

ANALIZE PODATAKA U BIOLOŠKIM ISTRAŽIVANJIMA

2P + 3V

Izv. prof. Anamaria Štambuk, Biološki odsjek, PMF astambuk@biol.pmf.hr

Izv. prof. Toni Safner, Agronomski fakultet tsafner@agr.hr, tsafner@gmail.com

Izv. prof. Rosa Karlić, Biološki odsjek, PMF rosa@bioinfo.hr

Lucija Kanjer, Biološki odsjek, PMF lucija.kanjer@biol.pmf.hr

Opcija 1 – kontinuirano praćenje:

- Studenti trebaju položiti **dva kolokvija** tijekom semestra. Svaki kolokvij sastoji se od zadataka s podacima koje treba analizirati na računalu. Pitanja uključuju i interpretaciju rezultata (predavanja!). Na kolokviju je moguće **koristiti sve mrežne izvore i literaturu**. Iz oba kolokvija potrebno je ostvariti minimalnu uspješnost 70%. Ocjena odgovara ukupnom postotku savladanog gradiva kroz oba kolokvija.

Opcija 2 (polaganje ispita u ispitnim rokovima):

- Ispit se sastoji od projektnog zadatka koji student treba riješiti prije ispita, te na usmenom ispitu prezentirati i obrazložiti postupak i rješenje.

Izvrstan (5)	90-100%
Vrlo dobar (4)	80-89%
Dobar (3)	70-79%
Dovoljan (2)	60-69%

ANALIZE PODATAKA U BIOLOŠKIM ISTRAŽIVANJIMA

Ciljevi kolegija?

- Matematičko razumijevanje statističkih modela
- Programiranje u R-u
- Odvedene biostatističke metode

Ishodi učenja

- 1) Raspoznati vrste bioloških podataka i varijabli
- 2) Razumjeti osnovne postavke prikupljanja podataka u biološkim istraživanjima
- 3) Primijeniti odgovarajuće osnovne statističke analize podataka u biološkim istraživanjima
- 4) Koristiti se programskim okruženjem R za prikaz i statističku analizu podataka u biološkim istraživanjima
- 5) **Interpretirati rezultate osnovnih deskriptivnih i statističkih metoda u biologiji**

ANALIZE PODATAKA U BIOLOŠKIM ISTRAŽIVANJIMA

Ciljevi kolegija?

- Percepcija **bioloških pitanja i eksperimenata kroz varijable**
- Koncepti eksperimentalnog/istraživačkog pristupa kroz postavljanje **hipoteze**
- Dizajniranje istraživanja s obzirom na **hipotezu, uzorkovanje i varijable**
- Testiranje hipoteza kroz osnovne **biostatističke** metode
- Grafički prikaz podataka
- Snalaženje u radu u okruženju R

ANALIZE PODATAKA U BIOLOŠKIM ISTRAŽIVANJIMA

Kvalitativna i kvantitativna istraživanja

Kvalitativni i kvantitativni podaci

~~ANALIZE~~ PODATAKA U BIOLOŠKIM ISTRAŽIVANJIMA

Kvalitativna istraživanja - opisna/deskriptivna istraživanja

- Istraživanja u kojima ne analiziramo podatke kvantitativno
 - Nema testiranja hipoteze
-
- Popis vrsta ptica selica na području Parka Prirode Kopački rit
 - Sekvenca varijante gena aldehid dehidrogenaza 2 (ALDH2*2) u Azijskoj populaciji
 - Popis lokaliteta invazivne biljne vrste žljezdasti nederak *Impatiens glandulifera* Royle na području Zagrebačke županije
 - Taksonomija i rasprostranjenost uholaža (Insecta: Dermaptera) u Hrvatskoj

ANALIZE PODATAKA U BIOLOŠKIM ISTRAŽIVANJIMA

Kvantitativna istraživanja – barem dio podataka je kvantitativan

- Istraživanja u kojima analiziramo podatke kvantitativno
- Često (i poželjno) testiramo hipoteze
- Položaj kromosoma 1 i 9, obilježenih pomoću tehnologije CRISPR/Cas9, u odnosu na jezgrinu ovojnicu i centrosome
- Neofobija juvenilnih jedinki iz pokusa križanja populacija primorske gušterice (*Podarcis siculus*)
- Mutacijski potpisi u topološki asocirajućim domenama uvealnog i kožnog melanoma čovjeka
- Morfološke značajke riba roda *Romanogobio* u dunavskom slijevu Hrvatske

