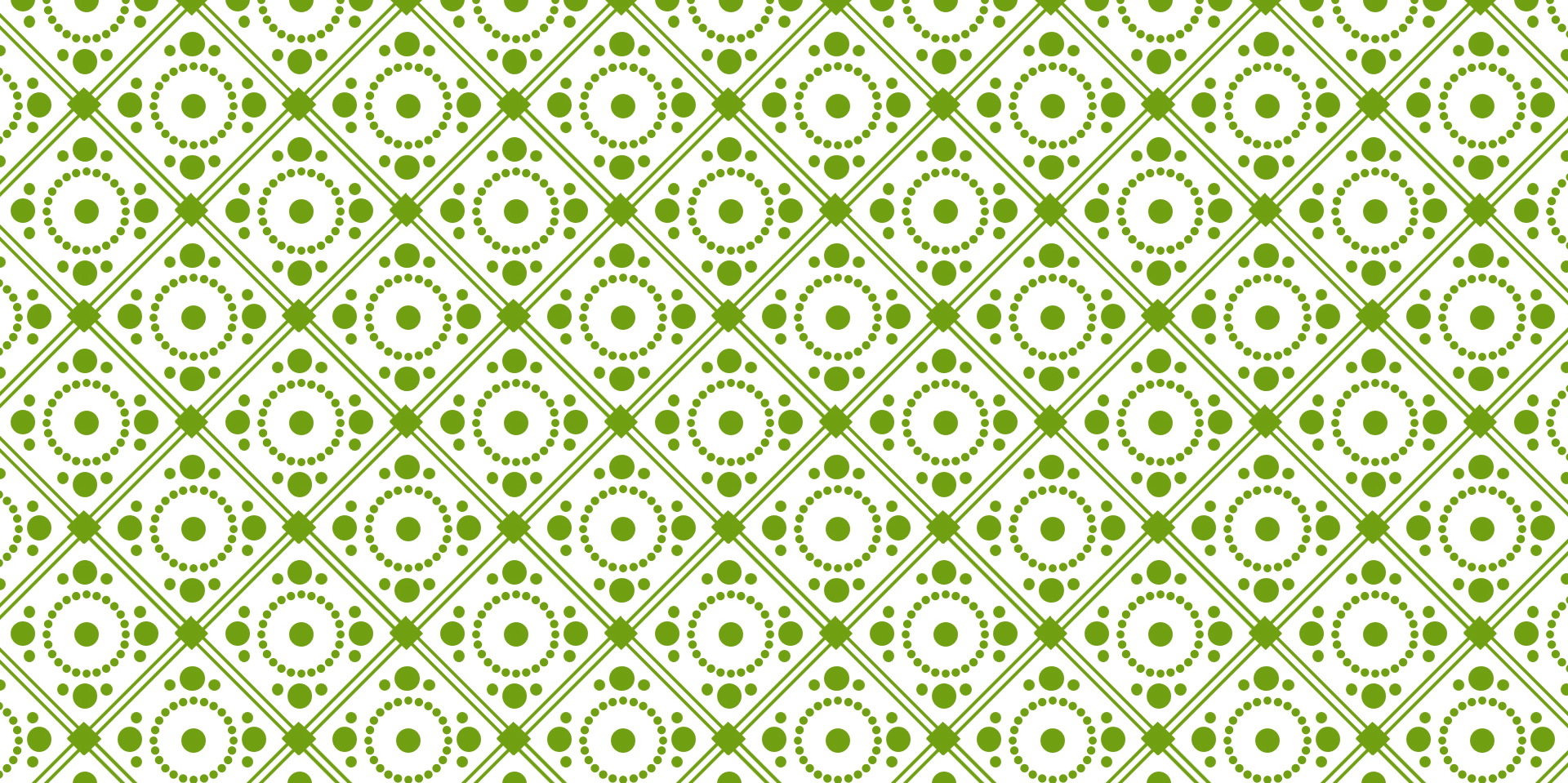




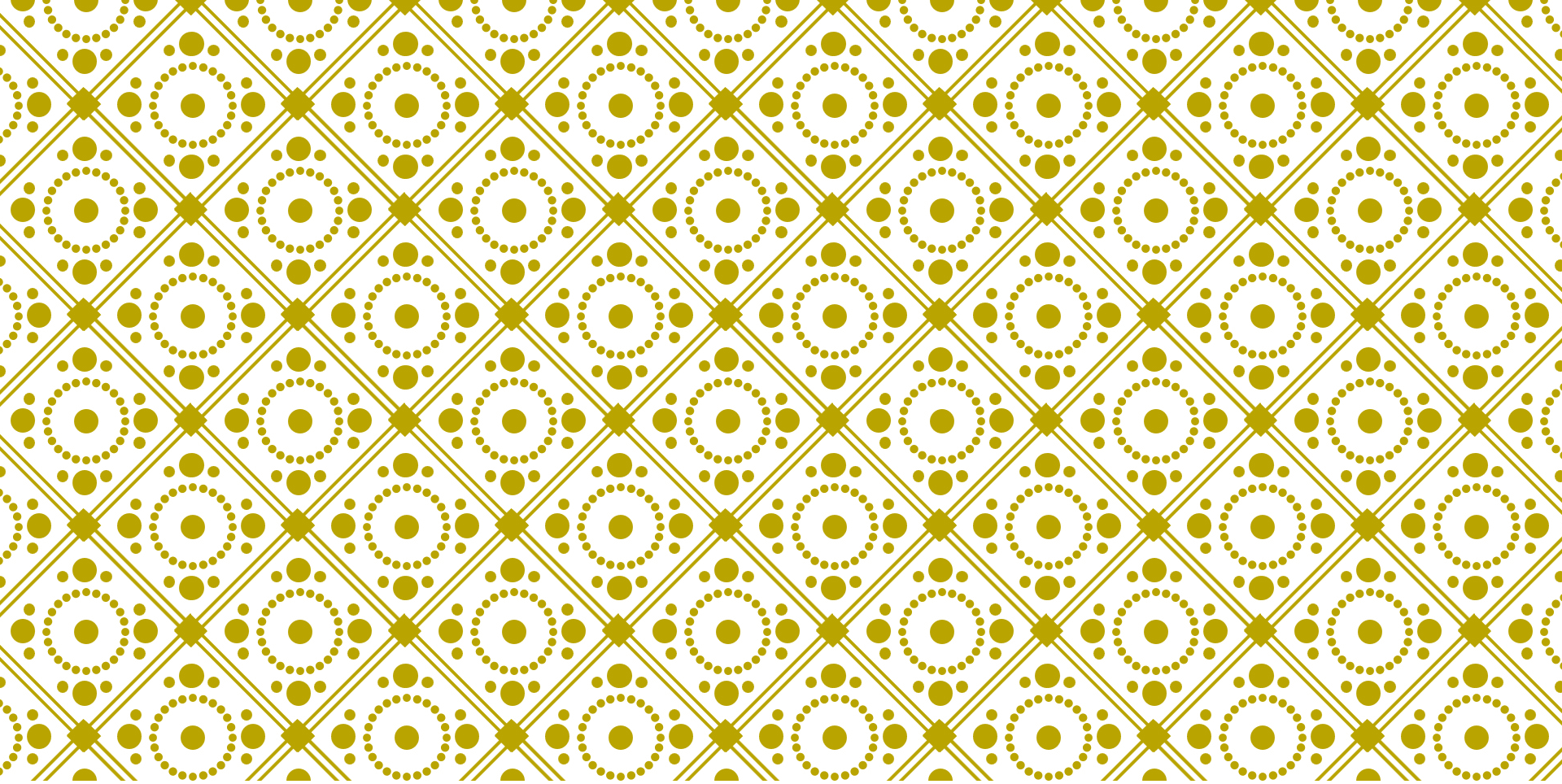
STATISTIČNA ANALIZA PODATKOV

Petra Lazić
FMF - UL | petra.lazic@fmf.uni-lj.si

»**FMF seminar za učitelje
matematike 26/27**
»25. septembra 2026

O ČEM BOMO GOVORILI?

- Statistika je povsod okoli nas!
- Ali lahko učence in dijake motiviramo, da se ukvarjajo s statistiko?
- Viri možnih napak pri sklepanju
- Spletne ankete



NEKAJ O STATISTIKI

- ❖ Zanima nas določen **pojav**
(volitve, rezultati izpitov, krvne preiskave, učinkovitost zdravil, širjenje bolezni, plače, cene, družbena omrežja ...)
- ❖ Zasnujemo **poskus**
(**eksperiment**)
- ❖ Zbiramo **podatke**
- ❖ Izpeljujemo **sklepe** o pojavu, ki ga opazujemo.



