Summary & Questions

Exam

We will have a multiple choice exam with 2 parts

Part 1: only 1 response option is correct (1 point if correct; 0 points when wrong or more than 1 option is ticked).

Part 2: at least one option is correct; ticking correct ones: 1 point (Hit), not ticking incorrect ones: 1 point (CR); otherwise 0 points

There will be 27 Qs covering all chapters

Exam

Bring a black or dark blue pen

Bring correction fluid or tape

No electronic devices and no materials are allowed

We will deliver a blank sheet for notes

Mobiles need to be off and cannot be on the table

You cannot take mobiles to wash room

You will leave one by one

Content

Quiz

The HIV test has a sensitivity of 0.9999 and a specificity of 0.9999. The incidence rate is 0.0001 in the normal population.
Now, a person is tested and the test is positive.

What is the probability that the person is infected by HIV?
It is 50%!

$$p(A|B) * p(B) = p(B|A) * p(A) \rightarrow p(A|B) = \frac{p(B|A) * p(A)}{p(B)}$$

$$p(HIV \mid T+) = \frac{p(T+ \mid HIV) * p(HIV)}{p(T+)}$$

$$p(HIV \mid T+) = \frac{p(T+ \mid HIV) * p(HIV)}{p(T+ \mid HIV) * p(HIV) + p(T+ \mid \sim HIV) * p(\sim HIV)}$$

$$p(HIV \mid T+) = \frac{0.9999 * 0.0001}{0.9999 * 0.0001 + (1-0.9999) * 0.9999} = 0.5$$

Quiz

What about a second test?

Say, the sample is 10.000 people and the first test is positive. Then, we have found by accident either the 1 person who indeed is infected (0.0001) or the test went wrong by accident (0.0001; false positive). Now, we do the test a second time on the very same person.

What is the chance to obtain a second HIT or second false positive? 0.9999 and 0.0001, respectively.

Hence, we are looking for

$$p(\text{HIV} \mid 2T+) = \frac{0.0001 * 0.9999^2}{(0.0001 * 0.9999^2 + 0.0001^2 * 0.9999)}$$

$$= 0.9999/(0.9999 + 0.0001)$$

$$= 0.9999$$

Surprisingly, actual numbers do not match. Why?

Odds Ratios & Simpsons Paradox

	Smokers	Non-smokers
Heart attack	7	1
No Heart attack	100	100

$$Ratios = \frac{\frac{7}{100}}{\frac{1}{100}} = \frac{7}{1} * \frac{100}{100} \cong 7$$

Odds Ratios & Simpsons Paradox

Question 1. What is the correct Odds Ratio for this example?

	Smokers	Non-smokers
Heart attack	7	1
No Heart attack	100	100

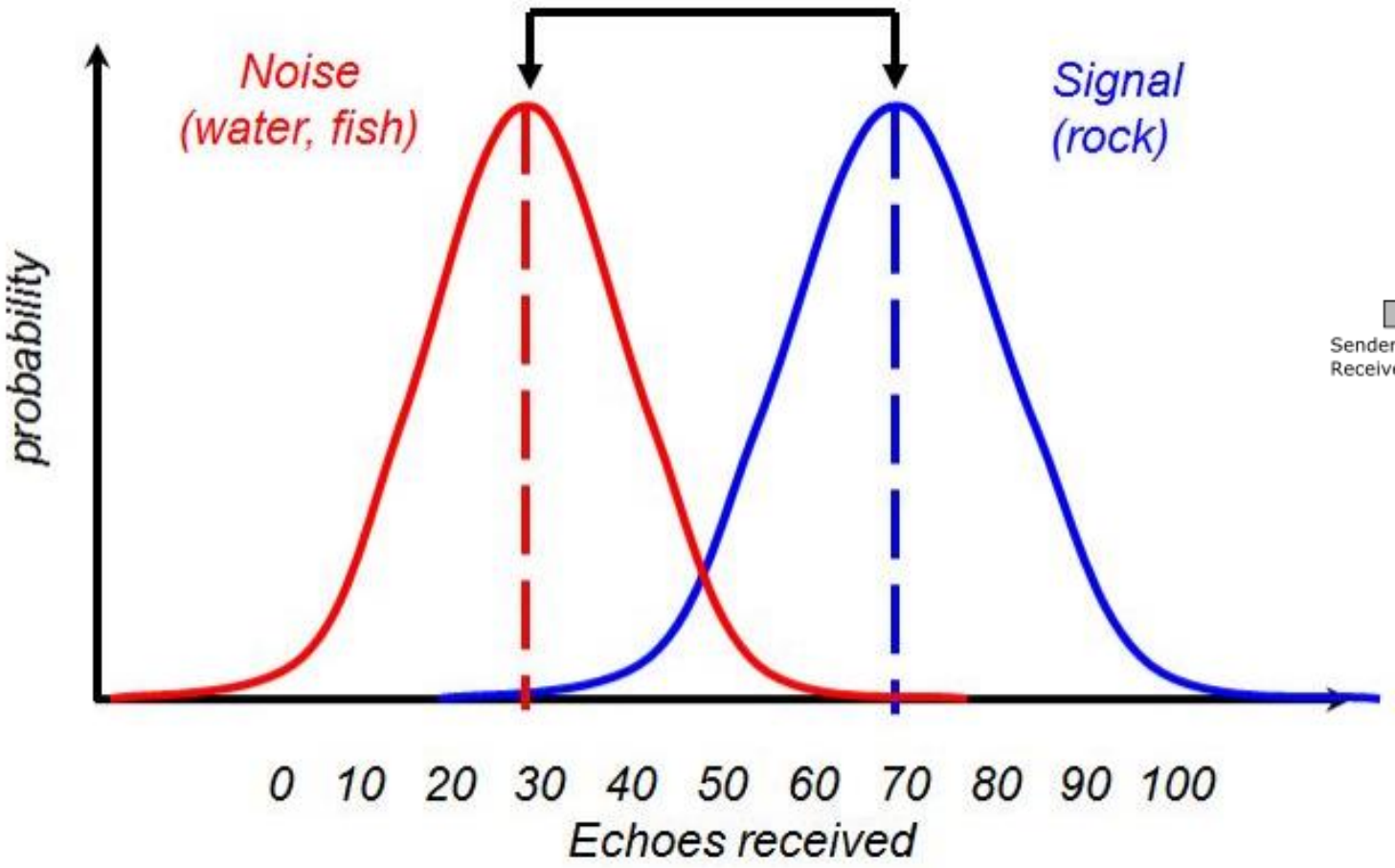
- 1
- 3
- 5
- 4
- 7

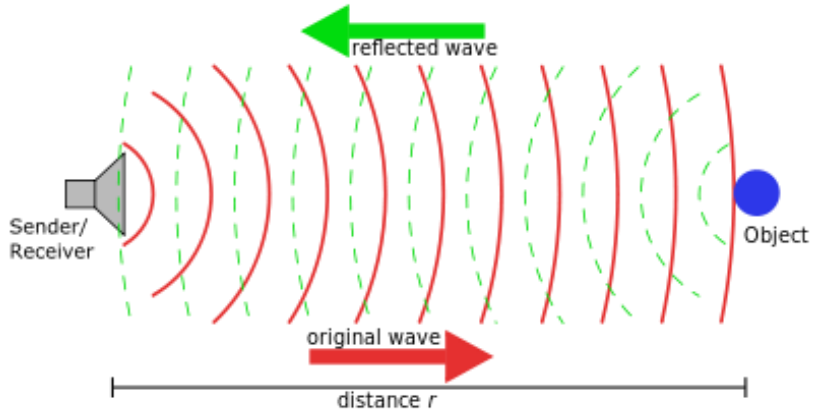
Content

1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

Conditional Probability: Noise (fish, water)

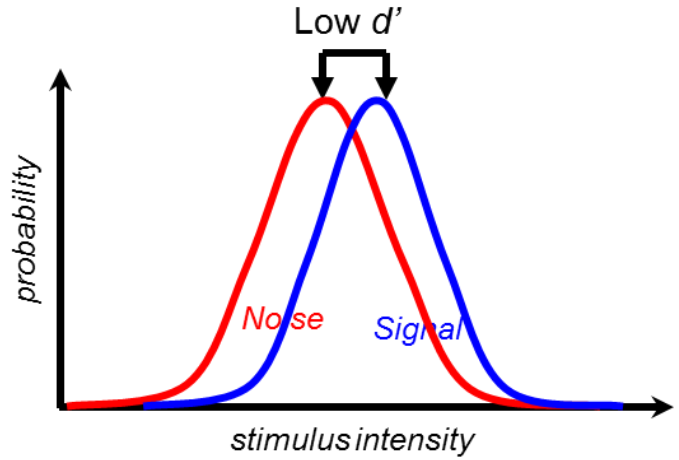
$$Discriminability\ (d') = \frac{|\mu_S - \mu_N|}{\sigma}$$



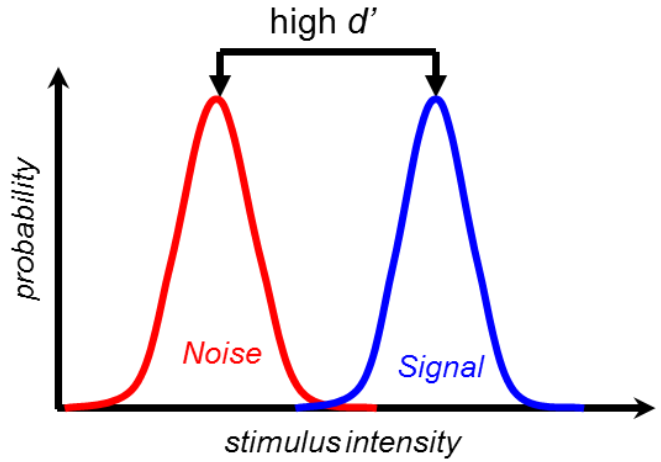


Interpreting d'

