

Review - uhuru data set

Luna L Sanchez Reyes

2022-12-08

A function for the UHURU data set

The UHURU experiment in Kenya has conducted a survey of *Acacia drepanolobium* among each of their ungulate exclosure treatments. Data for the survey is available here in a tab delimited (“`\t`”) format. Each of the individuals surveyed were measured for branch circumference (CIRC) and canopy width (AXIS1) and was identified for the associated ant-symbiont species present (ANT).

The following function takes a subset of the data for a given ANT symbiont and evaluates the linear regression (`lm()`) for a given relationship, returning the symbiont species used for the subset and the `r2` of the model.

```
report_rsquared <- function(data, species, formula){
  subset <- dplyr::filter(data, ANT == species)
  test <- lm(formula, data = subset)
  rsquared <- round(summary(test)$r.squared, 3)
  output <- data.frame(species = species, r2 = rsquared)
  return(output)
}
```

Explain what each line of code in the body of the function is doing. Add the explanations to your Rmd file as comments, before each line of code.

Execute the function using the UHURU data and specifying species = “CM” and formula = “AXIS1~CIRC”.

```
uhuru_data <- read.csv(file = "https://esapubs.org/archive/ecol/E095/064/ACACIA_DREPANOLOBIIUM_SURVEY.txt")
head(uhuru_data)
```

##	SURVEY	YEAR	SITE	BLOCK	TREATMENT	PLOT	ID	HEIGHT	AXIS1	AXIS2	CIRC
## 1	1	2012	SOUTH	1	TOTAL	S1TOTAL	581	2.25	2.75	2.15	20
## 2	1	2012	SOUTH	1	TOTAL	S1TOTAL	582	2.65	4.10	3.90	28
## 3	1	2012	SOUTH	1	TOTAL	S1TOTAL	3111	1.5	1.70	0.85	17
## 4	1	2012	SOUTH	1	TOTAL	S1TOTAL	3112	2.01	1.80	1.60	12
## 5	1	2012	SOUTH	1	TOTAL	S1TOTAL	3113	1.75	1.84	1.42	13
## 6	1	2012	SOUTH	1	TOTAL	S1TOTAL	3114	1.65	1.62	0.85	15
##	FLOWERS	BUDS	FRUITS	ANT							
## 1	0	0	10	CS							
## 2	0	0	150	TP							
## 3	2	1	50	TP							
## 4	0	0	75	CS							
## 5	0	0	20	CS							
## 6	0	0	0	E							

```
report_rsquared(data = uhuru_data, species = "CM", formula = "AXIS1~CIRC")
```

```
## species r2
## 1      CM 0.866
```

3. Modify the function so that it also determines if() the rsquared is significant based on a given threshold. The modified function should return() the species, rsquared and a significance value of “S” for a relationship with an rsquared > threshold or “NS” for an rsquared < threshold.

```
report_rsquared_sign <- function(data, species, formula, threshold){  
  
  subset <- dplyr::filter(data, ANT == species)  
  test <- lm(formula, data = subset)  
  rsquared <- round(summary(test)$r.squared, 3)  
  if (rsquared > threshold) {  
    print("Significant")  
    sign <- "S"  
  } else if (rsquared < threshold) {  
    print("Non significant")  
    sign <- "NS"  
  }  
  
  output <- data.frame(species = species, r2 = rsquared, significance = sign)  
  return(output)  
}
```

4. Execute your modified function for species of “CM”, “CS”, and “TP” given a threshold = 0.667.

```
report_rsquared_sign(data = uhuru_data, species = "CM", threshold = 0.667, formula = "AXIS1~CIRC")  
  
## [1] "Significant"  
##   species    r2 significance  
## 1      CM 0.866           S
```