Sadržaj

1 Pravila nasleđivanja **5**

2 Populaciona genetika **11**

3 Osnovi statistike **15**

4 Maksimizacija očekivanja **25**

5 Markovljevi lanci **33**

6 Algoritmi poravnanja **43**

7 Indeksiranje teksta i nizova karaktera **51**

8 Sklapanje sekvenci **55**

9 MinHash algoritam **69**

Glava 1

Pravila nasleđivanja

Zadatak 1.1 Alel za braon boju (B) kod slepih kučića je dominantan u odnosu na alel za belu boju (b). Recimo da imamo jednog mužjaka slepog kučeta braon boje.

a) Koji su mogući genotipi na alelu koji određuje boju ove jedinke?

b) Ukoliko imamo ženku bele boje da li i kako možemo da utvrdimo genotip braon mužjaka?

c) Pretpostavimo da je genotip mužjaka Bb. Ako posmatramo potomstvo ove dve jedinke, koji odnos genotipa i fenotipa bi mogli da očekujemo?

Rešenje

a) Mogući aleli su BB i Bb.

b) Možemo ukrstiti mužjaka i ženku (koja mora biti bb).

c) Ukoliko je mužjak BB sve jedinke u F1 generaciji će biti braon boje. Ukoliko je mužjak Bb očekujemo 50% belih i 50% braon jedinki.

Zadatak 1.2 Šta očekujete da bi se dobilo ukoliko bi ukrstili biljku zevalicu koja je homozigotno ("čistokrvno", tj "true breeding") crvena (RR) sa homozigotnom belom biljkom?

Rešenje

Sve jedinke u F1 generaciji će biti ružičasto. U pitanju je primer nepotpune dominacije, gde oba alela utiču na fenotip. U nekim slučajevima oba alela doprinose jednako (jedan alel proizvodi crveni, a drugi beli pigment, pa bi rezultat bio ružičast). U drugim slučajevima jedan od alela je nefunkcionalan, tj. označava "pokvaren"gen, a fenotip zahteva dve funkcionalne kopije (biljka proizvodi samo pola potrebne količine crvenog pigmenta, pa je rezultat ružičast).

Zadatak 1.3

a) Pomoću Punitove tabele [3] odrediti moguće genotipe u F1 generaciji

b) Nacrtati Punitovu tabelu za sve kombinacije iz F2 generacije

c) Odrediti procente različitih genotipa u F2 generaciji.

Rešenje

a)

	A	a
A	*AA*	*Aa*
a	*Aa*	*aa*

Mogući genotipi: AA (25%), aa (25%) i Aa (50%).

b)

	AA	Aa	aA	aa
AA	*AAAA*	*AAa*	*AAa*	*Aaa*
Aa	*AAa*	*AaAa*	*Aaa*	*aAaa*
aA	*AAa*	*Aaa*	*aAaA*	*aaAa*
aa	*Aaa*	*aAaa*	*aaAa*	*aaaa*

c)

Genotip	Broj	Procenat
aAbB	4	25
AAbB	2	12.5
aABB	2	12.5
aAbb	2	12.5
aabB	2	12.5
AABB	1	6.3
AAbb	1	6.3
aaBB	1	6.3
aabb	1	6.3

Zadatak 1.4 Dobili smo priliku da izučavamo genetiku drekavca. Saznali smo da je prisustvo kandži (KK ili Kk) dominantno u odnosu na odsustvo kandži (kk). Takođe, znamo da je prisustvo rogova (RR ili Rr) dominantno u odnosu na odsustvo rogova (rr).

Ukrstili smo mužjaka koji ima kandže i rogove sa ženkom koja ima kandže, ali nema rogove. Svih 18 jedinki u F1 generaciji ima rogove, 14 ima kandže, dok 4 nema kandže. Odrediti genotipe roditelja.