PRIMJER: IZBOR ZA NAJBOLJŠI BOŽIČNI SEJEM V EVROPI

- ❖ Pojav = kdo bo zmagal?
- ❖ Eksperiment = virtualna anketa z vprašanjem: „Za koga bi glasovali? Ljubljano, Berlin, ...?“
- ❖ Vzorec = zbranih 5000 glasov
 1. 2000 za Berlin
 2. 500 za Ljubljano
 3. ...
 4. 50 neodločenih ali jih to ne zanima
- ❖ Zaključek = ?
 1. Možnosti za zmago LJU su $500/5000 = 10\%$
- ❖ Težave???
 1. Kdo je izpolnil anketo??
 2. Kolikšno je skupno število volivcev??
 3. Ali smo prepričani o stabilnosti rezultatov??

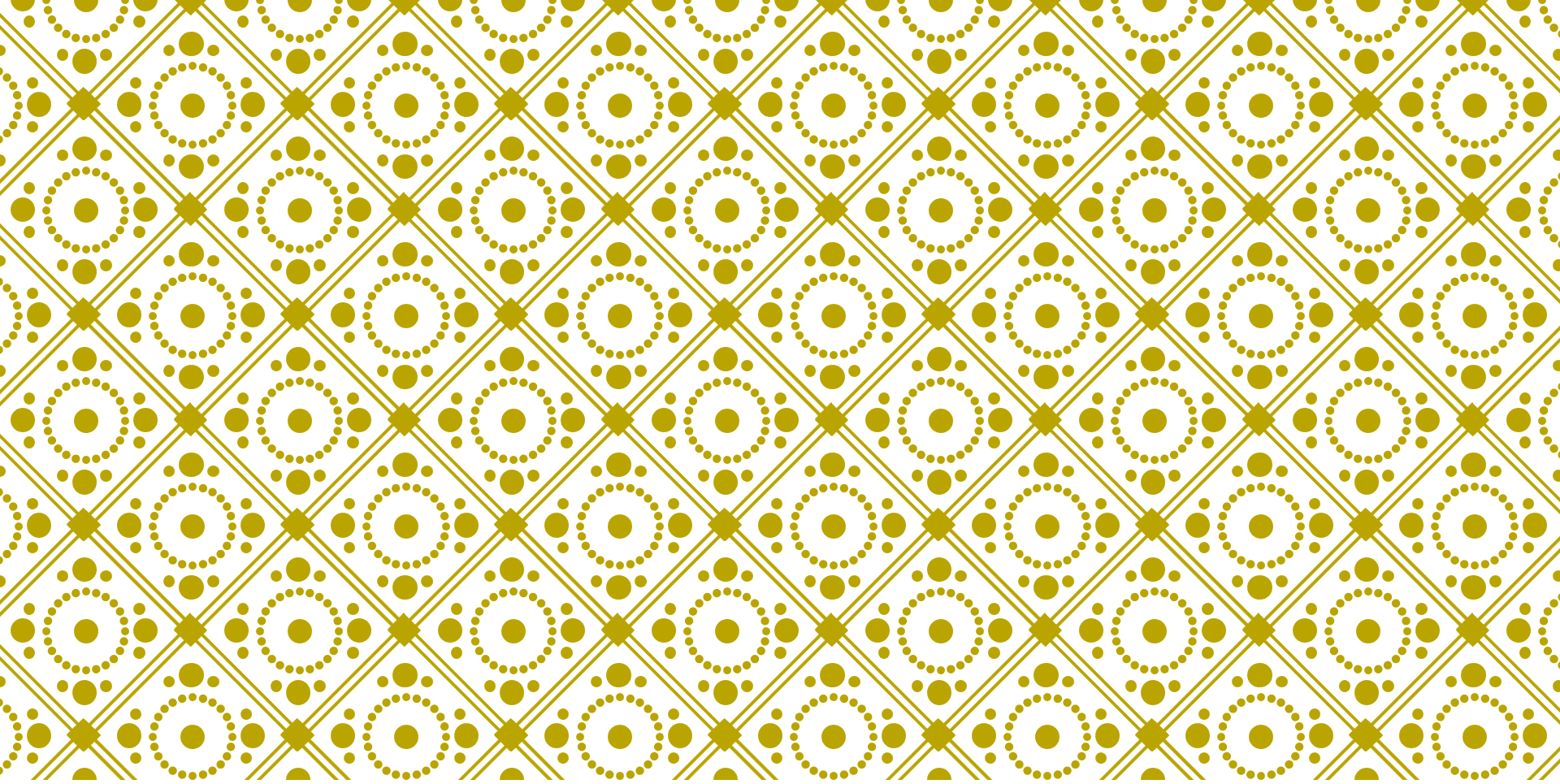


KJE SE STATISTIKA UPORABLJA?

- ❖ **Medicinske vede** (izpopolnjevanje diagnostičnih postopkov; vpliv zdravil na pooperativno bolečino; natančnost medicinskih pripomočkov ...), **farmacevtska industrija** (učinkovitost novih zdravil in neželeni učinki)
- ❖ **Podatkovna znanost** (pomembnost in natančnost napovedi, analiza delovanja algoritmov, verjetnostni algoritmi, ...)
- ❖ **Ekonomija** (tržni in borzni trendi, izračun tveganja, izračun zavarovanja, rezerve ...), **finance**
- ❖ Druge **vede** (eksperimenti, članki)
- ❖ **Šport** (spletne stavnice, prestopi igralcev, analize F1)
- ❖ **Družbene vede** (psihologija, muzikologija, sociologija, ...)

STATISTIČNA ANALIZA FINANČNIH PODATKOV

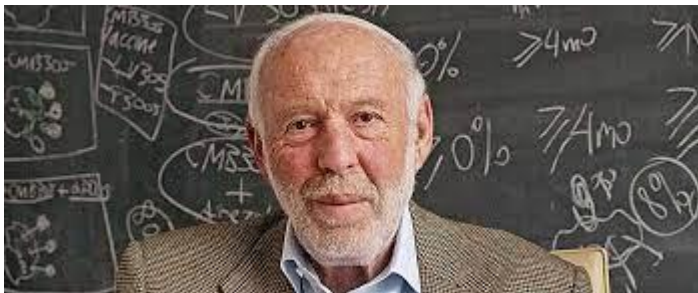
- ❖ Valutna inflacija
- ❖ Cene delnic in obveznic
- ❖ Različni kazalniki gospodarske rasti



NEKAJ O STATISTIKIH

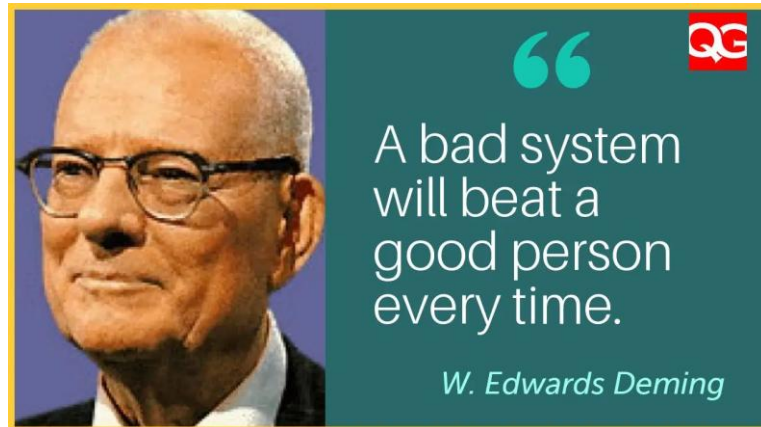
Jim Simons **"It's mostly statistics. It's mostly statistics and some probability theory."**

- (2018) #23 najbogatejša oseba po Forbesovi lestvici, #1 hedge fund upravljavec



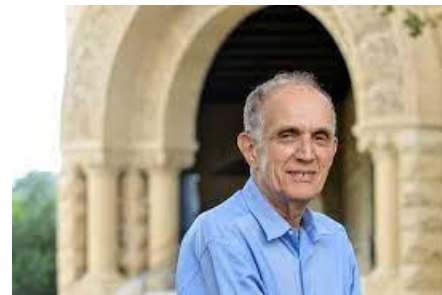
H.G.Wells **"Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."**

- Angleški avtor zf knjig



W. Edwards **"Without data, you're just another person with an opinion."**

- Ameriški ekonomist, teoretik, matematik, ekonomist, inženir ...



