

Informative hypotheses evaluation

Rebecca M. Kuiper

Department of Methodology & Statistics
Utrecht University

ANOVA Example: Comparisons of 3 Means

Example Palmer and Gough

Palmer and Gough (2007) examined the difference in the attribution of importance of defective education as an explanation for criminal behavior between three types of “offenders”:

(1) non-offenders, (2) property offenders, and (3) person offenders.

ANOVA

$$y_j = \mu_1 d_{1j} + \mu_2 d_{2j} + \mu_3 d_{3j} + \epsilon_j,$$

where μ_i ($i = 1, 2, 3$) is the mean in group i ,
 d_{ij} is a dummy variable denoting group membership,
 ϵ_j is an error term and $\epsilon_j \sim N(0, \sigma^2)$.

$(n_i = 20, 20, 31)$

NHST

Example Palmer and Gough

```
PandG_data <- read.table("Data_PalmerAndGough.txt",
                          header=TRUE)
PandG_data$group <- factor(PandG_data$group)
pairwise.t.test(PandG_data$Importance, PandG_data$group,
                p.adj = 'bonferroni')

##
## Pairwise comparisons using t tests with pooled SD
##
## data: PandG_data$Importance and PandG_data$group
##
##      1      2
## 2 0.191 -
## 3 0.011 1.000
##
## P value adjustment method: bonferroni
```


Confirmatory methods

Most researchers are able to specify “order-restricted” / “informative” / “theory-based” hypotheses, like $H_1 : \mu_1 > \mu_2 > \mu_3$.
Use prior knowledge and/or expertise in hypothesis.

Methods to evaluate theory-based hypotheses

- Hypothesis testing: F_{bar} (\bar{F}) test
(renders p-value and can test only one theory-based hypothesis)
- Confirmatory model selection using information criteria:
GORIC and GORICA
- (Confirmatory) Bayesian model selection (BMS)

Note: 'model' refers to hypothesis.

GORIC

Example Palmer and Gough

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

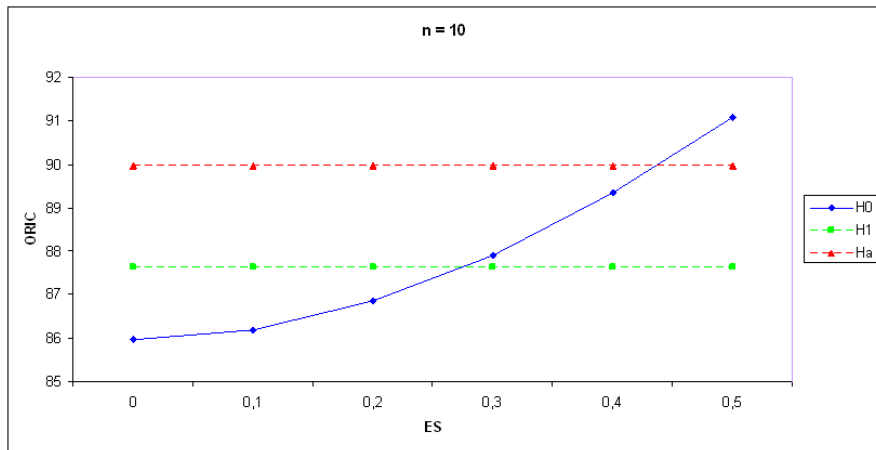
$$H_u : \quad \mu_1, \mu_2, \mu_3.$$

GORIC

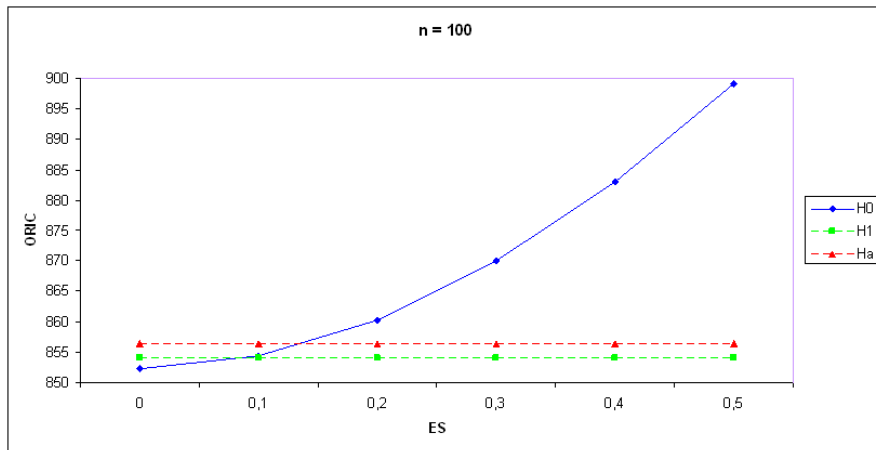
Model	Fit	Complexity	GORIC	GORIC weights
H_0	-196.36	2.00	396.71	0.02
H_1	-191.89	2.81	389.41	0.75
H_u	-191.89	4.00	391.79	0.23

Confirmatory methods have more “power” than their exploratory counterparts (cf. one-sided vs two-sided testing).

