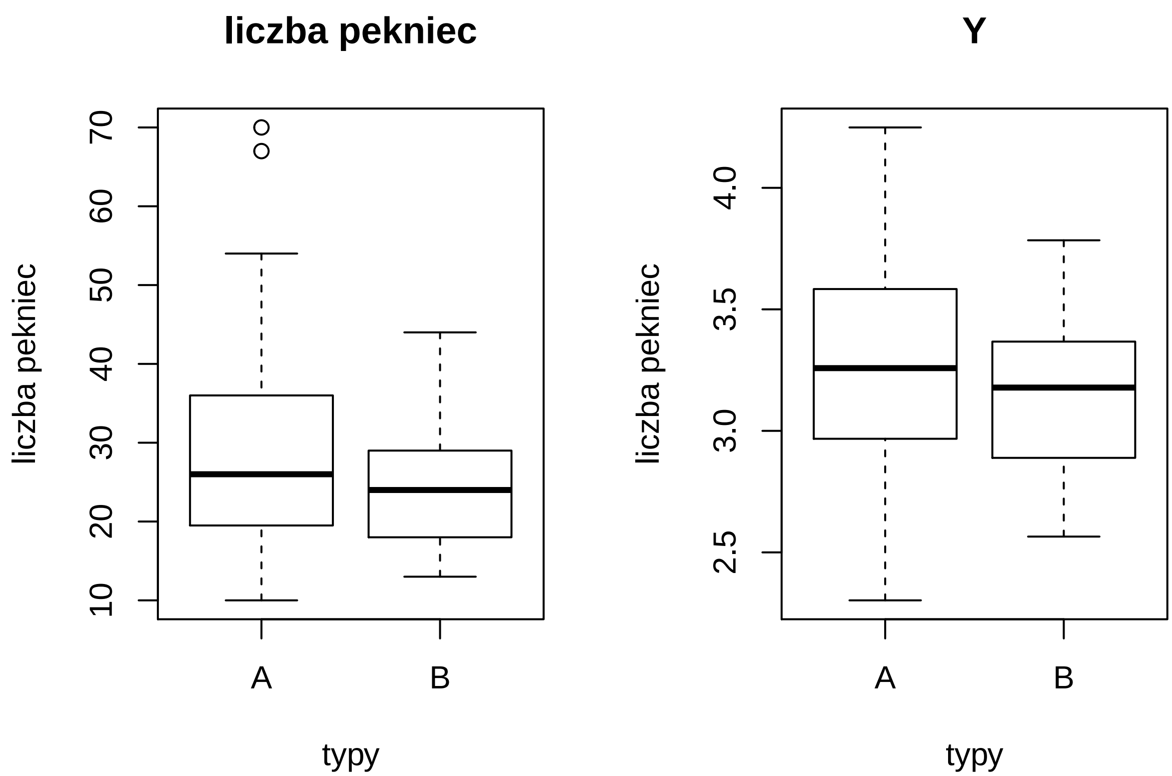
Lista 4

Aneta Przydróżna

Praca nad zbiorem danych, dotyczących zależności między liczbą pęknięć w osnowie przędzy a typem wełny oraz siły naprężenia.

Poniższe boxploty dodtyczą tej samej zależności, ale drugi ma nałożony logarytm na liczbę pęknięć, co sprawia, że pudełka mają regularniejszy kształt i są bardziej skondensowane. Ogólnie dla typu wełny A, dane są bardziej rozrzucone, zdarzają się też obserwacje mocno odstające. Wyniki dla typu B mają medianę niewiele niżej.

