

Pruebas de Hipotesis

Camila Silva

2025-01-20

PRUEBA DE HIPOTESIS

```
library("rlang")
library("phyloseq")
library("ggplot2")
library("vegan")

## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.6-8
library("RColorBrewer")
library("stringi")
library("dplyr")

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
library("plyr")

## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
##
## Attaching package: 'plyr'
## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
setwd("/home/camila/GIT/Tesis_Maestria/Data/fresa_solena/Data1")
outpath = "/home/camila/GIT/Tesis_Maestria/Analisis_Comparativo/Fresa_Solena/Results_img"
```

```

### Cargado de datos originales
fresa_kraken <- import_biom("fresa_kraken.biom")
colnames(fresa_kraken@tax_table@.Data) <- c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus", "Species")
fresa_kraken@tax_table@.Data <- substr(fresa_kraken@tax_table@.Data,4,100)
colnames(fresa_kraken@otu_table@.Data) <- substr(colnames(fresa_kraken@otu_table@.Data),1,6)
metadata_fresa <- read.csv2("/home/camila/GIT/Tesis_Maestria/Data/fresa_solena/Data1/metadata.csv",header=1)
fresa_kraken@sam_data <- sample_data(metadata_fresa)
fresa_kraken@sam_data$Sample<-row.names(fresa_kraken@sam_data)
colnames(fresa_kraken@sam_data)<-c('Treatment','Samples')
samples_to_remove <- c("MP2079","MP2080","MP2088","MP2109","MP2137")
fresa_kraken_fil <- prune_samples(!(sample_names(fresa_kraken) %in% samples_to_remove), fresa_kraken)
percentages_fil <- transform_sample_counts(fresa_kraken_fil, function(x) x*100 / sum(x) )
percentages_df <- psmelt(percentages_fil)

OTU <- fresa_kraken_fil@otu_table@.Data
SAM <- fresa_kraken_fil@sam_data

## UNIR TABLA DE ABUNDANCIAS CON METADATA DE SANOS Y ENFERMOS

## Calculamos la diversidad Shannon

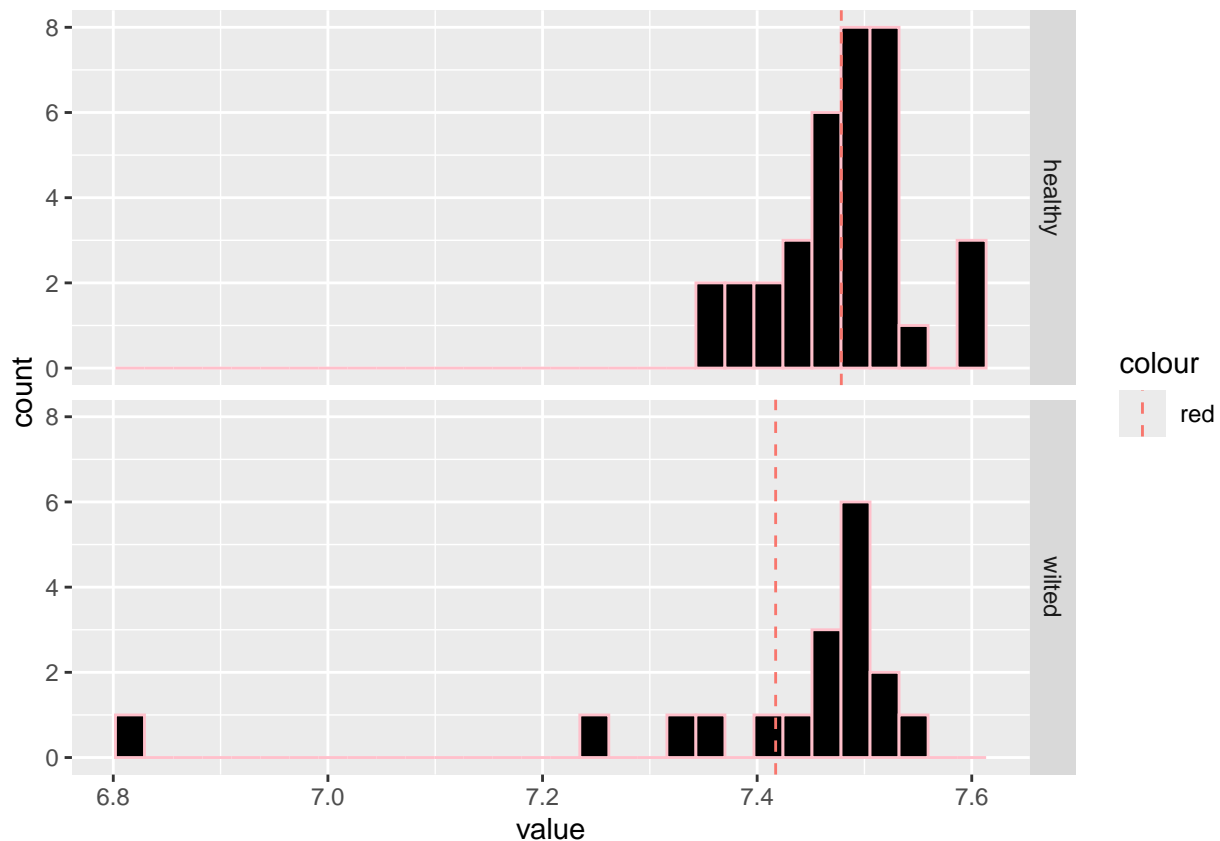
## Se usa la funcion diversidad del paquete vegan para calcular el indice Shannon
## Se realiza el dataframe del indice de Shannon
OTU <- t(OTU)
Shannon_OTU <- diversity(OTU, "shannon")
Shannon_OTU_df <- data.frame(sample=names(Shannon_OTU),value=Shannon_OTU,measure=rep("Shannon", length(Shannon_OTU)))
total <-cbind(Shannon_OTU_df,SAM)

# media por grupos
mu <- ddply(total, "Treatment", summarise, grp.mean=mean(value))

p<-ggplot(total, aes(x=value))+
  geom_histogram(color="pink",fill="black")+
  facet_grid(Treatment ~ .)
p+geom_vline(data=mu, aes(xintercept=grp.mean, color="red"),
  linetype="dashed")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