Confirmation more power: 1 data set. GORIC values for 3 groups, effect size ES , and $n = 10$ observations per group



Confirmation more power: 1 data set. GORIC values for 3 groups, effect size ES , and $n = 100$ observations per group



Confirmatory methods - some of my references

- Kuiper, R. M., and Hoijtink, H. (2010). Comparisons of Means Using Exploratory and Confirmatory Approaches. *Psychological Methods*, 15(1), 69–86.
- Kuiper, R. M., Klugkist, I., and Hoijtink, H. (2010). A Fortran 90 Program for Confirmatory Analysis of Variance. *Journal of Statistical Software*, 34(8), 1–31.
- Kuiper, R.M., Hoijtink, H. and Silvapulle, M.J. (2011). An Akaike type information criterion for model selection under inequality constraints. *Biometrika*, 98, 495-501. (GORIC)
- Kuiper, R.M., Nederhof, T., and Klugkist, I. (2015). Properties of hypothesis testing techniques and (Bayesian) model selection for exploration-based and theory-based (order-restricted) hypotheses. *British Journal of Mathematical and Statistical Psychology*, 68(2), 220 – 245.
- Altınışık, Y., Van Lissa, C. J., Hoijtink, H., Oldehinkel, A. J., and Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalization of the AIC. *Psychological Methods*, 26(5), 599–621.
<https://doi.org/10.1037/met0000406> (GORICA)

Table of Contents

Example Null hypothesis (H_0) vs Informative hypothesis (H_m)

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

Failsafe/Safeguard hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

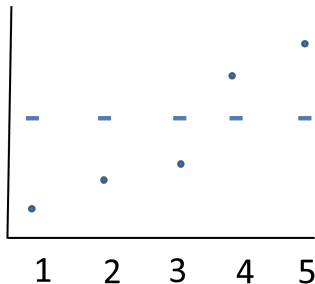
Information criteria (ICs)

IC, like AIC and GORIC, balances fit and complexity.

Describe data as good as possible (fit)
with fewest number of parameters (simplicity / non-complexity).

Intermezzo: Balance Fit and Complexity (2/5)

Example 5 means



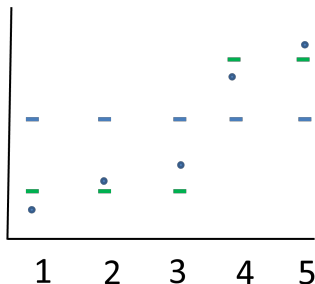
$$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$$

parameters: 1 mean parameter (i.e., μ)

Fit: Bad fit

Intermezzo: Balance Fit and Complexity (3/5)

Example 5 means



$$\mu_1 = \mu_2 = \mu_3 = \mu_{\text{low}}$$

$$\mu_4 = \mu_5 = \mu_{\text{high}}$$

```
# parameters:
```

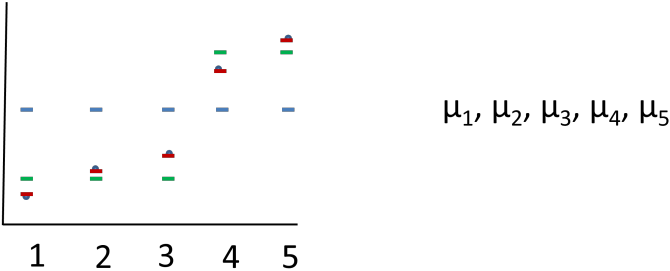
2 mean parameters (i.e., μ_{low} and μ_{high})

Fit:

Better than with one parameter, even good fit.

Intermezzo: Balance Fit and Complexity (4/5)

Example 5 means



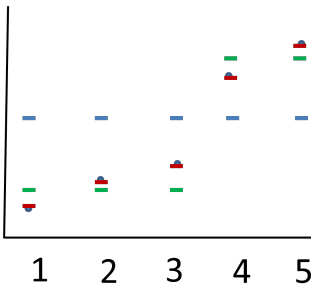
parameters: 5 mean parameter (complex)
Fit: Better than with two parameters, even best possible fit.

Thus: Best fit (= highest likelihood), but also most complex (= highest penalty).

In this example: Two means may be best trade-off between fit & complexity.

Intermezzo: Balance Fit and Complexity (5/5)

From “Example 5 means” to ‘Example 2 means’



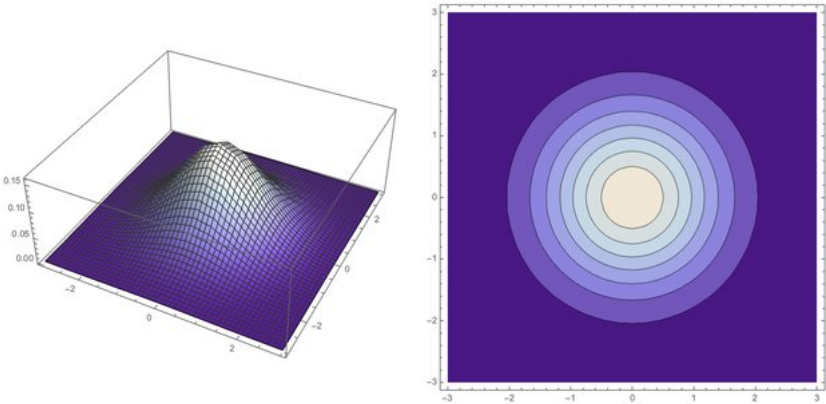
Now, we also incorporate order-restrictions (e.g., $\mu_1 < \mu_2$).

