Permutation

Jeff Leek

@jtleek

www.jtleek.com

Idea: Permute labels to “break relationship”

OPEN ACCESS Freely available online

PLOS MEDICINE

An Erythroid Differentiation Signature Predicts Response to Lenalidomide in Myelodysplastic Syndrome

Benjamin L. Ebert[1,2,3]☯, Naomi Galili[4]☯, Pablo Tamayo[1], Jocelyn Bosco[1,2], Raymond Mak[1,2], Jennifer Pretz[1,2], Shyam Tanguturi[1], Christine Ladd-Acosta[1], Richard Stone[2,3], Todd R. Golub[1,2,5,6], Azra Raza[4]*

1 Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **2** Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, United States of America, **3** Brigham and Women's Hospital, Department of Medicine, Boston, Massachusetts, United States of America, **4** St. Vincent's Comprehensive Cancer Center, New York, New York, United States of America, **5** Childrens's Hospital, Boston, Massachusetts, United States of America, **6** Howard Hughes Medical Institute, Chevy Chase, Maryland, United States of America

Response	R	R	•••	NR	NR

	Patient 1	Patient 2	•••	Patient n-1	Patient n
Gene 1	-1.64	-0.42	•••	-1.39	-0.38
Gene 2	-3.12	-3.60	•••	-3.80	-2.82
⋮	⋮	⋮	•••	⋮	⋮
⋮	⋮	⋮		⋮	⋮
⋮	⋮	⋮	•••	⋮	⋮
⋮	⋮	⋮		⋮	⋮

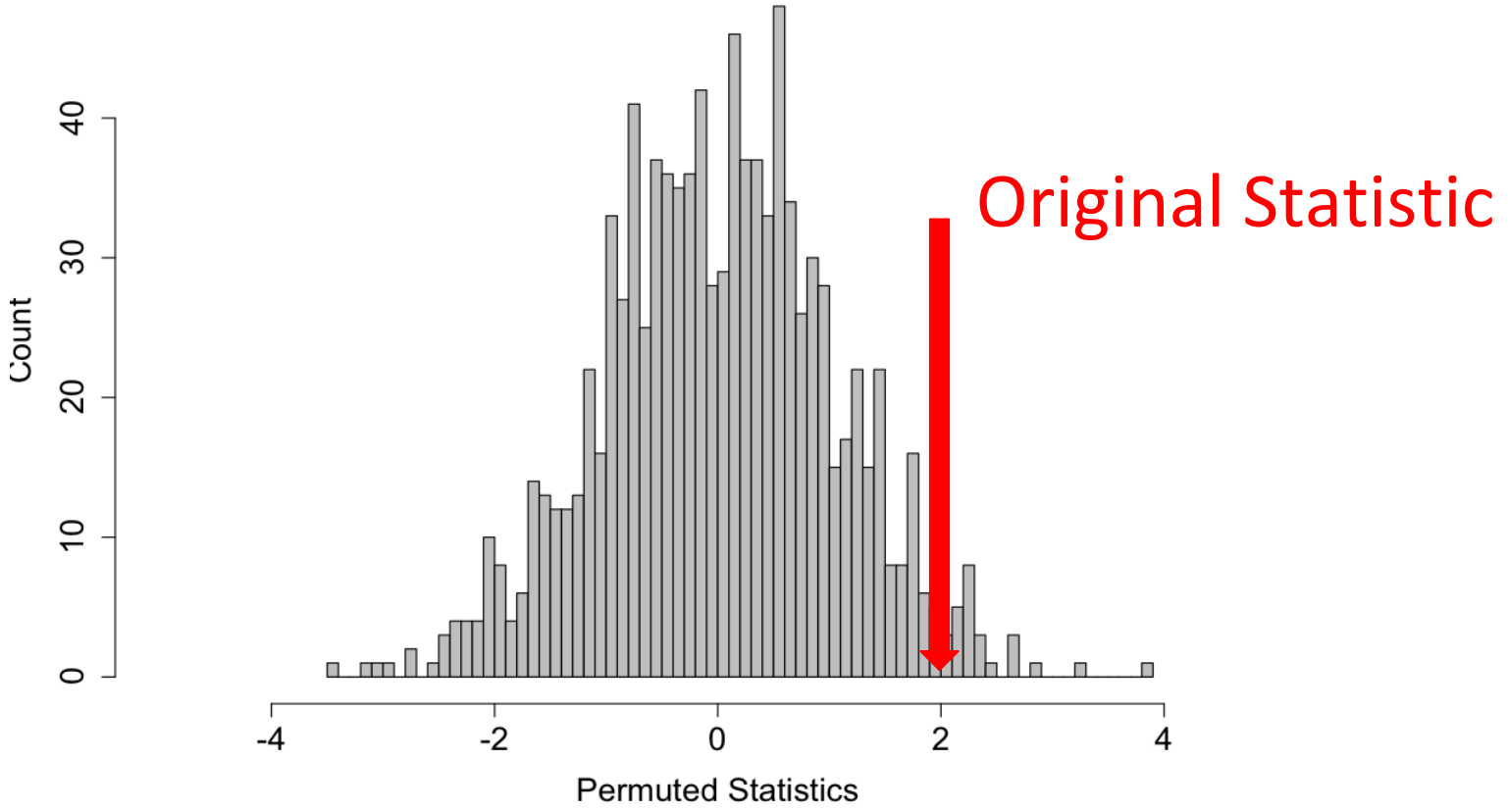
https://en.wikipedia.org/wiki/T-statistic

$$\frac{\overline{Y} - \overline{X}}{\sqrt{\frac{s_Y^2}{N} + \frac{s_X^2}{M}}}$$

Response	R	R	•••	NR	NR
	Patient 1	Patient 2	•••	Patient n-1	Patient n
Gene 1	-1.64	-0.42	•••	-1.39	-0.38
Gene 2	-3.12	-3.60	•••	-3.80	-2.82
:	:	:	•••	:	:
:	:	:		:	:
:	:	:	•••	:	:
:	:	:		:	:
Gene m-1	-2.34	-0.22	•••	-1.22	-2.76
Gene m	4.53	3.23	•••	0.29	3.11

Response	NR	R	•••	NR	R
	Patient 1	Patient 2	•••	Patient n-1	Patient n
Gene 1	-1.64	-0.42	•••	-1.39	-0.38
Gene 2	-3.12	-3.60	•••	-3.80	-2.82
:	:	:	•••	:	:
:	:	:		:	:
:	:	:	•••	:	:
:	:	:		:	:
Gene m-1	-2.34	-0.22	•••	-1.22	-2.76
Gene m	4.53	3.23	•••	0.29	3.11

Leaves the relationship between genes unchanged.


Original Statistic
Count
0
10
20
30
40
-4
-2
0
2
4
Permuted Statistics

Notes and further reading

- Permutation is used all the time!
- Assumes that if you switch the labels the data come from the exact same distribution - *not a comparison of means!*
- More information:
  - https://www.edx.org/course/statistics-r-life-sciences-harvardx-ph525-1x