TTEST and ANOVA

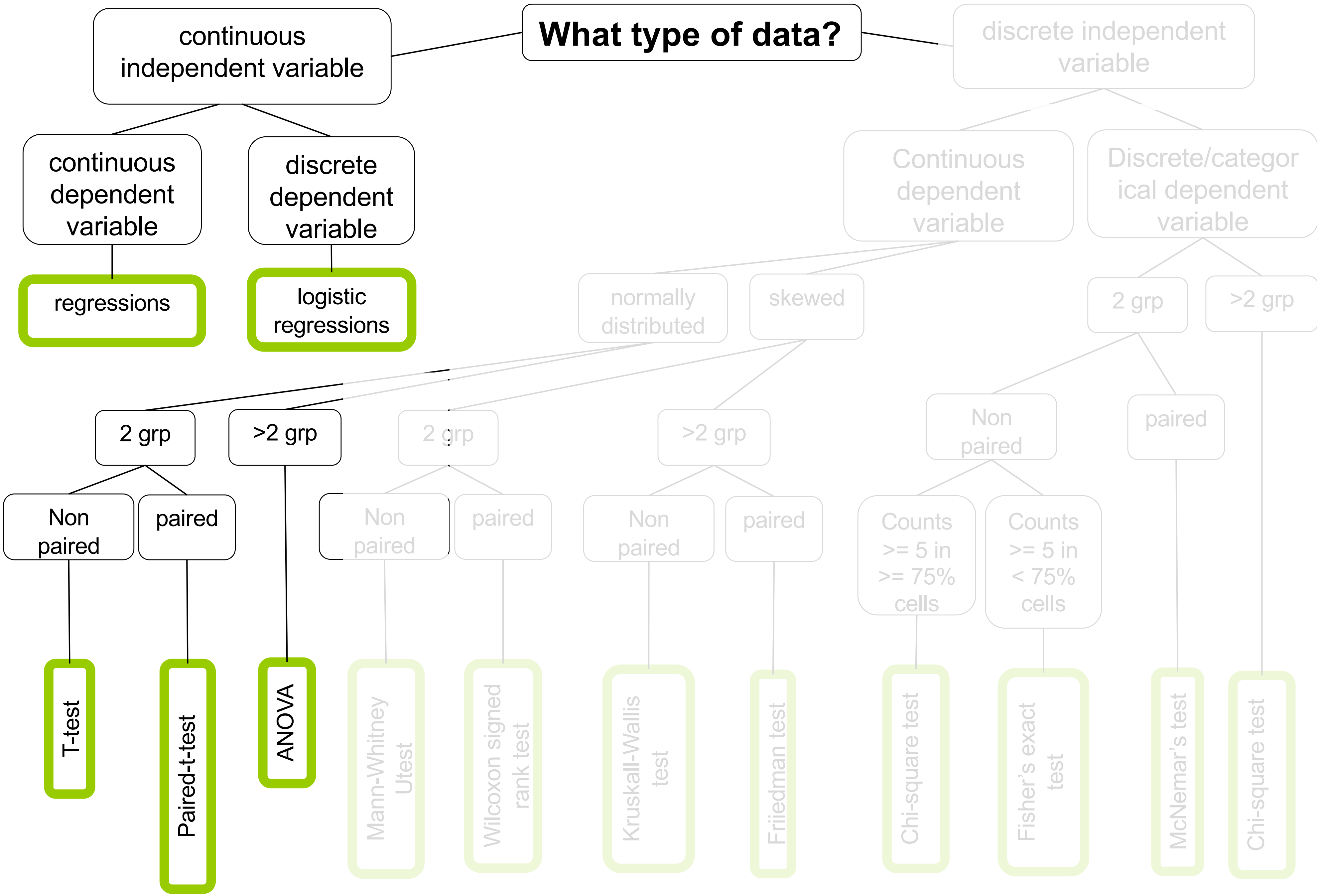
the theory

15

Probability and Statistics

COMS10011

Dr. Anne Roudaut
csxar@bristol.ac.uk

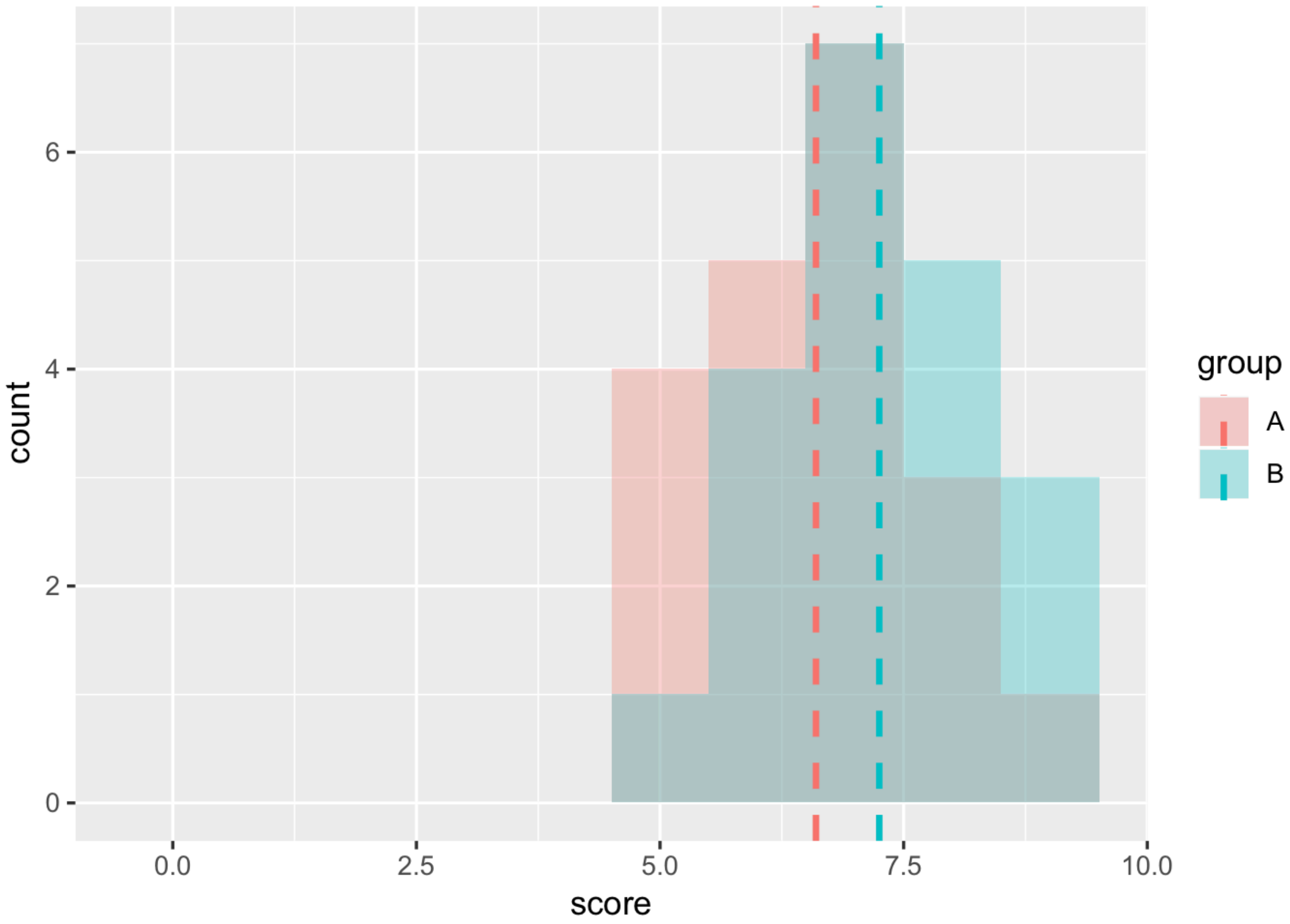


What type of data?
continuous independent variable
continuous dependent variable
discrete dependent variable
regressions
logistic regressions
2 grp
>2 grp
Non paired
paired
T-test
Paired-t-test
ANOVA
discrete independent variable
Continuous dependent variable
Discrete/categorical dependent variable
normally distributed
skewed
2 grp
>2 grp
Non paired
paired
Mann-Whitney Utest
Wilcoxon signed rank test
Kruskall-Wallis test
Friedman test
Counts >= 5 in >= 75% cells
Counts >= 5 in < 75% cells
Chi-square test
Fisher's exact test
McNemar's test
Chi-square test

