

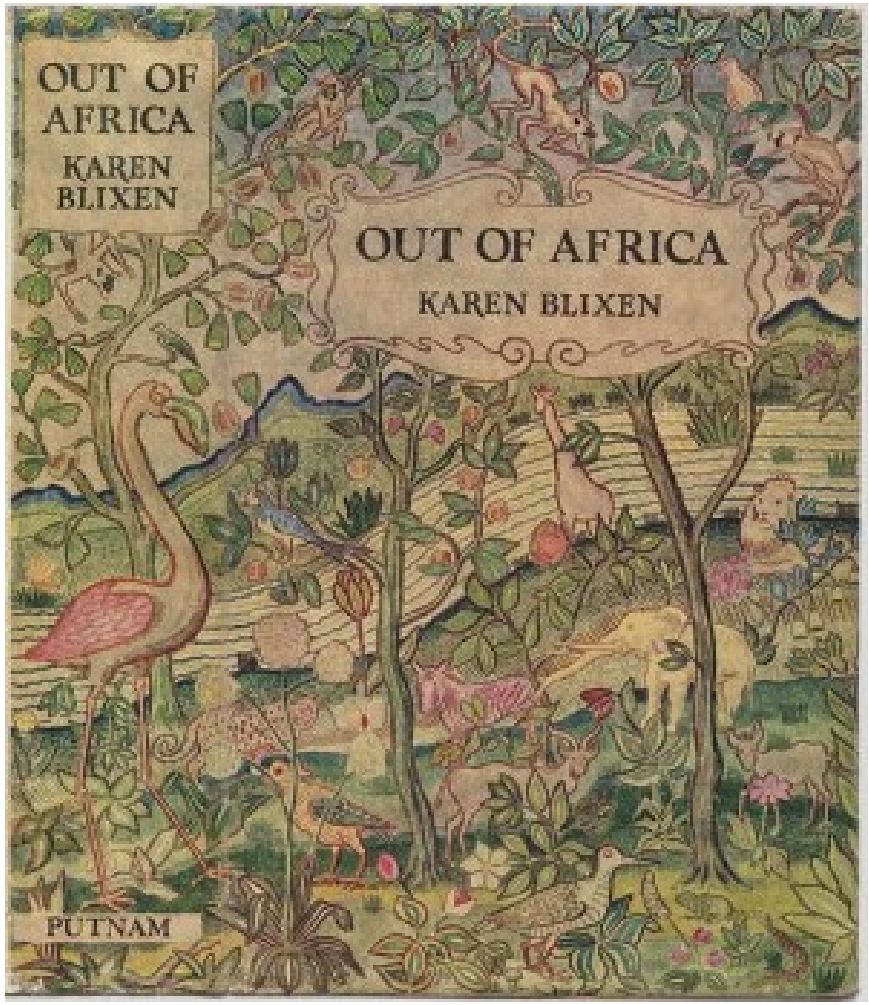
University of Zurich[UZH]

EBPI Epidemiology, Biostatistics and Prevention Institute

Out-of-Normal Land Transformations

Torsten Hothorn

KU Leuven, November 2026


OUT OF
AFRICA
KAREN
BLIXEN
OUT OF AFRICA
KAREN BLIXEN
PUTNAM

The Statistician's Paradise

$$\mathrm{N}_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Student 1908

VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Wilcoxon 1945

INDIVIDUAL COMPARISONS BY RANKING METHODS

FRANK WILCOXON
American Cyanamid Co.

Lehmann 1953

THE POWER OF RANK TESTS[1]

By E. L. Lehmann

Stanford University and University of California, Berkeley

suggests

$$F(y \mid \text{active}) = m(F(y \mid \text{placebo}), \beta)$$

as a model for the two distribution functions in a 2-way layout.

Maybe $m(p, \beta) = p^{\exp(-\beta)}$ ("Lehmann alternative")?

