

- Tijekom istraživanja priključuju se opažanja za određene subjekte iz populacije ili uzorka
- Ako se radi o uzorku on mora predstavljati populaciju
- Moramo imati replike
- Rezultati opažanja variraju
- Matematička/statistička procjena, testiranje hipoteze

Testiranje hipoteze mora osigurati

MORAMO

- Vidjeti ono što postoji
- Ne vidjeti ono što ne postoji

NE SMIJEMO

- Vidjeti ono što ne postoji
- Ne vidjeti ono što postoji

To moramo osigurati u eksperimentalnom dizajnu

Statističko testiranje hipoteze

PITANJE

- Gradonačelnik ste grada od milijun stanovnika
- 10000 su kriminalci
- U gradsko poglavarstvo želite ugraditi sigurnosni sustav koji će prepoznavati kriminalce
- Dobili ste ponudu za uređaj koji će se oglasiti ako u poglavarstvo uđe kriminalac, uređaj će uvijek identificirati kriminalca
- Biste li ugradili taj uređaj?

Vjerojatnost da ćemo se prevariti i krivo prepoznati poštene građane (kao kriminalce)

Međutim, uređaj ima mali problem. Ako pošteni građanin uđe u Poglavarstvo, postoji mala vjerojatnost (3%) da će se alarm oglasiti. Mijenja li to vašu odluku?

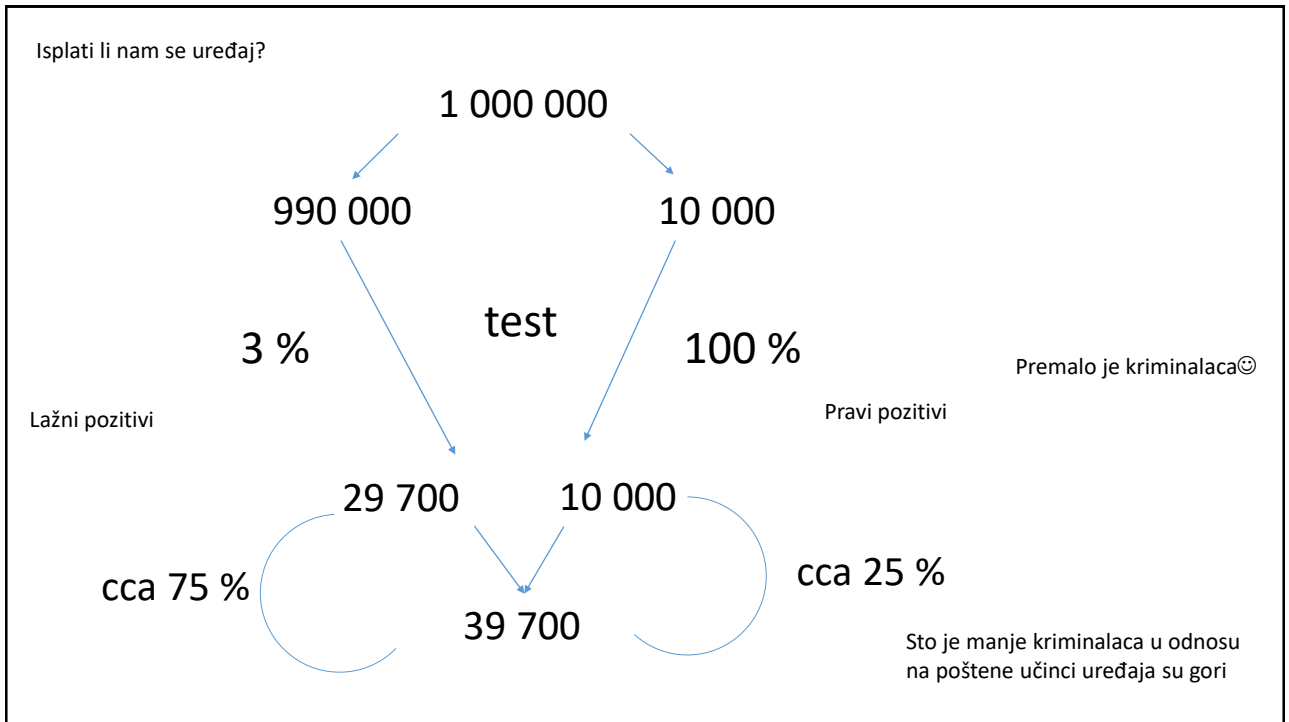
Vidjeti ćemo efekt kojeg u biti nema

PROBLEM

- Prije nego što se oglasio alarm, nismo mogli znati je li osoba koja ulazi kriminalac ili ne. Mogli smo samo znati da je ta vjerojatnost 1%
- Nakon što se oglasio alarm i dalje ne možemo znati je li ta osoba kriminalac. Međutim sad znamo da je ta vjerojatnost puno veća (25%), uvjetna vjerojatnost
- Tu vjerojatnost nazivamo uvjetnom: **vjerojatnost da je osoba kriminalac ako se alarm oglasio**
- Isto tako mogli bismo reći kolika je vjerojatnost da je osoba kriminalac ako znamo da se alarm nije oglasio: 0% (alarm će prepoznati svakog kriminalca 100%)

Vjerojatnost da ćemo **točno prepoznati kriminalce, vidimo efekt kojeg ima**

Vjerojatnost da ćemo se prevariti i **krivo prepoznati kriminalce** (kao poštene), **ne vidimo efekt koji postoji**



- DZ ako je ukupna populacija milijun nadjite minimalan broj kriminalaca da bi postotak laznih i pravih pozitivna bio 50:50

I u biološkim istraživanjima moramo biti oprezni

- vidimo li ono što stvarno postoji
- Ili vidimo da postoji ono što ne postoji
- Ili ne vidimo što postoji
- Ili ne vidimo ono što ne postoji

MORAMO MOĆI RAZLIKOVATI H_0 i H_A

HIPOTEZE

U znanosti uobičajeno je postaviti dvije hipoteze:

H_A alternativnu hipotezu, pretpostavku da je fenomen koji proučavamo stvaran,

H_0 nul-hipotezu, pretpostavku da je fenomen koji proučavamo samo plod naše mašte.