Rešenje

- Otac ima kandže i rogove, pa sledi da ima bar po jedan K i R alel, dok za druge alele nismo sigurni.
- Majka ima bar jedan K alel, i dva r alela (nema rogove).
- Kako svih 18 F1 jedniki ima rogove, otac mora biti RR (pa svi nasleđuju R od njega, te imaju Rr genontip).

- Pošto su neke F1 jedinke kk, znači da su neke i od oca i od majke nasledili k alel, pa su roditelji imaju Kk genotip. Ukoliko bi barem jedan roditelj imao KK genotip, onda bi svi potomci imali kandže.

Zadatak 1.5 Posmatramo zeca koji je heterozigot na hipotetička dva alela koji nisu povezani (tj. nisu na istom hromozomu). Jedan alel utiče na veličinu ušiju (AA ili Aa za veliki, odnosno aa za male uši), a drugi za veličinu očiju (BB i Bb za velike, a bb za sitne oči). Ako bi mogli da ukrstimo ovog zeca sa potpuno recesivnom individuom, koji bi bili rezultujući genotipi i fenotipi?

Rešenje Zec može imati veliki ili male uši i velike ili male oči (mora imati nešto od toga).

Heterozigot na dva alela, znači da će jedan alel odgovarati ušima, a drugi očima. Dakle, formiramo par alela koji odgovara uši-oči.

	ab	**ab**	**ab**	**ab**
AB	AaBb	AaBb	AaBb	AaBb
Ab	Aabb	Aabb	Aabb	Aabb
aB	aaBb	aaBb	aaBb	aaBb
ab	aabb	aabb	aabb	aabb

AaBb = 25%, Aabb = 25%, aaBb = 25%, aaBb = 25%,

Ukrštanje sa potpuno recesivnom jedinkom ("test-kros") služi za otkrivanje skrivenih recesivnih alela.

Ukrštanje sa heterozigotnom individuom rezultuje u svim mogućim kombinacijama u odnosu 1:1:1:1.

Zadatak 1.6 Posmatramo ponovo problem iz zadatka 1.5. Zamislite da smo za 4 moguća tipa F1 potomstava umesto odnosa 1:1:1:1 dobili odnose iz sledeće tabele. Na šta bi ukazivao ovakav rezultat?

Project	Genotip	Fenotip
48%	AaBb	Velike uši, velike oči
2%	Aabb	Velike uši, male oči
2%	aaBb	Male uši, velike oči
48%	aabb	Male uši, male oči

.

Rešenje

To bi značilo da su geni povezani (nalaze se na istom hromozomu). Retke kombinacije su posledica rekombinacije (kombinacija vezanih gena).

Zadatak 1.7 U AB0 [9] sistemu krvnih grupa, aleli A i B su ko-dominantni, a oba su dominantna nad alelom 0. Ukoliko je majka AB krvna grupa, a dete A krvna grupa, da li je moguće da otac bude (diskutovati mogućnosti):

a) A krvna grupa?

b) B krvna grupa?

c) AB krvna grupa?

d) 0 krvna grupa?

Rešenje

a) Moguće je, ako je otac AA ili A0, a dete AA ili A0.

b) Moguće je, ako je otac B0, a dete A0.

c) Moguće je, ako je dete AA.

d) Moguće je. ako je dete A0.

Zadatak 1.8 Alel za plavu boju očiju je recesivan (ako ovaj alel označimo sa b, da bi osoba imala plave oči neophodno je da njen genotip bude bb). Posmatrajmo osobu sa braon očima čiji otac ima plave oči.

a) Koji je genotip posmatrane osobe?

b) Da li dete posmatrane osobe može imati plave oči?

c) Da li dete dvoje ljudi koji imaju braon oči može imati plave oči? Kako?

Rešenje

a) Genotip osobe je Bb. Alel b mora bit nasleđen od oca, pošto otac ima plave oči, odnosno genotip bb. Alel B mora biti nasleđen od majke, da bi osoba imala braon oči.

b) Da.

c) Da, ako je genotip oba roditelja Bb.

