

STATISTIČKE ANALIZE

Statističke analize služe **procjeni** (*estimation*) parametara populacija te **testiranju hipoteza**.



Procjena – utvrđivanje kvantitativne vrijednosti parametra (varijable) populacije na temelju uzorka

- *Prosječna veličina srdele u populacijama južnog i sjevernog Jadrana*
- *Bazalna ekspresija CyP1A2 gena u stanici jetre zdravog muškarca*
- *Oksidativna oštećenja u mozgu nakon 2 sata korištenja mobilnog telefona na frekvenciji 900 MHz*

Hipoteza je tvrdnja o parametru (varijabli) populacije.

- *Srdele u južnom Jadranu su manje nego one u sjevernom*
- *Srdele u Jadranu pokazuju sezonalnost u veličini.*
- *Izloženost parama benzinskih motora u okolišno relevantnim koncentracijama tijekom 8 sati uzrokuje povišenu ekspresiju CyP1A2 gena u stanici jetre zdravog muškarca*
- *Inhibicija Tbx4 gena utječe na razvoj udova u embriju kokoške.*



Statističke metode ukazuju i na **podudarnost hipoteze o parametru populacije i podacima sakupljenim o uzorku**. Hipoteza je neophodna za biološka istraživačka pitanja i omogućuje utvrđivanje uzročno-posljedičnih odnosa, a testiranje hipoteze je informativnije od procjene.

TESTIRANJE HIPOTEZA

PITANJE – **HIPOTEZA/VARIJABLE** – DIZAJNIRANJE EKSPERIMENTA (UZORAK) – TESTIRANJE HIPOTEZE/ANALIZA PODATAKA



"I already wrote the paper.
That's why it's so hard to
get the right data."

TESTIRANJE STATISTIČKIH HIPOTEZA

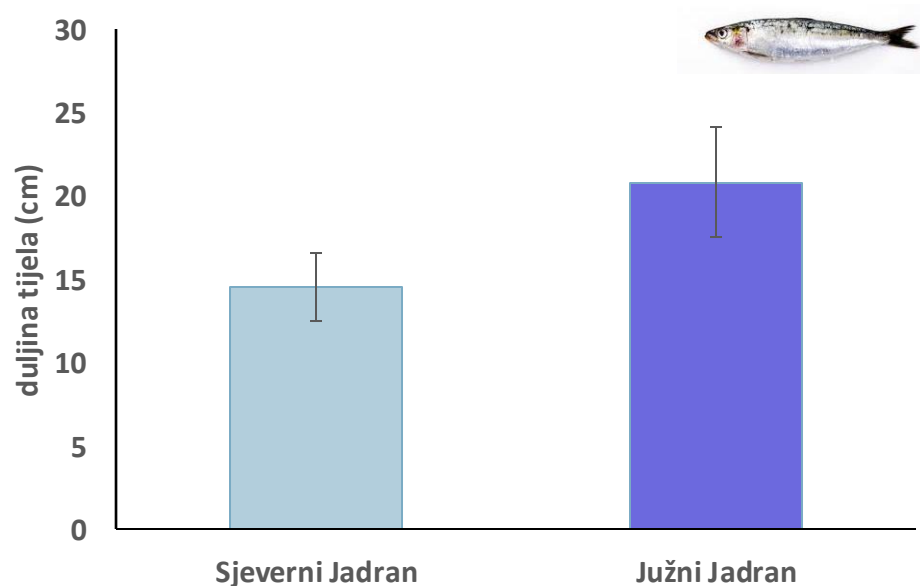
PITANJE – **HIPOTEZA/VARIJABLE** – DIZAJNIRANJE EKSPERIMENTA (UZORAK) – TESTIRANJE HIPOTEZE/ANALIZA PODATAKA

Nul hipoteza (H_0) – nema učinka među varijablama, nema razlike među populacijama

Nul hipoteza se može samo odbiti, ne potvrditi

Odbijanjem nul hipoteze (statističkim testom) zaključujemo da podaci podržavaju alternativnu hipotezu (H_A) – populacije se razlikuju ili nezavisna varijabla ima učinak na zavisnu

- *Inhibicija Tbx4 gena utječe na razvoj udova u embriju kokoške.*
Nul hipoteza je *Inhibicija Tbx4 gena ne utječe na razvoj udova u embriju kokoške.*
- *Srdele u južnom Jadranu su veće nego one u sjevernom.*
Nul hipoteza je *“Srdele u južnom i sjevernom Jadranu ne razlikuju se po veličini.”*



Procjena (*estimation*) nam daje uvid u magnitudu učinka.