KVALITATIVNI I KVANTITATIVNI PODACI

- Mjerenja ili opažanja podataka na setu uzoraka su **varijable**
- **Biološke varijable** su **mjerenja ili svojstva koja se razlikuju na bilo kojoj razini biološke organizacije** (molekula, stanica, tkivo, jedinka, populacija/vrsta, ekosustav..)

npr. DNA sekvenca promotora, broj mitohondrija, aktivnost enzima, veličina i broj makrofaga, ekspresija gena, prokrvljenost tumora, genotip, broj točki na krilima, duljina tibije, spol, dob, agresivnost, fekunditet, udio mužjaka, broj bakterija, raznolikost vrsta, brojnost predatora ...

- *mogu biti i okolišne (biotičke varijable)*

- U ekološkim, biomedicinskim, biokemijskim i mnogim molekularnim istraživanjima važne su i kemijske i fizikalne **varijable**

npr. konc. CO₂, temperatura, mineralni sastav, kemijski sastav otopine (uključujući koncentracije), elektromagnetska i ionizirajuća zračenja, brzina strujanja zraka/vjetra, udaljenost od kopna, izomerna konfiguracija molekule...

- *mogu biti okolišne ili intrinzičke (temp. okoliša ili temp. tijela)*

ANALIZE PODATAKA U BIOLOŠKIM ISTRAŽIVANJIMA

Kvantitativna istraživanja – uključuju kvantitativnu analizu podataka

Vrlo često biostatistika podrazumijeva utvrđivanje povezanost (*association*) dvije varijable, ili nalaženje razlika među grupama

Npr. *pokazuju li mladi poskoci brži agresivni odgovor na prijetnju?*

- Brzina reakcije ovisi o starosnoj kategoriji
- Starost i brzina reakcije su povezani



Npr. *Ima li najviše polifenola u grožđu sorte Plavac mali, sorte Zifandel ili Shiraz?*

- Količina polifenola ovisi o sorti grožđa
- Sorta grožđa i količina polifenola su povezani