Zadatak 1.9 Ukoliko bi za dati par gde oboje imaju braon oči utvrdili da sigurno postoji mogućnost da njihovo dete ima plave oči, koja bi bila verovatnoća da se to i dogodi? Poznato je sledeće:

- Muž i žena imaju braon oči.
- Muževljev otac ima plave oči.
- Ženin brat ima plave oči.

Izračunati verovatnoću da njihovo dete ima plave oči.

Rešenje

Muževljev otac ima plave oči, dakle muž je Bb, 50% šanse da dete nasledi plavi alel. Ženin brat ima plave oči, dakle oba ženina roditelja imaju po jedan placi alel; žena je BB, Bb ili bB, odnosno 66% šanse da je nasledila jedan plavi alel.

Verovatnoće da dete nasledi plavi alel od majke je onda 0.660.5=0.33. Ukupno, verovatnoća da je dete bb je 0.330.5 = 0.165*

Zadatak 1.10 Tigraste mačke imaju gen na X hromozomu čija dva alela kontrolišu boju:

- AA je crna ženka
- A'A' je žuta ženka
- AA' je tigrasta ženka
- A je crni mužjak
- A' je žuti mužjak.

Posmatramo crnu mačku koja je omacila 4 tigraste ženke i 2 crna mužjaka. Da li je otac mačor žute ili crne boje?

Rešenje Majka je svakako AA, dakle mačići su od oca nasledili A'. Kako otac ima samo jednu kopiju X hromozoma, ona mora nositi A' alel, te je otac žut.

Zanimljivo: Šare kod tigrastih mačaka verovatno nastaju usled procesa *inaktivacije X hromozoma*. Za razliku od mužjaka, ženke poseduju dve kopije X hromozoma, ali se u ćelijama nasumičnim odabirom jedna od dve kopije "gasi"(ceo hromozom se pakuje u gusto zamotan heterohromatin). Neke ćelije "ugase"kopiju koja nosi žuti alel, a neke onu koja nosi crni alel.

Zadatak 1.11 Česta forma daltonizma kod koje osoba ne razlikuje crvenu i zelenu boje nastaje usled recesivnog alela na X-hromozomu.

a) Da li dete čija su oba roditelja daltonisti može da ne bude daltonista?

b) Da li žena može da bude daltonista ukoliko nijedan od njenih roditelja nije daltonista?

c) Da li muškarac može da bude daltonista ukoliko nijedan od njegovih roditelja nije daltonista?

Rešenje

a) Ne, svi aleli oba roditelja su recesivni, prema tome dete ne može da nasledi alel koji kodira dominantni (ne-daltonistički) fenotip.

b) Ne, ukoliko otac nije daltonista, žena će sigurno naslediti ne-daltonistički alel.

c) Da, ukoliko je majka heterozigot.

Glava 2

Populaciona genetika

Zadatak 2.1 Frekvencije dva alela u populaciji su 0.19 (A) i 0.81 (a). Pretpostavite da je populacija u Hardi-Vajnberg ekvilibrijumu [6].

a) Izračunati procenat heterozigotnih jedinki u populaciji.

b) Izračunati procenat homozigotnih recesivnih jedinki u populaciji.

Rešenje

a) Prema HWE: $f(Aa) = 2pq$, gde su *p* i *a* frekvencije alela. Dakle: $f(Aa) = 2 \cdot 0.19 \cdot 0.81 = 0.31$ ili 31%

b) Prema HWE: $f(aa) = q^2$ Dakle: $f(aa) = 0.81^2 = 0.66$ ili 0.66%

Zadatak 2.2 Alel za belu vunu (V) je dominantan u odnosu na alel za crnu vunu (v). U uzorku od 900 ovaca 891 je belo, a 9 je crno. Izračunati frekvencije alela u ovoj populaciji, pretpostavljajući da je populacija u Hardi-Vajnberg ekvilibrijumu.

