

**ZADANIA**

1. Pliki `gryfy.txt` oraz `gryfygenealogia.txt` zawierają przykładowe dane dotyczące gryfa – tajemniczego zwierzęcia żyjącego gdzieś u krańców Śródziemia. Pliki te mają następującą strukturę:

```
> head(gryfy) # identyfikator, matka, rok urodz, plec, waga, skok
  ANIMAL MOTHER BYEAR SEX   BWT TARSUS
1   1029   1145   968   1 10.77  24.77
2   1299    811   968   1  9.30  22.46
3    643    642   970   2  3.98  12.89
4   1183   1186   970   1  5.39  20.47
5   1238   1237   970   2 12.12    NA
6    891    895   970   1    NA    NA

> head(gryfyped) # genealogia: identyfikator, ojciec, matka
  ID FATHER MOTHER
1 1306     NA     NA
2 1304     NA     NA
3 1298     NA     NA
4 1293     NA     NA
5 1290     NA     NA
6 1288     NA     NA
```

Sprawdź ile osobników nie zostało zmierzonych pod kątem długości skoku.

*Wskazówka 1:* zmienna NA nie jest ciągiem tekstowym ale wbudowanym typem zmiennej R, zapisujemy ją więc bez cudzysłowu.

*Wskazówka 2:* większość obiektów w R ma wbudowaną atrybut zawierający ilość elementów macierzy lub wektora. Aby dostać się do tego elementu można zastosować funkcję `length()`. Sprawdź co ta funkcja potrafi i czy możesz jej tutaj użyć.

2. Wykorzystując kilkadziesiąt złowionych osobników gryfów chciałbyś/chciałabyś założyć komercyjną hodowlę na potrzeby przemysłu baśniopisarskiego. Jest jednak jeden mały problem: dla osiągnięcia odpowiedniego efektu gryfy powinny być odpowiednio duże. Możesz oczywiście przeprowadzić w swojej hodowli odpowiednią selekcję pod kątem długości skoku – czy jednak zawsze da ona pożądaną efekt? Aby o tym się przekonać – musisz sprawdzić na ile dziedziczną cechą jest długość skoku.
  - a. Skonstruuj model za pomocą którego oszacujesz jak duża część zmienności w długości skoku ma komponentę genetyczną.
    - i. Stwórz odpowiedni prior (powinien zawierać efekt genetyczny oraz związany z wariancją resztkową).
    - ii. Uwzględnij w modelu dymorfizm płciowy wielkości ciała gryfów.
    - iii. Używając MCMCglmm oblicz odziedziczalność tej cechy, sprawdź też czy wariancja genetyczna jest istotnie większa od zera.
  - b. Czy model uruchomiony z domyślnymi ustawieniami daje poprawne wyniki?
 

*Wskazówka:* sprawdź czy łańcuch się dobrze miesza szukając oznak autokorelacji.

