Why this lesson?

- Going back to the foundations of our work: how are we advancing knowledge ?
- Trying to help us stand back and think
- Most of the material is from Z. Dienes, psychology as a science
- Disclaimer 1: I am no epistemiologist (but we need to think about epistemiology)
- Disclaimer 2: "The philosophy of science is as useful to scientists as ornithology is to birds" (Richard Feynman)

Karl Popper: 1902-1994

- Austrian, held chair in logic and scientific methods (London School of Economics)

Karl Popper: 1902-1994

- Austrian, held chair in logic and scientific methods (London School of Economics)
- Interested in the questions:
  - distinguishing science from pseudo science: "The demarcation pb"
  - how can we best grow the scientific knowledge (ie, make fastest progress)

Karl Popper: 1902-1994

- Austrian, held chair in logic and scientific methods (London School of Economics)
- Interested in the questions:
  - distinguishing science from pseudo science: “The demarcation pb”
  - how can we best grow the scientific knowledge (ie, make fastest progress)
- Background: Reacting against logical positivism:
  - some sentences are not verifiable (eg: “free will is an illusion”)
  - 2 problems to solve:
    - verify a specific statement: “this swan is white”
    - generalization: “all swans are white”
    - induction: seeing many examples of a fact leads to trust that this fact is “true”

Karl Popper: killing inductionisms / logical positivism

- Theory: in human being, the hippocampus is required for spatial navigation

Karl Popper: killing inductionisms / logical positivism

- Theory: in human being, the hippocampus is required for spatial navigation
- You find the example of an individual John whose hipp. and only it is destroyed and cannot do any spatial navigation
  - what can you say about the theory ?

Karl Popper: killing inductionisms / logical positivism

- Theory: in human being, the hippocampus is required for spatial navigation
- You find the example of an individual John whose hipp. and only it is destroyed and cannot do any spatial navigation
  - what can you say about the theory ?
- You find 12 individuals with the same destruction: their spatial navigation is very bad
  - what can you say ?

Karl Popper: killing inductionisms / logical positivism

- Theory: in human being, the hippocampus is required for spatial navigation
- You find the example of an individual John whose hipp. and only it is destroyed and cannot do any spatial navigation
  - what can you say about the theory ?
- You find 12 individuals with the same destruction: their spatial navigation is very bad
  - what can you say ?
- You find an individual with the same destruction, but their spatial navigation is good
  - what do you conclude about the theory ?

David Hume (1711 - 1776)

- Induction issue:
  - We cannot reason from past instances to future instance
    - Swan example

David Hume (1711 - 1776)

- Induction issue:
  - We cannot reason from past instances to future instance
    - Swan example
  - What about the probability of the theory ?
    - all future instances may disprove the theory

David Hume (1711 - 1776)

- Induction issue:
  - We cannot reason from past instances to future instance
    - Swan example
  - What about the probability of the theory ?
    - all future instances may disprove the theory
  - Has induction not already proved it self ?
    - requires to accept induction to prove it !

David Hume (1711 - 1776)

- Induction issue:
  - We cannot reason from past instances to future instance
    - Swan example
  - What about the probability of the theory ?
    - all future instances may disprove the theory
  - Has induction not already proved it self ?
    - requires to accept induction to prove it !
  - Example: Newtonian vs relativity and quantum physics

David Hume (1711 - 1776)

- Induction issue:
  - We cannot reason from past instances to future instance
    - Swan example
  - What about the probability of the theory ?
    - all future instances may disprove the theory
  - Has induction not already proved it self ?
    - requires to accept induction to prove it !
  - Example: Newtonian vs relativity and quantum physics
- Popper accepted all these arguments

Problem to be solved: choosing between theories

- Bertrand Russell (1872-1970): Induction is necessary to distinguish good from bad theories

Problem to be solved: choosing between theories

- Bertrand Russell (1872-1970): Induction is necessary to distinguish good from bad theories
- Popper argues back that induction is not needed:
    - a fact that complies with the theory does not prove the theory "All swans are white" - does a blue pen add support to the theory ?

