Applied Bayesian Statistics

Andrea Cognolato, July 2021
s281940@studenti.polito.it

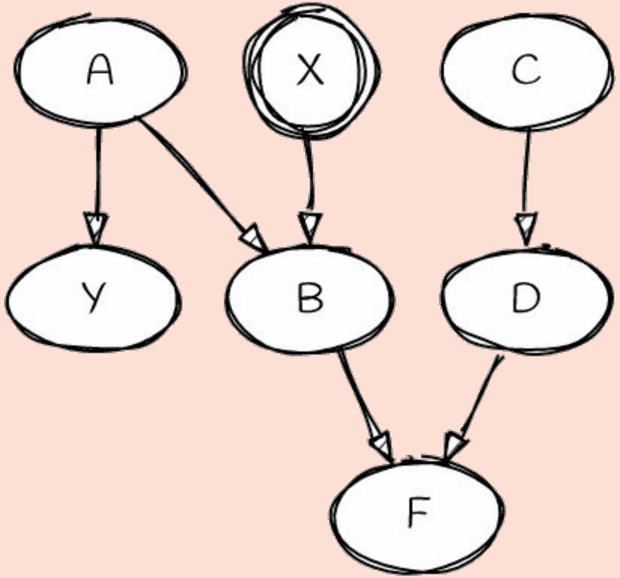
Mathematics in Machine Learning theoretical tesina

Probabilistic Programming

- Modeling uncertainty
- Flexibility
- Leverages Bayesian statistics

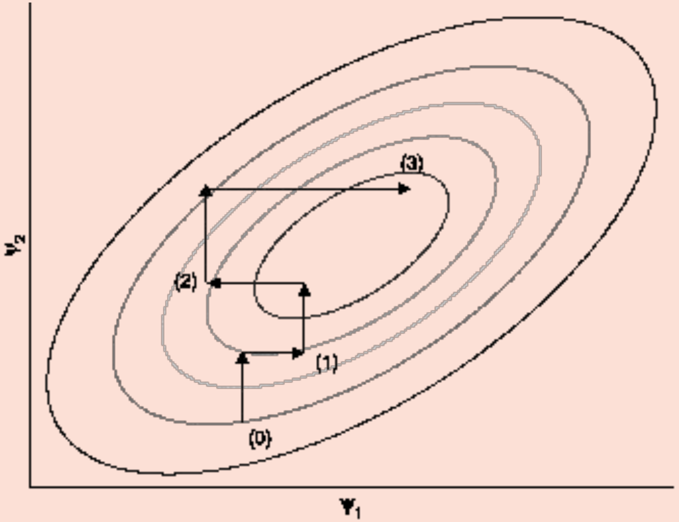
Directed Acyclic Graphs

- Representation of joint PDFs
- Conditional independence structure
- Belief Propagation, Conjugacy Detection, ...



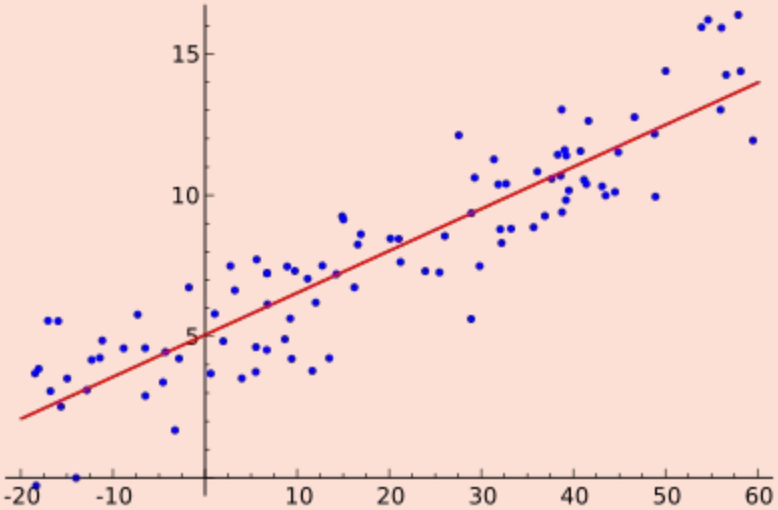
Gibbs sampling & BUGS

- Sampling with conditionals
- Axis-aligned moves
- WinBUGS/JAGS/OpenBUGS/...



