Crash Course in Statistics
ZNZ 2026
I

Christoph Luchsinger and Zofia Baranczuk

Based on Script by Daniel J. Stekhoven

About me

- Christoph Luchsinger, born 1968
- PhD in Mathematics (thesis on R_0...)
- Lecturer University of Zurich: Mathematics for Nature Science
- Managing Director of www.acad.jobs / www.math-jobs.com
- Columnist www.schweizermonat.ch/author/christophluchsinger

About you (10”)

- Each one …
  - Name
  - Research interest
  - Experience with statistics, theoretical
  - Experience with statistics, computational

How this is going to work

- Every day (with me) has the same structure

  0900-1100 Motivation and background on statistical concepts

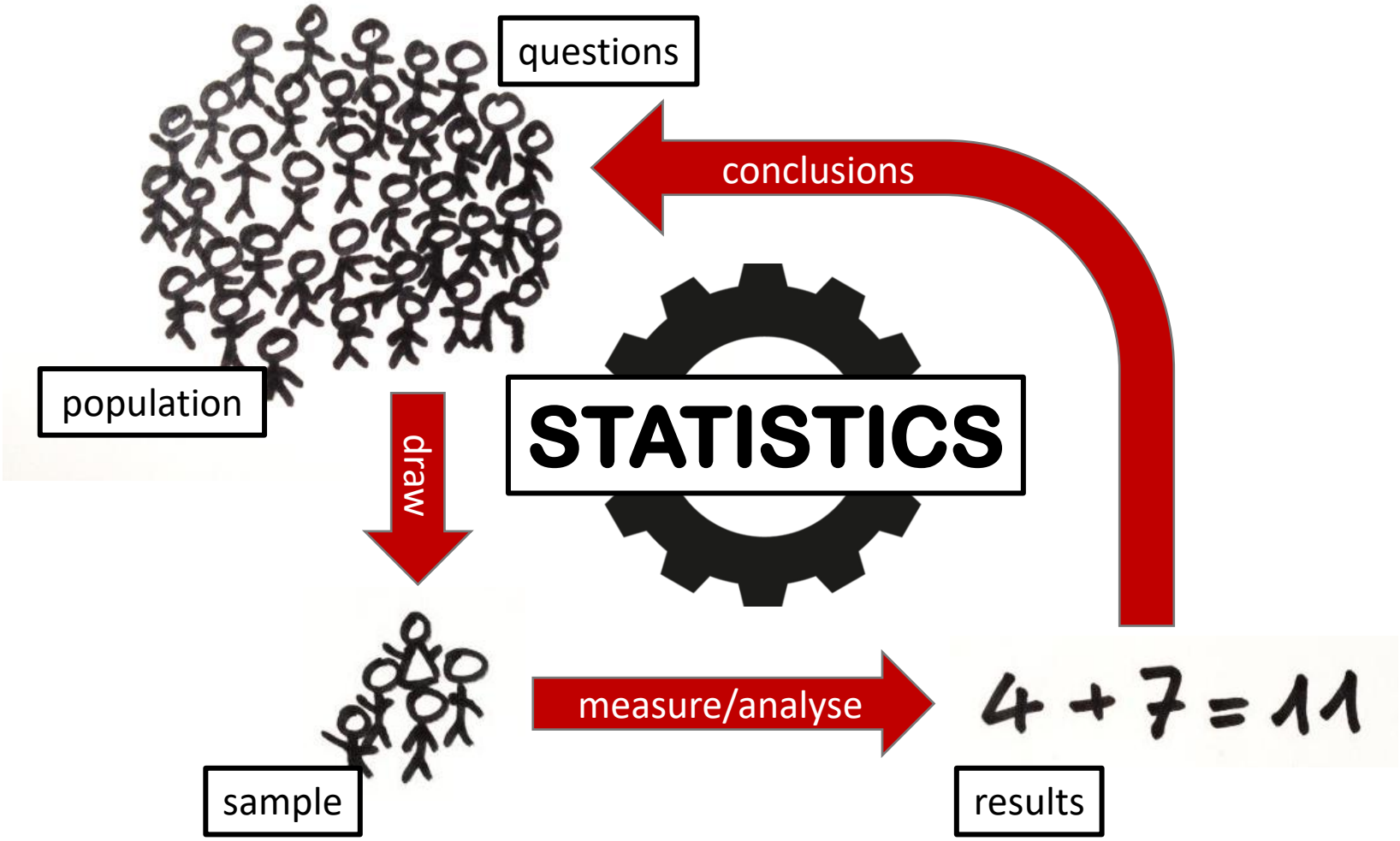
  1100-1200 Monday, Wednesday: Exercises

  1100-1200 Tuesday, Thursday, Friday: Theory until 1200 - Friday until finished

Some comments before we start

- If you miss more than 6 h (two half days) of Course, you don't get any CP's! In detail:
  - Max allowed absence for the theoretical part (Chris): One morning
  - Max allowed absence for "Applications in R" (Zofia): One morning OR 1 afternoon
- So in total, two half days absence is allowed, but not within the same section.
- You will need to finish the exercises and send your exercise to me / or show in class before 1200:
  - Christoph.Luchsinger@math.uzh.ch
  - This will be the basis for receiving the certification!