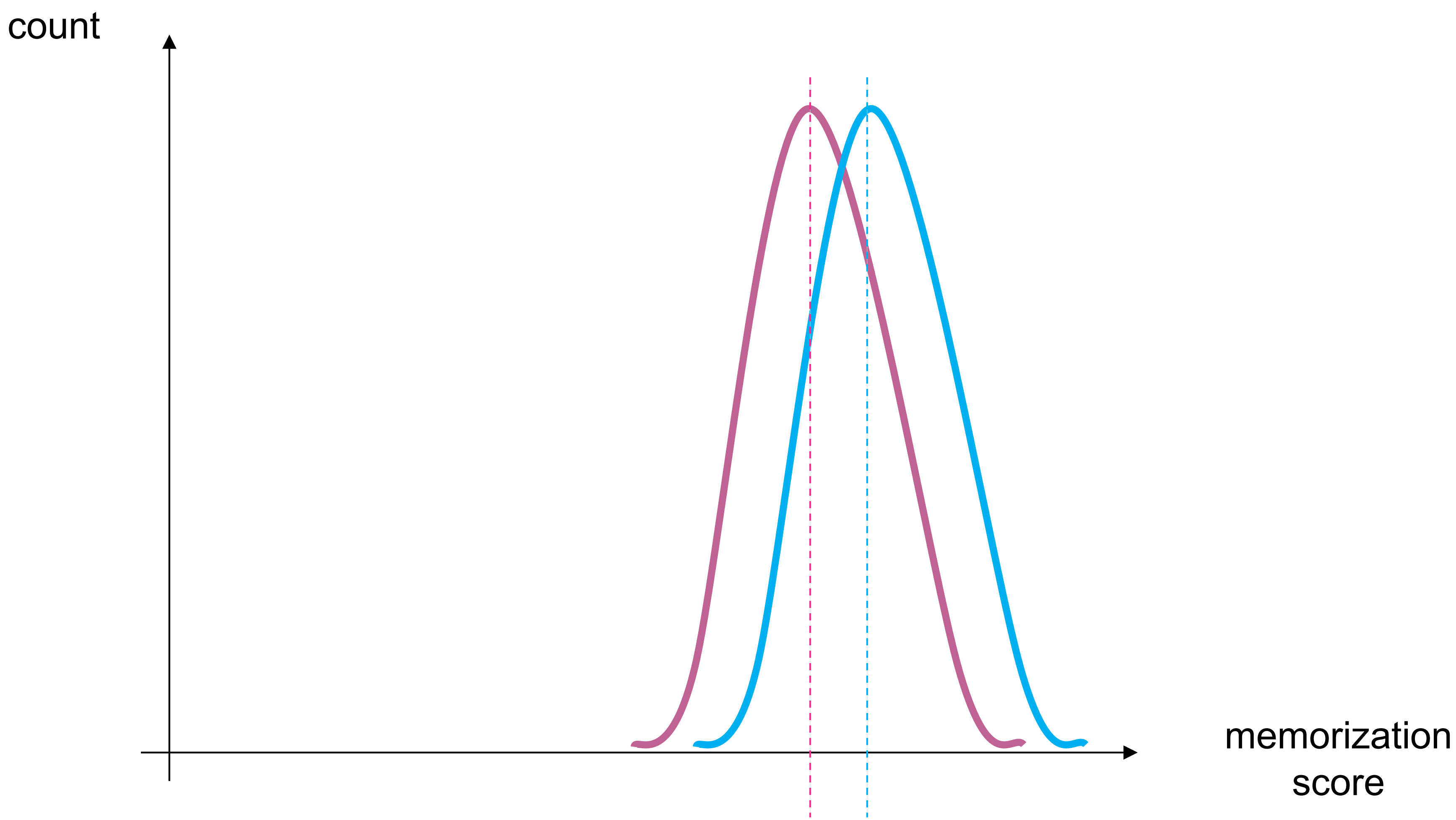
today::

look more into the math behind t-test and anova

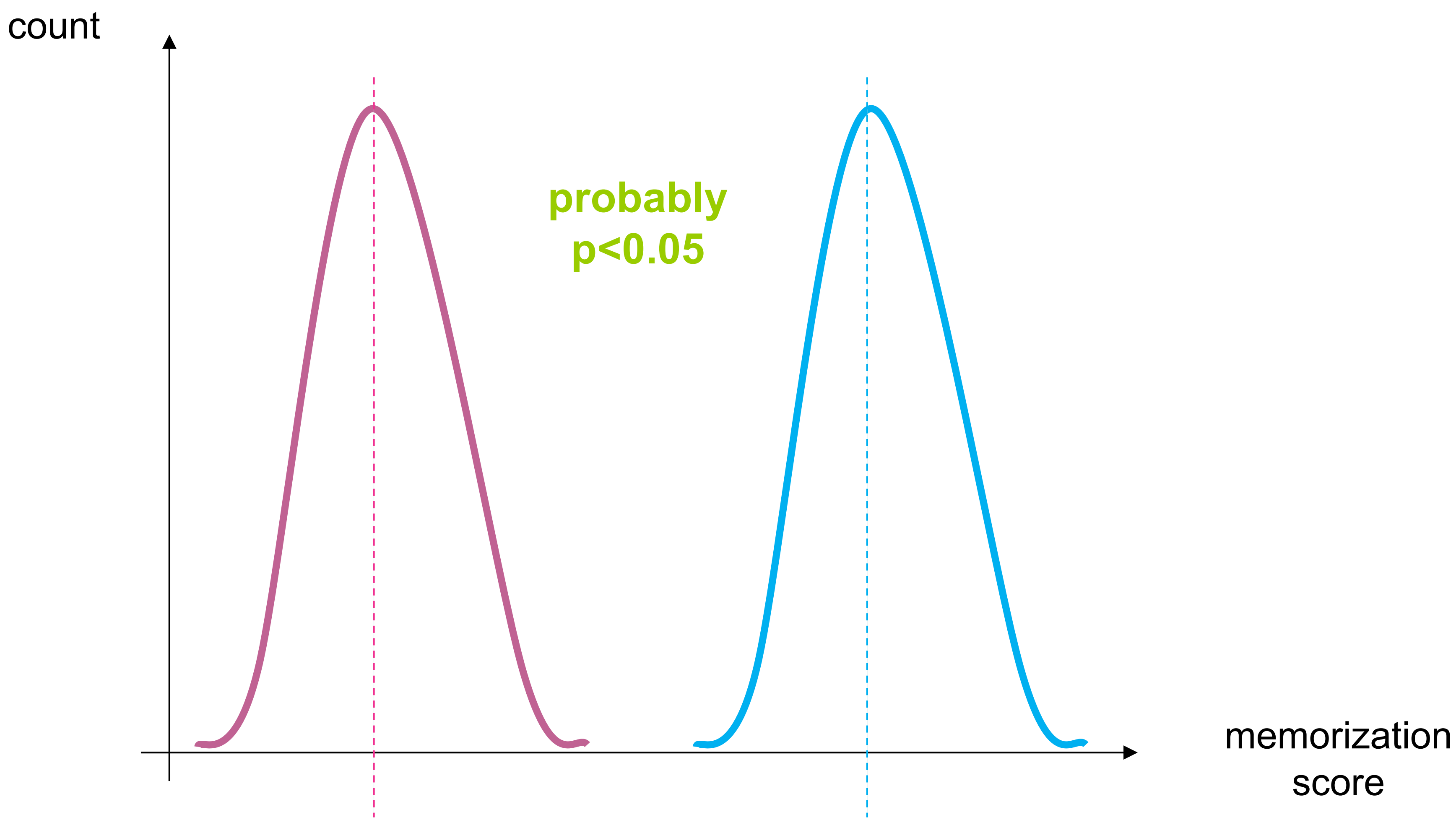
lets go back to
our memorization
experience