PITANJE – **HIPOTEZA/VARIJABLE** – DIZAJNIRANJE EKSPERIMENTA - UZORAK – TESTIRANJE HIPOTEZE

- ZAVISNE I NEZAVISNE VARIJABLE
- KATEGORIČKE I NUMERIČKE VARIJABLE
- GRUPIRAJUĆE VARIJABLE

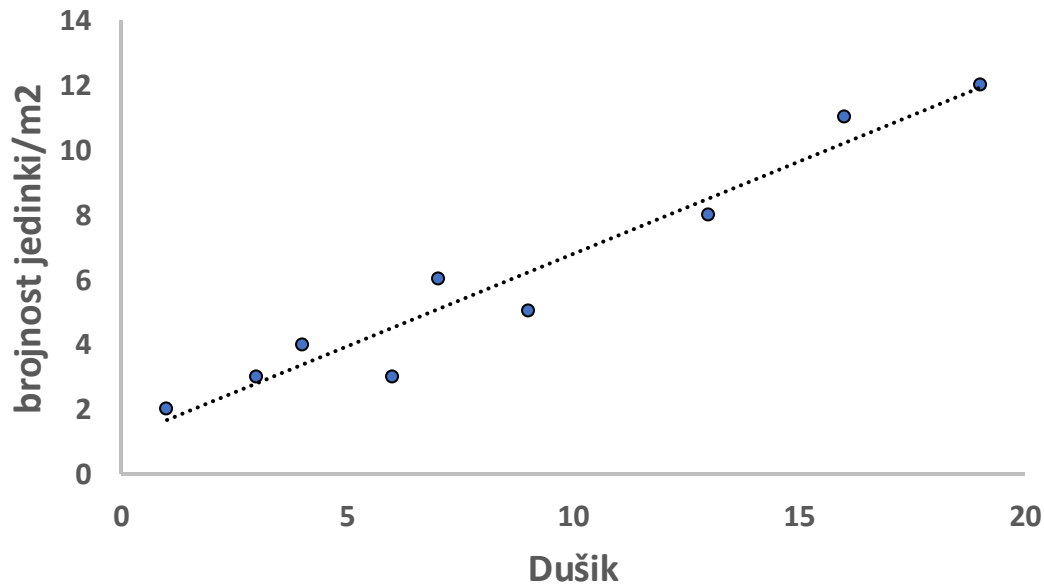
ZAVISNE I NEZAVISNE VARIJABLE

- **Nezavisne varijable (*predictor, explanatory*)** su one koje mogu uzrokovati promjenu u **zavisnoj varijabli (*outcome, response*)**
- Nezavisne varijable „manipuliramo“ u istraživanjima i često ih koristimo kao grupirajuće varijable
- *Pokazuju li mladi poskoci brži agresivni odgovor na prijetnju?*
- *Ima li najviše polifenola u grožđu sorte Plavac mali, sorte Zifandel ili Shiraz?*
- *Imaju li miševi s tumorom štitnjače višu temperaturu od zdravih?*
- *Induciraju li muški istraživači veći strah kod štakora?*
- *Jesu li maslačci na višim nadmorskim visinama niži?*
- *Spavaju li meduze duže po noći ili danu?*
- *Koja varijanta gena ALDH2 učinkovitije prerađuje acetaldehid?*



ZAVISNE I NEZAVISNE VARIJABLE

- **OPREZ!** Katkada je povezanost varijabli nejasna na prvi pogled!
- *Npr. sadržaj dušika u tlu i brojnost neke biljne vrste – ukoliko postoji povezanost (korelacija), potrebno je manipulirati jednu i drugu varijablu jer gustoća biljne populacije može utjecati na sadržaj dušika!*



KORELACIJA NIJE REGRESIJA I NE UKAZUJE NA UZORČNO POSLJEDIČNI ODNOS VARIJABLI

I USED TO THINK
CORRELATION IMPLIED
CAUSATION.



THEN I TOOK A
STATISTICS CLASS.
NOW I DON'T.



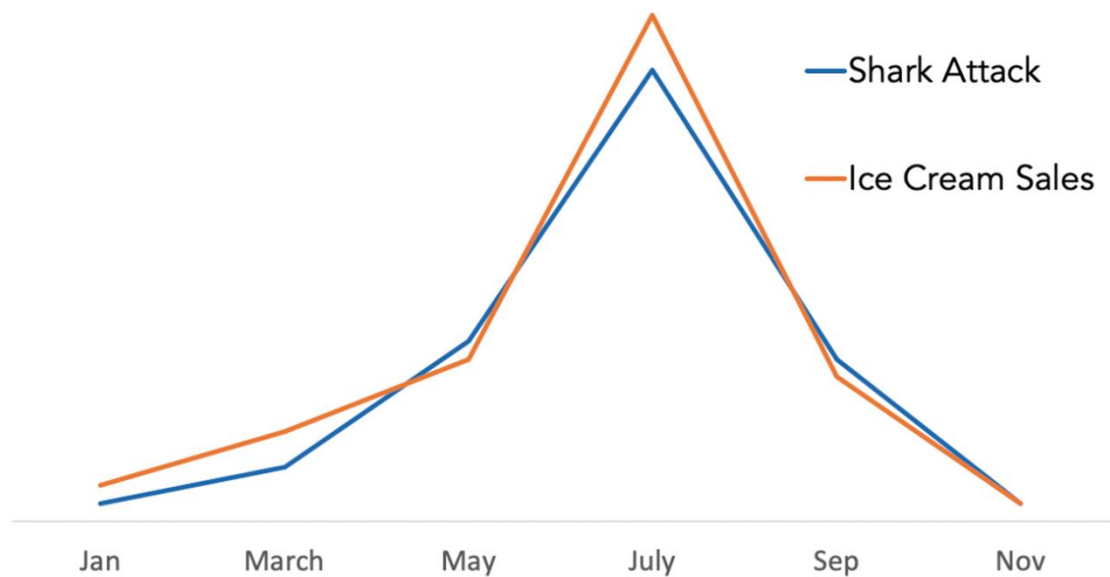
SOUNDS LIKE THE
CLASS HELPED.



ZAVISNE I NEZAVISNE VARIJABLE

- OPREZ! Katkada je povezanost varijabli netočna na prvi pogled!

