P-values, confidence intervals and point estimates. Inference in Statistics

Anne Lyngholm Sørensen

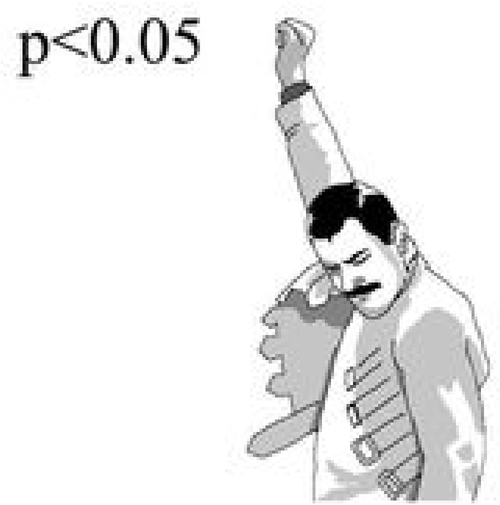
June 24, 2020

Outline

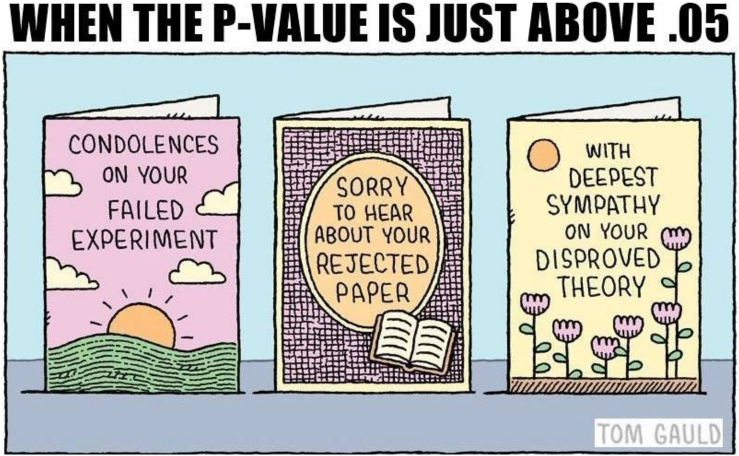
Plan for this talk

- We are going to talk about statistical inference.
  - Especially statistical inference in publications.
- This includes a discussion about whether p-values are the most important statistical metric in publications. And why it has so much attention.
- We will be using simulations for understanding p-values, point estimates and confidence intervals in more depth.

What is statistical inference? Is it this?



What is statistical inference? Is it this?



Different kinds of hypotheses

- "The hazard of death is different in the treatment group compared to the control group."
- "The level of testosterone is higher in males with XXX gene-mutation than other males"
- "The rate of infection per year is lower in people treated with medicine y compared with people treated with medicine x"

Different kinds of hypotheses

Two-sided

- H_0 : Hazard Ratio $= 1$ vs. H_A : Hazard Ratio $\neq 1$

One-sided

- $H_0 : Level_A > Level_B$ vs. $H_A : Level_A \leq Level_B$
  - Rewrite $H_0 : Level_A - Level_B > 0$.
- H_0 : Rate Ratio < 1 vs. H_A : Rate Ratio ≥ 1

What difference really matters?

We can test these hypotheses, but maybe we should think more in depth about what we are testing.
Say we do a hypothesis test of:

$$H_0 : Level_A - Level_B > 0 \text{ vs. } H_A : Level_A - Level_B \leq 0$$

If the true difference is 0.0001, we will with enough participants be able to reject this hypothesis. Is that true difference a relevant difference?

What difference really matters?

We can test these hypotheses, but maybe we should think more in depth about what we are testing.
Say we do a hypothesis test of:

$$H_0 : Level_A - Level_B > 0 \text{ vs. } H_A : Level_A - Level_B \leq 0$$

If the true difference is 0.0001, we will with enough participants be able to reject this hypothesis. Is that true difference a relevant difference?
A p-value below 0.05 (a rejected null hypothesis) does not equal relevance

Hypothesis for this talk

Are Australians the same height as New Zealanders?

$$H_0 : \text{Average Height}_{AUS} - \text{Average Height}_{NZ} = 0$$

vs.

$$H_A : \text{Average Height}_{AUS} - \text{Average Height}_{NZ} \neq 0$$

We are going to assume that the heights for our two populations are approximately normal distributed.

