

## Informative hypotheses evaluation

R. M. Kuiper

Department of Methodology & Statistics  
Utrecht University



# ANOVA Example: Comparisons of 3 Means

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

$(n_i = 20, 20, 31)$

# Hypotheses of interest

Palmer and Gough (2007) expect that

- non-offenders (1) attribute more importance to defective education for explaining crime than the other two offenders;  
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.,  $\mu_2 > \mu_3$ .

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

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



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

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 this data example):

$$\mu_1 \neq \mu_3, \mu_1 = \mu_2, \text{ and } \mu_2 = \mu_3,$$

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

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

# NHST on Palmer & Gough (2007) Data

```
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
```

Exploration (like in post hoc tests and possibly AIC)

Test 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: Compare 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:  $F_{\text{bar}} (\bar{F})$  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)

# GORIC on Palmer & Gough (2007) Data

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

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

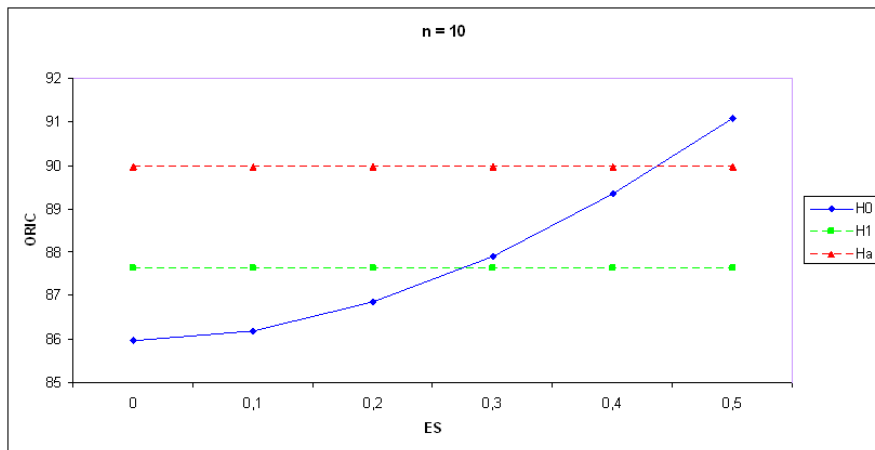
$$H_u : \mu_1, \mu_2, \mu_3.$$

## GORIC

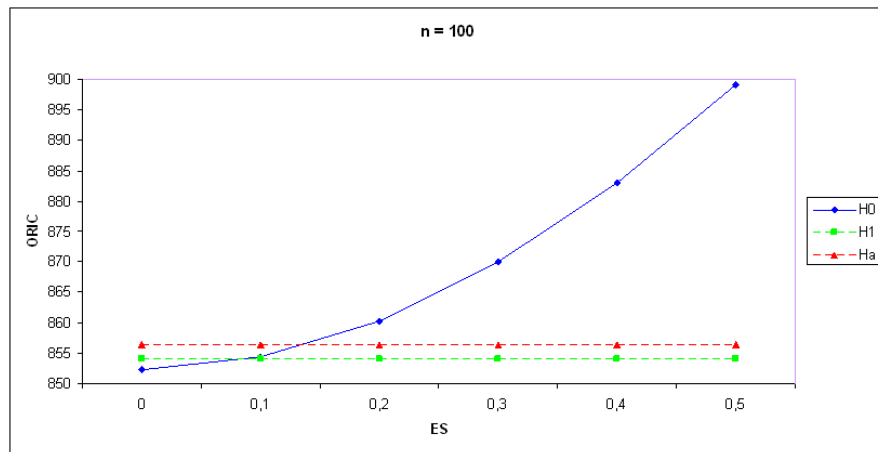
Model	Fit	Complexity	GORIC	GORIC weights
$H_0$	-196.36	2.00	396.71	0.02
$H_1$	-191.89	2.81	389.41	0.75
$H_u$	-191.89	4.00	391.79	0.23

Confirmatory methods have more “power” than their exploratory counterparts (cf. one-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





# 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 hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra



# Generalized Order-Restricted Information Criterion

## GORIC

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

Fit = Maximized order-restricted log likelihood

Maximized log likelihood based on parameters in agreement with  $H_m$ .