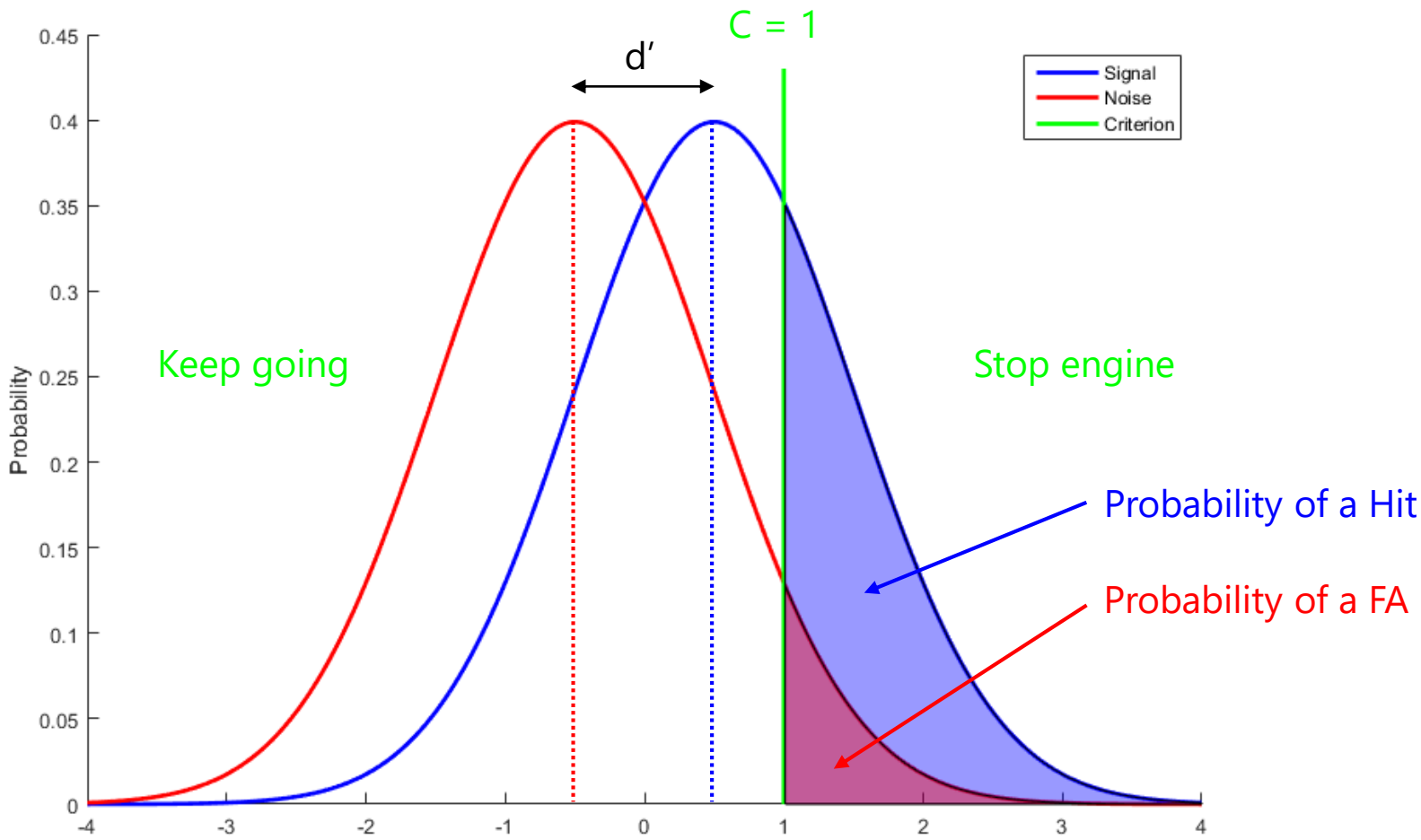
- If *d'* is low, discriminability is low.
  - The noise and stimulus distributions are highly overlapping.
  - *d'* = 0: chance level



- If *d'* is high, discriminability is high.
  - *d'* = 1: moderate performance
  - *d'* = 4.65: "optimal" (corresponds to hit rate=0.99, false alarm rate=0.01)



Discriminability & Criterion



Estimation of d'

- ***Theoretical*** $(d') = \frac{|\mu_S - \mu_N|}{\sigma}$
- d' can be estimated from the **experimentally** Hit and the False Alarm rate: ***d' = z*(Hit) - *z*(FA)**

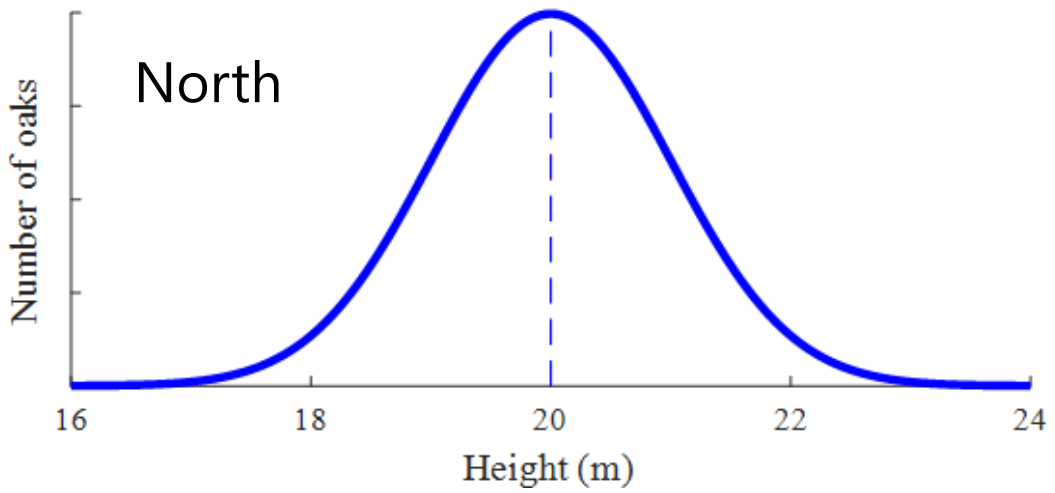
d’ is bias but not model free! It assumes:

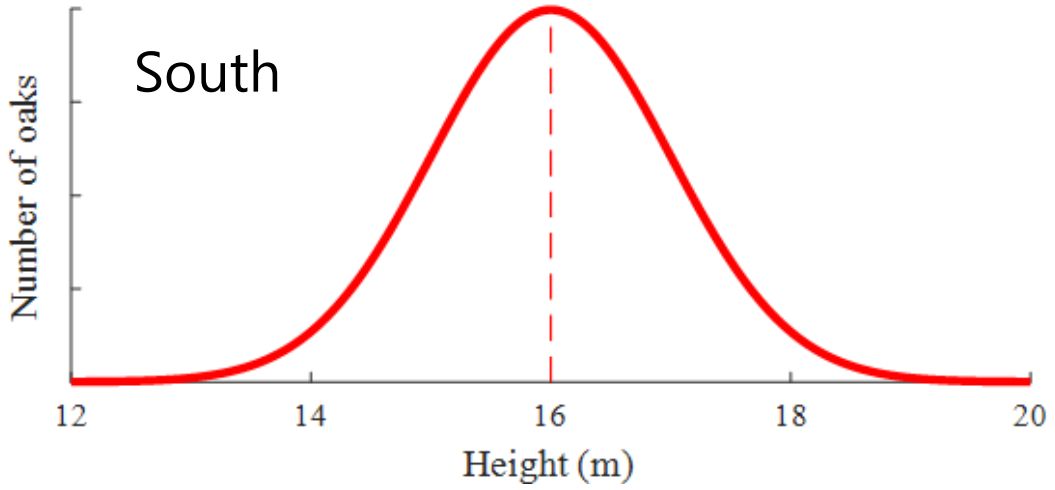
- Gaussian functions
- Equal Variance of $\sigma_n=\sigma_s$
- Criterion is constant

