

Partial Correlation

Instructor

Michał Kowalewski
Florida Museum of Natural History
University of Florida
kowalewski@ufl.edu



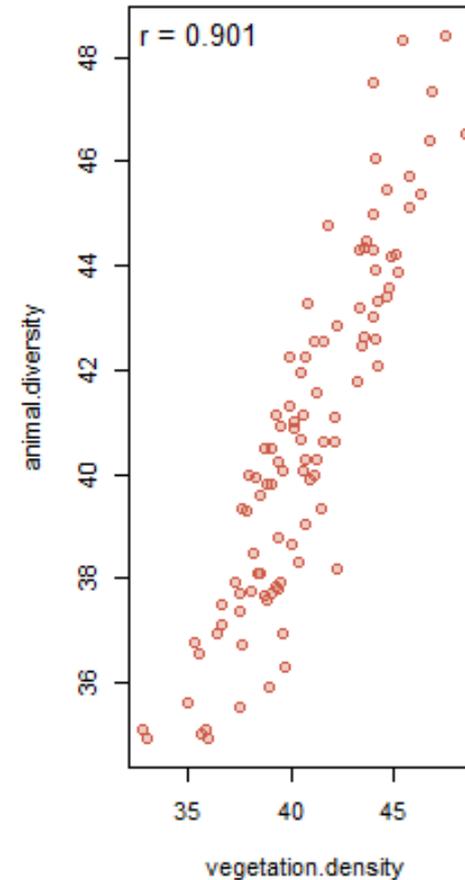
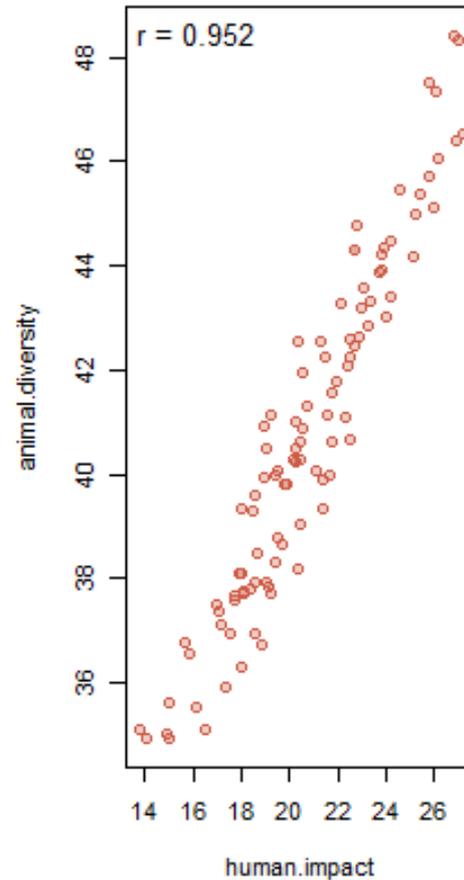
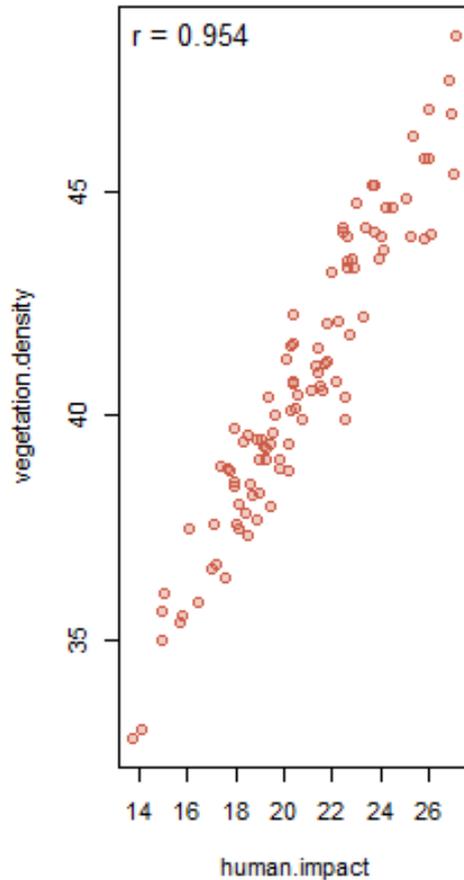
Partial Correlation

Partial correlation attempts to estimate relationship between two variables after accounting for other variables.