Then, helpful to look at likelihood (not scatter plot).

In presentation, I will use contour plots (in case of two means, not five).

Intermezzo: Contour plot (1/2)

Example 2 means

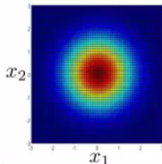
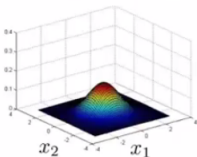


Intermezzo: Contour plot (2/2)

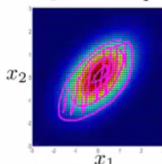
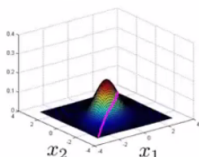
Example 2 means

Multivariate Gaussian (Normal) examples

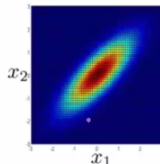
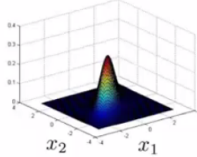
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

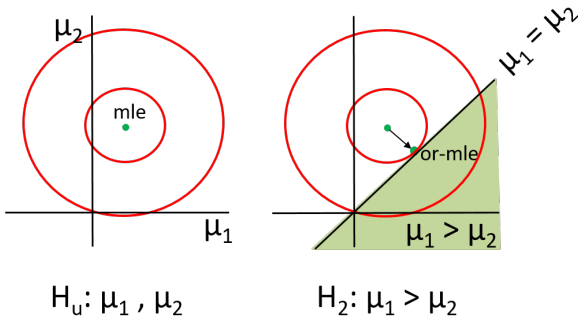


$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$



Idea fit: order-restricted maximum likelihood (or-ml)

Example 2 means



mle not in H_2 : Find highest likelihood in allowable (= green) space.
The resulting estimated mean $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2)$ is referred to as or-ml.

Idea fit

Example Palmer and Gough, that is, 3 means

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

$$H_2 : \mu_1 > \mu_2 < \mu_3,$$

$$H_3 : \mu_1 < \mu_2 < \mu_3,$$

$$H_u : \mu_1, \mu_2, \mu_3.$$

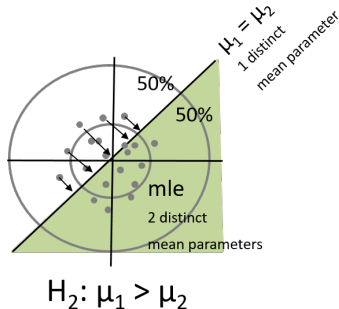
```
## group1 group2 group3
## 11.95  9.75  8.77
```

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_0	-196.36	2.00	396.71	0.01
H_1	-191.89	2.81	389.41	0.56
H_2	-192.34	3.19	391.05	0.25
H_3	-196.36	2.81	398.34	0.01
H_u	-191.89	4.00	391.79	0.17

Idea complexity: penalty (PT)

Example 2 means



Note: sampling from null distribution: $H_0: \mu_1 = \mu_2 = 0$.

complexity $H_2 = PT_2 = 1 + 0.5 \times 1 + 0.5 \times 2 = 2.5$.

PT = number of expected distinct parameters (under H_0).

Idea complexity

Example Palmer and Gough, that is, 3 means

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

$$H_2 : \mu_1 > \mu_2 < \mu_3,$$

$$H_3 : \mu_1 < \mu_2 < \mu_3,$$

$$H_u : \mu_1, \mu_2, \mu_3.$$

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_0	-196.36	2.00	396.71	0.01
H_1	-191.89	2.81	389.41	0.56
H_2	-192.34	3.19	391.05	0.25
H_3	-196.36	2.81	398.34	0.01
H_u	-191.89	4.00	391.79	0.17

GORIC: Lowest value is best

GORIC is like AIC expected distance from the truth (KL-distance). Hence, smallest value is best.

Or: One wants the smallest 'misfit + complexity'.
Note: $-2 \text{ fit} = \text{misfit}$.

GORIC

Example Palmer and Gough

$$H_0 : \quad \mu_1 = \mu_2 = \mu_3,$$

$$H_1 : \quad \mu_1 > \mu_2 > \mu_3,$$

$$H_u : \quad \mu_1, \mu_2, \mu_3.$$

GORIC

Model	Fit	Complexity	GORIC
H_0	-196.36	2.00	396.71
H_1	-191.89	2.81	389.41
H_u	-191.89	4.00	391.79

Hands-on/Demo (1a): GORIC

Let's practice.

- Go to <https://github.com/rebeccakuiper/Tutorials>:
 1. Click on green button called Code.
 2. Download zip (last option in list).
 3. Unzip it on your machine (that folder is now your working directory).
- Start Rstudio. Optional: make project.
- Open 'Tutorial_GORIC_restriktor_ANOVA.html', 'PalmerAndGough_and_Lucas.R', and/or '**Hands-on_1_GORIC_Unc_ANOVA_restriktor.R**' ('Hands-on files').
- Install packages and load them.
- Read and inspect data.
Use **Data_PalmerAndGough.txt** and/or Data_Lucas.txt.
- Run model (lm()).
- Specify hypotheses (make up your own).
Note: Use names used in the model.
- Run goric().
- Inspect and interpret output.

Note: 'GORIC weights' will be explained next.

