Modernizing *Intro to Statistics* with infer - Part 1: T-test

Zhaoshan Duan, Lydia Gibson

2021-01-24

In this blog series, we aim to advocate for the modernization of statistical inference in *“Intro to Statistics”* courses by using the [infer](https://infer.netlify.app/) package. The infer package features tidyverse-friendly wrapper functions to perform inferences and consolidates a [unified workflow framework](http://allendowney.blogspot.com/2016/06/there-is-still-only-one-test.html), which encompasses the essence of hypothesis testing.

Throughout the series, we plan to compare and contrast the classical approach of conducting hypothesis testing using **base R** functions with the [infer](https://infer.netlify.app/) workflow on some of the most popular statistical topics covered in introductory statistics classes. Our goal in each blog is to demonstrate how the infer framework adds additional workflow explainability and procedural clarity to the classical hypothesis testing approach.

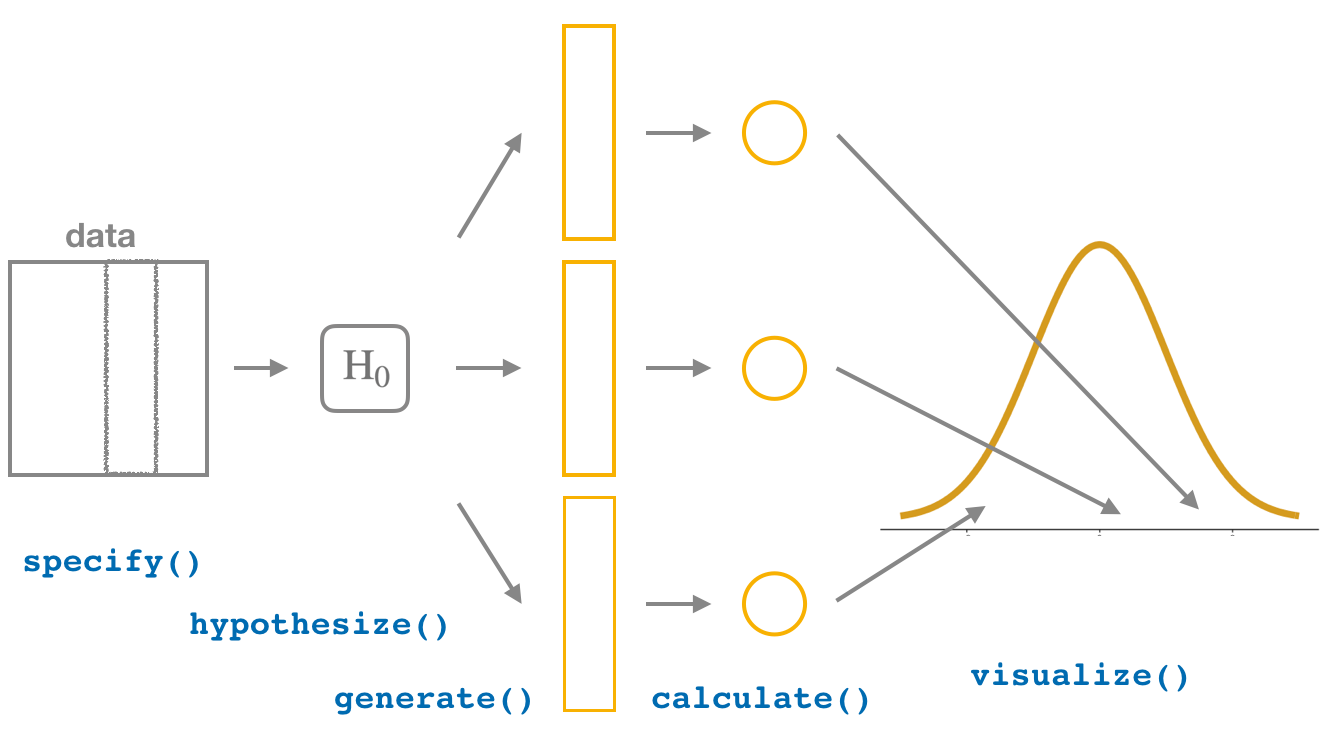


FIGURE 1.1 Hypothesis testing with the infer package

In this first blog, we are focusing on conducting **Two-sample T-test** and compute confidence interval using the ncbirths dataset from [*OpenIntro Statistics*](https://www.openintro.org/book/os/). We start with a quick overview of the dataset then proceed to a step-by-step analysis using the infer pipeline. To clearly demonstrate comparison, we include a brief discussion on the classical approach for each corresponding step of the infer framework.