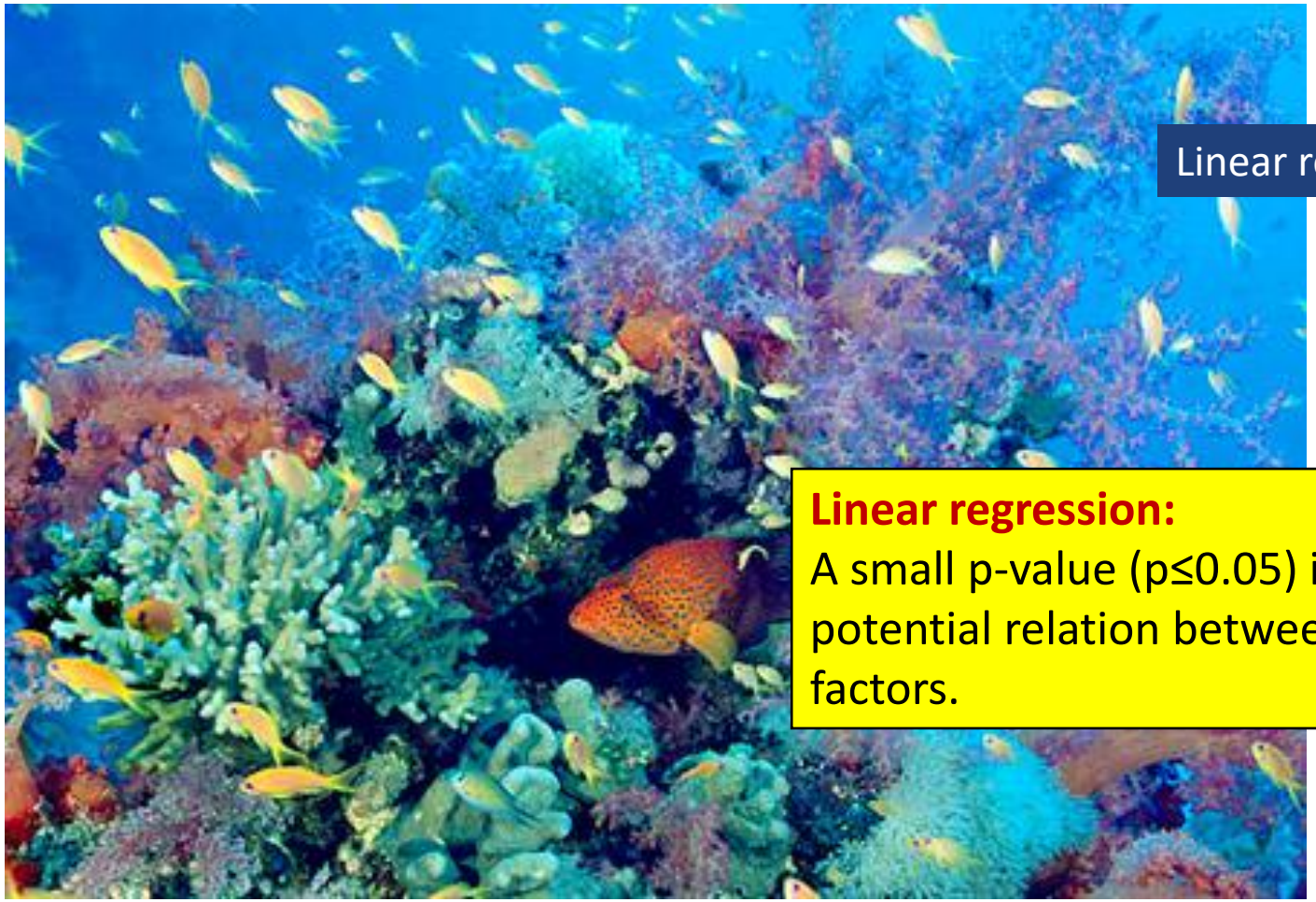
What is statistics?