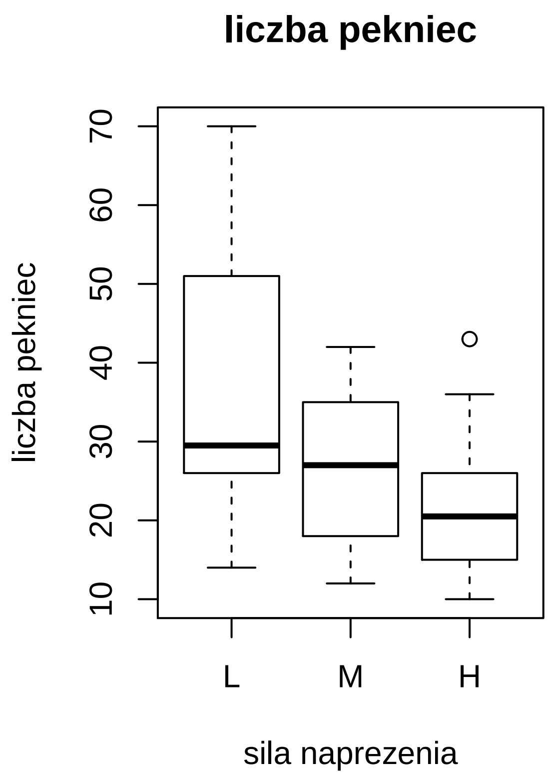


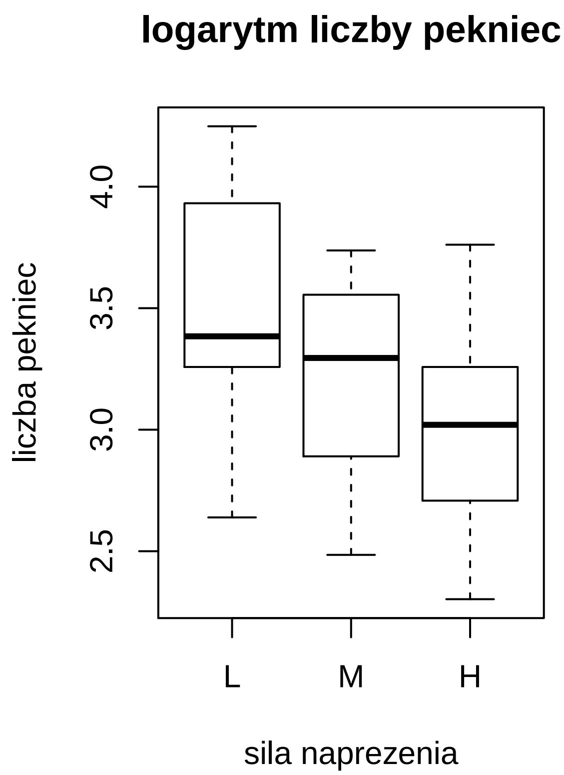
Testowanie hipotezy, że typ wełny ma wpływ na E(Y).

P-wartość (0.11) wykracza poza poziom istotności (0.05), więc nie mamy podstaw do odrzucenia hipotezy zerowej. Test nie dał konkretnej informacji, czy typ A różni się od typu B pod względem pęknięć przędzy.

Boxploty liczby pęknięć i logarytmu z liczby pęknięć dla różnych wielkości sił naprężenia.

Wyniki są zaskakujące, ponieważ okazuje się że im mocniej zwiążemy przędzę, tym mniej pęknięć powstanie. Największy rozrzut jest dla pęknięć przy słabym naprężeniu, wartości zaczynają się przy 15, a kończą na 70. Mediana wynosi 30. Dla pęknięć przy średnim naprężeniu mediana znajduje się nieco niżej, natomiast rozrzut liczb jest dwa razy mniejszy i kończy się blisko wartości 45. Dla pęknięć, przy mocnym naciągnięciu przędzy, mamy jedą wartość odstającą, mediana znajduje się blisko 20, a liczby w pudełku rozstawione są od 10 do 35. Wynika stąd, że przy większej sile naprężenia, otrzymamy mniej pęknięć przędzy. Po nałożeniu logarytmu na badaną własność okazuje się, że pudełka mają bardziej zbliżoną wielkość i znajdują się na mniejszej odległości.





