1. INVENT YOURSELF: GOOD GUESSES

- Team Switzerland
- Reporter: Luca Nashabeh
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.
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 $x_i$, $i = 1, 2, 3, \ldots, n$.

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**  
$$\text{med}(x_i) = \begin{cases} 
\frac{x_{n+1}}{2} & \text{for } n \text{ odd} \\
\frac{1}{2} \left( x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right) & \text{for } n \text{ even}
\end{cases}$$

of ordered set of data points

**mode**  
mode = most frequent data point
STATISTICS: BASICS (PART 2)

Set of data points or sample data $x_i$, $i = 1, 2, 3, \ldots, n$.

Empirical measures of spread

**range**

\[ x_{\text{max}} - x_{\text{min}} \]

**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

Population:
The Human Race

Sample:
Random selection of people in Switzerland

Conclusion about “good” guesses within population

Statistics about guessing the duration of one minute

Analyze Sample Data
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 ↑  ⇒  guess ↑

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

3. **Age groups and deviation**
   preschoolers  ⇒  “bad” guess
   school age    ⇒  “good” guess
   adults        ⇒  “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**

\[
PCC = \frac{\text{Covariance}(\text{age}, \text{time guessed})}{\text{Stand'\dev.(age)} \cdot \text{Stand'\dev.(time)}} = -0.15
\]
HYPOTHESES 1: RESULT

Hypothesis

1. Age versus time guessed

Statistics

Pearson Correlation

\[ \text{age} \uparrow \implies \text{guess} \uparrow \]

NOT CONFIRMED!
MANN-WHITNEY U TEST

Description

- Null hypothesis
- Comparing two independent samples
Deviations per gender

Number of people

Deviation from 60 seconds [s]

- male
- female
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

Seconds in groups [s]

Number of people

Preschoolers
School Age
Adults
HYPOTHESES 3: RESULTS

3. Age groups and deviation

Differences between age groups:

Mann-Whitney U Test:

- Preschool guesses worse than rest: Not Rejected
- School age guesses better than adults: Rejected

Statistic at confidence level 95%
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)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Farmers</th>
<th>My Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of participants</td>
<td>800</td>
<td>433</td>
</tr>
<tr>
<td>Percentage error of mean</td>
<td>0.1%</td>
<td>14.8%</td>
</tr>
<tr>
<td>Percentage error of median</td>
<td>0.8%</td>
<td>1.7%</td>
</tr>
</tbody>
</table>
**SOURCES**

- https://www.youtube.com/watch?v=aIx2N-viNwY
- https://www.youtube.com/watch?v=wwvdqDoqdM
- Time: Derek A. Muller, www.veritasium.com
THANK YOU FOR LISTENING
FARMERS
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]
\( x_n \) = \( | t_n - 60 | \) = deviation of 60 seconds (→ “good” guess!)

• Sample Size

\( n = 433 \)

gender
\( n_{female} = 210 \quad n_{male} = 223 \)

age group
\( n_{preschool} = 41 \quad n_{school} = 325 \quad n_{adult} = 67 \)
\( \text{(age 4 – 7)} \quad \text{(age 8 – 18)} \quad \text{(age 19 – 86)} \)
Frequency of Differences

Absolute Frequency (Number of people)

Difference from 60s [s]
Time guesses (ungrouped data)
# DATA SET: TIME GUESSED BY GENDER

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

**Coefficient of Variation** $v = \frac{s}{\mu}$ (relative standard deviation)
## Data Set: Time Guessed by Age

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

**Coefficient of variation**: \( v = \frac{s}{\bar{t}} \) (relative standard deviation)
DATA SET: **DEVIATION OF 60S BY GENDER**

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

**Coefficient of variation**  \( v = \frac{s}{\bar{x}} \) (relative standard deviation)
# Data Set: Deviation of 60s by Age

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

**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 $\rightarrow \frac{1}{6}$ (= probability of rolling a six) for $n \rightarrow \infty$

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 1’000’000 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 $H_0$: $Z \sim N(0,1)$
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%

  $Z \notin [-1.96, 1.96] \implies$ reject $H_0$

  $Z \in [-1.96, 1.96] \implies$ 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

\[
\text{PDF} = \left( \text{Probability of } X = x \right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2}
\]

\( X \sim \text{N} \left( 0, 1 \right) \)

Example:
Errors in physical measurements
PROPERTIES OF NORMAL DISTRIBUTION

\[ X \sim N(0, 1) \]
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

$$PDF = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$
Concept:
Use empirical data of a sample to estimate parameters of true probability distribution of the population.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Probability Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data</strong></td>
<td>Model, Assumption</td>
</tr>
<tr>
<td>data point, Counting</td>
<td></td>
</tr>
<tr>
<td><strong>Measure of Center</strong></td>
<td>Expected Value</td>
</tr>
<tr>
<td>Arithmetic Mean</td>
<td></td>
</tr>
<tr>
<td>$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$</td>
<td>$\mu = E[X] = \sum_{i=1}^{n} x_i \cdot p_i$</td>
</tr>
<tr>
<td>$\bar{x} = \sum_{i=1}^{n} x_i \cdot n_i$</td>
<td></td>
</tr>
<tr>
<td>(n_i = relative frequency of $x_i$)</td>
<td></td>
</tr>
<tr>
<td><strong>Measure of Spread</strong></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>Variance</td>
</tr>
<tr>
<td>$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$</td>
<td>$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i$</td>
</tr>
<tr>
<td>$s^2 = \frac{n}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot n_i$</td>
<td></td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td></td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$</td>
<td>$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \cdot p_i}$</td>
</tr>
<tr>
<td><strong>empirical</strong></td>
<td><strong>theoretical</strong></td>
</tr>
</tbody>
</table>
PEARSON CORRELATION COEFFICIENT

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

PCC Values

<table>
<thead>
<tr>
<th>Perfect Positive Correlation</th>
<th>High Positive Correlation</th>
<th>Low Positive Correlation</th>
<th>No Correlation</th>
<th>Low Negative Correlation</th>
<th>High Negative Correlation</th>
<th>Perfect Negative Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9</td>
<td>0.5</td>
<td>0</td>
<td>-0.5</td>
<td>-0.9</td>
<td>-1</td>
</tr>
</tbody>
</table>