Table of Contents

Example Null hypothesis (H_0) vs Informative hypothesis (H_m)

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

Failsafe/Safeguard hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

Illustration of the GORIC weights (w_m)

Example Palmer and Gough

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$

$$H_1: \mu_1 > \mu_2 > \mu_3,$$

$$H_u : \quad \mu_1, \mu_2, \mu_3.$$

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_0	-196.36	2.00	396.71	0.02
H_1	-191.89	2.81	389.41	0.75
H_u	-191.89	4.00	391.79	0.23

Table of Contents

Example Null hypothesis (H_0) vs Informative hypothesis (H_m)

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

Failsafe/Safeguard hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

R code: Hypotheses

Palmer & Gough

$$H_1 : \mu_1 > \mu_2 > \mu_3$$

$$H_2 : \mu_1 > \mu_2 < \mu_3$$

$$H_u : \mu_1, \mu_2, \mu_3$$

```
# Hypotheses
```

```
H1 <- 'group1 > group2 > group3'
```

```
H2 <- 'group1 > group2 < group3'
```

R code: GORIC

Palmer & Gough

$$H_1 : \mu_1 > \mu_2 > \mu_3$$

$$H_2: \mu_1 > \mu_2 < \mu_3$$

$$H_u : \quad \mu_1, \mu_2, \mu_3$$

```
# GORIC (using goric function in restriktor package)
library(restriktor)
set.seed(123) # Set seed value
goric.PandG <- goric(fit.PandG,
                     hypotheses = list(H1=H1, H2=H2))
#goric.PandG
#goric.PandG$ratio.qw
```

GORIC output and interpretation

Palmer & Gough

$$H_1 : \mu_1 > \mu_2 > \mu_3$$

$$H_2 : \mu_1 > \mu_2 < \mu_3$$

$$H_u : \mu_1, \mu_2, \mu_3$$

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-191.89	2.81	389.41	0.57
H_2	-192.34	3.19	391.05	0.25
H_u	-191.89	4.00	391.79	0.17

H_1 and H_2 are not weak (nl, better than H_u).

H_1 is $.57/.25 \approx 2.3$ times more supported than H_2 .

Another illustration: one-way ANOVA

Lucas: 5 groups

Lucas (2003) investigated difference between female and male leadership w.r.t. influence of the leader.

Five experimental groups:

1. a randomly selected male leader
2. a randomly selected female leader
3. male leader selected via task
4. female leader selected via task
5. female leader selected via task + institutionalized female leadership via movie

(Two informative) hypotheses of interest

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \text{ (include only when of interest)}$$
$$H_1 : \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$$
$$H_2 : \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2$$
 $H_u : \mu_1, \mu_2, \mu_3, \mu_4, \mu_5$ (included as failsafe)

Descriptive statistics of Lucas' Data

Group	Mean Influence	SD	n
1 (male, random)	2.33	1.86	30
2 (female, random)	1.33	1.15	30
3 (male, selected)	3.20	1.79	30
4 (female, selected)	2.23	1.45	30
5 (female, selected+)	3.23	1.50	30

R code: Hypotheses

Lucas

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$$

$$H_1: \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$$

$$H_2: \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2$$

$$H_u: \mu_1, \mu_2, \mu_3, \mu_4, \mu_5$$

```
# Hypotheses
#H0 <- 'group1 = group2 = group3 = group4 = group5'
H1 <- 'group5 = group3 > (group1, group4) > group2'
# Stated otherwise:
#H1 <- 'group5 = group3 > group1 > group2;
#           group3 > group4 > group2'
H2 <- 'group3 > group1 > group4 = group5 > group2'
```

R code: GORIC

Lucas

$$H_1: \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$$

$$H_2: \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2$$

$$H_u : \quad \mu_1, \mu_2, \mu_3, \mu_4, \mu_5$$

```
# GORIC (using goric function in restriktor package)
library(restriktor)
set.seed(123) # Set seed value
goric.lucas <- goric(fit.lucas,
                     hypotheses = list(H1=H1, H2=H2))
#goric.lucas
#goric.lucas$ratio.qw
```

GORIC output

Lucas

$$H_1: \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$$

$$H_2: \mu_3 > \mu_1 > \mu_4 = \mu_5 > \mu_2,$$

$$H_u : \quad \mu_1, \mu_2, \mu_3, \mu_4, \mu_5.$$

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-278.05	3.19	562.48	0.92
H_2	-281.76	3.14	569.79	0.02
H_u	-278.05	6.00	568.10	0.06

GORIC weights (w_m)

Lucas

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-278.05	3.19	562.48	0.92
H_2	-281.76	3.14	569.79	0.02
H_u	-278.05	6.00	568.10	0.06

H_1 is $.92/.02 \approx 39$ times more supported than H_2 .

Thus, there is quite some evidence that H_1 is the best of this set.

Note:

H_1 is $.92/.06 > 1$ times more supported than H_μ .

Hence, H_1 not weak and thus comparable to the other (weak and non-weak) informative hypotheses (as explained next).

Table of Contents

Example Null hypothesis (H0) vs Informative hypothesis (Hm)

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

Failsafe/Safeguard hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

Include “unconstrained” hypothesis

If set of hypotheses does not contain a reasonable/good one:
Select the best of set of weak hypotheses.
E.g.: $w_0 = .2$ and $w_1 = .8$.