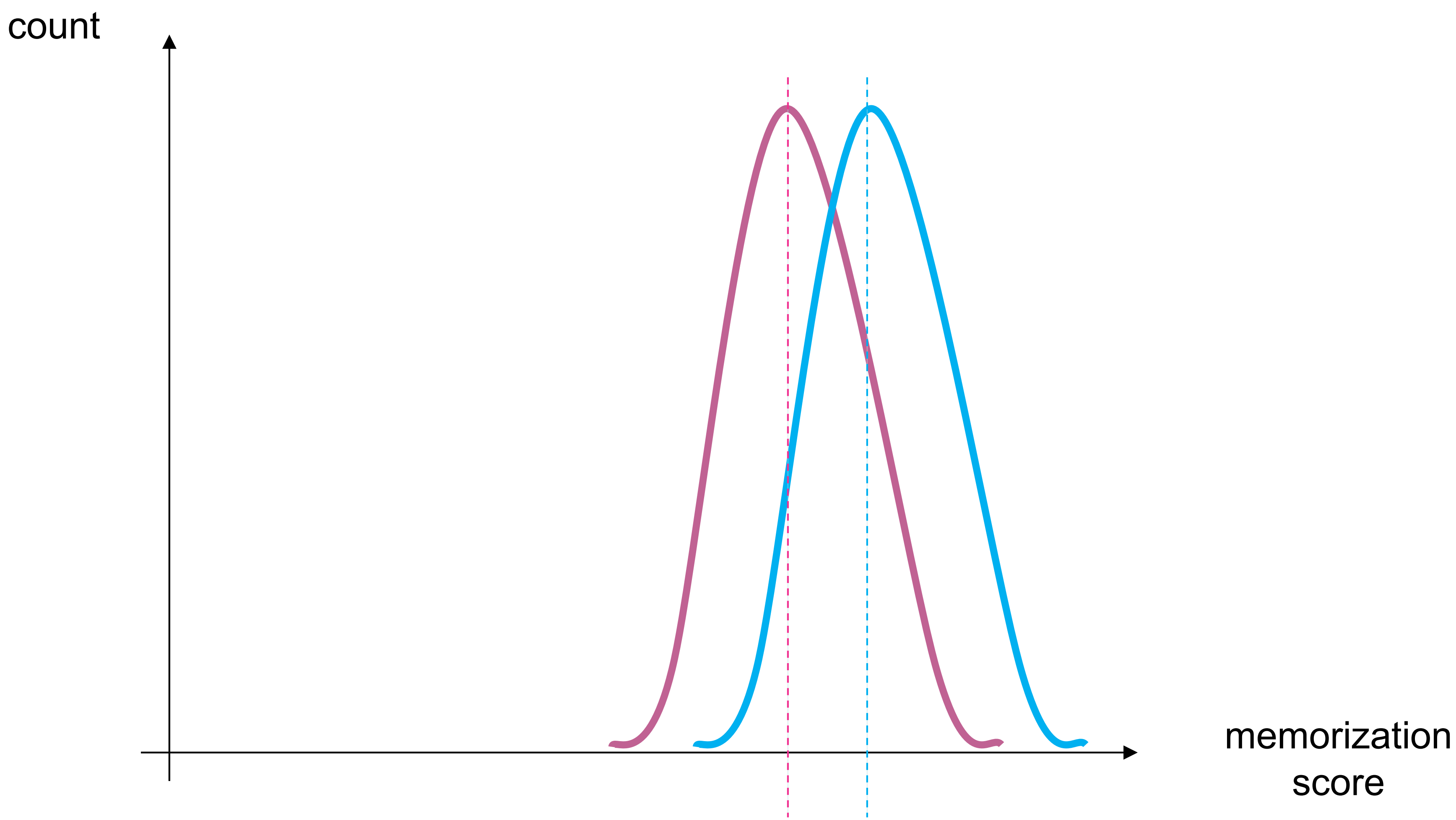
no evidences for chocolate vs. baseline (p>0.05)



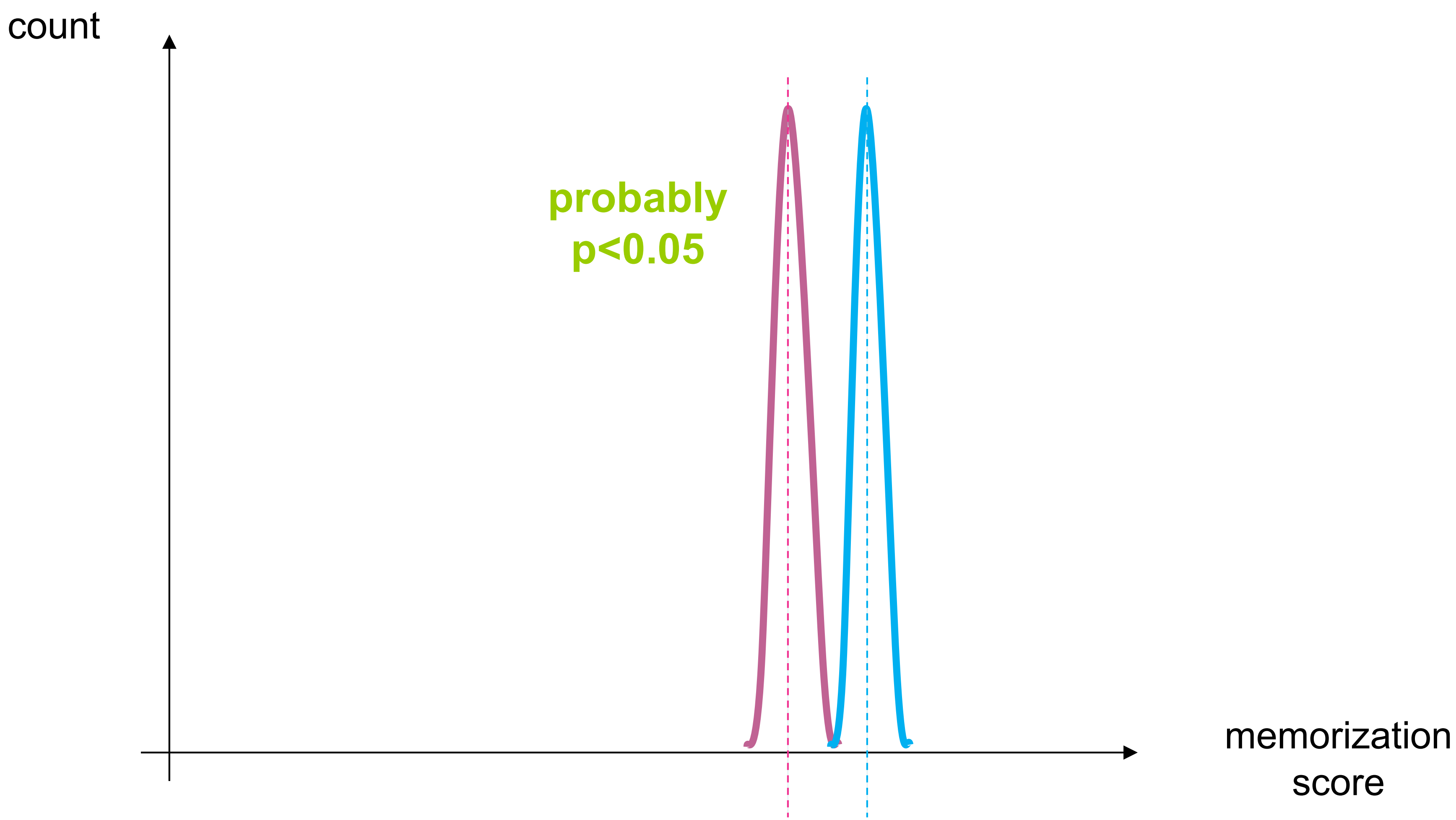
no evidences for chocolate vs. baseline (p>0.05)
(let’s just assume these are normally distributed)



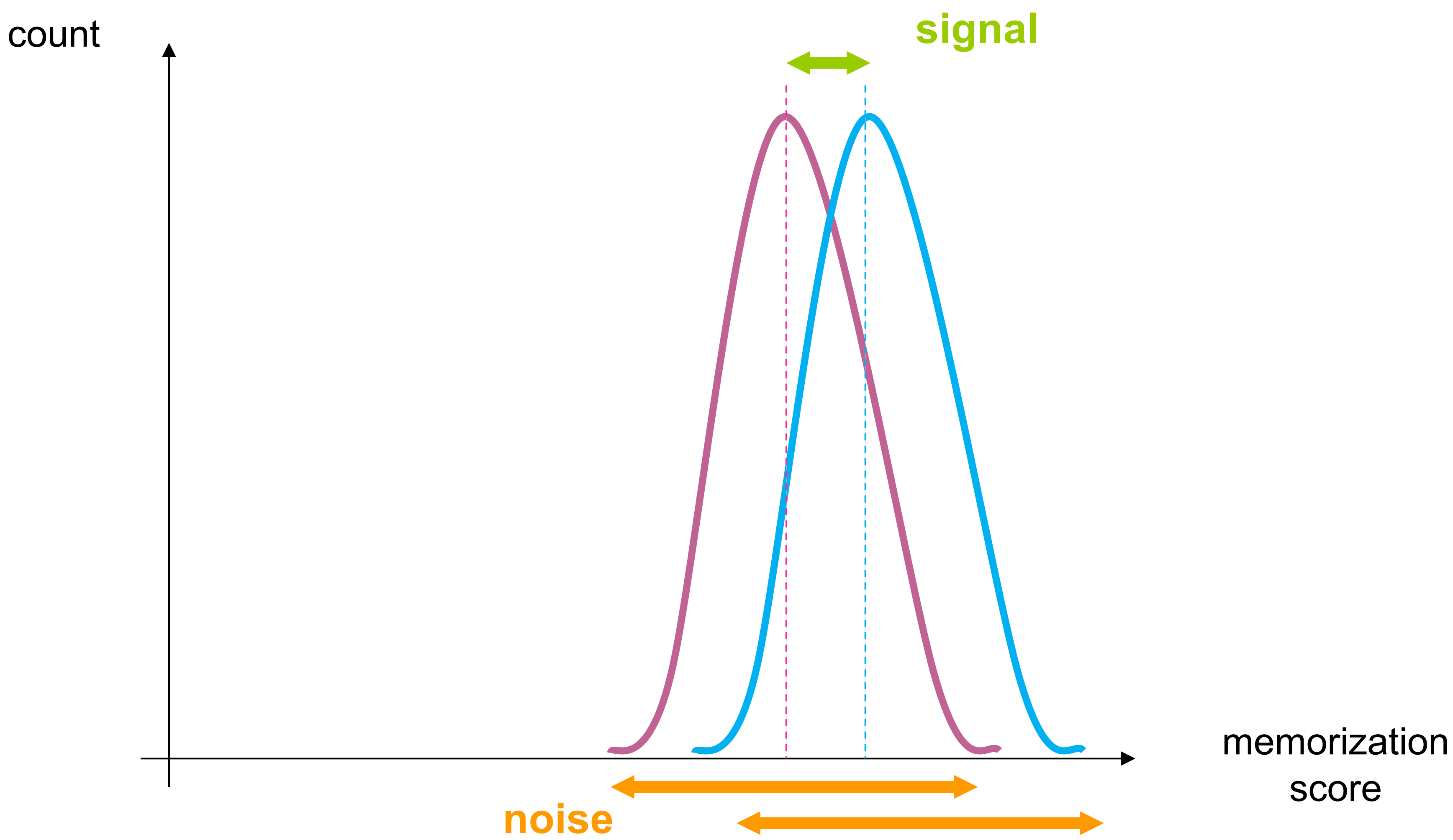
if the mean where further apart, it would increase our chances to have a $p<0.05$