Last Updated: January 25, 2022

# Data

Throughout this post, we make use of the ncbirths dataset from the openintro package.

library(openintro)

## Loading required package: airports

## Loading required package: cherryblossom

## Loading required package: usdata

data(ncbirths)

ncbirths records 1000 cases of North Carolina births in 2004, and has been of interest to medical researchers on studying the relation between habits and practices of expectant mothers and the birth of their children. The data looks like this:

str(ncbirths)

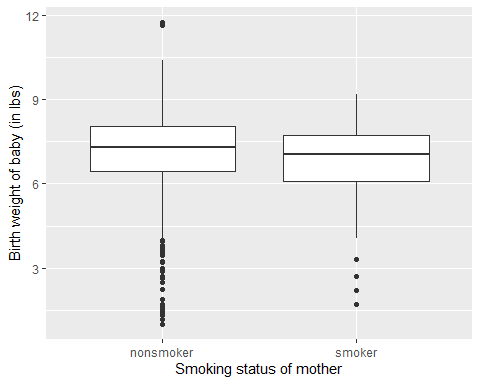
## tibble [1,000 x 13] (S3: tbl\_df/tbl/data.frame)  
## $ fage : int [1:1000] NA NA 19 21 NA NA 18 17 NA 20 ...  
## $ mage : int [1:1000] 13 14 15 15 15 15 15 15 16 16 ...  
## $ mature : Factor w/ 2 levels "mature mom","younger mom": 2 2 2 2 2 2 2 2 2 2 ...  
## $ weeks : int [1:1000] 39 42 37 41 39 38 37 35 38 37 ...  
## $ premie : Factor w/ 2 levels "full term","premie": 1 1 1 1 1 1 1 2 1 1 ...  
## $ visits : int [1:1000] 10 15 11 6 9 19 12 5 9 13 ...  
## $ marital : Factor w/ 2 levels "not married",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ gained : int [1:1000] 38 20 38 34 27 22 76 15 NA 52 ...  
## $ weight : num [1:1000] 7.63 7.88 6.63 8 6.38 5.38 8.44 4.69 8.81 6.94 ...  
## $ lowbirthweight: Factor w/ 2 levels "low","not low": 2 2 2 2 2 1 2 1 2 2 ...  
## $ gender : Factor w/ 2 levels "female","male": 2 2 1 2 1 2 2 2 2 1 ...  
## $ habit : Factor w/ 2 levels "nonsmoker","smoker": 1 1 1 1 1 1 1 1 1 1 ...  
## $ whitemom : Factor w/ 2 levels "not white","white": 1 1 2 2 1 1 1 1 2 2 ...

head(ncbirths)

## # A tibble: 6 x 13  
## fage mage mature weeks premie visits marital gained weight lowbirthweight  
## <int> <int> <fct> <int> <fct> <int> <fct> <int> <dbl> <fct>   
## 1 NA 13 younger~ 39 full ~ 10 not mar~ 38 7.63 not low   
## 2 NA 14 younger~ 42 full ~ 15 not mar~ 20 7.88 not low   
## 3 19 15 younger~ 37 full ~ 11 not mar~ 38 6.63 not low   
## 4 21 15 younger~ 41 full ~ 6 not mar~ 34 8 not low   
## 5 NA 15 younger~ 39 full ~ 9 not mar~ 27 6.38 not low   
## 6 NA 15 younger~ 38 full ~ 19 not mar~ 22 5.38 low   
## # ... with 3 more variables: gender <fct>, habit <fct>, whitemom <fct>

In the case of ncbirths, we study the mean difference between the expectant mothers’ smoking status by looking their babies’ weights. From observing the box plot below, we see that babies from mothers that smoke seem to weigh less than babies from mother that do not smoke. However, the difference can be due to sampling. To determine if this difference is statistically significant, we conduct **Two-sample T-test**.