Linear Regression

- The four assumptions
- Overfitting
- Outliers

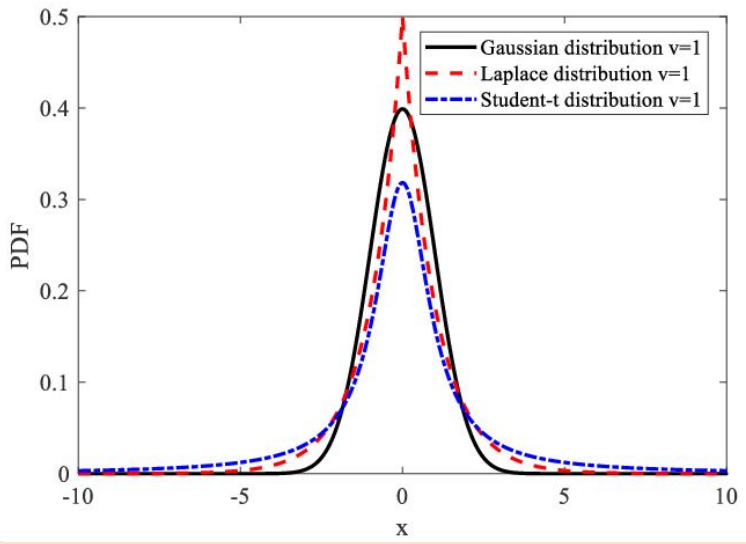


Regularization as Bayesian Occam's Razor

- Occam's Razor
- Prior knowledge
- Regularization

Alternative error structures

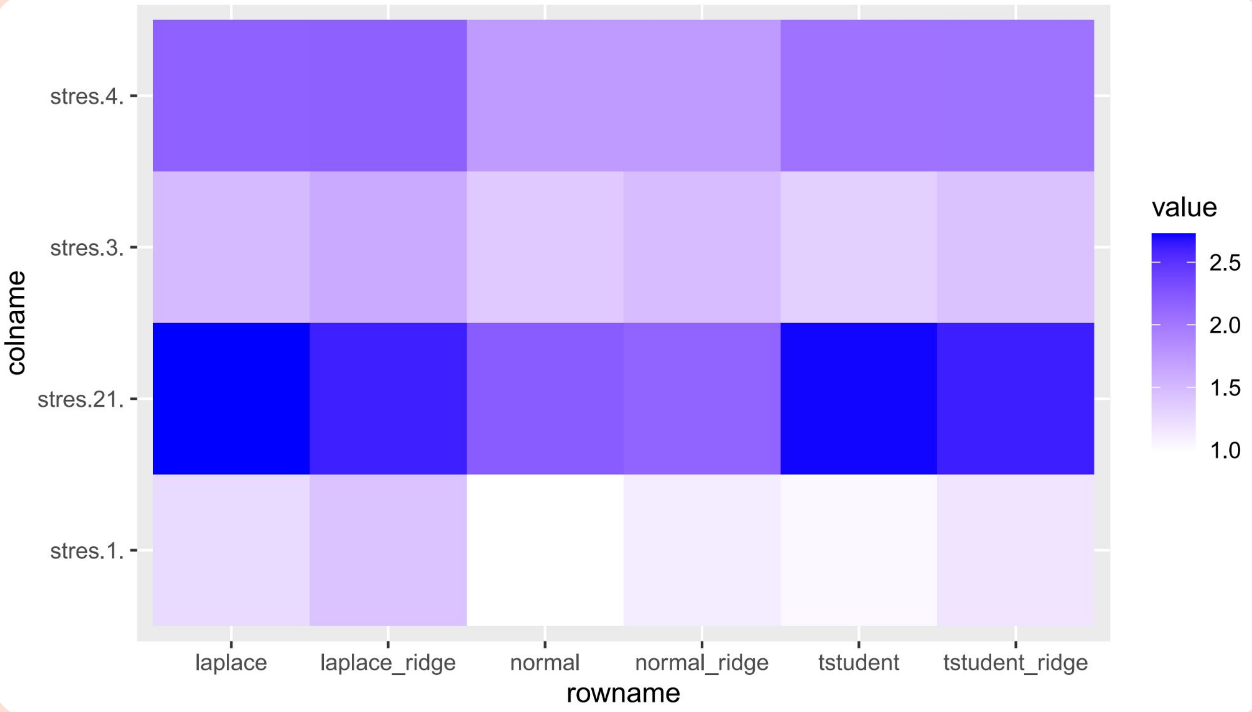
- Outliers
- Laplace
- t-student



Model

```
for (i in 1 : N) {
    Y[i] ~ dnorm(mu[i], tau)
    # Y[i] ~ ddexp(mu[i], tau)
    # Y[i] ~ dt(mu[i], tau, d)
    mu[i] ← beta0 + beta[1]*z[i, 1] +
                    beta[2]*z[i, 2] +
                    beta[3]*z[i, 3]
}
beta0 ~ dnorm(0, 0.00001)
for (j in 1 : p) {
    beta[j] ~ dnorm(0, 0.00001)
    # beta[j] ~ dnorm(0, phi)