<without changing the means, what else can we do to these data to make some more different?>



if the distributions where less spread out, it would increase our chances to have a $p<0.05$

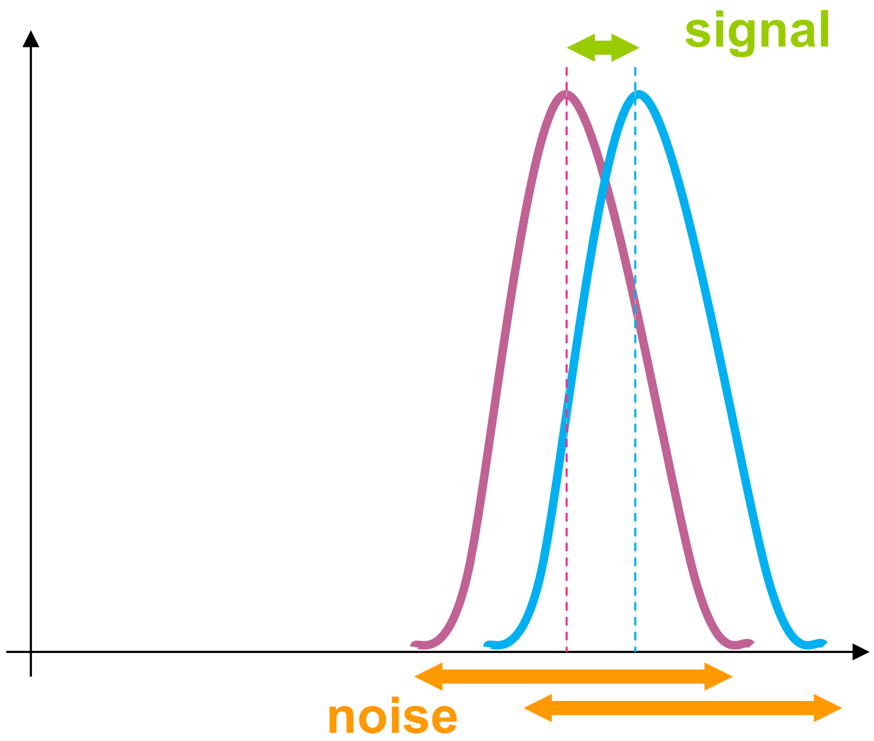


the goal of a study is to find
a signal in **a lot of noise**

any statistical tests ::

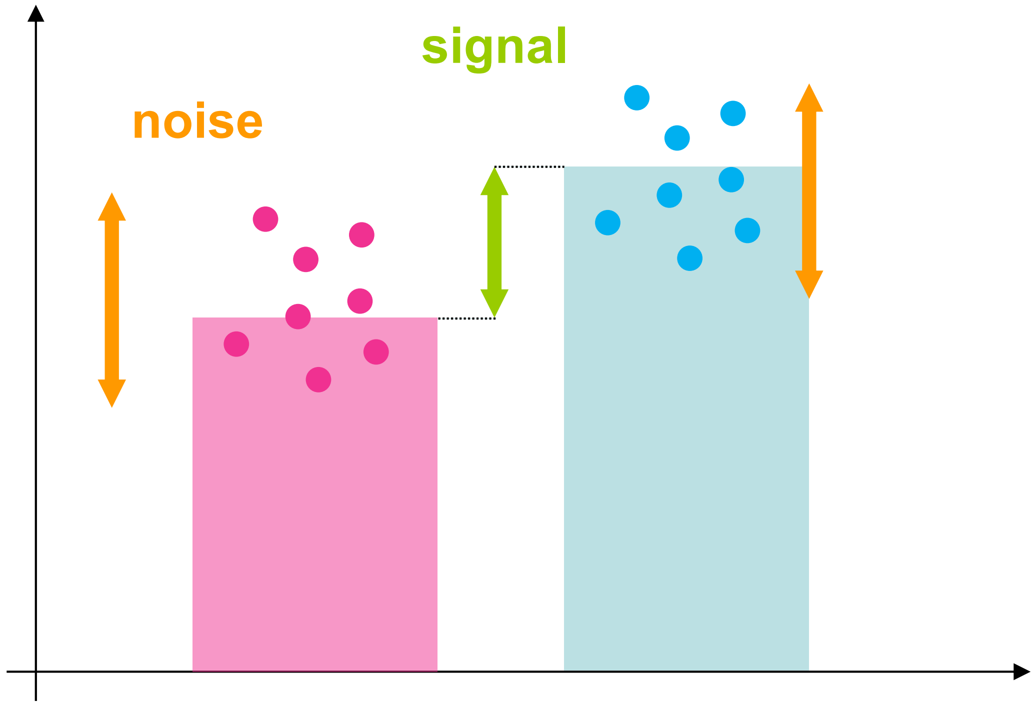
$$\frac{\text{signal}}{\text{noise}}$$

T-tests ::



$$\frac{\text{difference between group means}}{\text{variability of groups}}$$

T-tests ::



$$\frac{\text{difference between group means}}{\text{variability of groups}}$$

T-tests ::

Paired $$t = \frac{\overline{x1} - \overline{x2}}{s/\sqrt{n}}$$

Unpaired $$t = \frac{\overline{x1} - \overline{x2}}{s\sqrt{\left(\frac{1}{n1} + \frac{1}{n2}\right)}}$$

paired t-test

different between
group means
(to maximize)

$$t = \frac{\boxed{\overline{x1} - \overline{x2}}}{\boxed{s}/\sqrt{n}}$$

standard deviation
of the differences
(to minimize)

different between
group means

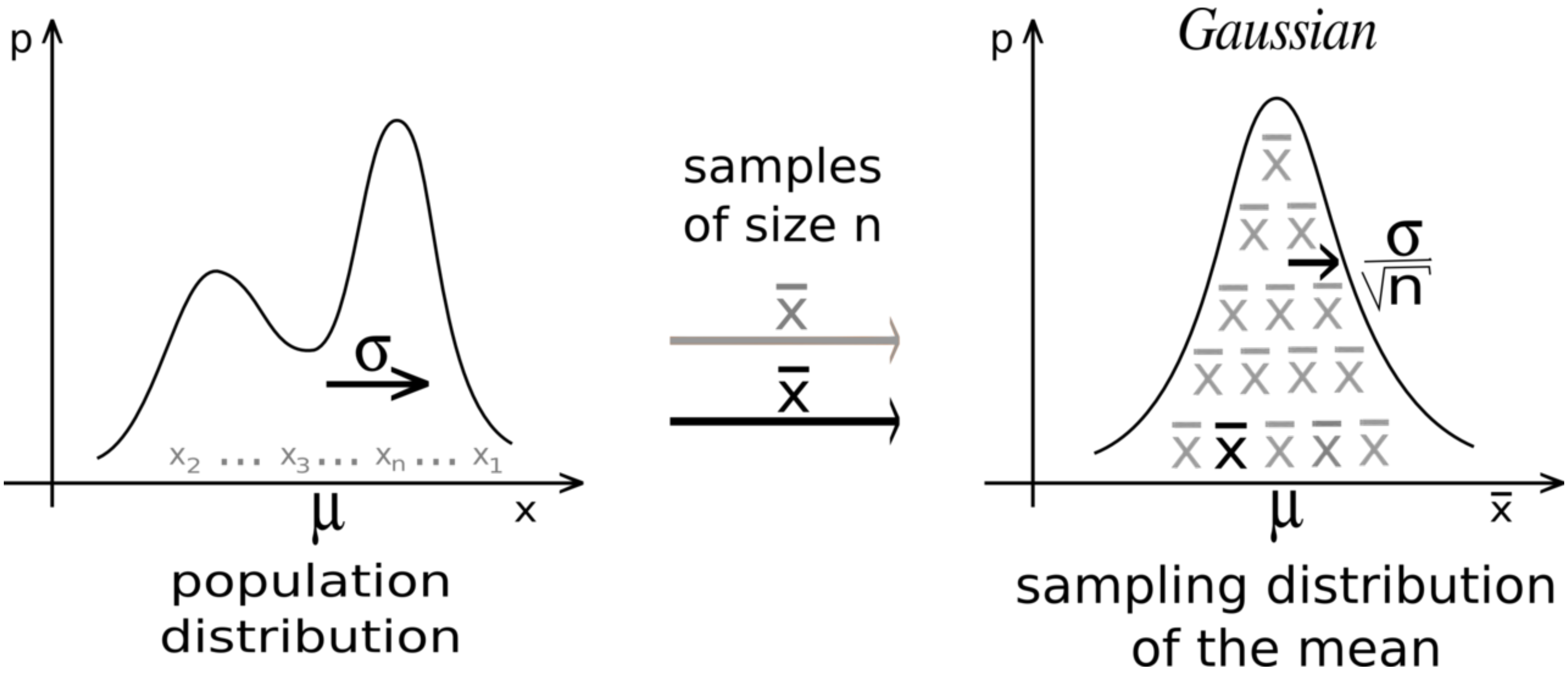
(to maximize)

$$t = \frac{\boxed{\overline{x1} - \overline{x2}}}{\boxed{s/\sqrt{n}}}$$

standard error of the mean

(to minimize)

but why do we need to divide by $\sqrt{n}$ (n = sample size)?



