flipscores test: Robust Inference in GLMs

Livio Finos
Dept of Statistical Sciences & Padova Neuroscience Center
University of Padova

BEYOND classical p-values
Statistical and AI methods for better inferences in Psychology
Rovereto, 26-31 May, 2025

Outline

Let's start from the end

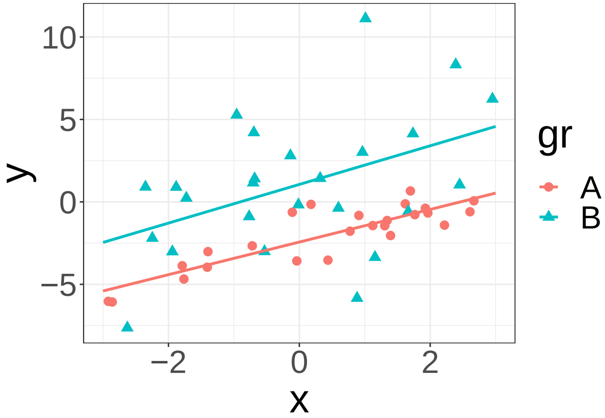
What I'm selling today:

- Robust inference on coefficients of Generalized Linear Models (GLM)
- No needs to worry about heteroscedasticity or variance mispecification
- An handy `R` package: `flipscores`

Let's see a couple of examples...

case 1: LM with heteroscedasticity?

Is `lm( y ~ x * gr, data = D)` OK?



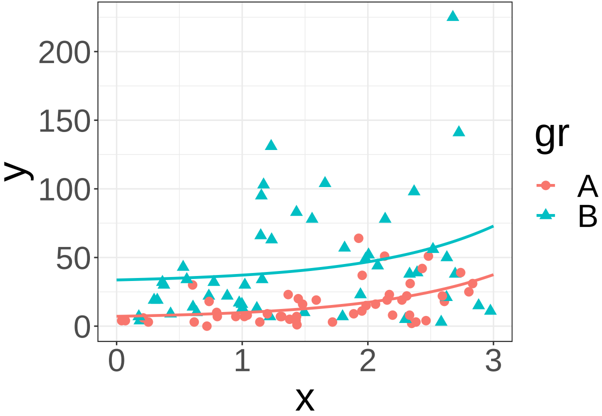
case 1: LM with heteroscedasticity?

```
Call:
flipscores(formula = y ~ x * gr, data = D)

 Coefficients:
             Estimate   Score  Std.Error  z value Part.Cor Pr(>|z|)
 (Intercept)  -0.688 -33.207   19.762  -1.680    -0.245   0.0946 .
 x             1.080 137.974   37.103   3.718     0.542   0.0010 ***
 gr1          -1.746 -84.262   22.763  -3.701    -0.540   0.0002 ***
 x:gr1        -0.091 -11.675   31.217  -0.374    -0.055   0.7116
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

case 2: GLM for count data: Poisson? Quasipoisson? Negative Binomial?

Is `glm( y ~ x * gr, data = D, Family= Poisson)` OK?



case 2: GLM for count data: Poisson? Quasipoisson? Negative Binomial?

```
Call:
flipscores(formula = y ~ x * gr, family = poisson, data = D)

Coefficients:
            Estimate    Score Std.Error   z value Part. Cor Pr(>|z|)
(Intercept)    2.423  317.066     7.422    42.716     0.464   0.0002 ***
x              0.496  656.862    38.178    17.205     0.340   0.0002 ***
gr1           -0.791 -280.629    20.805   -13.488    -0.281   0.0006 ***
x:gr1          0.138  165.308    35.349     4.676     0.102   0.2040
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Permutation tests

Usually

- require less assumptions
- exact control of Type I Error, even for small sample size
- converge to parametric counterpart
  (i.e. asymptotically same power)
- multivariate (multiplicity correction): easy and powerful

BUT the assumption of exchangeability of the observations (under the null hypothesis)

- makes hard to deal with counfounders
- doesn't allow to deal the case of hereroscedasticity

Permutation tests on GLM WithOut Confounders

Example: Poisson model ($g() = log()$ link function)

$$log(E(y_i)) = g(\mu) = \eta = \gamma_0 + \beta x_i, \;\; i = 1, \ldots, n$$

Tested Hypothesis: $H_0 : \;\beta = 0$
under H_0 the model reduces to

$$log(E(y_i)) = g(\mu) = \eta = \gamma_0, \;\; i = 1, \ldots, n$$

That is, observations are exchangeable, therefore

- Compute $S^{obs} = \hat{\beta}(y_i)$ (or any test stat)
- Permute y_i to get y_i^{*b} and compute $S^{*b} = \hat{\beta}(y_i^{*b})$
- repeat the step above (large) B times to the null distribution.
- p-value $= \frac{\#(|S^{*b}| \geq |S^{obs}|)}{B+1}$