PRIMJER BIOLOŠKOG PROBLEMA

- Ima li hrana za piliće GMO porijekla negativan utjecaj na prirast njihove mase?

IZMIŠLJENI PODACI



H_0 Razlika koju vidimo nema veze sa GMO (rezultat je slučajan)

H_A Razlika koju vidimo povezana je sa GMO

PROMJENA PITANJA

Kao i kod alarma, treba promijeniti pitanje:

- Početno pitanje: Ima li hrana sa GMO negativan utjecaj na prirast mase ili ne?
- **Drugačije pitanje: Kolika je vjerojatnost da ćemo vidjeti razliku koju imamo na grafu ako su hrana sa i hrana bez GMO jednakovrijedne?**

DISKRETNE I KONTINUIRANE VARIJABLE

- U pitanja s prethodnih slajdova uključene su diskretne varijable (načelno, one koje možemo izbrojiti).
- U većini slučajeva, međutim, bavimo se kontinuiranim varijablama (načelno, one koje ne možemo izbrojiti, ali ih svejedno možemo opisati brojem).

KONTINUIRANA VARIJABLA



- Slično pitanje: Kolika je vjerojatnost da se prisutnost GMO u hrani i izostanak GMO u hrani razlikuju kao na grafu **ako** su obje hrane jednako učinkovite?
- **Dopunjeno pitanje:** Kolika je vjerojatnost da se prisutnost GMO u hrani i izostanak GMO u hrani razlikuju kao na grafu **ili još više** ako su obje hrane jednako učinkovite?

VAŽNAN DIO EXP. DIZAJNA – VELIČINA UZORKA – BROJ REPLIKA

- **Veličina i kvaliteta uzorka određuje točnost procjene!**
- Primjer predizbornih anketa (broj ispitanika koji obećava točnost predikcije ovisi o mnogo faktora, i nije uvijek 30 niti je uvijek potreban isti broj ispitanika)

Prije eksperimenta – u svrhu procjene veličine uzorka (DESKRIPTIVNA STATISTIKA)

- Deskriptivna statistika opisuje prirodu uzorkovanih podataka

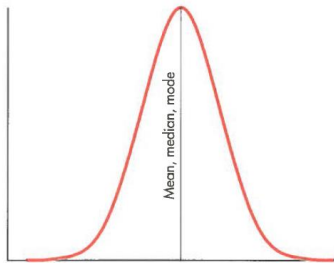
Glavni parametri deskriptivne statistike

- Procjena središnje tendencije (da li je opisuje srednja vrijednost ili median)
- Disperzija tj. varijacije mjerenja (standardna devijacija)
- Snaga efekta

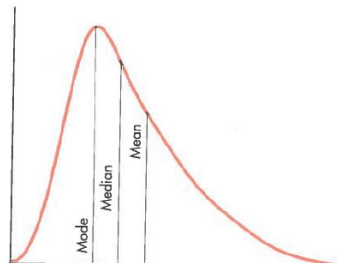
Distribucija podataka:

- Simetrični podaci
- Asimetrični podaci

Simetrična distribucija



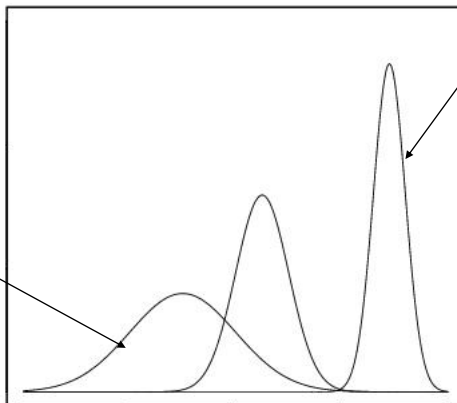
Asimetrična distribucija



Mjera disperzije podataka

Opisuju varijabilnost oko središnjeg dijela distribucije

Veliko rasap/visoka varijabilnost



Mali rasap/niska varijabilnost

- Standard deviation – variability around the mean

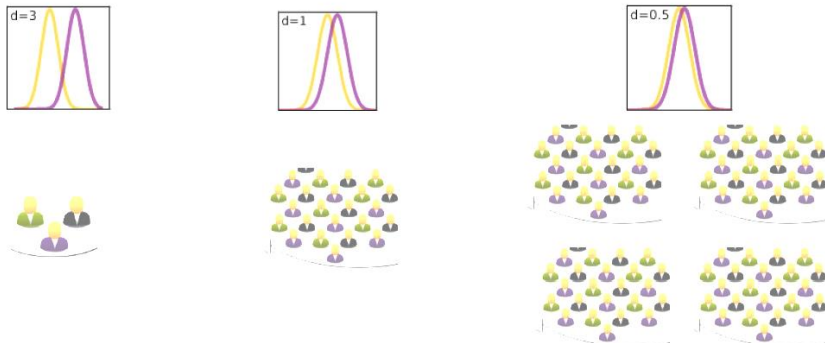
$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2}$$

- Variance = (standard deviation)²

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2$$

Snaga efekta koji nas zanima

- veličina efekta (veličina učinka)
- Koliki nam broj replika treba da bi mogli vidjeti efekt određene veličine (snage efekta, učinka)?
- Koliko velik/snažan efekt možemo vidjeti na uzorku određene veličine?