this comes from the **central limit theorem** you have seen before with Conor by dividing by $\sqrt{n}$, we add a "**penalty**" for using a sample instead of the entire population

penalty is large when sample is very small

as sample size increases, penalty diminishes …

… infinitely approaching point where sample = the population itself

$$\boxed{t} = \frac{\overline{x1} - \overline{x2}}{s/\sqrt{n}}$$

we get a t-value

(to maximize)

both signal and noise are in the units of your data

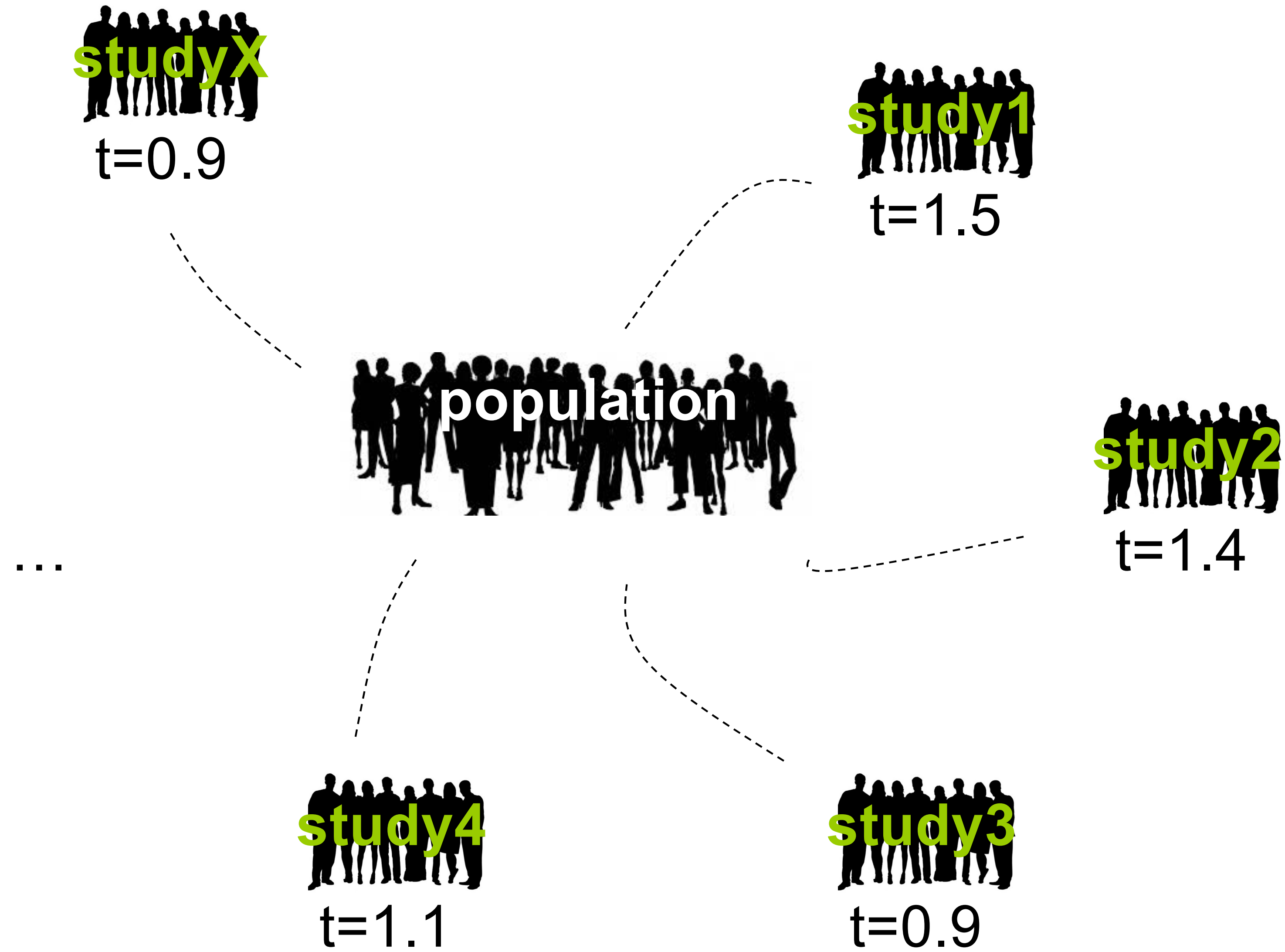
If signal = 6 and noise = 2, your t-value = 3, so the difference is 3 times the size of the standard error

If signal = 6 and noise = 6, your t-value = 1, the signal is at the same scale as the noise

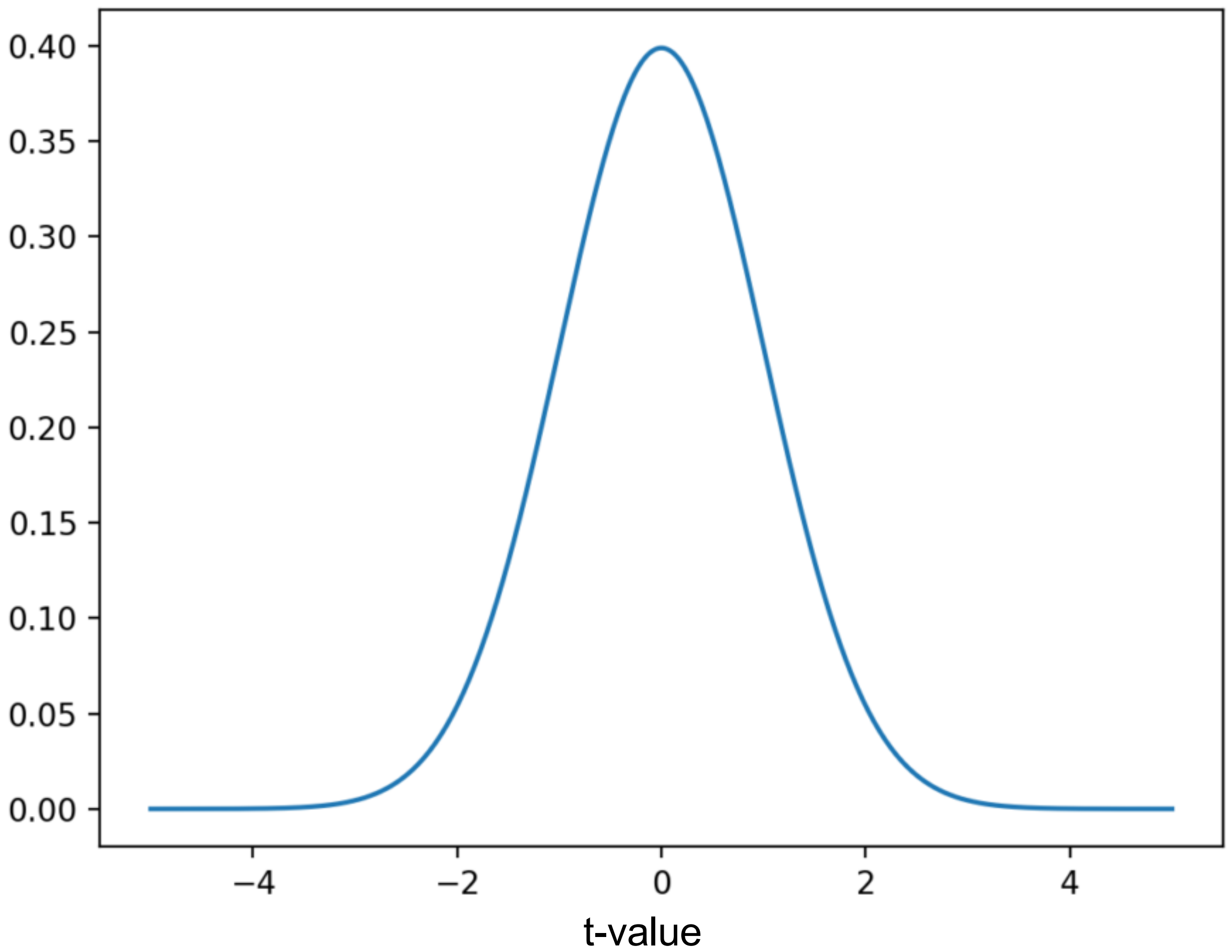
t-values = how distinguishable signal from noise

how do we know our t-value is any good, and how does this related to p-value?