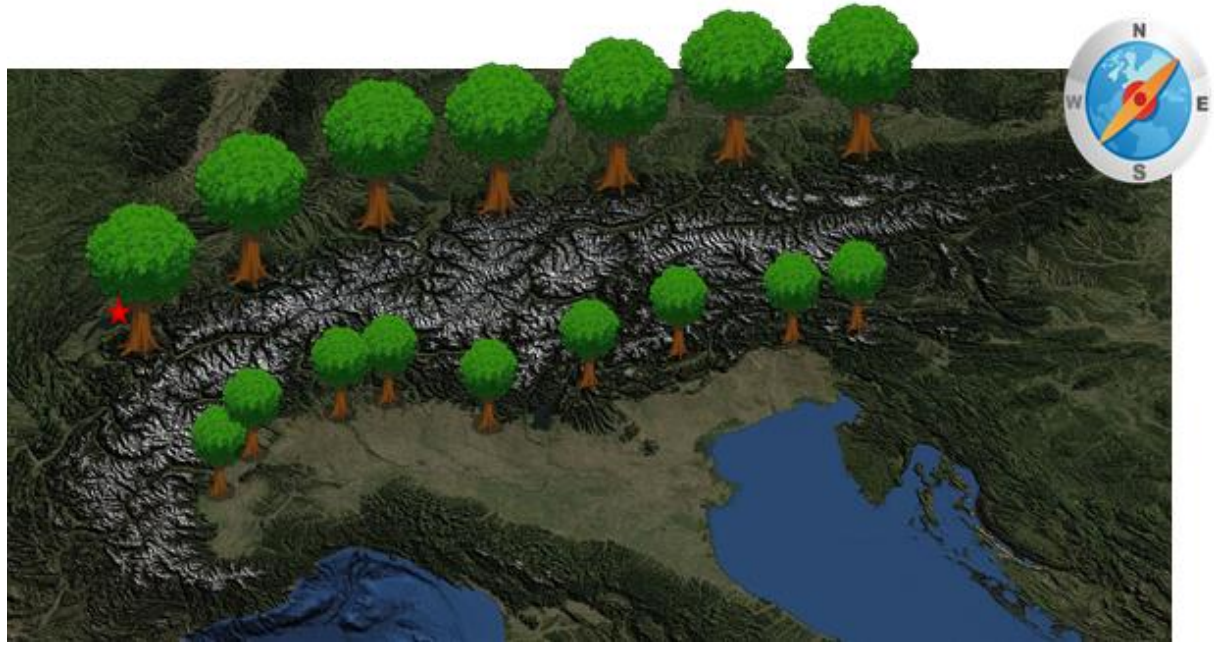
Content

1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

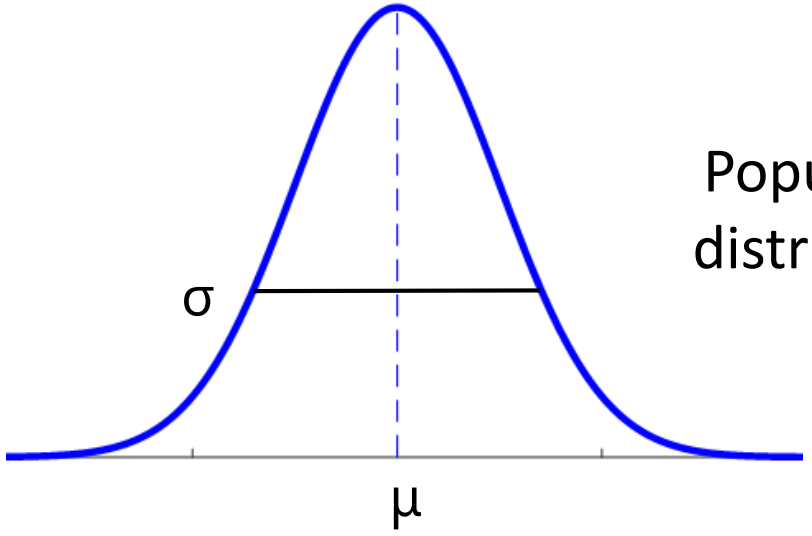
Step 1: Sampling & Sampling Error







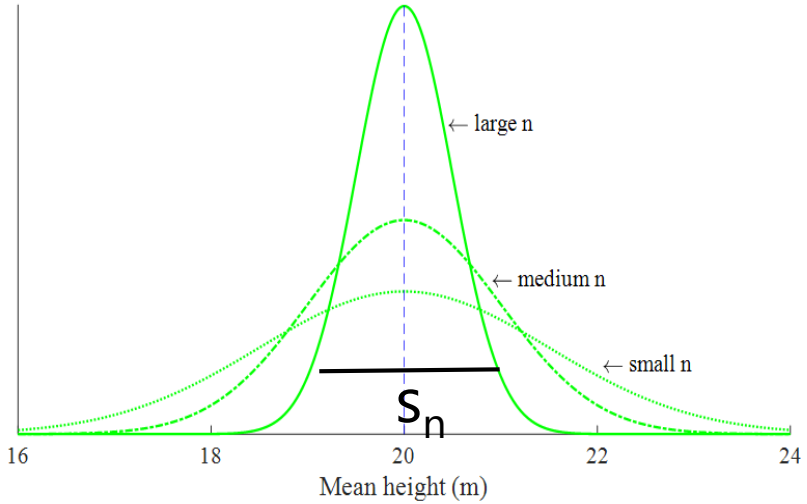
Many samples later…



Population distribution

The variances s_n of the sampling distributions become smaller, when n increases;

s_n is a good measure for the sampling error

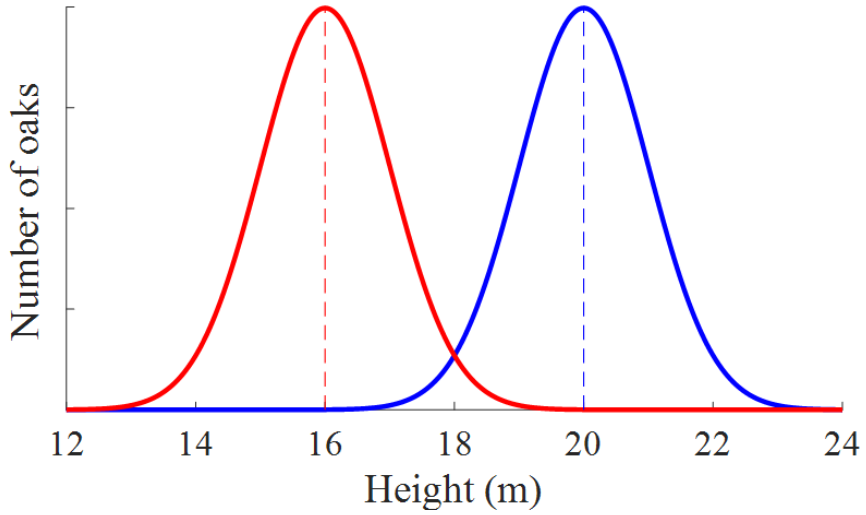


$$s_n = \frac{\sigma}{\sqrt{n}}$$

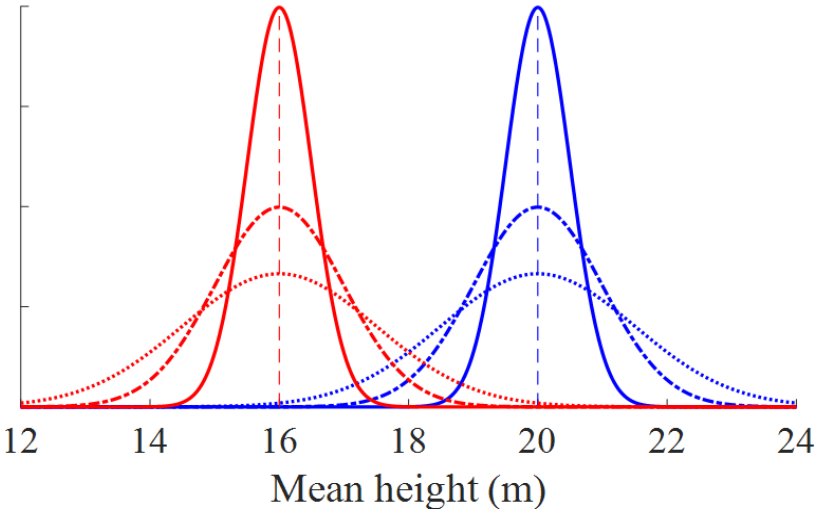
standard error of the mean

Step 2: Overlap

Population distributions



Sampling distributions Families (n)



Conclusion

$$t_n = \frac{|\bar{x}_{North} - \bar{x}_{South}|}{\hat{\sigma}} * \sqrt{n/2}$$

$$t_n = \hat{d}' * \sqrt{n/2}$$

The t-test confuses effect and sample size:

Partial information!

Cohen’s d

d= 0.2 small effect

d= 0.5 medium effect

d= 0.8 large effect

https://de.wikipedia.org/wiki/Effektst%C3%A4rke

Implications I

Implication Ib: For each d'≠0, the t-test is significant for a sufficiently large n_1 and non-significant for a smaller n_2

Is the effect significant? Sounds like a paradox but it is not.

$$t_n = \frac{|\bar{x}_{North} - \bar{x}_{South}|}{\hat{\sigma}} * \sqrt{n} = \hat{d}' * \sqrt{n/2}$$

Significance: $\hat{d}' * \sqrt{n} > a, \; e.g.\; 1.96 \; (depending \; on \; n)$

Implications III

Implication IIIb: Problems:
Missing power or Null result true?
One can never proof a Null result
Are there Null results at all?
Null results are easy to create

For these reasons Null results are hard to publish
This leads to precedence effects & False Positives survival

Null results are the Archilles Heel of the Frequentist approach

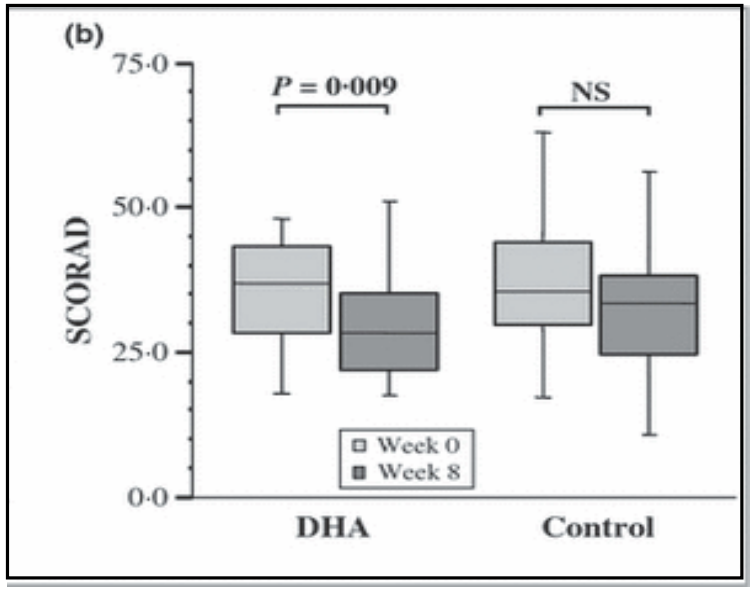
Where does the noise come from?

There are two terms:

$$x_i = \mu + v_i + e_i$$

where μ is the overall grand population mean across subjects, v_i is the deviation from the grand mean for subject i , and e_i is the noise for a particular trial for a particular subject.

Truth: if v is not 0.0, you cannot make conclusions on the individual level



Objective

To assess the risk of stroke and myocardial infarction (MI) after herpes zoster in a US community population of older adults.

Patients and Methods

We performed a community cohort study (January 1, 1986, to October 1, 2011) comparing the risk of stroke and MI in 4862 adult residents of Olmsted County, Minnesota, 50 years and older with and without herpes zoster and 19,433 sex- and age-matched individuals with no history of herpes zoster. Odds ratios are presented for MI and stroke at 3, 6, 12, and 36 months after index herpes zoster plus hazard ratios for long-term risk (up to 28.6 years).

Results

Individuals with herpes zoster had more risk or confounding factors for MI and stroke, suggesting that they had worse health status overall. When controlling for the multiple risk factors, those with herpes zoster were at increased risk for stroke at 3 months after herpes zoster compared with those without a history of herpes zoster (odds ratio, 1.53; 95% CI, 1.10-2.33; *P*=.04). The association between herpes zoster and MI at 3 months was not robust across analytic methods. Herpes zoster was not associated with an increased risk of stroke or MI at any point beyond 3 months. The risk increased by 53% from 0.4 to 0.7%.

