

ANOVAs

Pablo E. Gutiérrez-Fonseca

2023-10-02

Load libraries

```
library(vegan)
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.6-4
```

```
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(readxl)
```

```
# install.packages("ggplot2")
```

```
library(ggplot2)
```

Aquatic Macronvertebrates in two land uses (Agriculture and Urban streams)

```
data <- read_excel('data.xlsx', sheet = 'Sheet3')  
head(data,20)
```

```
## # A tibble: 20 x 4
```

```
##   site      sample  taxa      abundance
```

```
##   <chr>      <chr>  <chr>      <dbl>
```

```
## 1 Agriculture sample_3 Annelids 1
## 2 Agriculture sample_3 Tricoptera 1
## 3 Agriculture sample_2 Diptera 2
## 4 Agriculture sample_2 Decapoda 2
## 5 Agriculture sample_1 Tricoptera 4
## 6 Agriculture sample_2 Tricoptera 6
## 7 Agriculture sample_1 Plecoptera 15
## 8 Agriculture sample_3 Plecoptera 19
## 9 Agriculture sample_3 Ephemeroptera 21
## 10 Agriculture sample_1 Ephemeroptera 31
## 11 Agriculture sample_2 Coleoptera 33
## 12 Agriculture sample_2 Ephemeroptera 51
## 13 urban sample_1 Coleoptera 1
## 14 urban sample_3 Odonata 1
## 15 urban sample_3 Decapoda 1
## 16 urban sample_3 Sphaeriidae 1
## 17 urban sample_1 Diptera 2
## 18 urban sample_1 Tricoptera 2
## 19 urban sample_2 Diptera 2
## 20 urban sample_2 Megaloptera 2
```

```
# Assuming your data is stored in a data frame named 'data'
```

```
# Replace 'data' with your actual data frame name
```

```
# Perform ANOVA
```

```
anova_result <- aov(abundance ~ site + Error(sample), data = data)
```

```
# Display the ANOVA summary
```

```
summary(anova_result)
```

```
##
## Error: sample
##           Df Sum Sq Mean Sq F value Pr(>F)
## site       1 118.77  118.77   44.34 0.0949 .
## Residuals  1   2.68    2.68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## site       1  805.6   805.6   6.394 0.0184 *
## Residuals 24 3023.9   126.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Calculate total abundance and SD per site
```

```
abundance_summary <- data %>%
```