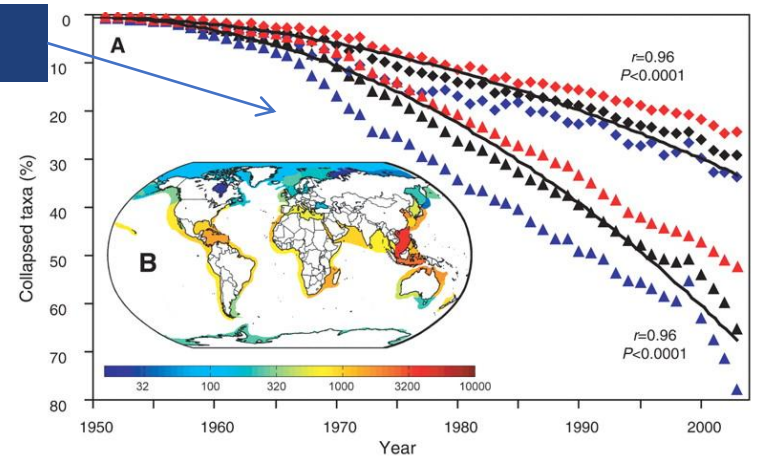
Regression in science

Impacts of Biodiversity Loss on Ocean Ecosystem Services

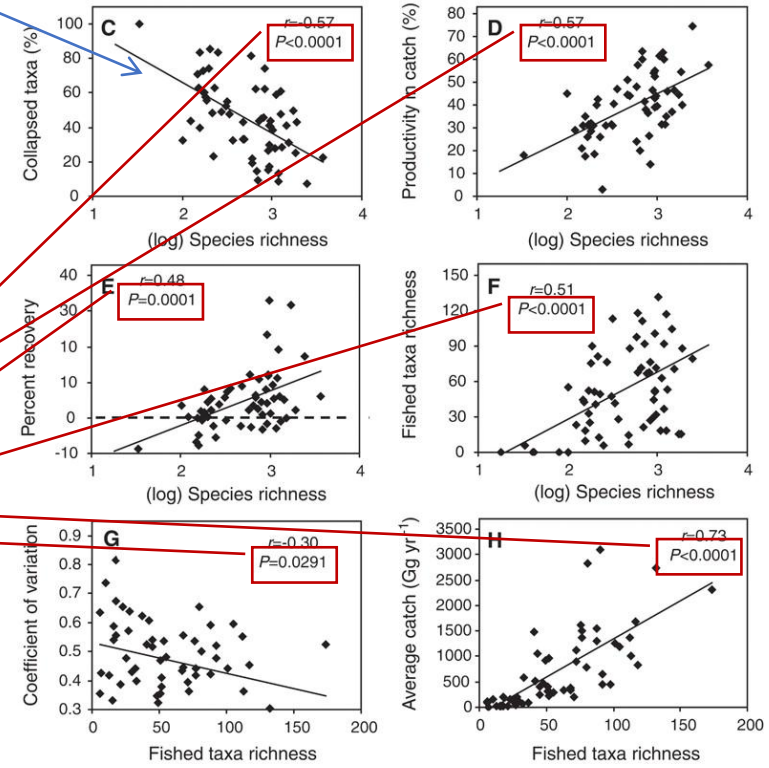
Boris Worm[1,*], Edward B. Barbier[2], Nicola Beaumont[3], J. Emmett Duffy[4], Carl Folke[5,6], Benjamin S. Halpern[7], Jeremy B. C. Jackson[8,9], Heike K. Lotze[1], Fiorenza Micheli[10], Stephen R. Palumbi[10], Enric Sala[8], Kimberley A. Selkoe[7], John J. Stachowicz[11], Reg Watson[12]