ncbirths %>%   
 select(habit, weight) %>%   
 drop\_na() %>%   
 ggplot(aes(x = habit, y = weight)) +  
 geom\_boxplot()+   
 labs(x = "Smoking status of mother", y = "Birth weight of baby (in lbs)")



# specify()

First, we need to specify the variables of interest. In this case, they are weight, weight of the baby at birth in pounds, and habit, status of the mother as a nonsmoker or a smoker. In **base R**, we can choose to manually subset the variables:

nc\_nosmoker<- subset(ncbirths ,habit == "nonsmoker")  
nc\_smoker<- subset(ncbirths ,habit == "smoker")

Using specify(), we can write:

ncbirths %>%   
 specify(weight ~ habit)

## Warning: Removed 1 rows containing missing values.

## Response: weight (numeric)  
## Explanatory: habit (factor)  
## # A tibble: 999 x 2  
## weight habit   
## <dbl> <fct>   
## 1 7.63 nonsmoker  
## 2 7.88 nonsmoker  
## 3 6.63 nonsmoker  
## 4 8 nonsmoker  
## 5 6.38 nonsmoker  
## 6 5.38 nonsmoker  
## 7 8.44 nonsmoker  
## 8 4.69 nonsmoker  
## 9 8.81 nonsmoker  
## 10 6.94 nonsmoker  
## # ... with 989 more rows

# hypothesize()

The next step is to declare the null hypothesis. For our research question, we are examining the weight difference between babies whose mothers do or do not smoke. We thus construct the following hypothesis:

In **base R**, this step is normally done on *actual* paper and assumed when we run t.test(). However, with infer, we can use hypothesize() and set null = "independence" since we have two samples.

ncbirths %>%   
 specify(weight ~ habit) %>%   
 hypothesize(null = "independence")

## Warning: Removed 1 rows containing missing values.

## Response: weight (numeric)  
## Explanatory: habit (factor)  
## Null Hypothesis: independence  
## # A tibble: 999 x 2  
## weight habit   
## <dbl> <fct>   
## 1 7.63 nonsmoker  
## 2 7.88 nonsmoker  
## 3 6.63 nonsmoker  
## 4 8 nonsmoker  
## 5 6.38 nonsmoker  
## 6 5.38 nonsmoker  
## 7 8.44 nonsmoker  
## 8 4.69 nonsmoker  
## 9 8.81 nonsmoker  
## 10 6.94 nonsmoker  
## # ... with 989 more rows

# generate()

After asserting our null hypothesis, we construct a null distribution via permutation. We will obtain 1000 observations that are “shuffled”.

ncbirths %>%   
 specify(weight ~ habit) %>%   
 hypothesize(null = "independence") %>%   
 generate(reps = 1000, type = "permute")

## Warning: Removed 1 rows containing missing values.

## Response: weight (numeric)  
## Explanatory: habit (factor)  
## Null Hypothesis: independence  
## # A tibble: 999,000 x 3  
## # Groups: replicate [1,000]  
## weight habit replicate  
## <dbl> <fct> <int>  
## 1 7.13 nonsmoker 1  
## 2 7 nonsmoker 1  
## 3 7.44 nonsmoker 1  
## 4 9.19 nonsmoker 1  
## 5 8.31 nonsmoker 1  
## 6 7.5 nonsmoker 1  
## 7 6.5 nonsmoker 1  
## 8 8.25 nonsmoker 1  
## 9 7.81 nonsmoker 1  
## 10 8.31 nonsmoker 1  
## # ... with 998,990 more rows

# calculate()

We next calculate the appropriate summary statistics of the null distribution, and the observed statistics.

null\_dist <- ncbirths %>%   
 specify(weight ~ habit) %>%   
 hypothesize(null = "independence") %>%   
 generate(reps = 1000, type = "permute") %>%   
 calculate(stat = "diff in means", order = c("nonsmoker", "smoker"))

## Warning: Removed 1 rows containing missing values.

observed\_statistic <-ncbirths %>%   
 specify(weight ~ habit) %>%   
 calculate(stat = "diff in means", order = c("nonsmoker", "smoker"))

