

ANOVA Example: Comparisons of 3 Means

Example Palmer and Gough (P&G)

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 $\epsilon_j \sim N(0, \sigma^2)$.

$(n_i = 20, 20, 31)$

Hypotheses of interest

Example Palmer and Gough (P&G)

Palmer and Gough (2007) expect that

- non-offenders (1) attribute more importance to defective education for explaining crime than the other two offenders (2 & 3):
i.e., $\mu_1 > \mu_2$ and $\mu_1 > \mu_3$,
- property offenders (2) attribute more importance to defective education than person offenders (3):
i.e., $\mu_2 > \mu_3$.

This leads to the theory-based hypothesis:

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

Note that ‘<’ denotes “smaller than” and ‘>’ denotes “larger than”.

ANOVA Example: Comparisons of 3 Means

Simple, hypothetical example

Examine the difference in happiness
between three types of “treatments”:
(1) new treatment, (2) current treatment, and (3) no treatment.

Theory-based hypothesis:

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

where “>” denotes “larger than”.

Testing the null hypothesis

Example 3 means

Test H_0 with ANOVA F test:

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

$$H_a : \quad \text{not } H_0.$$

Then, reject or not-reject ('accept') H_0 .

But, often not interested in H_0 !

Cannot say anything about $H_1 : \mu_1 > \mu_2 > \mu_3$.

Conclusions w.r.t. hypothesis of interest $H_1 : \mu_1 > \mu_2 > \mu_3$

Example 3 means

With post-hoc test (e.g., Bonferroni) and on basis of the sample means, we could say a bit more.

But, what if

- $\mu_1 \neq \mu_2, \mu_1 \neq \mu_3, \mu_2 = \mu_3,$
- $\bar{y}_1 > \bar{y}_2,$

which implies $\mu_1 > \mu_2 = \mu_3$.

But also not interested in this.

Furthermore, inconsistent results are possible (like in P&G example):

$\mu_1 \neq \mu_3, \mu_1 = \mu_2,$ and $\mu_2 = \mu_3,$

which is logically impossible, since latter two imply $\mu_1 = \mu_3$.

Using planned contrasts would overcome this, but still = restrictions then.

Exploration and Confirmation

Exploration (like in post hoc tests and possibly AIC)

Evaluate all possible pairs/subsets of means whether significant different (“,”) or not (“=”).

For example, when $k = 3$:

$$H_{0E} : \mu_1 = \mu_2 = \mu_3$$

$$H_{1E} : \mu_1 = \mu_2, \mu_3$$

$$H_{2E} : \mu_1, \mu_2 = \mu_3$$

$$H_{3E} : \mu_1 = \mu_3, \mu_2$$

$$H_a : \mu_1, \mu_2, \mu_3$$

When $k = 5$, there are even 52 hypotheses.

Confirmation

Limited set: Evaluate only prespecified hypotheses including order restrictions ($<$, $>$, but also $=$).

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: Fbar (\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, that is, 3 means