Problem to be solved: choosing between theories

- Bertrand Russell (1872-1970): Induction is necessary to distinguish good from bad theories
- Popper argues back that induction is not needed:
  - a fact that complies with the theory does not prove the theory "All swans are white" - does a blue pen add support to the theory ?
  - a fact that does not comply with the theory disproves it

Problem to be solved: choosing between theories

- Bertrand Russell (1872-1970): Induction is necessary to distinguish good from bad theories
- Popper argues back that induction is not needed:
  - a fact that complies with the theory does not prove the theory "All swans are white" - does a blue pen add support to the theory ?
  - a fact that does not comply with the theory disproves it
  - asymmetry of the problem: falsification is the right tool

Problem to be solved: choosing between theories

- Bertrand Russell (1872-1970): Induction is necessary to distinguish good from bad theories
- Popper argues back that induction is not needed:
  - a fact that complies with the theory does not prove the theory “All swans are white” - does a blue pen add support to the theory ?
  - a fact that does not comply with the theory disproves it
  - asymmetry of the problem: falsification is the right tool
- Criteria: falsification and simplicity
  - More falsification is better
    - more precise
    - applies to a broader range of situations

Problem to be solved: choosing between theories

- Bertrand Russell (1872-1970): Induction is necessary to distinguish good from bad theories
- Popper argues back that induction is not needed:
  - a fact that complies with the theory does not prove the theory "All swans are white" - does a blue pen add support to the theory ?
  - a fact that does not comply with the theory disproves it
  - asymmetry of the problem: falsification is the right tool
- Criteria: falsification and simplicity
  - More falsification is better
    - more precise
    - applies to a broader range of situations
  - A simpler theory is better

Probabilistic problems

- The DLPFC of patients suffering from Parkinson disease is *on average* greater than those of NC
  - not falsifiable ? would require all PD patients ?
  - probability theory allows for a fair coin to land 1M times on 'head'

Probabilistic problems

- The DLPFC of patients suffering from Parkinson disease is *on average* greater than those of NC
  - not falsifiable ? would require all PD patients ?
  - probability theory allows for a fair coin to land 1M times on ‘head’
- Leads to the creation of a severe test
  - Does this lead to classical statistical theory ?

Probabilistic problems

- The DLPFC of patients suffering from Parkinson disease is *on average* greater than those of NC
  - not falsifiable ? would require all PD patients ?
  - probability theory allows for a fair coin to land 1M times on 'head'
- Leads to the creation of a severe test
  - Does this lead to classical statistical theory ?
- See arguments that "NHST" is not "Popperian"

Probabilistic problems

- The DLPFC of patients suffering from Parkinson disease is *on average* greater than those of NC
  - not falsifiable ? would require all PD patients ?
  - probability theory allows for a fair coin to land 1M times on 'head'
- Leads to the creation of a severe test
  - Does this lead to classical statistical theory ?
- See arguments that "NHST" is not "Popperian" * to be Popperian, you would need to try to falsify H1 ? - instead : we try to falsify H0 to show that we cannot falsify H1 - but what is H0 ? in many instances, a very unlikely hypothesis
  - A statistical test never entirely falsifies !