Comparing distributions

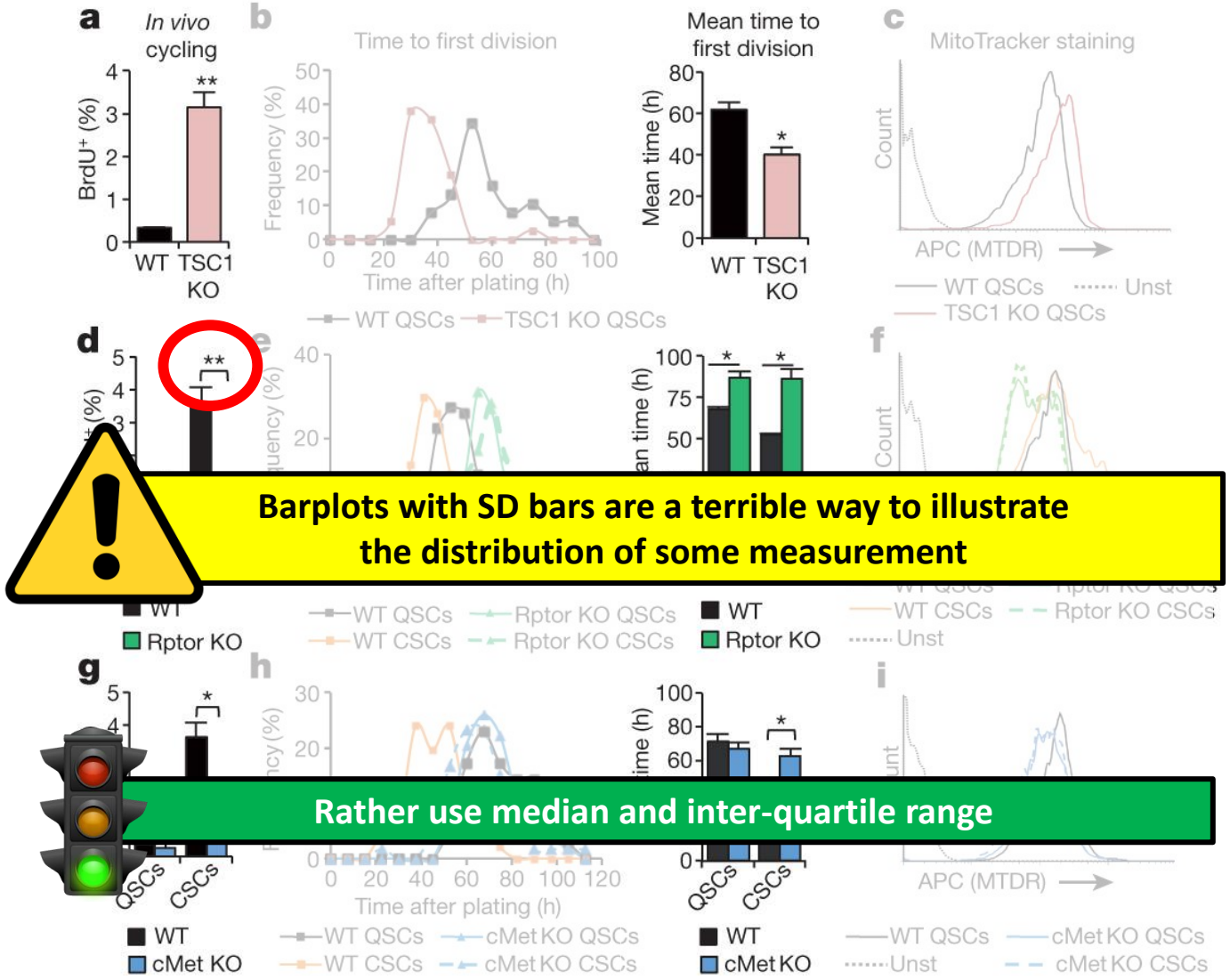


Linear regression



Linear regression:
A small p-value ($p \leq 0.05$) indicates a potential relation between the two factors.

Activation of mTORC1 is necessary and sufficient for the alert phenotype



nature

How to test a hypothesis

- An example...

FIFA WORLD CUP
RUSSIA
2018

Example inspired by Dr. Markus Kalisch, SfS, ETH Zurich

650 stickers

FIFA WORLD CUP RUSSIA 2018

Example inspired by Dr. Markus Kalisch, SfS, ETH Zurich

Blister

5 Stickers

Box

100 Blister = 500 Stickers

FIFA WORLD CUP
RUSSIA
2018

Example inspired by Dr. Markus Kalisch, SfS, ETH Zurich

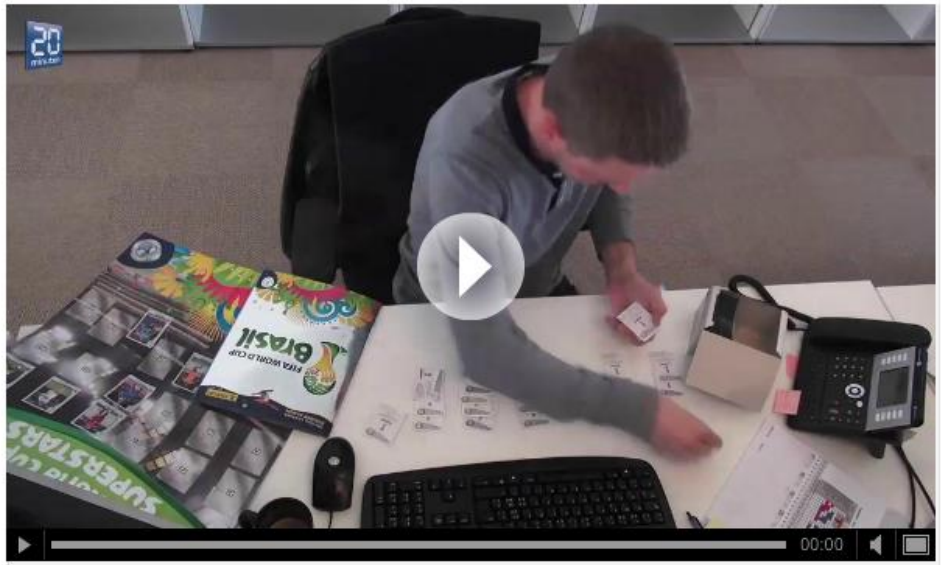
?

FIFA WORLD CUP
RUSSIA
2018

Example inspired by Dr. Markus Kalisch, SfS, ETH Zurich

Keine doppelten Bilder – aber Messi als Superstar

von Herbie Egli - Das Panini-Fieber grassiert: 20 Minuten hat innerhalb von zwei Stunden alle 500 Bilder einer Box ins Album geklebt – und vermisst die doppelten.



20 Minuten-Sportredaktor Herbie Egli klebte innerhalb von zwei Stunden alle 500 Bilder einer Panini-Box ins Album ein. (Video: 20 Minuten)

Im Grossformat auf dem Videoportal

ein aus i

17

Empfehlen

4

Morgens um 6 Uhr. Ich bin auf dem Weg zu meinem Frühdienst. Nie gehe ich an den Kiosk etwas besorgen, bevor ich den Zug besteige. Am 21. März muss ich jedoch einen Zwischenstopp einlegen: Die begehrten Panini-Bilder sind endlich im Handel.

Ich kaufe mir gleich eine ganze Box mit total 500 Bildern und

Bildstrecken

ZEIT ONLINE

Suche

Politik Gesellschaft **Wirtschaft** Kultur Wissen Digital Studium Karriere Entdecken Sport mehr ZEITmagazin

Panini

Das große teure Kleben

Panini – das sind Millionenumsätze, geheime Algorithmen und die Jagd nach dem letzten Sticker. Über das Geschäft mit der Sammelwut, dem das Internet nichts anhaben kann.