## Warning: Removed 1 rows containing missing values.

observed\_statistic

## Response: weight (numeric)  
## Explanatory: habit (factor)  
## # A tibble: 1 x 1  
## stat  
## <dbl>  
## 1 0.316

x <- t.test(weight ~ habit, data=ncbirths)

x\_1 <- x$conf.int[[1]]

We construct the hypothesis test, and use t.test() to examine whether the mean difference between the nonsmoker and smoker mothers is statistically significant, and calculate confidence interval for the difference between the two.

obs\_diff\_prop <- ncbirths %>%   
 specify(weight ~ habit) %>%   
 calculate(stat = "diff in means", order = c("nonsmoker", "smoker"))

## Warning: Removed 1 rows containing missing values.

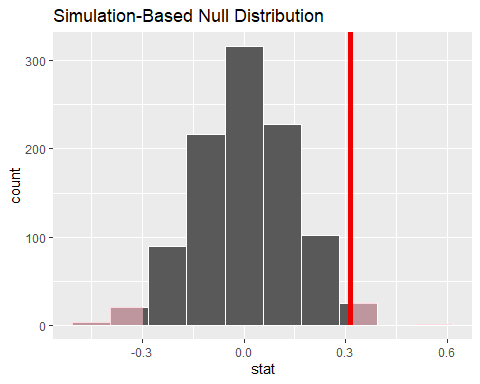
null\_dist %>%   
 get\_p\_value(obs\_stat = observed\_statistic, direction = "right")

## # A tibble: 1 x 1  
## p\_value  
## <dbl>  
## 1 0.015

# visualize()

visualize(null\_dist, bins = 10) +   
 shade\_p\_value(obs\_stat = observed\_statistic, direction = "two-sided")

## Warning: F usually corresponds to right-tailed tests. Proceed with caution.



# get\_confidence\_interval()

bootstrap\_distribution <- ncbirths %>%   
 specify(weight ~ habit) %>%   
 generate(reps = 1000, type = "bootstrap") %>%   
 calculate(stat = "diff in means", order = c("nonsmoker", "smoker"))

## Warning: Removed 1 rows containing missing values.

percentile\_ci <- bootstrap\_distribution %>%   
 get\_confidence\_interval(level = 0.95, type = "percentile")

Plot the mean difference

Alternative

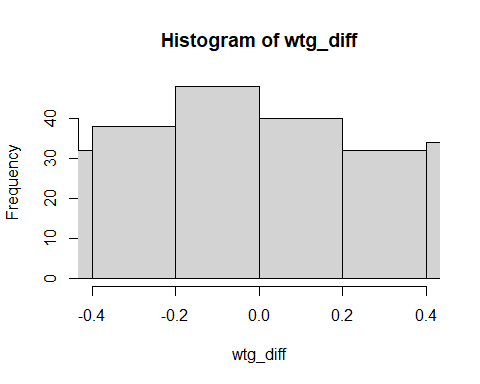
wtg\_diff <- nc\_nosmoker$weight - nc\_smoker$weight

## Warning in nc\_nosmoker$weight - nc\_smoker$weight: longer object length is not a  
## multiple of shorter object length

#wtg\_diff <- wtg\_diff[!is.na(wtg\_diff)] # remove missing entries (NA values)  
  
n <- length(ncbirths)  
  