- c. W okresie jednego z lat z których pochodzą pomiary odnotowano wyraźnie mniejszą podaż złotych rybek – podstawowego pokarmu gryfów. Jak skorygujesz swój model by uwzględnić to dodatkowe źródło zmienności?
  - d. Częstym problemem grafologii ilościowej jest nierozróżnialność efektu genetycznego i efektu matki na pewnym poziomie. Sprawdź, czy efekty matczyne mają istotny wpływ na długość skoku młodych gryfów. Jak to zrobisz? Co się wtedy stanie z wariancją genetyczną? Co to oznacza dla analiz nie uwzględniających efektu matki?
3. Gryfy nie są najłatwiejszymi w hodowli zwierzętami. Mierzenie ich skoku nie jest najbezpieczniejszym zadaniem. Na szczęście matka natura pomyślała za nas i niektóre cechy silnie skorelowała genetycznie z innymi. W takich cechach dobór działający na jedną z nich – będzie powodować odpowiednią zmianę w drugiej cesze. Ponieważ gryfa znacznie łatwiej zważyć niż zmierzyć mu skok – postanawiasz dokonywać selekcji rozmiarów ciała pod kątem wagi urodzeniowej. Czy słusznie?
  - a. Zbuduj model (opierając się na doświadczeniach z poprzedniego ćwiczenia uwzględnij w nim wszystkie konieczne czynniki) za pomocą którego zbadasz istnienie korelacji genetycznej między długością skoku i masą urodzeniową.
  - b. Jaką strukturę kowariancji zastosujesz dla wariancji resztkowej?  
*Wskazówka:* model uruchom z następującymi parametrami: `nitt = 200000`, `burnin = 40000`, `thin = 150`; unikniesz żmudnej diagnostyki modeli.
  - c. Oblicz korelację genetyczną oraz jej przedział ufności. Jakie wnioski wyciągniesz z takiej wartości  $r_g$ ?
  - d. W modelu mamy dodatkowy efekt „naśladujący” efekty genetyczne – efekt matczyny. Jak obliczysz korelację między badanymi cechami na poziomie kowariancji wynikającej z działania efektów matczynych? Jaką ta korelacja ma wartość? Co to oznacza?
  - e. Gdyby zmierzenie skoku lub zważenie gryfa zabijało go i powodowało natychmiastową dezintegrację zwłok (zupełnie jak zastosowanie najbardziej morderczej funkcji fazyzerów w Star Trek) – moglibyśmy mieć dane tylko dot. długości skoku albo tylko masy. Co by to zmieniło w specyfikacji naszego modelu?
4. ZADANIE DODATKOWE. Pewnym specyficznym typem kowariancji genetycznej jest interakcja genotyp-środowisko, której specjalnym rodzajem jest interakcja genotypu z płcią. Jak przekształcisz posiadane dane aby możliwe było zastosowanie ich do oszacowania takiej interakcji?  
*Wskazówka:* Aby uwidocznąć taką kowariancję należy zbudować model w którym wartość danej cechy (weźmy długość skoku) w danej płci to jedna cecha, zaś jej wartość w drugiej płci –to druga cecha. Należy więc stworzyć dwie dodatkowe kolumny, np MTARS i FTARS zawierające odpowiednie wielkości cechy dla właściwej płci, oraz NA dla cechy specyficznej dla płci przeciwnej, np (1 = samiec, 2 = samica):

	ANIMAL	MOTHER	BYEAR	SEX	BWT	TARSUS	MTARS	FTARS
1	1029	1145	968	1	10.77	24.77	<b>24.77</b>	<b>NA</b>
2	1299	811	968	1	9.30	22.46	<b>22.46</b>	<b>NA</b>
3	643	642	970	2	3.98	12.89	<b>NA</b>	<b>12.89</b>
4	1183	1186	970	1	5.39	20.47	<b>20.47</b>	<b>NA</b>
5	1238	1237	970	2	12.12	NA	<b>NA</b>	<b>NA</b>
6	891	895	970	1	NA	NA	<b>NA</b>	<b>NA</b>

Możesz tu wykorzystać funkcję `cbind()` która pobiera kolumny i łączy je wg wierszy.  
Np.:

```
> head(cbind(gryfy, kolumna1 = „puste”, kolumna2 = c(1,2,3,4)))
```

	ANIMAL	MOTHER	BYEAR	SEX	BWT	TARSUS	kolumna1	kolumna2
1	1029	1145	968	1	10.77	24.77	puste	1
2	1299	811	968	1	9.30	22.46	puste	2
3	643	642	970	2	3.98	12.89	puste	3
4	1183	1186	970	1	5.39	20.47	puste	4
5	1238	1237	970	2	12.12	NA	puste	1
6	891	895	970	1	NA	NA	puste	2

Jak łatwo i bezboleśnie osiągnąć zamierzony efekt i wygenerować potrzebny zestaw danych bez potrzeby żmudnego kopiowania i wklejania wartości poszczególnych cech?

Jeśli chcielibyście wiedzieć jak wejść w posiadanie danych na temat gryfów (lub jednorożców, smoków i tym podobnych) – polecam artykuł:

Clément et al. 2001. Simulation analysis to test the influence of model adequacy and data structure on the estimation of genetic parameters for traits with direct and maternal effects. *Genet. Sel. Evol.* 33: 369-95.