BIOL 220 Problem Set 13: Regression

Answer Key

Due Thursday, April 25, 2024 before class

Submit your answers via Google Classroom

Benefits of biodiversity

Figure 1: Ho'omaluhia Botanical Garden, an urban green space on O'ahu. Photo credit: Daniel Ramirez / Flickr

Just like in the problem set on correlation, we will again look at the data from Fuller *et al.* (2007) who measured the diversity of plant, bird, and butterfly species in 15 urban green spaces of varying size. They also asked 312 green-space users about their emotional "attachment" to green spaces to see if more biodiverse green space lead to greater emotional attachment from users.

Attachment	Area (ha)	Number of butterfly species	Number of bird species	ln(Number of plant species)
4.4	23.8	6	12	5.1

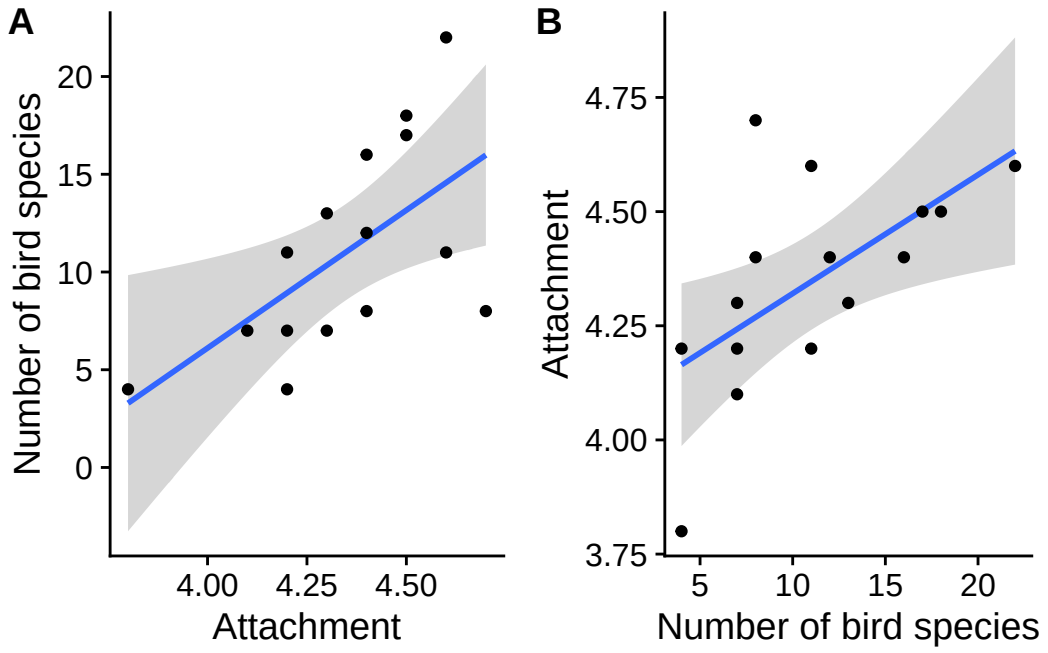
Attachment	Area (ha)	Number of butterfly species	Number of bird species	ln(Number of plant species)
4.5	16.0	14	18	5.5
4.7	6.9	8	8	6.4
4.5	2.3	10	17	4.7
4.3	5.7	6	7	5.3
3.8	1.2	5	4	4.6
4.4	1.4	5	8	4.5
4.6	15.0	7	22	5.5
4.1	3.1	9	7	5.2
4.2	3.8	5	4	4.6
4.6	7.6	10	11	4.5
4.2	12.9	9	11	5.0
4.3	4.0	12	13	5.0
4.4	5.6	11	16	5.6
4.2	4.9	7	7	5.4

Questions 1–5 refer to the prompt and data above

To get these data into *R* you can copy-paste and run this code:

```
fuller_etal_2017 <- data.frame(
    attachment = c(4.4, 4.5, 4.7, 4.5, 4.3, 3.8, 4.4, 4.6, 4.1,
                   4.2, 4.6, 4.2, 4.3, 4.4, 4.2),
    area_ha = c(23.8, 16.0, 6.9, 2.3, 5.7, 1.2, 1.4, 15.0, 3.1,
                3.8, 7.6, 12.9, 4.0, 5.6, 4.9),
    butterfly = c(6, 14, 8, 10, 6, 5, 5, 7, 9, 5, 10, 9, 12, 11, 7),
    bird = c(12, 18, 8, 17, 7, 4, 8, 22, 7, 4, 11, 11, 13, 16, 7),
    ln_plant = c(5.1, 5.5, 6.4, 4.7, 5.3, 4.6, 4.5, 5.5, 5.2,
                 4.6, 4.5, 5.0, 5.0, 5.6, 5.4)
)
```