Rešenje

Neka su frekvencije pojavljivanja alela date as: $f(V) = p,\ f(v) = q$.
Frekvencija recesivnih homozigota (crnih ovaca) je prema tome data sa q^2.
$q^2 = \frac{9}{900} = 0,01;\ q = 0.1\ p + q = 1;\ p = 1 - 0.1 = 0.9$

Zadatak 2.3 U populaciji koja je u Hardi-Vajnberg ekvilibrijumu frekvencija recesivnih homozigota neke osobine je 0.09. Izračunati procenat jedinki koje su homozigoti za dominantni alel.

Rešenje

$q^2 = 0.09; q = 0.3\ p = 1 - q = 0.7$ Prema tome, HWE procenat homozigota za dominantni alel je $p^2 = 0.49$ ili 49%

Zadatak 2.4 U populaciji koja je u Hardi-Vajnberg ekvilibrijumu 38% jedinki je homozigot za recesivni alel neke osobine. U populaciji od 14,500 jedinki odrediti broj jedinki koje su homozigoti za dominantni alel, kao i broj heterozigota.

Rešenje

$q = \sqrt{0.38} = 0.616; p = 1 - q = 0.384$

$homozigoti \approx 14500 * p * p = 2138$

$heterozigoti \approx 14500 * 2 * p * q = 3430$

Zadatak 2.5 Alel (T) za sposobnost da se oseti ukus određene hemikalije je dominantan u odnosu na alel t (za 'nesposobnost' da se oseti ukus date hemikalije). Četiri stotine studenata je učestvovalo u studiji u kojoj se za 64 njih pokazalo da ne mogu da osete ukus date hemikalije. Izračunati procenat heterozigota u studiji. Pretpostavite da je populacija u Hardi-Vajnberg ekvilibrijumu.

Rešenje

$N_{tt} = 64$

$q^2 = \frac{64}{400} = 0.16;\ q = 0.4$

$p = 1 - q = 0.6$

Procenat heterozigota: $2pq = 0.48$ ili 48%

Zadatak 2.6 Rh faktor se kod ljudi nasleđuje od roditelja nezavisno od AB0 alela za krvne grupe. Ljudi koji su Rh+ imaju Rh antigen na površini crvenih krvnih zrnaca, dok ga Rh- osobe nemaju. Postoje dva alela za Rh faktor: dominantni Rh alel rezultuje u Rh+ fenotipu, dok recesivni rh alel rezultuje u Rh- fenotipu.

Posmatramo populaciju koja je u Hardi-Vajnberg ekvilibrijumu. 160 od 200 individua ima Rh+ fenotip. Izarčunati frekvencije oba alela.

Rešenje

$N_{Rh-} = 200 - 160 = 40$

$q^2 = \frac{40}{200} = 0.2$

$q = \sqrt{0.2} = 0.45;\ p = 1 - 0.45 = 0.55$

Zadatak 2.7 Kod kukuruza boja zrna je pod kontrolom gena čiji dominantni alel (B) određuje belu boju, a recesivni alel (b) žutu. U uzorku od 100 zrna iz populacije koja je u Hardi-Vajnberg ekvilibrijumu 9 zrna je žuto, a 91 bele boje.

a) Izračunati frekvencije belih i žutih alela u ovoj populaciji

b) Izračunati procenat populacije koja je heterozigotna.

Rešenje

a) $q^2 = \frac{9}{100} = 0.09$

Žuti: $q = 0.3$, beli: $p = 0.7$

b) $f(Bb) = 2pq = 0.42$

Zadatak 2.8 Retka bolest koju izaziva recesivni alel (a) kada je homozigotan se pojavljuje u populaciji sa frekvencijom jedan u milion. Koliko nosilaca alela (a) možemo da očekujemo u gradu veličine 14 000? Pretpostaviti Hardi-Vajnberg ekvilibrijum.

Rešenje

Frekvencija pojavljivanja bolesti je $q^2 = 10^{-6}$,

Dakle, broj frekvancija alela koji izaziva bolest je: $q = 10^{-3}$, dok je frekvencija drugog ('zdravog') alela: $p = 0.999$.

