Chapter 12: What does a statistically significant results really tell us?

TXCL7565/PHSC7565

What This Chapter Covers

- Interpreting statistical significance

- Extreme skepticism

- Bayesian statistics

INTERPRETING STATISTICAL SIGNIFICANCE

Using a significance threshold of 5% does NOT imply that 5% of your 'significant' results are false.

Compare the following scenarios:

1. Early phase clinical trial - determine activity in humans of 200 candidate drugs found to be active in other mammals

2. Traditional herbal remedies - determine activity in humans of 200 candidates, none have any existing scientific evidence of activity

- Among the substances proven to be effective in other mammals, 90% are also genuinely effective in humans.

- Among traditional remedies, 10% are genuinely effective in humans.

- Both researchers carry out statistical testing using the usual standard of significance ($p < 0.05$)

- Both design their experiments to achieve 90% power

- All products are either completely inactive or they have a level of activity that exactly matches the figure used to plan experimental size.

look at all the previously available evidence as well as today's p-value

Two experiments may produce exactly the same p-value, but that does not mean that they necessarily lead to the same level of certainty that there is a true difference in outcome.

- If previous evidence (or basic scientific principles) already suggests a difference is very likely, a significant result will give a high level of confidence that there is a true difference.

- If prior information suggests that a real difference is unlikely, even a significant result will still leave considerable doubt.
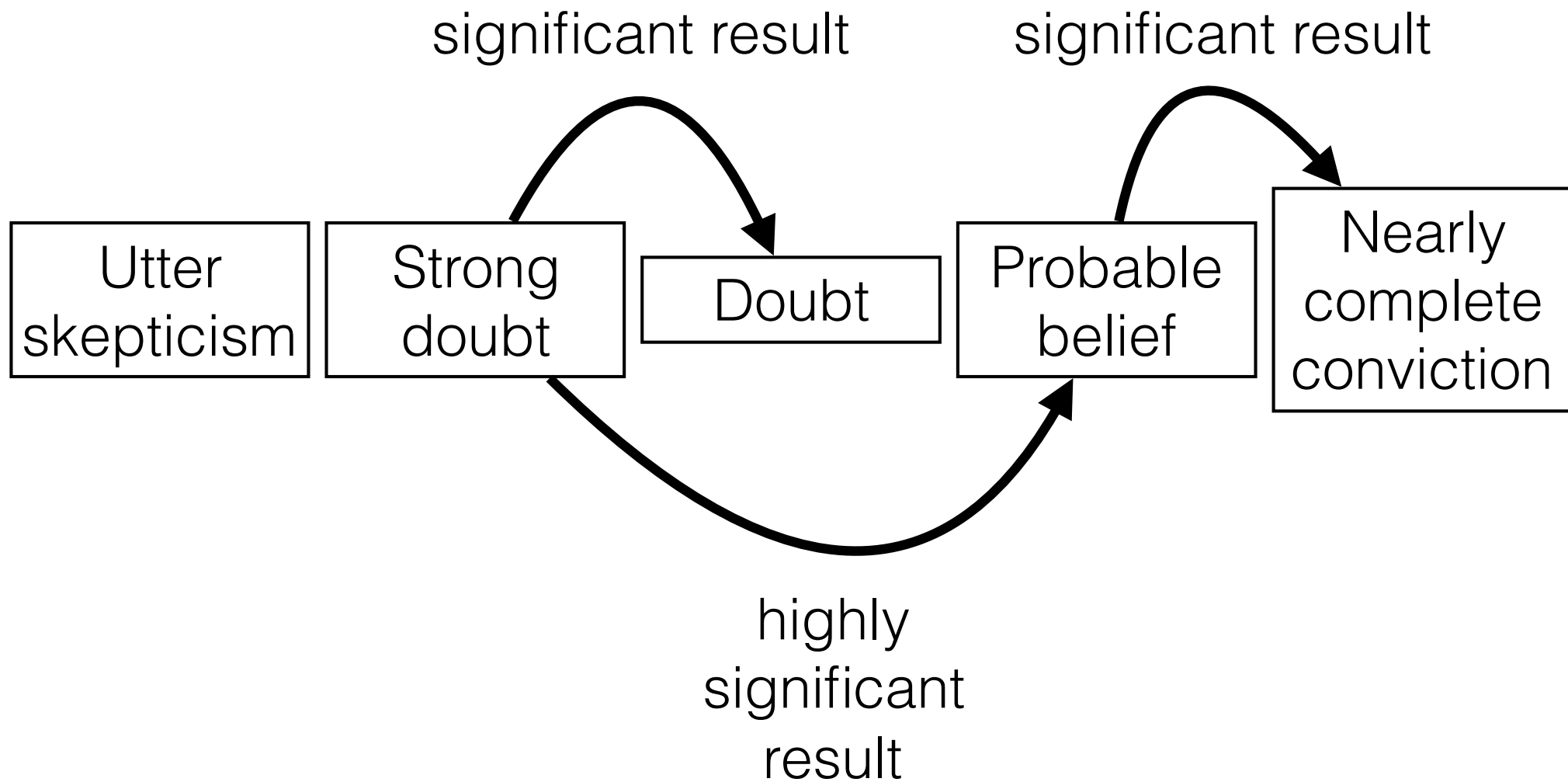
statistical results tell us how to modify our existing beliefs

A statistical results does not tell us what we should believe. It tells us how much we should change what we believe.

Non-significant: insufficient evidence to require any change

Significant (p<0.05): increase credence to a useful extent

Highly significant (p<0.001): increase credence markedly

BAYESIAN STATISTICS

the religions of statistics

Two 'religions' in statistics - Frequentist vs. Bayesian

- **Frequentist** - relies on p-values; most 'accepted'; taught in this class/book; easy to apply

- **Bayesian** - includes prior beliefs; relies on probability of being right (posterior probability) rather than probability of being wrong (p-value); requires advanced courses; harder to apply

Bayesian statistics

3 steps to a Bayesian analysis for the effectiveness of a product:

1. Determine the 'prior likelihood' that the product is effective.

2. Use the new evidence from the trial to calculate how much to modify our prior view

3. Use exact mathematical rules to combine the prior likelihood with the new evidence to produce a 'posterior likelihood'

What did we learn?

- The interpretation of statistical significance must involve not only looking at the p-value for the current experiment, but also taking stock of the previously available evidence.

- Statistical significant results from a single study should simply move us up a rung of the belief ladder and not necessarily shoot us all the way to 'nearly complete conviction'