Von **Anne-Sophie Lang**

6. Juni 2014, 10:27 Uhr / 28 Kommentare

Ein Panini-Mitarbeiter sortiert die Sticker in eine Maschine ein © Reuters

SPIEGEL ONLINE DER SPIEGEL SPIEGEL TV

Anmelden

WISSENSCHAFT

Schlagzeilen | Wetter | DAX 11.886,02 | TV-Programm | Abo

Nachrichten > Wissenschaft > Mensch > Fußball-EM 2016 > Panini-Sticker zur EM 2016: So kommen Sie günstig zum vollen Album

EM-Sticker

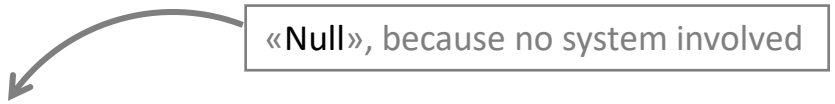
So kriegen Sie das Panini-Album günstig voll

Dreimal Ronaldo, aber kein Neuer: Sammelbildfans sind schon Wochen vor Anpfiff im EM-Fieber. Mit welcher Strategie füllt man sein Panini-Album möglichst günstig?

Von Holger Dambeck

Gut feeling (system 1) and hypotheses

- Complete box → less double stickers
- Buy blisters all over town → many doubles



- Nullhypothesis:
  - Stickers are filled **randomly** into the boxes
- Alternative hypothesis:
  - Stickers are filled **systematically** into the boxes such that there are less doubles

How can we decide between these two?

Hypothesis test

- I bought a box with 500 stickers and my god son was able to glue 493 stickers (so 7 doubles) into the album (650 possible pictures).

- If we assume that the null hypothesis $\mathcal{H}_0$ was true:
  - Is it plausible, to glue 493 stickers into the album?

- Do the null hypothesis $\mathcal{H}_0$: «*randomly packed*» and the observation *«493 stickers glued in*» fit together?

Problem: what is «*normal*»?

- If we can glue in much more than «*normal*», we probably conclude that they were not packed randomly.

- Given $\mathcal{H}_0$ is true, how many pictures would be «*normal*» and what would be «*much more*»?

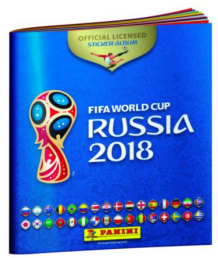
- Significance level α : how «*much more*» does an observation have to be, for us to believe that it is not «*normal*» anymore?
  - E.g. α = 1/1'000'000
    → we reject $\mathcal{H}_0$, if an observation occurs less than 1 in 1'000'000 cases.