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

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

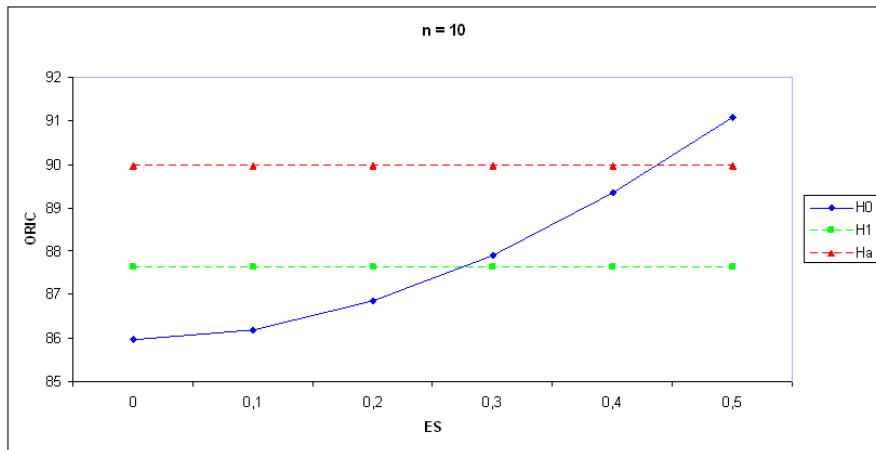
$$H_u : \mu_1, \mu_2, \mu_3.$$

GORIC

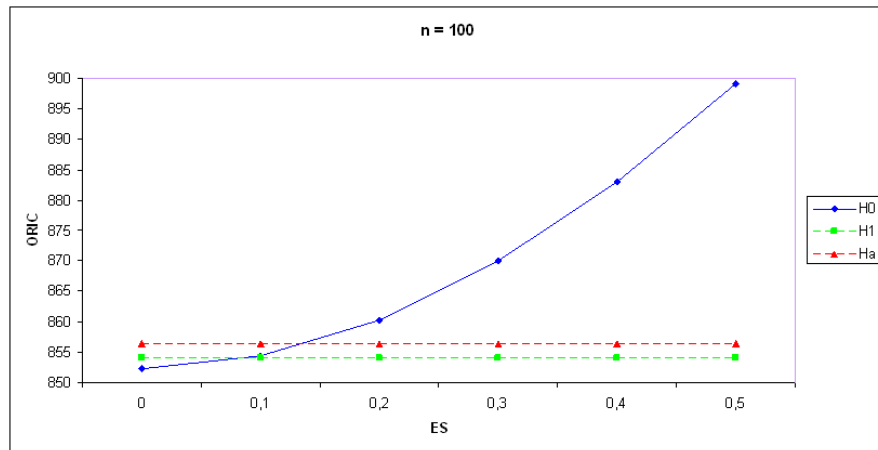
Model	Fit	Complexity	GORIC	GORIC weights
H_0	-196.36	2.00	396.71	0.02
H_1	-191.89	2.81	389.41	0.75
H_u	-191.89	4.00	391.79	0.23

Confirmatory methods (e.g., GORIC) have more “power” than their exploratory counterparts (e.g., AIC; cf. one- 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)

Your hypothesis of interest

If/When you have your own data

- What is your research question?
- What is your theory / expectation?
- What is your statistical hypothesis?
- Is there a competing statistical hypothesis?

Notably, think outside of the H_0 -box...

Table of Contents

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

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

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).

Generalized Order-Restricted Information Criterion

GORIC

$$'IC' = -2 \text{ fit} + 2 \text{ complexity}$$

Fit = Maximized order-restricted log likelihood

Maximized log likelihood based on parameters in agreement with H_m .

Complexity = Penalty

Represents: Expected number of distinct/free parameters.

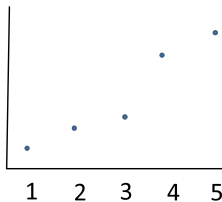
Here, expected number of distinct mean values plus 1 (because of the unknown variance term).

Details: Function of level probabilities.

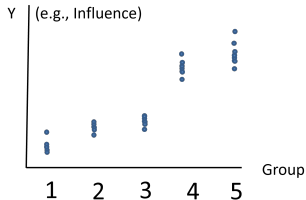
Intermezzo: Balance Fit and Complexity (1/6)

Example 5 means

Data for 5 groups – Oversimplified representation

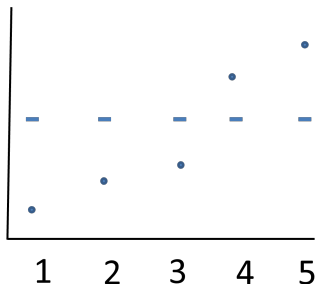


Data for 5 groups – Better representation (scatter of data points)