Brad Efron **"Those who ignore Statistics are condemned to reinvent it."**

- Profesor statistike na Stanfordu, nekdanji predsednik Ameriškega statističnega združenja

STATISTIK: POKLIC SEDANJOSTI IN PRIHODNOSTI

❖ US News:

- #4 Best Business Jobs
- #9 Best STEM Jobs
- #12 na seznamu 100 najboljših poklicev

➤ Zadovoljstvo pri delu:

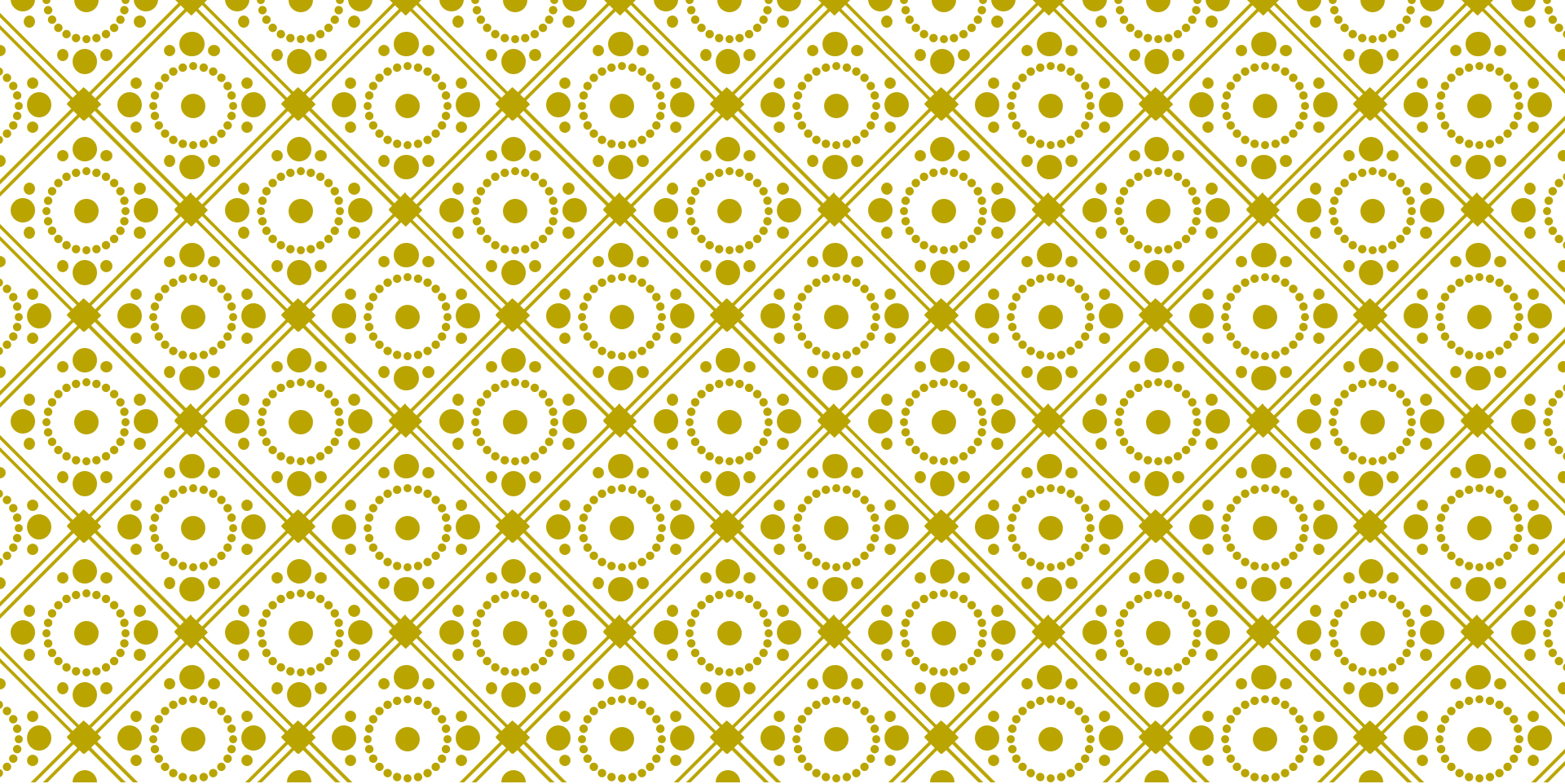
- **Napredovanje in plača:** nadpovprečen
- **Stres:** nadpovprečen
- **Ravnovesje med poklicnim in zasebnim življenjem:** slabše

❖ Ameriška služba za zaposlovanje:

- Povpraševanje po statistikah se bo do leta 2032 povečalo **več kot 30%**

❖ Fortune revija:

- Z diplomo iz statistike je po končanem študiju **najlažje najti zaposlitev**



VIRI MOŽNIH NAPAK PRI SKLEPANJU

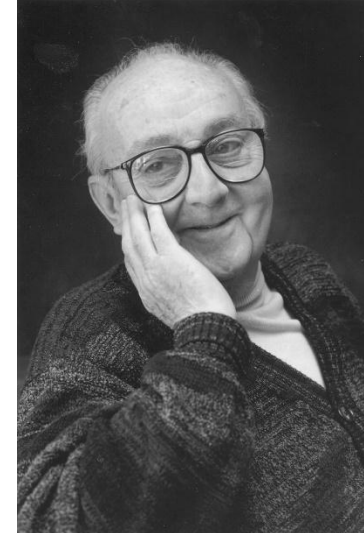
Barbara Doyle **"If you need statistics to prove it, it isn't true."**

➤ Glasbenica in pisateljica



Henry Clay **"Statistics are no substitute for judgment."**

➤ Ameriški senator iz 19. stoletja (nasvet iz 1842.)

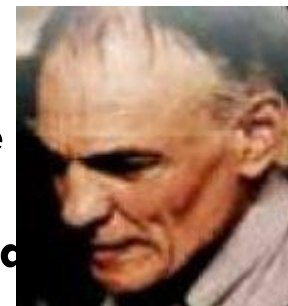


George Box **"All models are wrong, but some are useful."**

➤ Eden od velikih statističnih umov 20. stoletja

Aaron Levenstein **"Statistika je kao bikini. Ono što otkriva je zanimljivo, ali ono što sakriva je esencijalno."**

➤ Ameriški univerzitetni profesor in ekonomist



TEŽAVE Z VZORCEM

❖ Vprašanje: Ali je IQ povezan z velikostjo stopal?

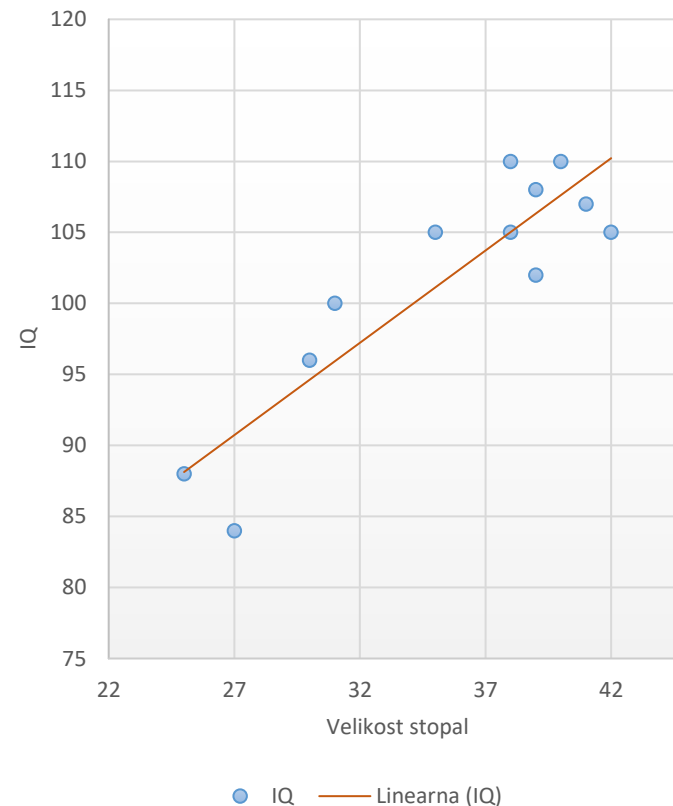
➤ Zbran vzorec 12 ljudi

❖ Zaključek: Ljudje z daljšimi stopali so pametnejši (?!)

❖ Problem: v študijo vključeni majhni otroci

➤ Za katero populacijo želimo izvesti raziskavo?

➤ VZOREC MORA BITI **REPREZENTATIVEN**



❖ **Vprašanje:** Kakšne so navade prebivalcev Ljubljane glede pitja kave?