Complexity = Penalty

Represents: Expected number of distinct parameters.

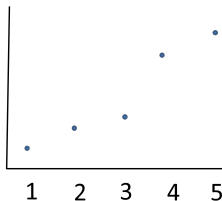
Here, expected number of distinct mean values plus 1 (because of the unknown variance term).

Details: Function of level probabilities.

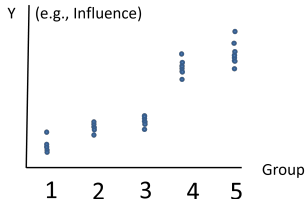


# Intermezzo: Balance Fit and Complexity

Data for 5 groups – Oversimplified representation

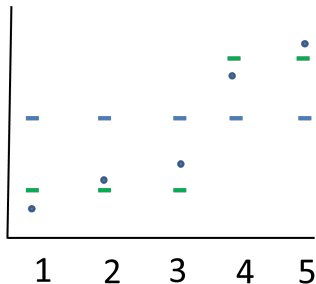


Data for 5 groups – Better representation (scatter of data points)





## Intermezzo: Balance Fit and Complexity



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

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

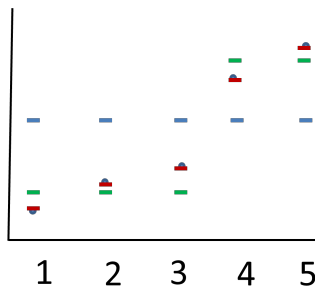
# parameters:

2 mean parameters (i.e.,  $\mu_{\text{low}}$  and  $\mu_{\text{high}}$ )

Fit:

Better than with one parameter, even good fit.

## Intermezzo: Balance Fit and Complexity



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

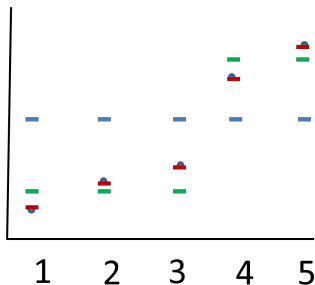
# parameters: 5 mean parameter (complex)

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

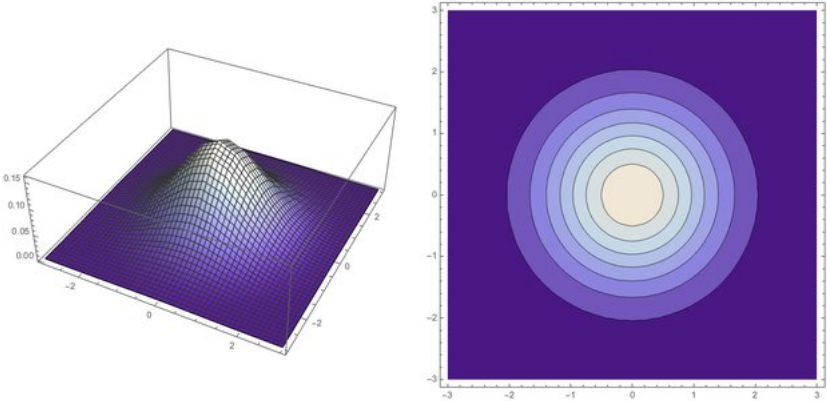


Now, we also incorporate order-restrictions (e.g.,  $\mu_1 < \mu_2$ ).

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

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

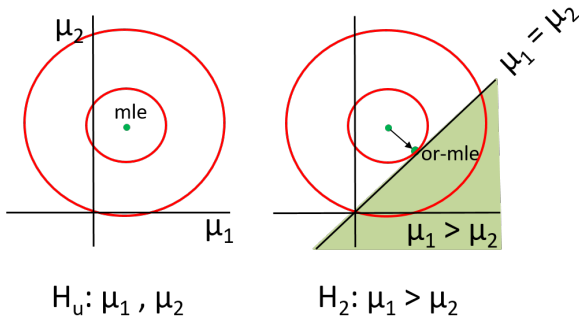
Intermezzo: Contour plot





# Idea fit

order-restricted maximum likelihood (or-ml)



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.



