Multiple testing: Intro & FWER[1]

Mark van de Wiel
mark.vdwiel@vumc.nl

Department of Epidemiology & Biostatistics,VUmc, Amsterdam

[1]Some slides courtesy of Jelle Goeman

VU medisch centrum

Practical notes

Study material

1. These slides (3 sets for multiple testing)

2. "*Tutorial in biostatistics: multiple hypothesis testing in genomics*" by Goeman & Solari. **Sections 4.3 & 4.4 are not compulsory study material.**

3. PDF handouts: technical details.

References

1. [hand]: refers to PDF handout
2. [tut]: refers to tutorial
3. [exer]: refers to exercise

Mutliple testing lectures

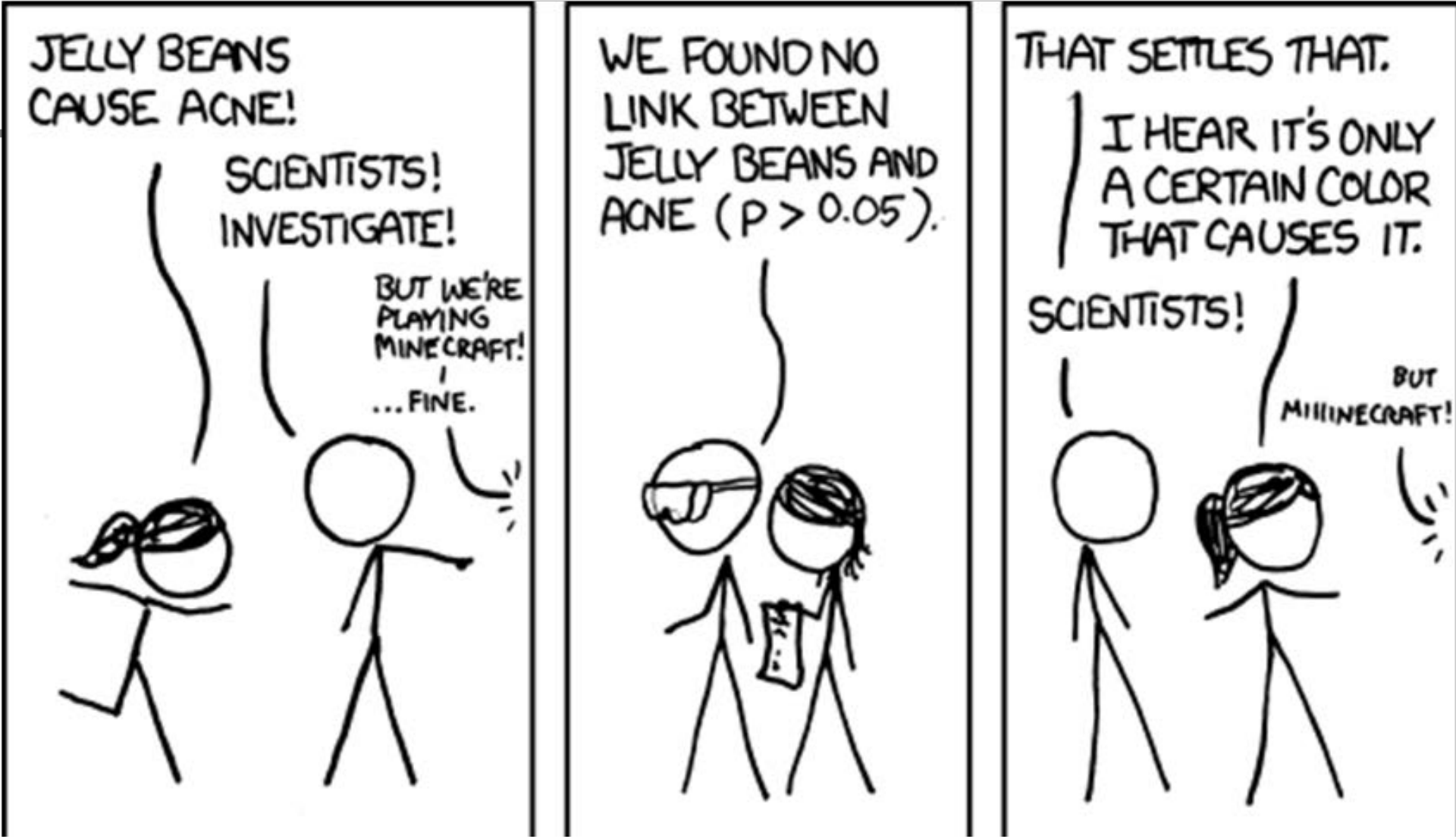
1. Introduction Multiple Testing & Family-wise error rate (FWER)

