Sample size calculation and optimal design for regression-based test norming

Francesco Innocenti, Frans Tan, Math Candel, & Gerard van Breukelen

Department of Methodology and Statistics
Maastricht University

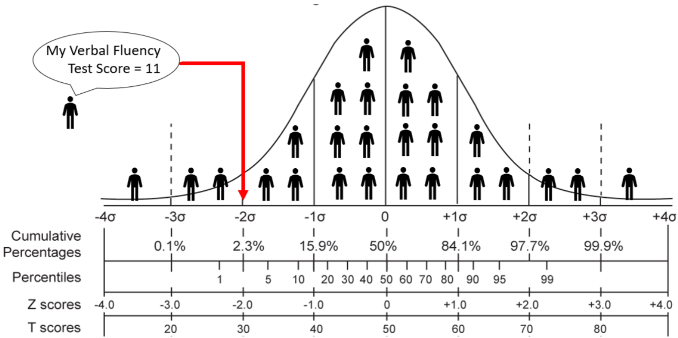
IMPS 2022 - Bologna

Outline

1. Introduction
2. Research question
3. Approaches to test norming
4. Optimal and robust design
5. Sample size calculation
6. Discussion

Normative data

- Norms facilitate the **interpretation of subjects' performance** on a test by directly comparing their scores with those of their peers



- Based on this information, decisions about individuals can be made (e.g. assignment to a treatment or remedial teaching)

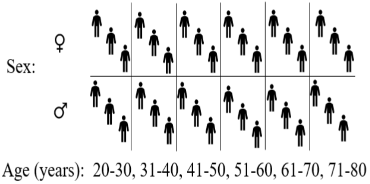
Research question

To prevent mistakes in the assessment of individuals, **norms should be precise**, that is, not being strongly affected by sampling error in the sample on which the norms are based. How to **minimize sampling error** and **maximize precision** of the norms?

1. Adopt an efficient approach to norming ⇒ **regression-based approach**
2. Find a sample composition (e.g. which age groups to include) that maximizes precision of estimation of the norms ⇒ the **optimal design**
3. Take a sufficiently large sample for the normative study ⇒ **sample size calculation formulas**

Traditional approach to norming

1. Split the sample drawn for norming into subgroups based on some relevant demographic factors (e.g. age and sex)
2. Compute the norm statistics of interest (e.g. mean and SD) within each subgroup



- Pros: No model assumptions
- Cons: (1) **Inefficient** (Oosterhuis et al. [2016]), (2) norms are not necessarily based on relevant predictors, (3) categorization of continuous predictors, (4) no optimal design

Maastricht University

Regression-based norming (1/2)

Several approaches (e.g. Lenhard et al. [2018]; Oosterhuis et al. [2016]; Sherwood et al. [2015]; Van Breukelen and Vlaeyen [2005]; Voncken et al. [019a,b]; Zachary and Gorsuch [1985]) but overall

- Pros: (1) **Norms based on the whole sample**, (2) it allows to identify relevant predictors, (3) categorization of continuous predictors not needed, and (4) optimal design
- Cons: The validity of the norms depends on model assumptions

Maastricht University

Regression-based norming (2/2)

Van Breukelen and Vlaeyen [2005]:

1. Sample N subjects from the reference population
2. Fit $\mathbf{y} = \mathbf{X}\beta + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$, thus obtaining $\hat{\beta}$ and $\hat{\sigma}$ from the normative sample

   To compare **a new individual** with the reference population:

3. Compute Z-score: $\hat{Z}_0 = \frac{Y_0 - \hat{Y}_0}{\hat{\sigma}} = \frac{Y_0 - \mathbf{x}_0^T \hat{\beta}}{\hat{\sigma}}$
4. Compute PR-score: $\hat{PR}_0 = \Phi\left(\hat{Z}_0\right) \times 100$

$\mathbf{x}_0$ = individual's scores on the predictors, $\Phi(.)$ = cdf of the standard normal distribution

Sampling variances

- Based on the delta method:

$$V\left(\hat{Z}_0\right) \approx \frac{d\left(\mathbf{X},\xi\right)}{N} + \frac{Z_0^2}{2\left(N-k-1\right)}$$

$$V\left(\hat{PR}_0\right) \approx V\left(\hat{Z}_0\right) \times \left(100 \times \phi\left(Z_0\right)\right)^2$$

where

$$d\left(\mathbf{X},\xi\right) = N\sigma^{-2}V\left(\hat{Y}_0\right) = N\mathbf{x}_0^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}_0$$

is the **standardized prediction variance**, ξ is the design of the normative sample, k = number of predictors, $\phi(.)$ is the pdf of the standard normal distribution, and $\mathbf{x}_0$ is the vector of the new individual's scores on the predictors.