PRIVIDNE (*spurious*) POVEZANOSTI



KORELACIJA NIJE REGRESIJA I NE UKAZUJE NA UZORČNO POSLJEDIČNI ODNOS VARIJABLI

ZAVISNE I NEZAVISNE VARIJABLE

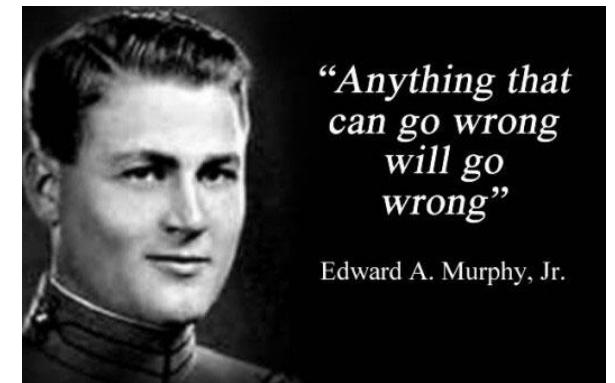
Posebna kategorija **nezavisnih varijabli** su **kontrolirane varijable** – važno je da su jednake kroz cijeli eksperiment (istraživanje)

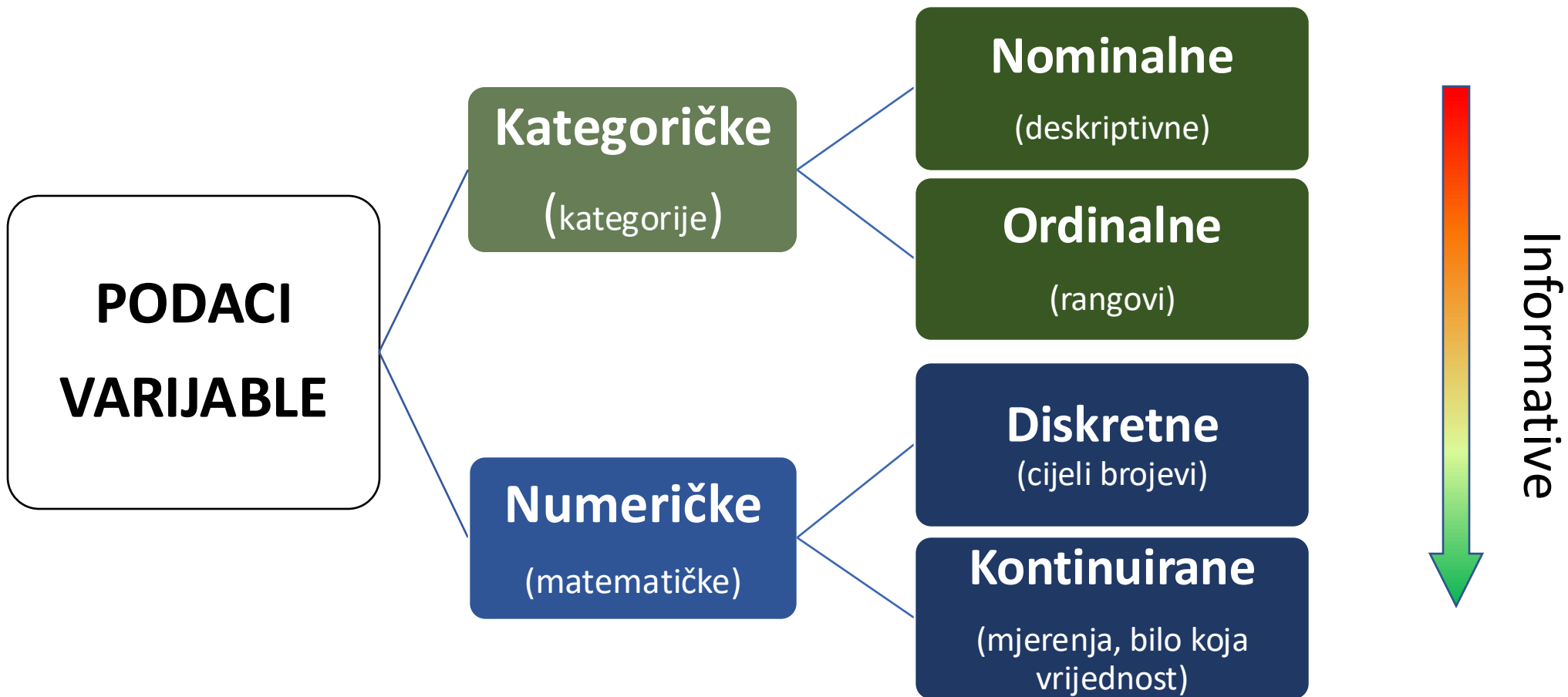
- Npr. *Pokazuju li mladi poskoci brži agresivni odgovor na prijetnju?*