Solution for normal: computer simulation

1

342

2

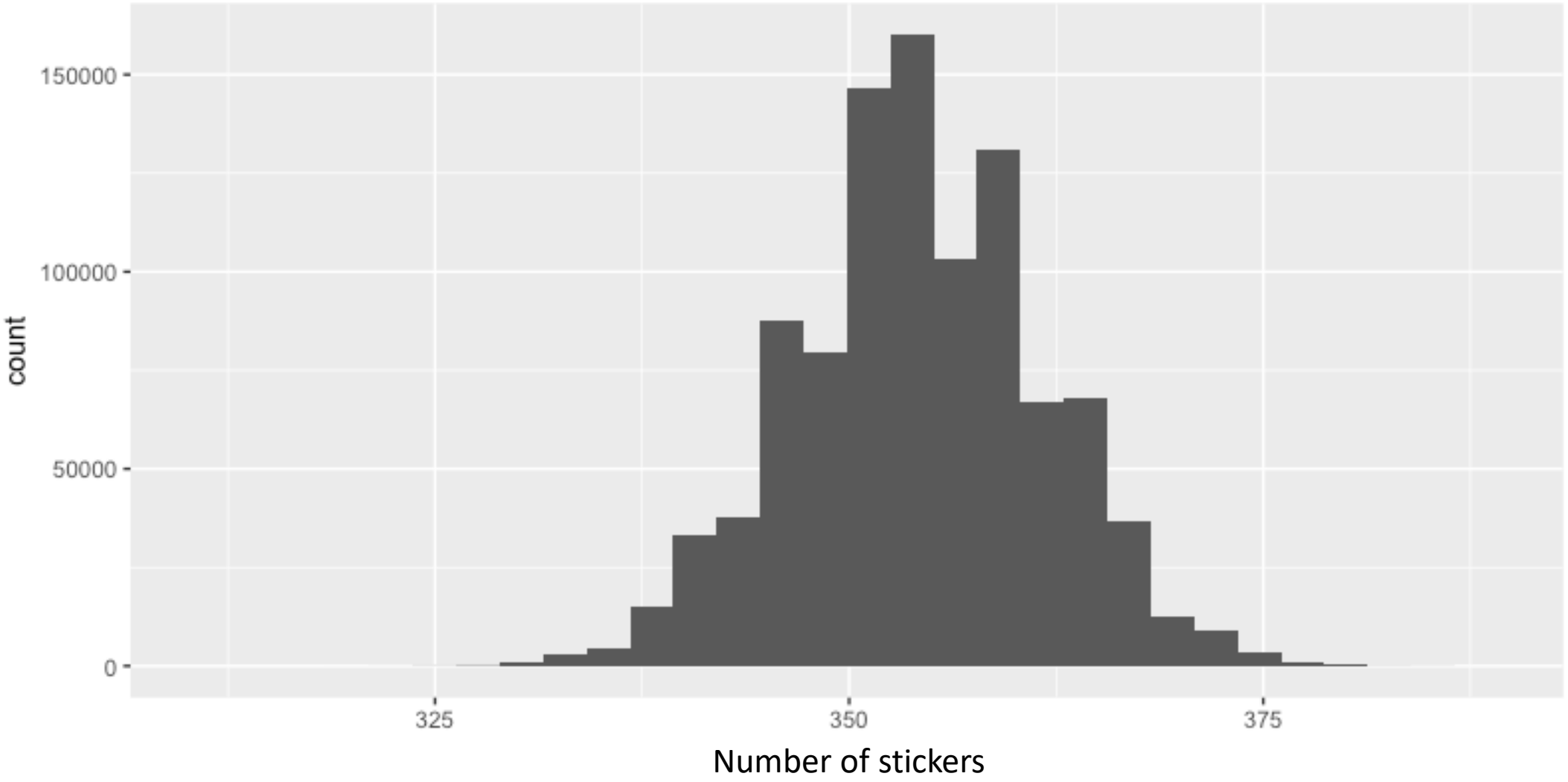


351

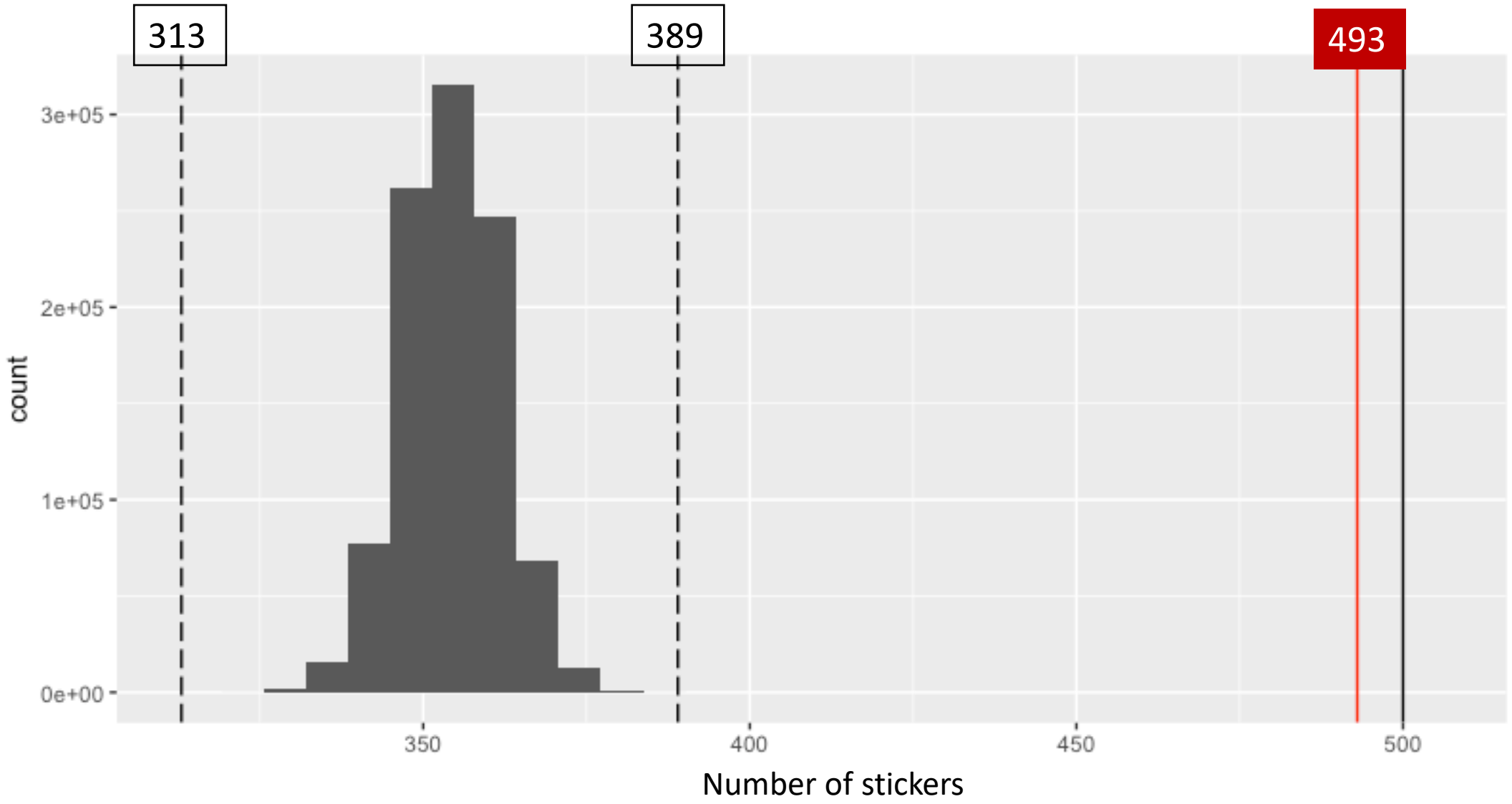
1 mio

337

Result of computer simulation



Normal and our observation

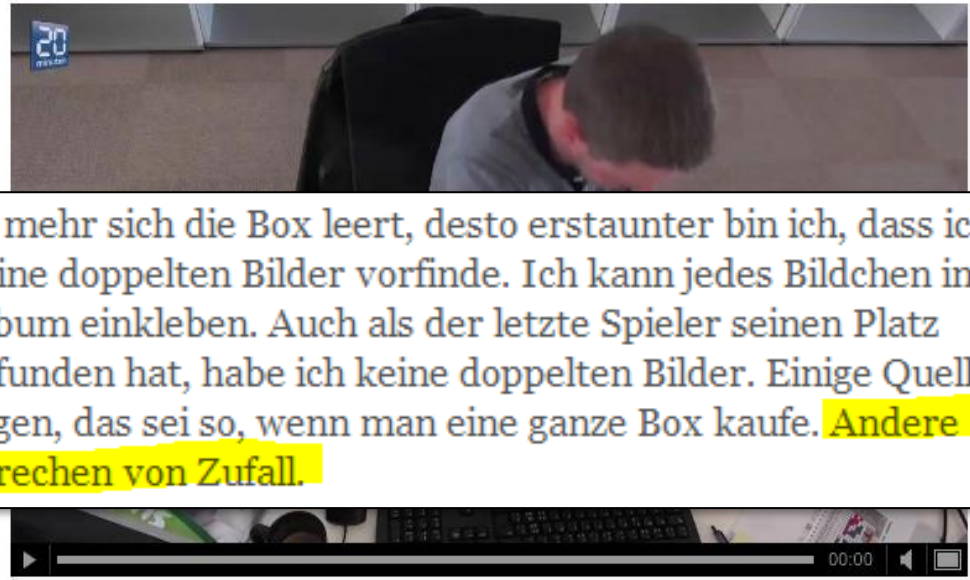


Conclusion

- If we assume that the stickers were packed randomly:
  - The probability to glue 493 or more stickers into the album would be less than 1 in a million!
  - ➢ Our observation and the null hypothesis do not fit together!
- Human being / decision:

Keine doppelten Bilder – aber Messi als Superstar

von Herbie Egli - Das Panini-Fieber grassiert: 20 Minuten hat innerhalb von zwei Stunden alle 500 Bilder einer Box ins Album geklebt – und vermisst die doppelten.



Je mehr sich die Box leert, desto erstaunter bin ich, dass ich noch keine doppelten Bilder vorfinde. Ich kann jedes Bildchen im Album einkleben. Auch als der letzte Spieler seinen Platz gefunden hat, habe ich keine doppelten Bilder. Einige Quellen sagen, das sei so, wenn man eine ganze Box kaufe. Andere sprechen von Zufall.

20 Minuten-Sportredaktor Herbie Egli klebte innerhalb von zwei Stunden alle 500 Bilder einer Panini-Box ins Album ein. (Video: 20 Minuten)

Im Grossformat auf dem Videoportal

Morgens um 6 Uhr. Ich bin auf dem Weg zu meinem Frühdienst. Nie gehe ich an den Kiosk etwas besorgen, bevor ich den Zug besteige. Am 21. März muss ich jedoch einen Zwischenstopp einlegen: Die begehrten Panini-Bilder sind endlich im Handel.

Ich kaufe mir gleich eine ganze

Summary: hypothesis test

1. Modell: draw 500 stickers from 650 possible slots
2. Null hypothesis $\mathcal{H}_0$: «*stickers are packed **randomly***»
   Alternative $\mathcal{H}_A$: «*stickers are packed **systematically** to avoid doubles*»