Conclusions

Herpes zoster was associated with only a short-term increased risk of stroke, which may be preventable with the prevention of herpes zoster.

Question: what type(s) of bad statistical practice is present in the above study:

- HARKing
- Multiple testing
- File drawer
- No Preregistration, if this is needed for this type of study
- OR are inconclusive since incidence rates are missing
- Confusion of 95%-CI with effect size
- None of them

Content

1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

Types of t-tests

- Two sample t-test

$$t_n = \frac{\bar{x}_{North} - \bar{x}_{South}}{\hat{\sigma}/\sqrt{n/2}}$$

- One sample t-test

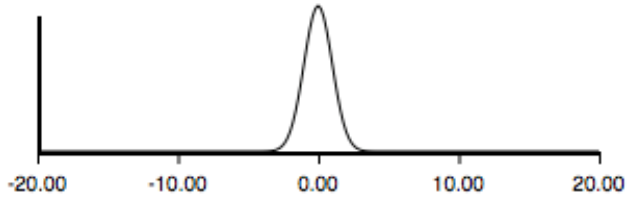
$$t_n = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{n}}$$

- Repeated measures t-test: a one sample t-test with μ = 0

Mean=0
SD: 5
Skewness: None
Sample size: 5

Population 1

Population 2



Mean=0
SD: 1
Skewness: None
Sample size: 25

Number of experiments:
10,000
Run simulation
Reset

	Significance criterion	
	.05	**.01**
Number significant:	3832	2432
Number not significant:	6168	7568
Type I error rate:	0.383	0.243

General Instructions

This demonstration allows you to explore the effects of violating the assumptions of normality and homogeneity of variance. When the simulation starts, you see the distributions of two populations. By default, they are both normally distributed and have means of 0 and standard deviations of 1. The default sample size for the simulations is 5 per group. If you press the "Run simulation" button, 2,000 simulated experiments are conducted. You can adjust the number of simulations from 2,000 to 10,000. A *t* test is computed for each experiment, and the number of tests that were significant and not significant for the .05 and .01 criteria are presented along with the Type I error rate (the proportion of significant tests) for each criterion.

Since the null hypothesis is true and all

Some words about Metrics

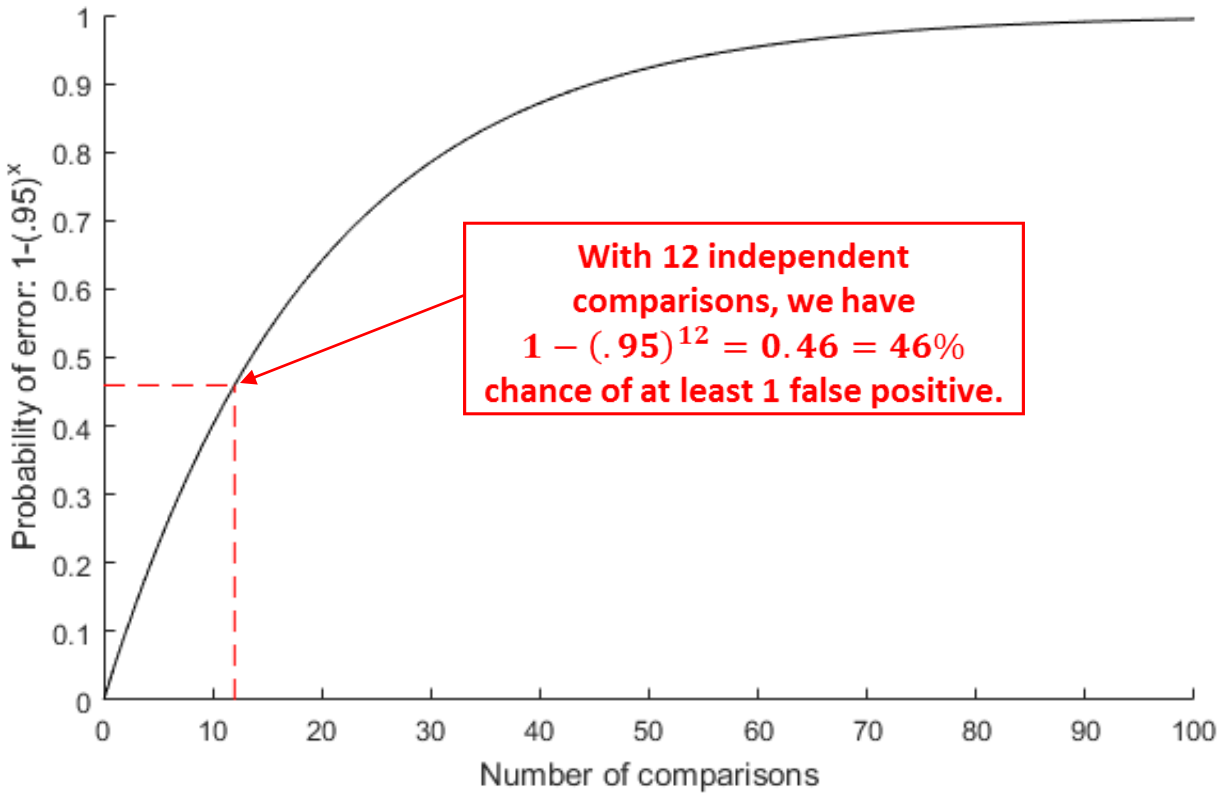
Metrics: Scaling is

- Nominal: no order (countries, genes)
- Ordinal: rank order (military ranks)
- Interval: only additive order (no zero), ratios make no sense (Calendar Dates such as 26th of Oct)
- Metric: additive & multiplicative order (with 0), ratios makes sense, *real numbers* (temperature)

Content

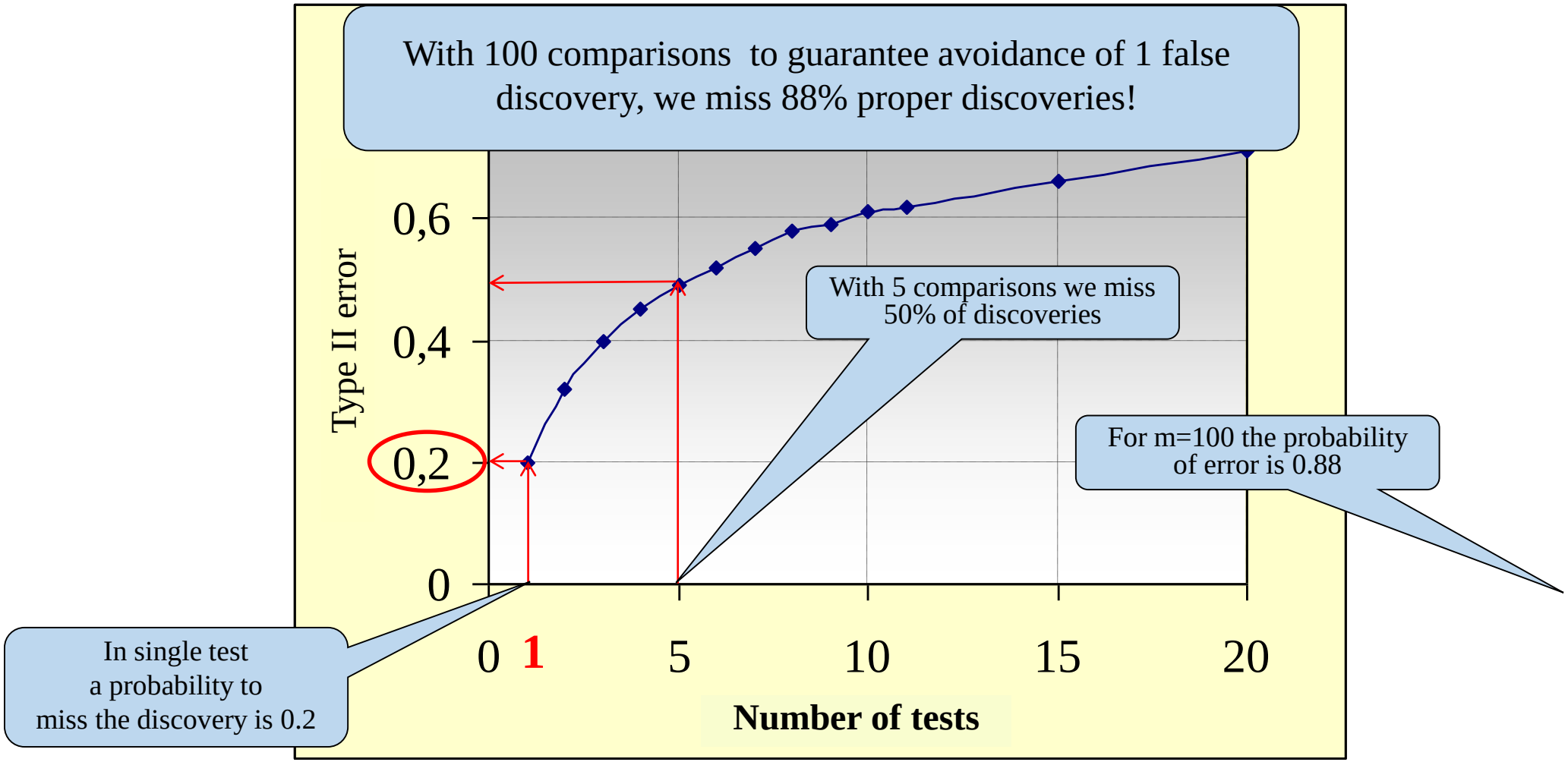
1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

Multiple comparisons



Dependence of Type II error on number of tests using the Bonferroni correction

Probability to miss gene with OR=2.7 with sample sizes 100 (case) and 100 (control)



FISH example

What if they are all correlated and you got a „bad" sample?
Tests are not (necessarily) independent