Pilot istraživanje

- Pilot istraživanje treba provesti prije prikupljanja svih podataka na manjem broju uzoraka, treba **procijeniti varijabilnost (disperziju izmjera) i snagu efekta te je povezati s brojem replika**
- Pilot istraživanje je korisno jer će poslužiti za provjeru **odabira statističkog testa kao i izračunavanje statističke snage**
- Snaga našeg testiranja je mjera da nećemo previdjeti razliku koja postoji

Odabir primjerenog broja replikata

- 1. **snaga efekta tj. veličina učinka** (npr. koliko će dodatak prehrani povećati masu pilića tj. razlika između srednje vrijednosti mase pilića hranjenih normalnom hranom i hranjenih isto + dodatak), **odražava biologiju sustava koji proučavamo, na to ne možemo utjecati**
- 2. **varijabilnost** - odražava biologiju sustava koji proučavamo, **na to ne možemo utjecati**
- 3. **buka u eksperimentu** (nastaje zbog svih ostalih čimbenika koji utječu na brzinu rasta mase pilića), **odražava biologiju sustava koji proučavamo, na to ne možemo utjecati (odnosno možemo djelomično) ali ne možemo potpuno kontrolirati**
- 4. dizajn pokusa – kako uzorkovati, dodatni kriteriji pri uzorkovanju npr. klasifikacija, grupiranje itd... (i odabrani statistički test)
- 5. broj neovisnih replika (veličina uzorka) – **potpuno u kontroli istraživača**

Primjer varijacija i replika

- **Utječe li primjena GM žitarica u ishrani na debljinu ljuske jajeta kokoši?**
- Problem varijabilnosti!
- Zadatak je vidjeti koliko je varijabilnost povezana s GM prehranom
- NULL hipoteza: debljina ljuske nije u vezi s GM prehranom

Suočavamo se s činjenicom da je debljina ljuske uvjetovana različitim faktorima od genetičkih, porijekla, starosti, količine vode, vlažnosti staništa, vremenom nošenje jaja. temperaturom i slicno. Vjerojatnost da ćemo dobiti istu vrijednost ako uzmemo dva jaja različitih kokoši hranjenih istom hranom je niska. Isto tako ista kokoš danas i sutra ne nese jaja iste debljine ljuske (varijabilnost). Stoga provodimo eksperiment na istoj liniji kokoši, isto starih, mjerimo jaja snesena u isto vrijeme itd. Još uvijek ćemo imati varijabilnost zbog koje moramo imati replikate, no djelomično ćemo stvoriti buku.

- Kontrola eksperimenta (identični uvjeti uz razliku u dodatku GM žitarica)
- Veći broj jaja

Ako imamo perfektnu Gaussovu distribuciju tada je 95% svih vrijednosti unutar dvije SD vrijednosti, a 68% unutar 1 SD

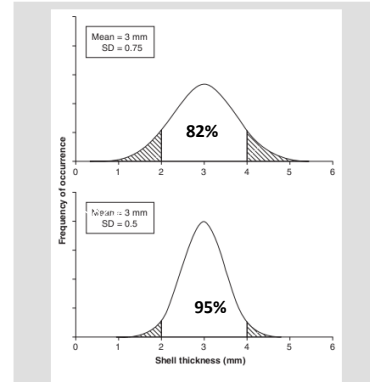
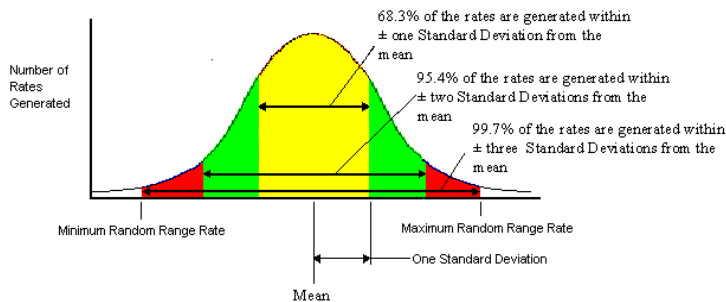


Figure 3.1 Two Gaussian or normal distributions to illustrate the kind of variation we are likely to see if we were to measure the shell thickness of many hundreds of thousands of eggs from hens fed a standard diet. Both distributions cluster around the same mean value (3 mm), but the distribution in the upper panel has a larger standard deviation, because the variation of the values around the mean value is greater: in the upper distribution, approximately 18% of the eggs have shells that are thicker than 4 mm or thinner than 2 mm, whilst in the lower distribution this is only 5%.

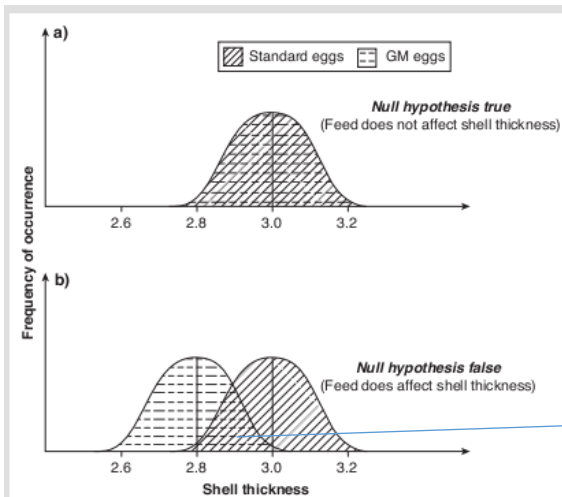


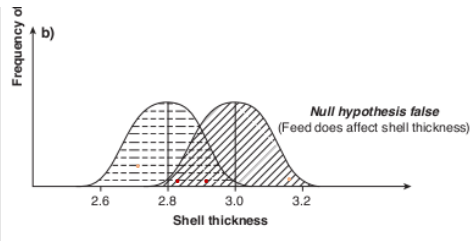
Figure 3.2 The distributions of egg shell thicknesses under two different scenarios. Under both scenarios shell thicknesses follow Gaussian distributions but in the top panel, the two populations have the same mean because the treatment is having no effect on shell thickness (i.e. the null hypothesis is true), whilst in the bottom panel the treatment is affecting shell thickness (i.e. the null hypothesis is false) and the means of the distributions differ.