A slide on Khun

- Best known for “The structure of scientific revolutions”

A slide on Khun

- Best known for “The structure of scientific revolutions”
- Scientific truth cannot be established solely by objective criteria

A slide on Khun

- Best known for “The structure of scientific revolutions”
- Scientific truth cannot be established solely by objective criteria
- Science must account for subjective perspectives
    - and yet we are after something that is not only in our mind

A slide on Khun

- Best known for “The structure of scientific revolutions”
- Scientific truth cannot be established solely by objective criteria
- Science must account for subjective perspectives
  - and yet we are after something that is not only in our mind
- Normal science: when a paradigm is established (and we try to fit results into a common framework)
  - “The abandonment of critical discourse marks the transition to science”

A slide on Khun

- Best known for "The structure of scientific revolutions"
- Scientific truth cannot be established solely by objective criteria
- Science must account for subjective perspectives
  - and yet we are after something that is not only in our mind
- Normal science: when a paradigm is established (and we try to fit results into a common framework)
  - "The abandonment of critical discourse marks the transition to science"
- The concept of a "paradigm shift" : science progress with new paradigms
  - the earth rotation in the solar system

A slide on Khun

- Best known for "The structure of scientific revolutions"
- Scientific truth cannot be established solely by objective criteria
- Science must account for subjective perspectives
  - and yet we are after something that is not only in our mind
- Normal science: when a paradigm is established (and we try to fit results into a common framework)
  - "The abandonment of critical discourse marks the transition to science"
- The concept of a "paradigm shift" : science progress with new paradigms
  - the earth rotation in the solar system
- Paradigm shifts are prompted by too many inconsistencies in the current paradigm

But - What is a probability ?

- 3 axioms:
  - Axiom 1 and 2 : P(an event A) $>=$ 0; P(all events) $=$ 1
  - Axiom 3: If A and B are *mutually exclusive*, then P(A or B) $=$ P(A) $+$ P(B)
  - hence: P(A or B) $=$ P(A) $+$ P(B) - P(A and B)
  - Often part of the definition: P(A and B) $=$ P(A)P(B$|$A)

But - What is a probability ?

- 3 axioms:
  - Axiom 1 and 2 : P(an event A) $>=$ 0; P(all events) $=$ 1
  - Axiom 3: If A and B are *mutually exclusive*, then P(A or B) $=$ P(A) $+$ P(B)
  - hence: P(A or B) $=$ P(A) $+$ P(B) - P(A and B)
  - Often part of the definition: P(A and B) $=$ P(A)P(B$|$A)
- Subjective : a degree of belief ("Evidential probability")

But - What is a probability ?

- 3 axioms:
  - Axiom 1 and 2 : P(an event A) $>= 0$; P(all events) $= 1$
  - Axiom 3: If A and B are *mutually exclusive*, then P(A or B) $=$ P(A) $+$ P(B)
  - hence: P(A or B) $=$ P(A) $+$ P(B) - P(A and B)
  - Often part of the definition: P(A and B) $=$ P(A)P(B$|$A)
- Subjective : a degree of belief ("Evidential probability")
- Objective/physical : a property of the nature or system
  - example : probability of the gaz molecules to hit the wall

But - What is a probability ?

- 3 axioms:
  - Axiom 1 and 2 : P(an event A) $>=$ 0; P(all events) $=$ 1
  - Axiom 3: If A and B are *mutually exclusive*, then P(A or B) $=$ P(A) $+$ P(B)
  - hence: P(A or B) $=$ P(A) $+$ P(B) - P(A and B)
  - Often part of the definition: P(A and B) $=$ P(A)P(B|A)
- Subjective : a degree of belief ("Evidential probability")
- Objective/physical : a property of the nature or system
  - example : probability of the gaz molecules to hit the wall
- Frequentist interpretation: limit of the relative frequency of an event in a large number of random trials

But - What is a probability ?

