
CENTRO DE CIENCIAS MATEMÁTICAS

CINVESTAV

ENES Análisis univariado de comunidades

Capítulo 8

Andrés Arredondo (email_1@gmail.com)
Adriana Haydé Contreras Peruyero (haydeeperuyero@gmail.com)
David Alberto García Estrada (@gmail.com)

Morelia

Septiembre de 2022

Índice

1. Introducción	3
1.1. Análisis univariado de comunidades	3

1. Introducción

1.1. Análisis univariado de comunidades

Dividimos el estudio de la composición de las comunidades microbianas en dos grandes componentes: (a) Evaluación de hipótesis sobre diversidad taxonómica, OTU y Taxones (b) Análisis de diferencias entre grupos. El primer componente pertenece principalmente a análisis univariado de comunidades. El segundo puede ser dividido en varias técnicas multivariadas, como lo son “clustering” y “ordinations”, y la evaluación de hipótesis de análisis multivariado de dissimilitudes.

##8.1 Comparaciones de diversidades entre dos grupos

En nuestro estudio con ratones Vdr-/-, uno de los propósitos es probar la diferencia de diversidades entre dos grupos (Vdr-/- y ratones de tipo salvaje) en sitios fecales y cecales. Aquí ilustraremos el análisis de la comunidad univariante, y compararemos la diversidad de Shannon (calculado anteriormente en Cap. 6 en las muestras fecales) usando varias estadísticas de prueba.

##8.1.1 Prueba t de Welch para dos muestras

El estadístico t fue introducido en 1908 por William Sealy Gosset. Una prueba t de dos muestras se utiliza para probar que las medias de dos poblaciones son iguales. Se aplica más comúnmente cuando el estadístico de prueba seguiría una distribución normal. Si los dos grupos tienen la misma varianza, el estadístico t se puede calcular de la siguiente manera:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

donde, s_p y s^2 son el estimador imparcial de la varianza de las muestras 1 y 2, respectivamente. Cuando las dos muestras tienen varianzas desiguales y tamaños de muestra desiguales, la prueba t de Welch se considera más confiable (Ruxton 2006). Por lo tanto, aquí usamos la prueba t de Welch para nuestros datos de ratón Vdr-/- . Primero, cargue y transponga el conjunto de datos:

```
# IMPORTANTE
```

```
# Ajustamos path de trabajo según tu PC
```

```
path <- "C:/Users/ANDRESARREDONDOCRUZ/"
```

```
path <- paste0(path, "Equipo4/")
```

```
setwd(path)
```

```
abund_table = read.csv(paste0(path, "data/VdrGenusCounts.csv"), row.names=1, check.names=FALSE)
abund_table <- t(abund_table)
```

Para incorporar la información del grupo del conjunto de datos directamente a la comparación, necesitamos administrar los datos. En el conjunto de datos, la información de id de muestra y grupo está en una franja de caracteres. Primero los extraemos de allí.

```
grouping <- data.frame(row.names=row.names(abund_table), t(as.data.frame(strsplit(row.names(abund_table), "_"),
grouping$Location <- with(grouping, ifelse(X3 %in% "drySt-28F", "Fecal", "Cecal"))
grouping$Group <- with(grouping, ifelse(as.factor(X2) %in% c(11,12,13,14,15), c("Vdr-/-"), c("WT")))
grouping <- grouping[,c(4,5)]
grouping
```

	Location	Group
5_15_drySt-28F	Fecal	Vdr-/-
20_12_CeSt-28F	Cecal	Vdr-/-

	Location	Group
1_11_drySt-28F	Fecal	Vdr-/-
2_12_drySt-28F	Fecal	Vdr-/-
3_13_drySt-28F	Fecal	Vdr-/-
4_14_drySt-28F	Fecal	Vdr-/-
7_22_drySt-28F	Fecal	WT
8_23_drySt-28F	Fecal	WT
9_24_drySt-28F	Fecal	WT
19_11_CeSt-28F	Cecal	Vdr-/-
21_13_CeSt-28F	Cecal	Vdr-/-
22_14_CeSt-28F	Cecal	Vdr-/-
23_15_CeSt-28F	Cecal	Vdr-/-
25_22_CeSt-28F	Cecal	WT
26_23_CeSt-28F	Cecal	WT
27_24_CeSt-28F	Cecal	WT