Prevent choosing a weak hypothesis

Include unconstrained hypothesis H_u (or H_a):

$$\begin{aligned} H_0 : & \quad \mu_1 = \mu_2 = \mu_3, \\ H_1 : & \quad \mu_1 > \mu_2 > \mu_3, \\ H_u : & \quad \mu_1, \mu_2, \mu_3, \\ & \quad (\text{i.e., no restrictions}). \end{aligned}$$

H_u highest fit but also most complex, thus failsafe/safeguard.

E.g.: $w_0 = .02$, $w_1 = .08$, and $w_\mu = .90$.

What if another hypothesis is true?

What if another informative hypothesis is true?

The failsafe hypothesis will be preferred over the informative hypotheses in the set, if the sample size is large enough.

E.g.: $w_0 = .02$, $w_1 = .08$, and $w_\mu = .90$.

Use of H_u

Palmer & Gough

$$H_0 : \mu_1 = \mu_2 = \mu_3,$$

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

$$H_u : \mu_1, \mu_2, \mu_3.$$

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-191.89	2.81	389.41	0.57
H_2	-192.34	3.19	391.05	0.25
H_u	-191.89	4.00	391.79	0.17

If at least one informative hypothesis not weak ($w_m > w_u$ or $w_m/w_u > 1$), then compare informative hypotheses.

Hence: H_u is only a failsafe, not another hypothesis of interest.

H_1 vs H_u
Palmer & Gough

What if only one informative hypothesis:

$$H_1: \mu_1 > \mu_2 > \mu_3,$$

$$H_u : \quad \mu_1, \mu_2, \mu_3.$$

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-191.89	2.81	389.41	0.77
H_u	-191.89	4.00	391.79	0.23

H_1 is $.77/.23 \approx 3.3$ times more supported than H_u .

BUT: H_{II} includes H_1 .

So, support for H_μ contains support for H_1 .

Alternative failsafe: Complement of H_m

Alternatively (in the case of one hypothesis of interest)

Evaluate hypothesis of interest against its complement; that is, all other possible hypotheses.

More powerful than against the unconstrained
if H_m has maximum fit.

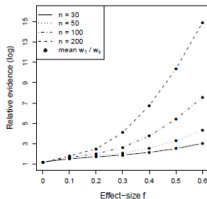
Reference:

Vanbrabant, L., Van Loey, N., and Kuiper, R. M. (2020). Evaluating a Theory-Based Hypothesis Against Its Complement Using an AIC-Type Information Criterion With an Application to Facial Burn Injury. *Psychological Methods*, 25(2), 129-142.
<https://doi.org/10.1037/met0000238>

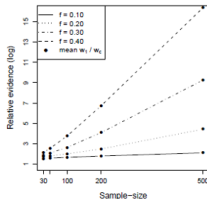
Alternative failsafe: Complement of H_m

Based on simulation

vs complement

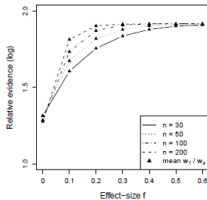


(a)

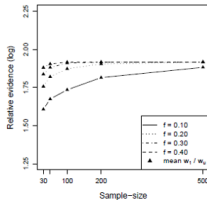


(b)

vs unconstrained



(c)



(d)

R code: complement

Palmer & Gough

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

$$H_c : \quad \text{not } H_1.$$

GORIC

```
H1 <- 'group1 > group2 > group3'
# GORIC (using goric function in restriktor package)
library(restriktor)
set.seed(123) # Set seed value
goric.PandG_C <- goric(fit.PandG,
                        hypotheses = list(H1),
                        comparison = 'complement')
```

H_1 vs H_c
Palmer & Gough

$$\begin{aligned} H_1 : & \quad \mu_1 > \mu_2 > \mu_3, \\ H_c : & \quad \text{not } H_1. \end{aligned}$$

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H1	-191.89	2.81	389.41	0.79
complement	-192.34	3.69	392.05	0.21

H_1 is $.79/.21 \approx 3.8$ times more supported than its complement, that is, any other hypothesis (i.e., any other ordering).

Hands-on/Demo (2): GORIC - H_m vs complement

Let's practice.

- If needed: Start Rstudio again (optional: make project) and then also load packages again.
- Optional: Open 'Tutorial_GORIC_restriktor_ANOVA.html', 'PalmerAndGough_and_Lucas.R', and/or '**Hands-on_2_GORIC_Compl_ANOVA_restriktor.R**'.
- Read and inspect data.
Use Data_PalmerAndGough.txt and/or **Data_Lucas.txt**.
- Run model (lm()).
- Specify hypotheses (make up your own). E.g., for Lucas:

$$H_1 : \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$$

Note: Use names used in the model.

- Run goric(); now, add:
`comparison = "complement"`
- Inspect and interpret output.

Table of Contents

Example Null hypothesis (H_0) vs Informative hypothesis (H_m)

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

Examples with R code

Failsafe/Safeguard hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

GORICA

GORIC: Normal linear models

GORIC can easily be applied to normal linear models (e.g., ANOVA models or regression models).

GORIC: Other statistical models

In case of other statistical models (e.g., a SEM model), more cumbersome to calculate maximized order-restricted log likelihood and thus GORIC.

GORICA: All statistical models

Therefore, GORICA: asymptotic expression for GORIC.
Can be used for all types of statistical models.