Prema tome, broj nosilaca je: $N_{Aa} + N_{aa}$ $N_{Aa} = 14000 \cdot 2pq \approx 28$; $N_{aa} = 14000 \cdot q^2 = 14000 \cdot 10^-6 \approx 0$ $N_{Aa} + N_{aa} \approx 28$

Glava 3

Osnovi statistike

Zadatak 3.1 Identifikujte svaki od sledećih skupova podataka kao populaciju ili uzorak.

a) Prosek svih studenata na fakultetu.

b) Prosek nasumično odabranog grupe studenata na fakultetu.

c) Starosti ministara Vlade Republike Srbije 1. januara 2020 godine.

d) Pol svakog drugog kupca koji ulazi u bioskop.

e) Dužina somova uhvaćenih na ribolovnom izletu.

Rešenje

a) Populacija.

b) Uzorak.

c) Populacija.

d) Uzorak.

e) Populacija.

Zadatak 3.2 Istraživač želi da proceni prosečnu potrošnju po osobi od posetilaca zabavnog parka. On uzima nasumični uzorak od četrdeset posetilaca i dobija prosek od 2000 dinara po osobi.

a) Definisati populaciju koja je predmet istraživanja.

b) Definisati parametar koji je predmet istraživanja.

c) Na osnovu ovog uzorka, da li znamo prosečan iznos koji posetilac parka potroši? Objasniti detaljno.

Zadatak 3.3 Istraživač želi da proceni procenat svih odraslih koji imaju mobilni telefon. On uzima nasumični uzorak od 1.572 odraslih; 1.298 od njih ima mobilni telefon, dakle $1298/1572 = 0.83$ ili oko 83% odraslih ima mobilni telefon.

a) Definisati populaciju koja je predmet istraživanja.

b) Definisati parametar koji je predmet istraživanja.

c) Definisati statistiku koju je istraživač koristio.

d) Na osnovu ovog uzorka, da li možemo odrediti procenat svih odraslih koji imaju mobilni telefon? Objasnite detaljno.

Zadatak 3.4 Sociolog želi da proceni procentualni udeo svih odraslih osoba u određenom regionu koje nikada nisu sklopile brak. U nasumičnom uzorku od 1.320 osoba, 145 nikada nisu sklopile brak, dakle $145/1320 = 0.11$ ili oko 11%.

a) Definisati populaciju koja je predmet istraživanja.

b) Definisati parametar koji je predmet istraživanja.

c) Definisati statistiku koju je istraživač koristio.

d) Na osnovu ovog uzorka, da li možemo tačno odrediti udeo svih odraslih osoba koje nikada nisu sklopile brak? Objasnite detaljno.

Rešenje

a) Svi odrasli u regiji.

b) Procenat odraslih u regiji koji nikada nisu bili u braku.

c) Procenat izračunat iz uzorka.

d) Ne možemo tačno, ali znamo približnu vrednost odnosa.

Zadatak 3.5 Kutija sadrži 10 belih i 10 crnih kugli.

a) Definisati skup mogućih uzoraka koje možemo dobiti prostor nasumičnim odabiranjem dve kugle sa zamenom.

b) Odrediti verovanoće sledećih rezultata

- Uzeta je barem jedna kugla svake boje.
- Nije uzeta nijedna bela kugla.

Rešenje

a) S= {*cc*, *cb*, *bc*, *bb*}, pri čemu *c* označava da je izvučena crna kugla, dok *b* označava da je izvučena bela kugla.

b)
- $P(\text{cb or bc}) = P(\text{cb}) + P(\text{bc}) = 0.5 \cdot 0.5 + 0.5 \cdot 0.5 = 0.5$
- $P(\text{cc}) = P(\text{c}) \cdot P(\text{c}) = 0.25$

Zadatak 3.6 Kutija sadrži 5 crvenih, 8 žutih i 8 zelenih kugli.

a) Definisati skup mogućih uzoraka koje možemo dobiti prostor nasumičnim odabiranjem dve kugle sa zamenom.

b) Odrediti verovanoće sledećih rezultata

- Uzete su dve kugle iste boje.
- Nije uzeta nijedna žuta kugla.