TESTIRANJE HIPOTEZA

PITANJE – **HIPOTEZA/VARIJABLE** – DIZAJNIRANJE EKSPERIMENTA (UZORAK) – TESTIRANJE HIPOTEZE/ANALIZA PODATAKA

Testiranjem statističke hipoteze uspoređujemo koliko podaci odstupaju od obrazaca koje očekujemo ako je nulta hipoteza točna!

Znanstvena hipoteza zapravo je pravo biološko pitanje, i znanstvenici postavljaju pretpostavke koje onda testiraju kroz statističke hipoteze.

Npr. znanstvena hipoteza bi bila da se srdele u Jadranu razlikuju po veličini, s pretpostavkom da su veće u južnom Jadranu zbog manje intenzivnog izlova

H_0 - *srdele u južnom i sjevernom Jadranu **ne razlikuju** se po veličini*

H_{A1} - *srdele u južnom i sjevernom Jadranu **razlikuju se** po veličini*

H_{A2} - *srdele u južnom Jadranu su veće od srdela u sjevernom Jadranu*

Alternativna hipoteza može biti jednostrana (H_{A2}) ili dvostrana (H_{A1}) (*one and two tailed*)

Dvostrana hipoteza uključuje vrijednosti na obje strane distribucije (srdele u južnom Jadranu su veće od srdela u sjevernom Jadranu i srdele u južnom Jadranu su manje od srdela u sjevernom Jadranu). Ukoliko postoji mogućnost za obje vrijednosti uvijek se koriste dvostrane hipoteze!

TESTIRANJE HIPOTEZA

- Parametri koje testiramo mogu biti srednje vrijednosti, varijance, korelacije, proporcije...
- Iznosimo najčešće **vrijednost statističkog testa, veličine uzoraka, deskriptivnu statistiku (zbog procjene učinka) i P -vrijednost**

Primjer: imaju li krastače veću preferencu korištenju jednog prednjeg uda?

18 krastača donešeno u labos, zavezan im je balon za glavu i bilježeno je s kojom prednjom nogom miču balon s glave



www.shutterstock.com - 1074389033

Bisazza et al. 1996. Right-pawedness in toads. *Nature* 379: 408.

H_0 : učestalost korištenja prednjih udova **je jednaka** u populaciji (i.e., $p = 0.5$).

H_A : učestalost korištenja prednjih udova **nije jednaka** u populaciji (i.e., $p \neq 0.5$).

Dvostrani test – biološki je značajno i ako preferencijalno koriste lijevu i desnu nogu
(frekvencija p za desnu nogu < 0.5 , frekvencija p za desnu nogu > 0.5)

H_0 – 9 krastača koristi desnu, devet lijevu nogu ($p = 0.5$)

