

Half-Normal Plots and Overdispersed Models in R: The hnp Package

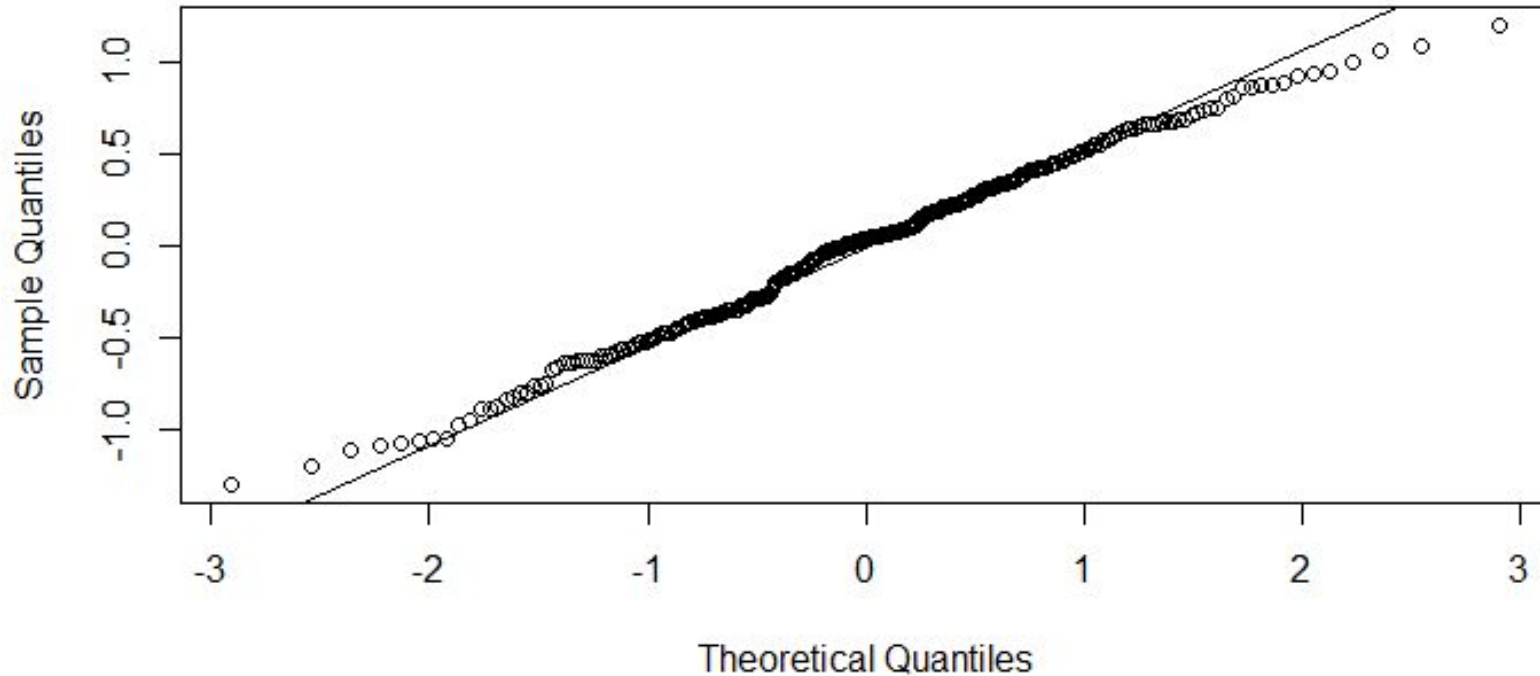
Alicja Gosiewska

Wydział Matematyki i Nauk Informacyjnych
Politechniki Warszawskiej

Moral, R., Hinde, J., & Demétrio, C. (2017). Half-Normal Plots and Overdispersed Models in R: The hnp Package. *Journal of Statistical Software*, 81(10), 1 - 23.