this is where **t-distributions** come in



now let's take all these possible values and …

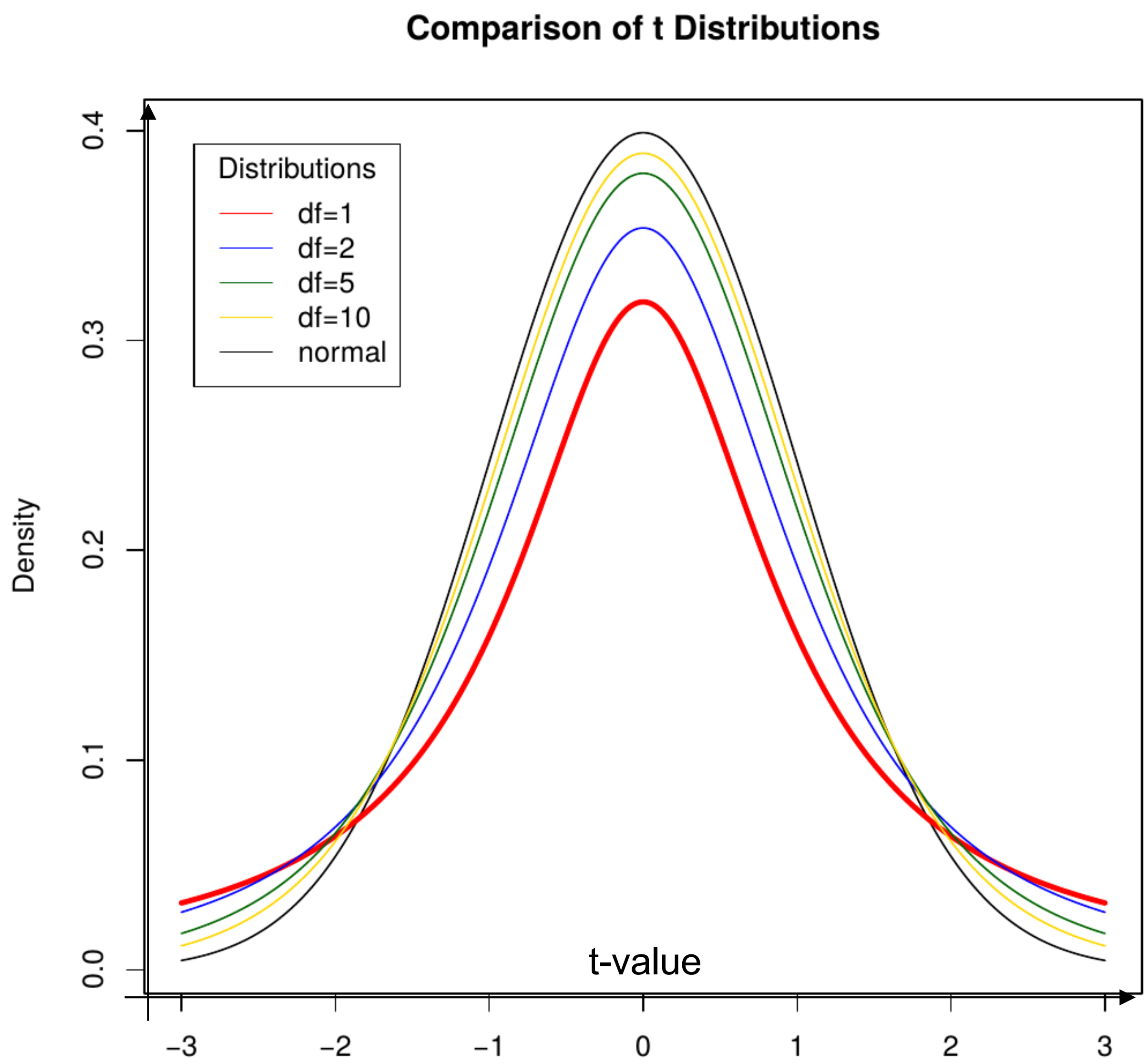


… plot a distribution of them

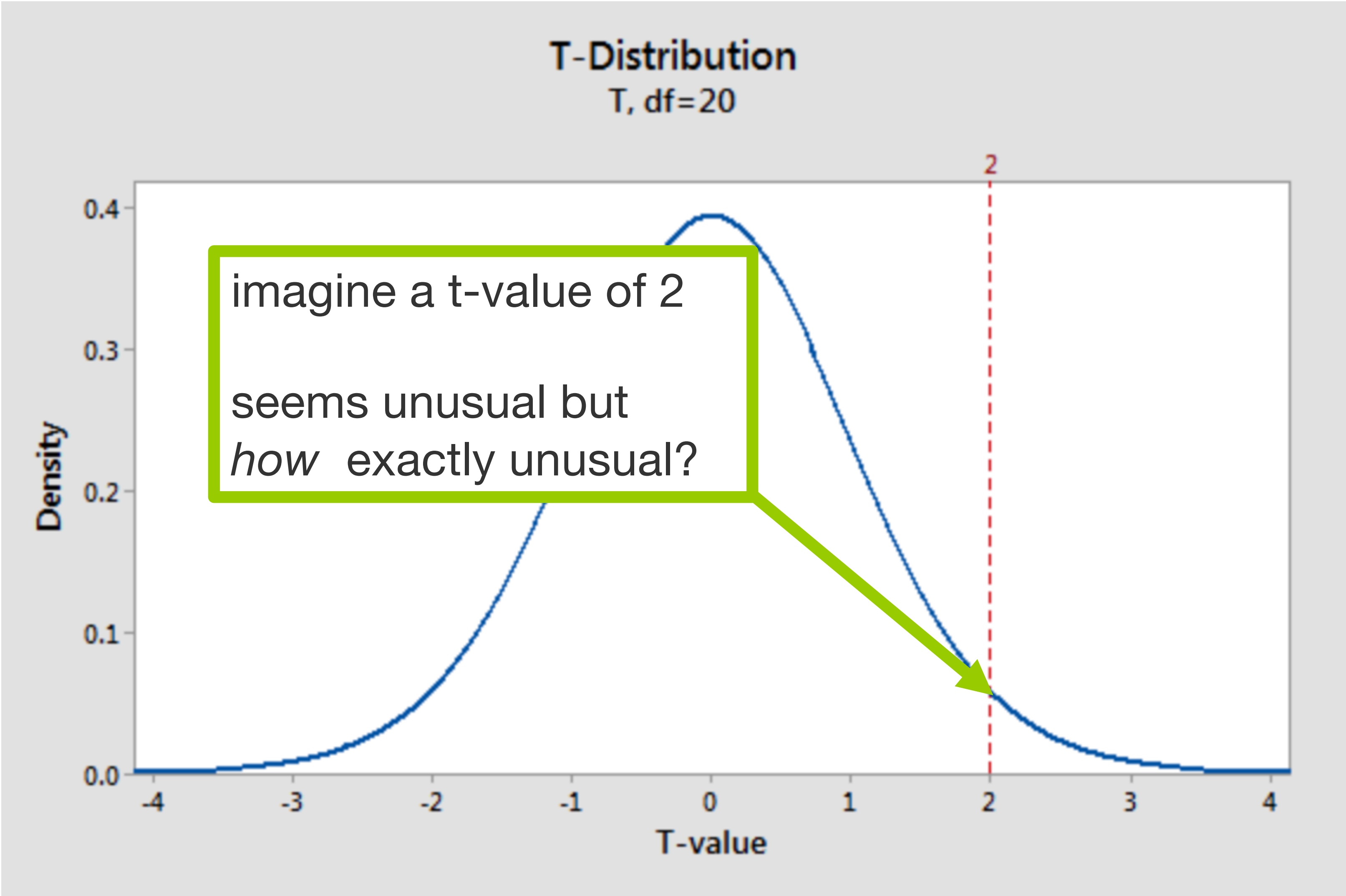
this type of distribution is a **sampling distribution** fortunately, the properties of t-distributions are well understood in statistics, so we can plot them without having to collect many samples!

a specific t-distribution is defined by its **degrees of freedom** (DF), a value closely related to sample size (here n-1)