wtg\_diff

## [1] 0.07 1.00 0.82 0.56 -0.68 -2.68 1.00 -3.19 2.62 5.25 0.69 1.87  
## [13] -2.38 -0.18 -6.44 -0.13 0.00 -1.25 0.50 0.75 1.32 2.00 -1.88 -1.56  
## [25] 0.06 2.18 1.93 4.31 -3.13 -4.81 0.00 -1.88 -6.75 2.38 -2.63 1.00  
## [37] 1.06 0.07 -0.44 0.12 -1.13 1.44 -0.06 1.62 -1.75 2.00 0.12 -1.00  
## [49] -0.32 -0.31 1.56 1.75 -0.94 -1.18 0.88 2.93 1.56 -1.07 0.00 -1.44  
## [61] 1.43 -1.31 -0.56 -1.06 -0.50 0.57 -1.56 1.94 -0.56 1.43 4.75 -0.31  
## [73] -1.37 0.37 -1.07 -0.37 0.81 -1.00 -1.00 1.75 -0.57 -0.81 -2.06 -2.18  
## [85] 0.57 -1.50 -4.12 1.13 -6.87 -0.94 -0.69 -2.69 3.19 6.63 -0.38 -0.06  
## [97] -2.19 0.75 0.37 1.75 -1.06 -0.13 2.00 1.19 0.69 -3.13 1.19 -1.12  
## [109] 0.43 1.94 1.56 3.00 -0.31 -0.38 1.13 1.00 0.75 1.75 1.62 0.31  
## [121] -1.56 -1.44 5.94 0.32 0.38 1.38 0.88 1.43 1.44 -1.38 -1.31 -1.31  
## [133] 0.62 1.12 0.06 5.37 1.25 1.37 -0.63 0.13 0.50 -1.44 -1.69 -1.87  
## [145] 0.94 -0.75 -1.37 -1.37 0.31 -0.68 -0.57 0.62 5.43 4.87 -0.25 -0.06  
## [157] -0.44 -2.13 -1.37 1.88 -2.25 1.69 1.38 0.57 0.87 1.00 0.69 1.75  
## [169] -2.37 0.12 -0.32 0.44 1.06 -0.62 -0.07 -0.43 3.31 -0.18 0.63 -0.31  
## [181] 2.13 1.56 -1.75 -2.32 -2.63 2.50 -0.32 -5.00 0.94 1.75 -1.37 -0.06  
## [193] -2.56 2.25 0.25 0.87 3.88 0.25 -0.43 -1.63 0.37 1.13 0.56 1.25  
## [205] 0.25 1.44 -1.69 -1.31 0.69 0.13 0.00 -1.69 -1.56 2.25 -0.56 2.75  
## [217] 0.25 1.06 1.82 3.75 -0.19 2.87 -1.31 2.31 0.06 -2.00 1.50 -0.50  
## [229] 0.81 2.38 1.94 -1.88 1.31 -2.06 -3.44 -1.00 -0.06 -1.75 -7.19 -1.94  
## [241] 2.57 2.37 1.12 -0.25 1.06 3.00 -0.62 0.56 4.50 1.82 -0.87 3.50  
## [253] 0.19 -0.82 1.19 -2.06 0.25 0.38 0.19 0.56 0.69 6.69 1.63 0.69  
## [265] -0.57 0.88 -0.63 -0.88 1.06 -1.12 1.12 3.44 -5.12 -1.62 1.81 -2.81  
## [277] -5.94 3.18 1.00 2.75 -1.00 -1.94 0.06 -3.25 -1.06 3.13 -2.69 1.63  
## [289] 0.44 2.25 1.18 -0.57 1.25 1.31 -0.18 1.31 0.62 -0.43 0.00 2.25  
## [301] -1.63 1.07 -2.00 0.13 1.63 0.25 0.56 0.12 1.31 0.18 -0.32 -0.25  
## [313] -1.19 0.00 2.19 1.07 0.63 -0.25 -1.87 1.00 1.31 1.37 2.75 0.50  
## [325] -1.50 -1.63 1.00 -2.25 0.00 -1.50 1.50 0.69 -0.75 -0.31 1.25 -1.25  
## [337] -0.06 -0.63 0.06 3.13 -0.56 0.50 -0.12 -1.19 3.88 5.44 0.87 1.69  
## [349] -4.75 3.87 2.25 1.38 -0.44 0.56 1.00 2.06 1.06 -2.31 1.75 0.57  
## [361] 1.25 3.56 1.69 2.44 -2.56 -0.75 2.69 2.31 2.31 -0.06 1.37 -1.69  
## [373] -5.81 -2.06 4.69 1.13 -0.81 3.50 -0.31 0.31 1.13 -0.31 0.50 0.19  
## [385] 0.94 -1.13 -0.81 5.44 2.00 0.81 0.06 0.69 0.25 -0.44 1.19 1.06  
## [397] 2.19 2.81 0.75 -0.94 0.75 -0.75 -1.50 2.06 3.31 4.06 -0.25 0.06  
## [409] 0.18 -2.13 -0.62 2.94 -6.38 2.69 -2.81 -1.31 -2.63 0.62 1.06 2.00  
## [421] -1.18 0.37 -0.94 -0.06 -0.57 -0.87 -3.57 1.19 3.13 -1.06 -0.44 0.88  
## [433] 0.56 2.93 