Content

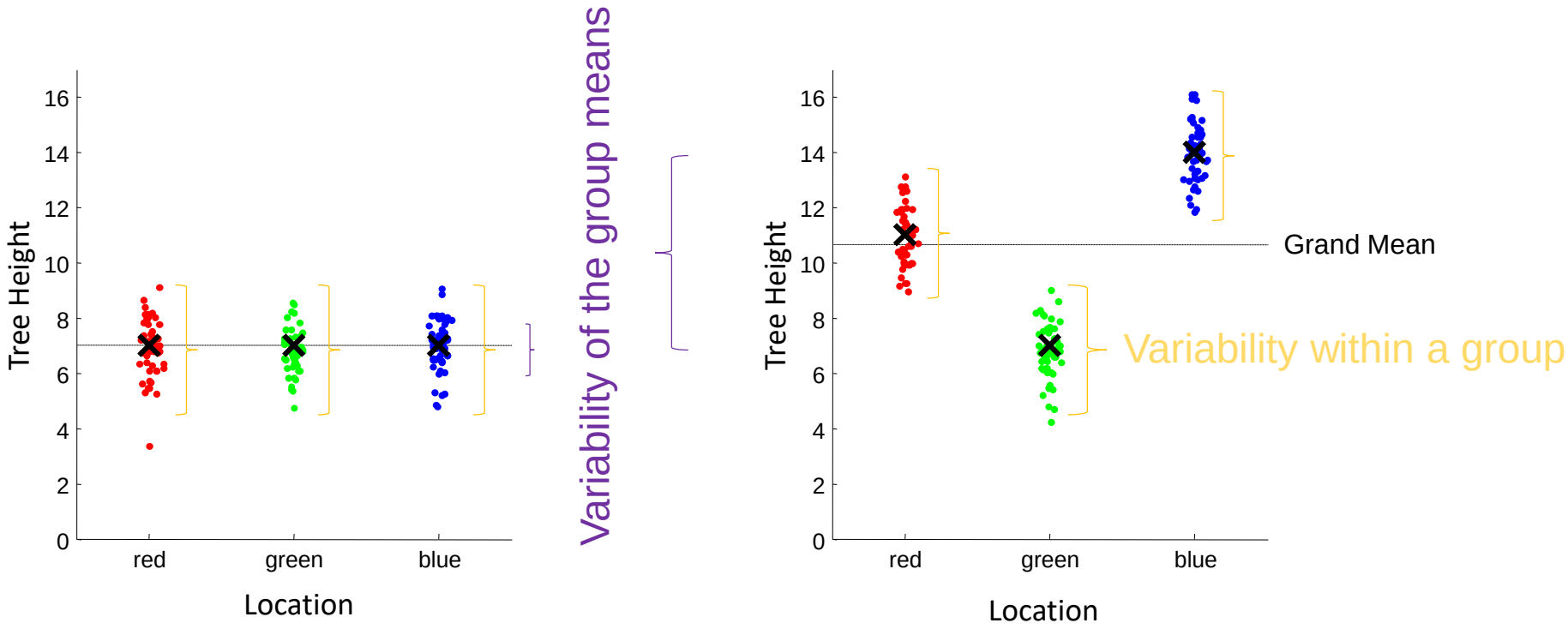
1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

Strategy

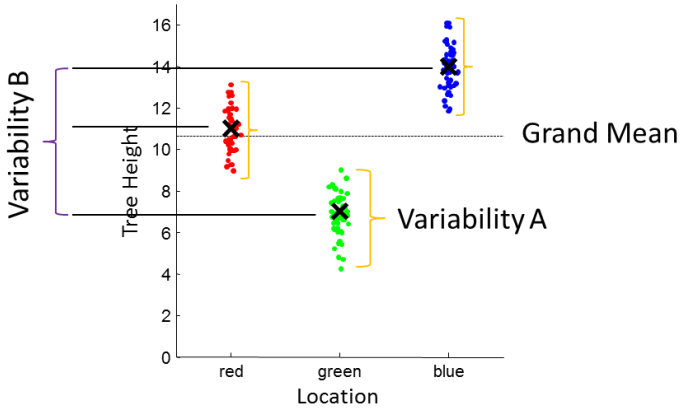
$$\frac{\text{Variability of the group means}}{\text{Variability within a group}}$$

$$\frac{\text{Small number}}{\text{Large number}}$$

$$\frac{\text{Large number}}{\text{Small number}}$$



Question 5. In the graph below, what are true statements about “Variability A”?

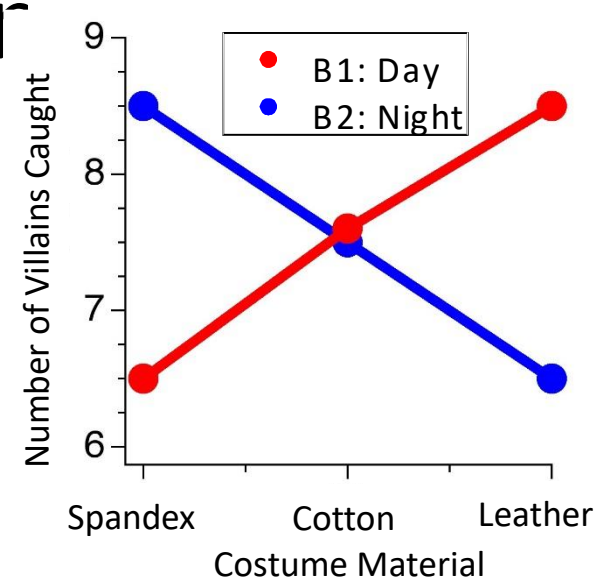


- It is the Variability within groups
- It is the Variability between groups
- It is the Variability within treatments
- It is the same as Variability B

The Data

	Spandex	Cotton	Leather
 Day	18	10	3
	10	8	5
	16	12	1
	12	6	7
	14	14	9
 Night	5	6	15
	7	14	13
	3	10	17
	9	8	11
	1	12	19

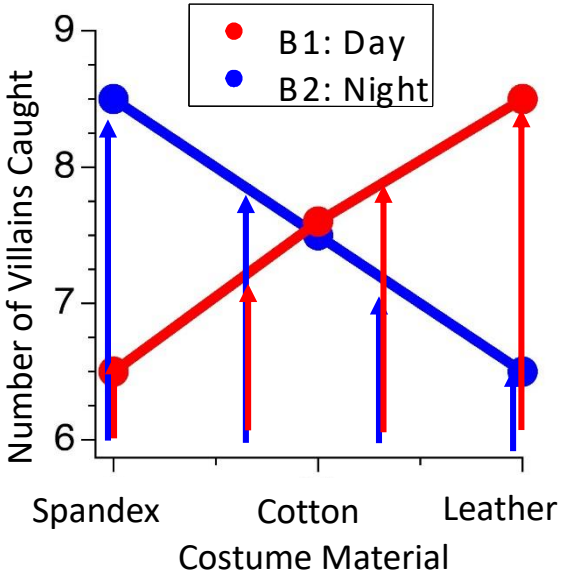
Sor



✓ **Interaction!**

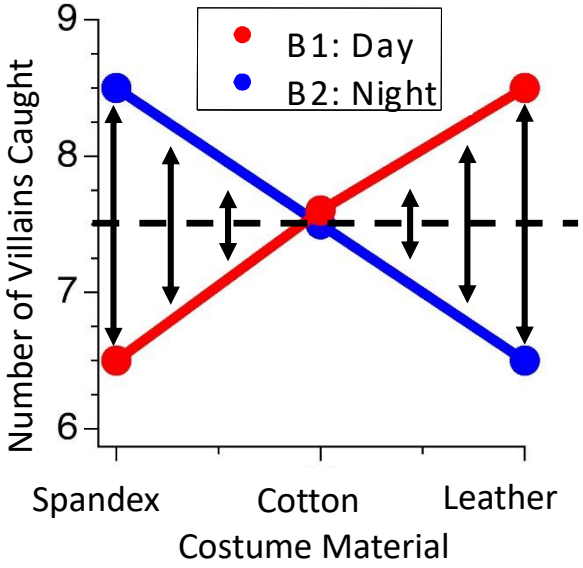
✓ The levels of one factor influence the effect of the other factor.

✓E.g. Time influences whether costume material leads to less or more villains caught.



✓ No main effect of costume.

✓ It just happens that, in this example, costume material has exactly <u>opposite</u> effects at day and at night -> **Interaction!**



✓ No main effect of time.

✓ It just happens that, in this example, time has exactly opposite effects across costume types -> **Interaction!**

- A large variety of people are not able to copy statistical values (t, p, df) correctly from the statistics program to the publication.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	216.425	2	108.212	.449	.649
Within Groups	2891.863	12	240.989		
Total	3108.288	14			

