1. INVENT YOURSELF: GOOD GUESSES

- Team Switzerland
- Reporter: Luca Nashabeh

1. INVENT YOURSELF: GOOD GUESSES

In 1906, Francis Galton observed a contest where 800 farmers guessed an animal's weight. To his surprise, the median of the guesses was within 0.8% of the true measured weight. What is the chance of obtaining such a good match by coincidence? Select an interesting and important parameter, measure it directly, and give a group of human observers the task to guess the value of the parameter. Discuss the results of your experiments.

INTERPRETATION

Propose an experiment similar to the one in the task. Conduct the proposed experiments and analyze its results by stating own hypotheses and testing these statistically.

STATISTICS: BASICS (PART 1)

Set of data points or sample data x_i , i = 1, 2, 3, …, n.

Empirical measures of central tendency

mean $\bar{x} = \frac{\text{sum of all data points}}{\text{number of data points}} = \frac{1}{n}\sum_{i=1}^{n} x_i$

median $\text{med}(x_i) = \begin{cases} x_{\frac{n+1}{2}} & \text{for n odd} \\ \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \text{for n even} \end{cases}$ of ordered set of data points

mode mode = most frequent data point

STATISTICS: BASICS (PART 2)

Set of data points or sample data x_i , i = 1, 2, 3, …, n.

Empirical measures of spread

range $x_{max} - x_{min}$

sample standard deviation $s = \sqrt{\frac{\text{sum of squared deviations from mean}}{\text{number of data values}-1}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

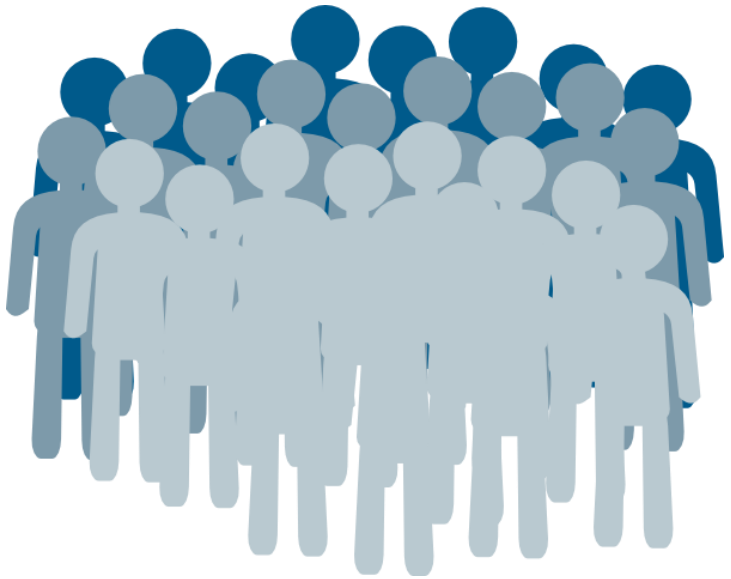
800 FARMERS – EXPERIMENT

- Ox weighed **543.4 kg**
- Guessed values mean =**542.9 kg** off by **0.1%**

  median ∈ [**539.1 kg , 547.7 kg**] off by **0.8%**

“good” match: median and mean relatively close to true value

800 FARMERS – ANALYSIS

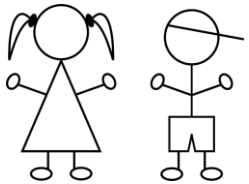


- Reasons for a good match:
  - large group → crowd wisdom
  - experienced participants → not representable for normal crowd

- Chance of obtaining a good match by coincidence:
  - very large crowd → not a coincidence
  - coincidentally choosing experienced or lucky participants → small chance

MY INVESTIGATION: TIME PERCEPTION

- **Interesting parameter** How do people perceive time?
- **Important parameter**



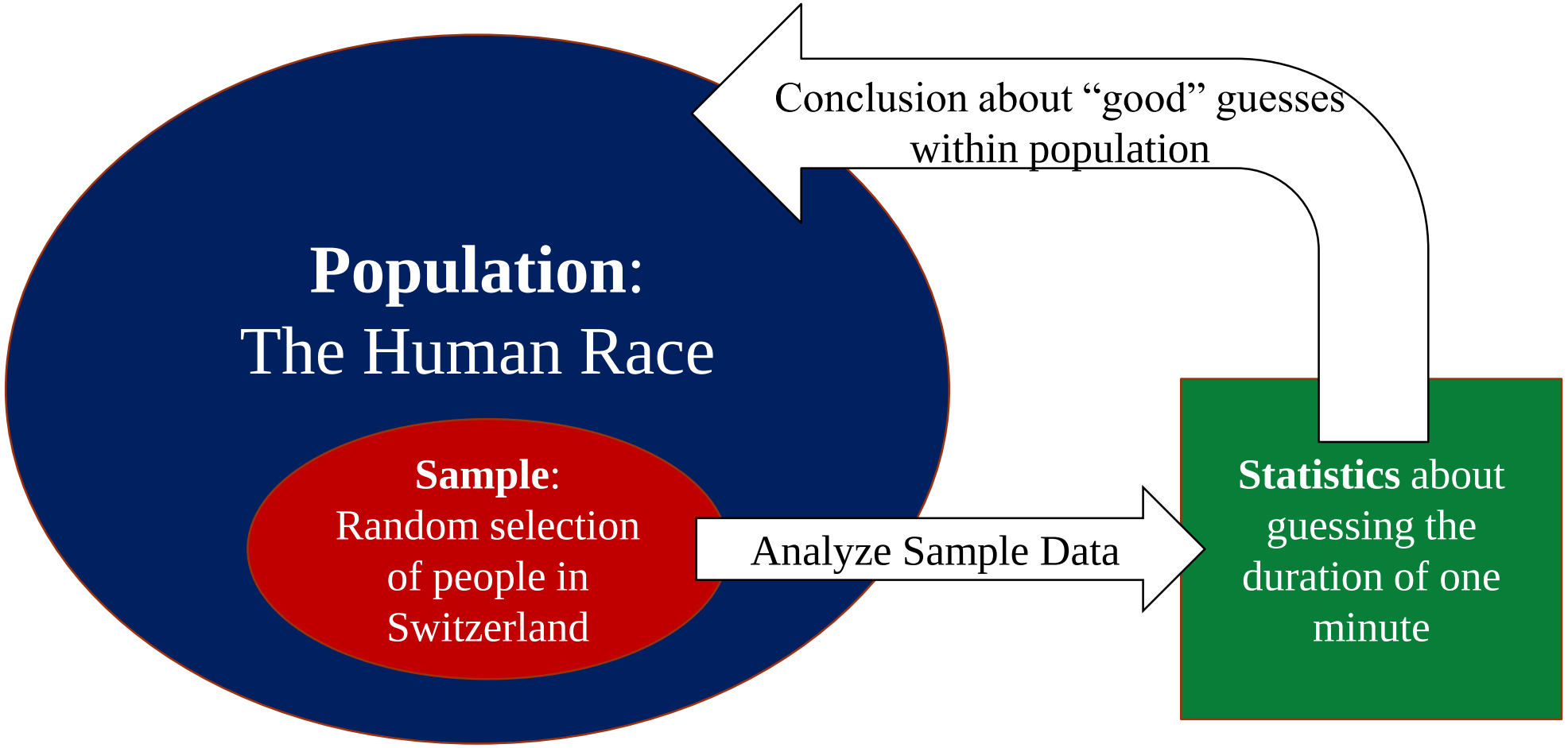
around age 5:

perception: time passes slowly
effect: overestimate elapsed time

getting older:

perception: time speeds up
effect: underestimate elapsed time

MY INVESTIGATION: GOAL



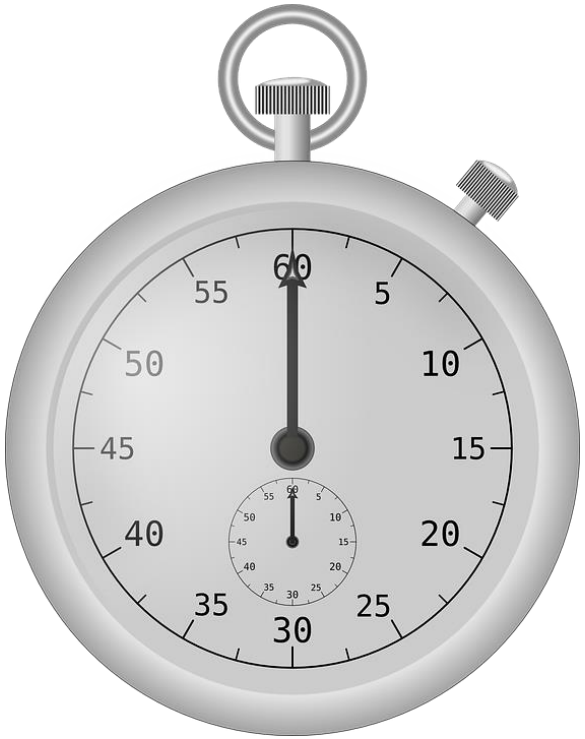
SURVEY CONDUCTION

Same survey condition for all participants

- Anonymous
- At school or at work
- Time was started and stopped on the participants cue
- Counting not allowed

data points (= survey data collected)