- Simulation study: for $N \geq 300$ relative bias of $V\left(\hat{Z}_0\right) \in (-3\%, +3\%)$; for $N \geq 1600$ relative bias of $V\left(\hat{PR}_0\right) \in (-5\%, +5\%)$

Optimal design: Theory

- Design ξ = joint distribution of the predictors in the normative sample given N, e.g. sex distribution and age distribution per sex level in the sample
- **Optimal design** ξ^* = the joint distribution of the predictors in the normative sample that **minimizes** $V\left(\hat{Z}_0\right)$ and $V\left(\hat{PR}_0\right)$ given N
- But $V\left(\hat{Z}_0\right)$ and $V\left(\hat{PR}_0\right)$ depend on ξ only through

$$d\left(\mathbf{X}, \xi\right) = N\mathbf{x}_0^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}_0$$

  so a safe approach to find ξ^* is to **minimize the maximum** of $d\left(\mathbf{X}, \xi\right)$ over $\mathbf{x}_0$ (G-optimality, see e.g. Schwabe [1996])
- This is a safe approach because it **minimizes the maximum** of $V\left(\hat{Z}_0\right)$ and $V\left(\hat{PR}_0\right)$ **over all possible combinations of the predictors** $\left(\mathbf{x}_0\right)$

Maastricht University

Optimal design: Results

Let $\epsilon_i \sim N\left(0, \sigma^2\right)$

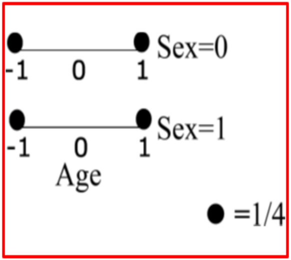
$$Y_i = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \epsilon_i \quad (1)$$

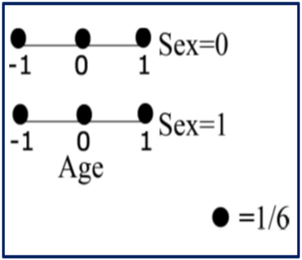
$$Y_i = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 Age_i^2 + \epsilon_i \quad (2)$$

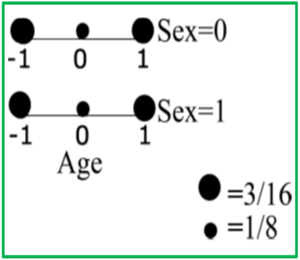
$$Y_i = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_4 Age_i Sex_i + \epsilon_i \quad (3)$$

$$Y_i = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 Age_i^2 + \beta_4 Age_i Sex_i + \epsilon_i \quad (4)$$

$$Y_i = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 Age_i^2 + \beta_4 Age_i Sex_i + \beta_5 Age_i^2 Sex_i + \epsilon_i \quad (5)$$

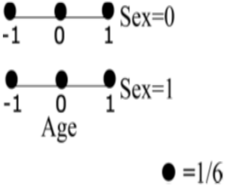






Maximin design

- The optimal design ξ^* depends on the assumed model, but at the design phase there is **uncertainty about the "true" model** (i.e. best fitting polynomial)
- A solution: Find the **most robust** design against misspecification of the model. Two alternative criteria:
  - **Relative Efficiency** (RE): ratio of $V\left(\hat{Z}_0\right)$ or $V\left(\hat{PR}_0\right)$ under ξ^* to $V\left(\hat{Z}_0\right)$ or $V\left(\hat{PR}_0\right)$ under ξ, given N ⇒ **RE maximin design** $=$ highest minimum relative efficiency across all plausible models
  - **Efficiency**: $\left(V\left(\hat{Z}_0\right)\right)^{-1}$ or $\left(V\left(\hat{PR}_0\right)\right)^{-1}$ ⇒ **Absolute maximin design** $=$ highest minimum efficiency across all plausible models



Sample size calculation

- Sample size calculation formulas for the **optimal or maximin design** and for a subject with scores on the predictors $(\mathbf{x}_0)$ such that $V\left(\hat{Z}_0\right)$ and $V\left(\hat{PR}_0\right)$ are maximum
- In practice, norms are used **to classify subjects' performance** on a test $(Z_t$ or $PR_t)$ relative to a chosen cut-off point $(Z_c$ or $PR_c)$ as "average" versus "below" or "above average", to make decisions
- This classification problem can be expressed as H_0 : "average" performance $Z_t = Z_c$ versus H_1 : "below average" performance $Z_t < Z_c$:

  N^* = **to detect the smallest clinically relevant difference** between subject's norm value and the cut-off point for decision making, **given a pre-specified Type I error rate and statistical power**

Maastricht University

Sample size calculation: Power

1. Choose: (i) the model for norming with k predictors, (ii) the cut-off point for decision making (Z_c or PR_c) , (iii) the smallest clinically relevant difference δ between subject's norm value (Z_t or PR_t) and the cut-off point, (iv) the Type I error rate α and statistical power $1-\beta$
2. For Z-scores, the required sample size is