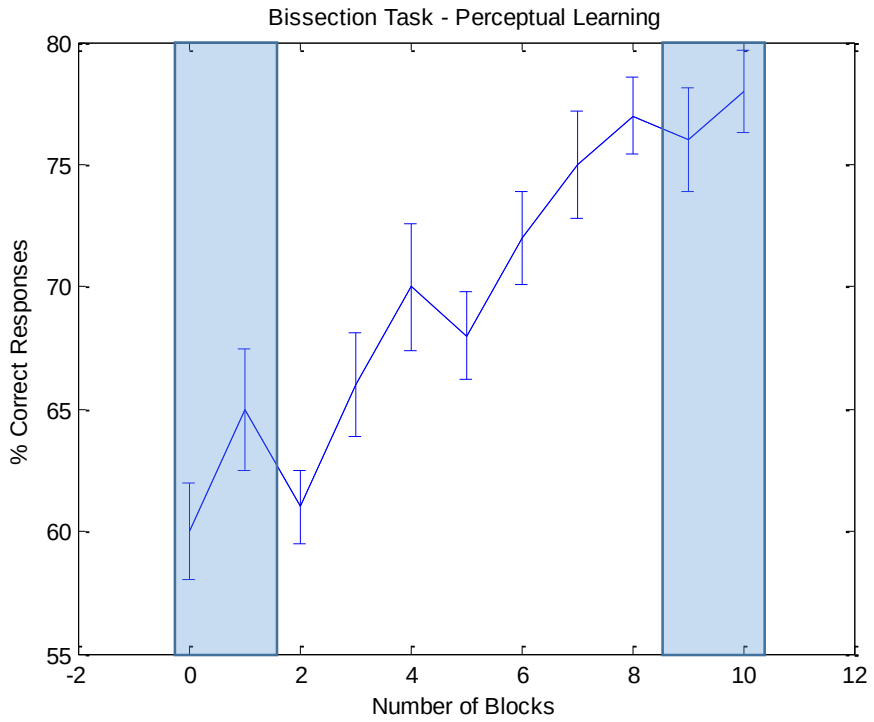
Content

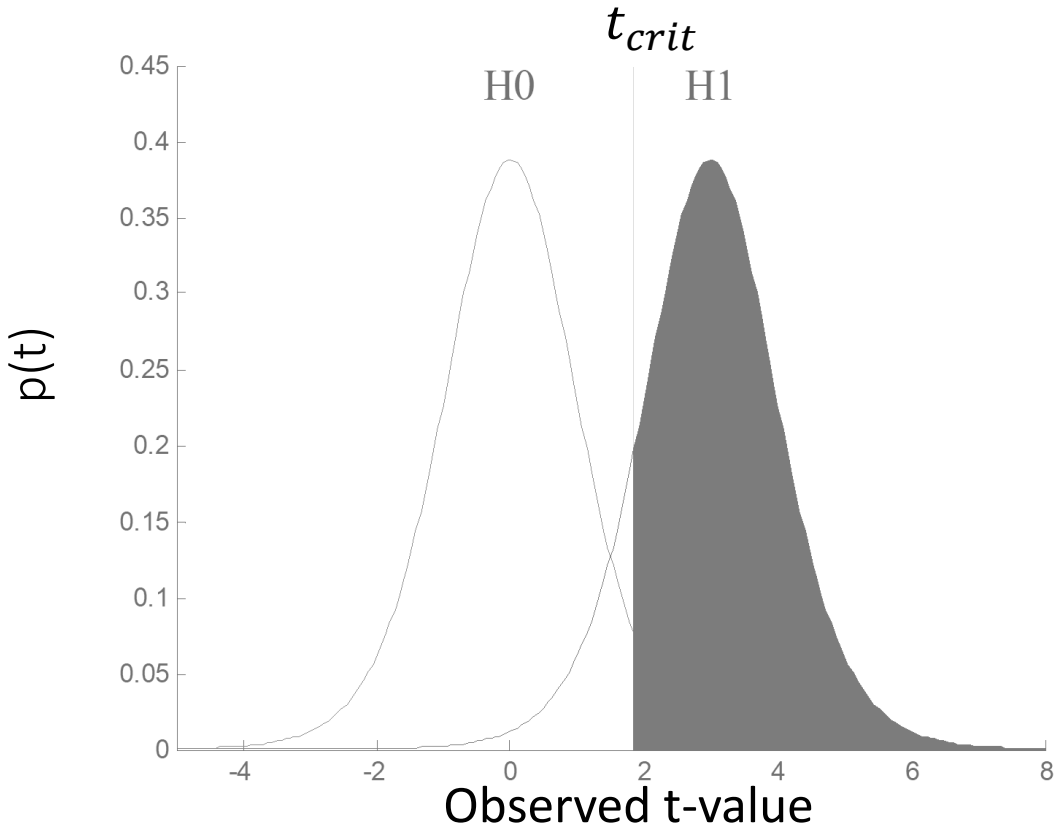
1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

Pre-post testing

Bisection Task

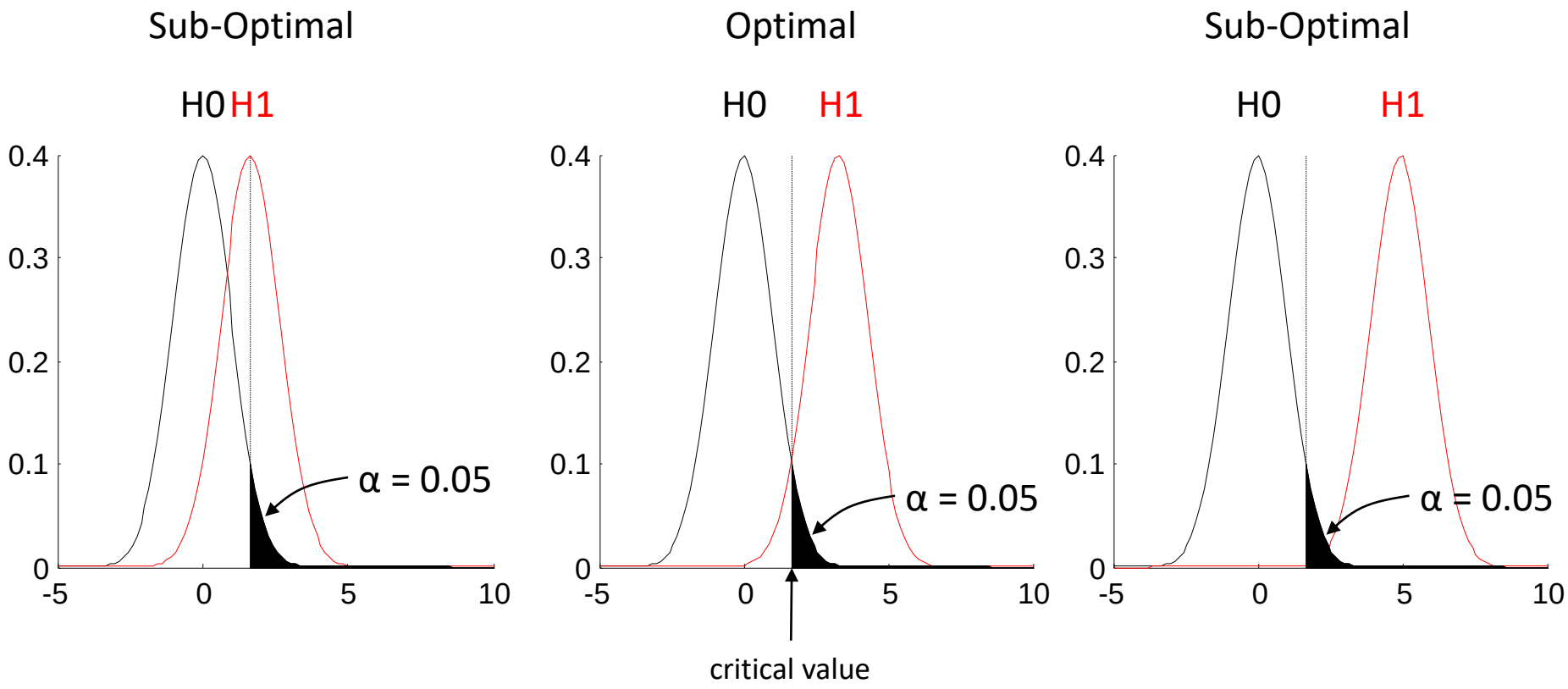
| |





	H0 True	H1 True
We Say H0 True	1-α	B
We Say H1 True	α	Power (1-β)

When Does NHST Make Optimal Decisions?



Null-hypothesis significance testing only makes optimal decisions when the intersection between the H0 and the H1 distributions overlaps with the critical value.

