This is only a draft! (some rearrangements will occur, references must be added)

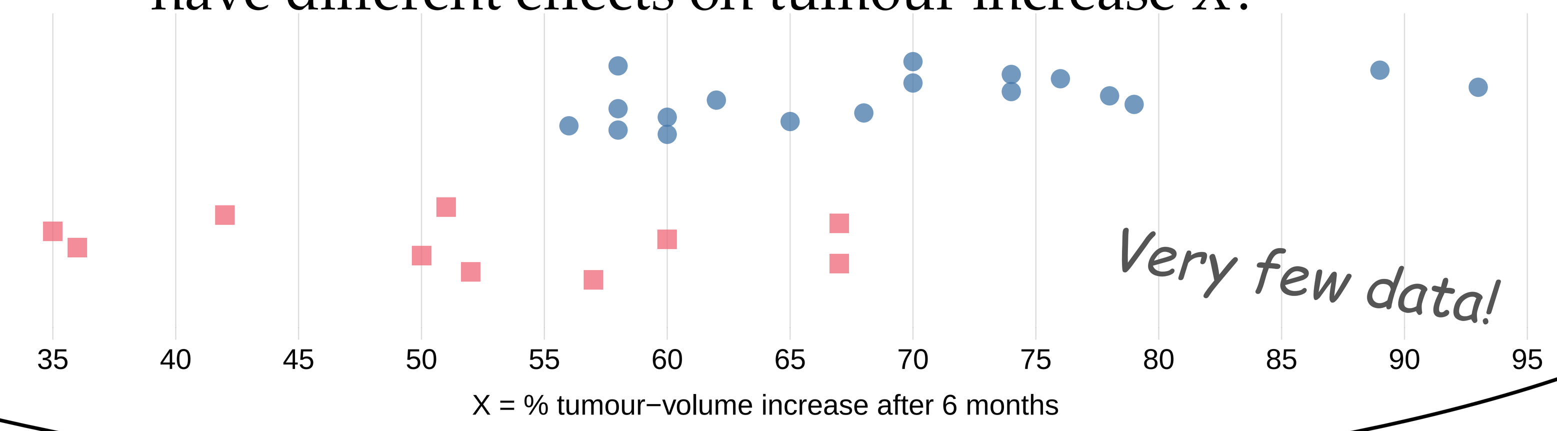
USER-FRIENDLY SOFTWARE FOR BAYESIAN ANALYSIS OF MEDICAL DATA

PierGianLuca Porta Mana, Kjetil Dyrland, Alexander Selvikvåg Lundervold

Dept of Computer Science, Electrical Engineering, and Mathematical Sciences, HVL; MMIV, HUS

Western Norway University of Applied Sciences

Simple example question:
do treatment A and treatment B
have different effects on tumour increase X?



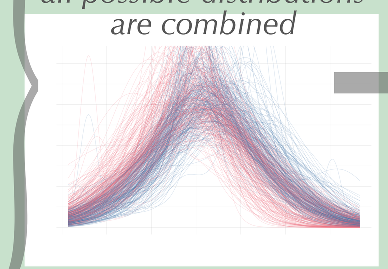
Bayesian theory

old sampling theory

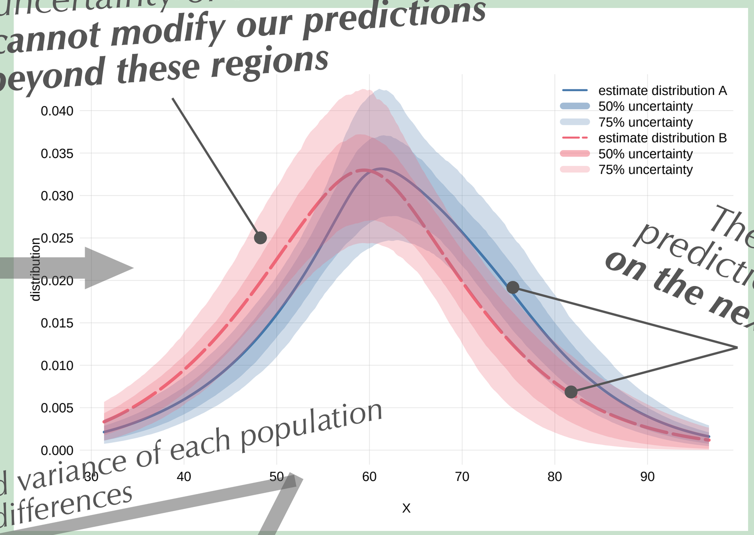
The Bayesian calculation is always the same
– no matter how many data we have
– no matter how the sample sizes were chosen
and it does not need any corrections