Intermezzo: Balance Fit and Complexity (2/6)

Example 5 means



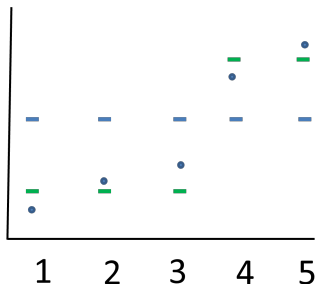
$$\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu$$

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

Fit: Bad fit

Intermezzo: Balance Fit and Complexity (3/6)

Example 5 means



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

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

parameters:

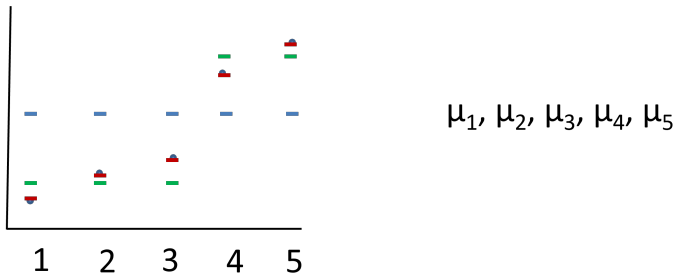
2 mean parameters (i.e., μ_{low} and μ_{high})

Fit:

Better than with one parameter, even good fit.

Intermezzo: Balance Fit and Complexity (4/6)

Example 5 means



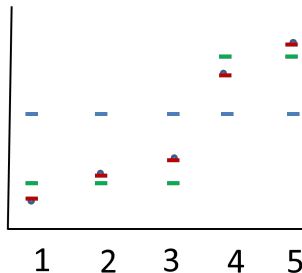
parameters: 5 mean parameter (complex)

Fit: Better than with two parameters, even best possible fit.

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

Intermezzo: Balance Fit and Complexity (5/6)

Example 5 means



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

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

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

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

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

Intermezzo: Balance Fit and Complexity (6/6)

Order-restrictions

Previous slides demonstrated balancing fit and complexity for equality ('=') restrictions.

For inequality/order ('<', '>', and possible some '=') restrictions, the calculation of the fit and complexity is more complex :-), but the idea of balancing is the same.

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: loose interpretation

Example Palmer and Gough, that is, 3 means

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

contains 1 ordering of three means, 1-2-3.

Thus, not complex (i.e., parsimonious).

$$H_2 : \mu_1 > (\mu_2, \mu_3)$$

contains 2 orderings of three means: 1-2-3 and 1-3-2.

Thus, more complex (less parsimonious).

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

contains all six possible orderings of three means.

Thus, is most complex one (least parsimonious).

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

$$IC_m = -2 \text{ fit}_m + 2 \text{ complexity}_m$$

Broad type of restrictions

More or less: any linear restriction.

e.g., the interaction $H_1 : \mu_1 - \mu_2 < \mu_3 - \mu_4$.

Note

If no inequalities ($<$ and/or $>$), then (G)ORIC = AIC.

Reference:

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: Lowest value is best

GORIC is like AIC expected distance from the truth (KL-distance).

Or: One wants the smallest 'misfit + complexity'.

Note: $\text{GORIC} = -2 \text{ fit} + 2 \text{ complexity} = 2 \text{ misfit} + 2 \text{ complexity}$

Hence, smallest value is best.

GORIC

Example Palmer and Gough

H_0 : $\mu_1 = \mu_2 = \mu_3$ (include only when of interest)

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

H_2 : $\mu_2 > \mu_1 > \mu_3$ (if competing theory/hypothesis)

H_u : μ_1, μ_2, μ_3 .

GORIC

Model	Fit	Complexity	GORIC
H_0	-196.36	2.00	396.71
H_1	-191.89	2.81	389.41
H_2	-193.70	2.81	393.03
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 dir.).
- 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

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

GORICA

GORIC(A) in JASP

End & Extra

Interpretation: GORIC weights

GORIC values

GORIC values cannot be interpreted, only compared:
Smallest is best.