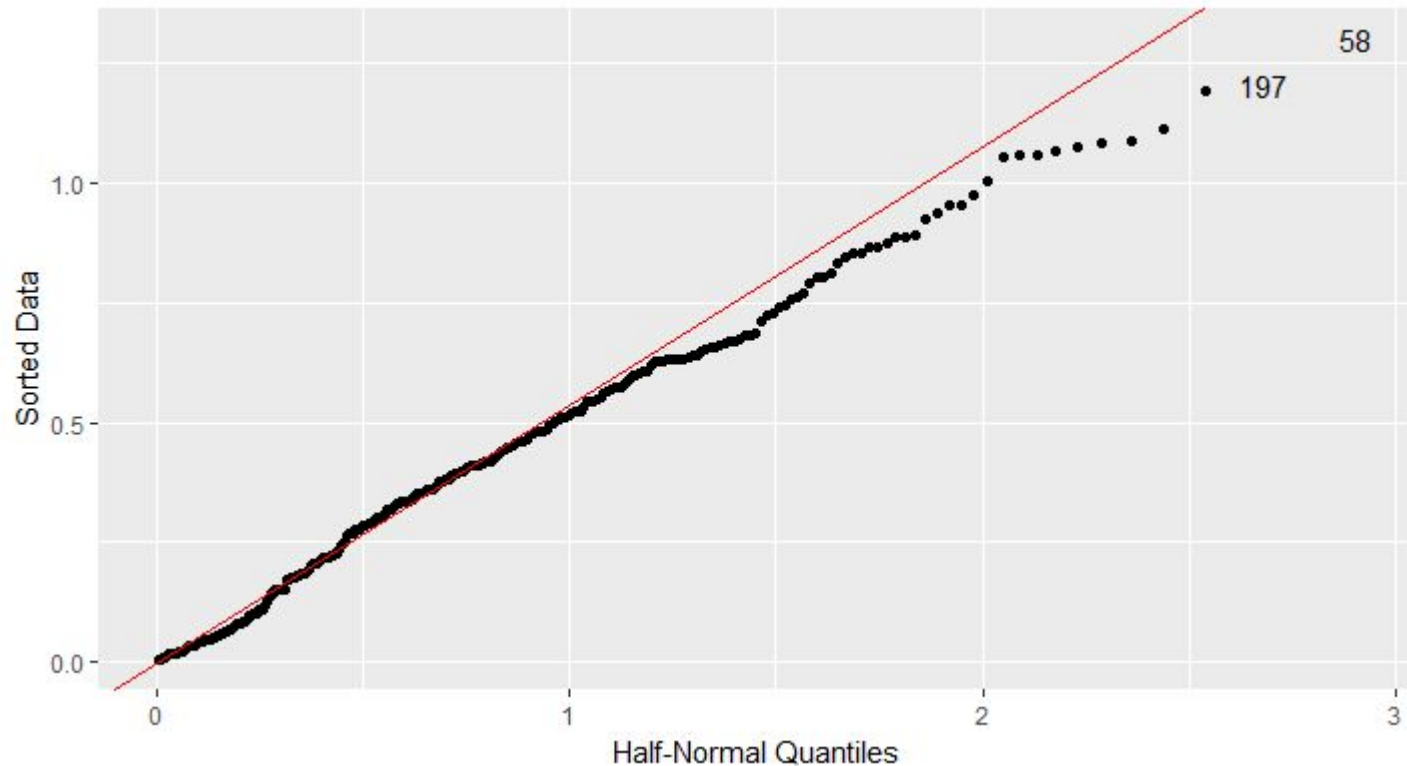
Wykres kwantylowy normalny

Normal Q-Q Plot

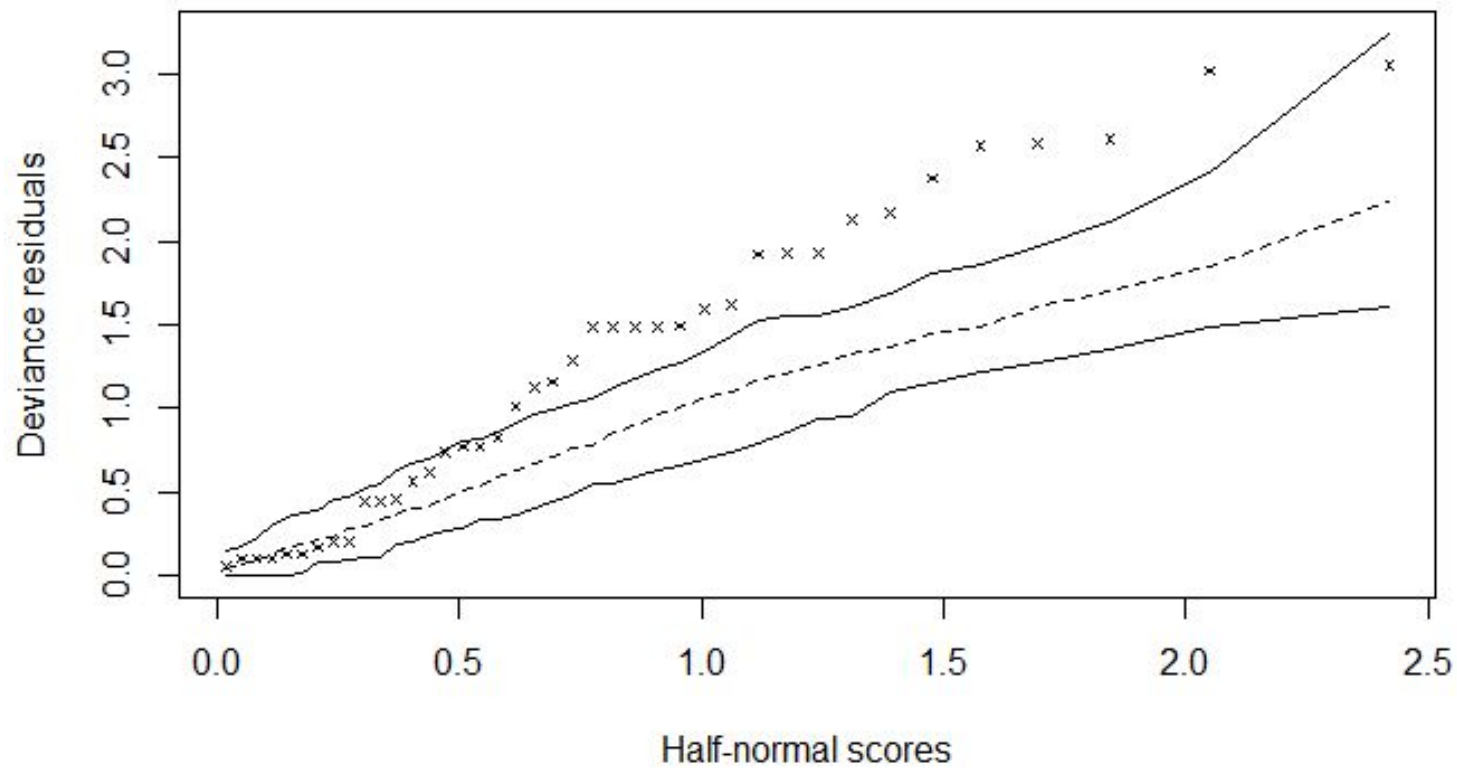


Half-normal plot

```
> gghalfnorm::gghalfnorm(resid)
```



hnp



Przykład 1 - wołek kukurydzowy



Flickr: Sitophilus granarius (Linné, 1758)



Annona mucosa

extract	factor o poziomach odpowiadających Źródłom ekstraktu: <i>leaf, branch, seed, control</i>
m	liczba ziaren kukurydzy na szalce
y	liczba uszkodzonych ziaren kukurydzy

```
> head(corn)
  extract  m  y
1   leaf 35 26
2   leaf 36 25
3   leaf 38 21
4   leaf 38 18
5   leaf 39 30
6   leaf 38  8
```

Regresja logistyczna

$$Y_i \sim B(n_i, \pi_i) \quad X_i = \begin{pmatrix} x_{1,i} \\ x_{2,i} \\ \vdots \\ x_{p,i} \end{pmatrix}$$

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_p x_{p,i}$$

Dewiancja i rezydua dewiancyjne

l^{max} - maksymalna wartość logarytmu wiarogodności dla dopasowanego modelu logistycznego

l_{full}^{max} - maksymalna wartość logarytmu wiarogodności dla dopasowanego modelu wysyconego

$$D = 2(l_{full}^{max} - l^{max}) \sim \chi_{N-p-1}^2$$

$$D = \sum_{i=1}^N (r_i^D)^2$$

Przykład 1 - model regresji logistycznej

```
> fit1_b <- glm(cbind(y, m - y) ~ extract, family = binomial, data = corn)
> anova(fit1_b, test = "Chisq")
```

Analysis of Deviance Table

Model: binomial, link: logit

Response: cbind(y, m - y)