Ukoliko se poskoci razlikuju u brzini agresivnog odgovora po spolu, idealno bi bilo da se odvojeno testira za mušjake i ženke (u nekim slučajevima dovoljno je da obje skupine sadrže jednaku zastupljenost po spolovima)

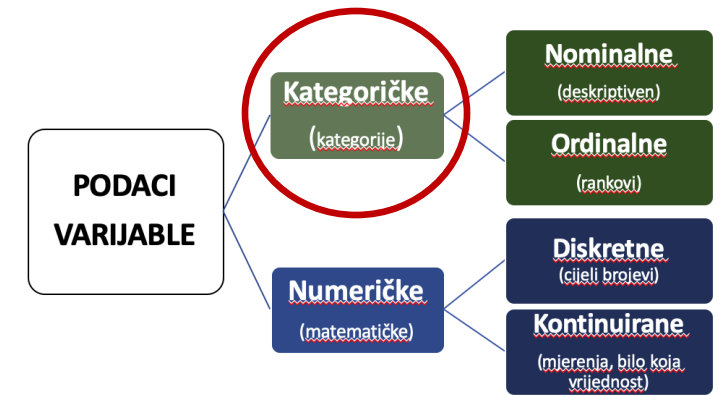
Starost i spol su samo neke od varijabli koje su često kontrolirane! – bitno je razmisliti koje varijable mogu još utjecati na zavisnu varijablu (**interni i eksterni, abiotski i biotski mikrokoliš**)

- *Npr. položaj jedinke ili posude u prostoriji, u staništu, pušački status, prehrana, utjecaj istraživača, režim svjetlosti, temperatura, reproduktivni status, mikrobiom, iste kemikalije korištene*