- V ljubljanski kavarni so anketirali 20 ljudi: 18 mlajših in 2 starejši osebi
- Starejši osebi kavo pijeta največ 5 minut, mlajše vsaj pol ure

❖ **Zaključek:** Mlajši ljudje potrebujejo več časa, da spijejo svojo kavo, kot starejši(?!)

❖ **Problem:** 20 ljudi ne more natančno predstavljati navad vseh prebivalcev

- Za katero populacijo želimo izvesti raziskavo?
- VZOREC MORA BITI **DOVOLJ VELIK**



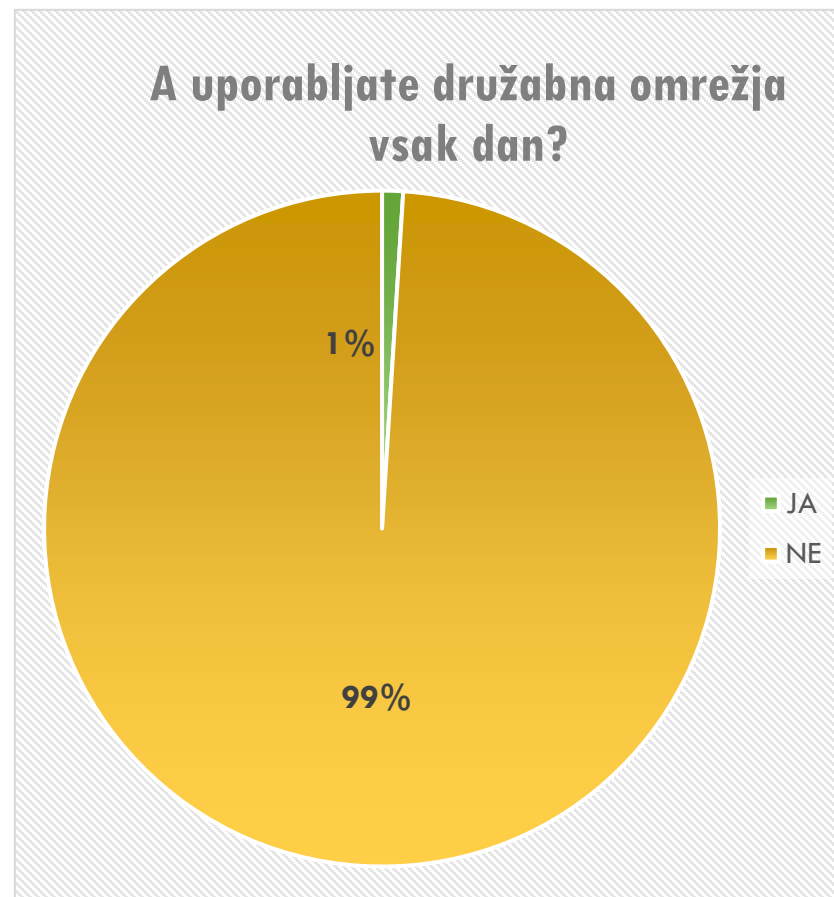
❖ **Hipoteza: Slovenci ne preživijo veliko časa na družbenih omrežjih**

- Anketiranih je bilo 10000 starejših odraslih (vsi starejši od 65 let)

❖ **Zaključek: Slovenci ne uporabljajo veliko družbenih omrežij (?!)**

❖ **Problem: selektivno izbiranje anketirancev, ki podpira našo hipotezo**

- Za katero populacijo želimo izvesti raziskavo?
- Kdo je zbiral podatke in kako?
- VZOREC MORA BITI **REPREZENTATIVEN**



TEŽAVE Z VPRAŠANJI

❖ **Preveliko število vprašanj**, ki odvrnejo anketiranca (kdo si želi 30 minut izpolnjevati vprašalnik?)

❖ **Intimna vprašanja**, na katera ljudje ne želijo iskreno odgovoriti (npr. ne verjamejo v zasebnost podatkov)

❖ V kakšni obliki naj bo vprašanje zastavljeno, da bo mogoče podatke analizirati?

➤ Npr. Starost kot številka (18, 32, ...) ali razpon (20-30, 30-40, ...)

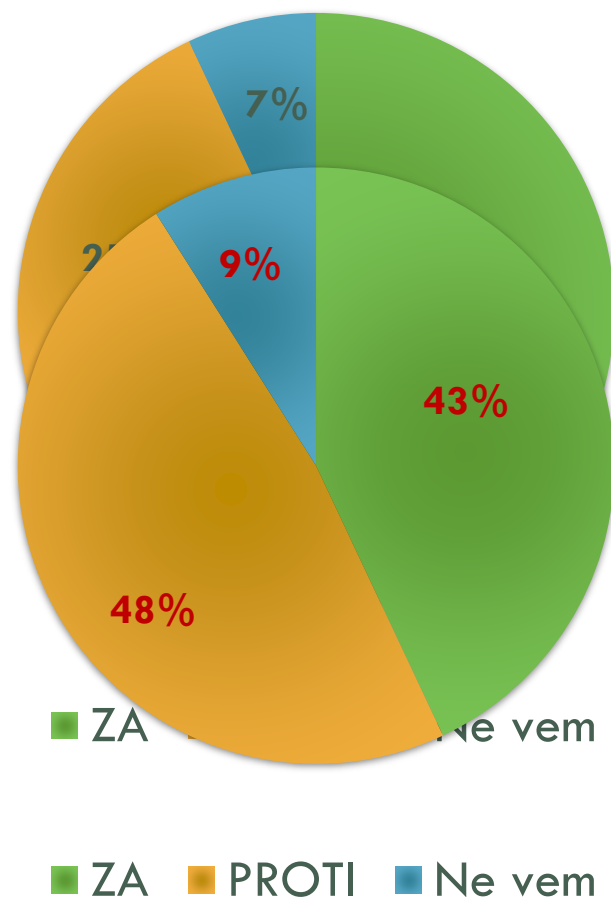
TUDI VPRAŠANJ SI NI LAHKO IZMISLITI.

❖ **Slaba vprašanja = slabi podatki = napačni zaključki**

❖ Psihologija v vprašanjih

- Odgovori so odvisni od načina, kako je vprašanje zastavljeno
- **Različica 1:** Ste **ZA** ali **PROTI** varčevanju v bitcoinu, ki je priljubljen zaradi možnosti visokega zaslužka?
- **Različica 2:** Ste **ZA** ali **PROTI** varčevanju v bitcoinu, ki je priljubljen zaradi možnosti visokega zaslužka, čeprav gre za eno najbolj tveganih metod zaradi nestanovitnosti kriptovalut?

❖ LJUDJE NISO RACIONALNI



❖ Vprašanje za dve skupini zdravnikov glede odločitve o zdravljenju bolnika s tumorjem

➤ Različica 1: Operacija ili radioterapija:

Operacija je **dolgoročno** uspešnejša, vendar prinaša **verjetnost smrti 10%** v prvem mesecu po operaciji.

✓ **50%** za operacijo

➤ Različica 2: Operacija ili radioterapija:

Operacija je **dolgoročno** uspešnejša, z **verjetnostjo preživetja 90%** en mesec po operaciji.

✓ **84%** za operacijo

Tudi zdravniki so samo **LJUDJE** ...