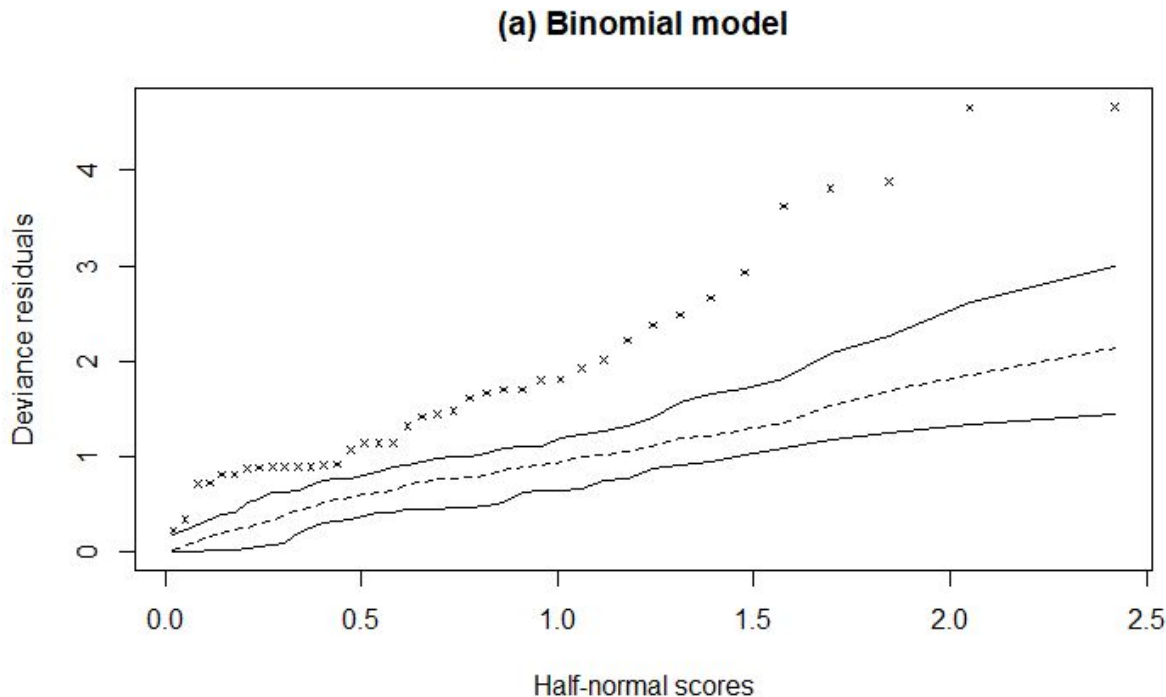
Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			39	801.96	
extract	3	636.04	36	165.92	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

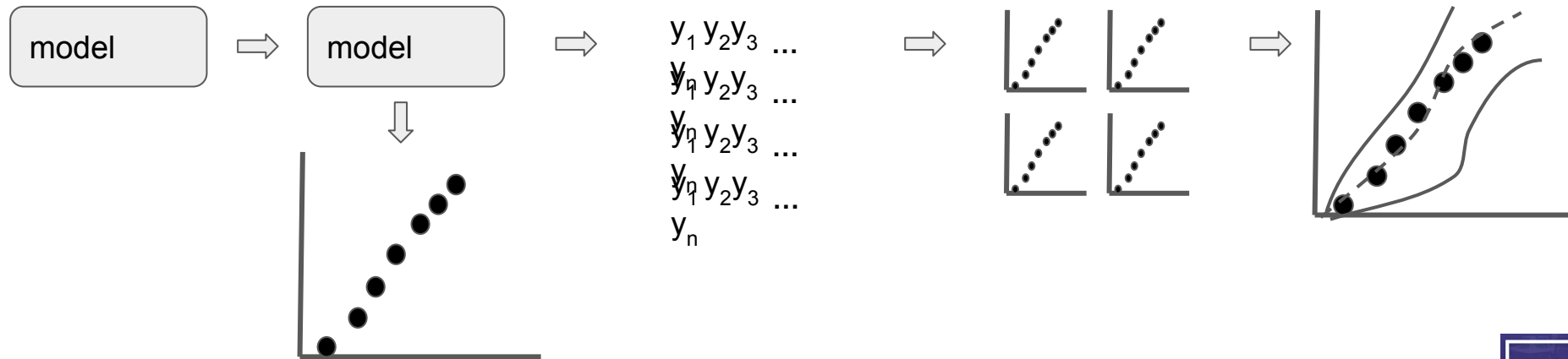
Przykład 1 - wołek kukurydzowy

```
> hnp(fit1_b, xlab = "Half-normal scores", ylab = "Deviance residuals",  
+     main = "(a) Binomial model", pch = 4)
```