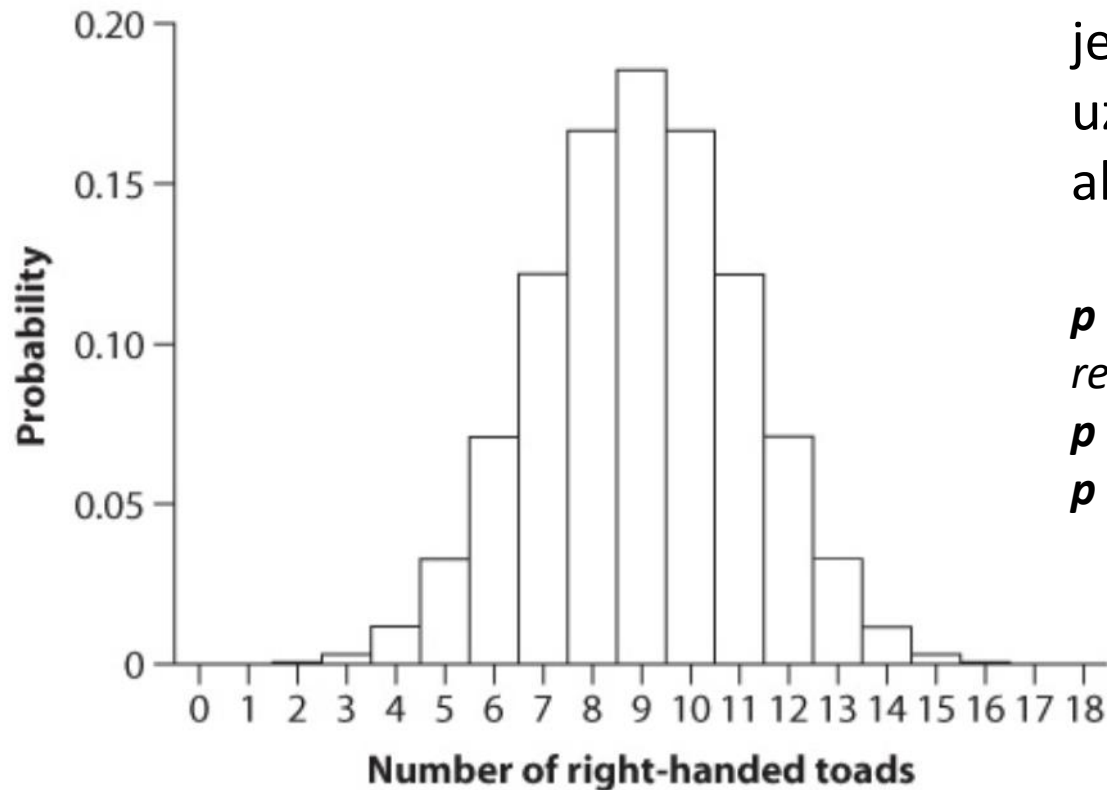
Rezultat mjerenja – 14 krastača koristilo je desnu nogu!
JE LI OVO SLUČAJAN ILI ZNAČAJAN REZULTAT?



Rezultat mjerenja – 14 krastača koristilo je desnu nogu! JE LI OVO SLUČAJAN ILI ZNAČAJAN REZULTAT



Potrebno je utvrditi distribuciju mogućih uzoraka za H_0 (nul distribucija)



Slučajnim uzorkovanjem 18 jedinki iz populacije krastača s jednakim preferencijama korištenja udova moguće je u uzorku dobiti od 0 do 18 jedinki koje koriste desnu nogu, ali je vjerojatnost svakog od tih događaja različita !

$p(0) = 0.000004$ (4 slučajna uzorkovanja od 1.000.000 daje ovaj rezultat)

$p(9) = 0.1855$ (18,55 slučajnih uzorkovanja na 100 daje ovaj rezultat)

$p(18) = 0.000004$

Figure 6.2-1

Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015 W. H. Freeman and Company