3. Test statistic: Number of stickers glued into an album, when buying a box of 500 stickers.
   Distribution of *test statistic* under the null hypothesis: computer simulation
4. Significance level α = 1/1'000'000 (you choose, the journal, pharma, regulator (FDA)
5. Critical region of the *test statistic*:
   In 1 mio runs the computer never saw more than 389 stickers
   → *critical region*: K={389, 390, …, 500} – entire tail of the distribution
6. Decision: The observed value (493) is inside the *critical region* of the *test statistic*.
   → The *null hypothesis* will be rejected on a *significance level* of α.

Models for probability

Three components:

1. Space Ω → *Universe*, elementary event ω_i → *"Atoms"*
2. Events *A*, *B*, *C*, …, Subsets of Ω → *"Molecules"*
3. Probability for an event $P(A)$, $P(B)$, $P(C)$, … → *Chance to find molecule*

Using events to compute stuff

- Operations from Set Theory (dt Mengenlehre) can be directly applied:

$$A \cup B \Leftrightarrow A \text{ or } B$$

$$A \cap B \Leftrightarrow A \text{ and } B$$

$$A^C \Leftrightarrow \text{not } A$$

- Example
  - $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - $\omega_1 = 1, \omega_2 = 2, \dots, \omega_6 = 6$
  - Event E: throw an even number → $E = \{2, 4, 6\}$
  - Probability for this event: $P(E) = 0.5 = 50\%$

Axioms for probability

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. $P(A \cup B) = P(A) + P(B)$, if $P(A \cap B) = 0$ (disjoint, disjunkt/elementfremd)

From this we can conclude…

$$1 = P(\Omega) = P(A \cup A^C) = P(A) + P(A^C)$$

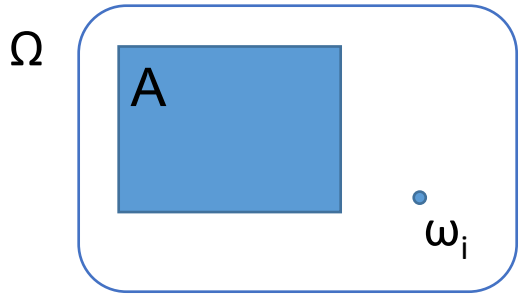
$$\Rightarrow P(A^C) = 1 - P(A)$$

…or much simpler…

Venn Diagrams

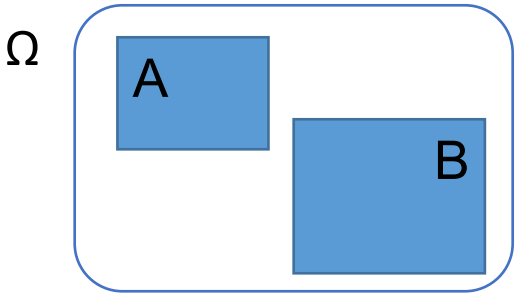
- Treat probability as an area (as you learned in set theory)…