GM hranjene:

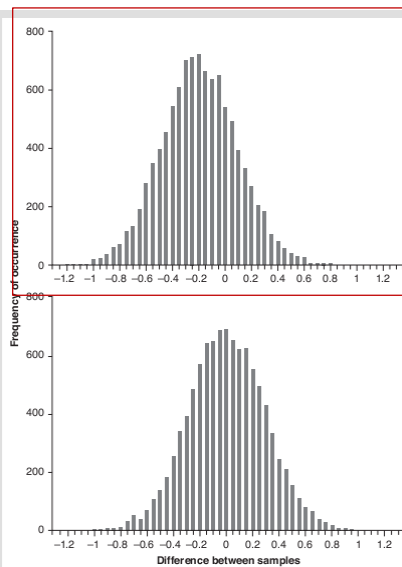
- ista distribucija varijabilnosti
- ako nema razlike – oko iste srednje vrijednosti
- ako ima razlike – oko razl. srednje vrijednosti

U preklapajućoj zoni su vrijednosti koje nam pokazuju da efekta nema

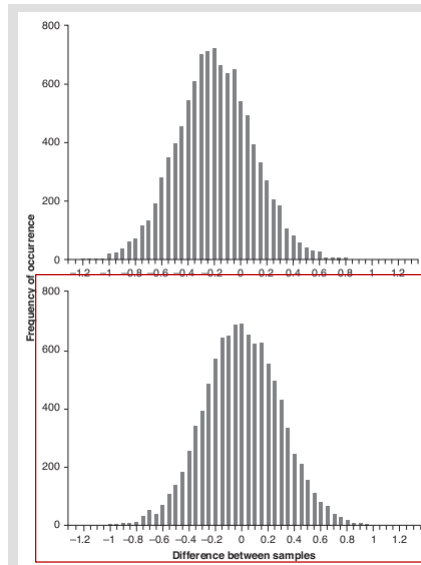
Kada bismo izmjerili sva jaja u populacijama ne bi nam trebala statistika međutim u svakom drugom slučaju potrebna je statistika da pokazuje značaj razlike od 0,2 mm



- Pojedinačna usporedba jednog GM i jednog normalnog jajeta – daje različite razlike
- Ako napravimo veliki broj (10000) nasumičnih usporedba i mjerimo razlike – dobijamo grafove (sljedeći slajd)

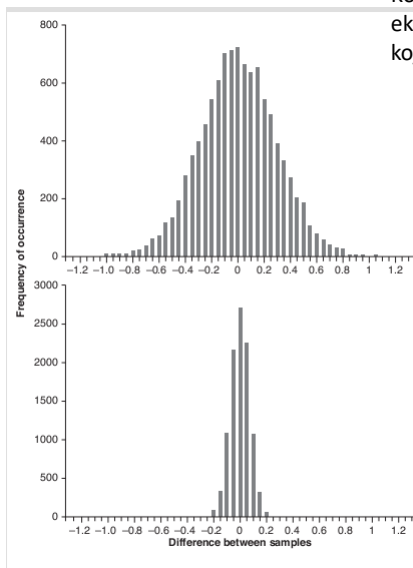


- Većina vrijednosti na gornjem grafu klasterira oko -0,2, no još uvijek ih ima puno koji pokazuju da GM hrana povećava debljinu ljuske (2427) dok 2351 eksperiment pokazuje da GM hrana smanjuje debljinu za više od 0,4mm .



- Većina vrijednosti na donjem grafu klasterira oko 0 ukazujući da nema razlike između GM ljsaka i neGM
- Unatoč tome puno pojedinačnih usporedbi moglo bi dati drugačije rezultate u bilo kojem smislu, npr razliku -0.32 i veću daje 1274

Odnos varijabilnosti klase/populacije i razlikovanja H_a i H_0



Količina varijacija u usporedbama koju vidimo u opisanim eksperimentima ovisi o tome koliko su varijabilne populacije koje uzorkujemo

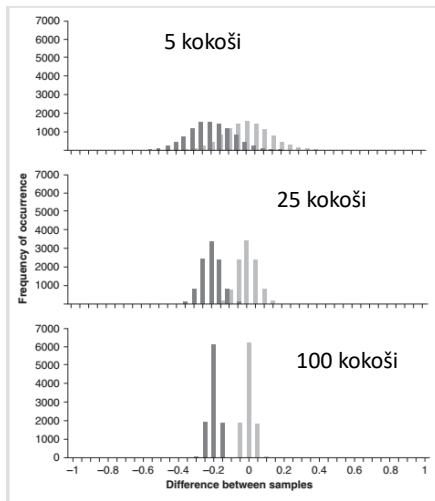
- Razliku od 0.32 možemo lako očekivati u mjerjenjima u gornjem eksperimentu (SD 0,2)
- Kada je SD 0,05 ova se razlika uopće ne može pojaviti

Od 10 000 imaginarnih eksperimenata, niti jedan nije predložio smanjenje debljina ljske u GM uzorcima za 0,32 mm ili više.

Tu ćemo lakše razlikovati H_0 i H_a a to je snaga statistike

No kako povećati snagu, budući da je SD karakteristika sustava kao i snaga efekta

BROJ REPLIKA



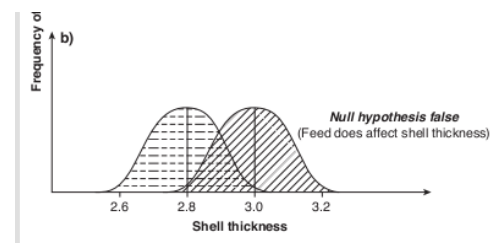
Snaga efekta procijenjena iz 10 000 imaginarnih eksperimenata u svijetu nul hipoteze (lijevo na svakom grafu) i alternativne hipoteze (desno), ali je veličina uzorka različita.

Kako veličina uzorka raste (broj replikata) diskriminacija dva svijeta je jasnija.

Razlikovanje između nulte i alternativne hipoteze – snaga statistike

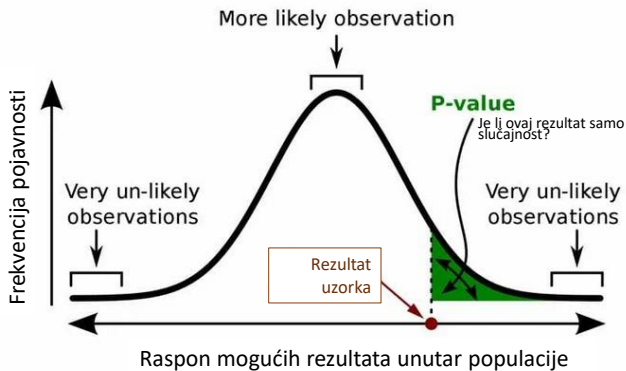
Zašto je kokoš u središtu pažnje, a ne jaje?

