V. Storme May 2019

VARIA

Outcome Variable	Are the observations independent or correlated?		Alternatives (assumptions violated)
	independent	correlated	
Continuous (e.g. pain scale, cognitive function)	Ttest ANOVA Linear correlation Linear regression	Paired ttest **Repeated-measures ANOVA** **Mixed models/GEE modeling**	Wilcoxon sign-rank test Wilcoxon rank-sum test Kruskal-Wallis test Spearman rank correlation coefficient
Binary or categorical (e.g. fracture yes/no)	Risk difference/Relative risks Chi-square test Logistic regression	McNemar's test Conditional logistic regression **GEE modeling**	Fisher's exact test McNemar's exact test
Count data	Poisson regression Negative binomial regression	**GEE modeling**	
Time-to-event (e.g. time to fracture)	Rate ratio **Kaplan-Meier statistics** **(Parametric regression)** **Cox regression**	Frailty model	Time-varying effects if PH assumption violated

Modelling counts
the Poisson distribution

- Famous example by von Bortkiewicz (1898): observe the number of soldiers in the Prussian army who got kicked by horses over a number of years and corps

# kicks (=k)	# soldiers	fraction	Expected fraction
0	109	0.545	0.543
1	65	0.325	0.331
2	22	0.110	0.101
3	3	0.015	0.021
4	1	0.005	0.003

- Average nr of horsekicks per soldier:

$$\bar{X} = \frac{0*109+1*65+2*22+3*3+4*1}{200} = 0.61$$

- The probability that the nr of kicks=k

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \hat{\lambda} = \bar{X}$$

Generalized linear models (GLM)

- A glm consists of 3 parts
  - A ***distribution***, specifying the conditional distribution of the the response Y given the predictor variables

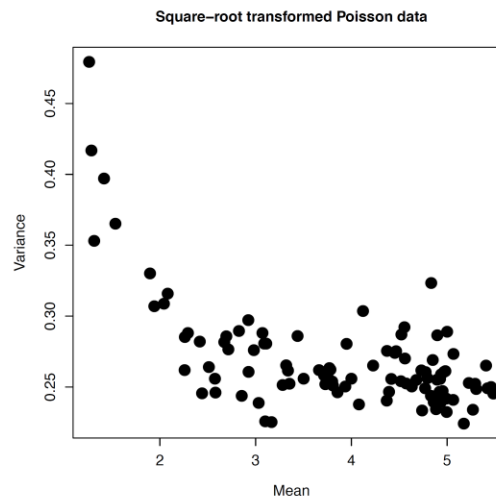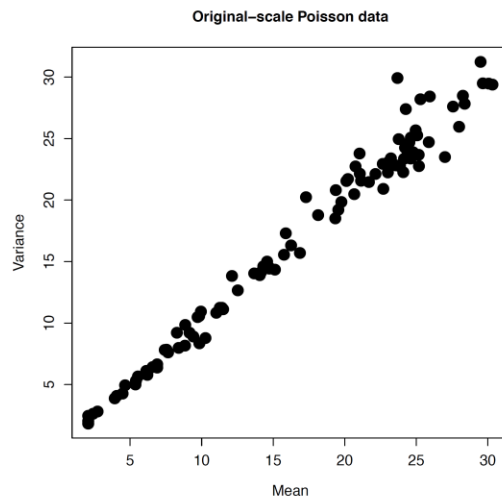 $$\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

  - A ***linear predictor***

  $$g(E[Y|X]) = \eta$$

  - A ***link function*** g, linking the conditional expected value of Y to η

Poisson regression: link function

- Distribution: Poisson

- Link function: log

  - The link function transforms the mean, not the observed values

  $$log(E[Y|X]) \neq E[log(Y|X)]$$

  - Transforming the observed values changes the association between mean and variance

offset

- Incorporate normalization factor as offset
  - Eg in lateral root data: count nr of lateral roots and normalize by root length

$$log(E[Y|X]) = \beta_0 + \beta_1 X_1 + log(rootlength)$$

- The counts are not explicitly scaled

$$log(E[\frac{Y}{rootlength}|X]) = \beta_0 + \beta_1 x_1$$

Longitudinal analysis

- Measurements on the same subject ifo time
- Measurements done on the same subjects are correlated

Logistic regression

- Binary response
- The residuals are non-normal and heteroscedastic
- Predictions can take values >1 of <0
- Logit model $\mathbf{ln[p/(1-p)]=\beta_0+\beta_1X+\varepsilon}$ solves this problem
- p the probability of Y=1 $\mathbf{P(Y=1|X)}$
- p/(1-p)=odds ratio
- ln[p/(1-p)]=ln(odds ratio)="logit"

Survival analysis

- Binary response
- Analyses time to event
- Censored data
- examples:
  - Time till death
  - Time till recidivism
  - Time until 70% of the machines break down
- The survival function is the probability to survive beyond S(t)
- Univariate
  - Kaplan-Meier curve
- Multivariate
  - Cox proportional hazards

Looking for courses or consulting

- @UGent
  - Faculty of Medicine: https://www.ugent.be/ge/en/services/biostatistics-unit
  - Faculty of Veterinary Medicine: contact Luc Duchateau
  - Department of Plant Biotechnology and Bioinformatics, contact: Veronique Storme
  - Faculty of Psychology and Educational Sciences - Statistics support
  - Other faculties: FIRE
- https://lstat.kuleuven.be/consulting
- https://www.uantwerpen.be/en/research-and-innovation/research-at-uantwerp/core-facilities/core-facilities/statua/who-we-are/

- https://www.flames-statistics.com/! Flames Summer School @UGent
- https://www.ugent.be/we/en/services/ICES

Useful R links

- https://onlinecourses.science.psu.edu/stat464/
- http://www.r-tutor.com/
- http://www.cookbook-r.com/
- https://www.datacamp.com/
- http://www.statmethods.net/index.html
- http://www.rdocumentation.org/
- https://www.zoology.ubc.ca/~schluter/R/
- http://www.sr.bham.ac.uk/~ajrs/R/index.html
- http://www.r-bloggers.com/
- http://r4stats.com/
- http://www.ats.ucla.edu/stat/r/
- http://manuals.bioinformatics.ucr.edu/home