Special Models

- $m_{\mathrm{L}}(p, \beta) = p^{\exp(-\beta)}$ (Lehmann alternatives, loglog link),
- $m_{\mathrm{PH}}(p, \beta) = 1 - (1 - p)^{\exp(-\beta)}$ (proportional hazards, cloglog link)
- $m_{\mathrm{PO}}(p, \beta) = \mathrm{expit}(\mathrm{logit}(p) - \beta)$ (proportional odds, logit link), and
- $m_{\mathrm{Cd}}(p, \beta) = \Phi(\Phi^{-1}(p) - \beta)$ (generalised Cohen's d, probit link)

The last three models correspond to log-rank (or Savage) tests, Wilcoxon tests, and van der Waerden normal scores tests testing $H_0 : \beta = 0$ *against these alternatives*.

The “Normal” Linear Regression Model

$$Y = \alpha + \boldsymbol{x}^\top \boldsymbol{\gamma} + \sigma Z, \quad Z \sim \mathrm{N}(0, 1)$$

- everything but “normal”
- most special and idealised case
- $Y \mid \boldsymbol{x} \sim \mathrm{N}(\alpha + \boldsymbol{x}^\top \boldsymbol{\gamma}, \sigma^2)$
- no way escaping normal land

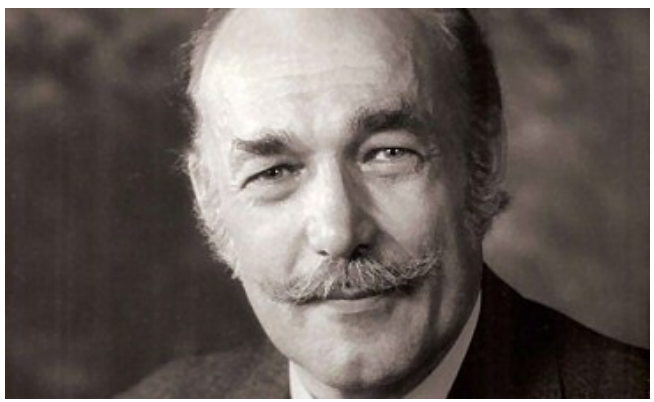
Generalisation I: GLMs (1972)

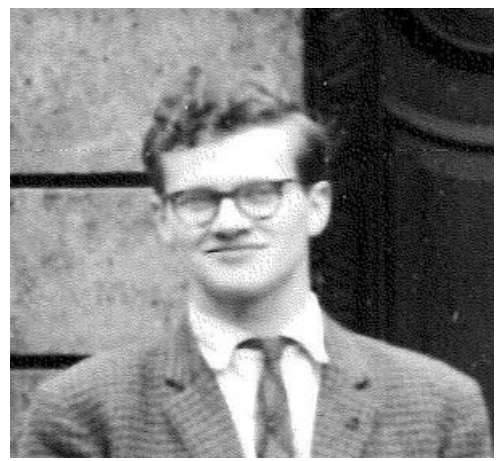
$$
\begin{aligned}
Y &= \alpha + \boldsymbol{x}^\top \boldsymbol{\gamma} + \sigma Z, && Z \sim \mathsf{N}(0,1) \\
\iff \mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x}) &= \alpha + \boldsymbol{x}^\top \boldsymbol{\gamma} && , \quad Y \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathsf{N}(,) \\
\hookrightarrow g(\mathbb{E}(Y \mid \boldsymbol{X} = \boldsymbol{x})) &= \alpha + \boldsymbol{x}^\top \boldsymbol{\gamma} && , \quad Y \mid \boldsymbol{X} = \boldsymbol{x} \sim \mathsf{ExpFam}(,)
\end{aligned}
$$

Generalized Linear Models

By J. A. NELDER and R. W. M. WEDDERBURN

Rothamsted Experimental Station, Harpenden, Herts





Generalisation II

$$
\begin{aligned}
Y &= \alpha + \boldsymbol{x}^\top \boldsymbol{\gamma} + \sigma Z, \quad Z \sim \mathrm{N}(0,1) \\
Y - \alpha &= \boldsymbol{x}^\top \boldsymbol{\gamma} + \sigma Z, \quad Z \sim \mathrm{N}(0,1) \\
\iff \frac{Y - \alpha}{\sigma} &= \boldsymbol{x}^\top \frac{\boldsymbol{\gamma}}{\sigma} + Z, \quad Z \sim \mathrm{N}(0,1) \qquad (1) \\
\iff \frac{Y - \alpha}{\sigma} &= \boldsymbol{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim \mathrm{N}(0,1) \\
\hookrightarrow h(Y) &= \boldsymbol{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim \mathrm{N}(0,1) \\
\hookrightarrow h(Y) &= \boldsymbol{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim F
\end{aligned}
$$