$P(A)$

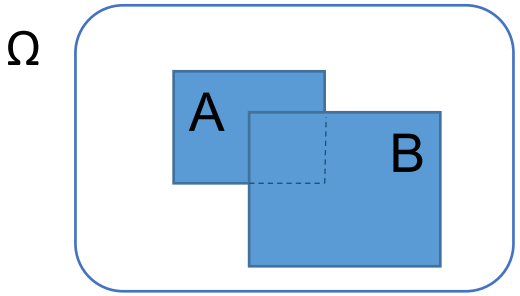


$P(\Omega) = 1$

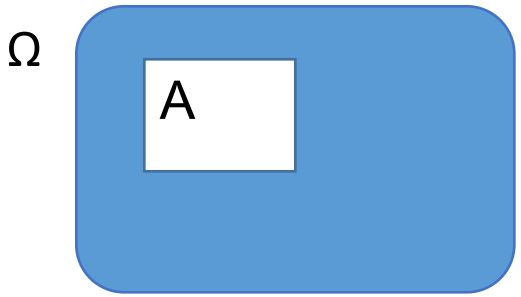
$P(A \cup B) = P(A) + P(B)$



$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

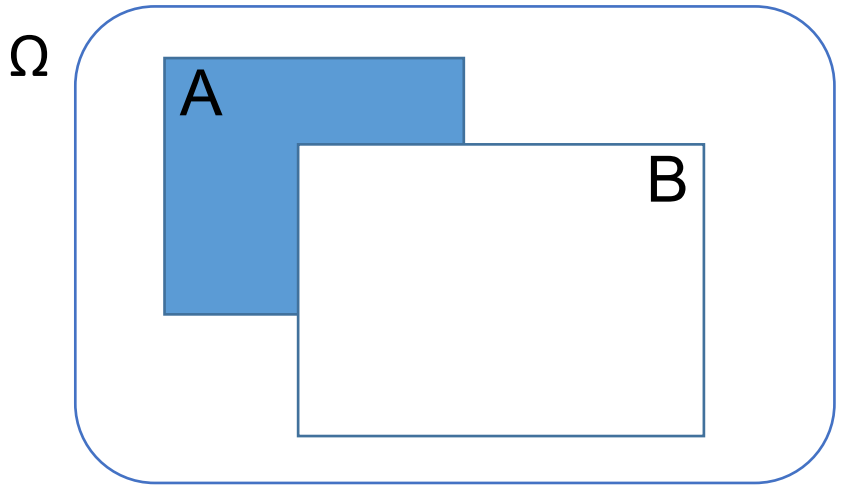


$P(A^C) = 1 - P(A)$



Exercise

$$P(A \backslash B) = \ldots$$



«A without B»

1. $P(A) - P(B)$
2. $P(A) + P(B) - P(A \cap B)$
3. $P(A) - P(A \cap B)$
4. $P(A) + P(B)$

Compute probabilities

1. Sum of elementary events

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

- Example: **Lotto**
  - $P(Winnings \geq 100) = P(\omega_3 \cup \omega_4)$
    $= P(\omega_3) + P(\omega_4)$
    $= 0.05 + 0.05$
    $= 0.1$

Wins	$P(\omega_i)$
$\omega_1 = 1$	0.8
$\omega_2 = 10$	0.1
$\omega_3 = 100$	0.05
$\omega_4 = 1000$	0.05

Compute probabilities

2. Laplace model:
   - All elementary events have the same probability

$$P(A) = \frac{\#\ favorable\ Events}{\#\ possible\ Events}$$

- Example: **Throw a W20**
  - Event V: results larger 17
    $\Rightarrow V = \{\omega_{18}, \omega_{19}, \omega_{20}\}$
  - $P(V) = \frac{|V|}{|\Omega|} = \frac{3}{20} = 0.15$

Results	$P(\omega_i)$
$\omega_1 = 1$	$1/20$
$\omega_2 = 2$	$1/20$
…	…
$\omega_i = i$	$1/20$
…	…
$\omega_{20} = 20$	$1/20$

Compute probabilities

3. Set operations

$$P(A^C) = 1 - P(A)$$

- Example: **Tombola** – 10 tickets bought
  - Event N: *not a single prize in 10 tickets*
  - Given: $P(N) = 0.03$
  - Wanted: $P(at\ least\ 1\ win) = P(N^C)$

    $P(N^C) = 1 - P(N) = 1 - 0.03 = 0.97$
  - Useful: A and B independent: $P(A \cap B) = P(A)P(B)$

Summary

- Statistics makes statements about a population by using a small sample
- Football stickers are packed into boxes systematically
- Three components:
  - Space made of elementary events
  - Events are sets of elementary events
  - Probability of events
- Probabilities can (and should) be computed