1.94 -1.07 1.37 0.37 1.12 -1.75 2.69 0.88 1.75 -0.50  
## [445] -1.06 0.62 3.13 1.12 3.06 -0.81 0.38 -7.75 -3.07 0.07 -1.82 -0.31  
## [457] -0.50 2.13 -1.07 -0.43 1.31 -0.18 1.44 -0.75 1.38 2.75 -0.25 0.69  
## [469] 0.31 -0.38 3.38 3.75 -0.19 1.56 -0.13 3.06 2.19 -0.75 -0.06 -1.07  
## [481] 2.06 0.44 1.06 -0.94 0.38 0.07 -2.57 2.13 1.81 2.81 -0.44 -1.00  
## [493] 1.88 2.06 0.06 -0.68 -3.50 -3.32 -1.25 -5.00 5.12 0.07 0.63 2.38  
## [505] 1.38 1.18 1.88 1.12 0.07 1.75 -0.63 0.43 2.25 5.25 1.75 0.94  
## [517] -0.44 -0.31 -0.13 -3.00 0.06 -2.06 1.19 1.25 0.69 -0.37 0.25 0.50  
## [529] -0.88 2.06 2.18 5.81 -0.38 -2.44 -0.94 -2.32 -2.00 4.69 -0.38 1.13  
## [541] 1.31 0.00 0.18 -1.44 1.00 1.38 -6.81 -0.38 -2.38 1.94 -1.07 0.06  
## [553] 0.06 1.32 1.00 0.82 -1.69 2.69 1.88 -1.57 1.56 -1.00 0.43 -1.75  
## [565] 1.31 0.38 2.25 5.94 0.88 0.19 -2.18 1.69 1.88 0.43 2.75 -3.81  
## [577] -0.25 -3.07 -4.19 -1.31 -1.88 0.94 0.75 0.56 -2.32 0.63 2.44 -0.12  
## [589] -0.43 -0.57 0.81 3.13 -0.25 -0.88 -0.87 -0.50 3.57 4.19 -1.38 -0.44  
## [601] -3.94 2.37 2.19 1.44 1.56 -0.25 0.75 0.38 3.00 -2.06 2.81 0.82  
## [613] -1.88 3.25 1.94 5.94 0.62 -2.63 0.69 1.75 2.00 0.13 1.68 -5.57  
## [625] -4.62 -2.19 3.75 -0.81 0.44 -0.75 -1.43 2.50 1.32 -0.06 0.38 -2.25  
## [637] 1.69 -1.63 2.31 9.94 0.00 -1.38 -2.38 -0.62 1.56 0.50 3.00 1.06  
## [649] 0.69 3.63 -0.12 1.31 -0.13 -1.68 -2.38 3.87 2.56 4.12 -4.69 -2.81  
## [661] -0.82 -4.69 -6.87 -3.56 -3.57 1.31 2.56 -1.25 0.12 1.06 0.37 0.44  
## [673] -3.87 1.68 0.87 1.07 0.25 1.06 0.68 -1.31 1.13 0.75 -0.81 0.00  
## [685] 1.13 3.50 0.69 0.75 0.93 0.00 0.37 -0.19 3.25 0.75 2.81 0.57  
## [697] -0.75 -0.06 2.13 2.18 3.38 1.63 2.75 -2.00 -0.32 -1.68 -0.82 -0.62  
## [709] 0.94 1.75 -0.25 -1.87 -0.37 0.94 -1.00 -0.50 0.56 2.50 -1.37 1.06  
## [721] 1.50 -3.19 4.69 3.13 -0.69 1.56 -0.44 3.62 1.50 -0.06 2.00 1.25  
## [733] 0.56 1.25 3.19 -2.88 0.75 -0.18 0.62 4.56 1.94 3.19 0.12 -1.25  
## [745] 1.25 -0.25 2.31 0.13 0.93 -2.19 -3.69 -5.31 7.94 1.00 1.81 1.44  
## [757] 1.25 1.56 2.32 0.19 1.38 0.82 0.81 0.00 1.12 6.19 0.75 -1.75  
## [769] -0.32 1.63 -0.25 -1.75 1.44 -0.69 -0.31 1.00 0.82 0.44 -0.94 -2.62  
## [781] 1.31 0.25 -1.88 4.06 -6.06 1.44 0.68 -0.32 -0.75 4.13 -2.44 -0.12  
## [793] 0.94 -1.06 0.75 -0.82 -0.06 1.00 0.75 1.81 -1.38 1.82 2.00 1.31  
## [805] -1.13 0.94 2.25 2.94 0.19 0.00 -2.50 -2.13 -3.06 -3.00 3.00 -1.19  
## [817] 0.68 1.25 4.31 1.19 0.69 1.32 -0.56 3.00 -0.19 0.12 2.50 -0.06  
## [829] -1.06 -1.19 1.18 1.19 1.00 -2.81 -1.31 -1.94 -5.25 0.19 0.25 -1.31  
## [841] 1.13 0.00 0.56 3.00 -1.00 2.69 -5.62 -2.44 4.00 4.50 -0.19 1.50  
## [853] -1.38 5.00 3.25 0.32 2.50 0.56 -0.69 1.81 0.19 -2.81 1.44 -1.18  
## [865] -0.44 2.81 0.62 1.25 0.25 -2.94 0.63 -1.88 1.50