Rešenje

a) S= $\{cc, cz, c\check{z}, zc, zz, z\check{z}, \check{z}c, \check{z}z, \check{z}\check{z}\}$, pri čemu c označava da je izvučena crvena kugla, z označava da je izvučena zelena kugla, a $\check{z}$ označava da je izvučena žuta kugla-

b)
- P(cc or zz or žž) = P(cc) + P(zz) + P(žž)

  $P(cc) = \frac{5}{5+8+8} \cdot \frac{5}{5+8+8} = 0.057$

  $P(zz) = \frac{8}{5+8+8} \cdot \frac{8}{5+8+8} = 0.145$

  $P(\check{z}\check{z}) = \frac{5}{5+8+8} \cdot \frac{5}{5+8+8} = 0.145$

  P(cc or zz or žž) = 0.347
- P(nijedna žuta) = $\frac{13}{5+8+8} \cdot + \frac{13}{5+8+8} = 0.383$

Zadatak 3.7 Neka je data tabela daje raspodelu stanovništva u određenom mestu prema starosti i posedovanju vozačke dozvole

Starost	Ima dozvolu	Nema dozvolu
Ispod 30	6000	17000
Preko 30	36000	41000

Ako nasumično odaberemo jednu osobu, odrediti verovatoću sledećih događaja:

a) Osoba ima vozačku dozvolu.

b) Osoba ima manje od 30 godina.

c) Osoba ima vozačku dozvolu i manje od 30 godina.

Rešenje

a) $P = \frac{6000+36000}{6000+36000+17000+41000} = 0.42$

b) $P = \frac{6000+17000}{6000+36000+17000+41000} = 0.23$

c) $P = \frac{6000}{6000+36000+17000+41000} = 0.06$

Zadatak 3.8 Data je tabela sa raspodelom procenta stanovništva u određenom mestu prema obrazovanju i konzumaciji cigareta.

Obrazovanje	Pušač	Nije pušač
Osnovna škola	0.16	0.18
Srednja škola	0.2	0.31
Fakultet	0.06	0.09

Ako nasumično odaberemo jednu osobu, odrediti verovatoću sledećih događaja:

a) Osoba je pušač.

b) Osoba ima završenu bar srednju školu.

c) Osoba je pušač i ima završenu samo osnovnu školu.

d) Osoba je pušač ako ima završenu samo osnovnu školu.

e) Osoba nije pušač ako ima završen fakultet.

Rešenje

a) $P = \frac{0.16+0.2+0.09}{0.16+0.18+0.2+0.31+0.09+0.06} = 0.45$

b) $P = \frac{0.2+0.31+0.06+0.09}{0.16+0.18+0.2+0.31+0.09+0.06} = 0.66$

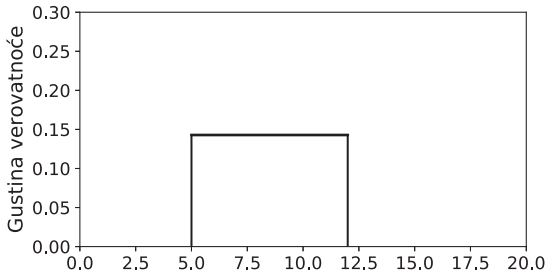
c) $P = \frac{0.16}{0.16+0.18+0.2+0.31+0.09+0.06} = 0.16$

d) $P = \frac{0.16}{0.16+0.18} = 0.471$

e) $P = \frac{0.09}{0.09+0.06} = 0.6$

Zadatak 3.8 Neka je data kontinualna slučajna varijabla sa uniformnom raspodelom na intervalu [5,12]. Nacrtati grafik funkcije gustine verovatnoće.

Rešenje



Zadatak 3.9 Neka je data kontinualna slučajna varijabla sa normalnom raspodelom na intervalu srednje vrednosti 42 i standadne devijacije 4. Nacrtati grafik funkcije gustine verovatnoće.

Rešenje