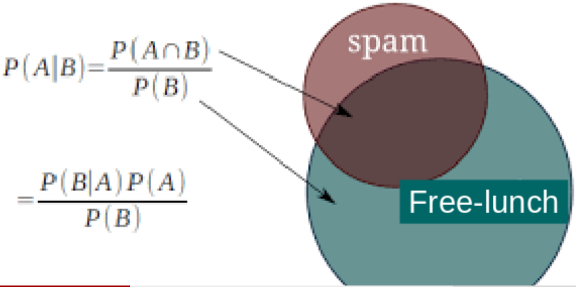
- 3 axioms:
  - Axiom 1 and 2 : P(an event A) $>=$ 0; P(all events) $=$ 1
  - Axiom 3: If A and B are *mutually exclusive*, then P(A or B) $=$ P(A) $+$ P(B)
  - hence: P(A or B) $=$ P(A) $+$ P(B) - P(A and B)
  - Often part of the definition: P(A and B) $=$ P(A)P(B$|$A)
- Subjective : a degree of belief ("Evidential probability")
- Objective/physical : a property of the nature or system
  - example : probability of the gaz molecules to hit the wall
- Frequentist interpretation: limit of the relative frequency of an event in a large number of random trials
- Objective/physical probability associated with a **collective**
  - the set of dice throws - a single event can be part of several collectives
  - prob(catch a cold) : Collective 1: people who live in cold climates. Collective 2: air conditioner full blast.

Bayes and likelihood functions

- Let's work with Bayesian probability - can still be related to objective ones
  - "as reasonable expectation representing a state of knowledge or as quantification of a personal belief"

Bayes and likelihood functions

- Let's work with Bayesian probability - can still be related to objective ones
    - "as reasonable expectation representing a state of knowledge or as quantification of a personal belief"
- Subjective/Evidential probabilities should still follow the physical probability axioms



Bayes and likelihood functions

- Derivation of Bayes theorem is easy: accept conditional probabilities

$$P(H, D) = P(H|D)P(D)$$

$$P(H, D) = P(D|H)P(H)$$

$$P(H|D) = \frac{P(D, H)}{P(D)} = \frac{P(D|H)P(H)}{P(D)}$$

$$P(H_1|D) = \frac{P(D, H_1)}{P(D)} = \frac{P(D|H_1)P(H_1)}{P(D)}$$

Bayes and likelihood functions : can we use this to choose between theories?

- Posterior: $P(H|D)$
- Prior: $P(H|D)$
- Likelihood: P(D | H) : Careful: not a frequentist probability !

Bayes and likelihood functions : can we use this to choose between theories?

- Posterior: $P(H|D)$
- Prior: $P(H|D)$
- Likelihood: P(D | H) : Careful: not a frequentist probability !

$$P(H_1|D) = \frac{P(D, H_1)}{P(D)} = \frac{P(D|H_1)P(H_1)}{P(D)}$$

$$P(H_0|D) = \frac{P(D, H_0)}{P(D)} = \frac{P(D|H_0)P(H_0)}{P(D)}$$

- Ratio: posterior odds = Bayes-factor $\times$ prior-odds

Bayes and likelihood functions : can we use this to choose between theories?

- Posterior: $P(H|D)$
- Prior: $P(H|D)$
- Likelihood: P(D | H) : Careful: not a frequentist probability !

$$P(H_1|D) = \frac{P(D, H_1)}{P(D)} = \frac{P(D|H_1)P(H_1)}{P(D)}$$

$$P(H_0|D) = \frac{P(D, H_0)}{P(D)} = \frac{P(D|H_0)P(H_0)}{P(D)}$$

- Ratio: posterior odds = Bayes-factor $\times$ prior-odds
- $BF = \frac{P(D|H_1)}{P(D|H_0)}$
- <1 : supports H0, 1-3: not worth mention , 3-10: substantial, 10-30: strong

Back to the basics: Effect size

What is the non standardized effect ?

Imagine 2 groups (1 and 2):

$$\mu = \bar{x}_1 - \bar{x}_2$$

What is the standardized effect ? (eg Cohen's d)

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sigma} = \frac{\mu}{\sigma}$$

"Z" : Effect accounting for the sample size

$$Z = \frac{\mu}{\sigma/\sqrt{n}}$$

Significance testing as perverse probabilistic reasoning

Consider a typical medical research study, for example designed to test the efficacy of a drug, in which a null hypothesis H_0 ('no effect') is tested against an alternative hypothesis H_1 ('some effect'). Suppose that the study results pass a test of statistical significance (that is P-value <0.05) in favor of H_1. What has been shown?