hist(wtg\_diff, breaks=100, xlim=c(-0.4,0.4))



### infer

In infer package provides a series of functions that provides more clarity to the process. Under the hood, these functions are wrapper functions to the base R functions we used in the previous [section](#base).

Null distribution:

set.seed(42)  
null <- ncbirths %>%   
 specify(weight ~ habit) %>%  
 hypothesize(null = "independence") %>%  
 generate(reps = 1000, type = "permute") %>%  
 calculate(stat = "diff in means", order = c("nonsmoker", "smoker"))

observed <- ncbirths %>%   
 specify(weight ~ habit) %>%  
 calculate(stat = "diff in means", order = c("nonsmoker", "smoker"))

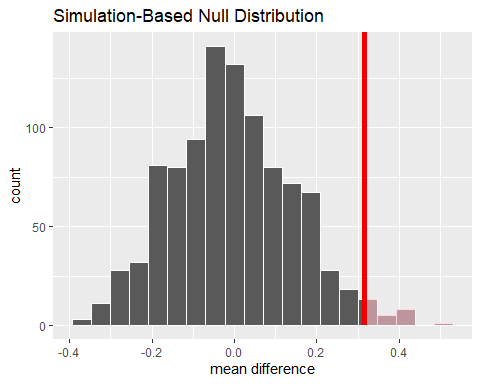
p-value:

null %>%   
 get\_p\_value(obs\_stat = observed, direction = "right")

## # A tibble: 1 x 1  
## p\_value  
## <dbl>  
## 1 0.022

Visualization p-value

visualize(null, bins = 20) +   
 shade\_p\_value(obs\_stat = observed, direction = "right") + labs(x="mean difference")



Or using wrapper function

t\_test(x = ncbirths,   
 formula = weight ~ habit,   
 order = c("nonsmoker", "smoker"),  
 alternative = "two-sided")

## # A tibble: 1 x 7  
## statistic t\_df p\_value alternative estimate lower\_ci upper\_ci  
## <dbl> <dbl> <dbl> <chr> <dbl> <dbl> <dbl>  
## 1 2.36 171. 0.0195 two.sided 0.316 0.0515 0.580

# Next Part

As students, we were overwhelmed by the use of p-values, various statistical tests, and the corresponding interpretations when hypothesis testing was first introduced. It only gets better once students acquire the realization and intuition of the [universal pattern](http://allendowney.blogspot.com/2016/06/there-is-still-only-one-test.html) that statistical tests follow through advanced courses.

Moreover, simulation-based inferences are often more visual and intuitive than analytical inferences. While many educators include those numerical approaches in their teaching, it often feels segmented, especially in base R code.

We hope this blog post could provide students’ perspectives on statistical education and shed light on the discussion of modernizing “Intro to Statistics” courses.

Follow-up tidymodel.