- Age
- Gender
- Time guessed



Guessing for one minute

HYPOTHESES

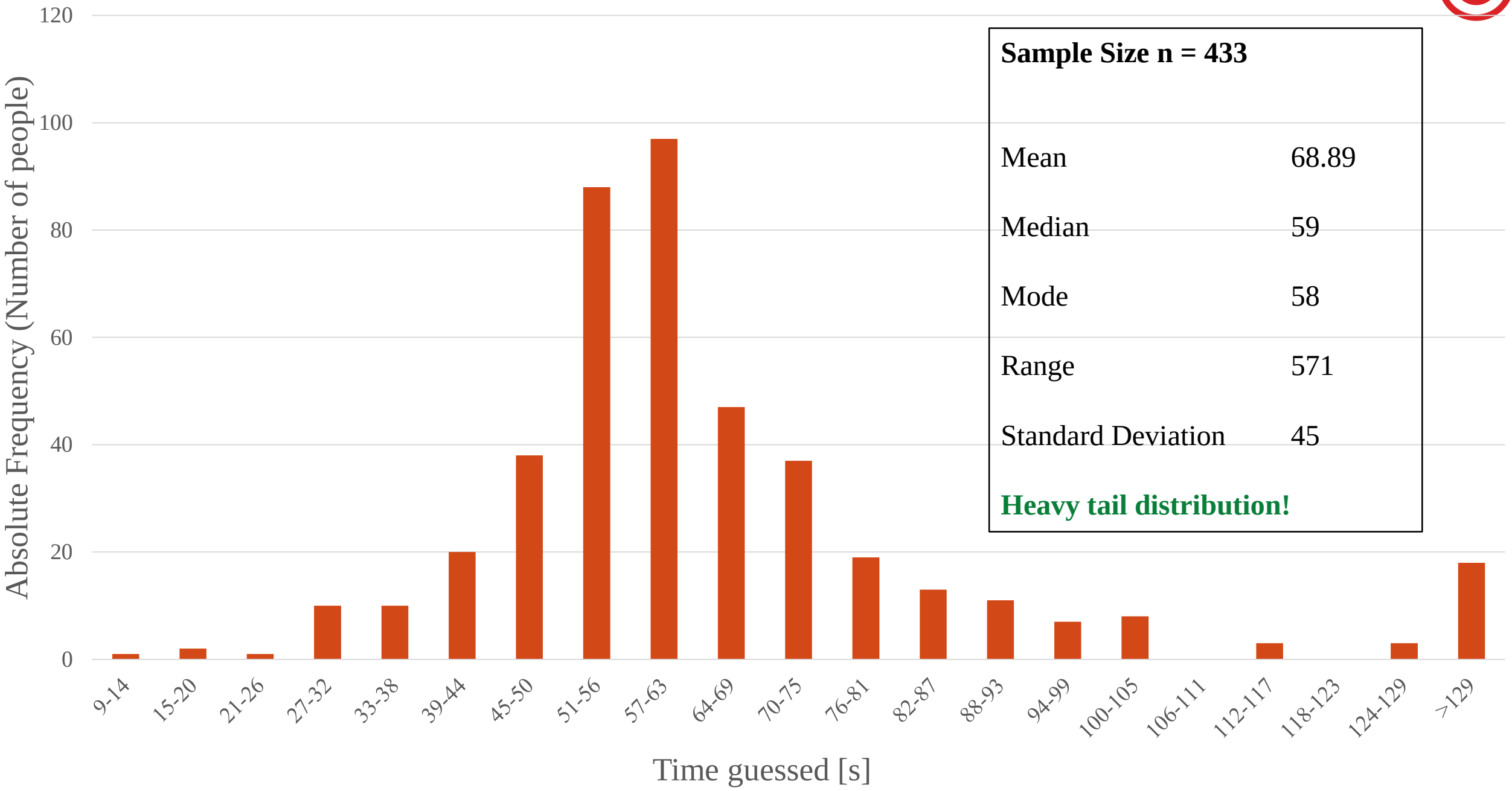
Hypothesis

1. **Age versus time guessed**
   age ⇡ $\Longrightarrow$ guess ⇡

2. **Gender and deviation**
   one gender guesses less accurate than the other

3. **Age groups and deviation**
   preschoolers $\Longrightarrow$ “bad” guess
   school age $\Longrightarrow$ “good” guess
   adults $\Longrightarrow$ “mediocre” guess