P VRIJEDNOST

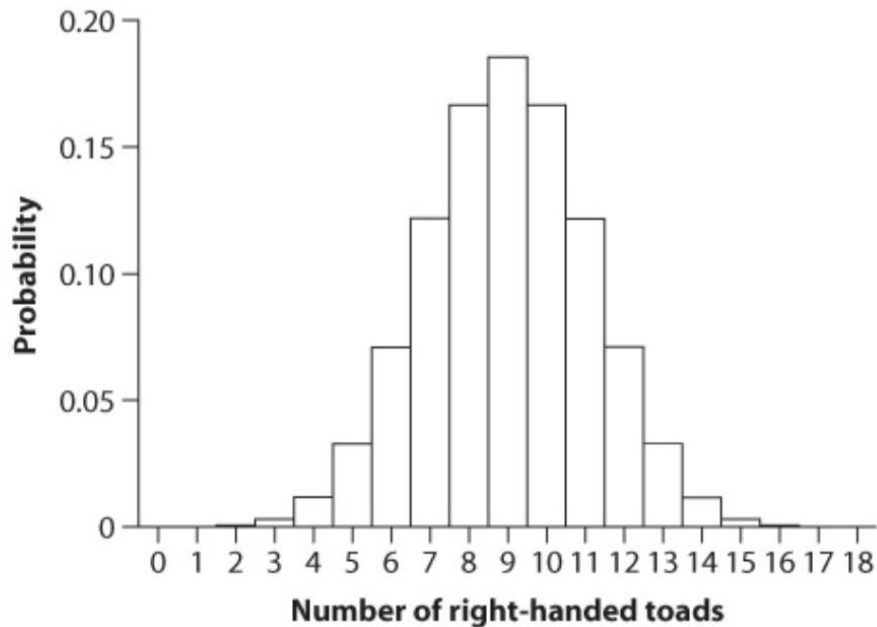