A major limitation: GLM with Confounders

Example: Poisson model ($g() = log()$ link function)

$$log(E(y_i)) = g(\mu_i) = \eta_i = \gamma_0 + \gamma_1 z_i + \beta x_i, \;\; i = 1, \ldots, n$$

$$H_0 : \; \beta = 0 \;\; \forall \, \gamma = (\gamma_0, \gamma_1)$$

When $H_0 : \; \beta = 0$ is true:

$$log(E(y_i)) = g(\mu_i) = \eta_i = \gamma_0 + \gamma_1 z_i, \;\; i = 1, \ldots, n$$

- Therefore observations are not exchangeable: $E(y_i) = e^{\gamma_0 + \gamma_1 z_i}$
- (outside the Linear Model) not only the mean, but even the variance is not equal among obs: e.g. $V(y_i) = e^{\gamma_0 + \gamma_1 z_i}$

Outline

The parametric score test

$$\begin{aligned} Score &= \left.\frac{\partial \ell(\beta|x,z,y)}{\partial \beta}\right|_{\beta=0} \\ &= \sum_{i=1}^{n} \frac{\partial}{\partial \beta} \log f_{\beta,\gamma,X_i}(Y_i)\Big|_{\beta=0,\gamma=\hat{\gamma}} = \\ &= \sum_{i=1}^{n} \nu_i \underset{H_0}{\dot\sim} N(0,\mathcal{I}) \ \ \text{(asymptotically, under } H_0) \\ & \quad \mathcal{I}\text{: Fisher Information Matrix} \end{aligned}$$

$\nu_i \sim NOTnormal(0, var(\nu_i))$ + Centr.Lim.Thm =
Approximated Type I Error control, but very good in practice.

Effective Score

$$Effective\ Score = \sum_{i}^{n} (\nu_i - \hat{\mathcal{I}}'_{XZ} \hat{\mathcal{I}}^{-1}_{ZZ} \boldsymbol{\nu}^{(k-1)}_{\hat{\gamma},i})$$

with $\boldsymbol{\nu}^{(k-1)}_{\hat{\gamma},i} = \frac{\partial}{\partial\gamma} \log f_{\beta,\gamma,X_i}(Y_i)\Big|_{\beta=0,\gamma=\hat{\gamma}} \in \mathbb{R}^{k-1},\ 1 \leq i \leq n$

and $\hat{\mathcal{I}}$ the Observed Fisher Information under H_0:

$$\hat{\mathcal{I}} = \begin{bmatrix} \hat{\mathcal{I}}_{XX} & \hat{\mathcal{I}}'_{XZ} \\ \hat{\mathcal{I}}_{XZ} & \hat{\mathcal{I}}_{ZZ}. \end{bmatrix}$$

Since $\sum_i^n \hat{\mathcal{I}}'_{XZ} \hat{\mathcal{I}}^{-1}_{ZZ} \boldsymbol{\nu}^{(k-1)}_{\hat{\gamma},i} = 0 \ \Rightarrow \ Effective\ Score = Score$

(Also named Efficient Score in Cox and Hinkley, 1979)

Effective Score in GLM

In GLM the effective score takes the following form:

$$
\begin{aligned}
S &= \sum_{i}^{n} (\nu_i - \hat{\mathcal{I}}'_{XZ} \hat{\mathcal{I}}_{ZZ}^{-1} \boldsymbol{\nu}_{\hat{\gamma},i}^{(k-1)}) = \\
&= \sum_{i=1}^{n} \left(x_i - X^\top WZ(Z^\top WZ)^{-1} z_i \right) \frac{(y_i - \hat{\mu}_i) d_i}{v_i} = \\
&= X^T W^{1/2} (I - H) V^{-1/2} (y - \hat{\mu})
\end{aligned}
$$

where $H = W^{1/2} Z (Z^T WZ)^{-1} Z^T W^{1/2}$;
$W = DV^{-1}D$, $D = \mathrm{diag}\left\{\frac{\partial \mu_i}{\partial \eta_i}\right\}$, $V = \mathrm{diag}\{\mathrm{Var}(y_i)\}$.