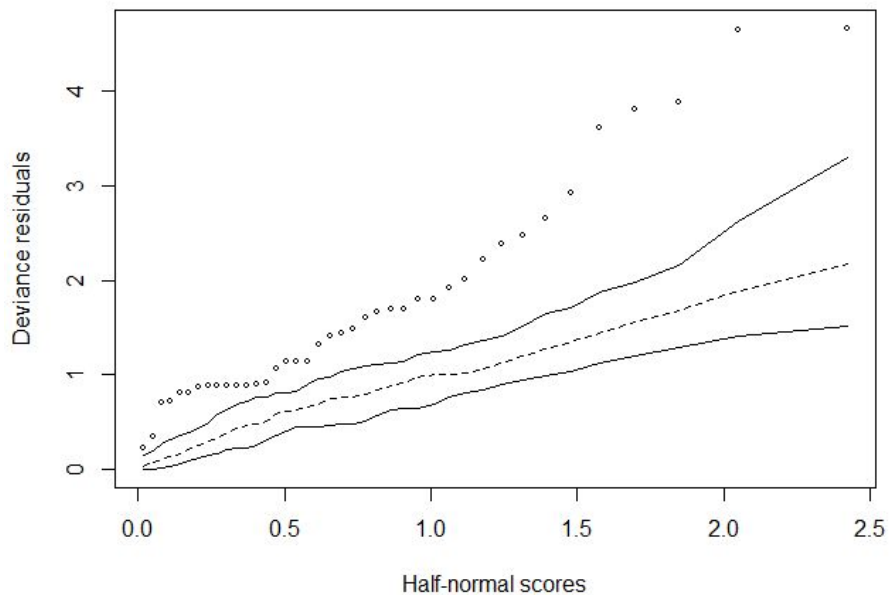
Otrzymywanie obwiedni

1. Dopasowanie modelu.
2. Wyliczenie wartości diagnostycznych.
3. Zasymulowanie zmiennych odpowiedzi posługując się macierzą modelu, rozkładem błędów oraz wyestymowanymi parametrami wejściowego modelu
4. Dopasowanie modelu do każdej z zasymulowanych odpowiedzi oraz wyliczenie wartości diagnostycznych.
5. Wyliczenie wybranych percentyli (np 2.5 i 97.5) wartości diagnostycznych uzyskanych z symulacji oraz użycie ich do utworzenia obwiedni.

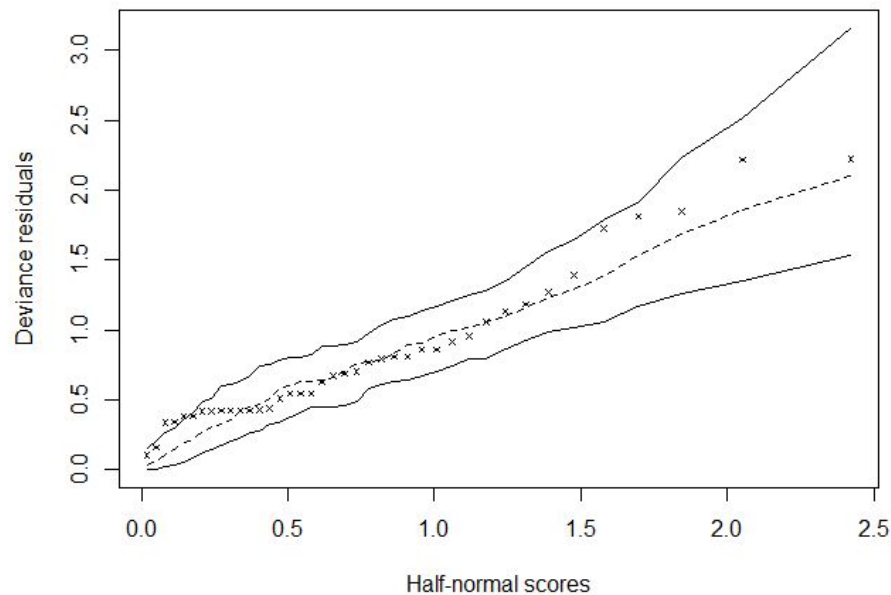


Przykład 1 - wołek kukurydzowy

(a) Binomial model



(b) Quasi-binomial model



Nowe klasy modeli

```
> fit_nbI <- gamlss(embryos ~ block + poly(dose, 2) * sugar, family = NBII(), data = orange)

> d.fun <- function(obj) resid(obj)
> s.fun <- function(n, obj) rNBII(n, obj$mu.fv, obj$sigma.fv)
> f.fun <- function(y.) gamlss(y. ~ block + poly(dose, 2) * sugar, family = NBII(), data = orange)

> hnp(fit_nbI, newclass = TRUE,
+     diagfun = d.fun, simfun = s.fun, fitfun = f.fun,
+     xlab = "Half-normal scores", ylab = "z-scores",
+     main = "", pch = 4, cex = 1, cex.lab = .8, cex.axis = .8)
```