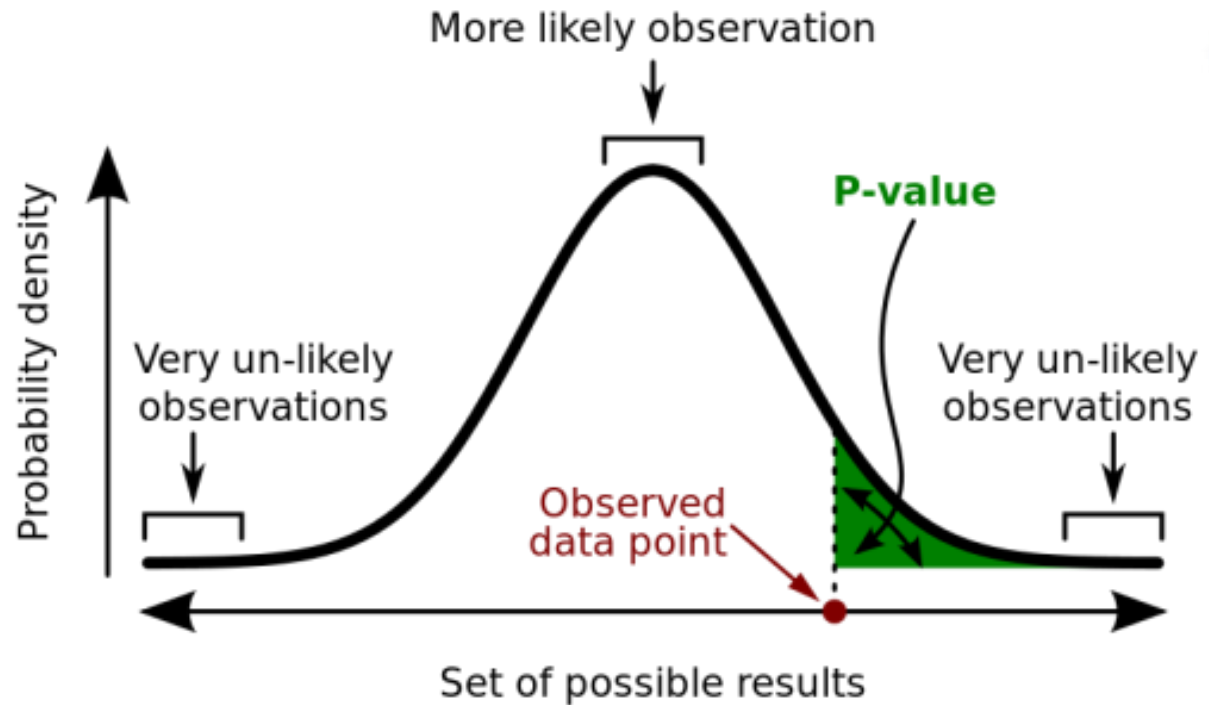
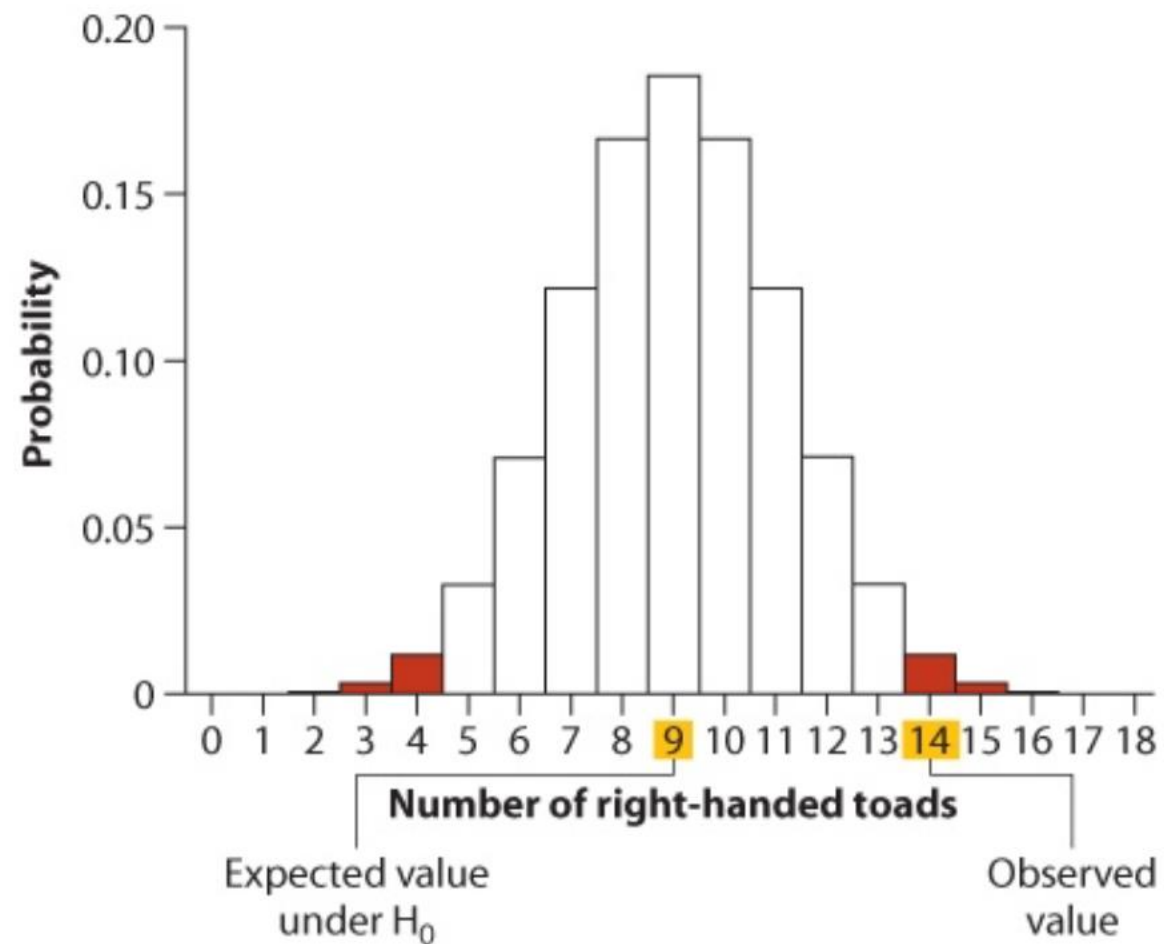