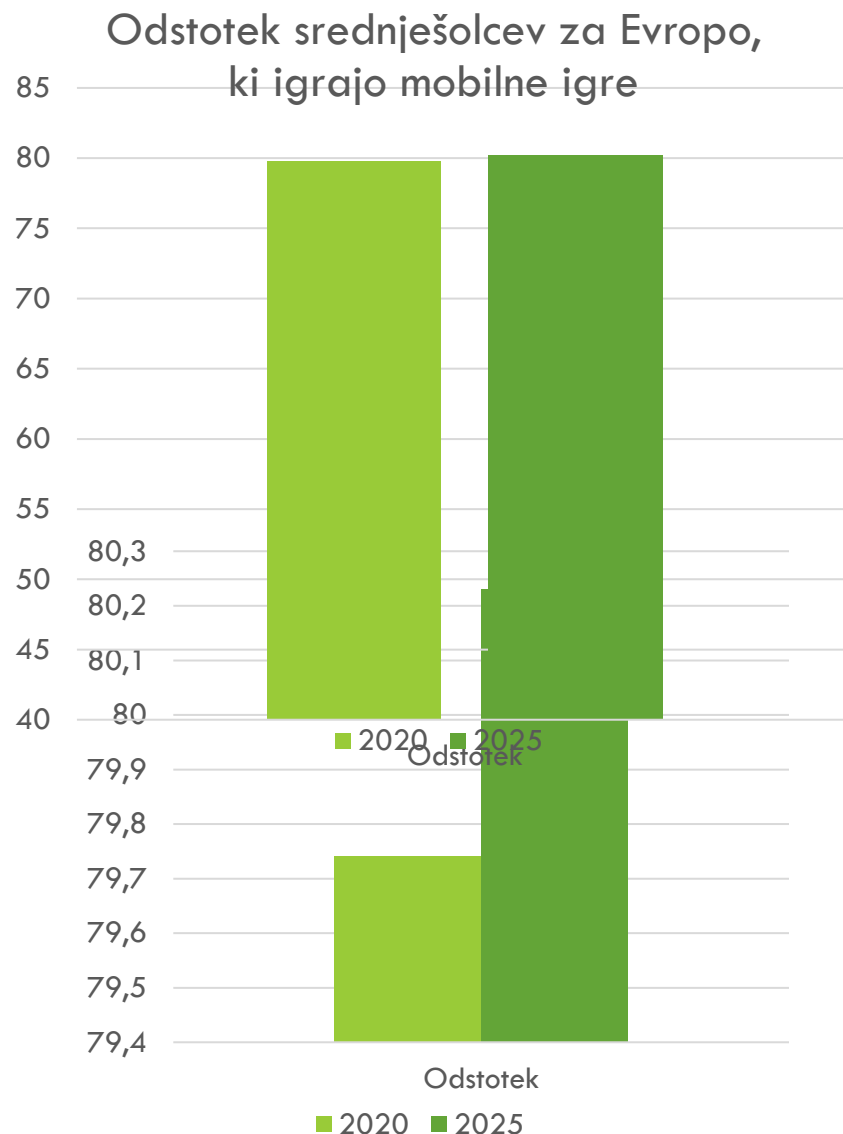
TEŽAVE Z GRAFI

❖ **Vprašanje:** ali se je znatno povečalo število srednješolcev, ki igrajo mobilne igre?

❖ **Zaključek:** JA

❖ **Problem:** y-os ni prikazana od 0

➤ Razpon med osema je **preveč zmanjšan**, pa se izgubi prava slika števil

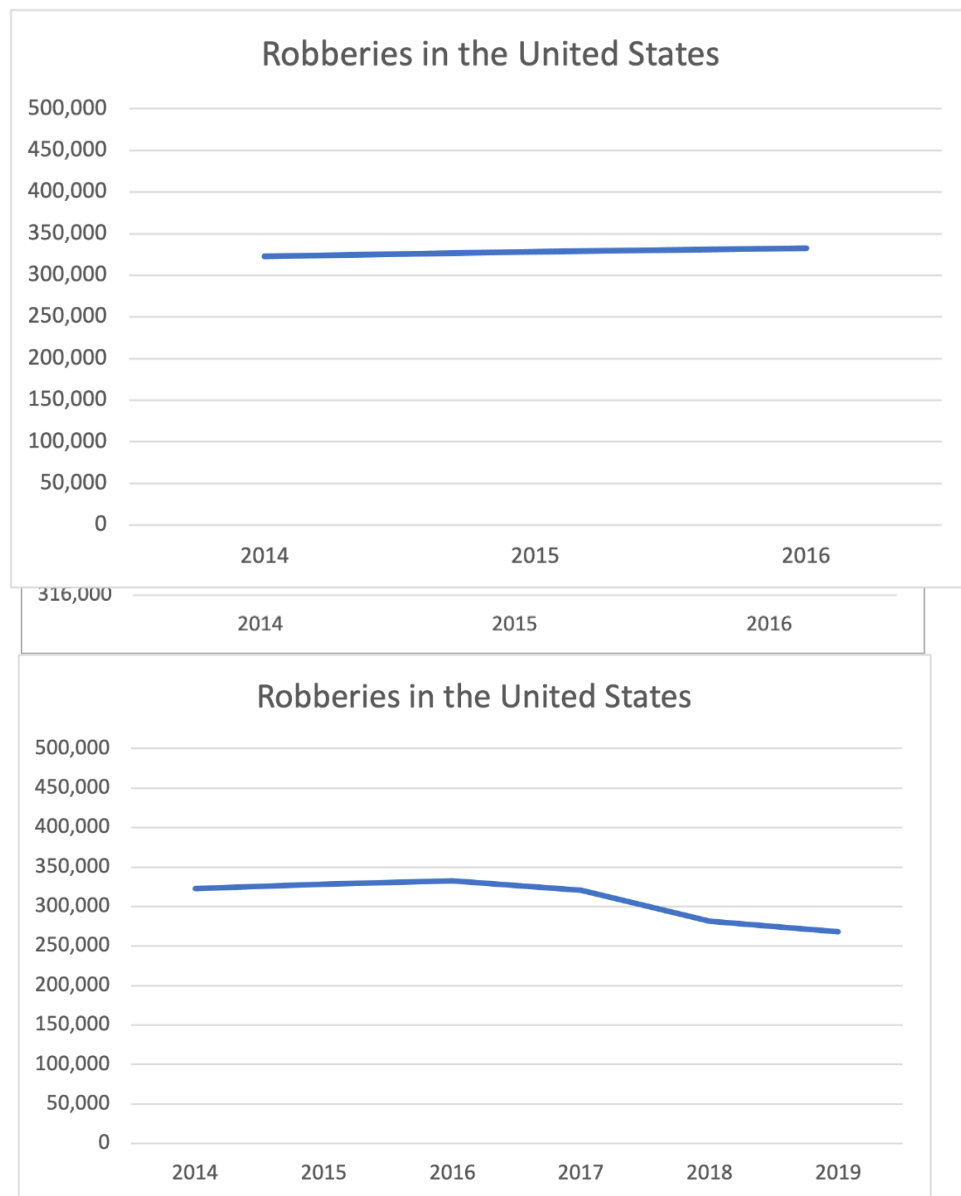


❖ **Vprašanje: Ali je v ZDA od leta 2014 prišlo do porasta tatvin?**

❖ **Zaključek: JA**

❖ **Problem:** y-os ni prikazana od 0, na x-osi prikazan le del podatkov

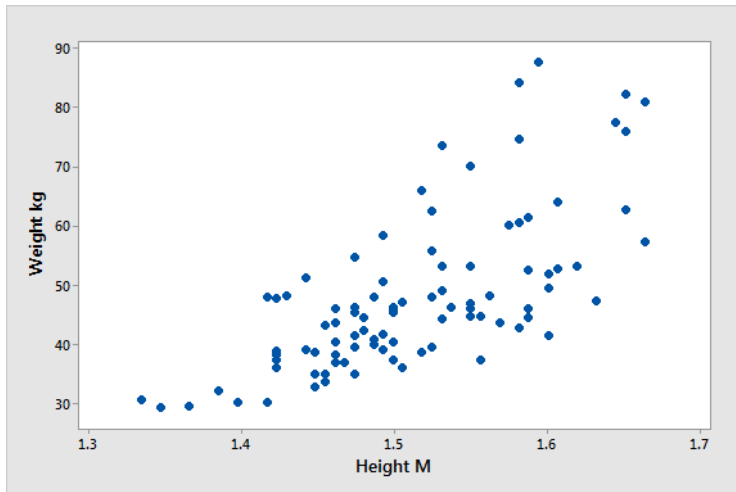
- Razpon med osema je **preveč zmanjšan**, zato se izgubi prava slika števil (razpon moramo pogledati v kontekstu podatkov)
- **Izpuščanje določenih podatkov** ustvarja zavajajočo sliko o trendu



VZROČNOST VS. KORELACIJA

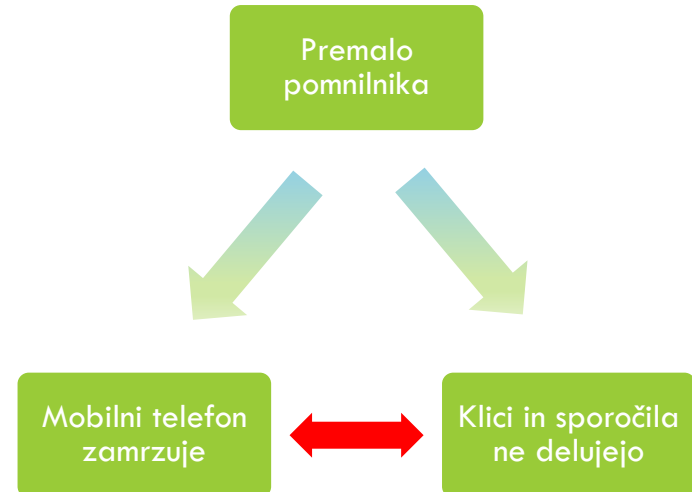
❖ **Korelacija:** dve spremenljivki se gibljeta skupaj