GORIC weights (w_m)

w_m quantifies relative support for H_m versus others in the set.
Values between 0 and 1, and they sum to 1.
The bigger, the better.

GORIC weight ratios ($w_m/w_{m'}$)

$w_m/w_{m'}$ quantifies relative support of H_m vs $H_{m'}$.
Values between 0 and infinity:
* If < 1 , lack of support for H_m (nl. $w_{m'}/w_m$, support for $H_{m'}$).
* If > 1 , support for H_m .

Illustration of the GORIC weights (w_m)

Example Palmer and Gough

H_0 : $\mu_1 = \mu_2 = \mu_3$ (include only when of interest)

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

H_2 : $\mu_2 > \mu_1 > \mu_3$ (if competing theory/hypothesis)

H_u : μ_1, μ_2, μ_3 .

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H_1	-191.89	2.81	389.41	0.68
H_2	-193.70	2.81	393.03	0.11
H_u	-191.89	4.00	391.79	0.21

H_1 is $0.68 / 0.11 \approx 6.11$ times more supported than H_2 .

Hands-on/Demo (1b): GORIC weights

Let's practice.

- If needed: Start Rstudio again (optional: make project) and then also load packages again.
- Read and inspect data.
- Run model (`lm()`).
- Specify hypotheses (make up your own).
Note: Use names used in the model.
- Run `goric()`.
- Inspect and interpret output: Focus on GORIC weights.

Table of Contents

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

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.: $w_1 = .8$ and $w_2 = .2$.

Prevent choosing a weak hypothesis

Include unconstrained hypothesis H_u (or H_a):

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

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

$$H_u : \quad \mu_1, \mu_2, \mu_3, \\ \text{(i.e., no restrictions).}$$

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

E.g.: $w_1 = .08$, $w_2 = .02$, and $w_u = .90$.

Use of H_u

Palmer & Gough

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

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

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

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
H1	-191.89	2.81	389.41	0.68
H2	-193.70	2.81	393.03	0.11
unconstrained	-191.89	4.00	391.79	0.21

If at least one informative hypothesis not weak (i.e., $w_m > w_u$ 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 ctd.

Palmer & Gough

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

Unconstrained is always true, so will always receive support.
Support for H_u contains support for H_1 (if any).

Now: Fit H_1 and H_u the same;
thus, GORIC weights are based on only complexity.

Consequently, $0.77 / 0.23 \approx 3.27$ is an upper bound.

H_1 vs H_u : Upper bound

Based on simulation

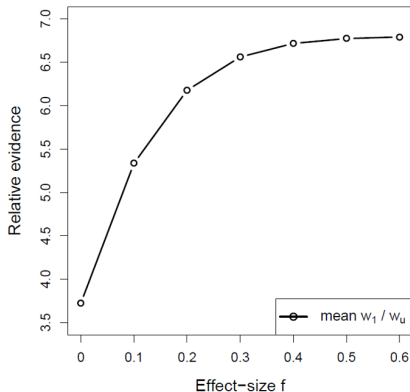


Figure: IC weights can have an upper bound, when informative hypothesis has maximum fit (i.e., is fully in agreement with the data).

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.

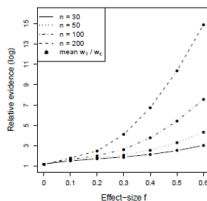
Reference:

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>

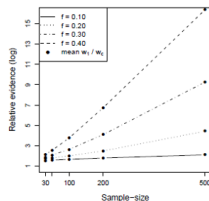
Alternative failsafe: Complement of H_m

Based on simulation

vs complement

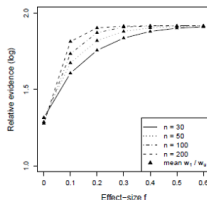


(a)