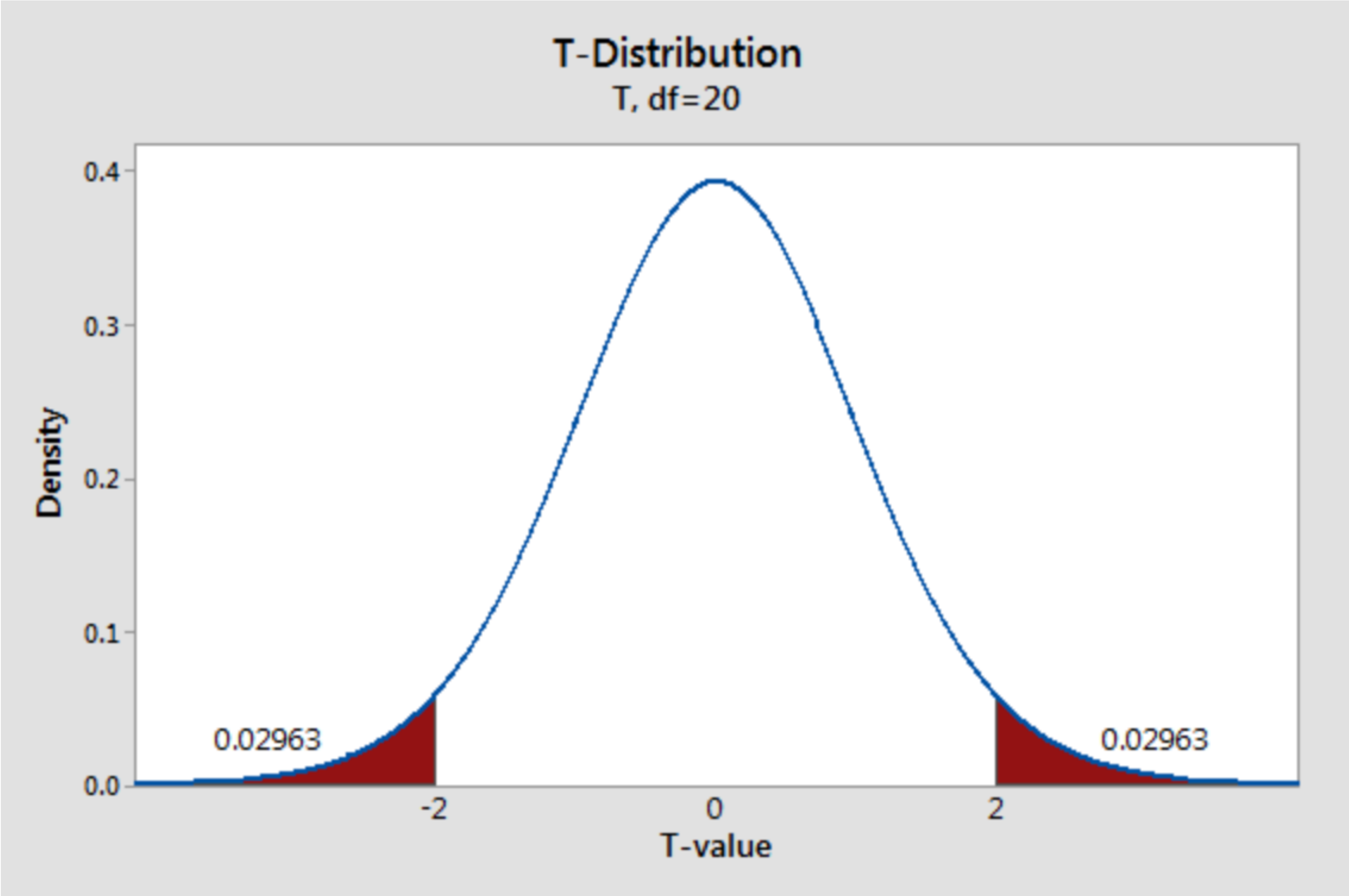
different t-distributions exist for every sample size


Comparison of t Distributions
Distributions
df=1
df=2
df=5
df=10
normal
Density
0.0
0.1
0.2
0.3
0.4
t-value
−3
−2
−1
0
1
2
3

t-distributions assume that you draw repeated random samples from a population where the null hypothesis is true. You place the t-value from your study in the t-distribution to determine how consistent your results are with the null hypothesis.



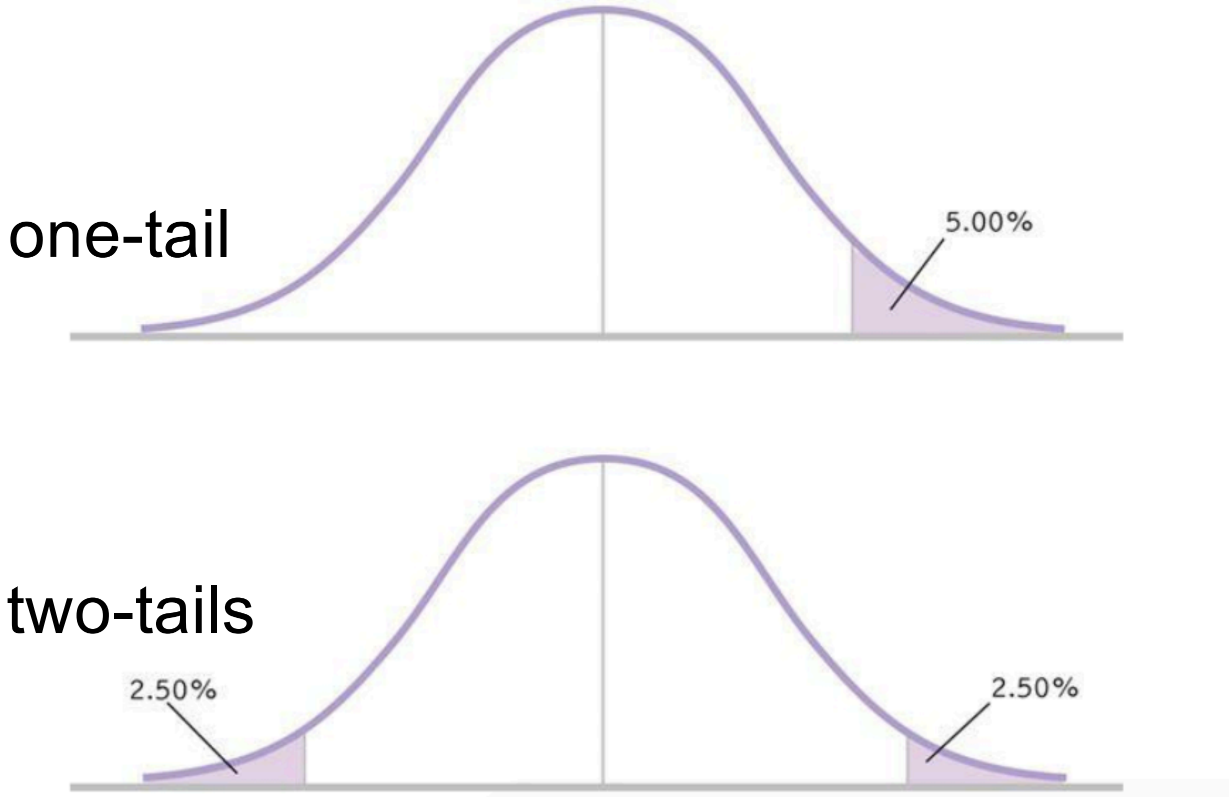
e.g. here a t-distribution (DF =20 which means a sample size of 21). It plots the probability density function (PDF), which describes the likelihood of each t-value.



shade the area of the curve with t-values >2 and <-2

each regions has a probability of 0.02963, which sums to a total probability of 0.05926 **… this is our pvalue!**

$$t = \frac{\overline{x1} - \overline{x2}}{s/\sqrt{n}}$$



it also does explain **one-tail vs. two tails** t-tests: one-tail only case about t=2 (not -2 or oppositely), so multiply pvalue by two.

at this point you understand more the **three reasons** of a low p_value (or t_value)

1. difference not large enough

(what you are searching for, your signal is weak)