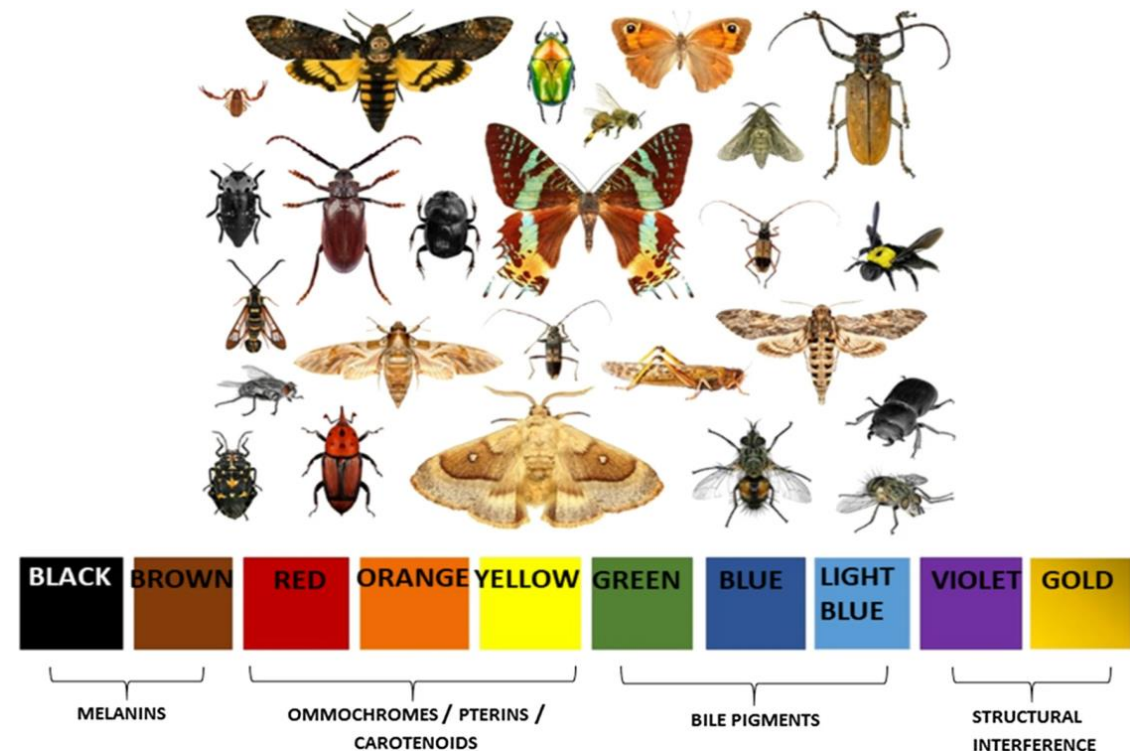




KATEGORIČKE (KVALITATIVNE) VARIJABLE

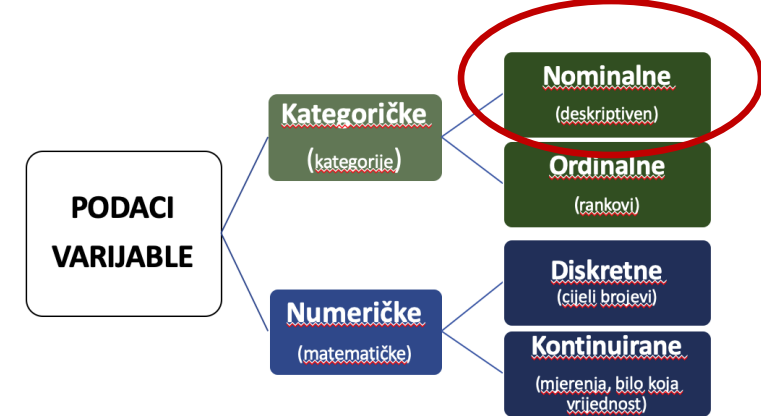


- Sve koje nisu izmjerene
- Svojstva ili kategorije
- Konačni broj mogućih vrijednosti



Nominalne (kategoričke) varijable

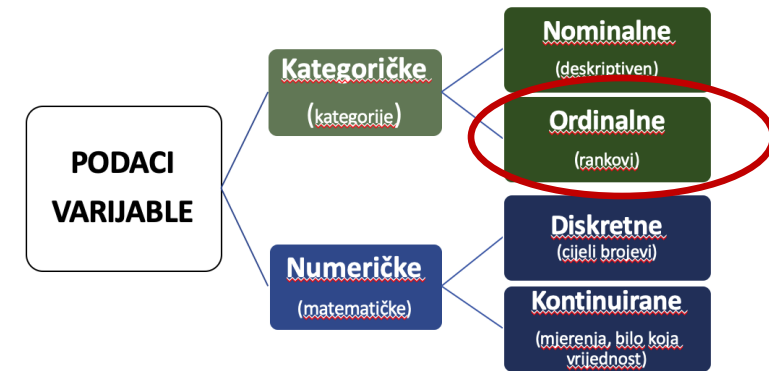
- Opisne, kvalitativne (bez vrijednosti)
- Često se koriste kao grupirajuće varijable
- *Populacija, tip stanica, spol, vrsta, genotip (npr. XX, XY), aminokiselina na 52. mjestu fosfomukoglutaze vinske mušice (alanin ili valin), boja, krvna grupa, stanište (šuma, polje, livada..), tip imunomodulatora..*
- Mogu biti dihotomske: DA/NE, ŽIV/MRTAV, 0/1



Ime uzorka je isto nominalna varijabla, npr. Miš_1, Miš_2. Miš_3, pacijent Pero Perić, XX_AB_27

Ordinalne (kategoričke) varijable

- **Kvalitativne (često opisne) ali se mogu rangirati**
- *Životni stadiji: jaje, ličinka, kukuljica, odrasli; juvenilni, odrasli, stari.*
- *Kategoriziranje kvantitativnih vrijednosti: spor, srednje brz, brz; nizak HDL, normalan HDL, visok HDL*
- *Svjetlo, sjena, mrak*
- *Nezrel plod, zrel, prezrel, truo*
- *Malen, srednje velik, velik*