2. False Discovery Rate (FDR) + Bayesian perspective on multiple testing

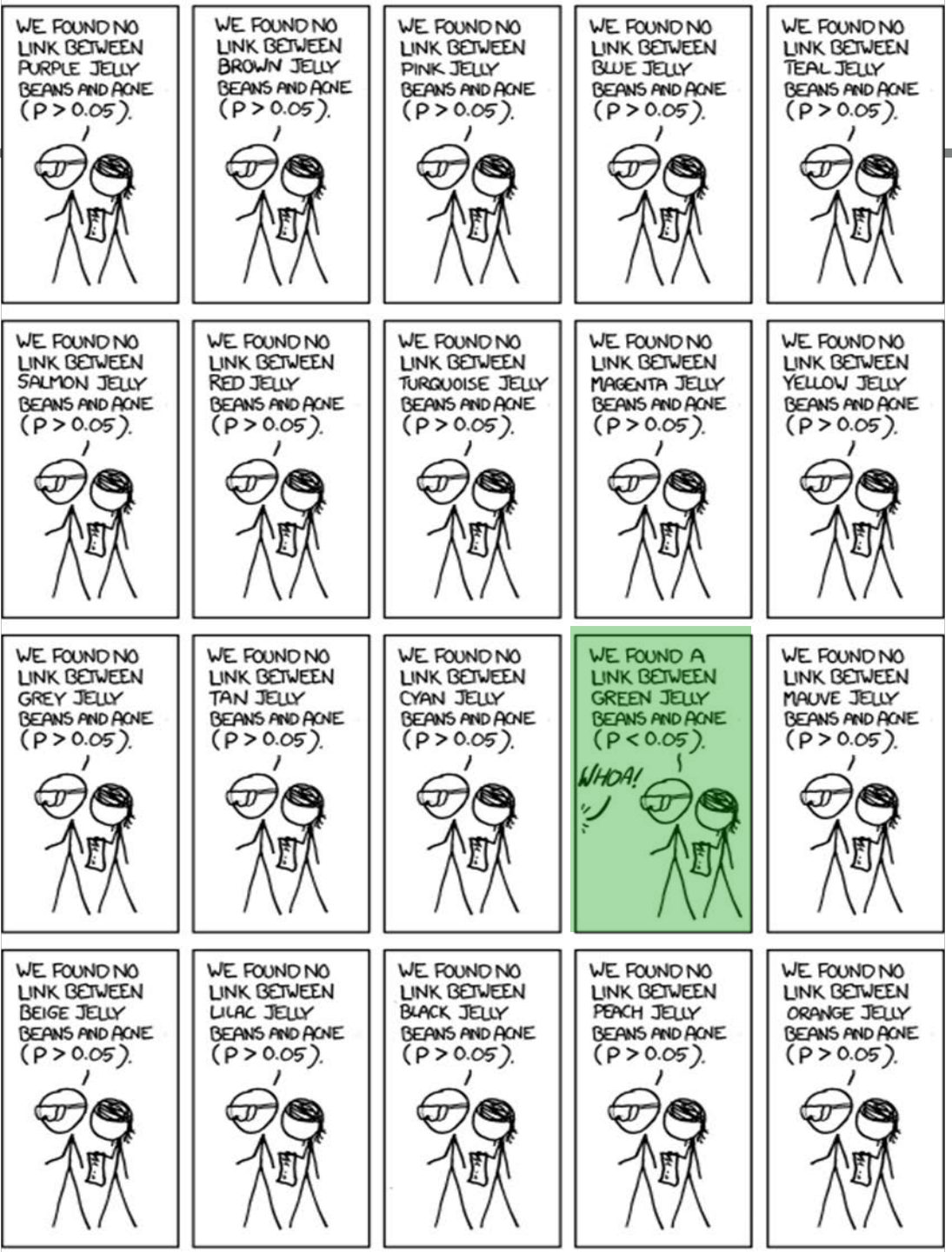
Content

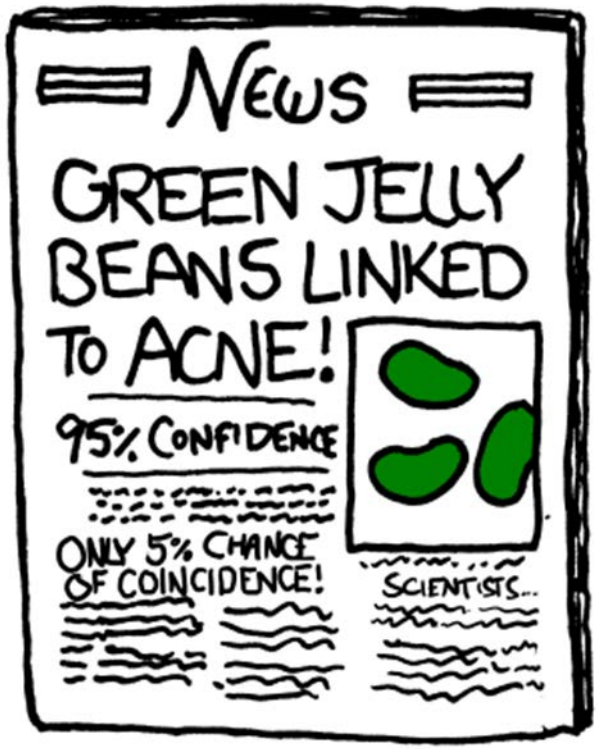
1. Introduction Multiple testing
2. FWER
3. Bonferroni & Holm
4. Permutations
5. R-code
6. Summary

