Biostats Lecture 5: Statistical Hypothesis Testing

Public Health 783

Ralph Trane University of Wisconsin–Madison

Fall 2019



Recap



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



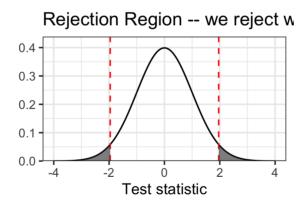
(For simplicity, assume we're testing $H_0: \mu=17$ vs. $H_A: \mu \neq 17$ at a lpha=0.05 level of significance.)

Get a sample

Calculate
$$t_{obs}=rac{ar{x}-17}{\mathrm{SD}(ar{x})}$$

Find $P(T > |t_{obs}| | H_0$ 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?

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

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

Type II Errors



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



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, \ldots, H_k . Then

```
P(	ext{no type I error}) = P(	ext{do not reject any hypotheses}|	ext{all hypotheses are true}) = P(	ext{do not reject } H_1|H_1 	ext{ is true}) \cdots P(	ext{do not reject } H_k|H_k 	ext{ is } = 0.95 \cdots 0.95 = 0.95^k.
```

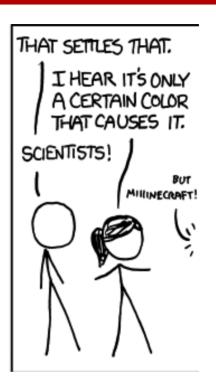
Probability of rejecting at least one hypothesis *IF* they are all true: 1 - P(do not reject any hypotheses|all hypotheses are true).







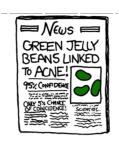






WE FOUND NO LINK BETWEEN PURPLE JELLY BEAUS AND AGNE (P>0.05)	WE FOUND NO LINK BETWEEN BROWN TELLY BEANS AND ACKE (P > 0.05)	WE FOUND NO LINK GETWEEN PINK JELLY BEAKS AND ACKE (P>0.05).	WE FOUND NO LINK GETWEEN BLUE JELLY BEANS AND ACKE (P>0.05).	WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND AONE (P>0.05)
WE POUND NO LINK BETWEEN SALMON JELLY BEFWS AND ACNE (P>0.05).	WE POUND NO LINK BETVEEN RED JELLY BEAKS AND ADIE (P>0.05).	WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEAKS AND AGIE (P>0.05).	WE FOUND NO LINK GETWEEN MAGENTA JELLY BEAKS AND AGIE (P>0.05).	WE FOUND NO LINK BETWEEN YELLOW JELLY BEAKS AND AGNE (P>0.05)
WE POUND NO LINK BETWEEN GREY JELLY BEFWS AND ACNE (P>0.05)	WE POUND NO LINK BETWEEN TAN JELLY BEAUS AND ADIE (P>0.05).	WE FOUND NO LINK BETWEEN CYAN JELLY BEAKS AND ACIE (P>0.05).	WE FOUND A LINK BETWEEN GREEN JELLY BEAKS AND AGIE (P < 0.05).	WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACKE (P>0.05).
WE FOUND NO LINK BETWEEN BEIGE TELLY BEAUS AND ACKE (P > 0.05).	WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ADIE (P>0.05).	WE FOUND NO LINK BETWEEN BLACK JELLY BEAKS AND AGNE (P > 0.05).	WE ROUND NO LINK BETWEEN PEACH JELLY BEANS AND ADIE (P > 0.05).	WE FOUND NO LINK BETWEEN ORANGE TELLY BEANS AND ACNE (P > 0.05).





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.....