Reference:

Altınışık, Y., Van Lissa, C. J., Hoijtink, H., Oldehinkel, A. J., and Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalization of the AIC. *Psychological Methods*, 26(5), 599–621.

<https://doi.org/10.1037/met0000406>

GORICA

Similarities with GORIC

- Form: $GORICA_m = -2 \text{ fit} + 2 \text{ complexity}$.
- Broad type of restrictions.

Differences compared to GORIC

- Uses asymptotic expression of the likelihood (is a normal):
can therefore be easily applied to all types of statistical models.
Disadvantage: might work less well in case of small samples.
- Does not need data set; mle's and their covariance matrix suffice.
- Can leave out nuisance parameters (i.e., not part of hypotheses).

Note

In case of normal linear models and/or not too small samples:
GORICA weights = GORIC weights.

R code: GORICA

Palmer & Gough

$$H_1 : \mu_1 > \mu_2 > \mu_3,$$

$$H_c : \text{not } H_1.$$

GORICA: type = "gorica"

```

H1 <- 'group1 > group2 > group3'
# GORICA (using goric function in restriktor package)
library(restriktor)
set.seed(123) # Set seed value
gorica.PandG_C <- goric(fit.PandG,
                        hypotheses = list(H1),
                        comparison = 'complement',
                        type = "gorica")
    
```


Hands-on/Demo (3): GORICA - H_m vs complement

Let's practice.

- If needed: Start Rstudio again (optional: make project) and then also load packages again.
- Optional: Open 'Tutorial_GORIC_restriktor_ANOVA.html', 'PalmerAndGough_and_Lucas.R', and/or '**Hands-on_3_GORICA_UncAndCompl_ANOVA_restriktor.R**'.
- Read and inspect data.
Use Data_PalmerAndGough.txt and/or **Data_Lucas.txt**.
- Run model (lm()).
- Specify hypotheses (make up your own). E.g., for Lucas:

$$H_1 : \quad \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$$

Note: Use names used in the model (or overwrite those).

- Run goric(); now, add:
`type = "gorica"`
- Inspect and interpret output.

GORIC in JASP: instructions

Lucas

- Open JASP and 'Data_Lucas_JASP.txt'. Check measurement levels.
- Go to ANOVA and denote variables of interest.
- Go to Order Restricted Hypotheses tab.
- Specify hypotheses. For example,

$$H_1: \mu_5 = \mu_3 > \{\mu_1, \mu_4\} > \mu_2$$

JASP : $group5 = group3 > group1 > group2$
 $group3 > group4 > group2$

- Press Ctrl+Enter. Inspect and interpret output.

GORIC and GORICA in JASP

	GORIC R	GORIC JASP	GORICA R	GORICA JASP
ANOVA	✓	✓	✓	
RM-ANOVA			✓	✓
Regression	✓	Not yet	✓	
SEM		Future?	✓	Future?
Other model		Future?	✓	Future?

GORIC and GORICA in JASP

GORIC or GORICA in the following statistical models (as part of that model)

GORIC

- Order Restricted Hypotheses
- Post Hoc Tests
- Descriptives Plots
- Raincloud Plots
- Marginal Means

ANOVA

Cases	Sum of Squares	df	Mean Square	F	p
Note: Type III Sum of Squares					

GORIC

- Order Restricted Hypotheses
- Post Hoc Tests
- Descriptives Plots
- Raincloud Plots
- Marginal Means

ANCOVA

Cases	Sum of Squares	df	Mean Square	F	p
Note: Type III Sum of Squares					

GORICA

- Assumption Checks
- Contrasts
- Order Restricted Hypotheses
- Post Hoc Tests
- Descriptives Plots
- Raincloud Plots
- Marginal Means
- Simple Main Effects

Repeated Measures ANOVA

Cases	Sum of Squares	df	Mean Square	F	p
Note: Type III Sum of Squares					

ANCOVA

- Dependent Variable
- Fixed Factors
- Covariates
- WLS Weights

Display

- Descriptive statistics
- Estimates of effect size
- Work-Save maximum p-value
- Model
- Assumption Checks
- Contrasts
- Order Restricted Hypotheses

The End

GORIC(A)

Thanks for listening!

Are there any questions?

Websites

<https://github.com/rebeccakuiper/Tutorials>
www.uu.nl/staff/RMKuiper/Software
www.uu.nl/staff/RMKuiper/Websites%20%2F%20Shiny%20apps
informative-hypotheses.sites.uu.nl/software/goric/

E-mail

r.m.kuiper@uu.nl

What's next

Depending on time and wishes:

- Some notes
- Demo in R
- Demo in JASP
- Updating hypotheses
- Evidence synthesis / Support aggregation

H0 vs Hm

GORIC

Weights
○○○

Example R

A diagram showing a 3x5 grid of circles. The top row has 4 circles, the middle row has 5 circles, and the bottom row has 5 circles. The circle at the top-right position (row 1, column 5) is missing, leaving a gap.

Failsafe

GORICA
OOOOOO

JASP
○○○○○
○○

End & Extra

Notes

Complement not always higher weight

1. H_m does not have maximum fit (but much lower penalty).
2. Support for H_m is less when evaluating it against its complement (than H_u).

This is of course a good thing, since H_m is not true.

Explanation:

Penalty of H_c is smaller than that of H_u .