Repetimos el calculo de la diversidad de Shannon para est tabla, igual que en el capitulo 6.

```
#library(vegan)
H<-diversity(abund_table, "shannon")
```

Luego combinamos dataframes de diversidad y agrupación para crear un nuevo dataframe.

```
df_H<-data.frame(sample=names(H),value=H,measure=rep("Shannon",length(H)))
df_G <-cbind(df_H, grouping)
rownames(df_G)<-NULL
df_G
```

sample	value	measure	Location	Group
5_15_drySt-28F	2.460729	Shannon	Fecal	Vdr-/-
20_12_CeSt-28F	2.339725	Shannon	Cecal	Vdr-/-
1_11_drySt-28F	2.228023	Shannon	Fecal	Vdr-/-
2_12_drySt-28F	2.734405	Shannon	Fecal	Vdr-/-
3_13_drySt-28F	2.077282	Shannon	Fecal	Vdr-/-
4_14_drySt-28F	2.466830	Shannon	Fecal	Vdr-/-
7_22_drySt-28F	1.777171	Shannon	Fecal	WT
8_23_drySt-28F	1.999559	Shannon	Fecal	WT
9_24_drySt-28F	1.971996	Shannon	Fecal	WT
19_11_CeSt-28F	1.344813	Shannon	Cecal	Vdr-/-
21_13_CeSt-28F	2.016113	Shannon	Cecal	Vdr-/-
22_14_CeSt-28F	1.955432	Shannon	Cecal	Vdr-/-
23_15_CeSt-28F	1.614456	Shannon	Cecal	Vdr-/-
25_22_CeSt-28F	1.958839	Shannon	Cecal	WT
26_23_CeSt-28F	2.270818	Shannon	Cecal	WT
27_24_CeSt-28F	2.002195	Shannon	Cecal	WT

A continuación, creamos subconjuntos de datos fecales del nuevo dataframe.

```
Fecal_G<- subset(df_G, Location=="Fecal")
Fecal_G
```

	sample	value	measure	Location	Group
1	5_15_drySt-28F	2.460729	Shannon	Fecal	Vdr-/-
3	1_11_drySt-28F	2.228023	Shannon	Fecal	Vdr-/-
4	2_12_drySt-28F	2.734405	Shannon	Fecal	Vdr-/-
5	3_13_drySt-28F	2.077282	Shannon	Fecal	Vdr-/-
6	4_14_drySt-28F	2.466830	Shannon	Fecal	Vdr-/-
7	7_22_drySt-28F	1.777171	Shannon	Fecal	WT
8	8_23_drySt-28F	1.999559	Shannon	Fecal	WT
9	9_24_drySt-28F	1.971996	Shannon	Fecal	WT

Ahora los datos están listos para el análisis estadístico. Antes de realizar la prueba de hipótesis, exploremos la distribución de los valores de diversidad de Shannon usando la función `ggplot()`

```
#library(ggplot2)
#dividimos el gráfico en dos paneles usando facet_grid.
p<-ggplot(Fecal_G, aes(x=value))+
  geom_histogram(color="black", fill="black")+
  facet_grid(Group ~ .)
```