Transformation models, $Z \in \mathbb{R}$ with absolute continuous log-concave density $f = F'$, $h : \mathbb{R} \to \mathbb{R}$ nondecreasing

(1) sacrifices collapsibility ($\boldsymbol{\beta}$ is Cohen's d)

An Analysis of Transformations (1964)



An Analysis of Transformations

By G. E. P. Box and D. R. Cox

University of Wisconsin *Birkbeck College, University of London*

[Read at a Research Methods Meeting of the Society, April 8th, 1964, Professor D. V. Lindley in the Chair]

Conceptually more powerful but, at the time, hard to compute and thus restricted to

$$h(y \mid \lambda) = \begin{cases} \frac{y^\lambda + 1}{\lambda} & \lambda > 0 \\ \log(y) & \lambda = 0 \end{cases} \quad \text{with } Z \sim \mathrm{N}(0, \sigma^2)$$

“Box-Cox” power transformation

Out-of-Normal Lands

Normal land

$$\begin{aligned} Y &= \alpha + \boldsymbol{x}^\top \boldsymbol{\gamma} + \sigma Z, \quad Z \sim \mathsf{N}(0,1) \\ \iff \mathbb{P}(Y \leq y \mid \boldsymbol{x}) &= \Phi\left(\frac{y - (\alpha + \boldsymbol{x}^\top \boldsymbol{\beta})}{\sigma}\right) \\ \iff Y \mid \boldsymbol{x} &\sim \mathsf{N}(\alpha + \boldsymbol{x}^\top \boldsymbol{\beta}, \sigma^2) \end{aligned}$$

Out-of-normal land

$$\begin{aligned} h(Y) &= \boldsymbol{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim \mathsf{N}(0,1) \\ \Rightarrow \mathbb{P}(Y \leq y \mid \boldsymbol{x}) &= \Phi\left(h(y) - \boldsymbol{x}^\top \boldsymbol{\beta}\right) \end{aligned}$$

Note: Last expression works for at least ordered outcomes.

Specials: GLMs and Cumulative Probability Models

E.g. binary $Y \in \{0, 1\}$:

$$\mathbb{P}(Y \leq y \mid \boldsymbol{x}) \quad = \quad \Phi\left(\alpha - \boldsymbol{x}^\top \boldsymbol{\beta}\right)$$

a.k.a. binary GLM with probit link Φ^{-1}

Ordered $Y \in \{y_1, \ldots, y_K\}$:

$$\mathbb{P}(Y \leq y \mid \boldsymbol{x}) \quad = \quad \Phi\left(\alpha_k - \boldsymbol{x}^\top \boldsymbol{\beta}\right)$$

a.k.a. cumulative probability model with probit link

Other Paradises

Changing the distribution function F of the error term Z changes the link function.

The complementary loglog link ($\log(-\log(1-p))$)

$$
\begin{aligned}
h(Y) &= \boldsymbol{x}^\top \boldsymbol{\beta} + Z, \quad Z \sim \text{Gompertz} \\
\Rightarrow \mathbb{P}(Y \leq y \mid \boldsymbol{x}) &= 1 - \exp\left(-\exp\left(h(y) - \boldsymbol{x}^\top \boldsymbol{\beta}\right)\right) \\
\iff 1 - \mathbb{P}(Y \leq y \mid \boldsymbol{x}) &= \exp\left(-\exp\left(h(y) - \boldsymbol{x}^\top \boldsymbol{\beta}\right)\right) \\
\iff -\log(1 - \mathbb{P}(Y \leq y \mid \boldsymbol{x})) &= \exp\left(h(y) - \boldsymbol{x}^\top \boldsymbol{\beta}\right) \\
\iff \Lambda(y \mid \boldsymbol{x}) &= \exp(h(y)) \exp\left(-\boldsymbol{x}^\top \boldsymbol{\beta}\right) \\
\iff \lambda(y \mid \boldsymbol{x}) &= \lambda_0(y) \exp\left(-\boldsymbol{x}^\top \boldsymbol{\beta}\right)
\end{aligned}
$$

leads to proportional hazards models; $h(y) = \vartheta_1 + \vartheta_2 \log(y)$ gives the Weibull model.