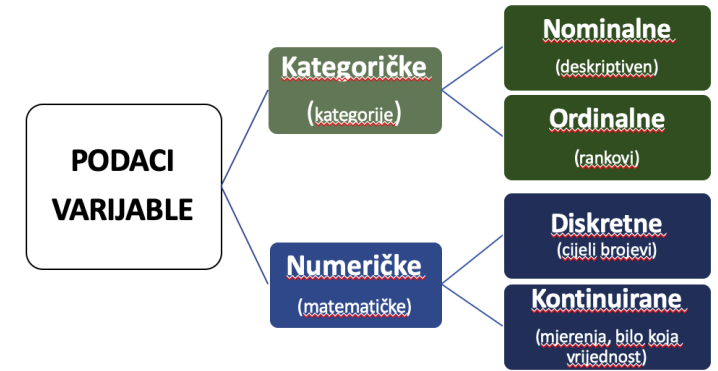
U slučaju kada su brojčane vrijednosti, nemaju numeričke odnose kao korišteni brojevi

- Npr. pratite vremenski redoslijed izlijevanje mladih kornjača po spolu

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
F	F	F	F	M	F	M	M	M	M



Spol nije ordinalna varijabla!

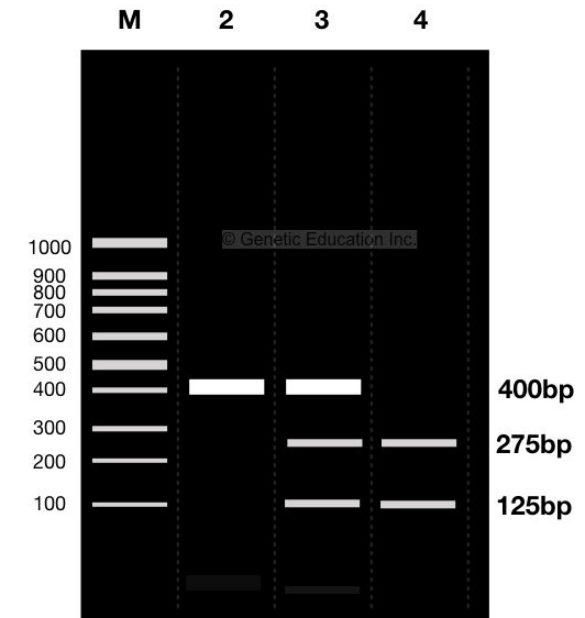
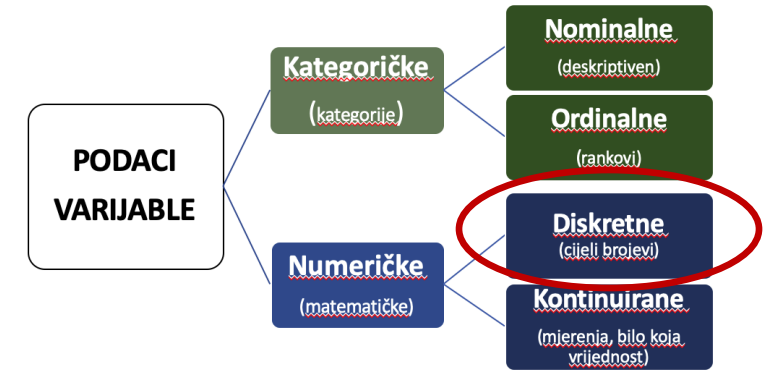