# GORIC on Palmer & Gough (2007) Data

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

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

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

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

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

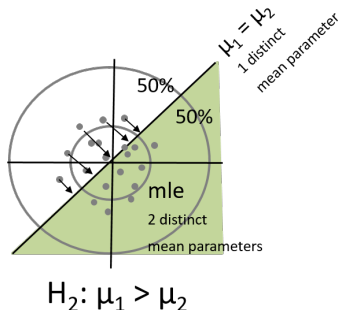
```
## group1 group2 group3
## 11.95  9.75  8.77
```

## 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)



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

loose interpretation

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

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).

# GORIC on Palmer & Gough (2007) Data

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

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

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

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

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

## 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.

# GORIC on Palmer & Gough (2007) Data

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

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

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

## 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

# Hands-on/Demo (1): GORIC

Start Rstudio and 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).
- Optional: Open 'Tutorial\_GORIC\_restriktor\_ANOVA.html'.
- Start R (optional: make project).  
Optional: open 'PalmerAndGough\_and\_Lucas.R'.
- Install packages and load them.
- Read and inspect data. Use Data\_PalmerAndGough.txt.
- Run model (lm()).
- Specify hypotheses (make up your own).  
Note: Use names used in the model, or overwrite those.
- Run goric().
- Inspect and interpret output.  
Note: 'GORIC weights' will be explained later on.



# 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 hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

# Interpretation: GORIC weights

## GORIC values

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

## GORIC weights ( $w_m$ ) and ratios ( $w_m/w_{m'}$ )

$w_m$  quantifies how much  $H_m$  is more supported than others in set.  
 $w_m/w_{m'}$  quantifies relative support of  $H_m$  vs  $H_{m'}$ .  
The bigger, the better.

Reference:

Kuiper, R.M., Hoijtink, H. and Silvapulle, M.J. (2012). Generalization of the order restricted information criterion for multivariate normal linear models. *Journal of Statistical Planning and Inference*, 142, 2454-2463.

# Illustration of the GORIC weights ( $w_m$ )

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

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

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

## 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

## 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.

### Example

$$w_m = .8$$

$$w_{m'} = .2$$

$H_m$  is  $.8/.2 = 4$  times more supported than  $H_{m'}$ .

If  $H_{m'}$  complement of  $H_m$ :

There is an error probability of  $(1 - .8) * 100\% = 20\%$  that  $H_m$  is true.

# 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 hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

## Palmer & Gough

```
fit.PandG <- lm(Importance ~ group - 1, data = PandG_data)
```



R code: GORIC

Palmer &amp; 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 (using goric function in restriktor package)
library(restriktor)
set.seed(123) # Set seed value
goric.PandG <- goric(fit.PandG,
                     hypotheses = list(H0, H1))
```



# GORIC on Palmer & Gough Data

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

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

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

## 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

# 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

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$  (include only when of interest),

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

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

$H_u : \mu_1, \mu_2, \mu_3, \mu_4, \mu_5$  (includes 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: Data and fit object

Lucas

```
# Data  
lucas_data <- read.table("Data_Lucas.txt", header=TRUE)  
lucas_data$group <- factor(lucas_data$group)  
  
# fit object (needed as input for goric function)  
fit.lucas <- lm(Influence ~ group - 1, data = lucas_data)
```



R code: GORIC

Lucas

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

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

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

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