Simulation

Let us simulate some data!

- Everything that exists is an object.
- Everything that happens is a function call.

P-values and hypothesis testing

To be able to reject a hypothesis, we need the following three things:

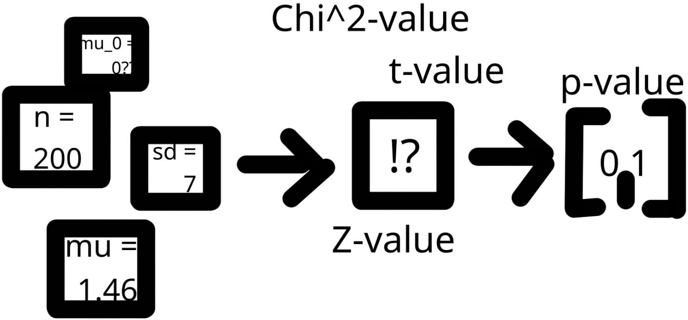
1. A hypothesis to test.
2. A metric to evaluate our data in relation to the hypothesis.
3. Some cut-off value related to when we reject or can not reject the hypothesis.

P-values and hypothesis testing

To be able to reject a hypothesis, we need the following three things:

1. A hypothesis to test.
2. A metric to evaluate our data in relation to the hypothesis.
3. Some cut-off value related to when we reject or can not reject the hypothesis.

The *p* value is the metric that we most often use. And the most common cut-off value is 0.05.


mu_0 =
0?
n =
200
sd =
7
mu =
1.46
Chi^2-value
t-value
!?
Z-value
p-value
[0,1]

Z-value and P-value

The p value is between 0 and 1. And it tells the probability of observing something at least as extreme as a specific value in relation to the null hypothesis.
To evaluate if the value we have measured is close to the null hypothesis, we need to understand which values are likely under the null hypothesis.

Z-value and P-value

The p value is between 0 and 1. And it tells the probability of observing something at least as extreme as a specific value in relation to the null hypothesis.
To evaluate if the value we have measured is close to the null hypothesis, we need to understand which values are likely under the null hypothesis.

$$Z = \frac{\hat{\mu} - \mu_0}{s^2/\sqrt{n}}$$

- $\hat{\mu}$ is the estimate
- μ_0 is the null hypothesis value
- s^2 is the variance estimate
- n is the number of observations

When is the Z value large?

What can give you a large Z value?

- The larger the distance between $\hat{\mu} - \mu_0$.
- The larger the number of observations n.

So do not get fooled by very small p-values. There could be testing of a too loose hypothesis. And p-values can be really small by having a really large number of observations.

When is the Z value small?

What can give you a small Z value?

- The smaller the distance between $\mu - \mu_0$.
- The smaller the number of observations n.

So do not get fooled by large p-values. They could test too strict a hypothesis. And they can become significant by getting a larger number of number of observations.

Height example

We are going to use the height example to illustrate some of the points.

$$t = \frac{\hat{\mu}_{AUS} - \hat{\mu}_{NZ}}{\sqrt{s^2 \cdot (1/n_1 + 1/n_2)}}$$

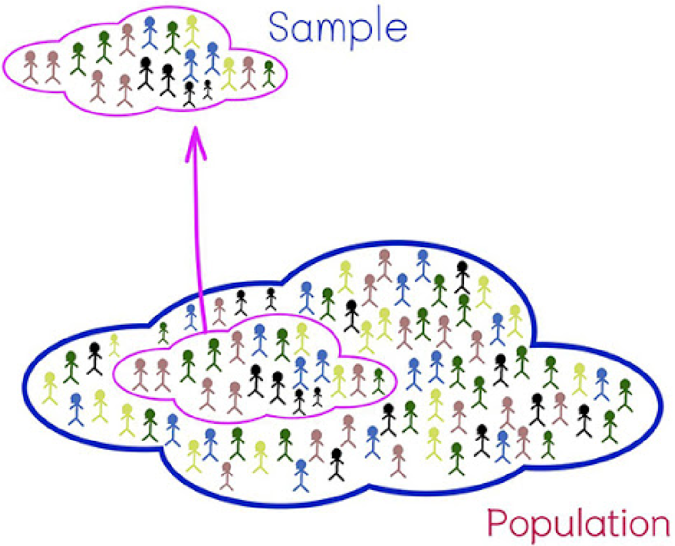
Simulation

Let us do some simulations to get smarter!

