

Biostats Lecture 5: Statistical Hypothesis Testing

Public Health 783

Ralph Trane
University of Wisconsin–Madison

Fall 2019



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Population of interest.

Get a representative sample.

Make assumptions about the way the sampling works (i.e. the distribution of the data).

Set up hypothesis.

Ask: "If the hypothesis is true, what is the probability of seeing data that are this far from the hypothesis?"

If probability is small, reject hypothesis. If probability large, do not reject hypothesis.

(**NOTE:** we NEVER accept the null hypothesis.)

Recap: Statistical Hypothesis Testing



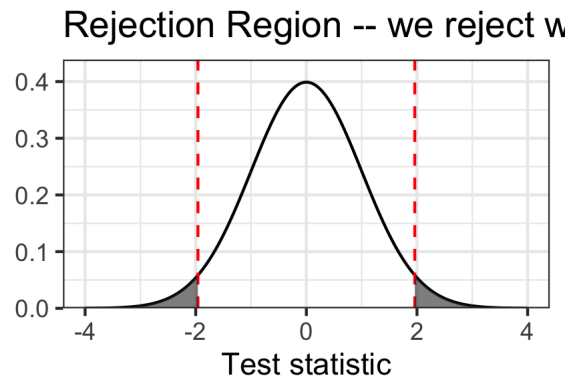
(For simplicity, assume we're testing $H_0 : \mu = 17$ vs. $H_A : \mu \neq 17$ at a $\alpha = 0.05$ level of significance.)

Get a sample

$$\text{Calculate } t_{obs} = \frac{\bar{x} - 17}{\text{SD}(\bar{x})}$$

Find $P(T > |t_{obs}| \mid H_0 \text{ is true})$. This is our p-value. Compare to significance level. If smaller, reject. If larger, do not reject.

Can also ask: what value of T will give us p-value smaller than α ?



What if H_0 is true, but we end up in grey anyway? Or what if H_0 is false, but we end up in white?

Errors we might make



Whenever you perform a hypothesis test, you might make a mistake

- Reject the null hypothesis when you shouldn't have
- Do not reject the null hypothesis when you should have

The first type of error is called type I error.

The second type of error is called type II error.

Type I Errors



When you reject the null hypothesis, but in fact the null hypothesis is true.

What is the probability of this happening?

$$\begin{aligned} P(\text{type I error}) &= P(\text{"far away"} | H_0 \text{ is true}) \\ &= P(T > |\text{critical value}| | H_0 \text{ is true}) \\ &= \alpha \end{aligned}$$

I.e. setting α means setting probability of type I error. So why don't we pick a super, super, super small α ?

Type II Errors



Adjusting the probability of type I error affects the probability of type II error.

Multiple Testing Problem



Probability of Type I error if we test one hypothesis: 0.05 (or, in general, α)

What if we test more hypotheses?

Assume they're independent, and we test k hypotheses H_1, H_2, \dots, H_k . Then

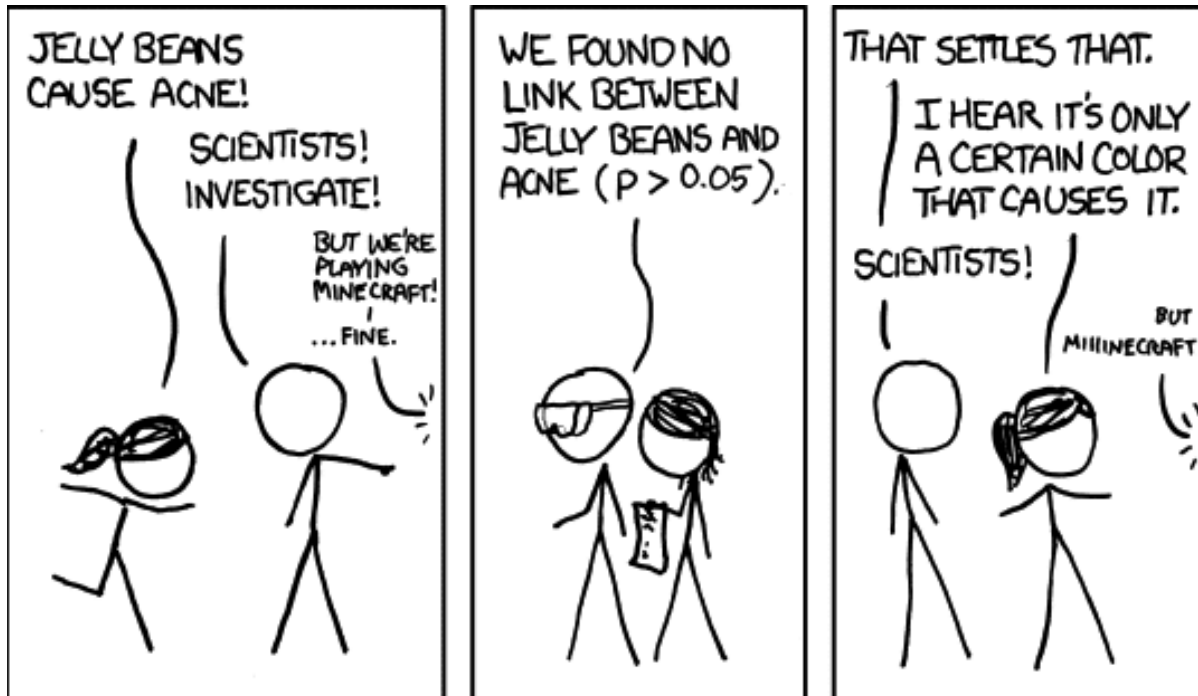
$$\begin{aligned} P(\text{no type I error}) &= P(\text{do not reject any hypotheses} | \text{all hypotheses are true}) \\ &= P(\text{do not reject } H_1 | H_1 \text{ is true}) \cdots P(\text{do not reject } H_k | H_k \text{ is true}) \\ &= 0.95 \cdots 0.95 = 0.95^k. \end{aligned}$$

Probability of rejecting at least one hypothesis *IF* they are all true:
 $1 - P(\text{do not reject any hypotheses} | \text{all hypotheses are true})$.

Multiple Testing Problem



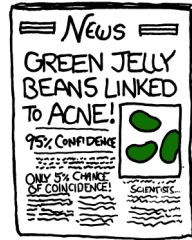
Multiple Testing Problem



Multiple Testing Problem



Multiple Testing Problem



Coming up next: Confidence Intervals



Imagine you test a range of different hypothesis.

Confidence interval: the values that are NOT rejected.

Hypothesis test: is it plausible this one value is the true value?

Confidence Interval: all plausible values.

So why would you ever perform a test, and report a p-value, but not a confidence interval? Good question.....