Introduction
Multiple Testing



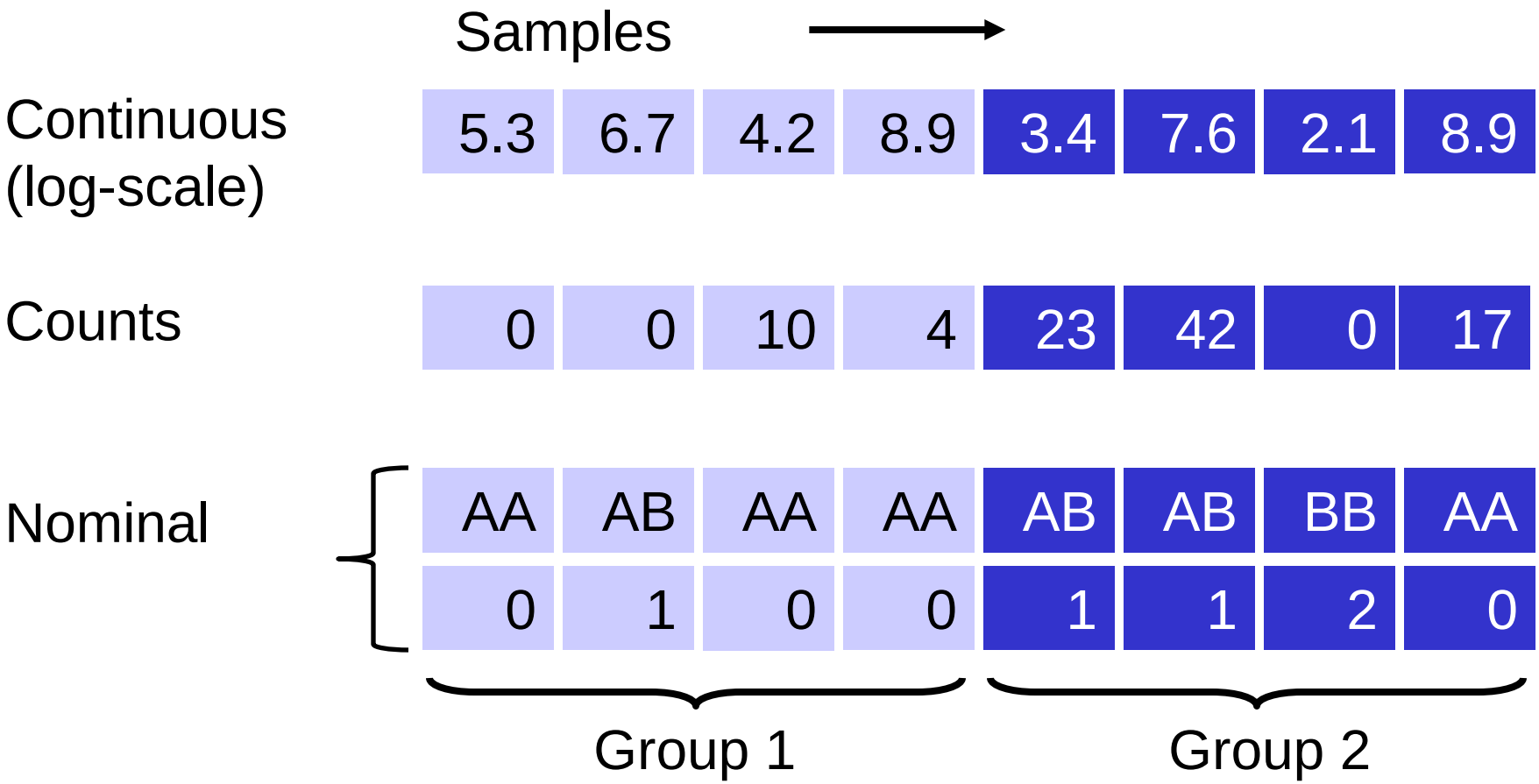
Source: http://flowingdata.com/2011/04/08/statistical-significance-on-xkcd/


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).


NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Differential expression

Data types

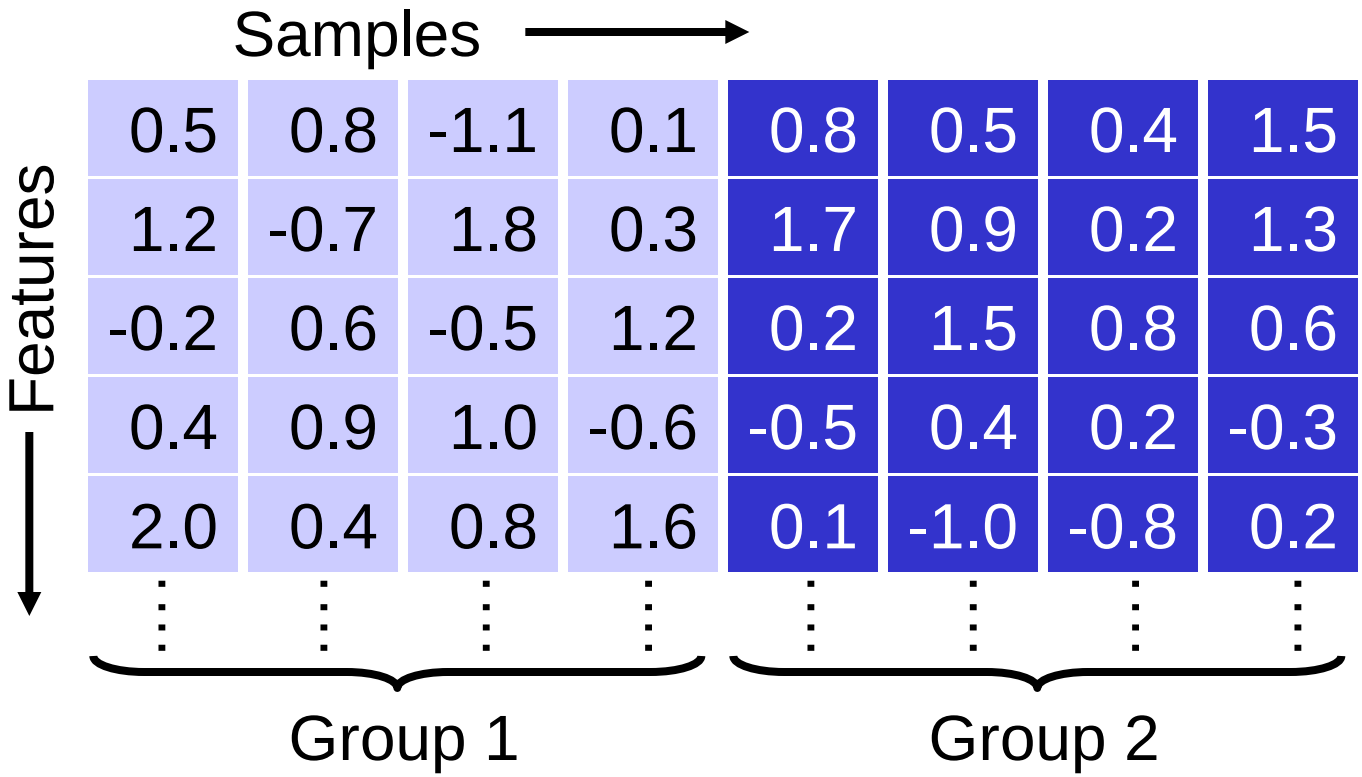


Which features are differential between groups?

Differential expression

Comparative microarray experiment

Given: matrix of (normalized) expression values



Differential expression

Top 5 genes out of 20,000

Gene	p-value
OCIAD2	5.5e-6
NEK3	6.7e-6
TAF5	7.1e-6
FOXD4L6	7.5e-6
ADIG	8.8e-6
⋮	⋮

Small p-value?

- Getting a p-value as small as 5.5e-6 is unlikely
- But is it also small if we admit that we tried 20,000 times?
- Can we reliably state that OCIAD2 is differentially expressed?
- What about NEK3?

Multiple testing

How does multiple testing differ from single testing?

1. p-values not sufficient to control false positive rate:

   $p \leq 0.05$ implies too many false positives

2. Different error control desirable when # tests large:

   control proportion of false positives instead of P(at least 1 false positive)

3. Opportunity to learn from other features:

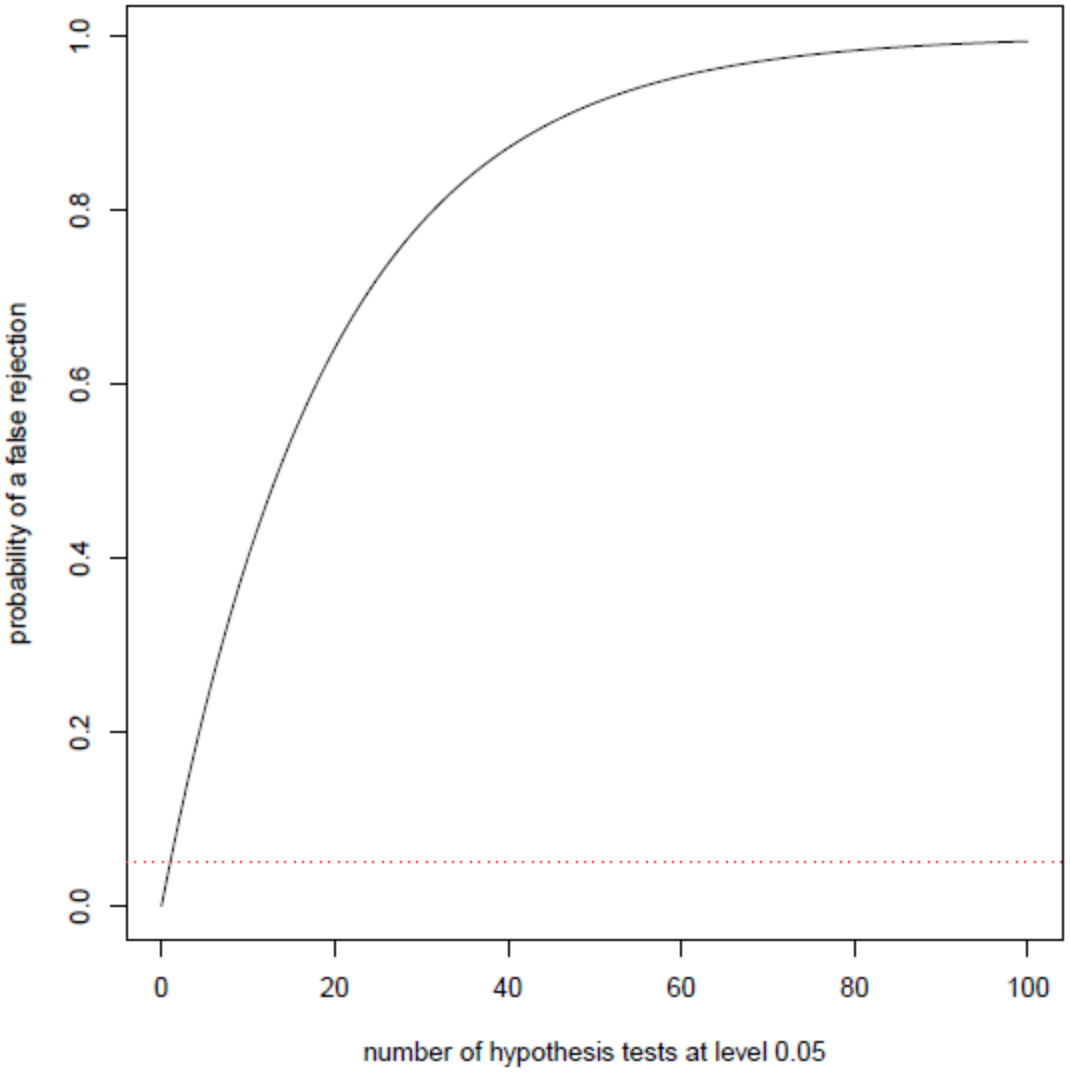
   When using a t-test: estimate the s.d. from all features (other lecture: limma)