An intuition: Effective score as a Partial Correlation

In Linear Models, a special case of GLM:

$$Y = Z\gamma + X\beta + \varepsilon$$

the Effective Score is a Double Residualization
(dropping some constant, e.g. the common variance of errors ε):
hat matrix simplifies to $H = Z(Z^T Z)^{-1} Z^T$ and $\hat{\mu} = Hy$ so that:

$$\begin{aligned} S &= X^T (I - H)\,(y - \hat{\mu}) \\ &= X^T (I - H)\,(I - H) y \\ &= \sum_{i=1}^{n} (x_i - \hat{x}_i)\;(y_i - \hat{y}_i) \end{aligned}$$

the partial correlation can be written as

$$\frac{X^T (I - H)\;(I - H) y}{\|(I - H) X\| \|(I - H) y\|} = \frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)\;(y_i - \hat{y}_i)}{\sqrt{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}}$$

Generalized Partial Correlation in GLM

$$\rho = \frac{X^T W^{1/2}(I - H)\, V^{-1/2}(y - \hat{\mu})}{\|(I - H)W^{1/2}X\| \|V^{-1/2}(y - \hat{\mu})\|}$$

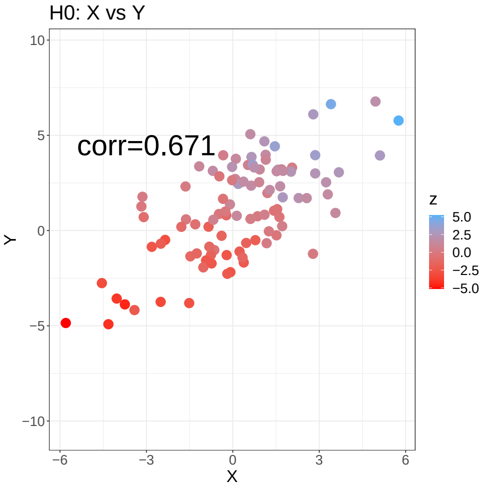
- It becomes the well known partial correlation coefficient in LM
- $-1 \leq \rho \leq 1$
- Easily extended to (Generalized) Determination Coefficient R^2 for multiple X.

Finos and Girardi, 2025(?)

A toy example, $\beta = 0$

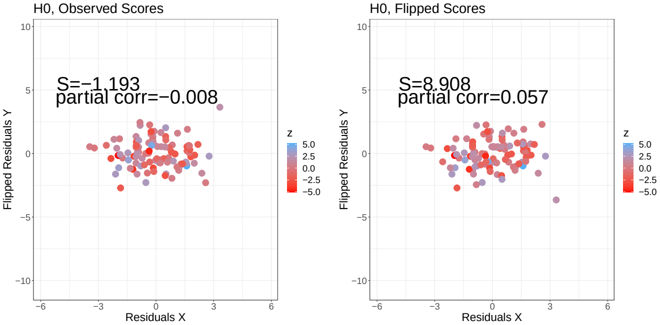
$$Y = Z\gamma + X\beta + \varepsilon$$

$$cor(X, Z) = 0.80, \;\; \gamma = 1, \beta = 0, \;\; \varepsilon \sim N(0, 1)$$



A toy example, $\beta = 0$

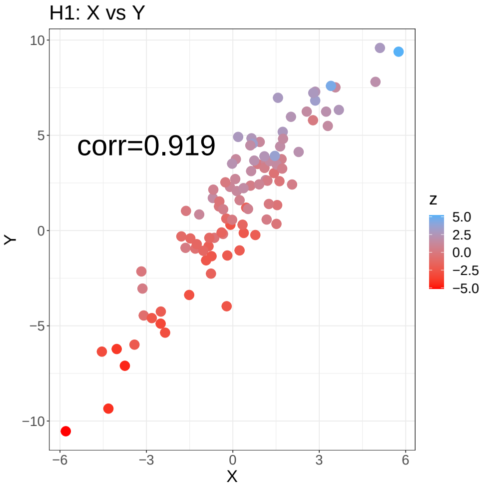
$$\sum_{i=1}^{n}(x_i - \hat{x}_i)(y_i - \hat{y}_i) \text{ Vs } \sum_{i=1}^{n}(x_i - \hat{x}_i) \pm (y_i - \hat{y}_i)$$



A toy example, $\beta \neq 0$

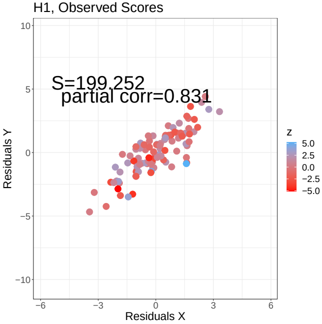
$$Y = Z\gamma + X\beta + \varepsilon$$

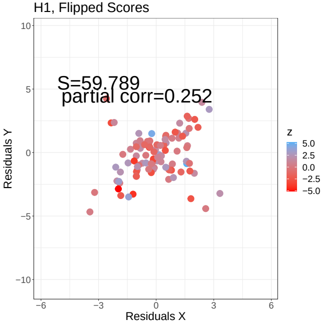
$$cor(X, Z) = 0.80,\ \gamma = 1,\ \beta = 1,\ \varepsilon \sim N(0, 1)$$