- Ko X narašča, narašča tudi Y
- Ko se X zmanjšuje, se zmanjšuje tudi Y
- Ko X narašča, se Y zmanjšuje



❖ **Vzročnost:** Sprememba ene spremenljivke povzroči spremembo druge spremenljivke

- Ko X narašča, povzroča, da Y narašča
- Ko X narašča, povzroča, da se Y zmanjšuje



KORELACIJA NE POMENI VZROČNOST

❖ Ugotovili smo, da sta spremenljivki **A in B korelirani**, možni **razlogi** vključujejo:

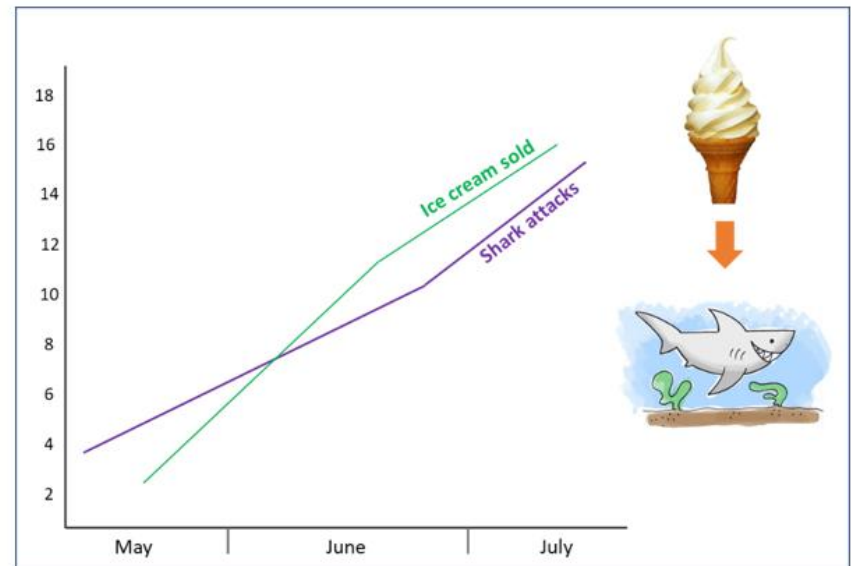
- A povzroči B
- B povzroči A
- A in B povzročata drug drugega
- A in B povzroča nek tretji dejavnik C
- B je povzročen od C, ki je koreliran z A
- A in B nista povezani, opažena korelacija je posledica naključja
- ...

❖ **Dokazano je, da je število zaužitih sladoledov močno povezano z napadi morskih psov.**

❖ **Zaključek:** več sladoleda kot ljudje pojedjo, več bo napadov morskih psov (?!)

❖ **Problem:** korelacija ne pomeni vzročnosti

- Oba dogodka **sta povzročena s tretjim dogodkom** – s poletno vročino
- Med spremenljivkami obstaja korelacija, ne pa tudi vzročnost

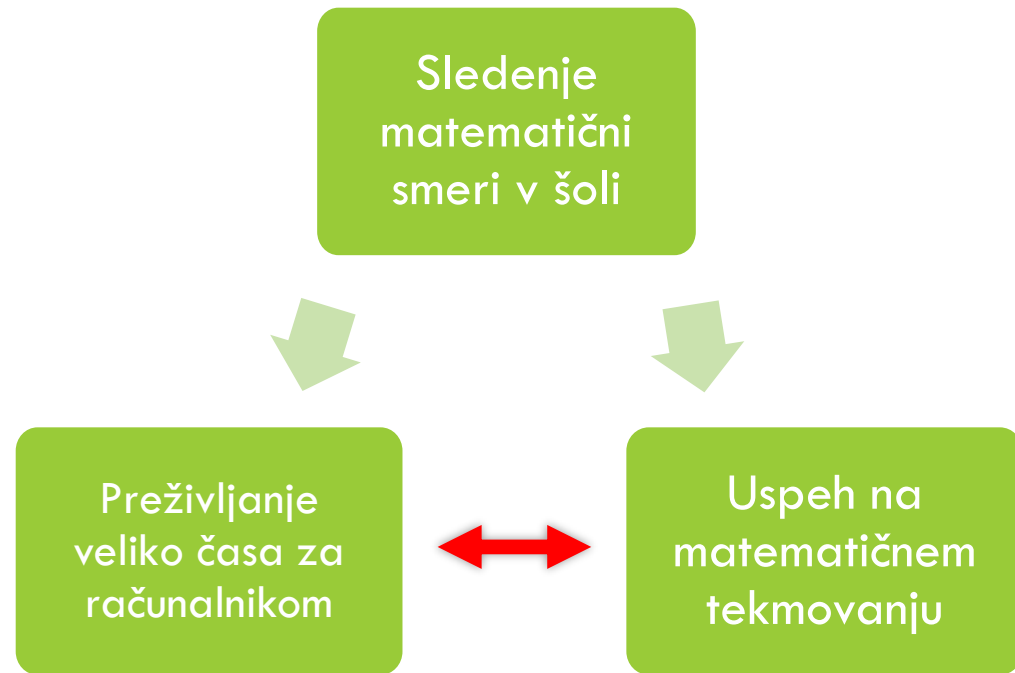


❖ **Velik delež uspešnih učencev na matematičnem tekmovanju je povedal, da veliko časa preživijo za računalnikom.**

❖ **Zaključek:** če bomo veliko časa preživeli za računalnikom, bomo imeli več možnosti, da se uvrstimo na matematično tekmovanje (?!)

❖ **Problem:** korelacija ne pomeni vzročnosti

- Oba dogodka **sta povzročena s tretjim dogodkom** – s sledenjem matematični smeri v šoli
- Med spremenljivkami obstaja korelacija, ne pa tudi vzročnost



❖ **Statistični podatki kažejo, da se zgodi zelo malo nesreč pri hitrosti vozil nad 150 kilometrov na uro.**

❖ **Zaključek:** moramo voziti zelo hitro (?!)

❖ **Problem:** korelacija ne pomeni vzorčnosti

- Majhno število nesreč **je povzročeno s tretjim dogodkom** – le redki avtomobili vozijo s hitrostjo nad 150 km/h zaradi omejitev hitrosti in kazni
- Med spremenljivkami obstaja korelacija, ne pa tudi vzročnost

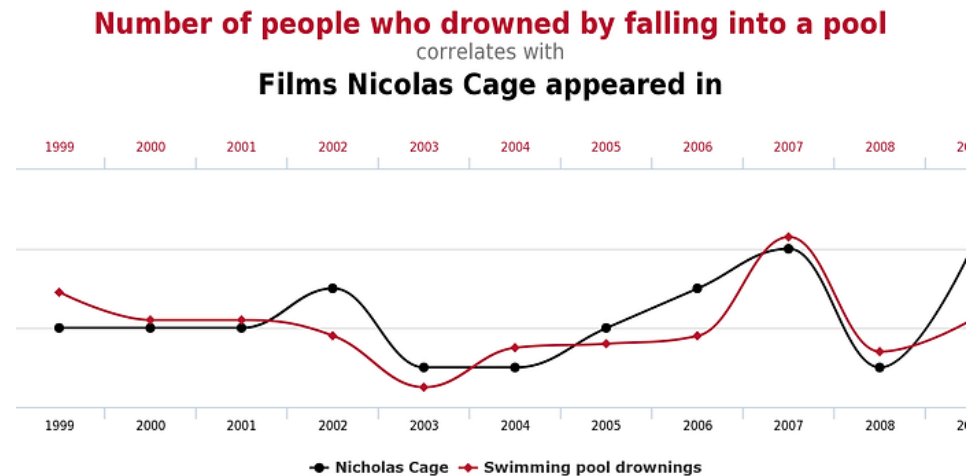


❖ Podatki kažejo, da so leta, v katerih je Nicolas Cage nastopil v več filmih, povezana z leti, v katerih je v bazenih utonilo veliko število ljudi.

❖ **Zaključek:** Nicholas Cage je kriv za tragične dogodke (?!)