Bayesian probability theory considers
all likely **distributions for future patients**
in the two groups, given the data

all possible distributions
are combined



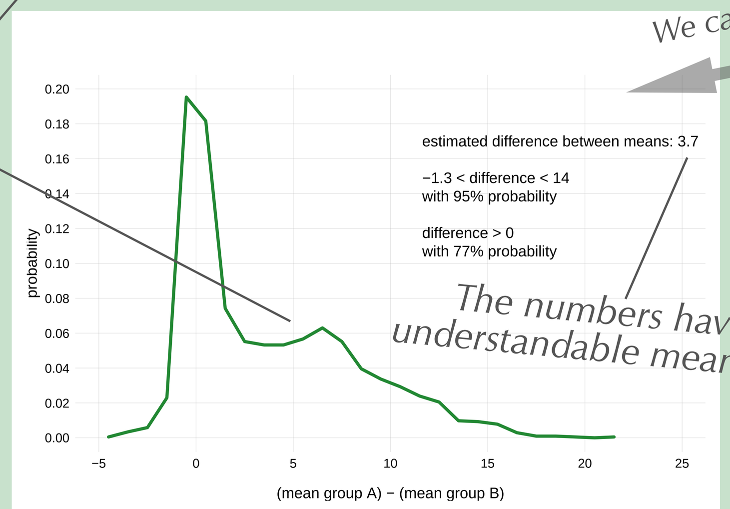
Estimates of full distributions
and uncertainty of estimates.
**New data cannot modify our predictions
beyond these regions**



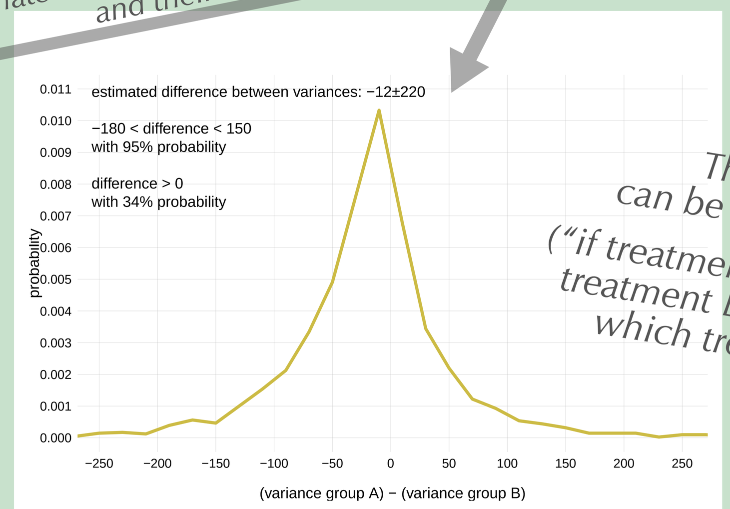
These curves are
predictions of the effect
on the next future patient