Against H_μ , you choose H_m 'sooner' because of low penalty for H_m .

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻ 81/96

Note on possibilities multiple studies

- Update GORIC(A) values & weights.
More data collected: (re-)calculate.
- Update hypotheses.
First data set (or a part of it) generates one or more hypotheses.
Other data set (or part) used to determine evidence / support.
Download 'Tutorial_GORIC_restriktor_UpdateHypo.html' and/or
'Hands-on_4_GORIC_UpdateHypo_restriktor.R' from
<https://github.com/rebeccakuiper/Tutorials>.
- Aggregate evidence for hypotheses.
Aggregate the support for theories (diverse designs allowed).
Bear in mind: Meta-analysis aggregates parameter estimates or effect sizes which need to be comparable (often same designs required).

Download 'Tutorial_GORIC_restriktor_evSyn.html' and/or
'Hands-on_4_GORIC_evSyn_restriktor.R' from
<https://github.com/rebeccakuiper/Tutorials>.

Extra material

- **Logistic Regression Modeling**
 - Article: <https://doi.org/10.1037/met0000406>
- **GORICA on SEM**
 - Article: <https://www.tandfonline.com/doi/full/10.1080/10705511.2020.1836967>.
 - R scripts: https://github.com/rebeccakuiper/GORICA_in_SEM.
- **GORICA on cross-lagged panel model (CLPM)** – Article: <https://doi.org/10.1111/bjep.12455>.
 - R scripts: https://github.com/rebeccakuiper/GORICA_in_SEM.
- **GORICA on Random-Intercept CLPM (RI-CLPM)**
 - Article: Sukpan, C. and Kuiper, R.M. (submitted 2023). How to evaluate causal dominance hypotheses in lagged effects models.
 - R scripts: <https://github.com/Chuenjai/Causal-dominance>.
- **GORICA on CTmeta**
 - Article: <https://doi.org/10.1080/10705511.2020.1823228>.
 - R scripts: https://github.com/rebeccakuiper/GORICA_on_CTmeta.
- **GORICA on Meta-analysis**
 - Article: <https://doi.org/10.3390/e24111525>.
 - R scripts: https://github.com/rebeccakuiper/GORICA_on_MetaAn.

Note: On github site, go to Code (green button) and download zip

Note on BF as an IC

The fit in GORICA refers to the maximum log likelihood.
The fit in BF refers to the maximum likelihood.

Notably, BF can be written as an IC (and vice versa):

$$\begin{aligned}
 -2 \log BF_{1u} &= -2 \log \frac{f_1}{c_1} \\
 &= -2 \log f_1 + 2 \log c_1 \\
 &= -2 \log \text{fit } H_1 + 2 \log \text{complexity } H_1.
 \end{aligned}$$

Note: complexity value in BF depends on prior.

Note on conditional error probability

using PMPs and GORIC(A) weights

H_m	weights
H_1 : Sex Match	.04
H_2 : Gender Role Match	.81
H_3 : Sex Mismatch	.01
H_4 : Gender Role Mismatch	.00
H_u :	.14

The conditional error probability for preferred hypotheses H_2 is $1 - .81 = .19$.

Note on conditional error probability

using PMPs and GORIC(A) weights

What if we compare many hypotheses?

H_m	weights
H_1 : Sex Match	.013
H_2 : Gender Role Match	.270
H_3 : Sex Mismatch	.003
H_4 : Gender Role Mismatch	.000
H_5 : Lets try this one too	.180
...	...
H_{12} : Don't miss something	.040
H_u :	.047

The conditional error probability for H_2 becomes $1 - .27 = .77$ (was .19).

Namely, included additional hypotheses may obtain some support.
Hence, the conditional error probabilities becomes larger.

Simulation study

Results Exploration vs Confirmatory ($k = 3$ and $n = 50$)

$$\begin{aligned}
 H_{30E} : \mu_1 = \mu_2 = \mu_3 & & H_{30C} : \mu_1 = \mu_2 = \mu_3 \\
 H_{31E} : \mu_1 = \mu_2, \mu_3 & & H_{31C} : \mu_1 < \mu_2 < \mu_3 \\
 H_{32E} : \mu_1, \mu_2 = \mu_3 & & H_{32C} : \mu_1 = \mu_2 < \mu_3 \\
 H_{33E} : \mu_1 = \mu_3, \mu_2 & & H_{33C} : \mu_1 < \mu_2 > \mu_3 \\
 H_{34E} : \mu_1, \mu_2, \mu_3 & & H_{34C} : \mu_1, \mu_2, \mu_3
 \end{aligned}$$

'ES'	Method	H_{30E}	H_{31E}	H_{32E}	H_{33E}	H_{34E}
0	PCIC: AIC	0.658	0.113	0.115	0.112	0.002
2	PCIC: AIC	0.201	0.356	0.358	0.028	0.057
5	PCIC: AIC	0.000	0.135	0.136	0.000	0.729

'ES'	Method	H_{30C}	H_{31C}	H_{32C}	H_{33C}	H_{34C}
0	ORIC	0.662	0.065	0.119	0.120	0.034
2	ORIC	0.148	0.506	0.311	0.029	0.006
5	ORIC	0.000	0.955	0.044	0.001	0.000