Synthetic Example

```
Z <- rnorm(100, 20, 3)
X <- rnorm(100, 20, 1) + Z
Y <- rnorm(100, 20, 1) + Z
```

```
# let Z ('human impact') be independent
# let X ('vegetation density') be strongly dependent on Z
# let Y ('animal diversity') be also strongly dependent on Z
```



Note that in this example X (vegetation density) and Y (animal diversity) are not interrelated directly
However, they appear strongly correlated because they are both strongly influenced by Z

One way to re-evaluate X and Y while minimizing effects of Z is partial correlation

Computing Partial Correlation

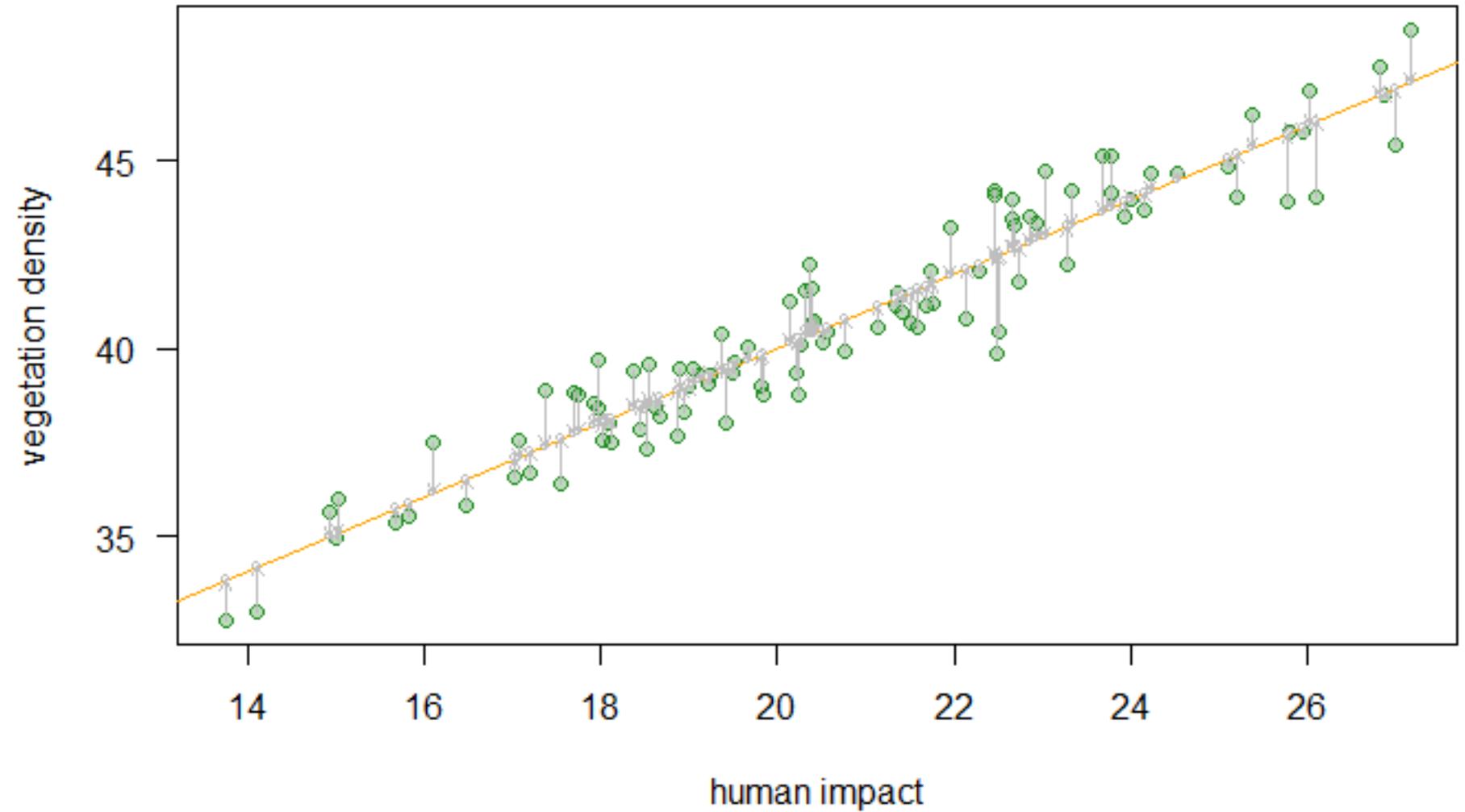
For a simplest case of three linearly related variables (x, y, z) **partial correlation of x and y given z** can be computed from residuals of OLS linear regression models $z \sim x$ and $z \sim y$