1. H_0 is false.
2. H_1 is true.
3. H_0 is probably false.
4. H_1 is probably true.
5. Both (1) and (2).
6. Both (3) and (4).
7. None of the above.

Figure 2: Westover, 2014

Significance testing as perverse probabilistic reasoning

Table 1 Quiz answer profile

Answer	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Number	8	0	58	37	6	69	12
Percent	4.2	0	30.5	19.5	3.2	36.3	6.3

Figure 3: Westover, 2014

- Westover, 2014

Back to statistics: p-values

- definition of a p-value
  - Probability of observing a **statistic** equal to the one seen in the data, or one that is more "extreme", when the null hypothesis is true

Back to statistics: p-values

- definition of a p-value
  - Probability of observing a **statistic** equal to the one seen in the data, or one that is more "extreme", when the null hypothesis is true
  - a statistics = a function of the data: s = f(Data)
  - define with common sense : eg difference between the means
  - but: what if there are different choices ? What if several could be biologically relevant?
  - can I observe several statistics ? Is that a problem ?
  - test can be "more or less powerful"

Back to statistics: p-values

- definition of a p-value
  - Probability of observing a statistic equal to the one seen in the data, or one that is more "extreme", when the null hypothesis is true
- Probability of observing a statistic **equal to the one seen in the data, or one that is more "extreme"**, when the null hypothesis is true
  - concept of repeating the same study in the same way an infinite number of times !
  - same study design
  - same sampling scheme

Back to statistics: p-values and power ?

- Probability of observing a statistic equal to the one seen in the data, or one that is more "extreme", **when the null hypothesis is true**
  - a statistics = a function of the data: s = f(Data)
  - How do we define the null ?
  - Is the null plausible ? or at least possible ?