Univariate Data: Frequency of **Time Guessed** in Seconds

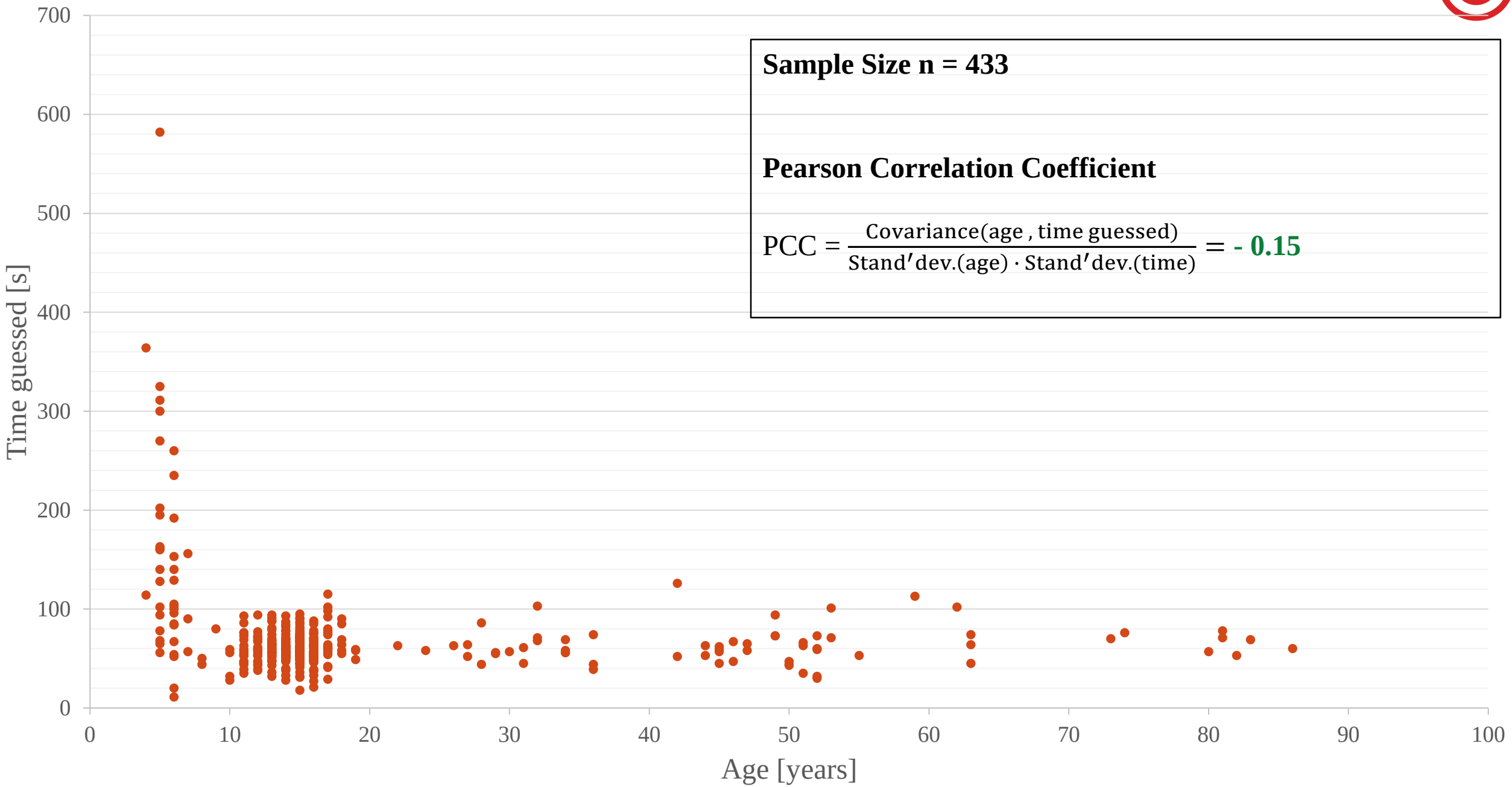


Sample Size n = 433

Mean	68.89
Median	59
Mode	58
Range	571
Standard Deviation	45

Heavy tail distribution!

Bivariate Data: **Time guessed versus Age**



Sample Size n = 433

Pearson Correlation Coefficient

$$\text{PCC} = \frac{\text{Covariance(age , time guessed)}}{\text{Stand'dev.(age)} \cdot \text{Stand'dev.(time)}} = \mathbf{-\ 0.15}$$

HYPOTHESES 1: RESULT

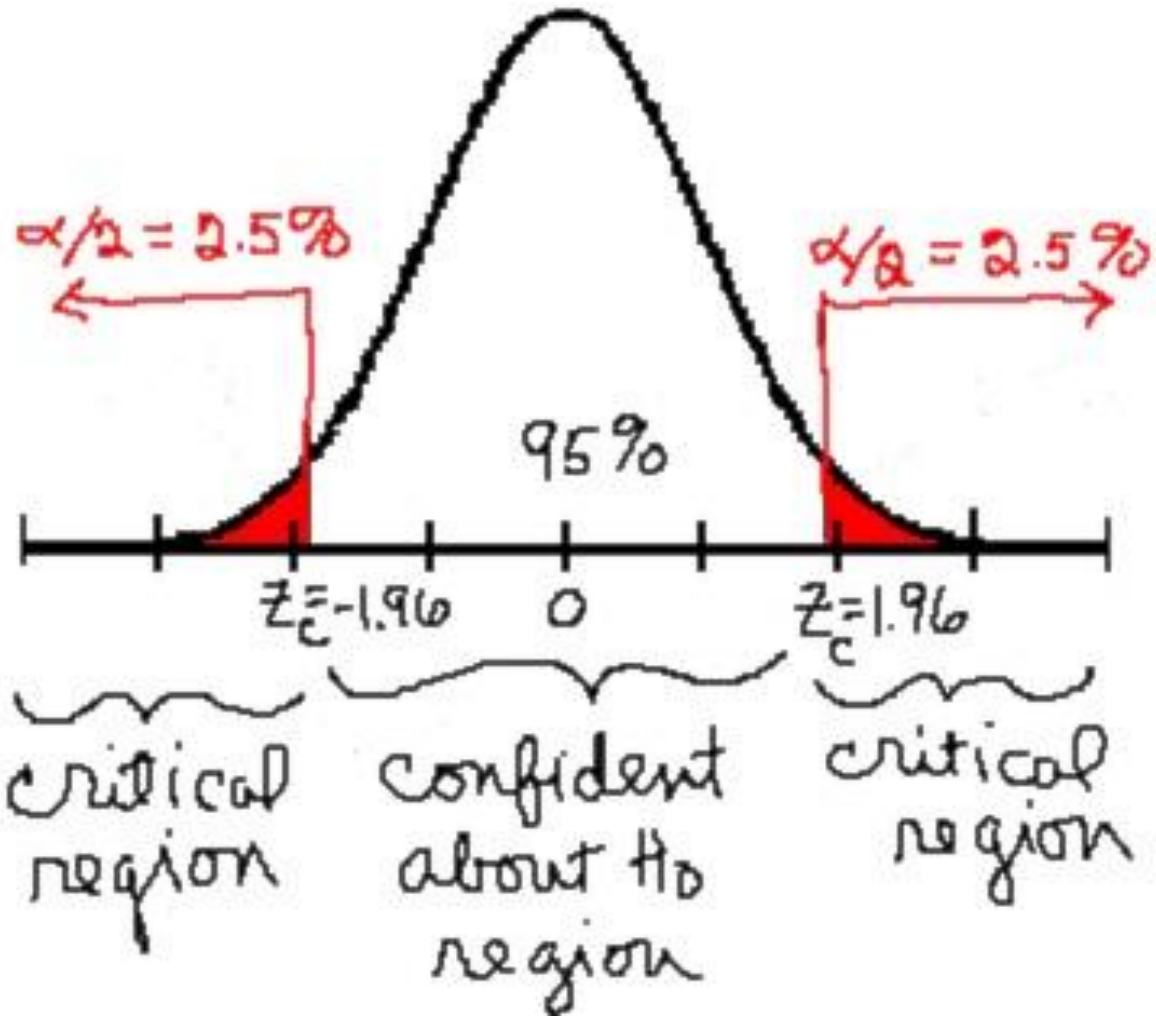
Hypothesis	**Statistics**
1. **Age versus time guessed**	**Pearson Correlation**
age $\Uparrow$ $\Longrightarrow$ guess $\Uparrow$	

NOT CONFIRMED!