Point Estimates

- A crucial information is the point estimate. It is the best guess in data about the metric of interest.
- There is not that many ways to go wrong with a point estimate. The most important thing to remember is that it is just an estimate.
- The more observations, the better and more precise estimate.

Point Estimtes



Simulation

Let us do some simulations to get smarter!

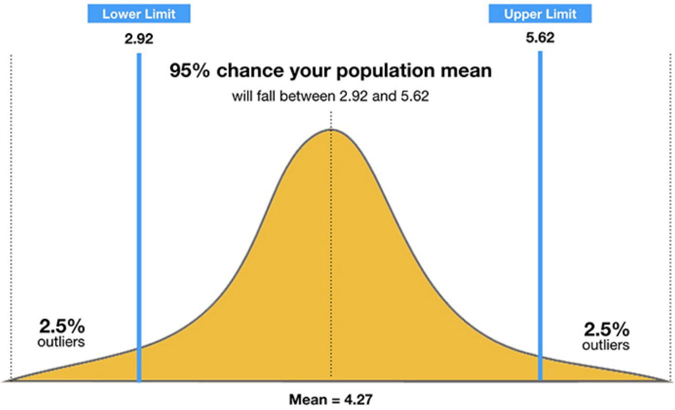
The good old formula

One sample:

$$t = \hat{\mu} \pm 1.96 \frac{s}{\sqrt{n}}$$

Our example:

$$95\ \text{CI} = \hat{\mu}_{AUS} - \hat{\mu}_{NZ} \pm 1.96 \frac{s}{\sqrt{1/n_{AUS} + 1/n_{NZ}}}$$


Lower Limit
2.92
Upper Limit
5.62
95% chance your population mean
will fall between 2.92 and 5.62
2.5%
outliers
2.5%
outliers
Mean = 4.27

What is confidence?

Forget about that picture - it is wrong!

What is confidence?

Forget about that picture - it is wrong!

The choice of the word "confidence" is actually quite crutial in understanding what a confidence interval is. And most importantly, confidence does not equal probability!

Misunderstandings

Given the data that you have, a confidence interval gives us α level confidence that our interval captured the true mean.

Misunderstandings

Given the data that you have, a confidence interval gives us α level confidence that our interval captured the true mean. There is not $1 - \alpha$ % chance that the true parameter is in the interval. There is a $1 - \alpha$ % chance that the interval contains the parameter.

Misunderstandings

Given the data that you have, a confidence interval gives us α level confidence that our interval captured the true mean. There is not $1 - \alpha$ % chance that the true parameter is in the interval. There is a $1 - \alpha$ % chance that the interval contains the parameter.

- This means that before the interval is computed, there is a 95% chance that the true value will be contained in the interval. But when it is computed, it will either contain the true value or not. And we will never know the true value.

Misunderstandings

Given the data that you have, a confidence interval gives us α level confidence that our interval captured the true mean. There is not $1 - \alpha$ % chance that the true parameter is in the interval. There is a $1 - \alpha$ % chance that the interval contains the parameter.

- This means that before the interval is computed, there is a 95% chance that the true value will be contained in the interval. But when it is computed, it will either contain the true value or not. And we will never know the true value.

- A 95% confidence level does not mean that 95% of the sample data lie within the confidence interval.

The non-fun tombola example

Simulation

Let us do the last round of simulation for understanding confidence better!

Summary

- P-values can be deceiving in relation to the relevance of ones findings.
- Always think about whether the point estimate and belonging confidence interval are showing something that can be relevant.