KONTINUIRANA VARIJABLA



- Slično pitanje: Kolika je vjerojatnost da se prisutnost GMO u hrani i izostanak GMO u hrani razlikuju kao na grafu **ako** su obje hrane jednako učinkovite?
- **Dopunjeno pitanje:** Kolika je vjerojatnost da se prisutnost GMO u hrani i izostanak GMO u hrani razlikuju kao na grafu **ili još više** ako su obje hrane jednako učinkovite?

p - vrijednost je vjerojatnost dobivenog rezultata (razlike) ili još ekstremnijeg uz pretpostavku da je nulta hipoteza točna



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

p-vrijednost je broj, izračunat iz statističkog testa koji se definira kao vjerojatnost da se dobije neki rezultat ili još ekstremniji ako je nulta hipoteza istinita.

P - vrijednost mjeri kolika je vjerojatnost da je uočena razlika između skupina rezultat slučajnosti.

Što je p vrijednost manja (slučajnost) to je dobiveni rezultat pouzdaniji.

p – VRIJEDNOST

Općenito, to je vjerojatnost da ćemo dobiti rezultate koje smo dobili ili još ekstremnije, ako fenomen koji istražujemo ne postoji (nulta hipoteza je točna).

Kritična vrijednost je točka na distribuciji vrijednosti za H_0 , koja opisuje skup vrijednosti koji zahtijevaju odbacivanje nulte hipoteze. **To je regija odbacivanja nulte hipoteze** (skup svih rezultata koji sugeriraju odbacivanje nulte hipoteze). **Taj se skup naziva kritična ili regija odbijanja** populacija

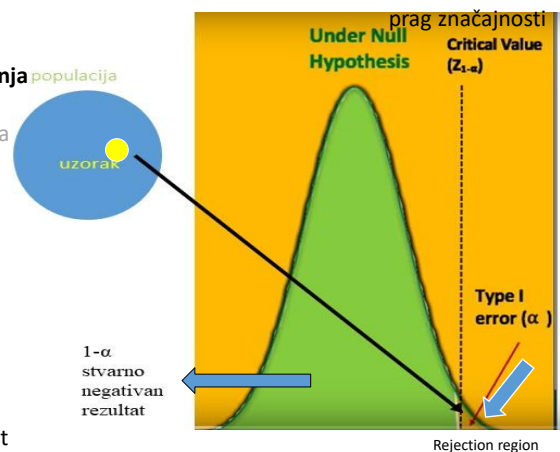
Kritične vrijednosti se određuju tako da je vjerojatnost da statistika testa ima vrijednost u području odbijanja testa kada je nulta hipoteza istinita jednaka razini značajnosti

Pogreška pri zaključivanju

Pogreška tipa I (α): ukazuje da učinak postoji, iako ga zapravo nema; stupanj ove pogreške određen je pragom značajnosti i prema njoj određujemo stat. značajnost.

Za rezultate s 95%-tnom razinom značajnosti : $\alpha = 0,05$ = vjerojatnost pogreške tipa I

Sami određujemo kritičnu vrijednost (prag značajnosti, signifikantnosti) ovisno o tome koliko smo kritični prema rezultatu i koliku razinu pogreške alfa možemo zanemariti (10%, 5%, 1%).



p – VRIJEDNOST

Statistička značajnost je usporedba između pogreška tipa I (α) i p vrijednosti proizašle iz stat. testa

p -vrijednost $\leq 0,05$: odbijamo H_0 odn. dobiveni rezultat je značajan

p -vrijednost $> 0,05$: nismo uspjeli odbiti H_0

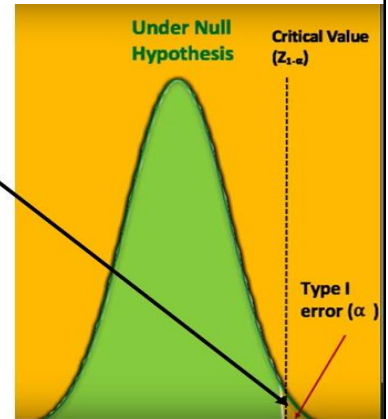
Prihvaćena kritična vrijednost u vezi je s razinom signifikantnosti (značajnosti) testa, a za nul-hipotezu se kaže da je *odbaćena na toj razini signifikantnosti*, ako je vrijednost p manja ili jednaka prihvaćenoj kritičnoj vrijednosti.

Za rezultate s 95%-tnom razinom pouzdanosti : $\alpha = 0,05$ = vjerojatnost pogreške tipa I