Figure 6.2-1
Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015 W. H. Freeman and Company



- **P** vrijednost nije vjerojatnost da je nul hipoteza točna ili kriva, nego MJERA NEKOMPTABILNOSTI PODATAKA SA STATISTIČKIM MODELOM (u širokom smislu vjerojatnost dobivanja tog rezultata ukoliko je nul hipoteza točna)
- **α** je statistička značajnost testa, u biologiji je često 0.05
- **P vrijednost** govori o snazi dokaza da je nul hipoteza kriva, ali ne daje informaciju o snazi učinka!!



$p < 0.05$
ODBACUJEMO NUL HIPOTEZU

Figure 6.2-2

Whitlock et al., *The Analysis of Biological Data*, 2e, © 2015 W. H. Freeman and Company

$V[14 \text{ ili više desnonogih}] = V[14] + V[15] + V[16] + V[17] + V[18] = 0.0155$ pomnožiti s 2 zbog dvostrukog testa, $V = 0.031$ **TO JE P- VRIJEDNOST**

RELEVANTNOST REZULTATA, ZNAČAJNOSTI I UČINKA

- provjeriti intervale pouzdanosti

95% CI za **vjerojatnost** (p) da je žaba desnonoga

$0.54 < p \text{ (desnonoge)} < 0.91$ – širok raspon, treba **povećati veličinu uzorka**



- **učinak (metoda procjene)** – npr. relativno povećanje varijable, 1%, 25%, 200 %
- **procjeniti biološku značajnost**
 - *npr. povećanje od 1 % bioraznolikosti u nekom području – kolika je ekološka značajnost!?*
 - *npr. povećanje od 2.5 % ekspresije gena Cyp1A1 – kolika je fiziološka značajnost*

GREŠKE TIP 1 I 2

Greška tipa 1 – odbijanje točne nul hipoteze

- “lažno pozitivni rezultat” (npr. nalazimo razliku koja ne postoji)

Greška tipa 2 – ne odbijanje netočne nul hipoteze (krivo prihvatanje H_0)

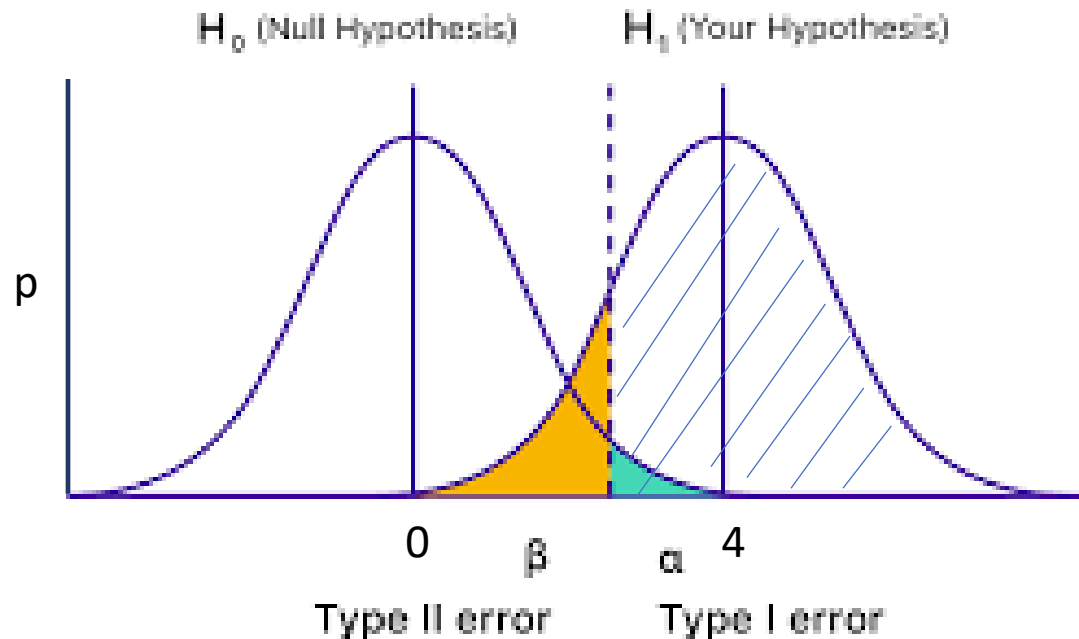
- “lažno negativni rezultat” (npr. ne nalazimo razliku koja postoji)

	H_0 True	H_0 False
Accept H_0 /Reject H_1	True	False: Type II error
Reject H_0 /Accept H_1	False: Type I error	True