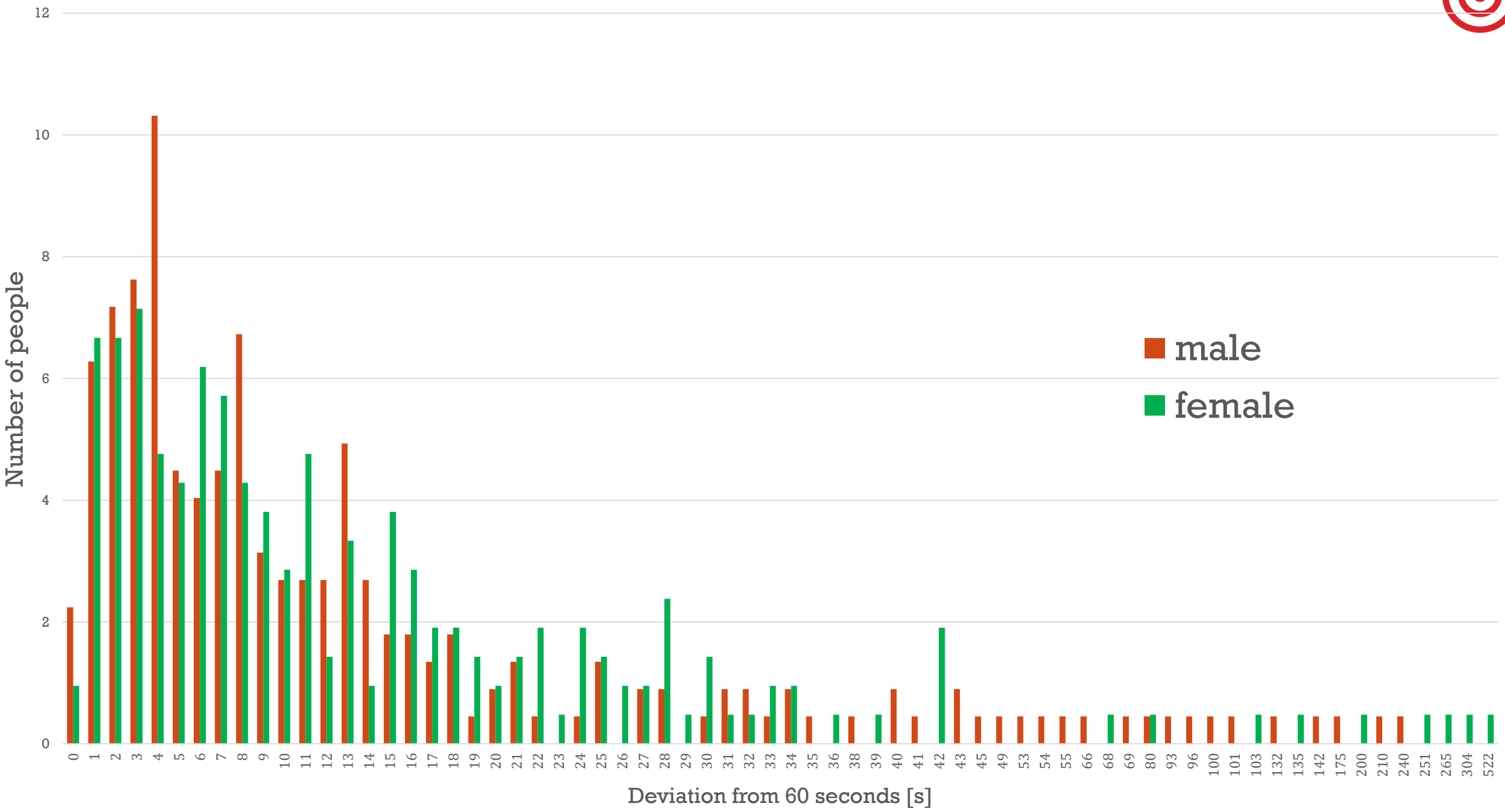
MANN-WHITNEY U TEST

Description

- Null hypothesis
- Comparing two independent samples



Deviations per gender



HYPOTHESIS 2: RESULT

Hypothesis about

2. Gender and deviation
One gender guesses less accurate than the other.

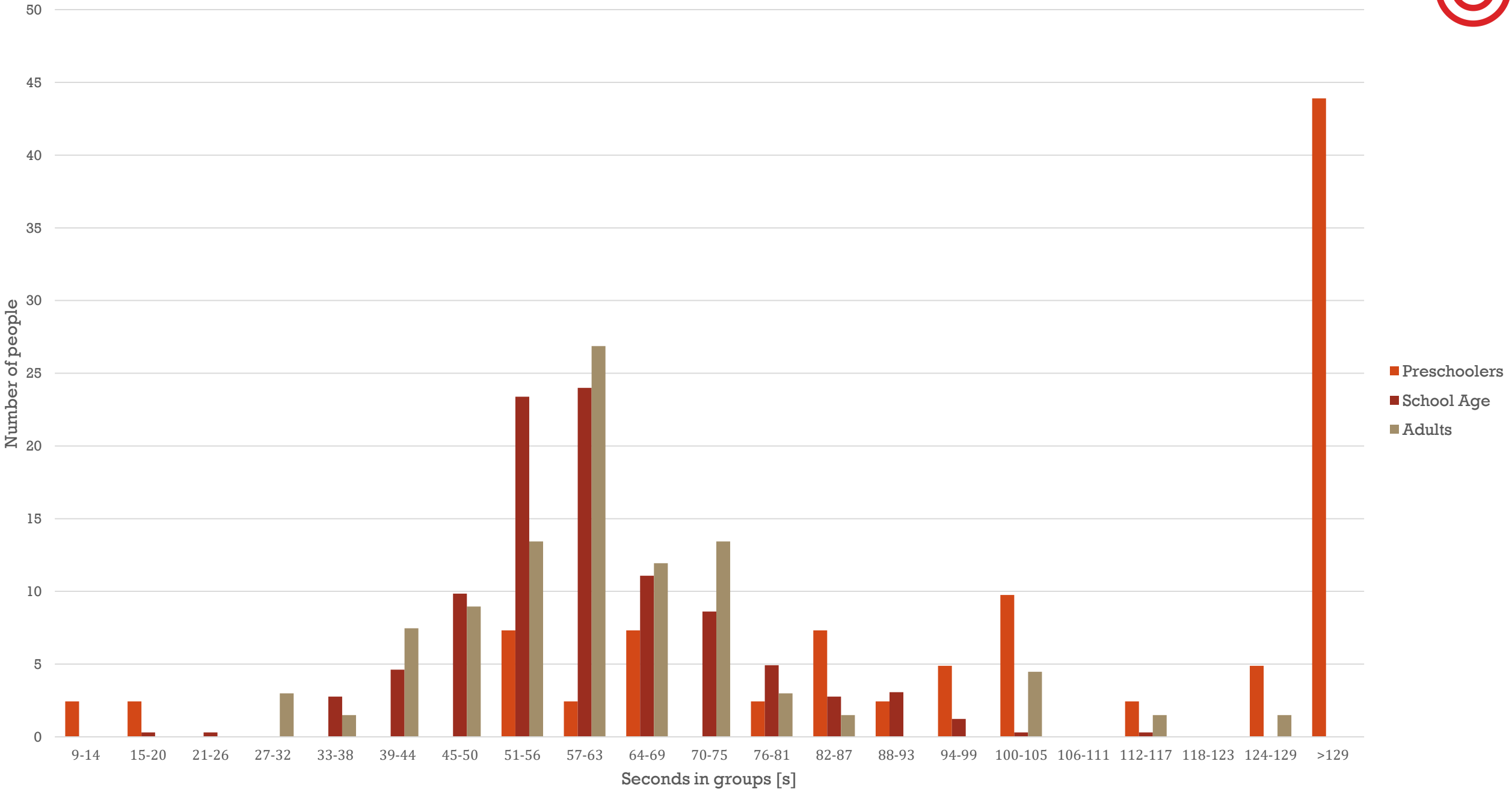
Mann-Whitney U Test:

Statistic at confidence level 95%

One gender guesses less accurate than the other

Rejected

Number of guesses per age per time



HYPOTHESES 3: RESULTS

3.Age groups and deviation

Differences between age groups:

Mann-Whitney U Test:	Statistic at confidence level 95%
Preschool guesses worse than rest	**Not Rejected**
School age guesses better than adults	**Rejected**

CONCLUSION (PART 1)

Wisdom of crowds

- measures of tendency got closer to the true value
- measures of spread decreased

➔ **theory confirmed for farmers and time is relative**

800 farmers

- “good guesses” can be explained by group selection (farmers knowledge and experience)

➔ **survey representative for farmer population only**

CONCLUSION (PART 2)

- **Correlation between age and time**
  - Hypothesis rejected
- **Difference between genders**
  - Hypothesis rejected
- **Difference between age groups**
  - Preschoolers significantly worth
  - Other age groups consistent

CONCLUSION (PART 3)

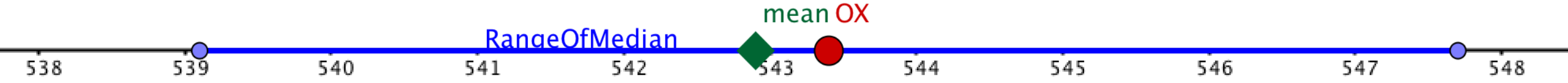
Statistic	Farmers	My Experiment
Number of participants	800	433
Percentage error of mean	0.1%	14.8%
Percentage error of median	0.8%	1.7%

SOURCES

- https://blogs.scientificamerican.com/mind-guest-blog/why-does-time-fly-as-we-get-older/
- http://www.chegg.com/homework-help/questions-and-answers/1906-800-people-attending-fair-outside-plymouth-england-took-part-competition-guess-weight-q15897955
- https://www.youtube.com/watch?v=aIx2N-viNwY
- http://journals.sagepub.com/doi/abs/10.1177/0961463X13478052
- https://www.youtube.com/watch?v=wwvkdqDoqdM
- http://www.socscistatistics.com/tests/mannwhitney/Default.aspx
- DMK Formulae Tables and Concepts, orell fuessli, ISBN 978-3-280-04084-3
- Mathematics HL Core, third edition, Haese Mathematics, ISBN 978-1-921972-11-9
- Mathematics HL Option: Statistics and Probability, Haese Mathematics, ISBN 978-1-921972-31-7
- ISO Standards, https://www.iso.org/obp/ui/?_escaped_fragment_=iso:std:44396:en
- Wilcoxon Rank Sum Test: https://www.graphpad.com, www.statisticsolutions.com, http://www.socscistatistics.com/tests/mannwhitney/Default3.aspx
- Time: Derek A. Muller, www.veritasium.com

THANK YOU FOR LISTENING

FARMERS


mean OX
RangeOfMedian
538
539
540
541
542
543
544
545
546
547
548

DATA: SYMBOLS AND SAMPLE SIZE

- Univariate data set

n	=	count of participants
$3n$	=	count of data points
t_n	=	time guessed of participant n [seconds]
x_n	=	$\lvert t_n - 60 \rvert$ = deviation of 60 seconds (→ "good" guess!)