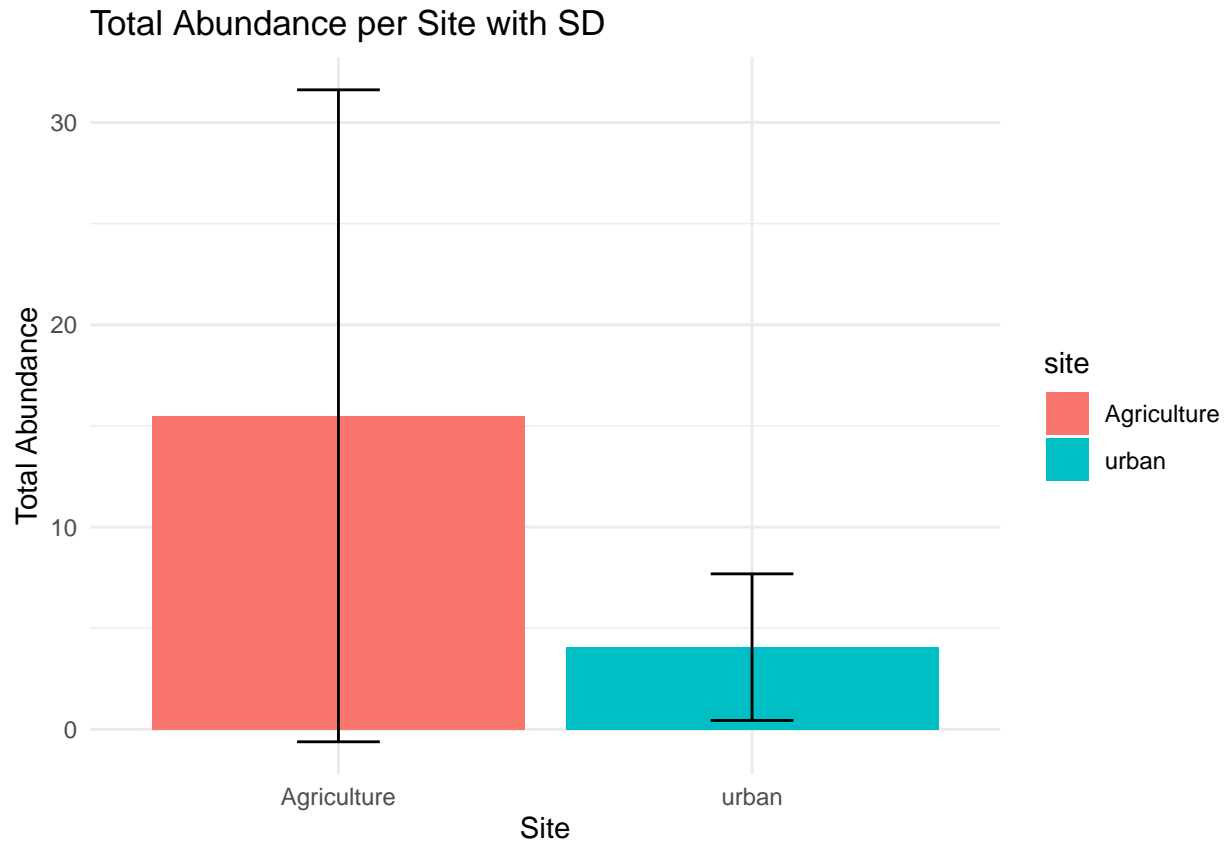
```
  group_by(site) %>%
```

```
  summarize(mean_abundance = mean(abundance),
            sd_abundance = sd(abundance))
```

```
# Create a bar plot with error bars
```

```
ggplot(abundance_summary, aes(x = site, y = mean_abundance, fill = site)) +
  geom_bar(stat = "identity") +
```

```
geom_errorbar(aes(ymin = mean_abundance - sd_abundance, ymax = mean_abundance + sd_abundance),
              width = 0.2, position = position_dodge(0.9)) +
labs(title = "Total Abundance per Site with SD",
     x = "Site",
     y = "Total Abundance") +
theme_minimal()
```



```
library(dplyr)
library(broom)

# Convert 'sample' column to a factor
data$sample <- as.factor(data$sample)

# Perform ANOVA
anova_result <- data %>%
  group_by(site, sample) %>%
  summarise(num_taxa = n_distinct(taxa)) %>%
  aov(num_taxa ~ site, data = .) %>%
  tidy()
```

`summarise()` has grouped output by 'site'. You can override using the
`.groups` argument.

```
# Print ANOVA results
print(anova_result)
```

```
## # A tibble: 2 x 6
##   term      df sumsq meansq statistic p.value
##   <chr>    <dbl> <dbl>  <dbl>    <dbl>   <dbl>
## 1 site      1  2.67   2.67     0.842   0.411
## 2 Residuals  4 12.7   3.17     NA      NA
```

```
# Calculate species richness per site and sample
species_richness <- data %>%
  group_by(site, sample) %>%
  summarise(species_count = n_distinct(taxa))
```

```
## `summarise()` has grouped output by 'site'. You can override using the
## `.groups` argument.
```

```
# Calculate average and SD per sample
summary_data <- species_richness %>%
  group_by(site) %>%
  summarise(mean_species_richness = mean(species_count),
            sd_species_richness = sd(species_count))
```

```
# Plot species richness by site
ggplot(species_richness, aes(x = site, y = species_count, fill = site)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Species Richness by Site",
       x = "Site",
       y = "Species Richness") +
  theme_minimal() +
  theme(legend.position = "none") + # To hide the legend for site
```

```
# Add average and SD per sample
geom_errorbar(data = summary_data, aes(x = site, y = mean_species_richness, ymin = mean_species_richness - 1.96 * sd_species_richness, ymax = mean_species_richness + 1.96 * sd_species_richness)) +
  geom_point(data = summary_data, aes(x = site, y = mean_species_richness), position = position_dodge(0.9))
```