}
tau ~ dgamma(1.0E-3, 1.0E-3)
```

Outlier analysis

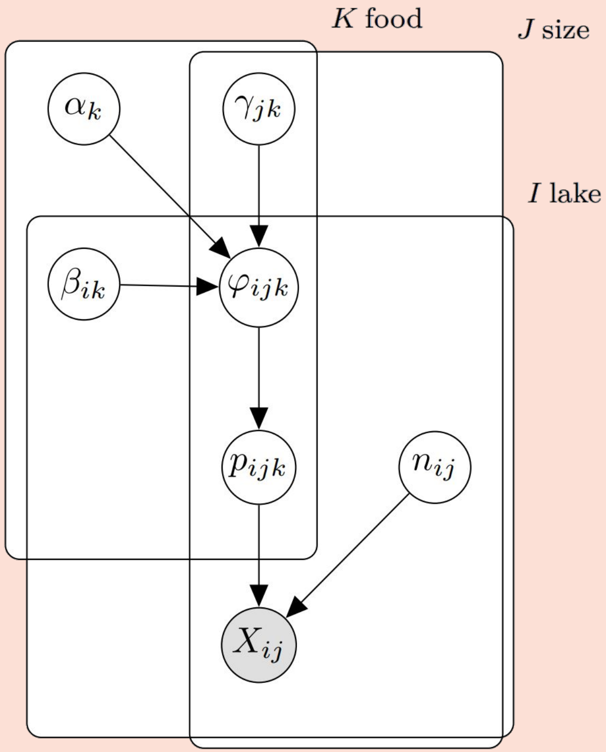


Classification

Distribution	#classes=2	#classes>2
#trials=1	Bernoulli	Categorical
#trials>1	Binomial	**Multinomial**

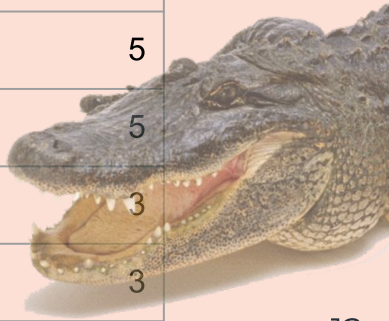
Hierarchical Modeling

- Multiple datasets
- Parameter sharing
- More modeling flexibility



Alligators

		Primary Food Choice				
Lake	**Size**	**Fish**	**Invertebrate**	**Reptile**	**Bird**	**Other**
Hancock	<=2.3	23	4	2	2	8
	>2.3	7	0	1	3	5
Oklawaha	<=2.3	5	11	1	0	3
	>2.3	13	8	6	1	0
Trafford	<=2.3	5	11	2	1	5
	>2.3	8	7	6	3	5
George	<=2.3	16	19	1	2	3
	>2.3	17	1	0	1	3