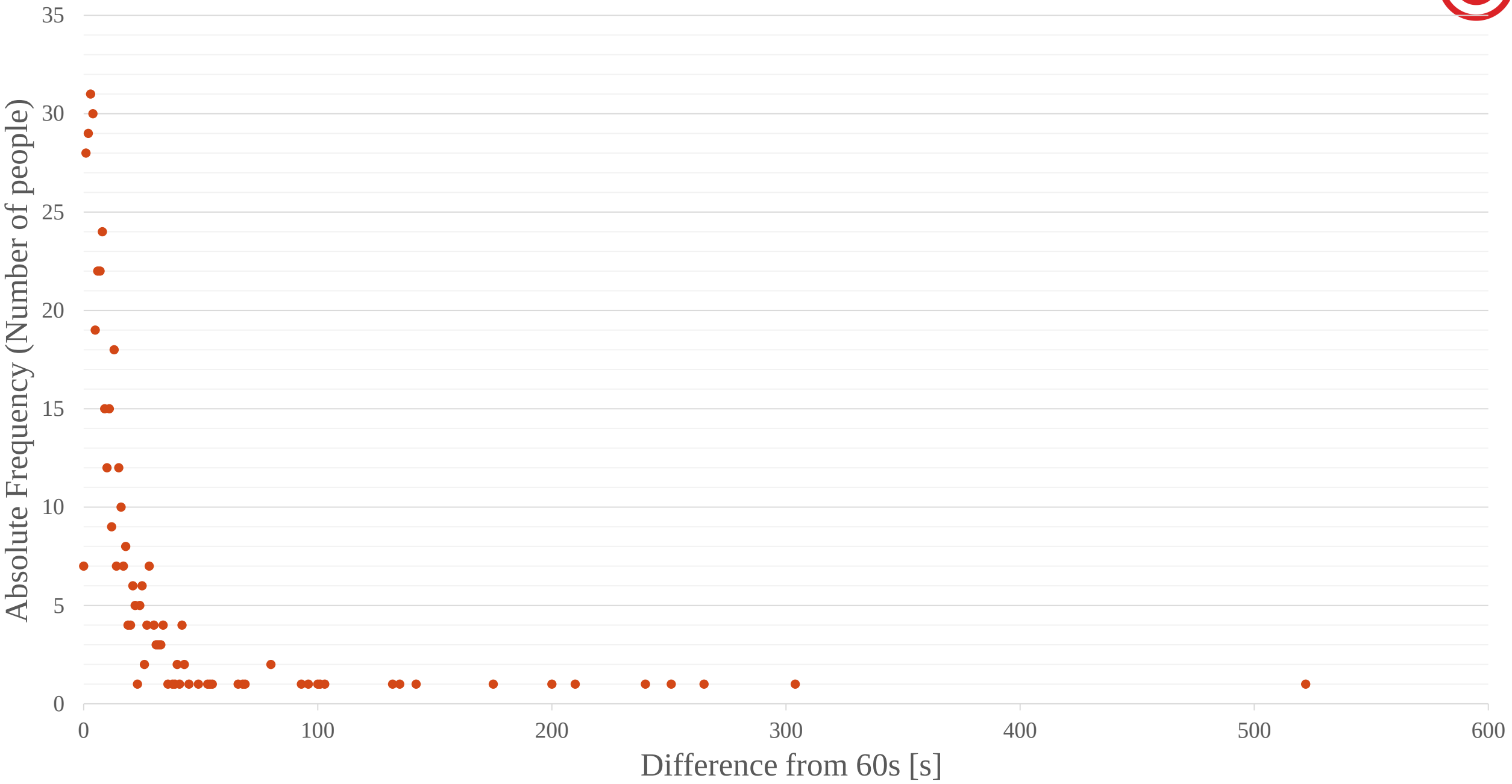
- Sample Size

n = 433

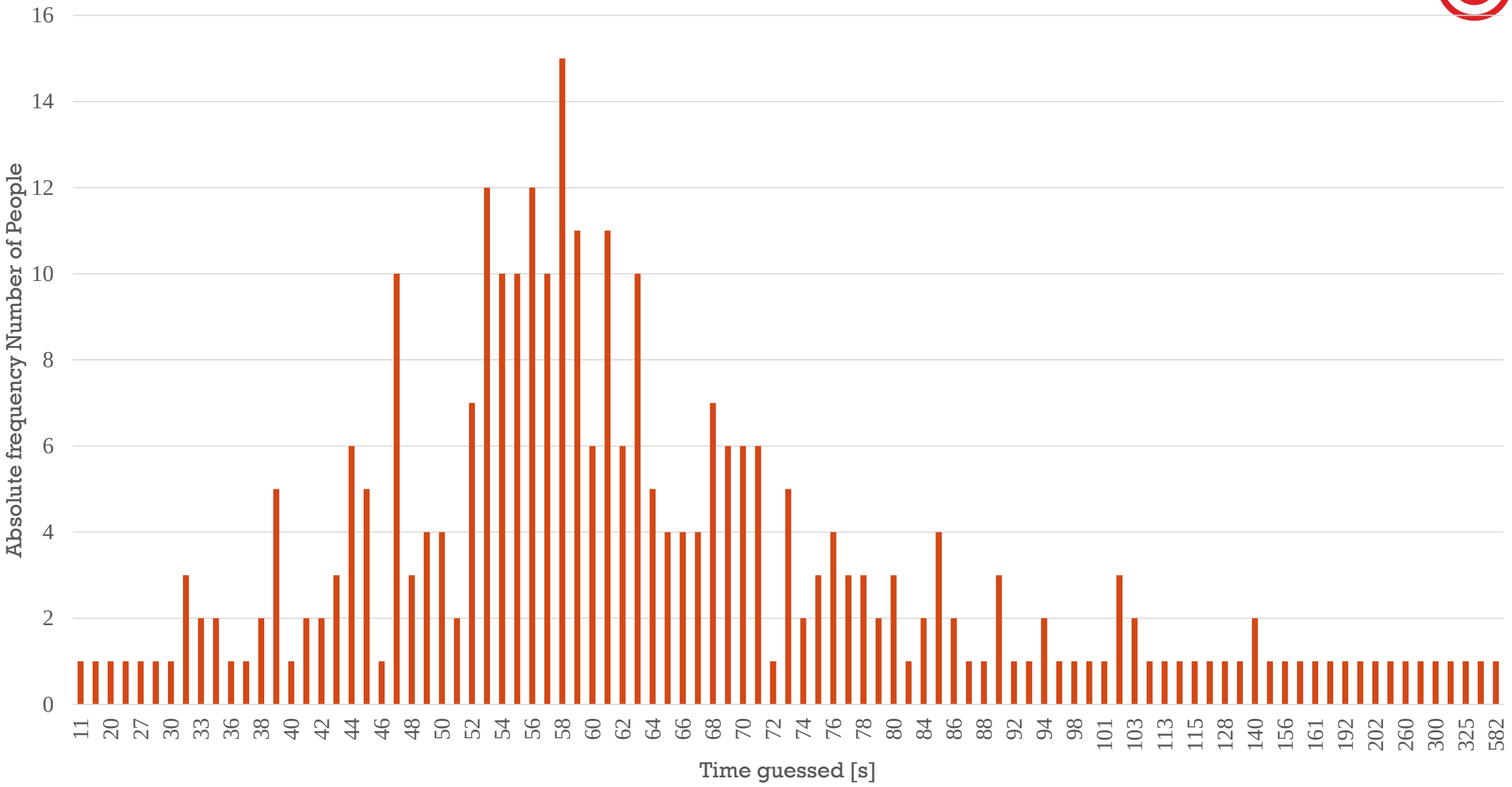
gender n_{female} = 210 n_{male} = 223

age group $n_{preschool}$ = 41 (age 4 – 7) n_{school} = 325 (age 8 – 18) n_{adult} = 67 (age 19 – 86)