p-values: Illustration with a normal null

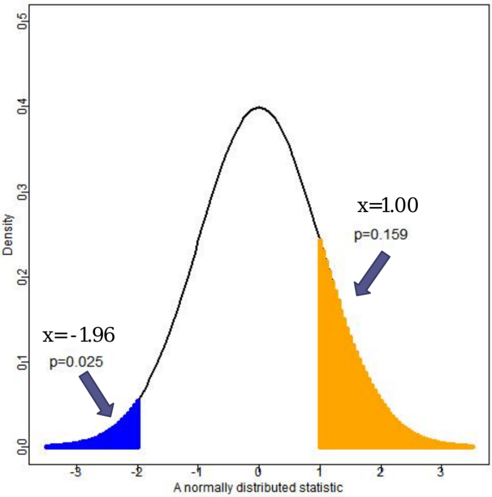


Figure 4: Illustration of p-value

p-values: Illustration with a Gamma null

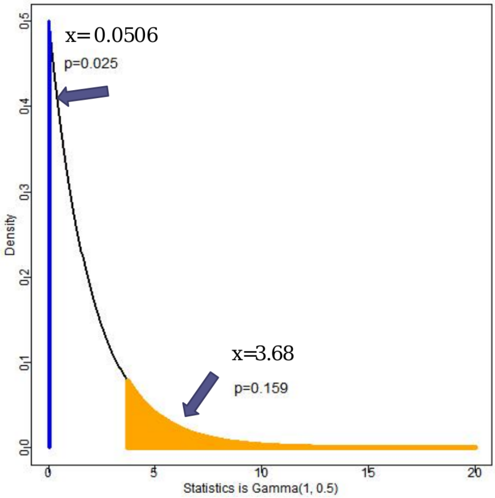


Figure 5: Illustration of p-value

p-values: Illustration with a uniform null

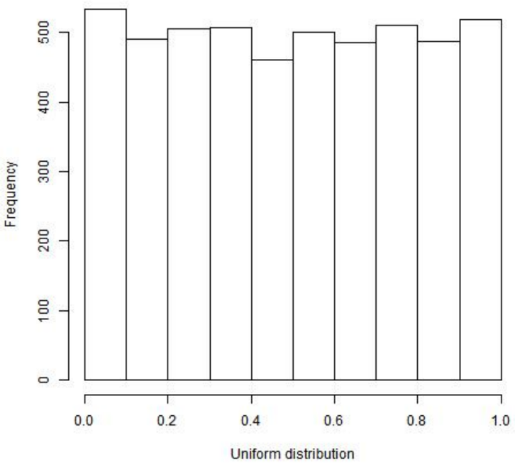


Figure 6: Illustration of p-value

Back to statistics: p-values and power

Decision/H	H0 True	H1 True
reject	α (type I)	$1 - \beta$ (Power)
not reject	1 - α	β (type II)

What is power ?

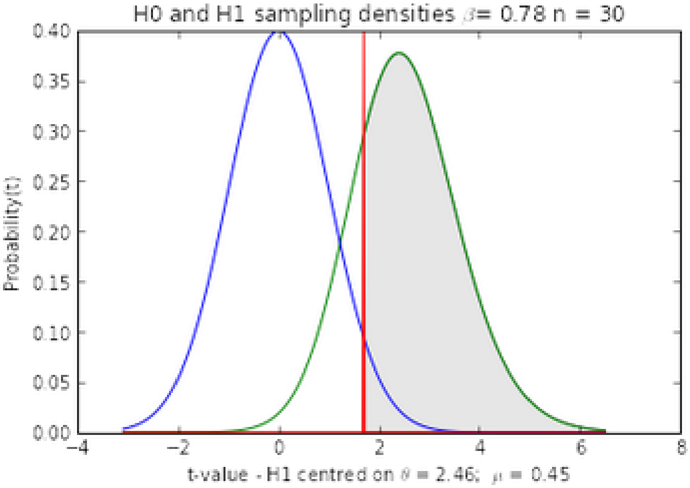


Figure 7: Illustration of Power

Power: experimentally

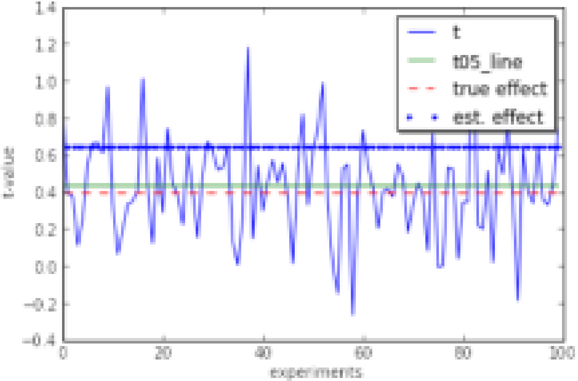


Figure 8: Illustration of Power

recall the definition

- Power will depend on **5** things:
  - The non standardized effect : μ
  - The standard deviation of the data : σ
  - The number of subjects : n
  - The type I risk of error : α
  - And on the distribution of the statistic under the alternative hypothesis.

Computing power in a simple case:

- We estimate the effect $\hat{\mu}$ under normal noise. Our statistic is:

$$t_{obs} = \frac{\hat{\mu}}{\hat{\sigma_{\mu}}} = \frac{\hat{\mu}}{S\hat{E}_{\mu}}$$

- Power is $P(t_{obs} > t_{.05})$, with $t_{.05}$ the t for $\alpha = 0.05$ **under the null**.
- We easily compute $t_{.05}$ because we know the null, but need to compute $P(t_{obs} > t_{.05})$.
- We therefore need the distribution of t_{obs} - and therefore we need to **know the signal / effect size** !
- Let's assume we know this and call it t_{nc}, and F_{nc} for the cumulative distribution (nc for non central).
- Power $= 1 - \beta = P(t_{obs} > t_{.05}) = 1 - F_{nc}(t_{.05})$