Model

$$\alpha_k \sim \mathcal{N}(0, 10^{-5})$$

$$\beta_{ik} \sim \mathcal{N}(0, 10^{-5})$$

$$\gamma_{jk} \sim \mathcal{N}(0, 10^{-5})$$

$$\varphi_{ijk} = \sigma_k(\alpha_k + \beta_{ik} + \gamma_{jk})$$

$$X_{i,j} \sim \text{Multi}(\varphi_{ijk}, n_{ij})$$

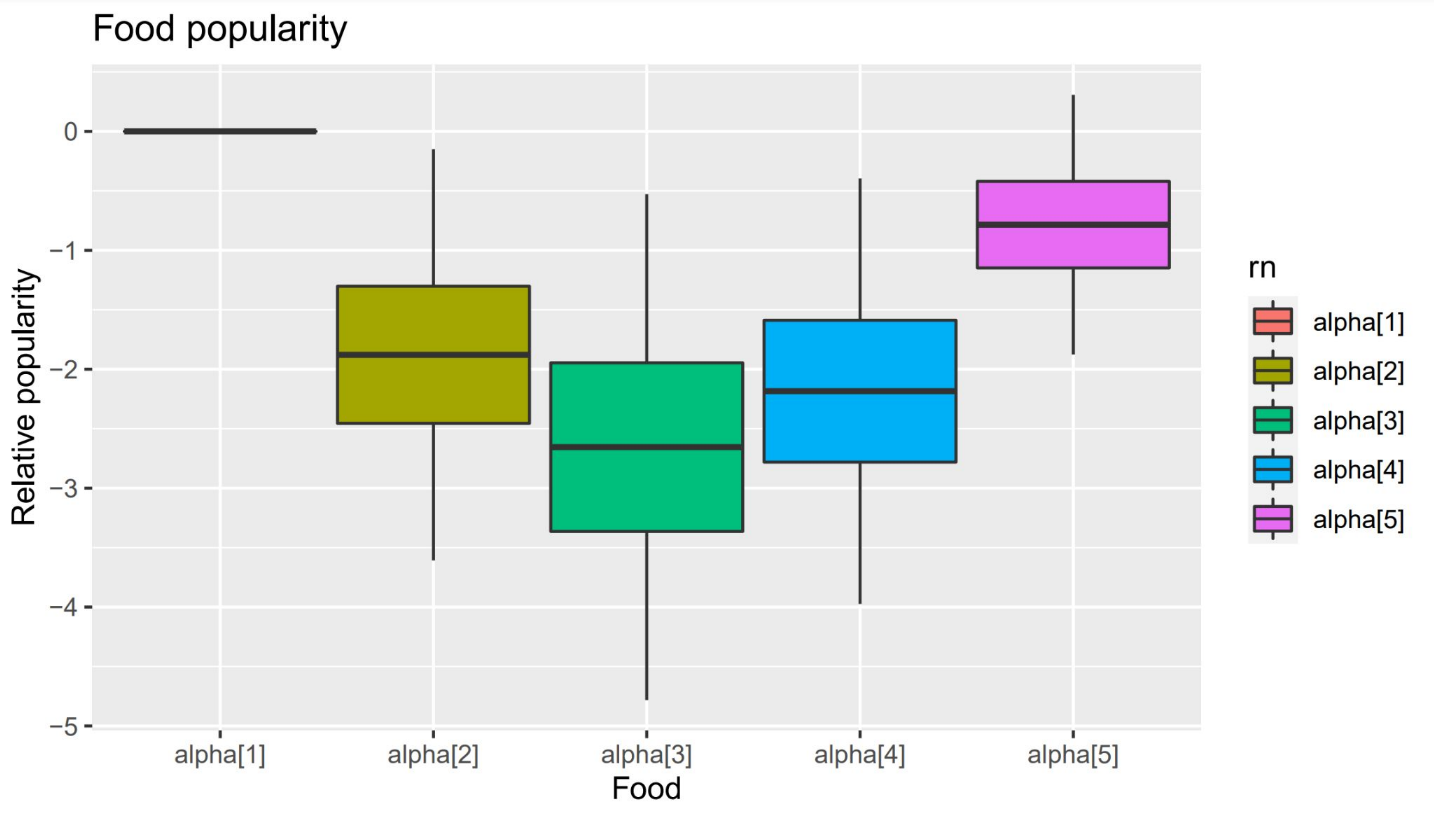
```
for (k in 1:K){
    alpha[k] ~ dnorm(0,0.00001);
}
for (i in 1:I) {
    for (k in 1:K) {
        beta[i,k] ~ dnorm(0,0.00001);
    }
}
for (j in 1:J) {
    for (k in 1:K){
        gamma[j,k] ~ dnorm(0,0.00001);
    }
}

for (i in 1:I) {
    for (j in 1:J) {
        for (k in 1:K) {
        p[i,j,k]          ← phi[i,j,k] / sum(phi[i,j,]);
        log(phi[i,j,k]) ← alpha[k] + beta[i,k] + gamma[j,k];
        }
        X[i,j,] ~ dmulti(p[i,j,] , n[i,j]);
    }
}
```

Goodness-of-fit statistic

- G-statistic, linked to G-test
- Likelihood ratio for multinomial RVs
- KL Divergence