P-values

- Test m null-hyposthesis: H_{01}, H_{02} , ..., H_{0m}
- Renders m p-values: p_1 ,...., p_m
- Property of p-values: $P_0(p_i \leq \alpha) = \alpha$ (or $\leq \alpha$) [proof: see hand-out]
- V: Number of false positives
- Under independence [tut]:

$$P(V>0) = 1-(1- \alpha)^{m}{}_{0}$$

Too many false positives



Multiple testing setting

	True Null	False Null	
Rejected	V	U	R
Non-rejected	$m_0 - V$	$m_1 - U$	m-R
	m_0	m_1	m

Family-wise error rate (FWER)

- FWER: probability of making at least one false rejection, so

$$FWER = P(V > 0)$$

- If m = 1 (one test), FWER reduces to type I error rate

Bonferroni

Bonferroni's solution

1. Multiply p-values with m, the number of tests
2. Reject null-hypothesis for feature i when

$$p_i^{bonf} = m \times p_i \leq \alpha \text{ (e.g. } \alpha = 0.05)$$

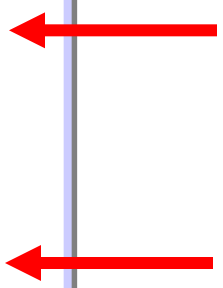
Under arbitrary dependency structure:

Reject H_{0i} when $p_i^{bonf} \leq \alpha$ implies FWER $\leq$ α [exer]

Bonferroni

- m = 5
- α = 0.05.

Gene	p-values p_i	Bonferroni p_i^{Bonf}	Reject H_0
1	0.006	$p_i \times m = 0.03$	Yes
2	0.372	$p_i \times m > 1$	No
3	0.012	$p_i \times m = 0.06$	No
4	0.000	$p_i \times m = 0.00$	Yes
5	0.811	$p_i \times m > 1$	No



Gene 1 and 4 are declared differentially expressed.

Bonferroni

- Now suppose m=1000
- Consider the same 5 genes
- $\alpha = 0.05$.

Gene	p-values	Bonferroni rejection level	Reject H_0
1	0.006	$p_i \times m > 1$	No
2	0.372	$p_i \times m > 1$	No
3	0.012	$p_i \times m > 1$	No
4	0.000	$p_i \times m = 0.00$	Yes
5	0.811	$p_i \times m > 1$	No

Only gene 4 is declared differentially expressed.

Holm

Holm's sequential procedure

1. Reject all null-hypotheses with p-value $p_i \leq \alpha/m$

2. If r null-hypotheses are rejected, reject all null-hypotheses with p-value $p_i \leq \alpha/(m-r)$

3. Continu until no rejections are done

Holm's sequential procedure, adjusted p-values

1. $p_{(1)}^{Holm} = m\, p_{(1)}$
2. $p_{(i)}^{Holm} = \max(p_{(i-1)}, (m-(i-1))\, p_{(i)})$
3. Reject when $p_{(i)}^{Holm} \leq \alpha$

Holm

Theorem

Under *arbitrary* dependency structure:

Reject H_{0i} when $p_i^{Holm} \leq \alpha$ implies FWER $\leq \alpha$

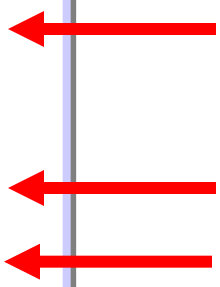
Proof

See hand-out

Holm

- m = 5
- α = 0.05.

Gene	p-values p_i	Holm p_i^{Holm}	Reject H_0
1	0.006	p_i x 4= 0.024	Yes
2	0.372	p_i x 2 > 0.744	No
3	0.012	p_i x 3= 0.036	Yes
4	0.000	p_i x 5 = 0.00	Yes
5	0.811	p_i x 1 > 0.811	No



Gene 1,3 and 4 are declared differentially expressed.

Holm

- Now suppose m=1000
- Consider the same 5 genes
- α = 0.05.

Gene	p-values	Rank	Holm p_i^{Holm}	Reject H_0
1	0.006	3	p_i x 998 > 1	No
2	0.372	455	p_i x 546 > 1	No
3	0.012	44	p_i x 957 > 1	No
4	0.000	1	p_i x m = 0.00	Yes
5	0.811	878	p_i x 123 > 1	No

Only gene 4 is declared differentially expressed.

Bonferroni vs Holm

- Bonferroni: very simple
- Holm is more powerful and valid under same assumptions,
- ... but for high-dimensional settings (large m): little gain by using Holm.