```
# GORIC (using goric function in restriktor package)
library(restriktor)
set.seed(123) # Set seed value
goric.lucas <- goric(fit.lucas,
                     hypotheses = list(H0, H1, H2))
```

# GORIC on Lucas' Data

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

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

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

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

## GORIC

Model	Fit	Complexity	GORIC	GORIC weights
$H_0$	-292.27	2.00	588.54	0.00
$H_1$	-278.05	3.20	562.49	0.92
$H_2$	-281.76	3.14	569.79	0.02
$H_u$	-278.05	6.00	568.10	0.06

## GORIC on Lucas' Data: GORIC weights ( $w_m$ )

GORIC

Model	Fit	Complexity	GORIC	GORIC weights
$H_0$	-292.27	2.00	588.54	0.00
$H_1$	-278.05	3.20	562.49	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  $.921/.024 \approx 38.6$  ( $\infty$ ) times more supported than  $H_2$  ( $H_0$ , resp.). Thus, there is quite some evidence that  $H_1$  is the best of this set.

Note:

$H_1$  is  $.921/.056 \approx 16.5 > 1$  times more supported than  $H_\nu$ .

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



## Hands-on/Demo (2): GORIC weights

Start Rstudio and let's practice.

- Optional: Open 'Tutorial\_GORIC\_restriktor\_ANOVA.html'.
- If needed: Start R again (optional: make project) and then also load packages again.  
Optional: open 'PalmerAndGough\_and\_Lucas.R'.
- Read and inspect data.  
Use Data\_Lucas.txt and/or Data\_PalmerAndGough.txt.
- Run model (lm()).
- Specify hypotheses (make up your own). Note: Use names used in the model, or overwrite those.
- Run goric().
- Inspect and interpret output: Focus on GORIC weights.

# 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 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.

E.g.:  $w_0 = .02$ ,  $w_1 = .08$ , and  $w_u = .90$ .



## Use of $H_u$ : Palmer & Gough (2007) Data

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

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

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

### 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

If at least one informative hypothesis not weak ( $w_1 > w_u$  or  $w_1/w_u > 1$ ), then compare informative hypotheses.

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

# $H_1$ vs $H_u$ : Palmer & Gough (2007) Data

What if only one informative hypothesis:

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

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

## GORIC

Model	Fit	Complexity	GORIC	GORIC weights
$H_1$	-191.89	2.81	389.41	<b>0.77</b>
$H_u$	-191.89	4.00	391.79	<b>0.23</b>

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

BUT:  $H_u$  includes  $H_1$ .

So, support for  $H_u$  contains support for  $H_1$ .

# $H_1$ vs $H_u$ : Palmer & Gough (2007) Data

## GORIC

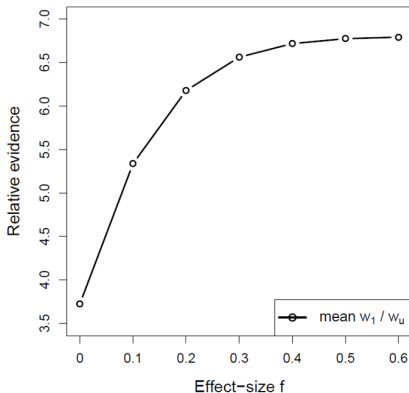
Model	Fit	Complexity	GORIC	GORIC weights
$H_1$	-191.89	2.81	389.41	0.77
$H_u$	-191.89	4.00	391.79	0.23

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,  $.77/.23 \approx 3.3$  is an upper bound.

# $H_1$ vs $H_u$ : Upper bound



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



# Alternative safeguard: Complement of $H_m$

## Alternatively (in 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.

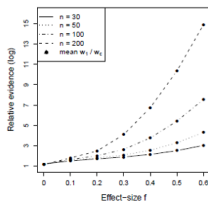
Reference:

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.

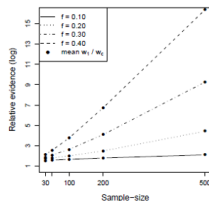
<https://doi.org/10.1037/met0000238>

## Alternative safeguard: Complement of $H_m$ using GORIC

vs complement

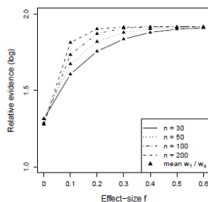


(a)

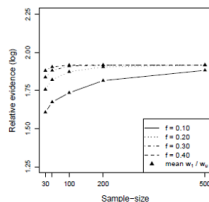


(b)

vs unconstrained



(c)



(d)



# Palmer & Gough (2007): $H_1$ vs $H_c$

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

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

## 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  $.79/.21 \approx 3.8$  times more supported than its complement, that is, any other hypothesis.

## Hands-on/Demo (3): GORIC - $H_m$ vs complement

Start Rstudio and let's practice.