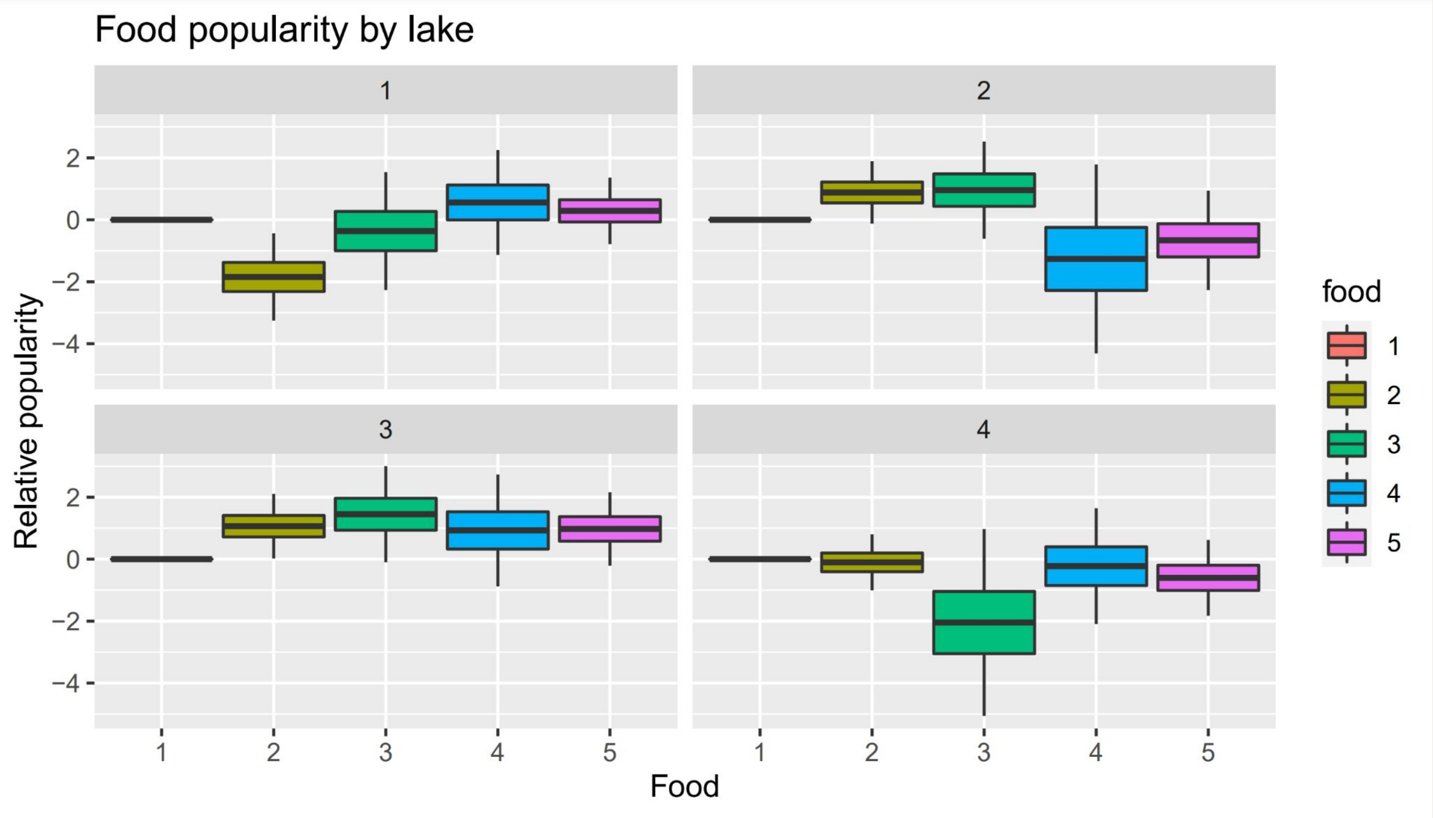
```
for (i in 1:I) {     # loop around lakes
    for (j in 1:J) {     # loop around sizes
       for (k in 1:K) {     # loop around foods
          E[i,j,k] <- p[i,j,k] * n[i,j];
          OlogOE[i,j,k] <- X[i,j,k] * log(X[i,j,k] / E[i,j,k]);
     }
  }
}
G2 <- 2 * sum(OlogOE[,,]);
```