Results Confirmation ($k = 3$ and $n = 50$)

$$H_{30C} : \mu_1 = \mu_2 = \mu_3$$

$$H_{31C} : \mu_1 < \mu_2 < \mu_3$$

$$H_{32C} : \mu_1 = \mu_2 < \mu_3$$

$$H_{33C} : \mu_1 < \mu_2 > \mu_3$$

$$H_{34C} : \mu_1, \mu_2, \mu_3$$

'ES'	Method	'Prior'	H_{30C}	H_{31C}	H_{32C}	H_{33C}	H_{34C}
0	ORIC		0.662	0.065	0.119	0.120	0.034
0	BMS	3	0.865	0.012	0.075	0.036	0.012
0	BMS	2	0.774	0.022	0.124	0.058	0.022
0	BMS	1	0.656	0.047	0.147	0.120	0.033
2	ORIC		0.148	0.506	0.311	0.029	0.006
2	BMS	3	0.299	0.299	0.371	0.031	0.001
2	BMS	2	0.252	0.328	0.381	0.040	0.000
2	BMS	1	0.125	0.494	0.321	0.063	0.000
5	ORIC		0.000	0.955	0.044	0.001	0.000
5	BMS	3	0.000	0.887	0.112	0.001	0.000
5	BMS	2	0.001	0.890	0.107	0.002	0.000
5	BMS	1	0.000	0.928	0.070	0.002	0.000

Results Confirmation ($k = 3$ and $n = 50$)

$H_{30C} : \mu_1 = \mu_2 = \mu_3$

$H_{31C} : \mu_1 < \mu_2 < \mu_3$

$H_{32C} : \mu_1 = \mu_2 < \mu_3$

$H_{33C} : \mu_1 < \mu_2 > \mu_3$

$H_{34C} : \mu_1, \mu_2, \mu_3$

1 – **Type I error** (here):

not preferring the true null hypothesis,
given the set of hypotheses (!).

'ES'	Method	'Prior'	H_{30C}	H_{31C}	H_{32C}	H_{33C}	H_{34C}
0	ORIC		0.662	0.065	0.119	0.120	0.034
0	BMS	3	0.865	0.012	0.075	0.036	0.012
0	BMS	2	0.774	0.022	0.124	0.058	0.022
0	BMS	1	0.656	0.047	0.147	0.120	0.033
2	ORIC		0.148	0.506	0.311	0.029	0.006
2	BMS	3	0.299	0.299	0.371	0.031	0.001
2	BMS	2	0.252	0.328	0.381	0.040	0.000
2	BMS	1	0.125	0.494	0.321	0.063	0.000
5	ORIC		0.000	0.955	0.044	0.001	0.000
5	BMS	3	0.000	0.887	0.112	0.001	0.000
5	BMS	2	0.001	0.890	0.107	0.002	0.000
5	BMS	1	0.000	0.928	0.070	0.002	0.000

Results Confirmation ($k = 3$ and $n = 50$)

$H_{30C} : \mu_1 = \mu_2 = \mu_3$

$H_{31C} : \mu_1 < \mu_2 < \mu_3$

$H_{32C} : \mu_1 = \mu_2 < \mu_3$

$H_{33C} : \mu_1 < \mu_2 > \mu_3$

$H_{34C} : \mu_1, \mu_2, \mu_3$

1 – **Type II error** (here):

not preferring the true hypothesis,
given the set of hypotheses (!).

'ES'	Method	'Prior'	H_{30C}	H_{31C}	H_{32C}	H_{33C}	H_{34C}
0	ORIC		0.662	0.065	0.119	0.120	0.034
0	BMS	3	0.865	0.012	0.075	0.036	0.012
0	BMS	2	0.774	0.022	0.124	0.058	0.022
0	BMS	1	0.656	0.047	0.147	0.120	0.033
2	ORIC		0.148	0.506	0.311	0.029	0.006
2	BMS	3	0.299	0.299	0.371	0.031	0.001
2	BMS	2	0.252	0.328	0.381	0.040	0.000
2	BMS	1	0.125	0.494	0.321	0.063	0.000
5	ORIC		0.000	0.955	0.044	0.001	0.000
5	BMS	3	0.000	0.887	0.112	0.001	0.000
5	BMS	2	0.001	0.890	0.107	0.002	0.000
5	BMS	1	0.000	0.928	0.070	0.002	0.000

Extra: Prior choice for BMS

Prior

$$P(\mu, \sigma^2) = P(\mu_1) \times \dots \times P(\mu_k) \times P(\sigma^2),$$

where $P(\mu_i) = P(\mu) \forall i = 1, \dots, k$.

$P(\mu)$ is a data-based normal distribution, with parameters β_0 and τ_0^2 , and $P(\sigma^2)$ an inverse chi-squared distribution (see Klugkist et al., 2005)

β_0 and τ_0^2 data-based hyperparameters

For each μ_i ($i = 1, \dots, k$) a credibility interval is computed:

estimated group mean \pm $PV \times$ *the group standard error*.

The credibility interval for μ_i is represented by " $[LB_i, UB_i]$ ".

Lowest lower bound: LB_{min} .

Highest upper bound: UB_{max} .

Then, $\beta_0 = \frac{LB_{min} + UB_{max}}{2}$ and $\tau_0^2 = \left[\frac{UB_{max} - LB_{min}}{2} \right]^2$.

Three types of prior vagueness: BMS PV 1, BMS PV 2 and BMS PV 3