We need a theoretical result

- With normal data, the $t = \frac{\hat{\mu}}{\hat{SE}_{\mu}}$ statistic follows *a non central t distribution* with non centrality parameter:

$$\theta = \mu\sqrt{n}/\sigma$$

  and $n-1$ degrees of freedom.
- We are done ! **assuming we know all these things . . . ;)**

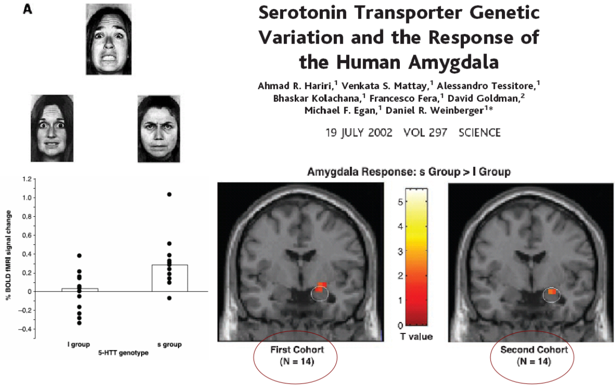
Some python code:

```
Inputs: sample_size (n), mu, sigma, alpha
Returns: power
"""
import scipy.stats as sst

# define H1
df = n-1
theta = np.sqrt(n)*mu/sigma
ncrv = sst.nct(df, theta)

# find the threshold value
t_alph_null = sst.t.isf(alpha, df)
power = 1 - ncrv.cdf(t_alph_null)
return power
```

Some first studies: small Ns



- Authors report $m_1 = .28, m_2 = .03, \mathrm{SDM}_1 = 0.08, \mathrm{SDM}_2 = 0.05, N_1 = N_2 = 14$
- How do we compute the effect size ?

Computing Effect size: practice

- First, compute the standard deviation of the data from the SDM
  - get σ from SDM : $\sigma = \sqrt{14-1} \times \mathrm{SDM}$
  - Combine the σ to have one estimation across the groups
    - formula easy to recompute or find
  - $\sigma = \sqrt{14-1} \times \mathrm{SDM}$, $d = \frac{m_1 - m_2}{\sigma} = 1.05$

Computing Effect size: practice

- First, compute the standard deviation of the data from the SDM
  - get σ from SDM : $\sigma = \sqrt{14-1} \times \mathrm{SDM}$
  - Combine the σ to have one estimation across the groups
    - formula easy to recompute or find
  - $\sigma = \sqrt{14-1} \times \mathrm{SDM}$, $d = \frac{m_1 - m_2}{\sigma} = 1.05$
  - What is the percentage of variance explained ?

Computing Effect size: practice

- First, compute the standard deviation of the data from the SDM
  - get σ from SDM : $\sigma = \sqrt{14-1} \times \mathrm{SDM}$
  - Combine the σ to have one estimation across the groups
    - formula easy to recompute or find
  - $\sigma = \sqrt{14-1} \times \mathrm{SDM}$, $d = \frac{m_1 - m_2}{\sigma} = 1.05$
  - What is the percentage of variance explained ?
  - Write the estimated model: $Y = [1 \dots 1]^t [m_1 - m_2] + \mathrm{residual}$
  - Compute the total sum of square $Y^t Y$, then the proportion:

Computing Effect size: practice

- First, compute the standard deviation of the data from the SDM
  - get σ from SDM : $\sigma = \sqrt{14-1} \times \mathrm{SDM}$
  - Combine the σ to have one estimation across the groups
    - formula easy to recompute or find
  - $\sigma = \sqrt{14-1} \times \mathrm{SDM}$, $d = \frac{m_1 - m_2}{\sigma} = 1.05$
  - What is the percentage of variance explained ?
  - Write the estimated model: $Y = [1 \ldots 1]^t [m_1 - m_2] + \mathrm{residual}$
  - Compute the total sum of square $Y^t Y$, then the proportion:
  - $V_e = \frac{(n_1+n_2)(m_1-m_2)^2}{n_1 s_1^2 + n_2 s_2^2 + (n_1+n_2)(m_1-m_2)^2} > 40\%$