Frequency of Differences



Time guesses (ungrouped data)



DATA SET: TIME GUESSED BY GENDER

Symbol	Description	All	Female	Male
n	number of data points	433	210	223
t_{min}	minimum data point	11	18	11
t_{max}	maximum data point	582	582	300
r	range	571	564	289
$\bar{t}$	mean	68.89	69.50	68
t_{med}	median	59	59.5	59
t_{mode}	mode	56	58	56
s	standard deviation	45	54	34
$v = s / \bar{t}$	variation	0.65	0.78	0.50

coefficient of variation $v = \frac{s}{\bar{x}}$ (relative standard deviation)

DATA SET: TIME GUESSED BY AGE

Symbol	Description	Preschool 4 – 7	School age 8 – 18	Adult 19 – 86
n	number of data points	41	325	67
t_{min}	minimum data point	11	18	30
t_{max}	maximum data point	582	115	126
r	range	571	97	96
$\bar{t}$	mean	150.02	59.91	62.85
t_{med}	median	114	58	60
t_{mode}	mode	140	56	58
s	standard deviation	110	14	18.02
$v = s / \bar{t}$	variation	0.73	0.24	0.29

coefficient of variation $v = \frac{s}{\bar{x}}$ (relative standard deviation)

DATA SET: DEVIATION OF 60S BY GENDER

Symbol	Description	All	Female	Male
n	number of data points	433	210	223
x_{min}	minimum data point	0	0	0
x_{max}	maximum data point	522	522	240
r	range	522	522	240
$\bar{x}$	mean	18.94	20.64	17.35
x_{med}	median	8	9	8
x_{mode}	mode	4	3	4
s	standard deviation	41	50	30
$v = s / \bar{x}$	variation	2.20	2.46	1.77

coefficient of variation $v = \frac{s}{\bar{x}}$ (relative standard deviation)

DATA SET: DEVIATION OF 60S BY AGE

Symbol	Description	Preschool 4 – 7	School age 8 – 18	Adult 19 – 86
n	number of data points	41	325	67
x_{min}	minimum data point	3	0	0
x_{max}	maximum data point	522	55	66
r	range	519	55	66
$\bar{x}$	mean	95.39	10.62	12.55
x_{med}	median	54	8	9
x_{mode}	mode	8	4	3
s	standard deviation	105	9	13
$v = s / \bar{x}$	variation	1.10	0.88	1.05

coefficient of variation $v = \frac{s}{\bar{x}}$ (relative standard deviation)

STATISTICS: ANALYSIS

Statistic	quantity calculated from sample data ➔ draw conclusions about a population
Hypothesis Test	calculating a statistic ➔ accept or reject a hypothesis
Mann-Whitney U Test	specific hypothesis test ➔ compare two independent samples

LAW OF LARGE NUMBERS

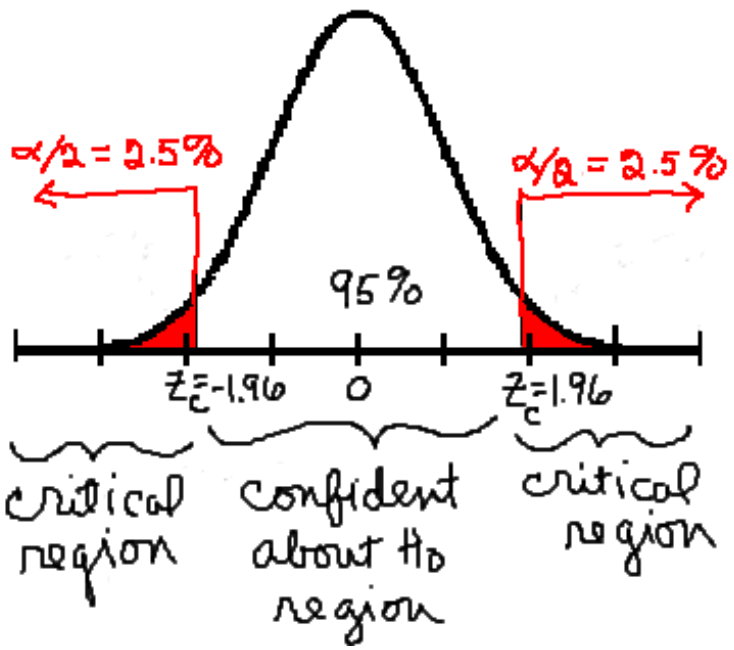
The relative frequency of an data point converges to the probability of the data point as the number of data points increases.

Example: Rolling a fair die n times

Law: relative frequency of rolling a six $\longrightarrow \frac{1}{6}$ (= probability of rolling a six) for n $\longrightarrow \infty$

Rolling a die once:	Observing a six is not very likely.
Rolling a die 100 times:	Observing a six is getting close to 1/6 of 100
Rolling a die 1’000’00 times	Observing a six will be really close to 1/6 of 1’000’000

HYPOTHESIS TESTING



- Specify the null hypothesis H_0 and the alternative hypothesis H_1.
- Calculate a test statistic Z from the sample data.
  (This test statistic has to be a random variable with a known distribution, e.g. standard normal distribution $Z \sim N(0,1)$)
- Significance level, e.g. $\alpha = 5\%$:
  Assuming that the null-hypothesis is true, it is likely that the test statistic lies within the 95% confidence interval.
- Result
  If the likely event, that the test statistics lies within the confidence interval, occurs, the null hypothesis is not rejected.
  If the unlikely event, that the test statistics lies outside the confidence interval, occurs, the null hypothesis is rejected.
- Error
  The probability of having made the “wrong” decision is 5%.

MANN-WHITNEY U TEST

Rank sum test step by step

- Order and rank data from both samples → list with $n_1 + n_2$ ranked data points
- Add ranks per sample: R_1 = sum of ranks for sample 1, R_2 = sum of ranks for sample 2
- H_0 : There is no difference between the distribution of two samples (ranks evenly spread)
  H_1 : There is a difference between the distribution of two samples
- Test statistic
  $U = \frac{n_1(n_1+1)}{2} + n_1 \cdot n_2 - R_1$ = maximum possible sum of ranks – observed sum of ranks
  Under H_0 : $U \sim N(\mu, \sigma)$
  Check: U(sample 1) + U(sample 2) = $n_1 \cdot n_2$
- Standardized test statistic
  $Z = \frac{U - \mu}{\sigma}$ with estimates for $\mu = \frac{n_1 \cdot n_2}{2}$ and for $\sigma = \sqrt{\frac{n_1 \cdot n_2(n_1 + n_2 + 1)}{12}}$ = standardized U
  Under H_0 : $Z \sim N(0,1)$