Cox 1972

Regression Models and Life-Tables

By D. R. Cox

Imperial College, London

The main problem considered in this paper is that of assessing the relation between the distribution of failure time and $\mathbf{z}$. This will be done in terms of a model in which the hazard is

$$\lambda(t; \mathbf{z}) = \exp(\mathbf{z}\boldsymbol{\beta})\,\lambda_0(t), \tag{9}$$

No word about his own 1964 paper!

Likelihoods

Suppose an observation is an interval $(\underline{y}, \bar{y}]$; as it always is, we just ignore it.

The log-likelihood contribution is then

$$\log\left(\mathbb{P}(Y \leq \bar{y} \mid \boldsymbol{x}) - \mathbb{P}(Y \leq \underline{y} \mid \boldsymbol{x})\right)$$

which is easy to evaluate in "semiparametric transformation models" of the form

$$\mathbb{P}(Y \leq y \mid \boldsymbol{x}) = F(h(y) - \boldsymbol{x}^\top \boldsymbol{\beta})$$

for some absolutely continuous $Z \sim F$ on $\mathbb{R}$.

h and $\boldsymbol{\beta}$ are the unknown parameters.

Maximum Likelihood

Empirical and interval-censored likelihoods covered.

For suitably parameterised functions $h(y \mid \boldsymbol{\vartheta})$, joint maximum likelihood estimation of $\boldsymbol{\vartheta}$ and $\boldsymbol{\beta}$ is a convex problem. Analytic gradients and Hessians available, so very fast and "simple" to compute.

New: `stats::free1way()` for semiparametrically efficient population- and permutation-based inference in distribution-free stratified K-sample oneway layouts ("simple" $\approx$ 2500 lines of R code).

Soon: `freelm()` (https://codeberg.org/thothorn/freelm) for distribution-free linear models.

No need to profile-out h (as done by the partial likelihood) $\Rightarrow$ standard inference.

Extensions

Mixed or additive models

$$\mathbb{P}(Y \leq y \mid \boldsymbol{x}) = F\left(h(y) - \sum_j g_j(x_j)\right)$$

implemented in package **tramME** feature transformations of response (h) and covariates (g_j).

Can be used to fit additive proportional hazards normal frailty models under interval-censoring (e.g., to disease-free survival).

Multivariate Normal Land

Fisher's Linear Discriminant Analysis

$$\boldsymbol{Y} \mid x = 0 \sim \mathrm{N}_J(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$$
$$\boldsymbol{Y} \mid x = 1 \sim \mathrm{N}_J(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$$

Quadratic Discriminant Analysis

$$\boldsymbol{Y} \mid x = 0 \sim \mathrm{N}_J(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$$
$$\boldsymbol{Y} \mid x = 1 \sim \mathrm{N}_J(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$$

Multivariate Out-of-normal Land

Transformation Discriminant Analysis

$$\begin{aligned} \boldsymbol{h}(\boldsymbol{Y} \mid x = 0) &\sim \mathrm{N}_J(\mathbf{0}, \boldsymbol{\Sigma}), \quad \mathrm{diag}(\boldsymbol{\Sigma}) = \mathbf{1} \\ \boldsymbol{h}(\boldsymbol{Y} \mid x = 1) &\sim \mathrm{N}_J(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned}$$

Transformation Quadratic Discriminant Analysis

$$\begin{aligned} \boldsymbol{h}(\boldsymbol{Y} \mid x = 0) &\sim \mathrm{N}_J(\mathbf{0}, \boldsymbol{\Sigma}_0), \quad \mathrm{diag}(\boldsymbol{\Sigma}_0) = \mathbf{1} \\ \boldsymbol{h}(\boldsymbol{Y} \mid x = 1) &\sim \mathrm{N}_J(\mathbf{0}, \boldsymbol{\Sigma}_1), \quad \mathrm{diag}(\boldsymbol{\Sigma}_1) = \mathbf{1} \end{aligned}$$