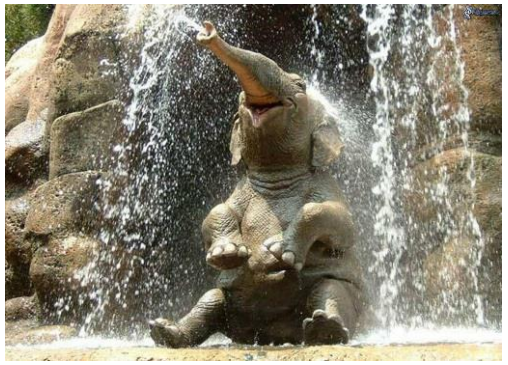
Content

1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

Covariation & Scaling

$$Cov(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{n-1}$$

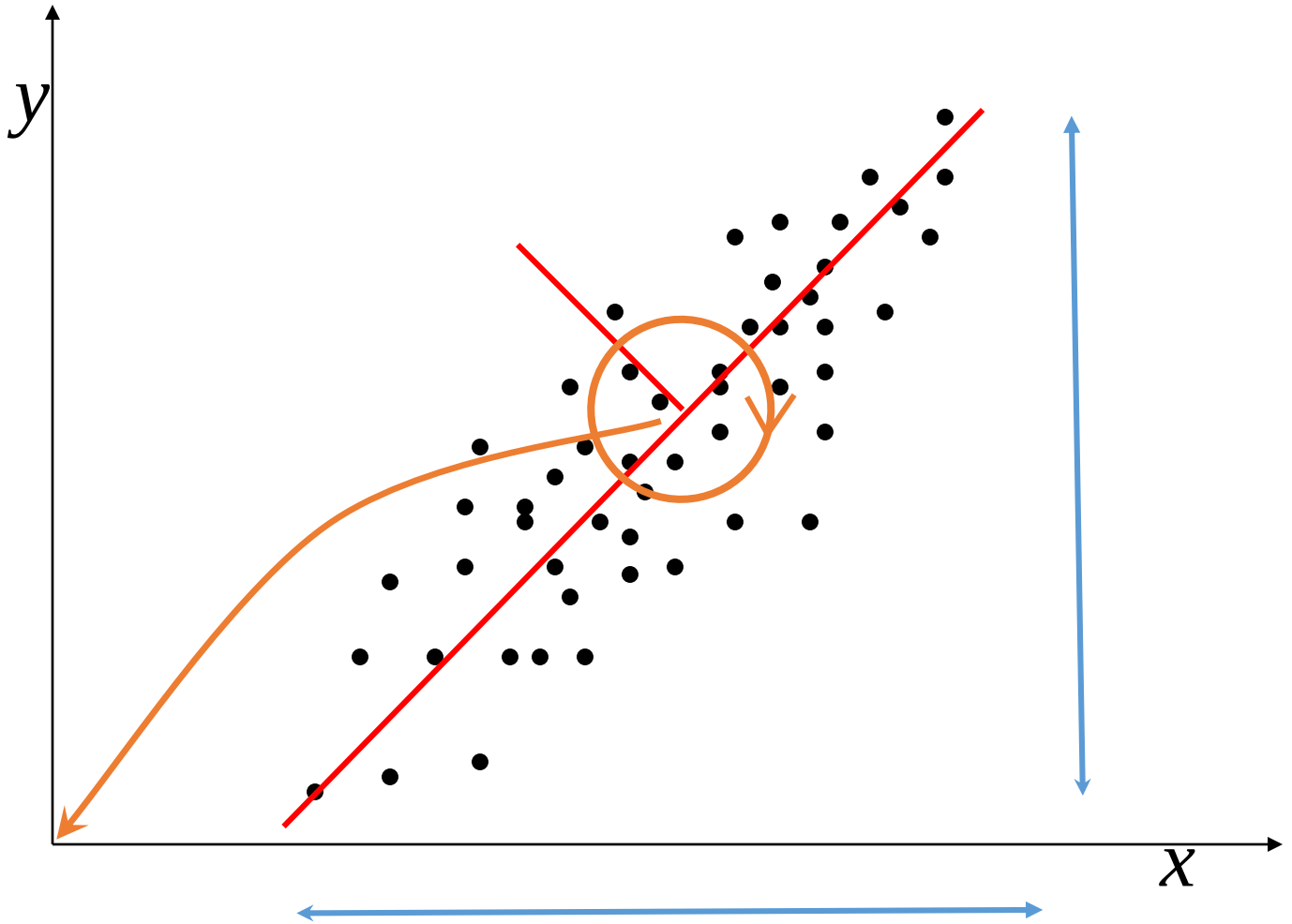
$$R = \frac{Cov(x,y)}{s_x \cdot s_y}$$

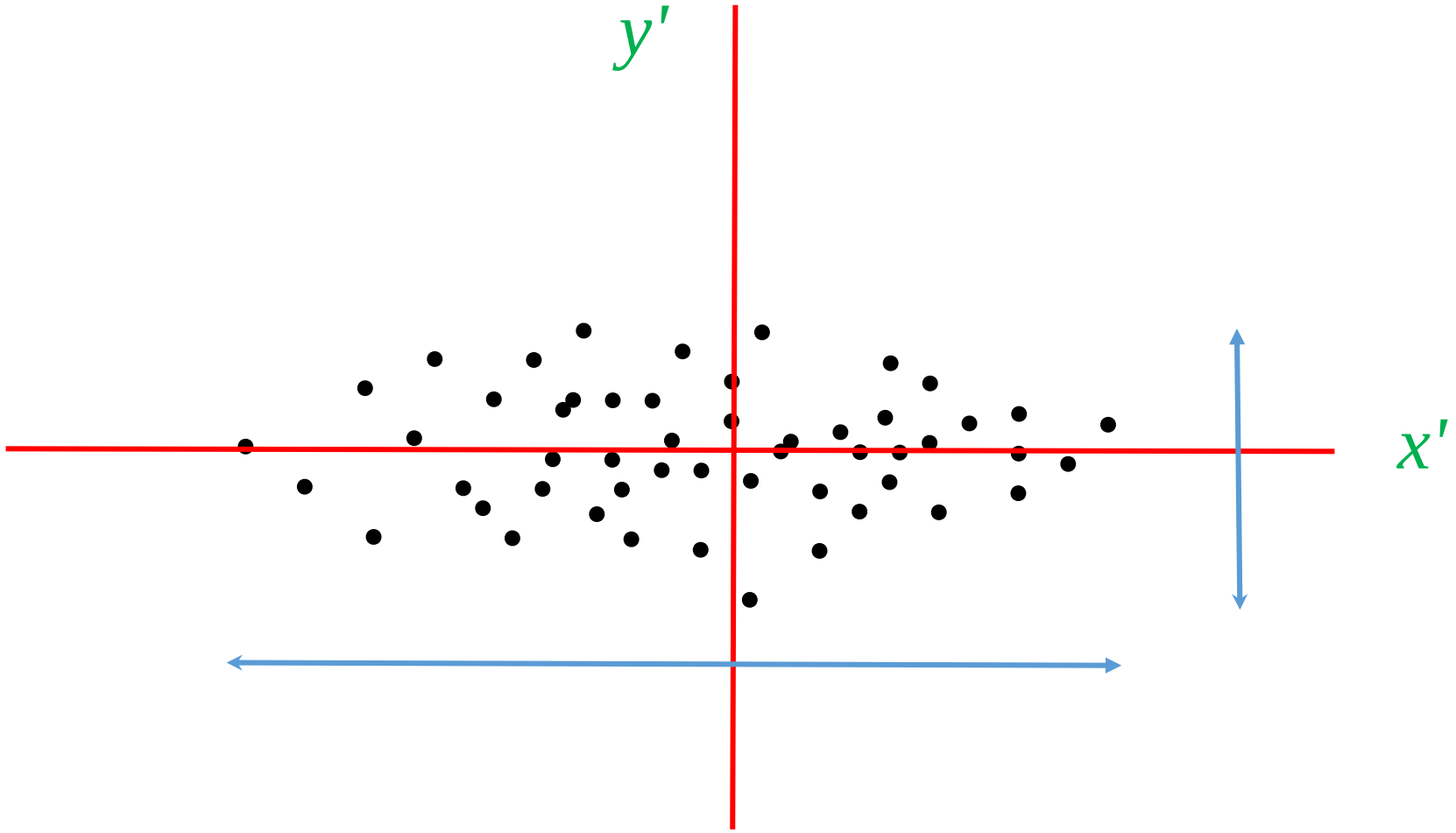


cm & m
Elephant and ant

Cohen's Rules Of Thumb For Effect Size

Effect size	Correlation coefficient	Difference between means
“Small effect”	r = 0.1	d = 0.2 standard deviations
“Medium effect”	r = 0.3	d = 0.5 standard deviations
“Large effect”	r = 0.5	d = 0.8 standard deviations


y
x


y′
x′

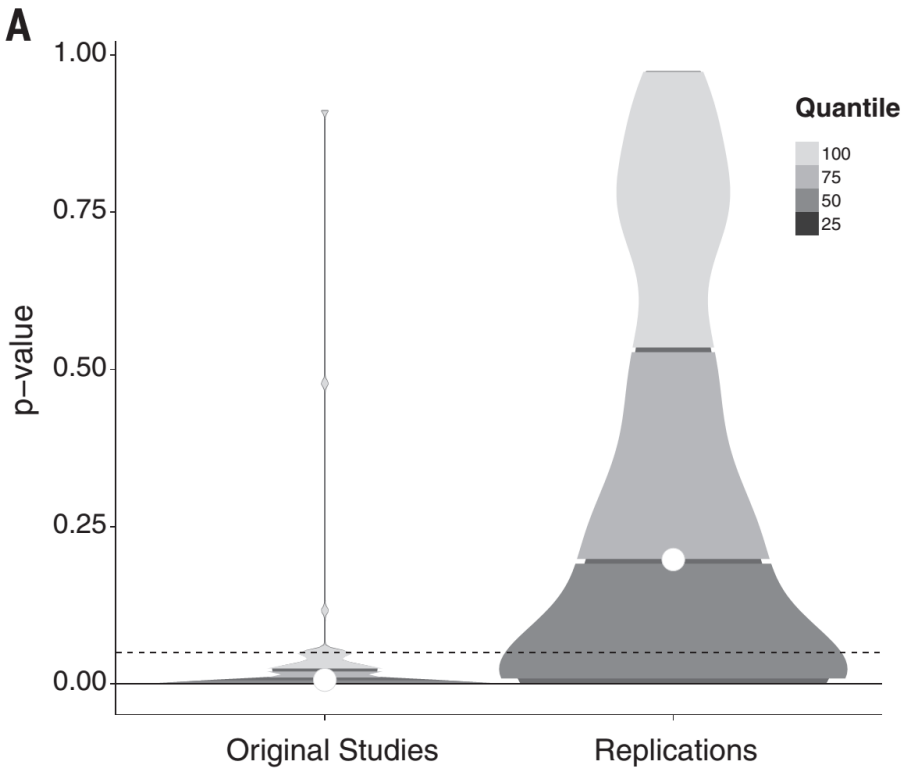
Content

1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science



ANALYSIS

Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

Problems with low-power studies:

1. The chance of discovering effects that are genuinely true is low.
2. The lower the power the lower the probability that an observed effect that passes p<.05 actually reflects a true effect.
3. Even when an underpowered study discovers a true effect, it is likely that the estimated effect size will be exaggerated.

- A random sample of 50 neuroscience meta-analyses showed that almost 50% of studies had an average power lower than 20%.
- The average statistical power of neuroscience studies is probably no more than between ~8% and ~31%, on the basis of evidence from diverse subfields within neuro-science.