MANN-WHITNEY U TEST

Rank sum test step by step → decision

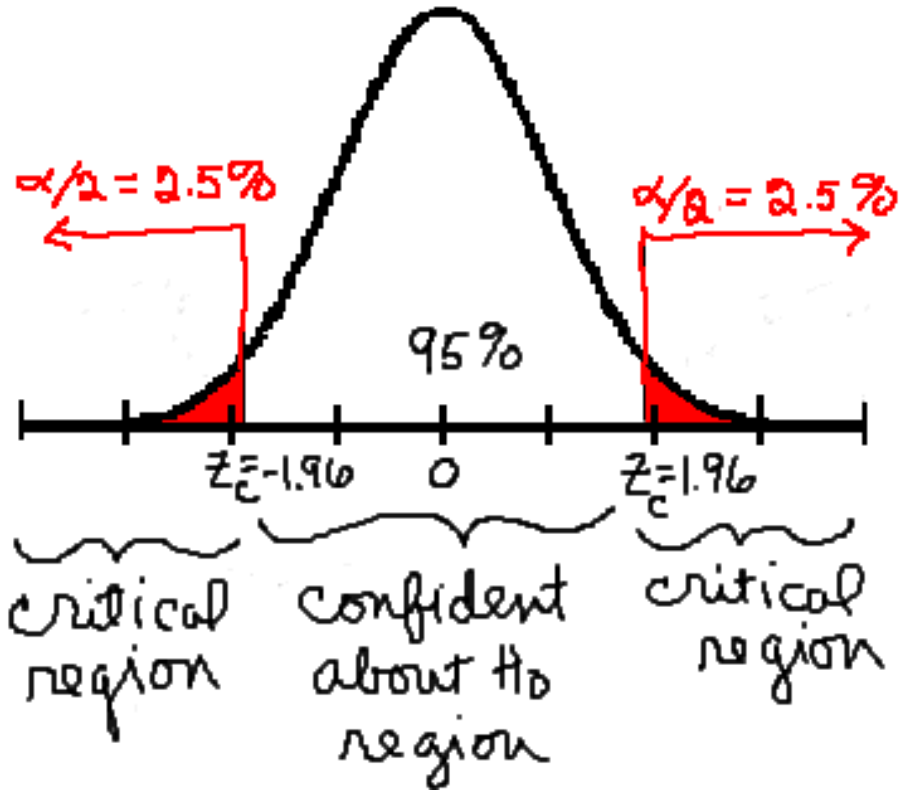
- Under H_0 : Z is a standard normal random variable

  $Z \sim N(0,1)$

- Decision for two-sided test at a significance level of 5%

  $Z \notin [\,-1.96\,,\,1.96\,] \quad \Longrightarrow \quad$ reject H_0

  $Z \in [\,-1.96\,,\,1.96\,] \quad \Longrightarrow \quad$ do not reject H_0



NORMAL DISTRIBUTION

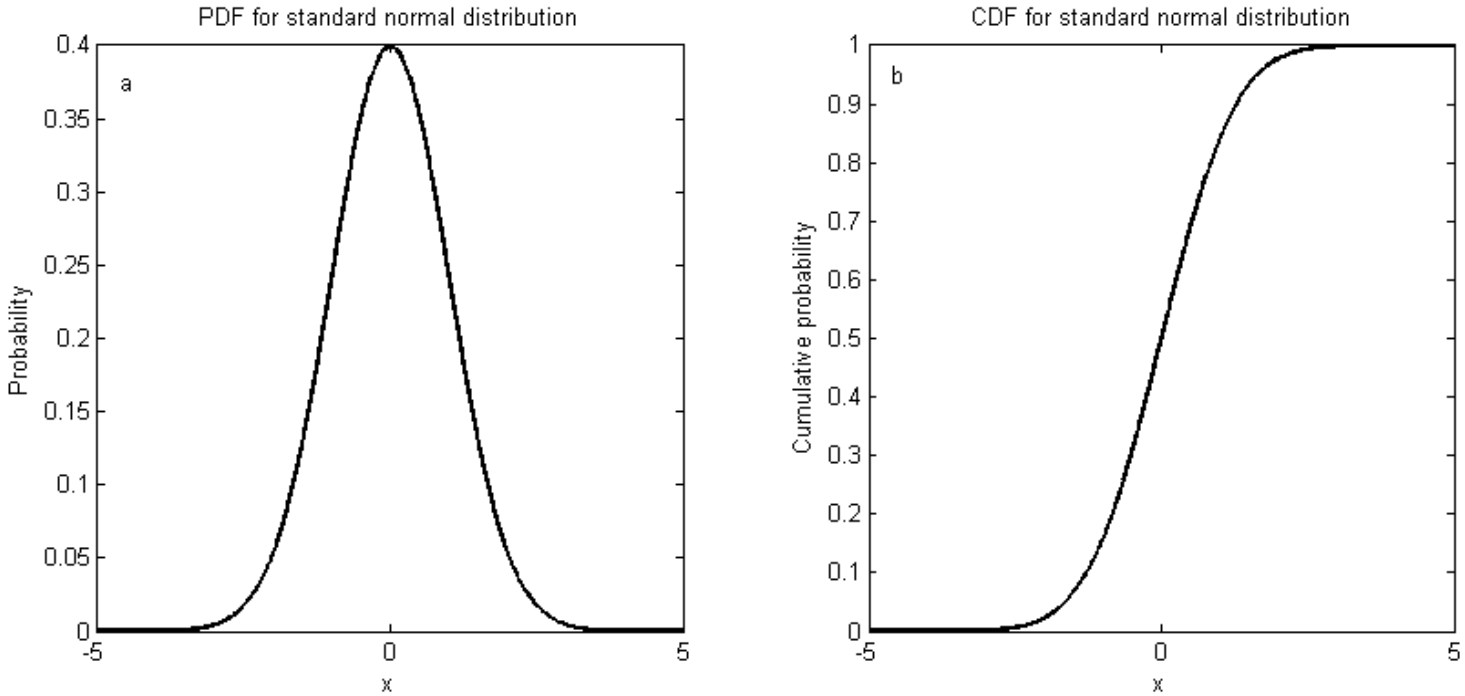
A **continuous** random variable X takes all values x in given interval of numbers.
X is **standard normally distributed** if its probability density function (probability distribution) is

PDF = (Probability of X = x)

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

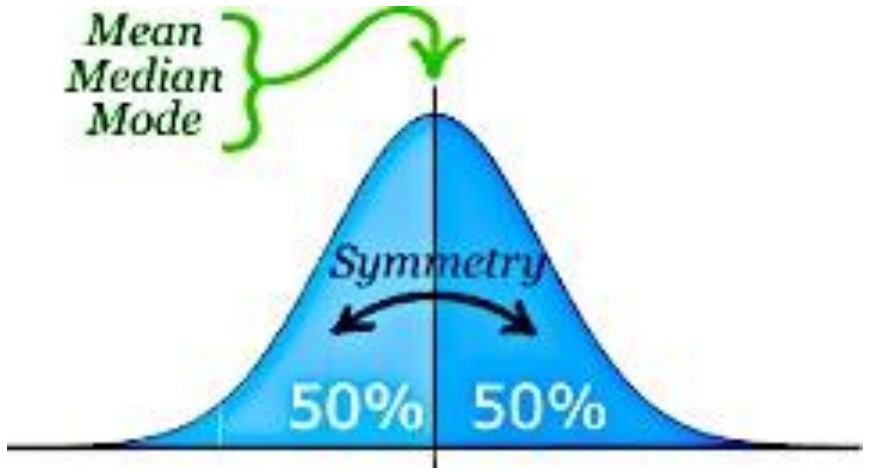
X ~ N (0 , 1)