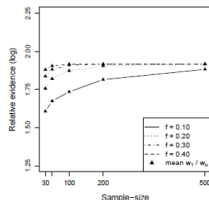


(b)

vs unconstrained



(c)



(d)

Note: unconstrained vs complement

H_1 contains 1 ordering of means:

1. $\mu_1 > \mu_2 > \mu_3$

H_c contains 5 orderings of means:

2. $\mu_1 > \mu_3 > \mu_2$
3. $\mu_2 > \mu_1 > \mu_3$
4. $\mu_2 > \mu_3 > \mu_1$
5. $\mu_3 > \mu_1 > \mu_2$
6. $\mu_3 > \mu_2 > \mu_1$

H_u combines H_1 and H_c , and thus consists of all 6 possible orderings.

R code: complement

Palmer & Gough

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

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

GORIC

```
H1 <- 'group1 > group2 > group3'  
# vs its complement (default in case of one hypothesis)  
#  
# GORIC (using goric function in restriktor package)  
library(restriktor)  
set.seed(123) # Set seed value  
goric.PandG_1c <- goric(fit.PandG,  
                        hypotheses = list(H1))
```

H_1 vs H_c

Palmer & Gough

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

H_c : not H_1 ,
complement of H_1 .

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 $0.79 / 0.21 \approx 3.73$ times more supported than its complement, that is, any other hypothesis (i.e., any other ordering).

Note on use of safeguard hypotheses (2/2)

My advise and default in goric function:

- If multiple informative, theory-based hypotheses:
Goal: Comparison of informative, theory-based hypotheses.
Safeguard: Unconstrained.
- If one informative, theory-based hypothesis (H_m):
Goal: Compare H_m with all other possibilities / theories.
Safeguard: Its complement.

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:

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

Note: Use names used in the model.

- Run goric()
- Inspect and interpret output.

Table of Contents

Generalized Order-Restricted Information Criterion (GORIC)

GORIC weights

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'  
# vs its complement (default in case of one hypothesis)  
#  
# GORICA (using goric function in restriktor package)  
library(restriktor)  
set.seed(123) # Set seed value  
gorica.PandG_C <- goric(fit.PandG,  
                        hypotheses = list(H1),  
                        type = "gorica") # needed if lm object
```

GORICA

Palmer & Gough

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

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

GORIC

Model	Fit	Complexity	GORICA	GORICA weights
H1	-1.96	1.81	7.55	0.79
complement	-2.39	2.69	10.15	0.21

H_1 is $0.79 / 0.21 \approx 3.67$ times more supported than its complement, that is, any other hypothesis.

Note: GORIC weights are the same (ratio may differ a bit).

More information GORIC(A) output

More information GORIC(A) output:

- 'Guidelines_output_GORIC.html'
- Possibly: 'Guidelines_GORIC-Benchmarks'

from <https://github.com/rebeccakuiper/Tutorials>.

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

Hands-on/Demo (3): GORICA

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 (since `lm` object):
`type = "gorica"`
- Inspect and interpret output.

GORIC in JASP: instructions

Palmer & Gough

- Extract data file 'Data_PalmerAndGough_JASP.txt' from 'data_JASP' subfolder (github.com/rebeccakuiper/Tutorials).
- Open JASP and 'Data_PalmerAndGough_JASP.txt'. Check measurement levels.
- Go to ANOVA and denote variables of interest.
- Go to Order Restricted Hypotheses tab.
- Specify hypotheses. For example,

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

$$JASP : \quad group1 > group2 > group3$$

- Press Ctrl+Enter. Inspect and interpret output.

Palmer & Gough

59/88

GORIC in JASP: instructions

Lucas

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

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

$$\begin{aligned} \text{JASP :} \quad & \text{group5} = \text{group3} > \text{group1} > \text{group2} \\ & \text{group3} > \text{group4} > \text{group2} \end{aligned}$$

- Press Ctrl+Enter. Inspect and interpret output.

GORIC in JASP: screenshot

Lucas

The screenshot shows the JASP software interface. The top toolbar includes icons for Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, Bain, SEM, and R (Beta). The main window is divided into two panes. The left pane, titled 'Order Restricted Hypotheses', contains instructions on how to enter restrictions and syntax settings. The right pane, titled 'Results', displays the ANOVA table and the Model Comparison table.

Order Restricted Hypotheses

Enter each restriction of one hypothesis on a new line, e.g.,
factorLow == factorMid
factorMid < factorHigh
where 'factor' is the factor (or covariate) name and 'Low'/'Mid'/'High' are the factor level names.
Click on the 'plus' icon to add more hypotheses.
Click the information icon for more examples.

Syntax settings

- ☐ Include intercept
- ☐ Show available coefficients

Set for all models

- ☐ Model summary
- ☐ Marginal means
- ☐ Informed hypothesis tests

Model 1