1. Compute residuals for x (dependent) \sim z (independent) model
2. Compute residuals for y (dependent) \sim z (independent) model
3. Compute correlation between the residuals

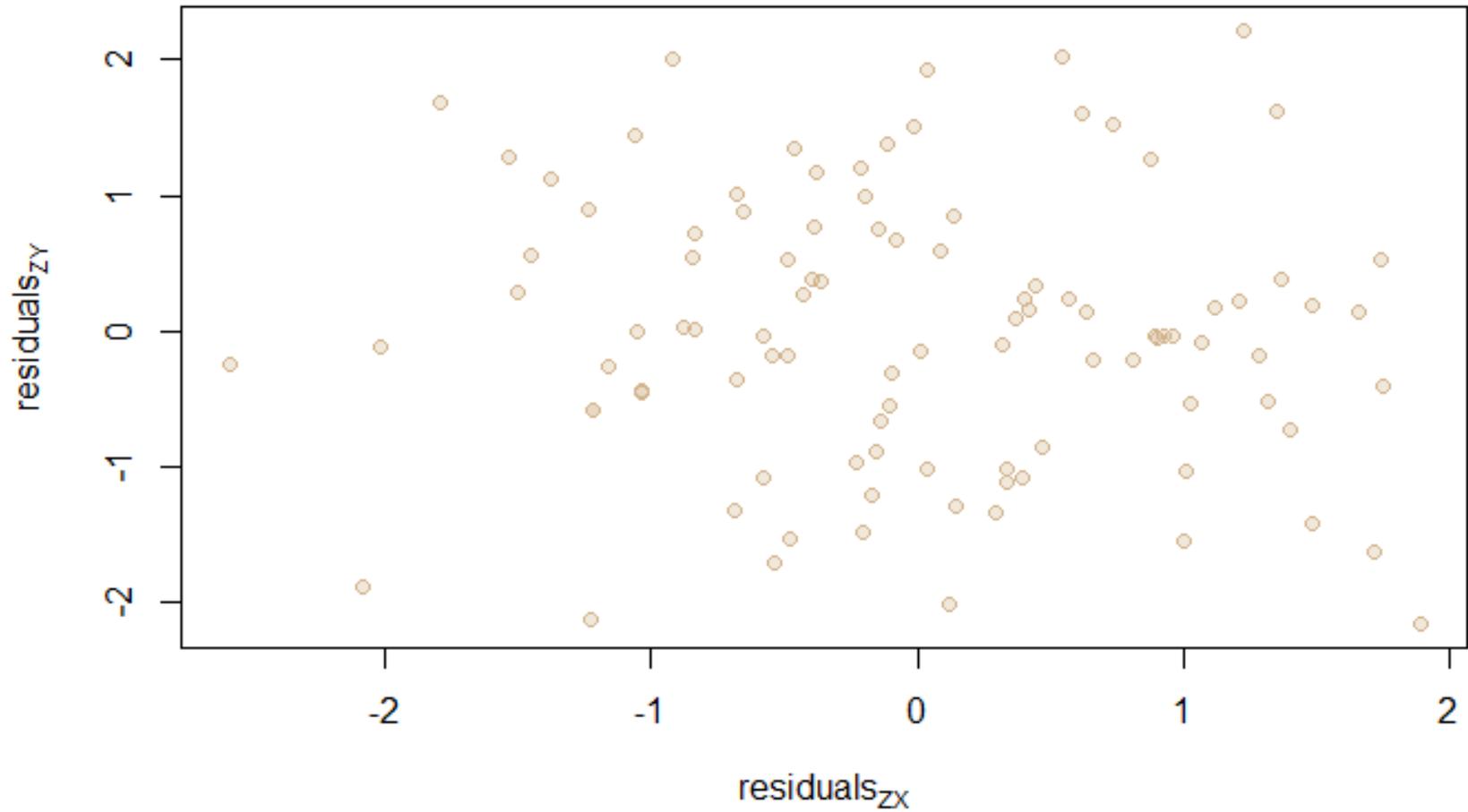
$$r_{xy_z} = \text{cor}(\text{lm}(X \sim Z)\$residuals, \text{lm}(Y \sim Z)\$residuals)$$

