

Two-Sample Permutation Tests in R

Math 445, Spring 2017

Beer Consumption ██████████ Human Attractiveness to Malaria Mosquitoes

Thierry Lefèvre^{1*}, Louis-Clément Gouagna^{2,3}, Kounbobr Roch Dabiré^{3,4}, Eric Elguero¹, Didier Fontenille², François Renaud¹, Carlo Costantini^{2,5}, Frédéric Thomas^{1,6}

1 Génétique et Evolution des Maladies Infectieuses, UMR CNRS/IRD 2724, Montpellier, France, **2** Caractérisation et Contrôle des Populations de Vecteurs, IRD/UR 016, Montpellier, France, **3** Institut de Recherche en Science de la Santé, Bobo-Dioulasso, Burkina Faso, **4** Laboratoire de Parasitologie et d'Entomologie Médicale, Centre Muraz, Bobo-Dioulasso, Burkina Faso, **5** Organisation de Coordination pour la lutte contre les Endémies en Afrique Centrale, Yaoundé, Cameroun, **6** Institut de Recherche en Biologie Végétale, Université de Montréal, Montréal, Canada

Abstract

Background: Malaria and alcohol consumption both represent major public health problems. Alcohol consumption is rising in developing countries and, as efforts to manage malaria are expanded, understanding the links between malaria and alcohol consumption becomes crucial. Our aim was to ascertain the effect of beer consumption on human attractiveness to malaria mosquitoes in semi field conditions in Burkina Faso.

Methodology/Principal Findings: We used a Y tube-olfactometer designed to take advantage of the whole body odour (breath and skin emanations) as a stimulus to gauge human attractiveness to *Anopheles gambiae* (the primary African malaria vector) before and after volunteers consumed either beer (n = 25 volunteers and a total of 2500 mosquitoes tested) or water (n = 18 volunteers and a total of 1800 mosquitoes). Water consumption had no effect on human attractiveness to *An. gambiae* mosquitoes, but beer consumption increased volunteer attractiveness. Body odours of volunteers who consumed beer increased mosquito activation (proportion of mosquitoes engaging in take-off and up-wind flight) and orientation (proportion of mosquitoes flying towards volunteers' odours). The level of exhaled carbon dioxide and body

Figure 1:

Results

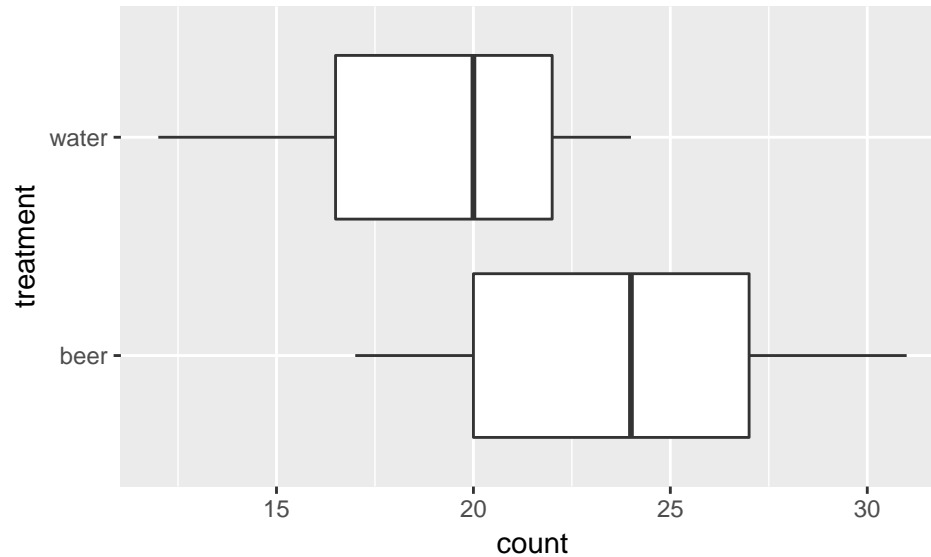
```
mosquitoes <- read.csv("../data/mosquitoes_beer.csv")
head(mosquitoes)
```

```
##   treatment count
## 1      beer    27
## 2      beer    19
## 3      beer    20
## 4      beer    20
## 5      beer    23
## 6      beer    17
```

```
summary(mosquitoes)
```

```
## treatment      count
## beer :25   Min.    :12.00
## water:18   1st Qu.:19.00
##           Median  :21.00
##           Mean    :21.77
##           3rd Qu.:24.00
##           Max.    :31.00
```

```
library(ggplot2)
ggplot(mosquitoes, aes(x = treatment, y = count)) +
  geom_boxplot() +
  coord_flip()
```



```
library(dplyr)
trt_means <-
  mosquitoes %>%
  group_by(treatment) %>%
  summarise(avg = mean(count))
trt_means
```

```
## # A tibble: 2 × 2
##   treatment      avg
##   <fctr>      <dbl>
## 1     beer 23.60000
## 2     water 19.22222
```

Key question:

Are the treatment groups really different or are the differences due to random chance?

Permutation test in R

I assume that you already have the data loaded along with `dplyr` and `ggplot2`...