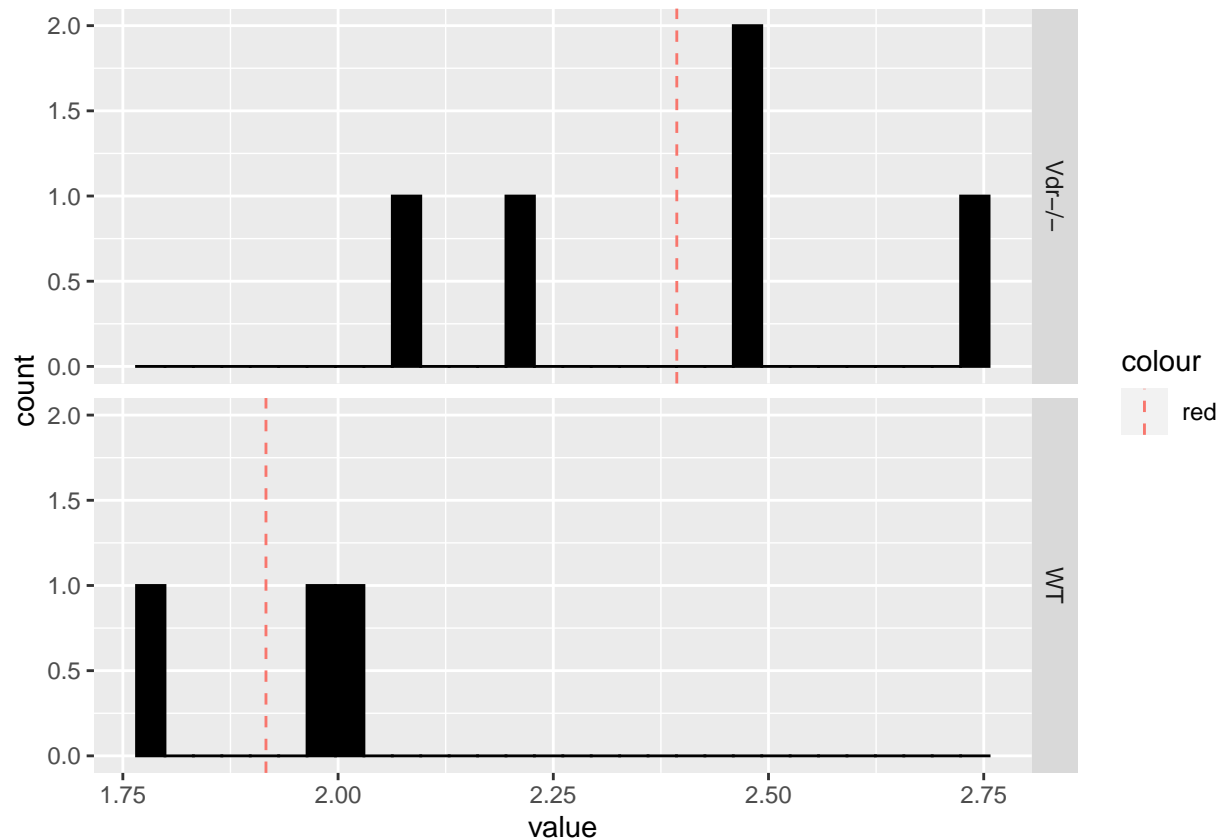
El paquete `plyr` se utiliza para calcular los valores promedio de diversidad de Shannon de cada grupo

```
#library(plyr)
#la función ddply toma input un df y arroja un df de output
mu <- ddply(Fecal_G, "Group", summarise, grp.mean=mean(value))
head(mu)
```

Group	grp.mean
Vdr-/-	2.393454
WT	1.916242

```
#Agregamos las líneas de la media a la gráfica
p+geom_vline(data=mu, aes(xintercept=grp.mean, color="red"),
  linetype="dashed")
```

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



En el histograma anterior la eliminación de Vdr de este grupo se desplaza hacia la derecha en relación con el grupo WT (hacia valores de diversidad más altos), lo que da como resultado una mayor diversidad.

#Usamos la prueba t de Welch para probar la hipótesis nula

```
fit_t <- t.test(value ~ Group, data=Fecal_G)
```

```
fit_t
```

```
##
```

```
## Welch Two Sample t-test
```

```
##
```

```
## data: value by Group
```

```
## t = 3.5999, df = 5.9206, p-value = 0.01163
```

```
## alternative hypothesis: true difference in means between group Vdr-/- and group WT is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## 0.1517841 0.8026392
```

```
## sample estimates:
```

```
## mean in group Vdr-/-      mean in group WT
```

```
##          2.393454          1.916242
```

Para probar la hipótesis nula de que no hay diferencia en la diversidad de Shannon, se utilizó una prueba t de Welch que dio como resultado un valor de $p = 0,01$ ($t = 3,6$, $df = 5,9$). Por lo tanto, rechazamos la hipótesis nula de no diferencia a favor de la alternativa de que las diversidades de Shannon son diferentes en los dos grupos.

##8.2 Comparisons of a Taxon of Interest between Two Groups