What are residuals?

OLS residuals for $y = \text{vegetation density}$



Partial correlation XY_Z is -0.081 (recall that r for XY was 0.901)



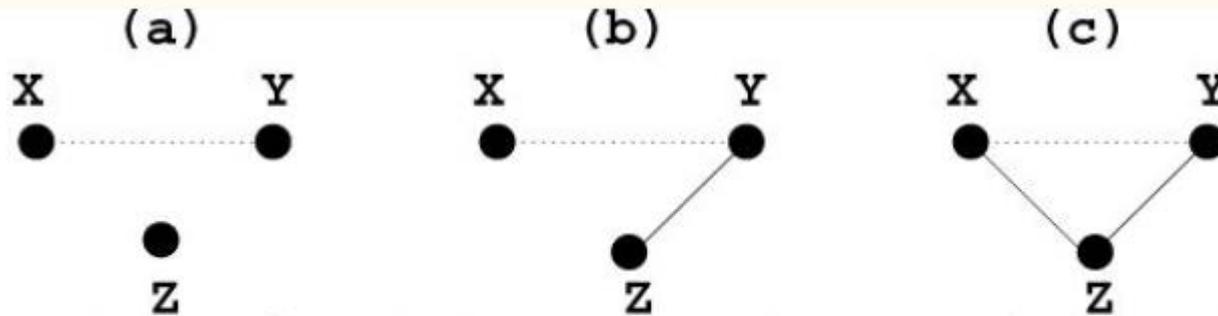
human. i mpact
human. i mpact
vegetati on. densi ty

versus
versus
versus

vegetati on. densi ty
ani mal . di versi ty
ani mal . di versi ty

parti al correlati ons
0. 72651264
0. 71126584
-0. 08147584

Partial and Semi-Partial Correlation



[Figure 1](#)

Graphical illustration of partial and semi-partial correlations among the three random variables X , Y , and Z

“The partial correlation can be explained as the association between two random variables after eliminating the effect of all other random variables, while the semi-partial correlation eliminates the effect of a fraction of other random variables, for instance, removing the effect of all other random variables from just one of two interesting random variables. The rationale for the partial and semi-partial correlations is to estimate a direct relationship or association between two random variables.”

{Kim, S. (2015) *Commun Stat Appl Methods* **22(6)**: 665–674.}

Partial and semi partial correlations can be computed using correlation coefficients

After: Kim, S. (2015) *Commun Stat Appl Methods* **22(6)**: 665–674.

Definition 1

The partial correlation of x_i and x_j given x_k is

$$r_{ij|k} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{1 - r_{ik}^2} \sqrt{1 - r_{jk}^2}} \quad (2.1)$$

and the semi-partial correlation of x_i with x_j given x_k is

$$r_{i(j|k)} = \frac{r_{ij} - r_{ik}r_{jk}}{\sqrt{1 - r_{jk}^2}}. \quad (2.2)$$

Parametric test for statistical significance of partial correlation

A slightly modified variant of t statistic used for standard Pearson's correlation

After: Kim, S. (2015) *Commun Stat Appl Methods* **22(6)**: 665–674.

$$t_{ij|S} = r_{ij|S} \sqrt{\frac{N - 2 - g}{1 - r_{ij|S}^2}}; \quad t_{i(j|S)} = r_{i(j|S)} \sqrt{\frac{N - 2 - g}{1 - r_{i(j|S)}^2}} \tag{2.8}$$

partial correlation semi-partial correlation

N – sample size, g – number of controlling variables

For computing p value P(t) set $df = N - 2 - g$

What is the difference between this test and standard t-test for correlation?