```
group5 = group3 > group1 > group2  
group3 > group4 > group2
```

Ctrl + Enter to apply. Click on the blue button above for help on the restriction syntax

Results

ANOVA

ANOVA - Influence

Cases	Sum of Squares	df	Mean Square	F	p
group	74.700	4	18.675	7.567	< .001
Residuals	357.835	145	2.468		

Note. Type III Sum of Squares

Order Restricted Hypotheses

Model Comparison

Model Comparison Table

Model	Log-likelihood	Penalty	GORIC	Weight	Weights ratio
Model 1	-278.051	3.197	562.495	0.930	13.357
Complement	-278.048	5.791	567.680	0.070	1.000

Note. Weights ratios indicate the relative weight for each model against the "complement" model. GORIC = Generalized Order-Restricted Information Criterion (Kuiper, Hoijtink, & Silvapulle, 2011).

GORIC and GORICA in JASP

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

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

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

Within Subjects Effects

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

Between Subjects Effects

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

ANCOVA

id
site
sex
age
viewcat
swelling
treat
prebody
prelet
preform
precarb
prevelad
preclast
postbody
postlet
postform
postcarb
postvelad
postclast
residual

Dependent Variable

Fixed Factors

Covariates

WLS Weights

Display

☐ Descriptive statistics

☐ Estimates of effect size

☐ Partial η^2

☐ Work-Save maximum p-value

Model

Assumption Checks

Contrasts

▶ Order Restricted Hypotheses

Your hypothesis of interest

If you have your own data

Before:

- What is your research question?
- What is your theory / expectation?
- What is your statistical hypothesis?
- Is there a competing statistical hypothesis?

Additionally:

- Are you able to specify your statistical hypothesis/-es?
- How will you evaluate it/them?

What's next

- Bayesian Model Selection (BMS)

Depending on time and wishes:

- Some extra information
- Demo in JASP (GORIC(A) and/or BMS)
- Demo in R (GORIC(A) and/or BMS)

We end with:

- Lab: [https://github.com/rebeccakuiper/
EMLaR---Informative-Hypothesis-Evaluation](https://github.com/rebeccakuiper/EMLaR---Informative-Hypothesis-Evaluation)

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/Extra2
informative-hypotheses.sites.uu.nl/software/goric/

E-mail

r.m.kuiper@uu.nl

Note on comparable estimates

Until now: comparing means.

Continuous predictors

If compare relative strength/importance of parameters (e.g., $\beta_1 > \beta_2$), then make sure comparable:
e.g., standardize continuous predictors.

Multiple outcomes

If compare parameters across outcomes, then (also) standardize outcomes.

Note on type of hypotheses

Absolute strength

Compare strength/importance of parameters: $\beta_1 > \beta_2$.

Compare absolute strength/importance of parameters:
 $abs(\beta_1) > abs(\beta_2)$.

About-equality

Equality: $\beta_1 = \beta_2$.

About-equality: $-0.01 < \beta_1 - \beta_2 < 0.01$.

Minimum difference

Difference: $abs(\beta_1) > abs(\beta_2)$.

Minimum difference: $abs(\beta_1) - abs(\beta_2) > 0.01$.

Note on hypotheses

1. Only include hypotheses with sound theoretical and/or empirical basis.
Often a null hypothesis is not of interest.
2. Keep the number of hypotheses included as small as possible.
3. This is a subjective endeavor, aim for inter-peer / inter-subjective agreement.

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 (1/2)