A toy example, $\beta \neq 0$

$$\sum_{i=1}^{n} (x_i - \hat{x}_i)(y_i - \hat{y}_i) \text{ Vs } \sum_{i=1}^{n} (x_i - \hat{x}_i) \pm (y_i - \hat{y}_i)$$





Effective Flip Score test

$$S^* = \sum_{i=1}^{n} \pm(\nu_i - \hat{\mathcal{I}}'_{XZ}\hat{\mathcal{I}}^{-1}_{ZZ}\boldsymbol{\nu}^{(k-1)}_{\hat{\gamma},i})$$

Algorithm

- Compute $S^{obs} = S$
- Compute S^{*b}
- repeat the step above (large) B times to the null distribution.
- p-value $= \frac{\#|S^{*b}| \geq |S^{obs}|}{B+1}$

REMARK:

- $\sum_i^n +\hat{\mathcal{I}}'_{XZ}\hat{\mathcal{I}}^{-1}_{ZZ}\boldsymbol{\nu}^{(k-1)}_{\hat{\gamma},i} = 0$ BUT
- $\sum_i^n \pm\hat{\mathcal{I}}'_{XZ}\hat{\mathcal{I}}^{-1}_{ZZ}\boldsymbol{\nu}^{(k-1)}_{\hat{\gamma},i} \neq 0$ (in general)
  this is why we need Effective Score

Properties

Asymptotically Exact

- 0-mean: $E(\nu_i) = 0 \; AND \; E(\boldsymbol{\nu}_{\hat{\gamma},i}^{(k-1)}) = 0 \Rightarrow E(\pm\nu_i) = 0$
- $y_i \perp\!\!\!\perp y_j \;\; i,j = 1,\ldots,n$
- constant variance (Asymptotically):
  $V(+(\nu_i - \hat{\mathcal{I}}'_{XZ}\hat{\mathcal{I}}_{ZZ}^{-1}\boldsymbol{\nu}_{\hat{\gamma},i}^{(k-1)})) = V(-(\nu_i - \hat{\mathcal{I}}'_{XZ}\hat{\mathcal{I}}_{ZZ}^{-1}\boldsymbol{\nu}_{\hat{\gamma},i}^{(k-1)})) \Rightarrow$
  $V(S^*) = V(S^{obs})$

Properties

It converges to parametric score test (i.e. asymptotically):

i. is normal $N(0, \mathcal{I})$

ii. is exact

iii. is locally most powerful (LMP)

iv. if the parametric S test is UMP (UMPU),
S^* is asymptotically UMP (UMPU)

Effective Flip Score is Asymptotically Exact

- We need: $\nu_i \perp\!\!\!\perp \nu_{i'}$.
- However, when we plug $\hat{\gamma}$ into $S_{\hat{\gamma}}$, the $\nu_{\hat{\gamma},i}$ become dependent (e.g. in linear model the effective d.f. are $n - \mathrm{rank}(Z)$).
- The correlation disappears when $n \to \infty$

We get **asymptotically exact** test (with any variance estimates).

Effective Flip Score is Asymptotically Exact

- We need: $\nu_i \perp\!\!\!\perp \nu_{i'}$.
- However, when we plug $\hat{\gamma}$ into $S_{\hat{\gamma}}$, the $\nu_{\hat{\gamma},i}$ become dependent (e.g. in linear model the effective d.f. are $n - \mathrm{rank}(Z)$).
- The correlation disappears when $n \to \infty$

We get **asymptotically exact** test (with any variance estimates).

We propose the **Standardized** Flip Score for small (and large) sample size!

Outline

Conclusion

Standardized Flip Score Test

- for GLM
- gooood control of the type I error

Effective Flip Score Test:

- general: any score statistics
- Asymtotically exact

Both

- As we will see: General Approach, many extensions
- assume the link function to be right ($E(\hat{\mu}) = \mu$)
- very **robust** (heteroscedasticity, overdispersion etc)
- quite **fast** to compute
- R package: `https://github.com/livioivil/flipscores`
- some relationship with solution of F. Pesarin (2001) to Behrens-Fisher problem

References

Jesse Hemerik, Jelle J Goeman, and Livio Finos (2020). Robust testing in generalized linear models by sign flipping score contributions. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.3, pp. 841–864.

Riccardo De Santis et al. (2025) Inference in generalized linear models with robustness to misspecified variances. *Journal of the American Statistical Association* just-accepted , pp. 1–16.