Zależność zmiennej Y od typu i naprężenia przędzy.

```
## (Intercept)       woolB    tensionM    tensionH
##    0.000000    0.158540    0.032048    0.000448
```

Z testów istotności wynika, że typ wełny nie jest istotny, natomiast największy wpływ na Y ma mocne napięcie przędzy.

Szacowana liczba pęknięć dla przędzy typu A ze średnim naprężeniem.

Okazuje się, że Y jest równy 3.29, zatem możemy spodziewać się exp(3.29) $\approx$ 26 pęknięć.

Liczba pęknięć, szacowana w regresji Poissona.

```
## (Intercept)       woolB    tensionM    tensionH
##     0.0e+00     6.5e-05     0.0e+00     0.0e+00
```

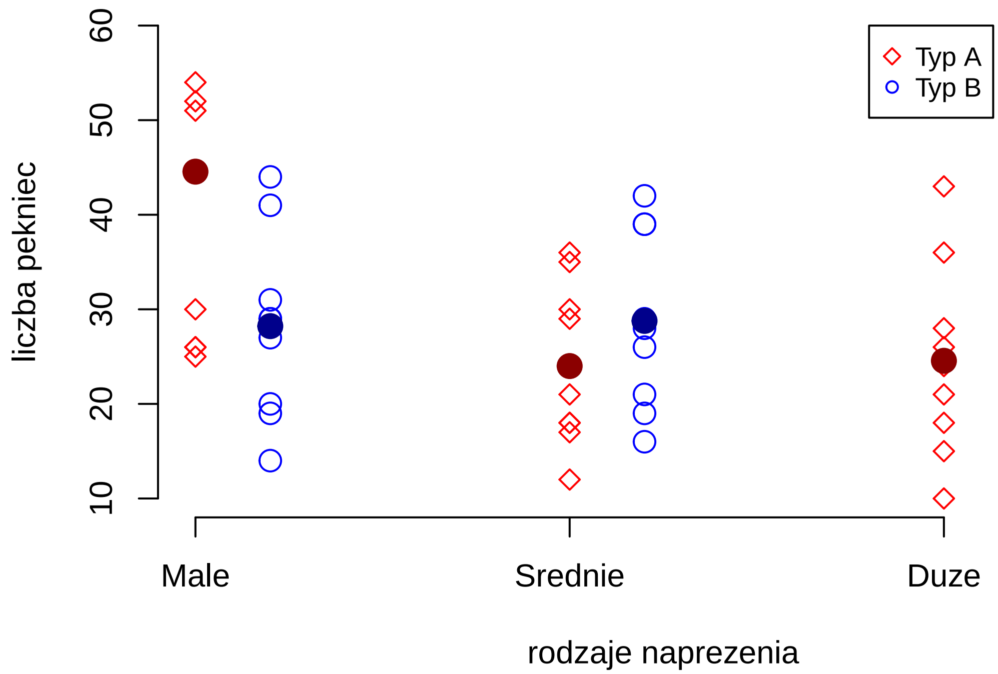
Wszystkie zmienne w modelu są istotne, a przy szacowaniu liczby pęknięć dla typu wełny A przy średnim naprężeniu Y=3.37, zatem pęknięć powinno być 29. Wyniki z poprzedniej analizy wydają się być bardziej wiarygodne, po zapoznaniu się ze zbiorem danych, gdzie średnia badanej własności w tym przypadku wynosi 24.