Results

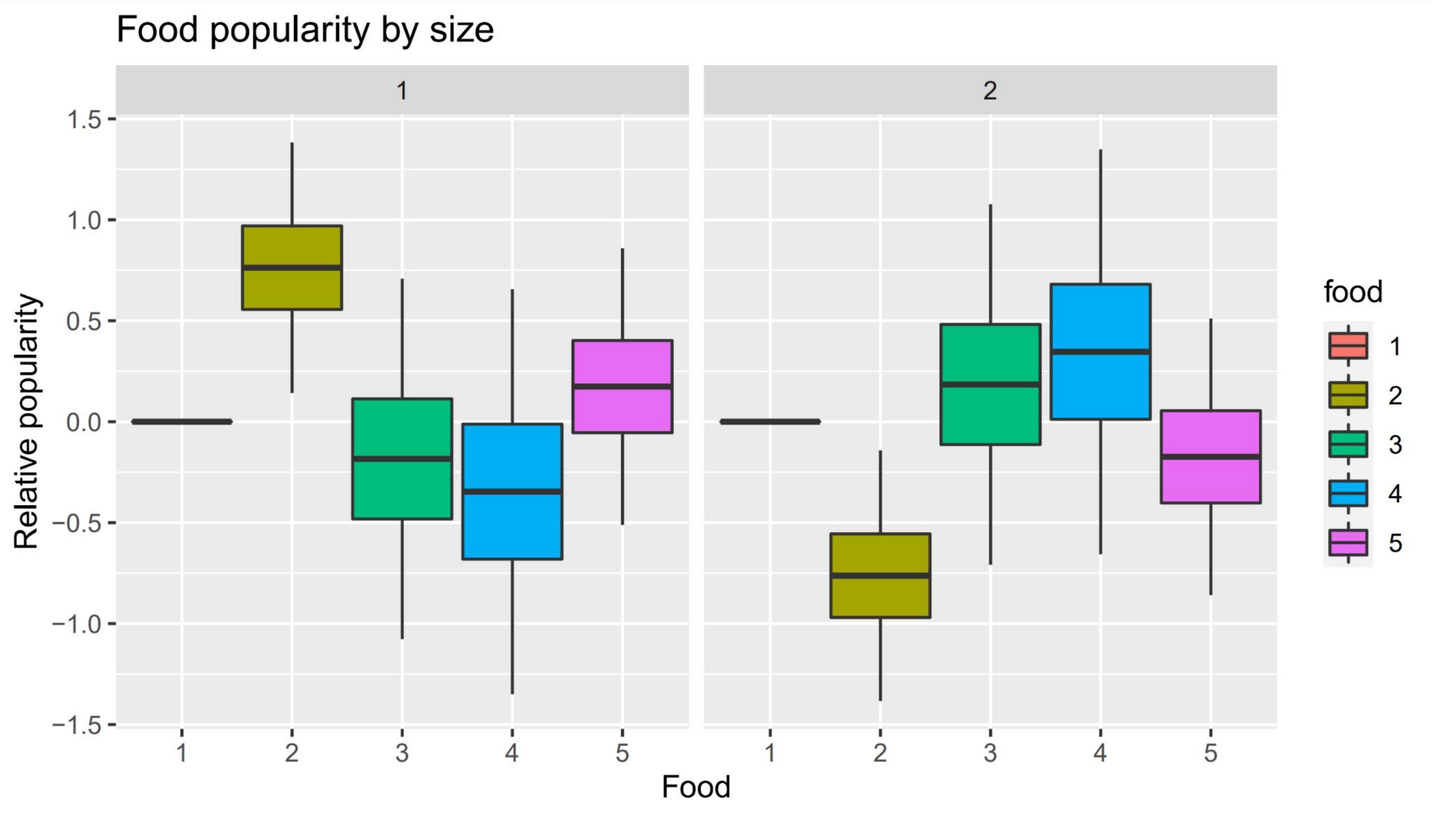


Food popularity
0
−1
−2
−3
−4
−5
Relative popularity
alpha[1]
alpha[2]
alpha[3]
alpha[4]
alpha[5]
Food
rn
alpha[1]
alpha[2]
alpha[3]
alpha[4]
alpha[5]

Results


Food popularity by lake
1
2
3
4
Relative popularity
2
0
−2
−4
Food
1
2
3
4
5
food
1
2
3
4
5

Results



Food popularity by size
1
2
Relative popularity
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
1
2
3
4
5
Food
food
1
2
3
4
5