NUMERIČKE (KVANTITATIVNE) VARIJABLE

- Mjerimo ih ili brojimo

Diskretne (numeričke) varijable

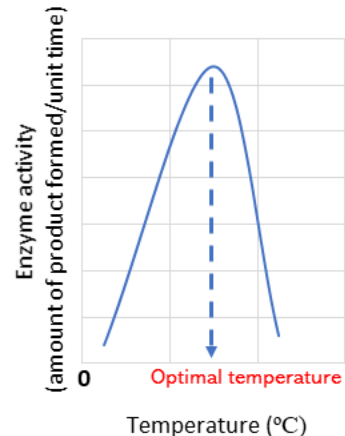
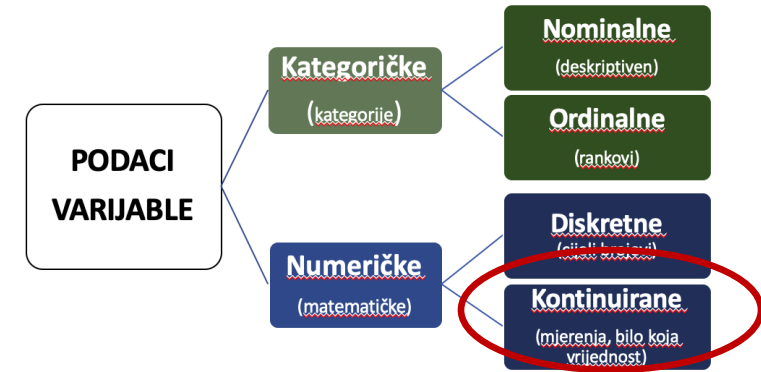
- Cijeli brojevi (nisu moguće sve numeričke vrijednosti)
- Uglavnom brojnosti
- *Broj sjemenki, duljina sekvence (bp), broj mladunaca, broj zamaha krilima*
- *Merističke značajke: broj trnaca na klijestima, šipčica u peraji*



U slučajevima kada je mali broj mogućih vrijednosti, bolje ih je tretirati kao ordinalne (nominalne) varijable

Kontinuirane (numeričke) varijable

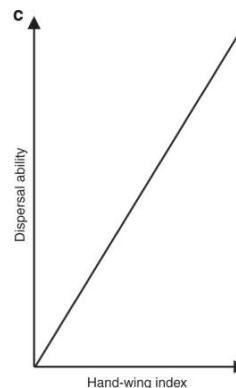
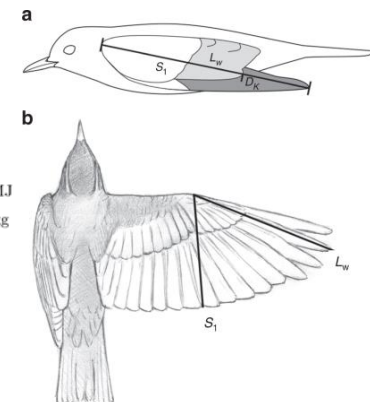
- Uglavnom mjerenja
 - Realni brojevi (najčešće racionalni)
 - Moguće sve numeričke vrijednosti u intervalu
 - Sadržaj informacije i moć statističke analize je mnogo veća!
 - Statistički testovi kompliciraniji jer je neophodno provjeriti distribuciju i druge pretpostavke
-
- *Morfometrijske veličine (masa, visina, širina), količina lipida u stanici, gustoća populacije, aktivnost enzima*
 - *Veličine kemijskih i fizikalnih čimbenika (valne duljine, pH, temperatura, koncentracija agensa u tretmanu; put koji pređe makrofag, energetska učinkovitost razmnožavanja)*

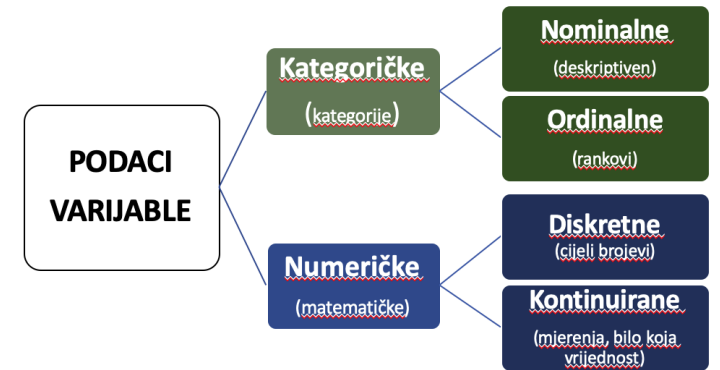


$$\begin{aligned} n &= 1 \\ \sum R_i &\approx 61 \text{ MJ} \\ \sum M_i &= 30 \text{ kg} \end{aligned}$$



$$\begin{aligned} n &= 37 \\ \sum R_i &\approx 61 \text{ MJ} \\ \sum M_i &= 74 \text{ kg} \end{aligned}$$





- Sva četiri tipa varijabli mogu biti i zavisne i nezavisne!