- Optional: Open 'Tutorial\_GORIC\_restriktor\_ANOVA.html'.
- If needed: Start R again (optional: make project) and then also load packages again.
- Read and inspect data.  
Use Data\_Lucas.txt and/or Data\_PalmerAndGough.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:  
`comparison = "complement"`

- Inspect and interpret output.



# GORICA

## GORIC: Normal linear models

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

## GORIC: Other models

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

## GORICA: All 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 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

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

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

GORICA: type = "gorica"

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



## Hands-on/Demo (4): GORICA - $H_m$ vs complement

Start Rstudio and let's practice.

- Optional: Open 'Tutorial\_GORIC\_restriktor\_ANOVA.html'.
- If needed: Start R again (optional: make project) and then also load packages again.
- Read and inspect data.  
Use Data\_Lucas.txt and/or Data\_PalmerAndGough.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: Palmer & Gough

- Open JASP and 'Data\_PalmerAndGough\_JASP.txt'. Check measurement levels.
- Go to ANOVA and denote variables of interest.
- Go to Order Restrictions tab.
- Specify hypotheses. For example,

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

*JASP* :  $group1 > group2 > group3$

- Press Ctrl+Enter. Inspect and interpret output.

More in (demo and) lab.

# GORIC in JASP: Palmer & Gough

## GORIC in JASP: 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,

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

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

- Press Ctrl+Enter. Inspect and interpret output.

# GORIC in JASP: Lucas

Data\_Lucas\_JASP
C:\WORK\Ow - 5000 - winter - Jasp on R\Lab\JASP labmeeting\JASP - GORIC slides

Descriptives

T-Tests

ANOVA

Mixed Models

Regression

Frequencies

Factor

Bain

SEM

R (Beta)

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

Set for all models

☐ Include intercept
☐ Model summary

☐ Show available coefficients
☐ 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

☐ Summary for Model 1
☐ Marginal means for Model 1
☐ Informed hypothesis tests for Model 1

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	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).

72/97



# 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?



# 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 hypotheses: Unconstrained & Complement

GORICA

GORIC(A) in JASP

End & Extra

## Promo: Publish in special issue

### Applying GORIC(A)?

If you collected data or will collect,  
if you have one or more a-priori hypotheses,  
then you can apply the GORIC(A).

### FYI: Special issue

Journal (mdpi): Mathematics (ISSN 2227-7390)

Special Issue:

*Evaluation of Theory-Driven Hypotheses: No Hypothesis, No G(L)ORIC*

Guest Editor: Rebecca M. Kuiper

[https://www.mdpi.com/journal/mathematics/special\\_issues/97C26430R3](https://www.mdpi.com/journal/mathematics/special_issues/97C26430R3)

Contact: [r.m.kuiper@uu.nl](mailto:r.m.kuiper@uu.nl)

Please contact me ([r.m.kuiper@uu.nl](mailto:r.m.kuiper@uu.nl)), if you want to explore the possibilities (and to obtain discount).

# The End

Thanks for listening!

Are there any questions?

## Websites

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
- Evidence synthesis / Support aggregation

We end with:

- Lab
- Discussion

## Notes





## Note on using complement

Complement not always higher weight

In case  $H_m$  is almost true, but not true:

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  $H_u$ , you choose  $H_m$  'sooner' because of low penalty for  $H_m$ .