populacija

uzorak

$1 - \alpha$
stvarno
negativan
rezultat



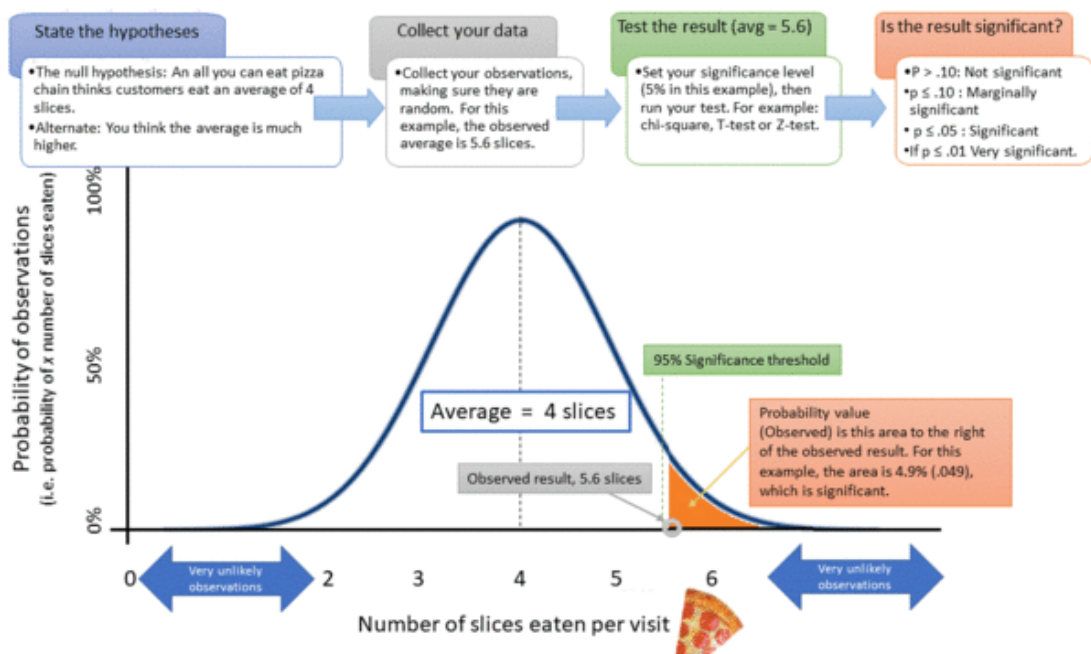
p -vrijednost, sažetak

- P -vrijednosti jedan je od načina testiranja je li rezultat eksperimenta statistički značajan.
- U biti, testira je li rezultat eksperimenta na podskupu populacije (uzorak) dovoljno snažan da donesemo zaključak o hipotezi.
- Budući da statistika nije egzaktna, o statističkim testovima možemo razmišljati kao o vrlo fino podešenom nagađanju.
- Kako je test statistike nagađanje, moramo znati koliko je naša hipoteza u skladu sa stvarnosti. To dovodi do pojma značajnosti.
- Statistička značajnost način je odgovora na pitanje utemeljenosti rezultata.

p-vrijednost, sažetak, KOJI REZULTAT SMATRAMO ZNAČAJNIM?

- Obično rezultat smatramo značajnim ako vrijedi $p < 0.05$
- I takav je rezultat objavljiv, no o samom problemu ovisi je li prihvatljiv.
- Sami odabiremo kriterij što smatramo značajnim (pitanje kriticnosti)
- Granicu značajnosti treba odabrati prije nego što počnemo s obradom podataka. Ona treba iskazivati naša uvjerenja o primjerenosti prema temi koju istražujemo.

- Shema za p-vrijednosti, na primjeru eksperimenta.



JE LI TO SVE ŠTO NAM TREBA?

Sjetimo se sigurnosnog sustava sa početka predavanja. Bio je opisan sa dva broja

- Vjerojatnošću da će krivo prepoznati poštene građane tj. vidjeti efekt iako ga nema (3%)
- Vjerojatnošću da će krivo prepoznati kriminalce (0%) odnosno ne vidjeti efekt koji postoji

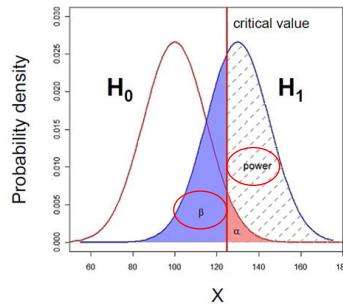
STATISTIČKA SNAGA

U odnosu na postavljenu hipotezu možemo napraviti:

- Pogrešku tipa I: test (slučajno) ukazuje da učinak postoji, iako ga zapravo nema (stupanj ove pogreške određen je pragom statističke značajnosti)
- Pogrešku tipa II: test ukazuje da učinak ne postoji, iako je on uistinu prisutan, povišenjem pogreške tipa I (α), umanjujemo stopu pogreška tipa II (koja je povezana sa STATISTIČKOM SNAGOM)

Stupanj pogreške tipa I je u potpunosti u kontroli istraživača. U prirodnim i biomedicinskim znanostima, ova stopa $p \leq 0.05$ se smatra prihvatljivom (kad bi ju još više snizili, povećala bi se vjerojatnost za činjenje pogreške tipa II, a nemoguće je istodobno smanjiti obje).

Što je snaga? Snaga i pogreška tipa 2 (β)



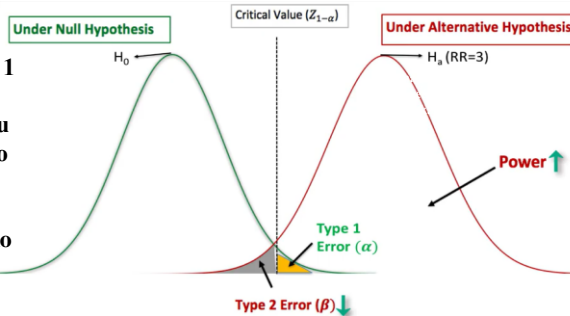
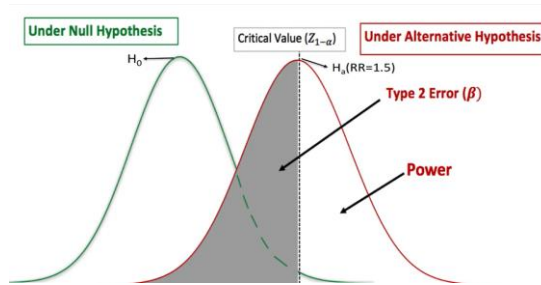
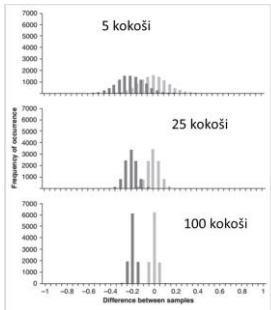
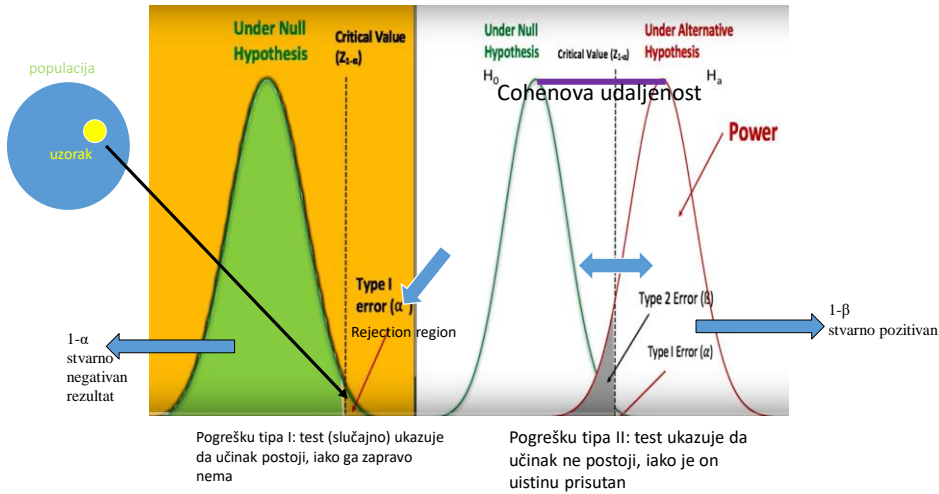
- Pogreška tipa II (β) je neuspjeh odbijanja H_0 koja nije točna
- Vjerojatnost da **nećemo vidjeti učinak koji stvarno postoji.**
- Snaga: **vjerojatnost otkrivanja učinka koji stvarno postoji**
- Izravni odnos između snage i pogreške tipa II:
- **Snaga = $1 - \beta$**