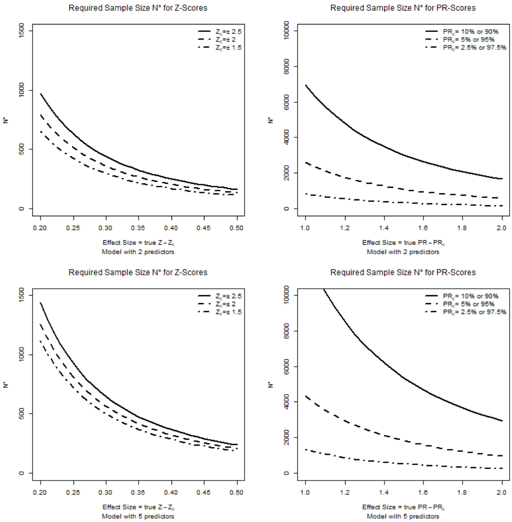
$$N^* = \left[\frac{z_{1-\alpha}\left(k+1+\frac{Z_c^2}{2}\right)^{1/2} + z_{1-\beta}\left(k+1+\frac{Z_t^2}{2}\right)^{1/2}}{\delta} \right]^2$$

For PR-scores, the required sample size is

$$N^* = \left[\frac{z_{1-\alpha}\cdot 100\cdot\phi\left(Z_{PR_c}\right)\left(k+1+\frac{Z_{PR_c}^2}{2}\right)^{1/2} + z_{1-\beta}\cdot 100\cdot\phi\left(Z_{PR_t}\right)\left(k+1+\frac{Z_{PR_t}^2}{2}\right)^{1/2}}{\delta} \right]^2$$

Sample size calculation: Results

Type I error rate $= 5\%$, Power $= 80\%$



Sample size calculation: Precision

Alternative approach: N^* = **half the confidence interval width equals the pre-specified margin of error**

1. Choose: (i) the model for norming with k predictors, (ii) the Z-score or PR-score of interest (e.g. $Z_0 = -2$ or $PR_0 = 5\%$) , (iii) the desired margin of error τ, (iv) the confidence level $1 - \alpha$
2. For Z-scores, the required sample size is

$$N^* = \left\lceil \frac{z_{1-\alpha/2} \left(k + 1 + \frac{Z_0^2}{2} \right)^{1/2}}{\tau} \right\rceil^2$$

For PR-scores, the required sample size is

$$N^* = \left\lceil \frac{z_{1-\alpha/2} \cdot 100 \cdot \phi\left(Z_0\right) \cdot \left(k + 1 + \frac{Z_0^2}{2} \right)^{1/2}}{\tau} \right\rceil^2$$

Conclusion

To **maximize** the **precision** of norms:

- A **regression-based approach** is recommended because **more efficient** than the traditional approach
- The **sample composition** should be as prescribed by the **optimal design**, if prior knowledge about best fitting polynomial is available
- If there is **uncertainty about the model**, **efficient robust designs** can be used instead of the optimal design
- Two approaches to determine the **required sample size** for the optimal/maximin design

Future research

Sample size calculation and optimal design for

- Multivariate regression-based norming (Van der Elst et al. [2017])
- Non-normality and heteroscedasticity: GAMLSS (Timmerman et al. [2021]; Voncken et al. [019a,b]), quantile regression (Sherwood et al. [2015]), cNORM (Lenhard et al. [2018]), Flexible discrete norming (Van der Ark et al., [2022])

References

Innocenti, F., Tan, F., Candel, M., and Van Breukelen, G. (2021). Sample size calculation and optimal design for regression-based norming of tests and questionnaires. *Psychological Methods*, Advance online publication.

Lenhard, A., Lenhard, W., Suggate, S., and Segerer, R. (2018). A continuous solution to the norming problem. *Assessment*, 25:112–125.

Oosterhuis, H., Van der Ark, L., and Sijtsma, K. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, 23:191–202.

Schwabe, R. (1996). *Optimum designs for multi-factor models*. Springer-Verlag, New York.

Sherwood, B., Zhou, A., Weintraub, S., and Wang, R. (2015). Using quantile regression to create baseline norms for neuropsychological tests. *Alzheimer's & dementia*, 2:12–18.

Timmerman, M., Voncken, L., and Albers, C. (2021). A tutorial on regression-based norming of psychological tests with gamlss. *Psychological methods*, 26:357–373.

Van Breukelen, G. and Vlaeyen, J. (2005). Norming clinical questionnaires with multiple regression: The pain cognition list. *Psychological Assessment*, 17:336–344.

Van der Elst, W., Molenberghs, G., Van Tetering, M., and Jolles, J. (2017). Establishing normative data for multi-trial memory tests: the multivariate regression-based approach. *The Clinical Neuropsychologist*, 31:1173–1187.

Voncken, L., Albers, C., and Timmerman, M. (2019a). Improving confidence intervals for normed test scores: Include uncertainty due to sampling variability. *Behavior Research Methods*, 51:826–839.

Voncken, L., Albers, C., and Timmerman, M. (2019b). Model selection in continuous test norming with gamlss. *Assessment*, 26:1329–1346.

Zachary, R. and Gorsuch, R. (1985). Continuous norming: Implications for the wais-r. *Journal of Clinical Psychology*, 41:86–94.

Thank you for your attention!

francesco.innocenti@maastrichtuniversity.nl