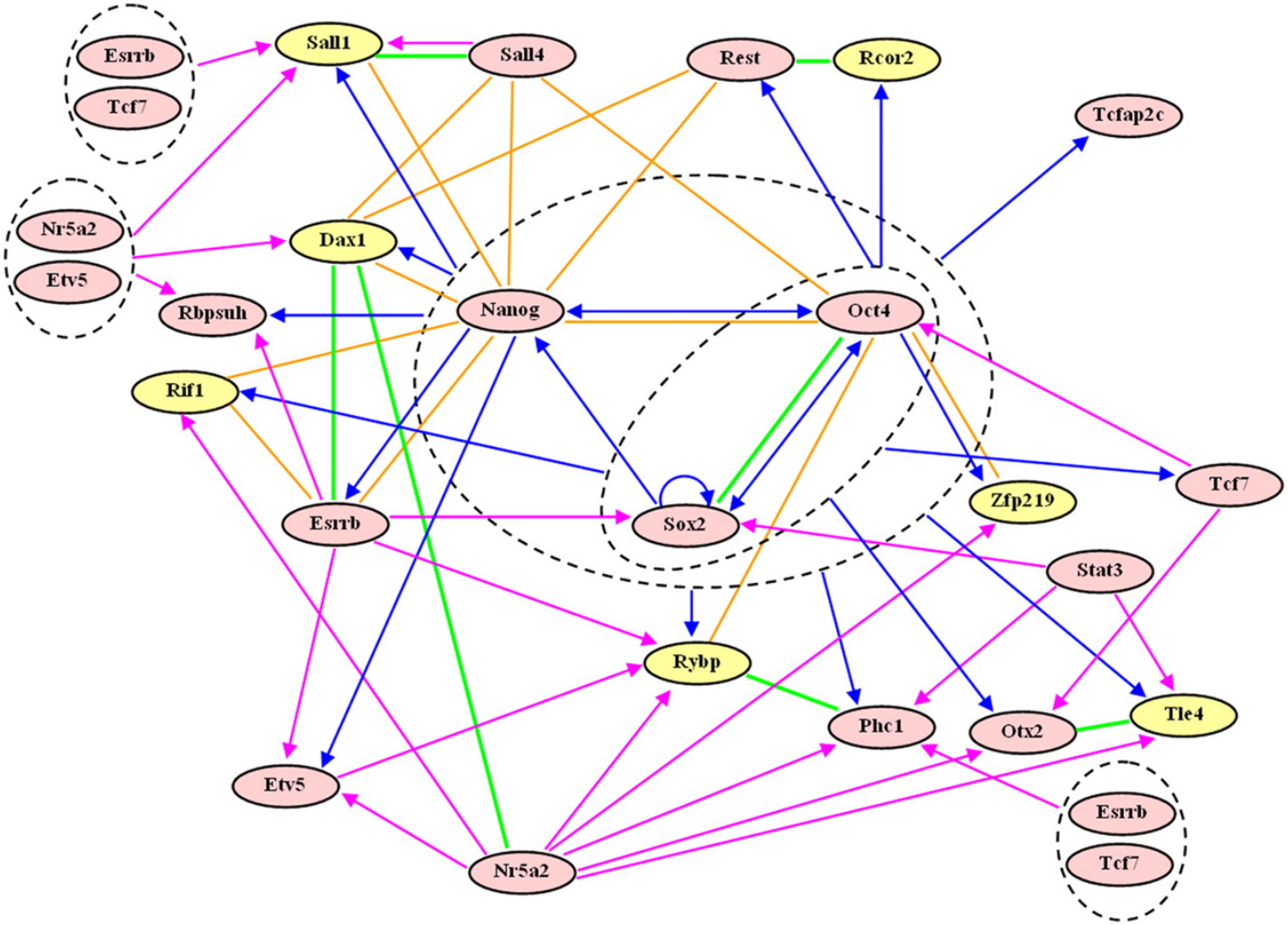
Bonferroni and Holm in R

```
load("C:\\Synchr\\Onderwijs\\HighDimensional\\Slides\\FDR\\Exercises\\pvals.Rdata")

pvaladjHolm <- p.adjust(pvals,method="holm")
pvaladjBonf <- p.adjust(pvals,method="bonferroni")

cbind(sort(pvaladjHolm)[1:50],
sort(pvaladjBonf)[1:50])
```