Meta-analysis

- Things can get complicated quite quickly
- If you have some between-subject designs and some within-subject designs, you need to be sure you use equivalent effect size measures
  - For the within-subject effect size, compensate for the correlation that is used to produce the *t* value
  - This gives a *d* that is "equivalent" to a between-subject's design
- Similar issues for ANCOVA

Cohen's *d*

$$d = \frac{t}{\sqrt{n}}\sqrt{2(1-r)}$$

Hedge's *g*

$$g = \left(1 - \frac{3}{4(n-1)-1}\right)d$$

Content

1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

Effect size

- Use meta-analytic techniques to pool the effect sizes across all ten experiments (Hedges & Olkin, 1985)
- Pooled effect size
  - g^*=0.1855

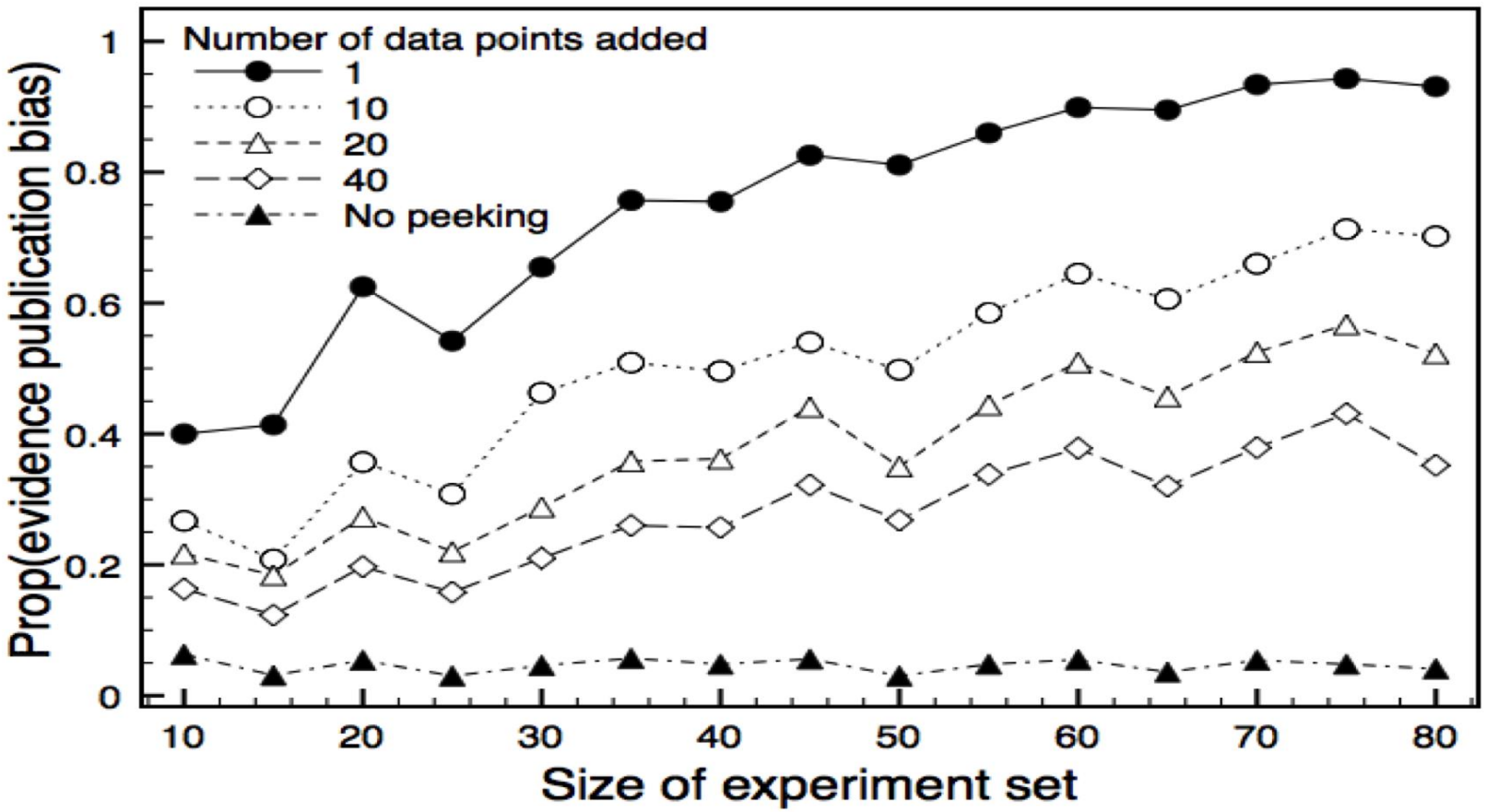
$$g^* = \frac{\sum_{i=1}^{M} w_i g_i}{\sum_{i=1}^{M} w_i}$$

  - w_i is the inverse variance of the effect size estimate

	Sample size	Effect size (g)
Exp. 1	100	0.249
Exp. 2	150	0.194
Exp. 3	97	0.248
Exp. 4	99	0.202
Exp. 5	100	0.221
Exp. 6 Negative	150	0.146
Exp. 6 Erotic	150	0.144
Exp. 7	200	0.092
Exp. 8	100	0.191
Exp. 9	50	0.412

Simulated Optional Stopping bias

- The test for publication bias works properly
- But it is conservative
- When the test indicates bias, it is almost always correct



Good News

- Many people get very concerned when their experimental finding is not replicated by someone else
- Lots of accusations about incompetence and suppositions about who is wrong
- But "failure" to replicate is *expected* when decisions are made with hypothesis testing
  - At a rate dependent on the experimental power
- Statisticians have an *obligation* to be **wrong** the specified proportion of the time

Content

1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

TES Analysis For PSCI

2010: 12 out of 14 articles have $P_{TES} \leq .1$

Authors	Short title	P_{TES}
Balcetis & Dunning	Wishful Seeing	.076
Bowles & Gelfand	Status and Workplace Deviance	.057
Damisch, Stoberock & Mussweiler	How Superstition Improves Performance	.057
de Hevia & Spelke	Number-Spacing Mapping in Human Infants	.070
Ersner-Hershfield, Galinsky, Kray & King	Counterfactual Reflection	.073
Gao, McCarthy & Scholl	The Wolfpack Effect	.115
Lammers, Stapel & Galinsky	Power and Hypocrisy	.024
Li, Wei & Soman	Physical Enclosure and Psychological Closure	.079
Maddux, Yang, Falk, Adam, Adair, Endo, Carmon & Heine	Culture and the Endowment Effect	.014
McGraw & Warren	Benign Violations	.081
Sackett, Meyvis, Nelson, Converse & Sackett	When Time Flies	.033
Savani, Markus, Naidu, Kumar & Berlia	What Counts as a Choice?	.058
Senay, Albarracín & Noguchi	Interrogative Self-Talk and Intention	.090
West, Anderson, Bedwell & Pratt	Red Diffuse Light Suppresses Fear Prioritization	.157

= Premium Content

THE CHRON® of Higher Education

Monday, November 12, 2012

HOME | NEWS | OPINION & IDEAS | FACTS & FIGURES | BLOGS | JOBS

Home > Blogs > Percolator

PERCOLATOR

Research that matters.

Previous ← MIT Report Says Clean Energy Winner Was 'Misleading' but Didn't Break Rules

Next Is Electricity the New Smart Drug? →

Is Psychology About to Come Undone?

April 17, 2012, 5:09 pm

By Tom Bartlett

Brian Nosek

If you're a psychologist, the news has to make you a little nervous—particularly if you're a psychologist who published an article in 2008 in any of these three journals: *Psychological Science*, the *Journal of Personality and Social Psychology*, or the *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

RENEW

man Origins | Living World | Environment | ronment

thinkocket Science ce

in psychology
e field

Diederek Stapel: 30+ fraudulent results over 20 years. Zero published failures to replicate!

Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

Raise standards for preclinical cancer research

C. **Glenn Begley** and **Lee M. Ellis** propose how methods, publications and incentives must change if patients are to benefit.

- 55 Landmark studies replicated by Amgen corporation; 11% were found to be reproducible.
- Similar results reported by Merck.

"It drives people in industry crazy. Why are we seeing a collapse of the pharma and biotech industries? One possibility is that academia is not providing accurate findings." (CNBC, 2012).

Optional stopping

- The real problem is not with adding subjects, but with stopping when you like the outcome (e.g., p<.05)
  - If you observe p=0.03 and add more subjects, you might get p>.05
- This means that the interpretation of your *p* value depends on what you **would do** if you observed *p*<.05 or *p*>.05
- If you **would have** added subjects when get *p*>.05, then even if you actually get *p*<.05, then you have an experiment with an inflated Type I error rate
- If you do not know what you would have done, then you do not know the Type I error rate of your hypothesis test

Data peeking

- Subjects are scarce, so researchers sometimes “peek” at the data to see if the experiment is working
- If the knowledge from such a peek changes their sample, then there is loss of Type I error
- Consider the following experiment plan
  - Data is gathered from $n_1=n_2=10$ subjects. A *p* value is computed
    - If $p<0.2$, additional data is gathered to produce $n_1=n_2=50$, and the results are reported
    - If $p>0.2$, the experiment is aborted and not reported
  - Among the *reported* experiments, the Type I error rate is 13%
- The effect is bigger when the peek occurs at what is closer to the final value
  - Data is gathered from $n_1=n_2=10$ subjects. A *p* value is computed
    - If $p<0.2$, additional data is gathered to produce $n_1=n_2=20$, and the results are reported
    - If $p>0.2$, the experiment is aborted and not reported
  - Among the *reported* experiments, the Type I error rate is 20%

Interpretation

- The number of times Bem (2011) rejected the H_0 is inconsistent with the size of the reported effect and the properties of the experiments
  - 1) Perhaps there were additional experiments that failed to reject H_0 but were not reported
  - 2) Perhaps the experiments were run incorrectly in a way that rejected the H_0 too frequently
  - 3) Perhaps the experiments were run incorrectly in a way that underestimated the true magnitude of the effect size
- The findings in Bem (2011) are *too good to be true*
  - Non-scientific set of findings
  - Anecdotal
- Note, the effect may be true (or not), but the studies in Bem (2011) give no guidance

Content

1. Basic Probability Theory
2. Signal Detection Theory (SDT)
3. SDT and Statistics I
4. SDT and Statistics II
5. Statistics in a nutshell
6. Multiple Testing
7. ANOVA
8. Experimental Design & Statistics
9. Correlations & PCA
10. Meta-Statistics: Basics
11. Meta-Statistics: Too good to be true
12. Meta-Statistics: How big a problem is publication bias?
13. Meta-Statistics: What do we do now?
14. Summary

Good Practice

- Specific research question
- Determine experimental and output variables
- Choose statistical test beforehand
- Choose alpha (0.05) and beta level (0.2)
- Choose effect size and compute sample size n
- Run the experiment without any changes
- Compute statistics
- No post-hoc hypothesis

The  Practice

- Measure as many variable as possible
- See what happens: interesting hot spots?
- Create Post-hoc hypothesis
- Remove “outliers” depending on post-hoc hypothesis
- Try various statistical tests

…….until significant

The  Practice

- Measure as many variable as possible
- See what happens: interesting hot spots?
- Create Post-hoc hypothesis
- Remove “outliers”
- Try various statistical tests

…….until significant

Actually this is all fine, if you treat your experiment as an exploratory experiment- and repeat…….

Good luck…….