Partial and Semi-Partial Rank Correlations

Also possible to compute Kendall and Spearman partial correlations

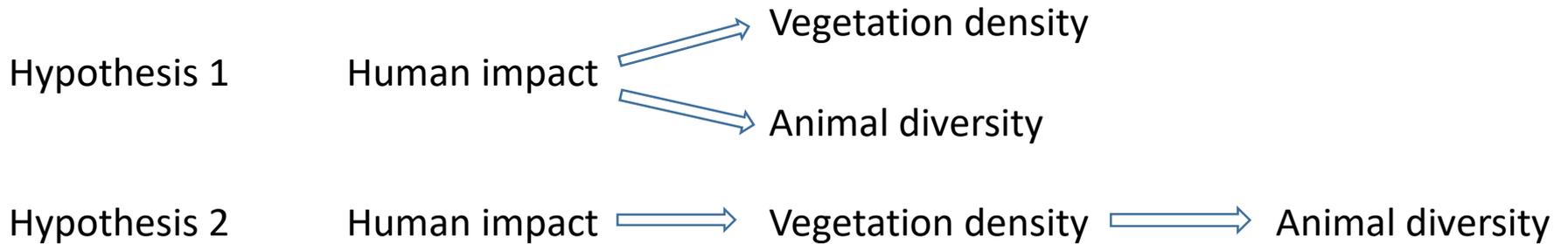
Statistical tests are also available for rank correlation coefficients

R package {ppcor} (Kim 2015) computes partial and semi-partial correlations and computes p values and test statistics.

Interpretation of Partial Correlations

Partial correlations attempt to assess direct interactions of two variables after accounting for other variables. Consequently, it can help us to evaluate alternative hypotheses that cannot be evaluated with standard correlations.

For example, consider the two alternative hypotheses:



Note that significant Pearson correlations are expected for all pairs of variables when standard correlation is applied.

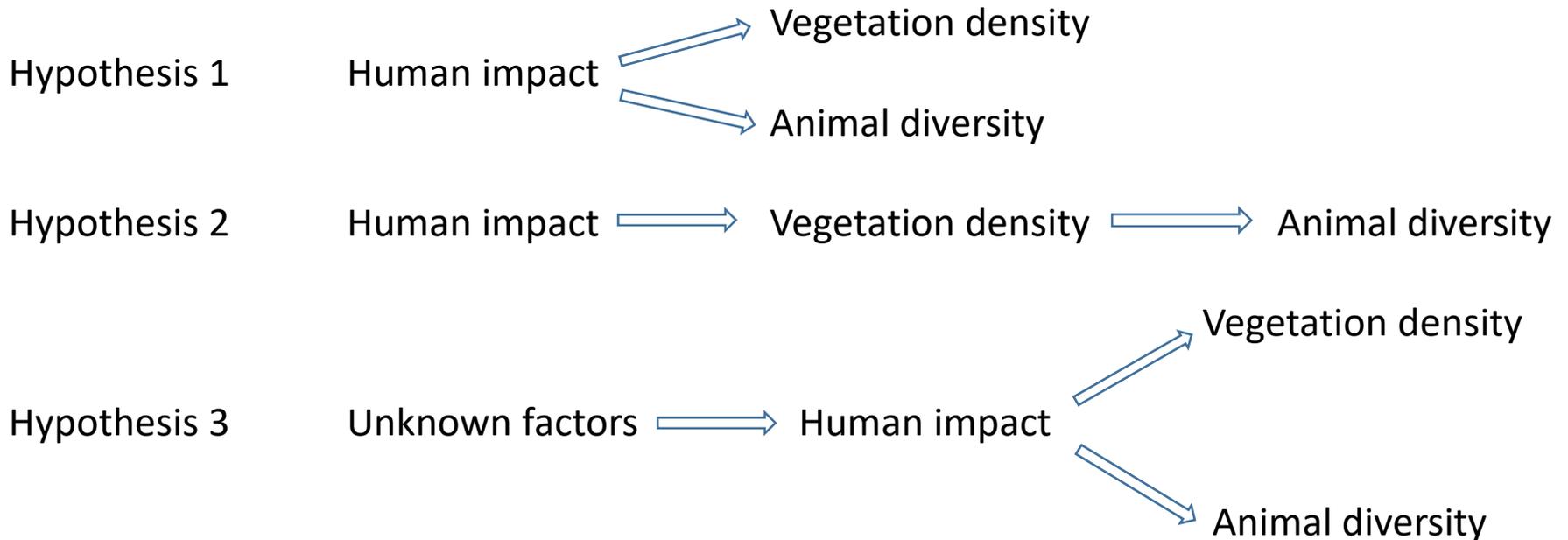
However, partial correlations should distinguish between those two hypotheses

