

Problem #1

V datoteki **basketball_shots.csv** so podatki o metih na koš. Vsak košarkaš je 60 krat vrgel na koš običajne velikosti, nato pa še 60 krat na koš z manjšim obsegom.

Zanima nas:

1) Primerjava med uspešnostjo košarkaša #1 ter košarkaša #2:

- *Kdo je boljši?*
- *Kako prepričani smo v to ugotovitev?*
- *Kolikšna je razlika v uspešnosti?*

2) Primerjava med metanjem na običajni ter manjši obroč za košarkaša #1.

Namigi:

Za modeliranje uspešnosti uporabi Bernoullijevo porazdelitev.

Predznanje v model lahko vnesemo s pomočjo Beta porazdelitve.

44

Naš prvi model – meti na koš

45

Naš prvi model – meti na koš

- **data**

n – število metov
y – rezultat meta (0 – neuspešen met, 1 – uspešen)

```
data {
  int<lower=1> n;
  int y[n];
}
```

- **parameters**

theta – parameter Bernoulli distribucije, ki ocenjuje uspešnost

```
parameters {
  real<lower=0, upper=1> theta;
}
```

- **model**

predznanje (prior)
opis modela

```
model {
  // prior
  theta ~ beta(1,1);

  for (i in 1:n) {
    y[i] ~ bernoulli(theta);
  }
}
```

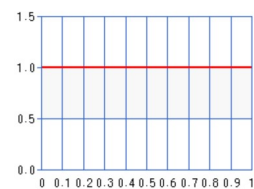
46

Prior za parameter theta

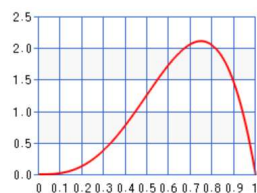
- **brez**

"flat" porazdelitev na intervalu $[-\infty, \infty]$

- **beta(1, 1)**



- **beta(4, 2)**



47

Problem #2

V datoteki **temperature.csv** so podatki o temperaturi, za Slovenijo imamo podatke o povprečni temperaturi za vsak mesec med 1901 in 2015.

Zanima nas, ali je bila temperatura julija (najbolj vroč mesec) med 1970 in 1985 nižja kot med 2000 in 2015?

- *Kako prepričani smo v to trditev?*
- *Za koliko je bila temperatura nižja?*

Namig:

Za modeliranje uspešnosti uporabi normalno porazdelitev – $N(\mu, \sigma)$.

48

Problem #3

V datoteki **temperature.csv** so podatki o temperaturi. Za Slovenijo imamo podatke o povprečni temperaturi za vsak mesec med 1901 in 2015, za Finsko pa med 2000 in 2015.

Zanima nas:

1) Ali julijska temperatura skozi čas na Finskem narašča?

- *Kakšno je naše zaupanje v napovedi modela, zakaj?*
- *Kako lahko ta problem rešimo?*

2) Ali julijska temperatura skozi čas v Sloveniji narašča?

3) Kakšna bo temperatura leta 2019, kakšna 2070? Kakšna je verjetnost, da bo leta 2019, oziroma 2070, pričakovana temperatura višja od 25°C?

Namig:

Za modeliranje uspešnosti uporabi normalno linearno regresijo – normalni linearni model, kjer je μ porazdelitve linearno odvisno ($a + bx$) od leta (x).

49

Problem #4

Novo nastalo zagonsko podjetje nas je najelo, da jim pomagamo pri dveh pomembnih odločitvah:
(ob predpostavki, da želijo maksimizirati svoj dobiček)

- 1) Kam vlagati sredstva (razvoj, marketing ali administracija) ?
- 2) Kje naj imajo svoje prostore (na voljo imajo Florido, Kalifornijo in New York)?

S pomočjo analiziranja podatkov iz datoteke **50_startups.csv** jim pomagaj pri odločitvi. V datoteki se nahajajo podatki o 50 zagonskih podjetjih (kako so vlagali denar, kje imajo pisarno ter koliko so zaslužili).

Namig:

Nadgradi linearno normalno regresijo iz prejšnjega problema.

Odvisna spremenljivka (profit) je odvisna od več atributov (research, marketing, ...) – vhod X je torej matrika.

Za vsak atribut želimo svojo beta (b) vrednost – parameter b je torej vektor.

Iz kategorične spremenljivke (state) naredimo več binarnih spremenljivk.