 - \alpha \Leftrightarrow P\left(\frac{1}{c_2} < \frac{\sigma^2}{(n-1)S^2} < \frac{1}{c_1}\right) = 1 - \alpha \Leftrightarrow P\left(\frac{(n-1)S^2}{c_2} < \sigma^2 < \frac{(n-1)S^2}{c_1}\right) = 1 - \alpha$$

Interval for σ: $P\left(\sqrt{\frac{(n-1)S^2}{c_2}} < \boldsymbol{\sigma} < \sqrt{\frac{(n-1)S^2}{c_1}}\right) = 1 - \alpha$

Example

Determine a 90%-CI for the standard deviation of the yearly returns on investment (see reader exercise 3.3.5)

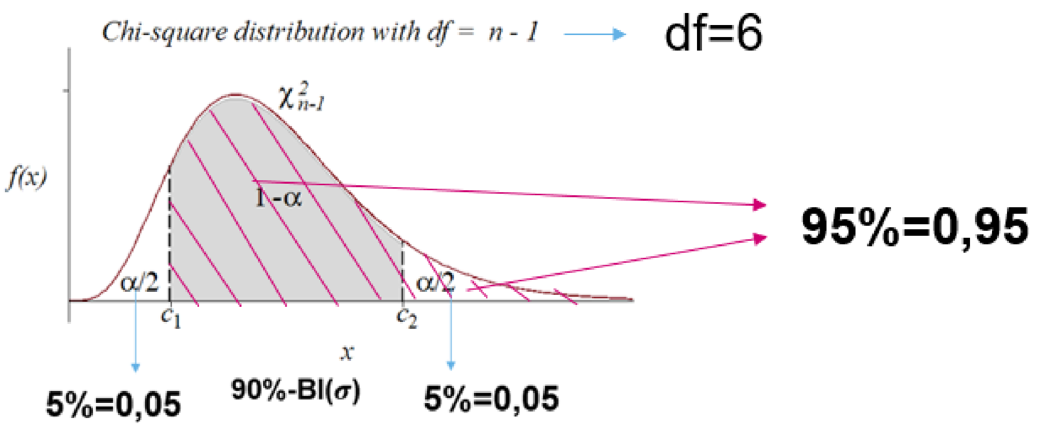
We will use the same numerical summary of the 7 returns:

$$\boldsymbol{n = 7,\ \bar{x} = 8.0} \text{ and } \boldsymbol{s \approx 5.84}$$

<u>Model</u> (statistical assumptions): The returns on investment, $X_1, \ldots, X_7$ are independent and **normally distributed** with **unknown μ** and ***unknown*** **σ^2.**

- **90%-CI(σ)** = $\left(\sqrt{\frac{(n-1)s^2}{c_2}}, \sqrt{\frac{(n-1)s^2}{c_1}}\right) \approx (\mathbf{4.0, 11.2})$

  since $P(\chi^2_6 \leq c_1) = 0.05$, so $c_1 = 1.64$
  and $P(\chi^2_6 \geq c_2) = 0.05$, so $c_2 = 12.59$



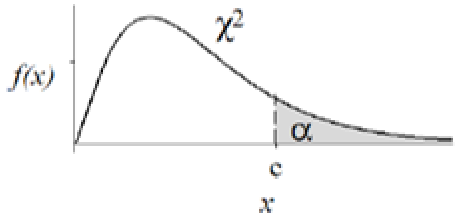
Critical values from the Chi-Square distribution table

Tab-3

Table Chi-square distribution

In the table you will find critical values c for the upper-tailed probabilities

$$P(\chi^2 \geq c) = \alpha$$



df = number of degrees of freedom

df	α = 0.995	0.990	0.975	0.95	0.90	0.75	0.25	0.10	0.05	0.025	0.010	0.005
1	0.00	0.00	0.00	0.00	0.02	0.10	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	10.22	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	5.90	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	13.70	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	29.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	30.43	35.56	38.89	41.92	45.64	48.29

$$90\%\text{-CI}(\sigma) = \left(\sqrt{\frac{(n-1)s^2}{c_2}}, \sqrt{\frac{(n-1)s^2}{c_1}}\right) \approx (4.0, 11.2)$$

"At a level of confidence 90%, the real standard deviation of the yearly returns on investment are between 4.0 and 11.2%".

Questions ?

Thank you!