HarvardX PH525.1x
Statistics and R

edX

Ayuda

Mguzmanfredes

Curso   Progreso   Fechas   Discusión   Notas   Data Analysis for Life Sciences Series   Preguntas Frecuentes

Curso / Week 3 / Inference I: P-values, Confidence Intervals and Power Calculations

Anterior   Siguiente

Confidence Intervals Exercises

Marcar esta página

Ocultar Notas

For these exercises we will load the babies dataset from `babies.txt` . We will use this data to review the concepts behind the p-values and then test confidence interval concepts.

```
url <- "https://raw.githubusercontent.com/genomicsclass/dagdata/master/inst/extdata/babie
filename <- basename(url)
download(url, destfile=filename)
babies <- read.table("babies.txt", header=TRUE)
```

This is a large dataset (1,236 cases), and we will pretend that it contains the entire population in which we are interested. We will study the differences in birth weight between babies born to smoking and non-smoking mothers.

First, let's split this into two birth weight datasets: one of birth weights to non-smoking mothers and the other of birth weights to smoking mothers.

```
bwt.nonsmoke <- filter(babies, smoke==0) %>% select(bwt) %>% unlist
bwt.smoke <- filter(babies, smoke==1) %>% select(bwt) %>% unlist
```

Now, we can look for the true population difference in means between smoking and non-smoking birth weights.

```
library(rafalib)
mean(bwt.nonsmoke)-mean(bwt.smoke)
popsd(bwt.nonsmoke)
popsd(bwt.smoke)
```

The population difference of mean birth weights is about 8.9 ounces. The standard deviations of the nonsmoking and smoking groups are about 17.4 and 18.1 ounces, respectively.

As we did with the mouse weight data, this assessment interactively reviews inference concepts using simulations in R. We will treat the babies dataset as the full population and draw samples from it to simulate individual experiments. We will then ask whether somebody who only received the random samples would be able to draw correct conclusions about the population.

We are interested in testing whether the birth weights of babies born to non-smoking mothers are significantly different from the birth weights of babies born to smoking mothers.

Confidence Intervals Exercises #1

1 punto posible (calificable)

Set the seed at 1 and obtain two samples, each of size N = 25, from non-smoking mothers (dat.ns) and smoking mothers (dat.s). If instead of CLT, we use the t-distribution approximation, what do we add and subtract to obtain a 99% confidence interval (use `2*N-2` degrees of freedom)?

Enviar   Ha realizado 0 de 5 intentos

Confidence Intervals Exercises #2

0/1 punto (calificado)

No matter which way you compute it, the p-value `pval` is the probability that the null hypothesis could have generated a t-statistic more extreme than than what we observed: `tval`. If the p-value is very small, this means that observing a value more extreme than `tval` would be very rare if the null hypothesis were true, and would give strong evidence that we should **reject** the null hypothesis. We determine how small the p-value needs to be to reject the null by deciding how often we would be willing to mistakenly reject the null hypothesis.

The standard decision rule is the following: choose some small value α (in most disciplines the conventional choice is $\alpha = 0.05$) and reject the null hypothesis if the p-value is less than α. We call α the *significance level* of the test.

It turns out that if we follow this decision rule, the probability that we will reject the null hypothesis by mistake is equal to α. (This fact is not immediately obvious and requires some probability theory to show.) We call the *event* of rejecting the null hypothesis, when it is in fact true, a *Type I error*, we call the *probability* of making a Type I error, the *Type I error rate*, and we say that rejecting the null hypothesis when the p-value is less than α, *controls* the Type I error rate so that it is equal to α. We will see a number of decision rules that we use in order to control the probabilities of other types of errors. Often, we will guarantee that the probability of an error is less than some level, but, in this case, we can guarantee that the probability of a Type I error is *exactly equal* to α.

Which of the following sentences about a Type I error is **not** true?

- ○ The following is another way to describe a Type I error: you decided to reject the null hypothesis on the basis of data that was actually generated by the null hypothesis.

- ● The following is the another way to describe a Type I error: due to random fluctuations, even though the data you observed were actually generated by the null hypothesis, the p-value calculated from the observed data was small, so you rejected it.

- ○ From the original data alone, you can tell whether you have made a Type I error.

- ○ In scientific practice, a Type I error constitutes reporting a "significant" result when there is actually no result.

✖

Enviar   Ha realizado 1 de 5 intentos

Confidence Intervals Exercises #3

1 punto posible (calificable)

In the simulation we have set up here, we know the null hypothesis is false -- the true value of difference in means is actually around 8.9. Thus, we are concerned with how often the decision rule outlined in the last section allows us to conclude that the null hypothesis is actually false. In other words, we would like to quantify the *Type II error rate* of the test, or the probability that we fail to reject the null hypothesis when the alternative hypothesis is true.

Unlike the Type I error rate, which we can characterize by assuming that the null hypothesis of "no difference" is true, the Type II error rate cannot be computed by assuming the alternative hypothesis alone because the alternative hypothesis alone does not specify a particular value for the difference. It thus does not nail down a specific distribution for the t-value under the alternative.

For this reason, when we study the Type II error rate of a hypothesis testing procedure, we need to assume a particular *effect size*, or hypothetical size of the difference between population means, that we wish to target. We ask questions such as "what is the smallest difference I could reliably distinguish from 0 given my sample size N?" or, more commonly, "How big does N have to be in order to detect that the absolute value of the difference is greater than zero?" Type II error control plays a major role in designing data collection procedures **before** you actually see the data, so that you know the test you will run has enough sensitivity or *power*. Power is one minus the Type II error rate, or the probability that you ⌨ Ocultar Notas reject the null hypothesis when the alternative hypothesis is true.

reject the null hypothesis when the alternative hypothesis is true.

There are several aspects of a hypothesis test that affect its power for a particular effect size. Intuitively, setting a lower α decreases the power of the test for a given effect size because the null hypothesis will be more difficult to reject. This means that for an experiment with fixed parameters (i.e., with a predetermined sample size, recording mechanism, etc), the power of the hypothesis test trades off with its Type I error rate, no matter what effect size you target.

We can explore the trade off of power and Type I error concretely using the babies data. Since we have the full population, we know what the true effect size is (about 8.93) and we can compute the power of the test for true difference between populations.

Set the seed at 1 and take a random sample of $N = 5$ measurements from each of the smoking and nonsmoking datasets. What is the p-value (use the `t.test()` function)?

Enviar   Ha realizado 0 de 5 intentos

‹ Anterior   Siguiente ›

edX

edX

Acerca de
Afiliados
edX para negocios
Open edX
Carreras
Noticias

Legal

Condiciones de Servicio y Código de Honor
Política de privacidad
Políticas de Accesibilidad
Política de marcas
Mapa del Sitio

Contáctanos

Blog
Contáctenos
Centro de Ayuda
Kit Multimedia

✎ Ocultar Notas

✏ Ocultar Notas

✏ Ocultar Notas