Sewak *et al.* (2026b)

Multivariate Out-of-normal Land

Multivariate transformation models (Klein *et al.*, 2022), a.k.a. “Gaussianisation”, “nonparanormal model”, or “normalising flow”:

$$\mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{y} \mid \boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{\Phi}_J(\boldsymbol{h}(\boldsymbol{y} \mid \boldsymbol{x}) \mid \boldsymbol{\Sigma}(\boldsymbol{x}))$$

Unfortunately, this is a beast when we need to actually evaluate probabilities.

Adding some Hygge

Write $\boldsymbol{\Sigma}(\boldsymbol{x}) = \boldsymbol{\Omega}(\boldsymbol{x})^{-1}\boldsymbol{\Omega}(\boldsymbol{x})^{-\top}$ with lower-triangular $\boldsymbol{\Omega}(\boldsymbol{x})$ and $\operatorname{diag}(\boldsymbol{\Omega}(\boldsymbol{x})) > \mathbf{0}$.

Write $\boldsymbol{\Omega}(\boldsymbol{x}) = \boldsymbol{\Lambda}(\boldsymbol{x}) \operatorname{diag}(\boldsymbol{\Lambda}^{-1}(\boldsymbol{x})\boldsymbol{\Lambda}^{-\top}(\boldsymbol{x}))^{1/2}$ with lower-triangular unit matrix $\boldsymbol{\Lambda}(\boldsymbol{x})$ containing $J(J-1)/2$ unconstrained off-diagonal parameters possibly depending on $\boldsymbol{x}$. Ensures $\operatorname{diag}(\boldsymbol{\Sigma}(\boldsymbol{x})) = \mathbf{1}$.

Write $\boldsymbol{h}(\boldsymbol{y} \mid \boldsymbol{x}) = (h_1(y_1 \mid \boldsymbol{x}), \ldots, h_J(y_J \mid \boldsymbol{x}))^\top$, for example with $h_j(y_j \mid \boldsymbol{x}) = h_j(y_j) - \boldsymbol{x}^\top \boldsymbol{\beta}_j$.

Likelihoods

With box observations $(\underline{y}_j, \bar{y}_j]$ for dimensions $j = 1, \ldots, J$, we compute the probability

$$\mathbb{P}\left(\bigcap_{j=1}^{J} \underline{y}_j < Y_j \leq \bar{y}_j \mid \boldsymbol{X} = \boldsymbol{x}\right) =$$

$$\mathbb{P}\left(\bigcap_{j=1}^{J} h_j(\underline{y}_j \mid \boldsymbol{x}) < h_j(Y_j \mid \boldsymbol{x}) \leq h_j(\bar{y}_j \mid \boldsymbol{x}) \mid \boldsymbol{X} = \boldsymbol{x}\right)$$

via the integral

$$(2\pi)^{-J/2} \det(\boldsymbol{\Omega}(\boldsymbol{x})) \int_{h_1(\underline{y}_1|\boldsymbol{x})}^{h_1(\bar{y}_1|\boldsymbol{x})} \cdots \int_{h_J(\underline{y}_J|\boldsymbol{x})}^{h(\bar{y}_J|\boldsymbol{x})} \exp\left(-\frac{1}{2}\|\boldsymbol{\Omega}(\boldsymbol{x})\boldsymbol{y}\|_2^2\right) \, d\boldsymbol{y}$$

where $\det(\boldsymbol{\Omega}(\boldsymbol{x})) = \prod_{j=1}^{J} \boldsymbol{\Omega}(\boldsymbol{x})_{jj}$

Computations

This is still nasty but, with the help of a new implementation of Genz' algorithm, manageable (package **mvtnorm**, with scores).

For short intervals, we can approximate log-probabilities by log-densities, which are very fast to compute for smooth h_j.

Joint maximum likelihood estimation of all model parameters ($\boldsymbol{h}$ and $\boldsymbol{\Omega}$) becomes possible but is no longer a convex problem (Hothorn, 2026).