(see our R markdown file)

Interpretation of Partial Correlations

While partial correlations may help resolve problems not resolvable using standard correlation, interpreting those results in terms of causalities is still risky.

This is because unknown variables that influence the study system can drive the observed patterns.



Paleontological Example

WOLFGANG EDER, JOHANN HOHENEGGER, and ANTONINO BRIGUGLIO, 2017, Depth-related morphoclines of megalospheric tests of *Heterostegina depressa* D'Orbigny: biostratigraphic and paleobiological implications. *Palaios* 32(1-2):110-117.

“MATERIAL AND METHODS

One hundred ninety-six (196) specimens of *H. depressa* from 5 to 90 m water depth (collected by J. Hohenegger; sample list provided in Online Supplemental File 1) were used for these analyses. All megalospheric specimens were collected offshore Sesoko-Jima (26°39' 5.134"N, 127°51'11.635"E and 26°39'38.776"N, 127°51'56.28"E) in 1993 and 1996. Samples to 40 m depth were taken by SCUBA in 10-m intervals, whereas those from 50 to 90 m water depth were dredged. “

	Pearson's r					
	P	X	S	d	D	K
Depth						
corr.	0.726	-0.819	0.572	0.613	0.639	0.541
p-value	1.03E-33	5.27E-49	9.29E-19	6.91E-22	3.35E-24	1.29E-16
P						
corr.		-0.789	0.692	0.893	0.886	0.517
p-value		3.78E-43	1.45E-29	2.22E-69	6.22E-67	4.19E-15
X						
corr.			-0.620	-0.746	-0.761	-0.516
p-value			1.55E-22	2.06E-36	1.18E-38	5.01E-15
S						
corr.				0.673	0.704	0.515
p-value				1.80E-27	6.13E-31	5.63E-15
d						
corr.					0.981	0.406
p-value					1.04E-140	1.72E-09
D						
corr.						0.559
p-value						7.91E-18
Partial correlation						
Control variable:						
depth	X	S	D	d	K	
P						
corr.	-0.491	0.490	0.824	0.798	0.215	
p-value	1.51E-13	1.78E-13	7.83E-50	1.49E-44	0.001	
X						
corr.		-0.322	-0.539	-0.539	-0.150	
p-value		2.20E-06	2.25E-16	2.19E-16	0.018	
S						
corr.			0.497	0.536	0.298	
p-value			7.58E-14	3.38E-16	1.19E-05	
D						
corr.				0.970	0.112	
p-value				1.16E-120	0.059	
d						
corr.					0.330	
p-value					1.24E-06	

Biometric character	Definition
P	the inner maximum diameter of the proloculus, omitting the wall
X	the number of undivided chambers until the first emergence of secondary septa, not counting embryonic chambers
S	the number of chamberlets at the 14th chamber. If it is an undivided chamber, the values were marked as 1, to avoid 0 values in shallower specimens
d	the diameter of the 1st whorl measured perpendicular to the axis of P
D	the diameter of the whorl at 1.5 revolutions
K	the index of spiral opening, which is calculated independently of P (Less et al. 2008)

Biometric variables vary predictably with depth and thus the correlations of individual biometric variables with each other may be partly reflecting their joint response to bathymetry.

We can remove the effect of depth on all biometric variables using partial correlation and then reevaluate interrelations between biometric variables.

Note for example that partial correlation of D and K becomes low and insignificant.

R-markdown Tutorial

Partial correlation