Liczba pęknięć, szacowana dla rozkładu ujemnego dwumianowego.

```
## (Intercept)       woolB    tensionM    tensionH
##    0.000000    0.065128    0.013965    0.000036
```

W tym przypadku nieistotny jest typ wełny, natomiast naprężenia tak, aczkolwiek w mniejszym stopniu niż poprzednio. Badana liczba pęknięć również wynosi 29. Rozkład Poissona jest jednym z przypadków rozkładu ujemnego dwumianowego, dlatego oba modele są do siebie podobne. Ponadto w badanych danych E(pęknięcia)=28.15, a Var(pęknięcia)=174.2, zatem wariancja jest ponad 6-krotnie większa od wartości oczekiwanej, co powoduje nadmierną dyspersję dla regresji Poissona, w której E(Y)=Var(Y). Dla rozkładu ujemnego dwumianowego $E(Y) = \mu$, a $Var(Y) = \mu + \frac{\mu^2}{r}$, dlatego lepiej użyć rozkładu NB, który lepiej dopasowuje się do takich danych.

Zależność średniej wartości zmiennej Y od rodzaju naprężenia, osobno dla dwóch rodzajów wełny.



Puste kształty przedstawiają pojedyncze próbki, natomiast wypełnione są ich średnią. Największa różnica między typami wełny pojawia się w słabo napiętych przędzach, kiedy pęknięć dla typu A jest średnio 45, a dla typu B niecałe 30. Podobna sytuacja dzieje się w trzecim przypadku, gdy naprężenie jest największe, ale rozrzut średnich jest, co najmniej, 3 razy mniejszy- od 20 do 25. Typ B okazał się gorszy jedynie w próbie ze średnio naprężoną przędzą, z różnicą około 4 pęknięć. Znowu można zauważyć spadkową tendencję dla rodzaju naprężenia. Jeżeli chodzi o typy wełny to lepiej wypadł typ B, którego średnie leżą niżej i są bardziej zbliżone do siebie niż w przypadku A. Najbardziej jednolite są wyniki dla mocno napiętej przędzy wełny typu B. Ma ona również najmniej pęknięć ze wszystkich kategorii.

Sprawdzanie zależności Y od wełny typu A i średniego naprężenia w regresji wielorakiej z interakcjami.

W zwykłym modelu liniowym nieistotna okazuje się zmienna: Typ B i naprężenie mocne. Intercept ma największe znaczenie. Y=3.12 dla typu wełny A oraz średniego naprężenia, zatem pęknięć powinno być około 22. To oznacza, że dodatkowa zmienna zależna spowodowała zbliżenie się oczekiwanej liczby pęknięć, dla konkretnych warunków, do prawdziwej liczby pęknięć w tym przypadku.

```
##    (Intercept)          woolB       tensionM       tensionH woolB:tensionM
##       0.000000       0.017094       0.001326       0.001345       0.015141
## woolB:tensionH
##       0.377494
```

W modelu regresji Poissona z interakcjiami istotne są wszystkie zmienne, a w rozpatrywanej sytuacji szacowane są dokładnie 24 pęknięcia.

```
##    (Intercept)          woolB       tensionM       tensionH woolB:tensionM
##       0.000000       0.621620       0.490828       0.491483       0.612760
## woolB:tensionH
##       0.884967
```

W modelu regresji dla rozkładu ujemnego dwumianowego z interakcjami, jedyną nieistotną zmienną, jest łączona: Typ B i naprężenie mocne. Można zauważyć, że tam gdzie występuje typ wełny B, istotność zmiennych spada. Natomiast szacowana liczba pęknięć jest taka sama jak powyżej i wynosi 24.

```
##    (Intercept)          woolB       tensionM       tensionH woolB:tensionM
##       0.000000       0.022553       0.001791       0.001818       0.018632
## woolB:tensionH
##       0.487021
```

Porównując modele ciążko wybrać czy najlepszy jest ten z rozkładem Poissona czy NB, aczkolwiek po poznaniu wyników dotychczasowych badań, można stwiedzić, że NB lepiej szacuje istotność zmiennych. Jeżeli weźmiemy pod uwagę residual deviance, równe dla regresji Poissona 210.4, z interakcją 182.3, a dla regresji NB 36.3 oraz 31.5. Jest to bardzo duża różnica świadcząca o korzystniejszym działaniu NB na tych danych.