- Logistic Regression Modeling
 - Article: doi.org/10.1037/met0000406
- GORICA on SEM
 - Article: www.tandfonline.com/doi/full/10.1080/10705511.2020.1836967.
 - R scripts: github.com/rebeccakuiper/GORICA_in_SEM.
- GORICA on cross-lagged panel model (CLPM)
 - Article: doi.org/10.1111/bjep.12455.
 - R scripts: github.com/rebeccakuiper/GORICA_in_CLPM.
- GORICA on Random-Intercept CLPM (RI-CLPM)
 - Article: Sukpan, C. and Kuiper, R.M. (2023). How to evaluate causal dominance hypotheses in lagged effects models.
 - R scripts: github.com/rebeccakuiper/GORICA%20in%20RI-CLPM.
- GORICA for EffectLiteR
 - R scripts: <https://github.com/rebeccakuiper/Tutorials/tree/main/GORICA%20for%20EffectLiteR>.

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

Extra material (2/2)

- GORICA on CTmeta
 - Article: doi.org/10.1080/10705511.2020.1823228.
 - R scripts: github.com/rebeccakuiper/GORICA_on_CTmeta.
- GORICA on Meta-analysis
 - Article: doi.org/10.3390/e24111525.
 - R scripts: 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 is related 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 GORIC weights vs BF and PMPs

ratio GORIC weights ($w_m/w_{m'}$) \sim Bayes factor ($BF_{mm'}$).

GORIC weight (w_m) \sim posterior model probability (PMP).

1 - w_m = conditional error probability.

Like PMP, w_m depends on set of hypotheses.

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.

Note on conditional error probability

using PMPS and GORIC(A) weights

Hypotheses that overlap can also share support (like H_u does).

My advise:

Only use error probabilities if one hypothesis versus its complement.
(Or when you are sure that there is no overlap in hypotheses)

H_m	weights
H_2 : Gender Role Match	.82
H_c : complement H_2	.18

Properties of Methods: Simulation study

1. All methods (e.g., (G)ORIC and BMS) programmed in Fortran.
2. Generate data sets with know structure (conditions).
3. Apply methods to all data sets (within one condition).
4. Summarize results (per condition):
proportion of (correct) chosen hypotheses.

Reference:

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.

Population Values in Simulated Data Sets when $k = 3$

Values of Population Means (μ_i) and Standard Deviations (σ_i) Used to Simulate 1000 Data Sets

$k = 3$	Case Number ('ES')	μ_1	μ_2	μ_3	ES (Cohen)
	0	0.00	0.00	0.00	
	1	-0.10	0.00	0.10	.1
	2	-0.20	0.00	0.20	.2
	3	-0.30	0.00	0.30	.3
	4	-0.40	0.00	0.40	.4
	5	-0.50	0.00	0.50	.5
		σ_1	σ_2	σ_3	
	0 - 5	1.00	1.00	1.00	

NB in Case 0 $H_{30C} : \mu_1 = \mu_2 = \mu_3$ is true
in Cases 1 to 5 $H_{31C} : \mu_1 < \mu_2 < \mu_3$ is true.

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

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

$$H_{31E} : \mu_1 = \mu_2, \mu_3$$

$$H_{32E} : \mu_1, \mu_2 = \mu_3$$

$$H_{33E} : \mu_1 = \mu_3, \mu_2$$

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

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

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

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

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

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

'ES'	Method	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):

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):
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

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

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