## 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) and GORIC(A) weights:  
More data collected, (re-)calculate the GORIC(A) weights.
- 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.  
*See this html tutorial and/or this R script tutorial.*
- 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).  
*See this html tutorial and/or this R script tutorial.*

## 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 GORIC weights vs BF and PMPs

$w_m/w_{m'} = \text{relative weight} \sim \text{Bayes factor } (BF_{mm'})$ .

$w_m \sim \text{posterior model probability (PMP)}$ .

$1 - w_m = \text{conditional error probability}$ .

Note:  $w_m$  depends on set of hypotheses.

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



## 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](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](https://github.com/rebeccakuiper/GORICA_in_SEM).
- **GORICA on Random-Intercept CLPM (RI-CLPM)**
  - Article: Chuenjai Sukpan and Rebecca M. Kuiper (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](https://github.com/rebeccakuiper/GORICA_on_CTmeta).
- **GORICA on cross-lagged panel model (CLPM)**
  - Article: <https://doi.org/10.3390/e24111525>.
  - R scripts: [https://github.com/rebeccakuiper/GORICA\\_on\\_MetaAn](https://github.com/rebeccakuiper/GORICA_on_MetaAn).

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



## 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 ( $\mu_i$ ) and Standard Deviations ( $\sigma_i$ ) Used to Simulate 1000 Data Sets

$k = 3$	Case Number ('ES')	$\mu_1$	$\mu_2$	$\mu_3$	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
		$\sigma_1$	$\sigma_2$	$\sigma_3$	
	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 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		<b>0.662</b>	0.065	0.119	0.120	0.034
0	BMS	3	<b>0.865</b>	0.012	0.075	0.036	0.012
0	BMS	2	<b>0.774</b>	0.022	0.124	0.058	0.022
0	BMS	1	<b>0.656</b>	0.047	0.147	0.120	0.033
2	ORIC		0.148	<b>0.506</b>	0.311	0.029	0.006
2	BMS	3	0.299	<b>0.299</b>	0.371	0.031	0.001
2	BMS	2	0.252	<b>0.328</b>	0.381	0.040	0.000
2	BMS	1	0.125	<b>0.494</b>	0.321	0.063	0.000
5	ORIC		0.000	<b>0.955</b>	0.044	0.001	0.000
5	BMS	3	0.000	<b>0.887</b>	0.112	0.001	0.000
5	BMS	2	0.001	<b>0.890</b>	0.107	0.002	0.000
5	BMS	1	0.000	<b>0.928</b>	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		<b>0.662</b>	0.065	0.119	0.120	0.034
0	<b>BMS</b>	<b>3</b>	<b>0.865</b>	0.012	0.075	0.036	0.012
0	BMS	2	<b>0.774</b>	0.022	0.124	0.058	0.022
0	BMS	1	<b>0.656</b>	0.047	0.147	0.120	0.033
2	ORIC		0.148	<b>0.506</b>	0.311	0.029	0.006
2	BMS	3	0.299	<b>0.299</b>	0.371	0.031	0.001
2	BMS	2	0.252	<b>0.328</b>	0.381	0.040	0.000
2	BMS	1	0.125	<b>0.494</b>	0.321	0.063	0.000
5	ORIC		0.000	<b>0.955</b>	0.044	0.001	0.000
5	BMS	3	0.000	<b>0.887</b>	0.112	0.001	0.000
5	BMS	2	0.001	<b>0.890</b>	0.107	0.002	0.000
5	BMS	1	0.000	<b>0.928</b>	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		<b>0.662</b>	0.065	0.119	0.120	0.034
0	BMS	3	<b>0.865</b>	0.012	0.075	0.036	0.012
0	BMS	2	<b>0.774</b>	0.022	0.124	0.058	0.022
0	BMS	1	<b>0.656</b>	0.047	0.147	0.120	0.033
2	ORIC		0.148	<b>0.506</b>	0.311	0.029	0.006
2	BMS	3	0.299	<b>0.299</b>	0.371	0.031	0.001
2	BMS	2	0.252	<b>0.328</b>	0.381	0.040	0.000
2	BMS	1	0.125	<b>0.494</b>	0.321	0.063	0.000
5	ORIC		0.000	<b>0.955</b>	0.044	0.001	0.000
5	BMS	3	0.000	<b>0.887</b>	0.112	0.001	0.000
5	BMS	2	0.001	<b>0.890</b>	0.107	0.002	0.000
5	BMS	1	0.000	<b>0.928</b>	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  $\beta_0$  and  $\tau_0^2$ , and  $P(\sigma^2)$  an inverse chi-squared distribution (see Klugkist et al., 2005)

### $\beta_0$ and $\tau_0^2$ data-based hyperparameters

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

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

The credibility interval for  $\mu_i$  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  $\tau_0^2 = [\frac{UB_{max} - LB_{min}}{2}]^2$ .

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