Dependence

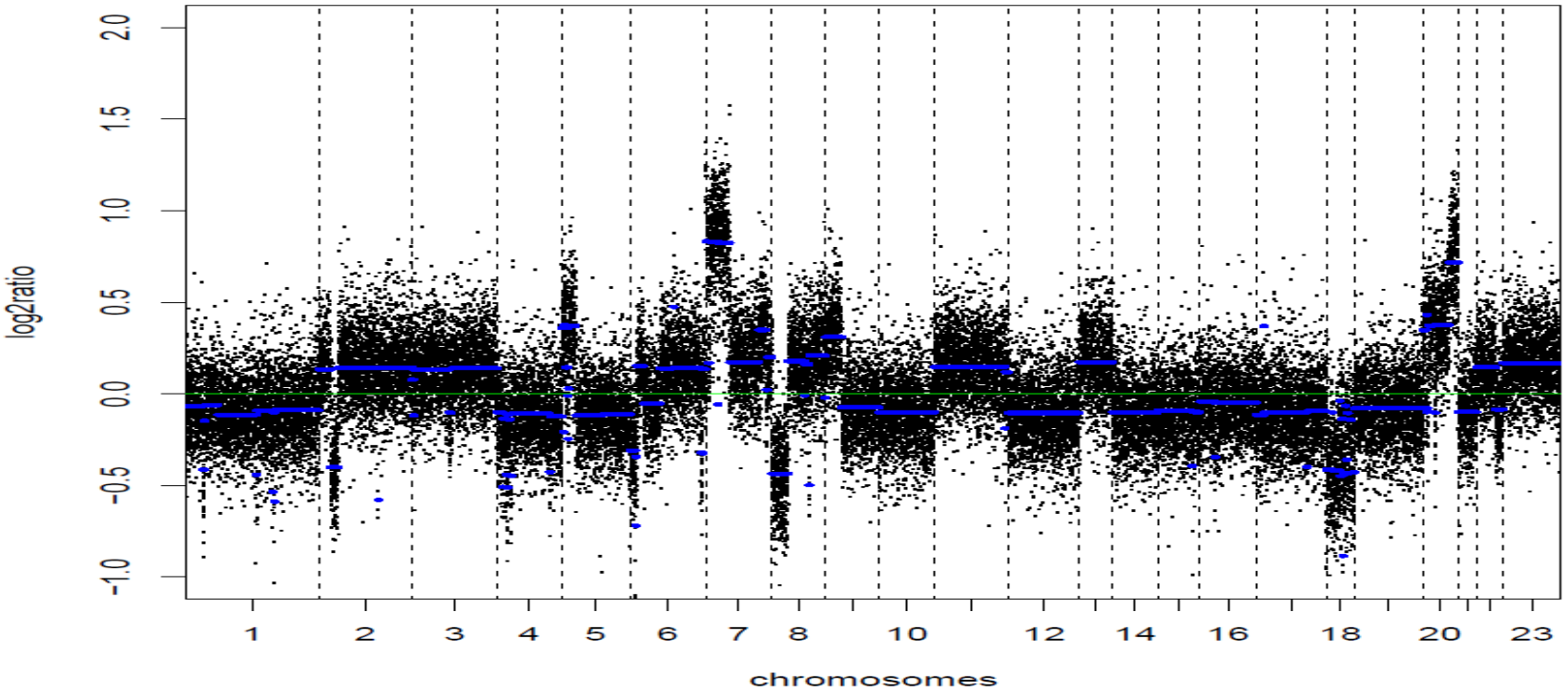


Genes are not independent, but collaborate in networks...

FWER under dependence

• Bonferroni and Holm are valid, but... likely to be conservative when many positive correlations are present.

• Very relevant in imaging data (neighboring pixels are extremely highly correlated), or DNA genomics data:



FWER under dependence

- For example, imagine m=100.000 variables, which are perfectly correlated withing blocks of 100.

- Bonferroni: $p_i^{bonf} = m \times p_i$,

- However, $p_i^{adj} = m/100 \times p_i$ would suffice for FWER control and is much less conservative

- m/100: effective dimension. Often hard to determine.

- Solution: permutation. Retains the correlation structure between hypotheses (genes).

FWER by permutation

X: gene matrix, Y: response (disease) labels

Original data (X,Y)

subject	gene1	gene2	disease
1	3.4	3.8	1
2	5.7	1.9	1
3	2.9	3.7	1
4	3.6	1.3	1
5	1.4	4.1	0
6	1.8	3.8	0
7	2.6	4.7	0
8	3.5	2.9	0

Permuted data (X,π(Y))

subject	gene1	gene2	disease
1	3.4	3.8	1
2	5.7	1.9	0
3	2.9	3.7	0
4	3.6	1.3	1
5	1.4	4.1	1
6	1.8	3.8	0
7	2.6	4.7	1
8	3.5	2.9	0

FWER by permutation

$FWER = P(V>0) = P_0(\min_{i=1,\ldots,m} p_i \leq \alpha')$

Find α' such that $FWER \leq \alpha$

→ Use permutations to find the distribution of

$$minP = \min_{i=1,\ldots,m} p_i$$

Exchangeability condition

The null-distribution of a test statistic T can be obtained by permutation *if*, under H_0, the distribution of (X, Y) is the same as for $(X,\pi(Y))$, where π is a random permutation of response labels Y.

FWER by permutation

Single-step algorithm, permutation-equivalent of Bonferroni

1. Initiate k=1.

   2. Iteration k: Randomly permute the labels Y
   3. Calculate new p-values for all tests based on permuted data
   4. Calculate the smallest p-value for permuted data: $\min P_k$

5. Repeat 2 - 4 many times (say, 10,000)
6. Calculate critical value α': the largest value of threshold t for which $P_0(\min P_k \leq t) \leq \alpha$
7. Reject all H_{0i} for which $p_i \leq \alpha'$.