1. If our question is "does greater bird diversity increase attachment?", which graph below best visualizes the data in a way consistent with this question? [1 point]



Answer

B

2. If our question is "does greater bird diversity increase attachment?" what is the best null hypothesis in a linear regression framework? [1 point]

   a. H_0: The slope of **(change in bird species richness)** versus **(change in attachment)** is 0
   b. H_0: The slope of **(change in attachment)** versus **(change in bird species richness)** is 0
   c. H_0: The slope of **(change in attachment)** versus **(change in bird species richness)** is not 0
   d. H_0: The slope of **(change in bird species richness)** versus **(change in attachment)** is less than 0

Answer

b. H_0: The slope of **(change in attachment)** versus **(change in bird species richness)** is 0

3. Calculate the estimated slope for each of the following relationships:

a. Attachment as predicted by Number of butterfly species [1 point]
b. Attachment as predicted by Number of bird species [1 point]
c. Attachment as predicted by Area [1 point]

Answers

a. Attachment as predicted by Number of butterfly species

```
mod_bfly <- lm(attachment ~ butterfly, data = fuller_etal_2017)
summary(mod_bfly)
```

```
Call:
lm(formula = attachment ~ butterfly, data = fuller_etal_2017)

Residuals:
     Min       1Q   Median       3Q      Max
-0.45156 -0.13257 -0.01359  0.13388  0.36110

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.10599    0.18737   21.91  1.2e-11 ***
butterfly    0.02911    0.02157    1.35      0.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.223 on 13 degrees of freedom
Multiple R-squared:  0.1229,    Adjusted R-squared:  0.05547
F-statistic: 1.822 on 1 and 13 DF,  p-value: 0.2001
```

The slope is 0.029

b. Attachment as predicted by Number of bird species

```
mod_bird <- lm(attachment ~ bird, data = fuller_etal_2017)
summary(mod_bird)
```

```
Call:
lm(formula = attachment ~ bird, data = fuller_etal_2017)

Residuals:
     Min       1Q   Median       3Q      Max
-0.36467 -0.08767 -0.02867  0.04633  0.43133

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.060667   0.115151 535.264 2.72e-14 ***
bird        0.026000   0.009476   2.744   0.0167 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1895 on 13 degrees of freedom
Multiple R-squared:  0.3667,    Adjusted R-squared:  0.318
F-statistic: 7.528 on 1 and 13 DF,  p-value: 0.01674
```

4. For the linear model predicting Attachment from Number of bird species, report the results for a t-test null hypothesis test. Specifically, report the t-statistic, degrees of freedom, and P-value. [1 point]

Answer

t value	df	Pr(>|t|)
2.743769	13	0.0167359

5. Use the `confint` function to calculate the 95% confidence interval for the slope of Attachment as predicted by Number of bird species. [1 point]

Answer

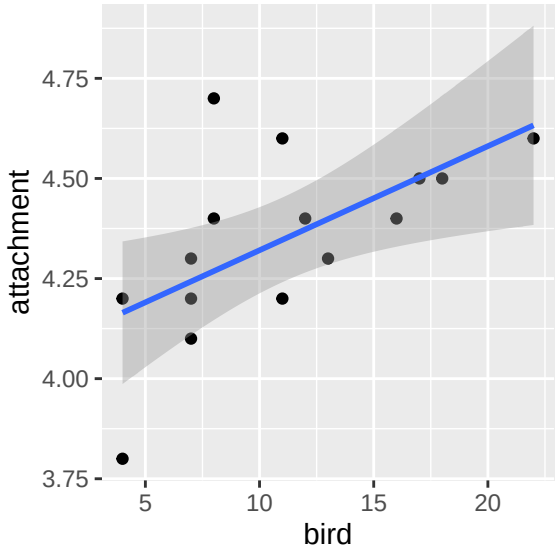
```
confint(mod_bird, "bird")
```

```
           2.5 %     97.5 %
bird 0.005528313 0.04647169
```

6. Use `ggplot` code to show the data for Attachment as predicted by Number of bird species, the regression line for this relationship, and the 95% confidence band. Copy and paste your code into the google form to answer this question. [1 point]

```
library(ggplot2) # not necessary for student to show library call

ggplot(fuller_etal_2017, aes(x = bird, y = attachment)) +
    geom_point() +
    geom_smooth(method = "lm")
```



7. Based on your findings from questions 2–6, do you reject the null hypothesis? What, if any, caveats should you state when interpreting the relationship between Attachment and Number of bird species? [2 points]

Answer

We reject the null. We can interpret number of bird species as having a significant positive relationship with attachment. Some caveats we could consider include (students can report any **one** of these, or another if it seems valid):

- $n = 15$ is a pretty small sample size
- perhaps because of small sample size, our 95% CI is close to 0
- correlation is not causation