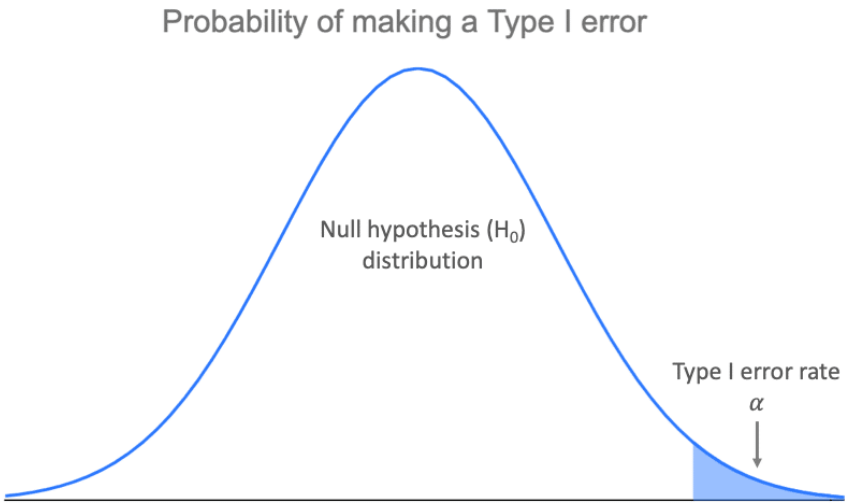
Multiple Testing Correction

Problem with running a large number of experiments, e.g. A/B Testing

Number of dice	Chance of getting 6
1	0.167
2	0.306
3	0.421
4	0.518
……	
25	0.999

If you repeat the experiment will the probability be the same?

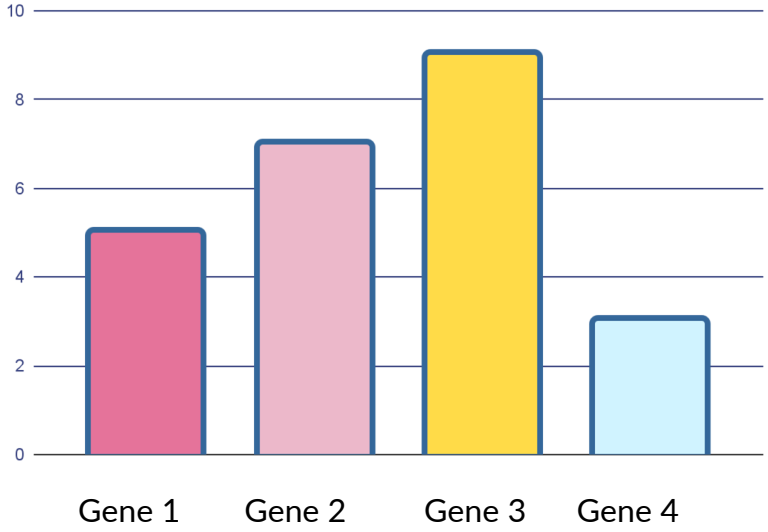


5% chance of error in 10,000 genes

=

500 genes of false positive

Multiple testing!



Gene	Gene 1	Gene 2	Gene 3
Gene 1			
Gene 2	√		
Gene 3	√	√	
Gene4	√	√	√

We need 6 tests!

Familywise Error Rate

Formula

$FWER = 1-(1-\alpha)^k$

FWER for 6 tests

$1-(1-0.05)^6 = 26.5\%$

FWER for 1 test

$FWER = 1-(1-0.05)^1 = 5\%$

It is a problem!

