Hypothesis testing

Statistical Inference

Brian Caffo, Jeff Leek, Roger Peng

## Hypothesis testing

* Hypothesis testing is concerned with making decisions using data
* A null hypothesis is specified that represents the status quo, usually labeled
* The null hypothesis is assumed true and statistical evidence is required to reject it in favor of a research or alternative hypothesis

## Hypothesis testing

* The alternative hypotheses are typically of the form , or
* Note that there are four possible outcomes of our statistical decision process

|  |  |  |
| --- | --- | --- |
| Truth | Decide | Result |
|  |  | Correctly accept null |
|  |  | Type I error |
|  |  | Correctly reject null |
|  |  | Type II error |

## Example

* Consider our example again
* A reasonable strategy would reject the null hypothesis if was larger than some constant, say
* Typically, is chosen so that the probability of a Type I error, , is (or some other relevant constant)
* = Type I error rate = Probability of rejecting the null hypothesis when, in fact, the null hypothesis is correct

## Discussion

* In general we don't convert back to the original scale
* We would just reject because the Z-score; which is how many standard errors the sample mean is above the hypothesized mean
* is greater than
* Or, whenever

## Notes

* We have fixed to be low, so if we reject (either our model is wrong) or there is a low probability that we have made an error
* We have not fixed the probability of a type II error, ; therefore we tend to say ``Fail to reject '' rather than accepting
* Statistical significance is no the same as scientific significance
* The region of TS values for which you reject is called the rejection region

## Example reconsidered

* Consider our example again. Suppose that (rather than ). Then consider that
* So that our test statistic is now $(32 - 30) / 10 = 0.8 $, while the critical value is .
* We now fail to reject.

## T test in R

library(UsingR); data(father.son)

## Loading required package: MASS

## Loading required package: HistData

## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, round.POSIXt, trunc.POSIXt, units

##   
## Attaching package: 'UsingR'

## The following object is masked from 'package:survival':  
##   
## cancer

t.test(father.son$sheight - father.son$fheight)

##   
## One Sample t-test  
##   
## data: father.son$sheight - father.son$fheight  
## t = 11.789, df = 1077, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 0.8310296 1.1629160  
## sample estimates:  
## mean of x   
## 0.9969728

## Exact binomial test

* Recall this problem, *Suppose a friend has children, of which are girls and none are twins*
* Perform the relevant hypothesis test.
* What is the relevant rejection region so that the probability of rejecting is (less than) 5%?

|  |  |
| --- | --- |
| Rejection region | Type I error rate |
| [0 : 8] | 1 |
| [1 : 8] | 0.9960938 |
| [2 : 8] | 0.9648438 |
| [3 : 8] | 0.8554688 |
| [4 : 8] | 0.6367187 |
| [5 : 8] | 0.3632813 |
| [6 : 8] | 0.1445313 |
| [7 : 8] | 0.0351563 |
| [8 : 8] | 0.0039062 |

## Notes

* It's impossible to get an exact 5% level test for this case due to the discreteness of the binomial.
* The closest is the rejection region [7 : 8]
* Any alpha level lower than 0.0039062 is not attainable.
* For larger sample sizes, we could do a normal approximation, but you already knew this.
* Two sided test isn't obvious.
* Given a way to do two sided tests, we could take the set of values of for which we fail to reject to get an exact binomial confidence interval (called the Clopper/Pearson interval, BTW)
* For these problems, people always create a P-value (next lecture) rather than computing the rejection region.