STATISTIČKA SNAGA kojom testiramo neku hipotezu je vjerojatnost da nećemo počinuti pogrešku tipa II (ne vidjeti učinak koji postoji)

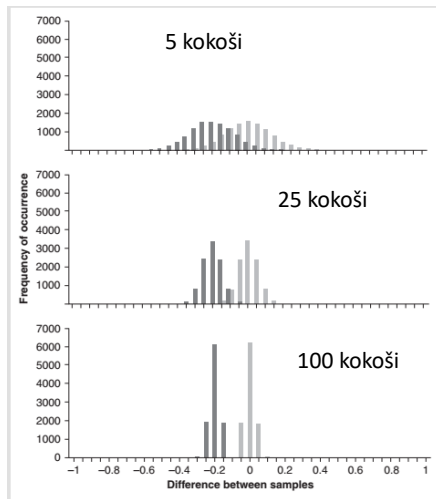
- Determinirana je
1. Pogreskom tipa 1 (alfa)
 - 2. Veličinom uzorka**
 3. Snagom efekta
 4. Varijabilnosti mjerenja

Što određuje snagu? (Veličina učinka)

Cohen's d is an effect size used to indicate the standardised difference between two means



Snaga testa će se povećati ako povećamo grešku tipa 1 (α) npr. ima li vitamin C kakvog učinka na prehladu - nećemo se toliko brinuti o lažno pozitivnima (tip 2 greška) jer znamo da je bezazlen, ali ako ispitujemo stetnost lijeka za trudnice na fetus pogreska tipa dva je opasna



Snaga efekta procijenjena iz 10 000 imaginarnih eksperimenata u svijetu nul hipoteze (lijevo na svakom grafu) i alternativne hipoteze (desno), ali je veličina uzorka različita.

Kako veličina uzorka raste (broj replikata) diskriminacija dva svijeta je jasnija.

Razlikovanje između nulte i alternativne hipoteze – snaga statistike

Kako “ručno” izračunati snagu planiranog istraživanja?

- **Utječe li GMO hrana na debljinu ljuske smanjujući je?**
- Mjerimo po pet jaja GMO i 5 standardnih 1. dan i uočavamo stanjenje ljuske u GMO uzorku
- Sljedećih 4 dana radimo isto i utvrđujemo isto, no peti dan rezultat je obrnut
- Nastavljamo mjeriti 100 dana i ako u 72 dana dobijemo rezultat smanjenja onda dokazujemo alternativnu hipotezu **snagom od 72%**

No to sigurno nećemo tako raditi.

Da imamo jaci efekt dobili bi veću snagu
Da imamo više jaja dobili bi veću snagu

Kako otkriti snagu planiranog istraživanja

Ključni čimbenici su:

- Veličina učinka
- Nasumična varijabilnost
- Dizajn i statistički test
- Broj replika

1. Da bi odredili snagu moramo pilot istraživanje odrediti **koji učinak imamo te definirati koji nam je učinak zadovoljavajući, odnosno bitan**

2. Nakon toga pilot istraživanjem moramo odrediti rasap vrijednosti

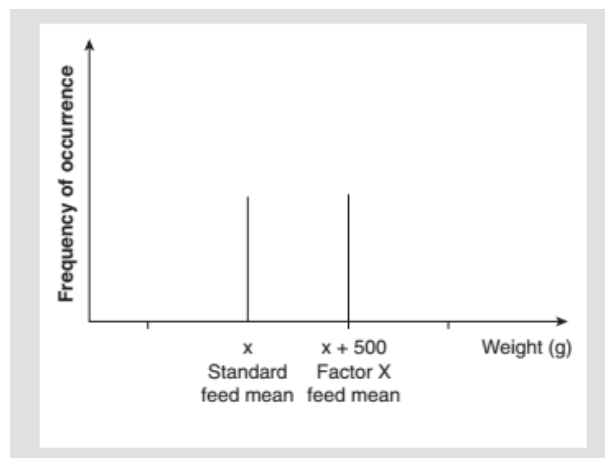
Kako otkriti snagu planiranog istraživanja

Uočite da ne mjerimo više debljinu ljuske nego masu pilića hranjenih sa dvije vrste hrane, no sve ostalo ostaje isto

Zanima nas je li obogaćena hrana zaista bolja

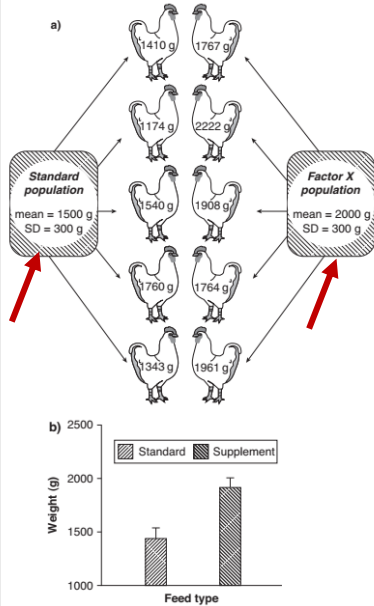
Npr. učinak koji smatramo relevantnim je za 500 g veća masa pilića hranjenih obogaćenom hranom u odnosu na standardnu hranu.