Power in practice:

- Harvard Medical School researchers show that broccoli reduces Parkinson patients symptoms X. 20 participants in each of 2 groups.
- The diffence in the measure of symptom A is significant with a t-test $p=0.02$
- You decide to replicate the study at the MNI. How many subjects should you test ?

What happens if . . . p is "significant" but study power is low ?

- Study in Button et al, 2013, more than half of the studies published have less than 30% power
- Low Positive Predictive Value P(H_A true | test significant)
- Inflated effect size
- Depends on the prior probability of H_A and H_0

PPV

- Define : $P_1 = P(H_1)$ and $P_0 = P(H_0)$

$$PPV = \frac{(1-\beta)P_1}{(1-\beta)P_1 + \alpha P_0}$$

- With $R = P_1/P_0$ and $W = 1-\beta$:

$$PPV = \frac{WR}{WR + \alpha}$$

- Wikipedia (ML): PPV = number-of-true-positive / (number-of-true-positive + number-of-false-positive)

Low PPV P(H_A is true | test signif.) with low power

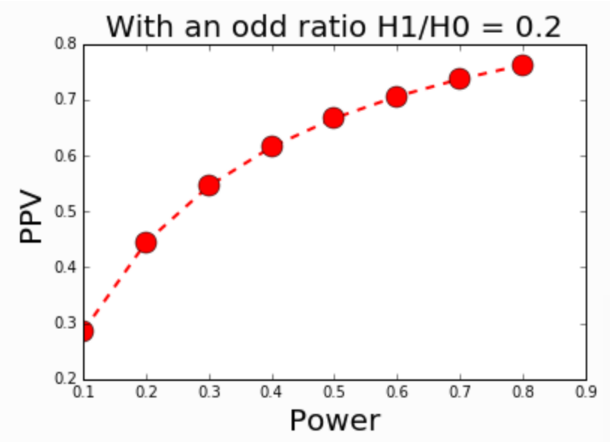


Figure 9: PPV = f(power), alph=0.05

Low PPV with high alpha : P(H_A is true | test signif.)

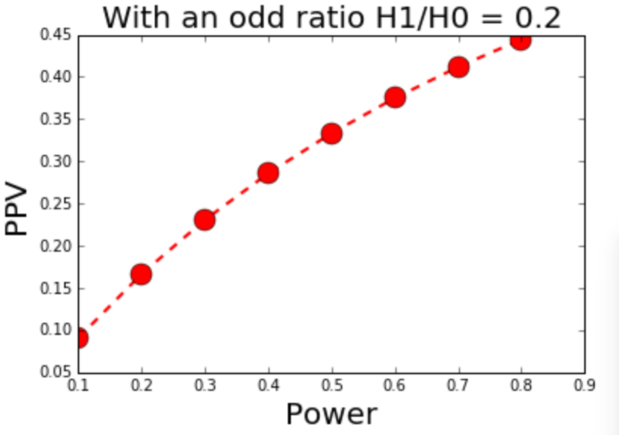


Figure 10: PPV = f(power), alph=0.2

Credits and References

- Zoltan Dienes: understanding psychology as a science
- Celia Greenwood for some slides
- Many others

Questions

- Questions

Acknowledgements

- Elizabeth Dupré (co-organizer)
- Jake Vogel (co-organizer)
- Andrew Doyle
- Félix-Antoine Fortin
- Peer Herholz
- Greg Kiar
- Liza Levitis
- Bratislav Misic
- Manjari Narayan
- Estefany Suarez
- Joe Viviano