❖ **Problem:** korelacija ne pomeni vzorčnosti

➤ Verjetno ni nobene razlage, povezava je **naključna**



SIMPSONOV PARADOKS

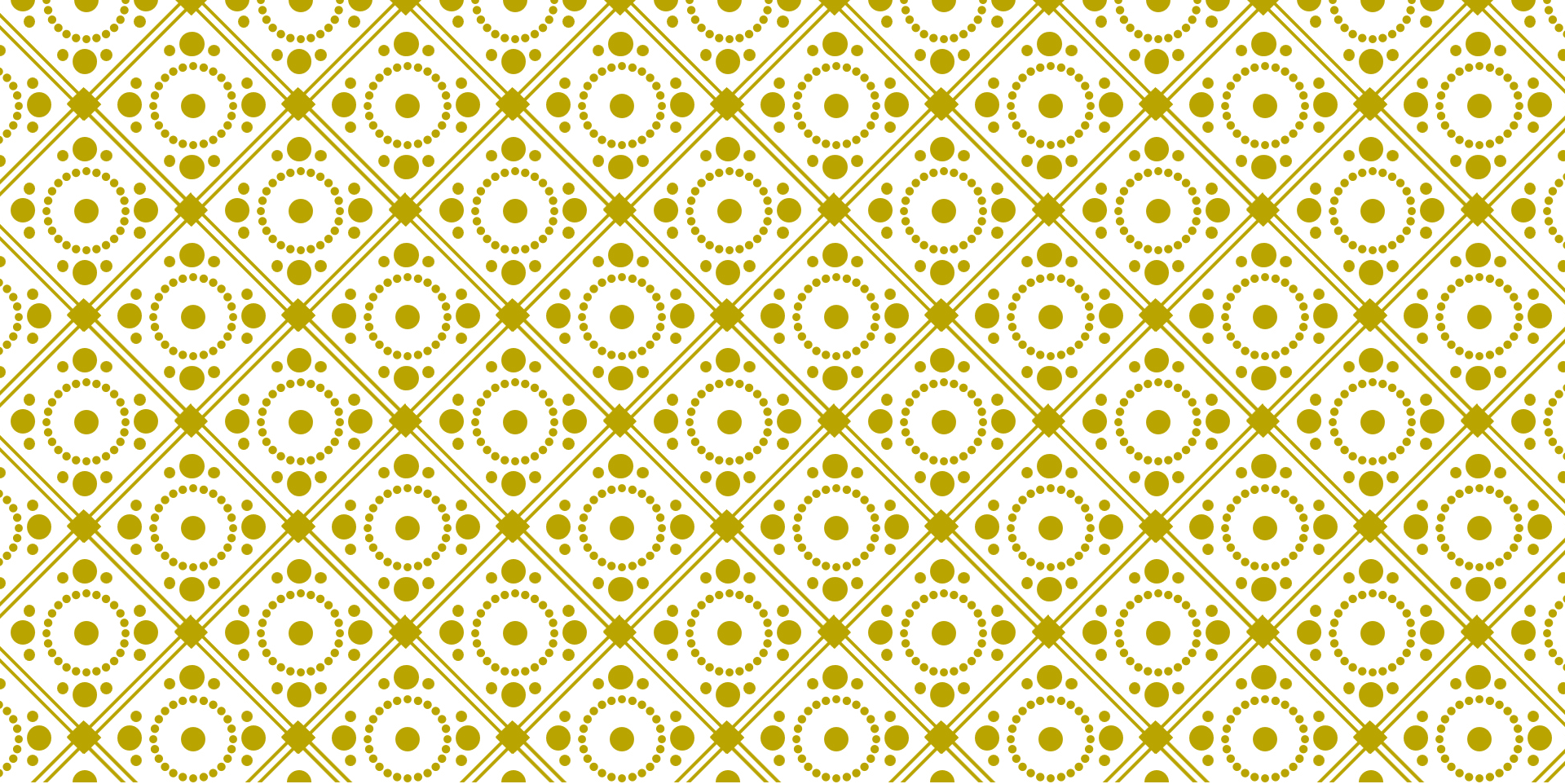
❖ Možne napake:

- **prekomerno posploševanje** (s podpopulacije na celotno populacijo – volitve)
- **pretirana specializacija** (od populacije k subpopulaciji)

❖ Primer: odstotek žensk v britanskem parlamentu

- Skupaj: **650** členov (*House of Commons*)

LETO	VSE STRANKE		KONZERVATIVCI		LABURISTI		LIBERALNI DEMOKRATI	
	Skupaj	Odstotek žensk	Skupaj	Odstotek žensk	Skupaj	Odstotek žensk	Skupaj	Odstotek žensk
2024	650	40%	121	24%	413	46%	73	44%
2019	650	34%	322	27%	260	40%	12	58%



NEKAJ O IZVAJANJU ANKET

SPLETNE ANKETE

❖ **Definiranje eksperimenta:** oblikovanje hipoteze, postavljanje vprašanj

❖ **Zbiranje podatkov:**

❖ Google forms

❖ 1KA (razvito na Fakulteti za družbene vede u Ljubljani - podpira ustvarjanje vprašalnikov s pomočjo umetne inteligence)

❖ Survio

❖ **Problemi:** kako postaviti vprašanje, kako oblikovati možne odgovore, da jih je enostavno analizirati (razpon, ...)

❖ **Statistična analiza podatkov:**

❖ Upoštevajte morebitne **težave s podatki ali z vprašanji** ter ustrezno prilagodite sklepe (Kdo je izpolnil vprašalnik? Kako prepričani smo v rezultate?)

❖ Skrbno izberite grafične prikaze (ne smejo ničesar prikriti ali voditi do napačnih rezultatov)

❖ Kritično vrednotenje informacij (upoštevati vse okoliščine, povezane z izvedbo poskusa ter metodo zbiranja in obdelave podatkov)

POSTAVLJANJE VPRAŠANJ

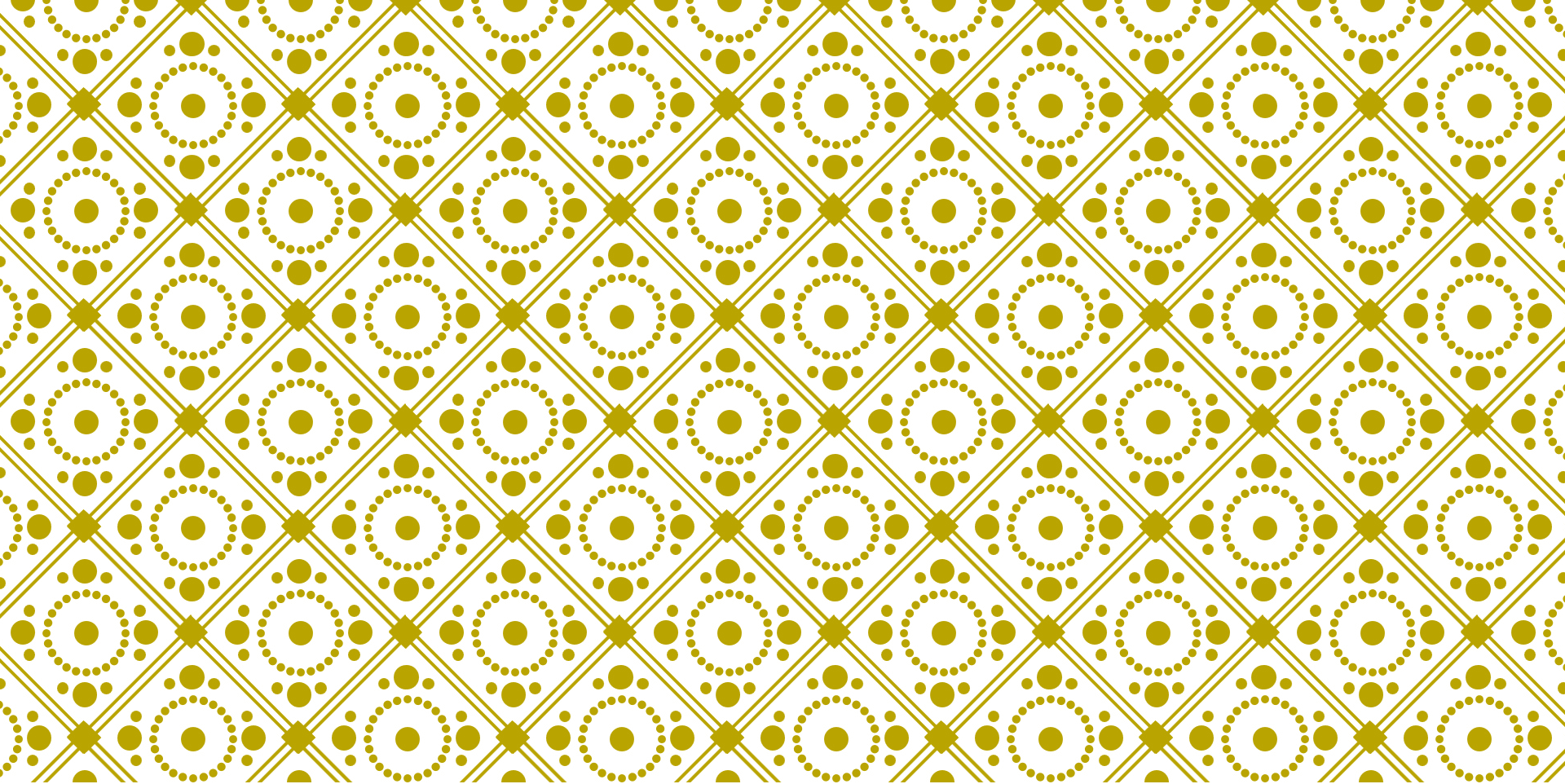
- ❖ **Jasnost in jedrnatost:** vprašanja morajo biti kratka, enostavna in nedvoumna. Izogibajte se strokovnemu žargonu.
- ❖ **Eno vprašanje hkrati:** ne združujte dveh tem v eno vprašanje (npr. „Kako ocenjujete naše cene in ljubeznivnost?“)
- ❖ **Nevtralnost:** vprašanja ne smejo napeljevati k določenemu odgovoru
- ❖ **Uravnotežene lestvice:** Pri ocenjevalnih lestvicah (npr. **Likertova lestvica: odgovori 1 - 7**) ponudite enako število pozitivnih in negativnih odgovorov ter nevtralno sredino
- ❖ **Bodite pozorni na pravne in etične vidike:** GDPR, anonimnost (če je anketa anonimna, ne smete zbirati nobenih podatkov), intimna vprašanja (vplivajo na iskrenost udeležencev)