`tram::Mmlt()` provides a proof-of-concept implementation.

Applications

- polychoric correlations (Jöreskog, 1994)
- competing risks (Czado and Van Keilegom, 2023): Y_j time to event j, all marginally Weibull
- dependent censoring (Deresa and Keilegom, 2024): Y_1 time to right-censoring marginally Weibull, Y_2 time to event with $h_2(y) = \Phi^{-1}(1 - \exp(-\exp(h(y_2) - \boldsymbol{x}^\top \boldsymbol{\beta})))$, i.e., a proportional hazards model
- (time-dependent) receiver operating characteristic curves for biomarker combinations (Sewak *et al.*, 2026b,a)
- copula regression (Lesaffre and Kaufmann, 1992; Song *et al.*, 2009; Masarotto and Varin, 2012; Barratt and Boyd, 2023)
- multivariate count data (Graz *et al.*, 2026)
- structural equation models
- …

Allows us to pick what we really want, not what is available off-the-shelf.

Nonparanormal Adjusted Marginal Inference

Consider a randomised trial with at least ordered outcome Y, $J-1$ baseline covariates $\boldsymbol{X}$ and a binary treatment w. The treatment effect is a marginal common odds ratio $\exp(\delta)$

$$\begin{aligned} \mathbb{P}(Y \leq y \mid w = 0) &= \text{expit}(h(y)) \\ \mathbb{P}(Y \leq y \mid w = 1) &= \text{expit}(h(y) - \delta) \end{aligned}$$

and the Wilcoxon test is the semiparametric score test for $H_0 : \delta = 0$ (with nonparametric h and empirical likelihood).

How can we adjust for $\boldsymbol{X}$ to increase power while maintaining interpretability?

Nonparanormal Adjusted Marginal Inference

The normal-land-move

$$\mathbb{P}(Y \leq y \mid w, \boldsymbol{x}) \quad = \quad \text{expit}(h(y) - w\delta - \boldsymbol{x}^\top \boldsymbol{\beta})$$

makes δ a *conditional* log-odds ratio b/c lack of collapsibility as in (1).

Go multivariate $(X_1, \ldots, X_{J-1}, Y)^\top$ and model

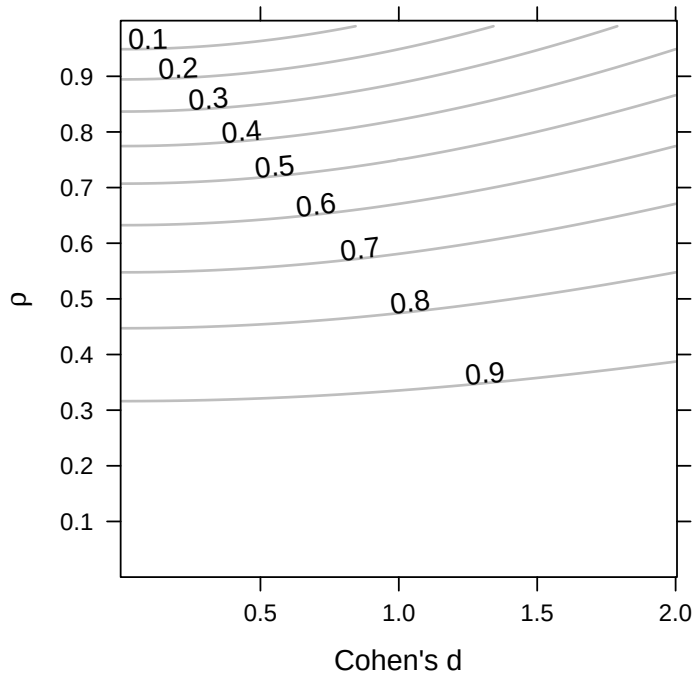
$$\begin{aligned}\mathbb{P}(X_1 \leq x_1, \ldots, X_{J-1} \leq x_{J-1}, Y \leq y \mid w) = \\ \boldsymbol{\Phi}_J \left(h_1(x_1), \ldots, h_{J-1}(x_{J-1}), h_J(y \mid w) \mid \boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}^{-\top}\right)\end{aligned}$$

with

$$h_J(y \mid w) = \Phi^{-1}(\text{expit}(h(y) - w\delta))$$

maintaining the marginal interpretability of δ ("NAMI", Dandl and Hothorn, 2026).