GREŠKE TIP 1 I 2

Greška tipa 1 – odbijanje točne nul hipoteze

- definirana s, smanjuje se smanjivanjem α
- smanjivanje α dovodi i do povećanja greške tipa 2
- npr. ako je $\alpha=0.05$ jednom u 20 puta činimo grešku tipa 1



	H_0 True	H_0 False
Accept H_0 / Reject H_1	True	False: Type II error
Reject H_0 / Accept H_1	False: Type I error	True

H_0 – distribucija srednjih razlika riba ako je H_0 točna (vrijednosti koje se mogu pojaviti slučajnim uzorkovanjima)
 H_1 – distribucija srednjih razlika uzoraka riba ako je H_1 točna



Greška tipa 2 – ne odbijanje netočne nul hipoteze (krivo prihvaćanje H_0)

- Snaga testa – $1 - \beta$ (vjerojatnost da slučajan uzorak dovede do greške tipa 2)
- **Povećana veličina uzorka smanjuje grešku tipa 2 i povećava snagu testa (uža distribucija H_1)**

GREŠKE TIPA 1 I 2

Greška tipa 1 – odbijanje točne nul hipoteze

- *npr. smatramo da je lijek učinkovit i kada nije*
- *npr. smatramo da su srdele različite u sjevernom i južnom Jadranu i ako nisu*
- *npr. smatramo da cjepiva uzrokuju autizam*

Greška tipa 2 – ne odbijanje netočne nul hipoteze (krivo prihvatanje H_0)

- *npr. prihvatimo hipotezu da lijek nije učinkovit i kada je*
- *npr. prihvatimo hipotezu da srdele nisu različite u S i J Jadranu i ako jesu*
- *npr. smatramo da geni za nikotinske acetilholinske receptore (nAChRs) ne utječu na ovisnost o nikotinu*

REVIEW

Four principles for improved statistical ecology

Gordana Popovic¹  | Tanya Jane Mason^{2,3}  | Szymon Marian Drobnia^{4,5}  |

Najčešće greške

- 1) POSTAVLJANJE HIPOTEZE NA OSNOVU REZULTATA
- 2) NE IZNOŠENJE NE-ZNAČAJNIH REZULTATA
- 3) TESTIRANJE NULTE HIPOTEZE ZA KOJU ZNAMO DA JE KRIVA
- 4) KRIVA INTERPRETACIJA NE-ZNAČAJNIH REZULTATA KAO NE POSTOJANJE UČINKA ILI POVEZANOSTI