FWER by permutation

Westfall & Young algorithm: permutation-equivalent of Holm

1. Start with all hypotheses

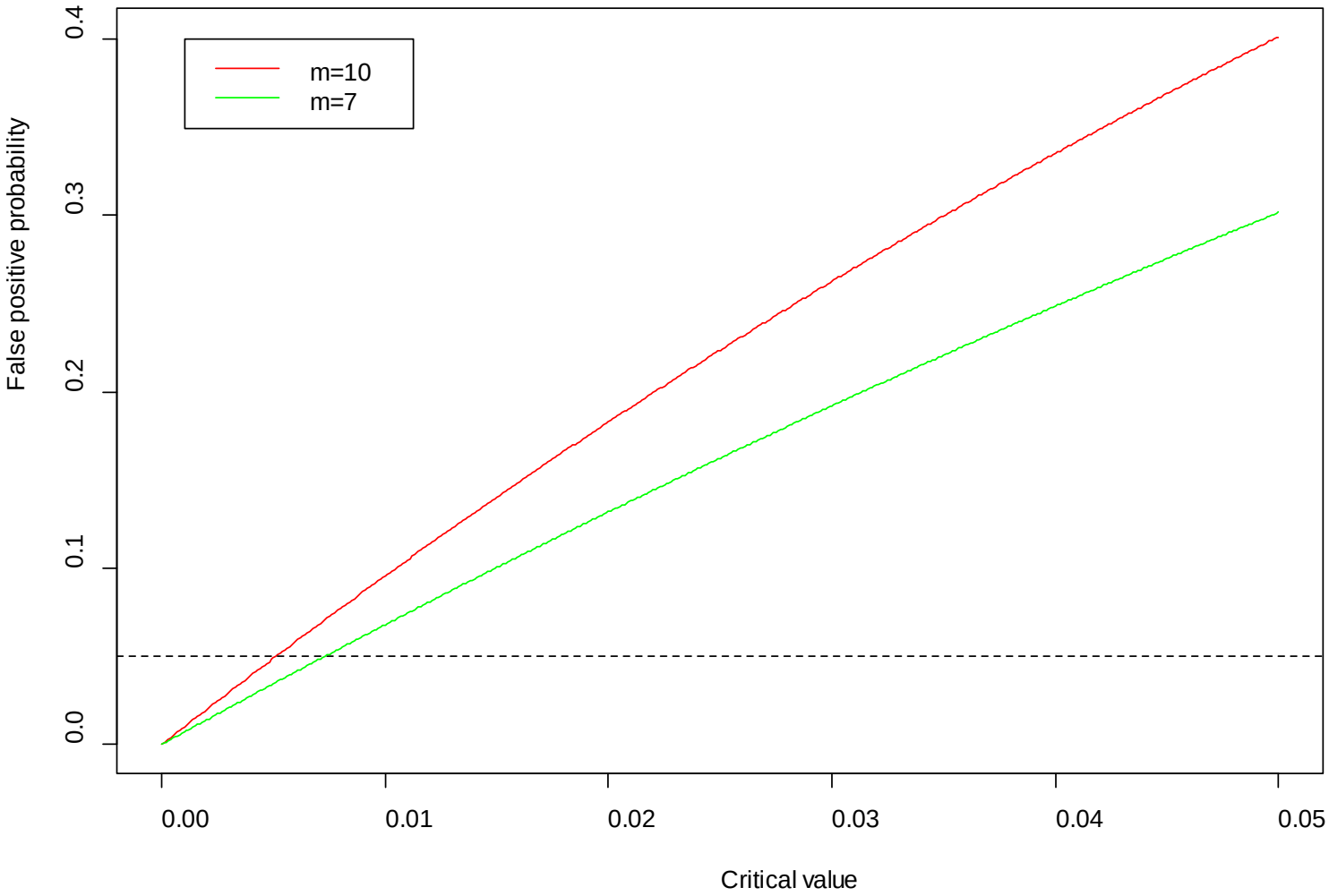
2. Repeat

    3. Single step minP to calculate α'
    4. Reject hypotheses with p-value $p_i \leq \alpha'$
    5. Remove rejected hypotheses

6. Stop when no more rejections occur

FWER by permutation

Why is W&Y more powerful than single-step?



FWER by permutation

Example

p_i
~~$p_1 = 0.001$~~
$p_2 = 0.009$
$p_3 = 0.604$
$p_4 = 0.211$
~~$p_5 = 0.002$~~
$p_6 = 0.843$
$p_7 = 0.402$
$p_8 = 0.041$

k=1	...	k=K=10^4
~~$p_{1,1}$~~	~~...~~	~~$p_{1,K}$~~
...	...	...
...	...	...
...	...	...
~~...~~	~~...~~	~~...~~
...	...	...
...	...	...
$p_{8,1}$	...	$p_{8,K}$
$minP_1$	...	$minP_K$

$minP_1$
.
.
.
$minP_K$

Sort →

$minP_{(1)}$
.
.
.
$minP_{(K)}$

→ Critical value α'= $minP_{(0.05K)}$ ↓ Say α' = 0.008

Note: Bonferroni gives α' = 0.05/8 = 0.00625

FWER by permutation

Example

p_i
Rejected
~~$p_1 = 0.009$~~
$p_2 = 0.604$
$p_3 = 0.211$
Rejected
$p_4 = 0.843$
$p_5 = 0.402$
$p_6 = 0.041$

k=1	...	k=K=10^4
$p_{1,1}$	...	$p_{1,K}$
...	...	...
...	...	...
...	...	...
...	...	...
$p_{6,1}$	...	$p_{6,K}$
$minP_1$	...	$minP_K$

$minP_1$
.
.
.
$minP_K$

Sort

$minP_{(1)}$
.
.
.
$minP_{(K)}$

Critical value α'= $minP_{(0.05K)}$

Say α' = 0.011

Summary

Familywise error rate (FWER)

- Generalizes Type I error to multiple hypotheses
- Limits the probability of any error among all inferences

FWER control methods

- Basic: Bonferroni
- Extension 1: Holm
- Extension 2: permutations
- Extension 1 & 2: Westfall & Young