```



```
ggsave("ShannonDistribucion.png", plot = last_plot(), path = "/home/camila/GIT/Tesis_Maestria/Analisis_")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

El gráfico muestra la distribución del índice de diversidad de Shannon para dos grupos de tratamiento diferentes: “healthy” (sano) y “wilted” (marchito), con líneas verticales indicando la media de cada grupo.

Análisis del gráfico:

Ejes del gráfico: Eje X (value): Representa los valores del índice de Shannon, que mide la diversidad de microorganismos. Valores más altos indican una comunidad microbiana más diversa. Eje Y (count): Indica la frecuencia (número de muestras) para cada rango de valores de Shannon. Paneles por tratamiento:

El gráfico se divide en dos paneles: Arriba: “healthy” (sano). Abajo: “wilted” (marchito). Distribución dentro de los tratamientos:

En el grupo “healthy”, la diversidad microbiana parece estar concentrada en valores más altos de Shannon (aproximadamente entre 7.2 y 7.6), indicando una mayor diversidad en muestras saludables. En el grupo “wilted”, los valores de Shannon están distribuidos en un rango más bajo (entre 6.8 y 7.4), sugiriendo que las muestras marchitas tienen menor diversidad microbiana. Líneas verticales:

Las líneas rojas punteadas representan la media del índice de Shannon para cada tratamiento: Para “healthy”, la media está cerca de 7.4. Para “wilted”, la media está más baja, alrededor de 7.2. Interpretación biológica: El índice de Shannon más alto en las muestras saludables podría indicar que una mayor diversidad microbiana está asociada con un mejor estado de salud de las plantas. Por otro lado, la menor diversidad en muestras marchitas podría estar relacionada con un desequilibrio o dominancia de ciertas especies microbianas perjudiciales.