We can estimate the mean and variance of each population
and their differences

No assumptions
of gaussianity



The numbers have
understandable meanings

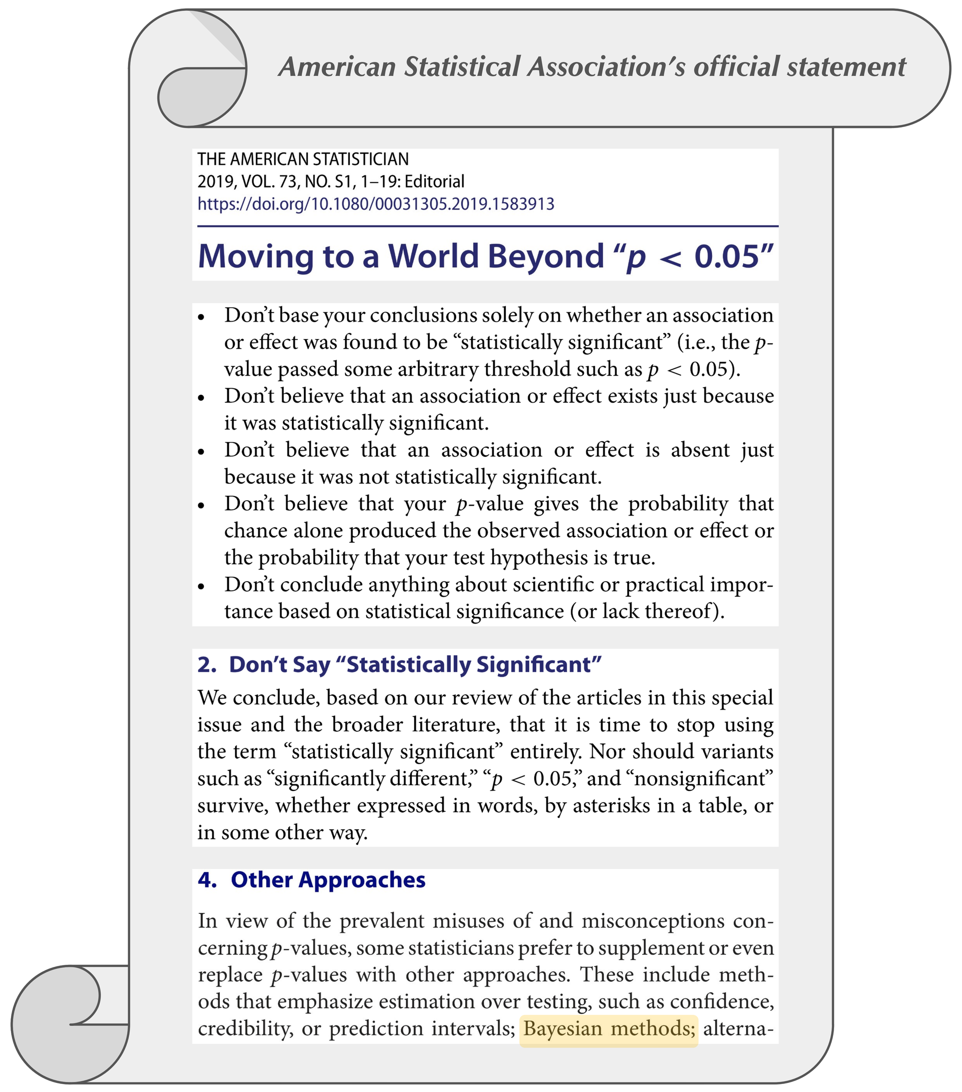


These detailed probabilities
can be used in a **cost/benefit analysis**
(“if treatment A costs \$ and gives Y% benefit,
treatment B costs \$\$ and gives Z% benefit,
which treatment should we choose?”)

“ Assuming that the two full-population distributions are smooth, we predict:

- The distributions of **future treatment outcomes** will be as in the plots, within the uncertainties shown
- Future patients under treatment A will have on average $X=63$, and $59<X<69$ with 95% probability
- Future patients under treatment B will have on average $X=59$, and $52<X<66$ with 95% probability
- The difference between the mean X of the treatments will be within –1.3 and 14 with 95% probability
- Average X under treatment A will be **larger** than under treatment B with 77% probability
- Variance of X under treatment A will be **smaller** than under treatment B with 66% probability ”

American Statistical Association’s official statement



THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 1–19: Editorial
https://doi.org/10.1080/00031305.2019.1583913

Moving to a World Beyond “$p < 0.05$”

- Don’t base your conclusions solely on whether an association or effect was found to be “statistically significant” (i.e., the p-value passed some arbitrary threshold such as $p < 0.05$).
- Don’t believe that an association or effect exists just because it was statistically significant.
- Don’t believe that an association or effect is absent just because it was not statistically significant.
- Don’t believe that your p-value gives the probability that chance alone produced the observed association or effect or the probability that your test hypothesis is true.
- Don’t conclude anything about scientific or practical importance based on statistical significance (or lack thereof).

2. Don’t Say “Statistically Significant”

We conclude, based on our review of the articles in this special issue and the broader literature, that it is time to stop using the term “statistically significant” entirely. Nor should variants such as “significantly different,” “$p < 0.05$,” and “nonsignificant” survive, whether expressed in words, by asterisks in a table, or in some other way.

4. Other Approaches

In view of the prevalent misuses of and misconceptions concerning p-values, some statisticians prefer to supplement or even replace p-values with other approaches. These include methods that emphasize estimation over testing, such as confidence, credibility, or prediction intervals; Bayesian methods; alterna-

*It’s actually more probable that the variances are **different**: non-gaussianity leads to unreliable F-test*

“ Assuming that the two full populations:
- are gaussian and independent
- have identical variances (F-test, $p=0.44$)

and that the sample sizes where decided before running the experiment, then:

This p-value is wrong if the sample sizes 18 & 10 were decided not in advance but by some other rule

The hypothesis that the population means are equal has a p-value 0.00043 (two-tail t-test: +4.0).

What does the number “0.00043” really mean?

The sample mean difference of X is17.6
The 95% confidence interval is [8.6, 26.7] ”

*This does **not** mean that 8.6 < X < 26.7 with 95% probability!
It means that this technique to construct the interval contains the true value in 95% of all imaginary datasets.
But what’s the best interval that contains X for **our** dataset?*

Compare the two analyses

The predictions of Bayesian theory are:
- detailed
- quantitative
- easy to understand

(eg, “fraction x% of population will have effect y”)

The statements of sampling theory are:
- vague
- obscure or misleading
- heavily dependent on tacit assumptions

This was just a simple example. Bayesian theory deals in the same way with multiple hypotheses, variates, and correlation questions

“Looks great, but there’s very little friendly software for doing this!”

**True! That’s why we are preparing
a user-friendly software to do Bayesian analysis
on (non-imaging) medical data**

The maths will be taken care of under the hood
The software will suggest meaningful questions to be asked (in line with ASA’s statement)
Works with continuous and categorical variables
- allows predictions about their correlations and relevance

No assumptions of gaussianity, linearity or other assumptions typical of sampling theory
No need for corrections of any kind

We’re already using a prototype version for drug-discovery research

Please get in touch if you want to test it and help us making a great sofware!