Calculating the observed difference in means

```
trt_means <-
  mosquitoes %>%
  group_by(treatment) %>%
  summarise(avg = mean(count))
trt_means
```

```
## # A tibble: 2 × 2
##   treatment    avg
##   <fctr>    <dbl>
## 1    beer 23.60000
## 2    water 19.22222

observed <- diff(trt_means$avg)
observed

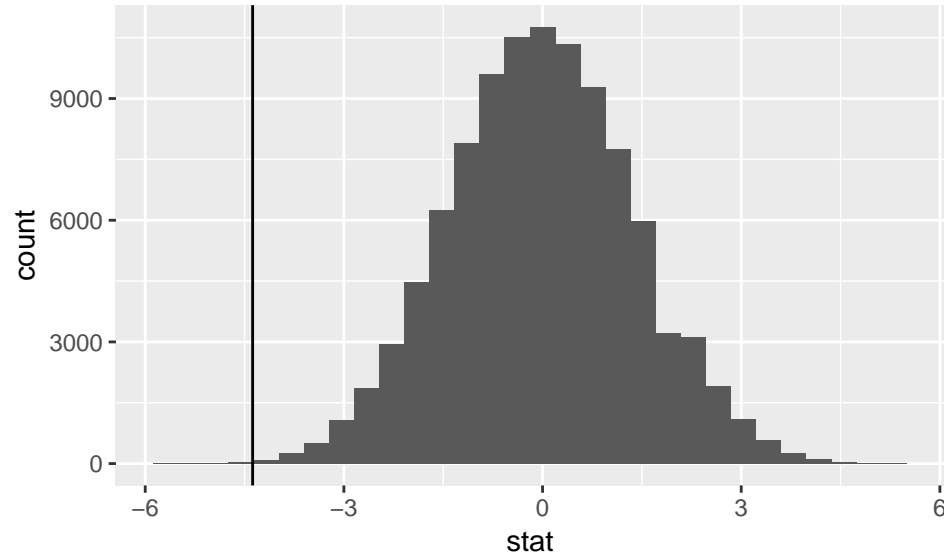
## [1] -4.377778
```

Option 1: Using a for loop to run the permutations

```
N <- 10^5 - 1
result <- numeric(N)
for(i in 1:N) {
  index <- sample(nrow(mosquitoes), size = 25, replace = FALSE)
  result[i] <- mean(mosquitoes$count[index]) - mean(mosquitoes$count[-index])
}
```

Plotting the null dsu

```
null_dsn <- data.frame(stat = result)
ggplot(data = null_dsn) +
  geom_histogram(mapping = aes(x = stat)) +
  geom_vline(xintercept = observed)
```



Calculating the p-value

```
(sum(result <= observed) + 1) / (N + 1)

## [1] 0.00056
```

Option 2: Using replicate to run the permutations

```
N <- 105 - 1
```

Writing a function that runs the permutations

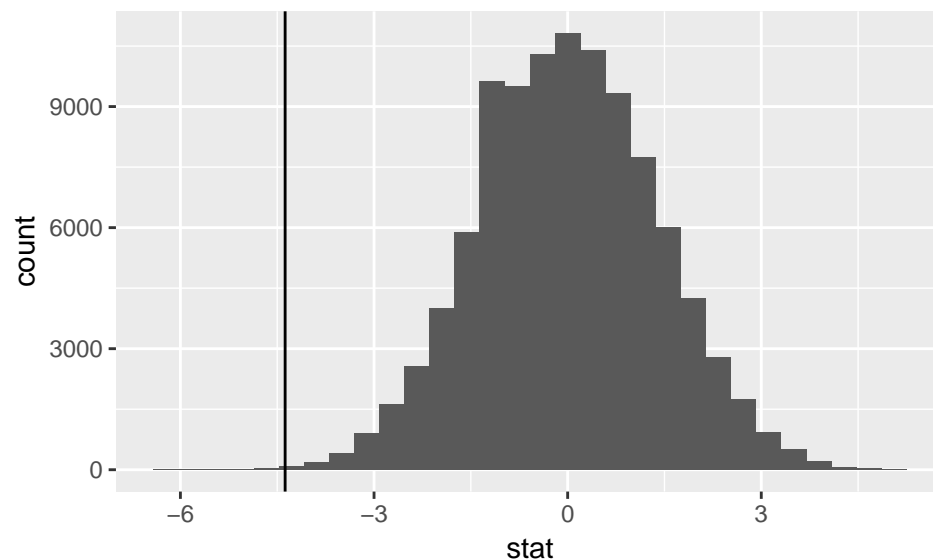
```
permute <- function(values, ngrp1) {  
  index <- sample(length(values), size = ngrp1, replace = FALSE)  
  stat <- mean(values[index]) - mean(values[-index])  
  return(stat)  
}
```

Running the permutations

```
result <- replicate(N, permute(values = mosquitoes$count, ngrp1 = 25))
```

Plotting the null dsn

```
null_dsn <- data.frame(stat = result)  
ggplot(data = null_dsn) +  
  geom_histogram(mapping = aes(x = stat)) +  
  geom_vline(xintercept = observed)
```



Calculating the p-value

```
(sum(result <= observed) + 1) / (N + 1)
```

```
## [1] 0.00063
```