To nam je minimalni prirast koji nam čini značaj za novu vrstu prehrane!



DEFINIRALI SMO UČINAK KOJI ŽELIMO VIDJETI

PILOT procjena srednje vrijednosti i varijabilnosti



Kako uočiti takav učinak, tj. koliko replikata uključiti u eksperiment



KOJA JE SADA ŠANSA DA ĆEMO TAJ UČINAK UOČITI?

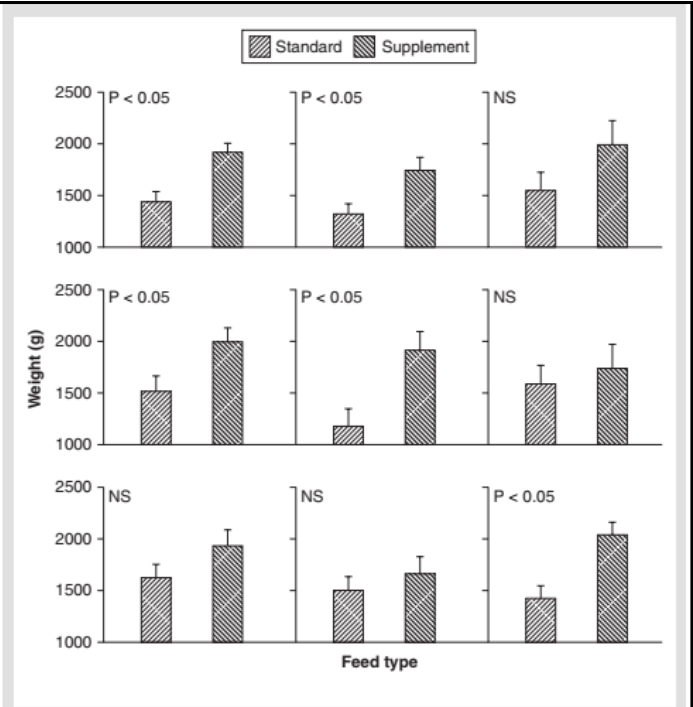
Procjena snage

ZAMIŠLJENI POKUS

1. Npr. 5 pilića standardno hranjenih kojima kompjuter dodjeljuje mase unutar 1500 ± 300 , te 5 pilića hranjenih obogaćenom hranom kojima kompjuter dodjeljuje mase unutar 2000 ± 300
2. Provede se odabran stat. test
3. Takvo dodjeljivanje ponovimo 9 puta

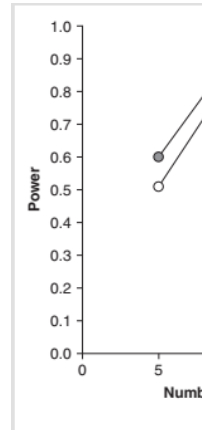
I rezultati pokazuju

U 4 pokusa s ovim učinkom razlika nije značajna



No kompjuter može izvršiti takav imaginarni pokus (umjesto 9) 1000 puta

- I tada u 507 pokusa imamo značajnost, što znači da nam je šansa uočavanja razlike sa 5 pilića 50 : 50



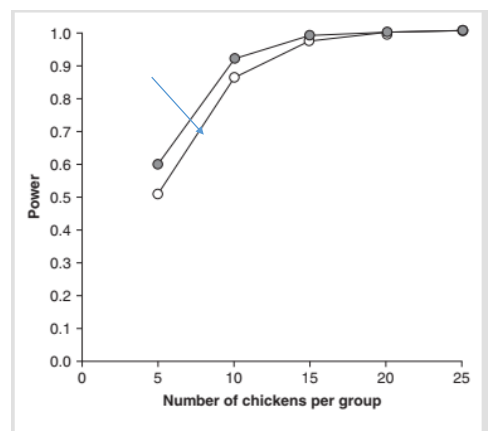
Promjena veličine uzorka i njen utjecaj na snagu

Ali ako povećamo broj replika

GRAF prikazuje predviđenu snagu pokusa za uzorke različite veličine (za istu snagu učinka). Ispitan je i efekt ugradnje klima, sivi kružići).

Povećanjem broja pilića snaga eksperimenta je raste, učinak je jasan!

Test nam ujedno i pokazuje da 10 do 15 pilića osigurava dovoljnu snagu i da nema potrebe ici na više!!!!



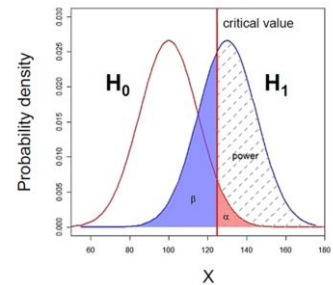
Kolika treba biti snaga? Snaga = 80%

- Pogreška tipa II (β) je neuspjeh odbijanja lažne H_0
- Vjerojatnost neuočavanja učinka koji stvarno postoji.
- **Power:** Vjerojatnost uočavanja učinka koji stvarno postoji.
 - Direktna je povezanost snage i pogreške tipa II :
 - Ako je **Snaga** = 0.8 tada je $\beta = 1 - \text{Snaga} = 0.2$ (20%)
- Stoga će stvarna razlika biti neuočena u 20% slučajeva
- **Opća konvencija za snagu je 80%**, ali može biti i više

Prema Cohen-u (1988):

- Za većinu istraživača: pogreške tipa I četiri su puta ozbiljnije od pogreške tipa II tako da je $0,05 \times 4 = 0,2$ (učinak koji ne postoji biti će uočen u 5% slučajeva)

Koji je tip pogreske stvarno opasniji?



kolokvij