hw\_7

Iastre

17 09 2020

Посмотрим на таблицу

library(readxl)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────────────────────────────────────────────────────────────────── tidyverse 1.3.0 ──

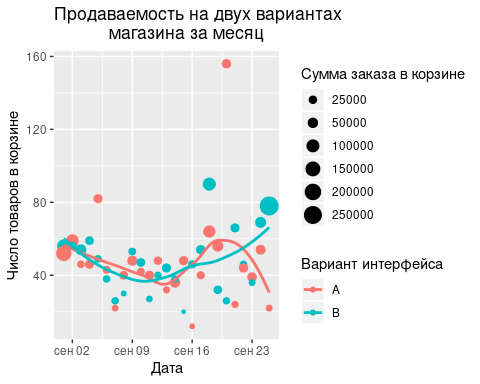
## ✓ ggplot2 3.2.1 ✓ purrr 0.3.3  
## ✓ tibble 2.1.3 ✓ dplyr 0.8.4  
## ✓ tidyr 1.0.2 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.4.0

## ── Conflicts ────────────────────────────────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

df <- read\_xlsx("Ювелирный магазин, данные.xlsx")  
df

## # A tibble: 50 x 4  
## date Variant `cnt(addToCart)` `sum(addToCartItems)`  
## <dttm> <dbl> <dbl> <dbl>  
## 1 2019-09-01 00:00:00 90 56 110247  
## 2 2019-09-01 00:00:00 10 52 163098  
## 3 2019-09-02 00:00:00 10 59 93317  
## 4 2019-09-02 00:00:00 90 56 26457  
## 5 2019-09-03 00:00:00 90 54 56418  
## 6 2019-09-03 00:00:00 10 46 16985  
## 7 2019-09-04 00:00:00 10 46 31152  
## 8 2019-09-04 00:00:00 90 59 27379  
## 9 2019-09-05 00:00:00 90 49 17503  
## 10 2019-09-05 00:00:00 10 82 32306  
## # … with 40 more rows

col = factor(df$Variant, labels=c("A","B"))  
ggplot(df, aes(y=`cnt(addToCart)`, x=date, color=col)) +   
 geom\_point(aes(size=`sum(addToCartItems)`)) +  
 geom\_smooth(se=FALSE, method="loess") +   
 scale\_size\_continuous(breaks = sort(as.numeric(c(25000,50000,100000,150000,200000, 250000)))) +   
 labs(y="Число товаров в корзине",  
 x = "Дата",  
 title = "Продаваемость на двух вариантах \n магазина за месяц",  
 color = "Вариант интерфейса",  
 size = "Сумма заказа в корзине")



A <- df$`sum(addToCartItems)`[df$Variant != 90]  
B <- df$`sum(addToCartItems)`[df$Variant == 90]  
suc.A <- c()  
suc.B <- c()  
for (x in 1:(length(df$`sum(addToCartItems)`)/2)){  
 if (A[x] > B[x]){  
 suc.A <- c(suc.A, 1)   
 suc.B <- c(suc.B, 0)  
 } else if (B[x] < A[x]){  
 suc.A <- c(suc.A, 0)   
 suc.B <- c(suc.B, 1)  
 } else {  
 suc.A <- c(suc.A, 0)   
 suc.B <- c(suc.B, 0)  
 }  
}  
sum.suc.A <- sum(suc.A)  
sum.suc.B <- sum(suc.B)  
len\_A <- length(A)  
len\_B <- length(B)  
prop.test(c(sum.suc.A,sum.suc.B), c(len\_A, len\_B), alternative="greater")

##   
## 2-sample test for equality of proportions with continuity correction  
##   
## data: c(sum.suc.A, sum.suc.B) out of c(len\_A, len\_B)  
## X-squared = 18.667, df = 1, p-value = 7.784e-06  
## alternative hypothesis: greater  
## 95 percent confidence interval:  
## 0.3988379 1.0000000  
## sample estimates:  
## prop 1 prop 2   
## 0.6 0.0

Пусть

N <- 1000  
  
conv\_A <- mean(suc.A)  
conv\_B <- mean(suc.B)  
p <- conv\_B - conv\_A  
cat("разница между выборками равна :", p, "\n")

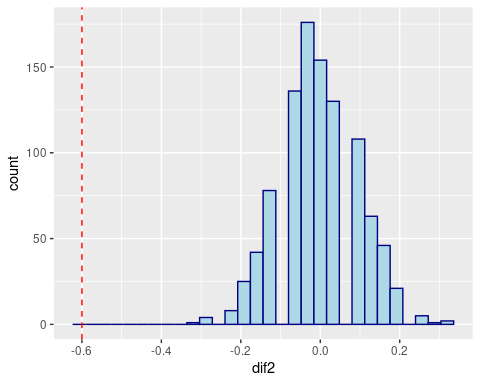
## разница между выборками равна : -0.6

differences <- rep(NA, N)  
for (i in 1:N){  
 s1 <- sample(suc.A, replace = TRUE)  
 s2 <- sample(suc.B, replace = TRUE)  
 p1 <- mean(s1)  
 p2 <- mean(s2)  
 dif <- p2 -p1  
 differences[i] <- dif  
}  
head(differences)

## [1] -0.52 -0.56 -0.56 -0.64 -0.72 -0.68

dif2 <- differences - mean(differences)  
dif\_df2 <- data.frame(dif2)  
ggplot(dif\_df2, aes(x=dif2)) + geom\_histogram(fill = "lightblue", color="navy") + geom\_vline(xintercept = p, color = "red", lty=2)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



conv\_A

## [1] 0.6