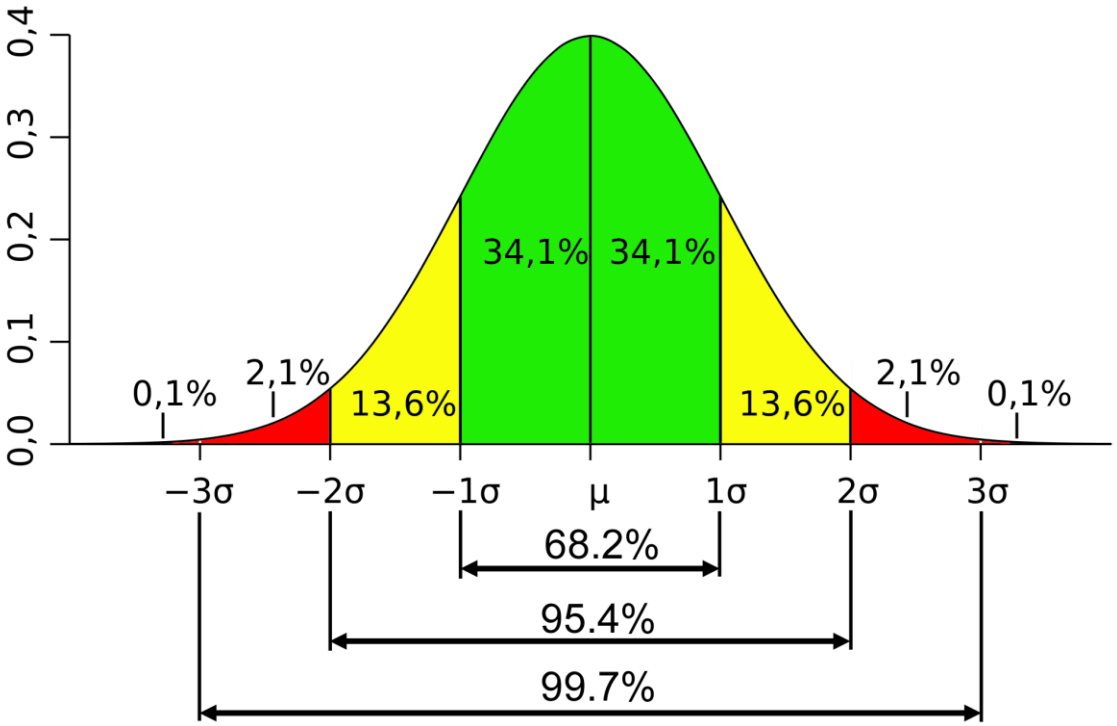
Example:
Errors in physical measurements



PROPERTIES OF NORMAL DISTRIBUTION

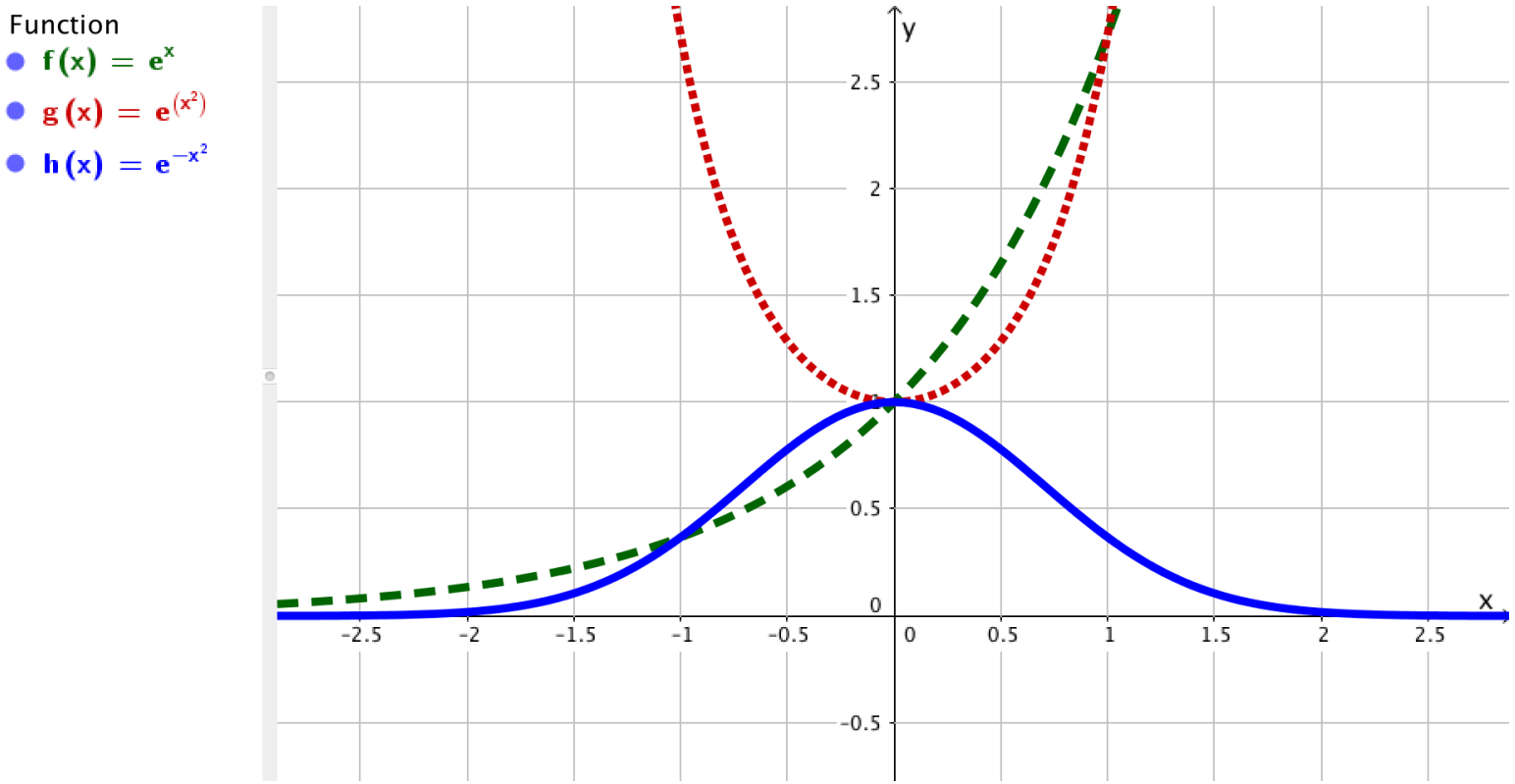
$X \sim N(0, 1)$





DERIVING THE FORMULA FOR THE STANDARD NORMAL DISTRIBUTION

Carl Friedrich Gauss came up with the idea of how to shape a normal curve (original idea: approximation of binomial distribution)



adjustments have to be made so that it is a probability distribution function, i.e. the area = 1

$$\text{PDF} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

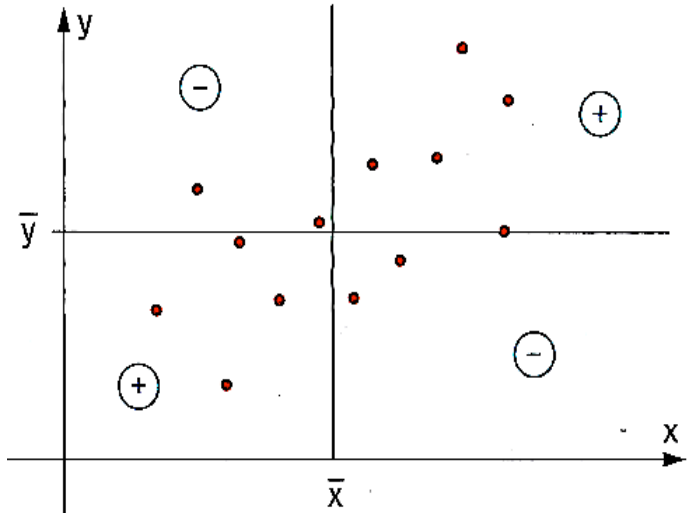
STATISTICS VERSUS PROBABILITY THEORY

Concept:
Use empirical data of a sample to estimate parameters of true probability distribution of the population.

	Statistics	**Probability Theory**
Data	data point, Counting	Model, Assumption
Measure of Center	Arithmetic Mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\bar{x} = \sum_{i=1}^{n} x_i \cdot n_i$ (n_i = relative frequency of x_i)	Expected Value $\mu = E[X] = \sum_{i=1}^{n} x_i \cdot p_i$
Measure of Spread	Variance $_{\text{empirical}}$ $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ $s^2 = \frac{n}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot n_i$	Variance $_{\text{theoretical}}$ $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \cdot p_i$
	Standard Deviation $_{\text{empirical}}$ $\mathbf{s} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$	Standard Deviation $_{\text{theoretical}}$ $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \cdot p_i}$

PEARSON CORRELATION COEFFICIENT

$$\text{PCC} = \frac{\text{Covariance(age , time guessed)}}{\text{Stand'dev.(age)} \cdot \text{Stand'dev.(time)}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$



PCC Values