TEHNIČNA IZVEDBA IN DOSTOPNOST

- ❖ **Prilagojenost za mobilne naprave**
- ❖ **Kratko trajanje**
- ❖ **Indikator napredka:** spodbuja uporabnike, da anketo izpolnijo do konca
- ❖ **Vidnost:** kdo ima dostop do ankete, kako motivirati uporabnike, kako in kje oglaševati anketo

TEŽAVE Z VZORCEM

- ❖ **Vzorec iz priročnosti:** izbira enot, ki so najbližje ali najbolj dostopne (npr. anketiranje učencev lastni šoli)
- ❖ **Samoizbira:** ljudje sami odločijo za sodelovanje (npr. spletne ankete v medijih). V takšnih anketah pogosto sodelujejo le ljudje z ekstremnimi ali zelo močnimi mnenji
- ❖ **Izključitev delov populacije:** nekateri deli populacije sploh nimajo možnosti, da bi bili izbrani
- ❖ **Težave z velikostjo vzorca:** premajhen ali prevelik
- ❖ **Bodite pozorni, kdo je izpolnil vprašalnik!**



NEKAJ O ZNANSTVENIH TEORIJAH

KAKO STATISTIKA FUNKCIONIRA?

❖ Postavljata se **dve hipotezi**

- H_0 : **osnovna hipoteza** - velja za pravilno, dokler podatki ne pokažejo, da je napačna
- H_A : zanjo iščemo dokaze - kadar obstajajo zadostni dokazi, se **alternativna hipoteza** šteje za pravilno

❖ Vprašanje: Ali je oseba storila kaznivo dejanje?

❖ Hipotezi:

- H_0 : oseba je nedolžna
- H_A : oseba je kriva

❖ Zaključek: privzete nedolžnosti – oseba je nedolžna **DOKLER** ji krivda ni dokazana onkraj razumnega dvoma

- Če ne zavrnilo H_0 , to ne pomeni, da je pravilna, ampak nimamo dovolj dokazov, da bi jo označili za nepravilno



Poskus
ugibanja

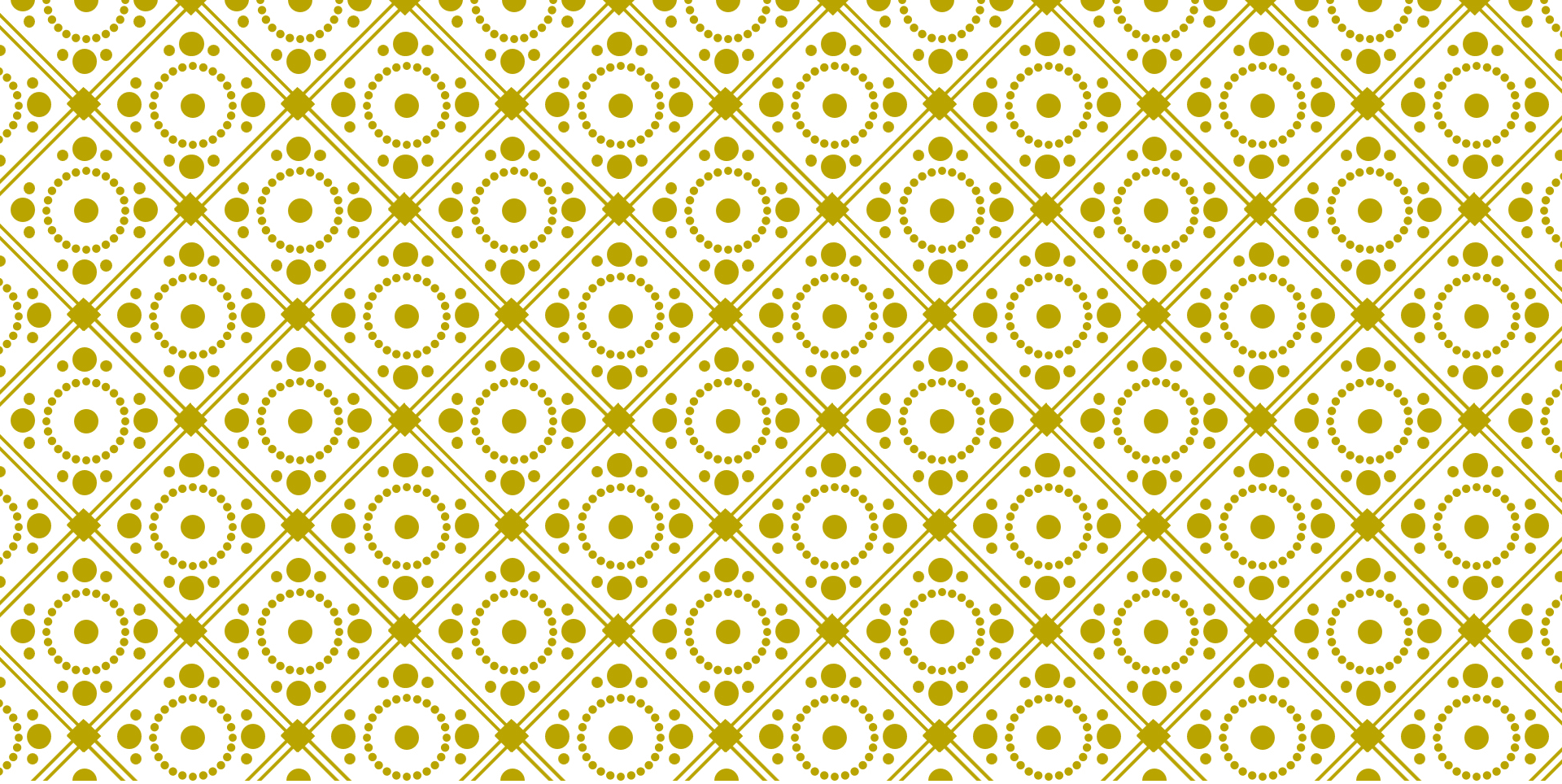
Lastnosti in
posledice

Eksperimenti

Teorija ŠE ni
bila
ovržena.

Teorija se je
izkazala za
napačno.

Ne moremo
biti prepričani,
da imamo
prav in samo
smo lahko
prepričani, da
se motimo!



**ALI JE STATISTIKA TOREJ
SMISELNA??**

❖ Brez statistike ne gre, **A** z obvezno uporabo **zdrave pameti**

❖ Statistika je nepogrešljiva veščina v vsakem poklicu in na vseh znanstvenih področjih

- Vedno želimo potrditi svoje hipoteze in teorije
- Ni brezposelnih statistikov

❖ Kaj smo se naučili?

- Pomembno je razumeti statistične koncepte in postopke
- Bodite pozorni na zbiranje podatkov
- Pri razlagi podatkov bodite previdni
- Iščemo dokaze, da se motimo, a ne moremo potrditi, da imamo prav