Potential Benefits



Ratio of adjusted to unadjusted squared standard errors in a normal model as a function of Cohen's *d* treatment effect and correlation ρ between covariate and outcome.

Nonparanormal Adjusted Marginal Inference

If (some of) the baseline covariates are (strongly) correlated with the outcome, the standard error of $\hat{\delta}$ decreases. If not, the standard error is about the same as the one obtained from the marginal estimate.

Model assumptions can be tested empirically.

Works for at least ordered and potentially censored outcomes with Lehmann-type treatment effects.

Dependent censoring: Treat time to censoring as “covariate” with marginal Weibull model.

Heterogeneous treatment effects can be estimated by making $\boldsymbol{\Omega}(w)$ treatment-dependent, https://doi.org/10.48550/arXiv.2605.23691.

Summary

- GLMs and conditional mean models look like paradise, but are a rabbit hole.
- Transformation models are much more versatile and fun.
- Discover connections between many classical and contemporary models.
- Stick to core principles (ML) and simplify your life.
- Some computational challenges in the multivariate world, but things are moving fast these days.

References

Barratt S, Boyd S (2023). "Covariance Prediction via Convex Optimization." *Optimization and Engineering*, **24**(3), 2045–2078. doi:10.1007/s11081-022-09765-w.

Czado C, Van Keilegom I (2023). "Dependent Censoring Based on Parametric Copulas." *Biometrika*, **110**(3), 721–738. doi:10.1093/biomet/asac067.

Dandl S, Hothorn T (2026). "Nonparanormal Adjusted Marginal Inference (with Discussion)." *Biometrics*, **82**(2), ujag065. doi:10.1093/biomtc/ujag065.

Deresa NW, Keilegom IV (2024). "Copula Based Cox Proportional Hazards Models for Dependent Censoring." *Journal of the American Statistical Association*, **119**(546), 1044–1054. doi:10.1080/01621459.2022.2161387.

Graz L, Barbanti L, Brandl R, Hothorn T (2026). "Joint Count Transformation Models with Covariate-dependent Correlations." *Ecology*. doi:10.48550/arXiv.2201.13095. Accepted for publication 2026-07-06.

Hothorn T (2026). "On Nonparanormal Likelihoods." *Peer Community Journal*, **6**, e105. doi:10.24072/pcjournal.814.

Jöreskog KG (1994). "On the Estimation of Polychoric Correlations and Their Asymptotic Covariance Matrix." *Psychometrika*, **59**(3), 381–389. doi:10.1007/BF02296131.

Klein N, Hothorn T, Barbanti L, Kneib T (2022). "Multivariate Conditional Transformation Models." *Scandinavian Journal of Statistics*, **49**, 116–142. doi:10.1111/sjos.12501.

Lesaffre E, Kaufmann H (1992). "Existence and Uniqueness of the Maximum Likelihood Estimator for a Multivariate Probit Model." *Journal of the American Statistical Association*, **87**(419), 805–811. doi:10.2307/2290218.

Masarotto G, Varin C (2012). "Gaussian Copula Marginal Regression." *Electronic Journal of Statistics*, **6**, 1517–1549. doi:10.1214/12-EJS721.

Sewak A, Inácio V, Wuu J, Benatar M, Hothorn T (2026a). "Likelihood-based Modeling of Covariate-Specific Time-Dependent ROC Curves." *Statistical Methods in Medical Research*, **35**(6), 1325–1346. doi:10.1177/09622802261445416.

Sewak A, Siegfried S, Hothorn T (2026b). "Transformation Discriminant Analysis for Constructing Optimal Biomarker Combinations." *Statistics in Medicine*, **45**(15–17), e70665. doi:10.1002/sim.70665.

Song PXK, Li M, Yuan Y (2009). "Joint Regression Analysis of Correlated Data Using Gaussian Copulas." *Biometrics*, **65**(1), 60–68. doi:10.1111/j.1541-0420.2008.01058.x.