Controlling the familywise error rate: Bonferroni correction

α/n

Example:
α/6 = = 0.05/6 =0.008

We only reject a null hypothesis if the p-value is less than 0.008

Familywise Error Rate

Formula

$FWER = 1-(1-\alpha)^k$

FWER for 6 tests

$1-(1-0.05)^6 = 26.5\%$

FWER for 1 test

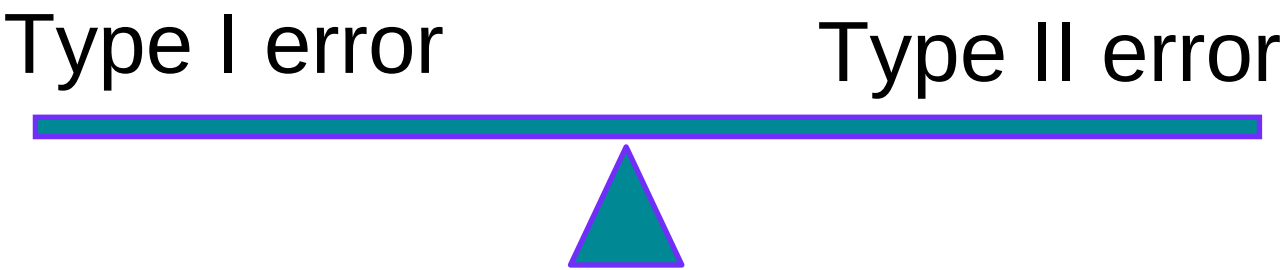
$FWER = 1-(1-0.05)^1 = 5\%$

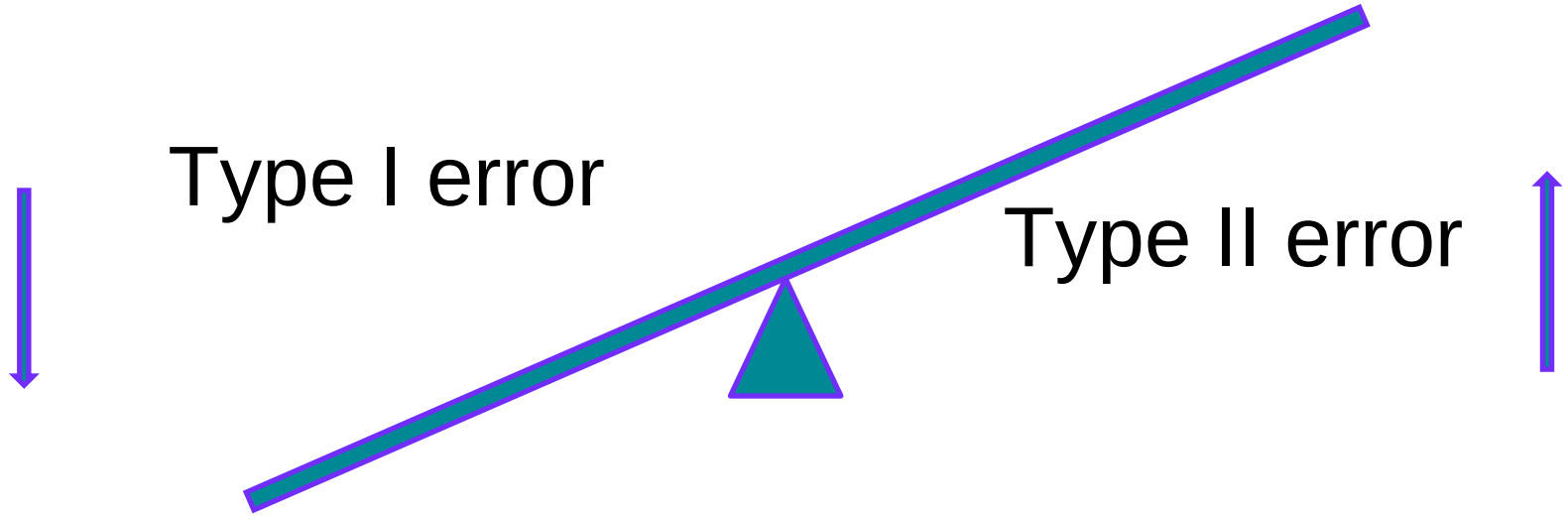
Corrected FWER for 6 tests

$1-(1-0.05/6)^6 = 4.9\%$

	Gene	α =0.05	α =0.008
1	Gene 1 vs Gene 2	0.98	0.348
2	Gene 1 vs Gene 3	.012	.111
3	Gene 1 vs Gene 4	.011	.003
4	Gene 2 vs Gene 3	.0.36	.493
5	Gene 2 vs Gene 4	0.02	.078
6	Gene 3 vs Gene 4	0.028	0.234

Do you think Bonferroni correction(α/n) is good for testing 10,000 genes?

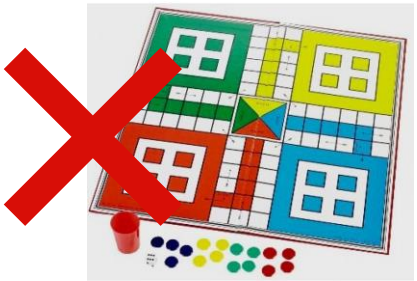

Type I error
Type II error


Type I error
Type II error

Controlling the false discovery rate: Benjamini-Hochberg procedure

i	variables	P value	(i/n)*Q
1	A	0.008	0.010
2	B	0.039	0.030
3	C	0.062	0.060
4	D	0.071	.0.071

Now, let's play!



Summary!

You cannot repeat your experiment

Bonferroni very strict
BH more Powerful

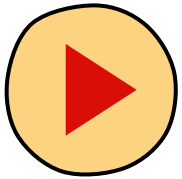
using suitable p-value adjustment method

you can skip p-value adjustment method

References

McDonald, J. H. (2009). *Handbook of biological statistics* (Vol. 2, pp. 6-59). Baltimore, MD: sparky house publishing.] Retrieved from Multiple comparisons - Handbook of Biological Statistics (biostathandbook.com)

Goldman, M. (2008). Spring 2008-stat c141/bioeng c141-statistics for bioinformatics. *Course Website*.] Retrieved from Section0402.pdf (berkeley.edu)



McNamara, A. (2020). *Correcting for multiple comparisons in R (STAT 320 week 10 video 5 of 5)* [Video]. Retrieved from https://www.youtube.com/watch?v=pevqgRKjlk4