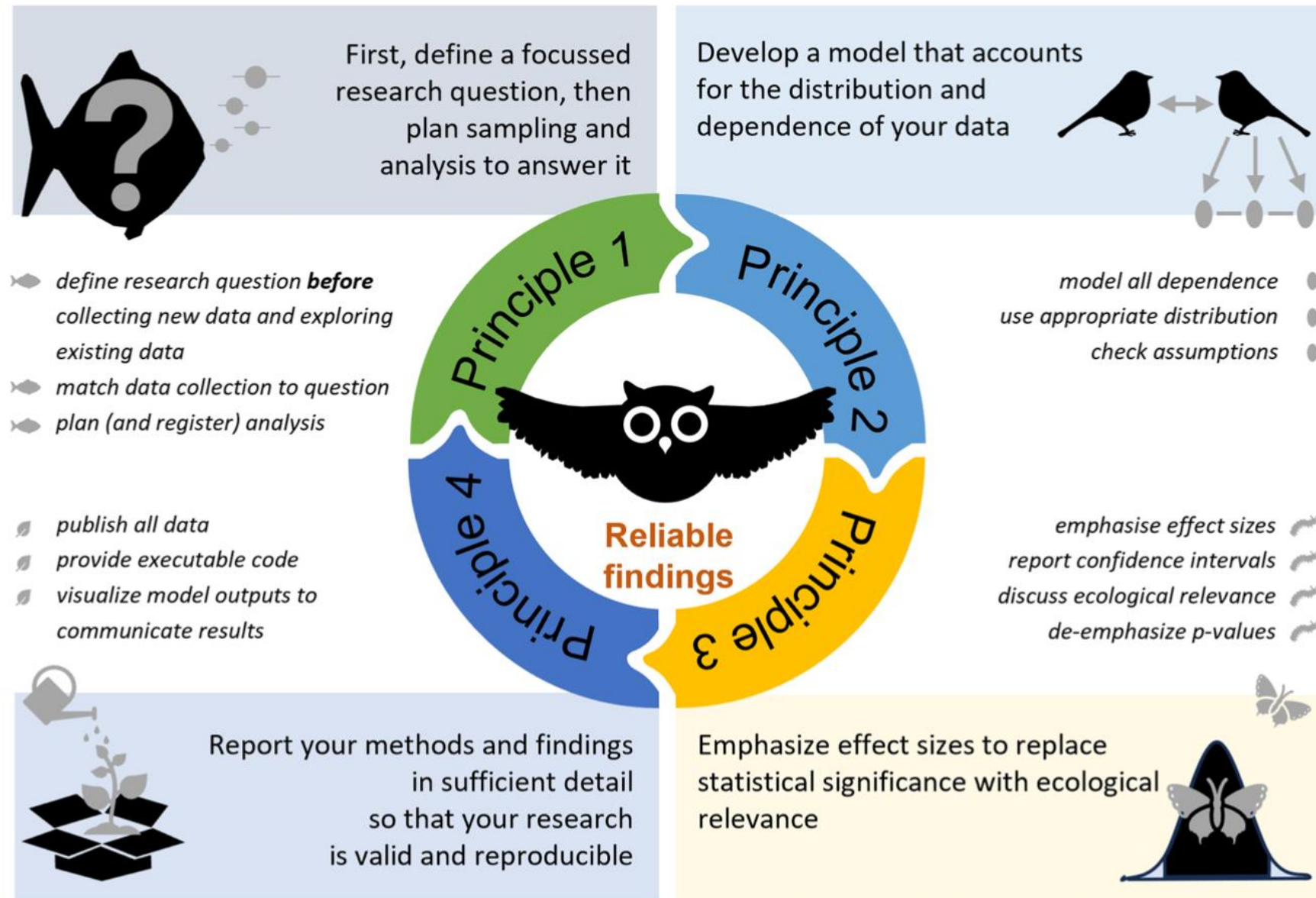
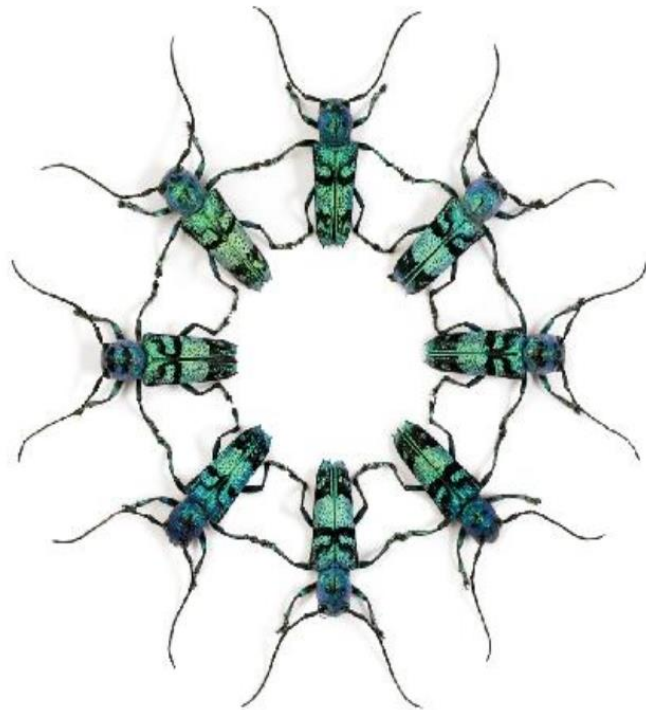


FIGURE 1 Four principles for improved statistical ecology.



The Analysis of Biological Data

WHITLOCK • SCHLUTER

SECOND EDITION