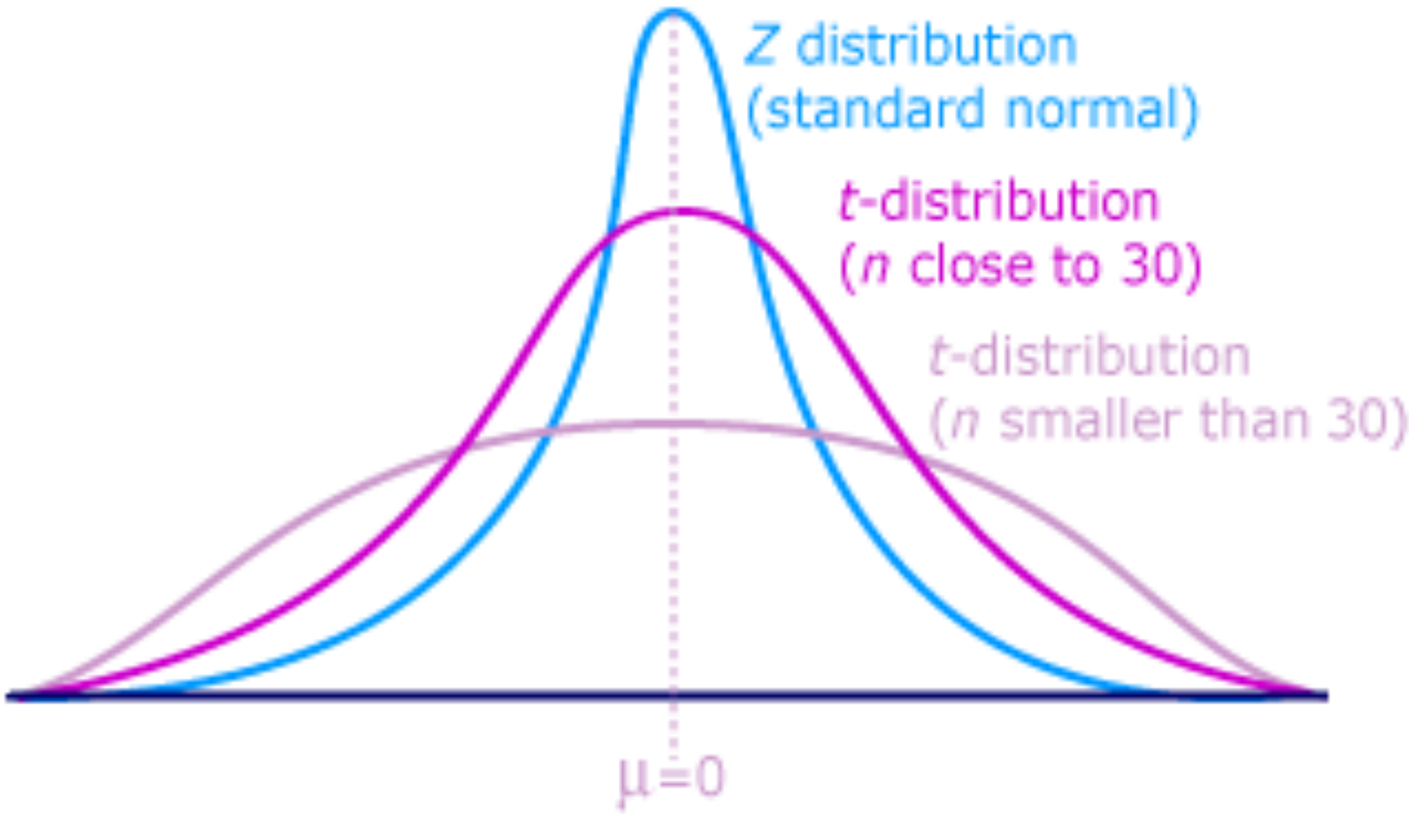
$$t = \frac{\boxed{\overline{x1} - \overline{x2}}}{\boxed{s}/\sqrt{\boxed{n}}}$$

2. too much noise

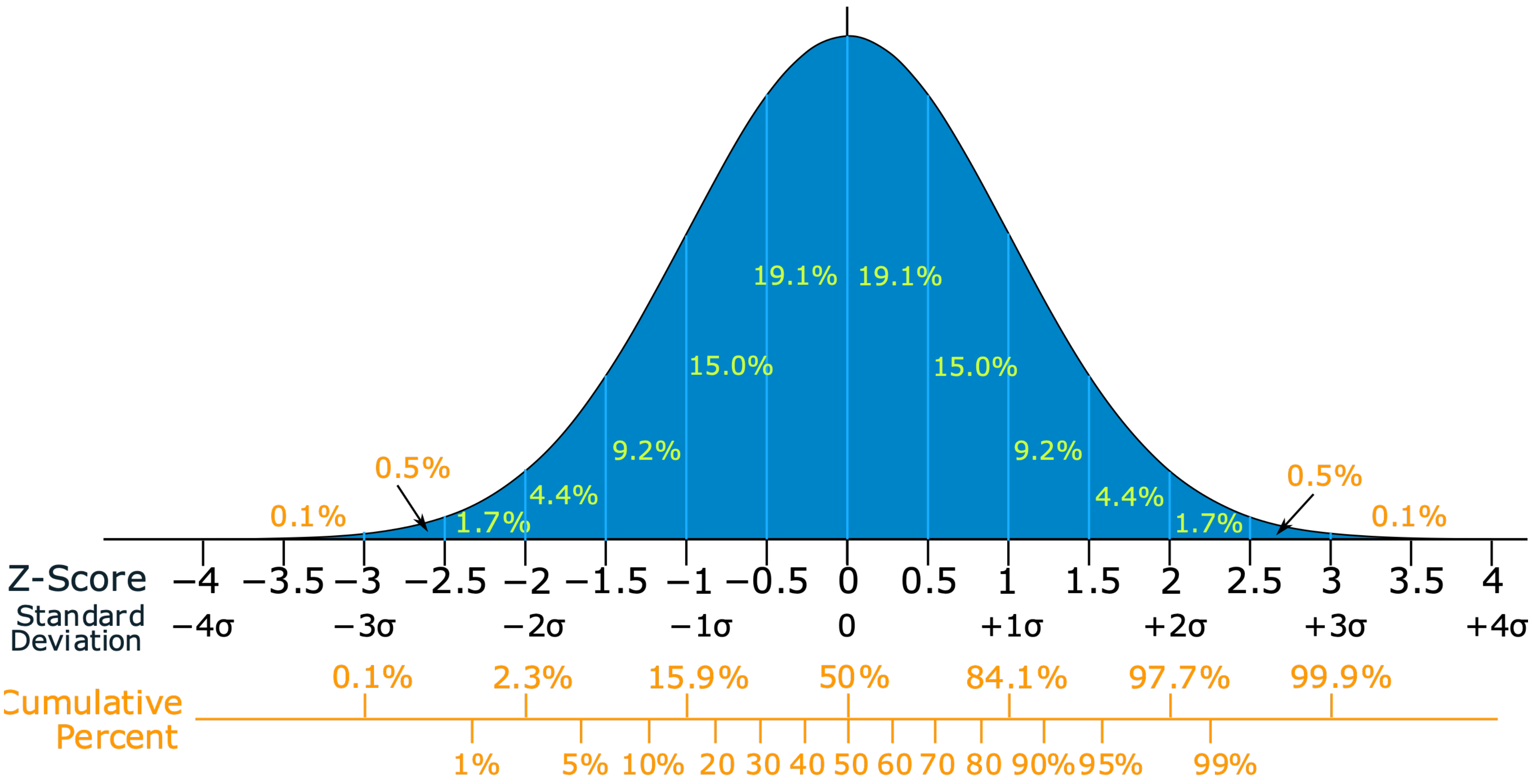
(could your experimental design introduce noise? Check it)

3. not enough data

(run more participants, gather more trials) how much data is enough?



the larger the sample size, the more t-distributions become a z-distribution (at around n=30), the less the area under the curve to reach a low pvalue.



mean that more sample = back to using a Z-score (i.e. number of standard deviations from the mean on normal distribution) so why using a t-test then?

well there are cases when we want to use less sample to speed up the evaluation

this was the case of **William Sealy Gosset** …

Employee of Guiness, Gosset developed a **small sample** method to select the best yielding varieties of barley.

Biometricians like Pearson typically had hundreds of observations.

Guinness allowed him to publish his method under the name “Student” to prevent disclosure of confidential information.

Where do t-distrubitions came from?

https://www.youtube.com/watch?v=NvUDvmrd6fo&feature=youtu.be

now it does not mean that 2 sample is enough … a rule of thumb for a simple within experiment is 12-16 participants for example (and twice more for between experiment).

… this is our pvalue!

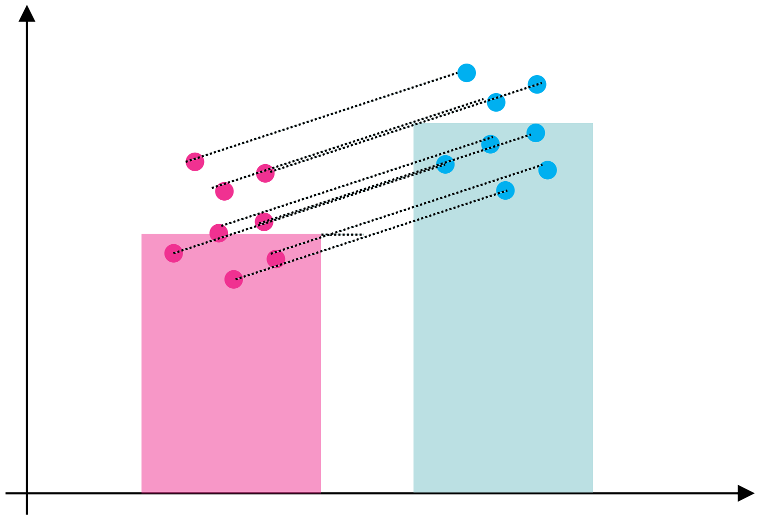
… we will actually look at how to know if you sample size is good in the last week unpaired t-test

T-tests ::

Paired $t = \frac{\overline{x1} - \overline{x2}}{s/\sqrt{n}}$

Unpaired $t = \frac{\overline{x1} - \overline{x2}}{s\boxed{\sqrt{\left(\frac{1}{n1} + \frac{1}{n2}\right)}}}$

similar than paired t-test except this

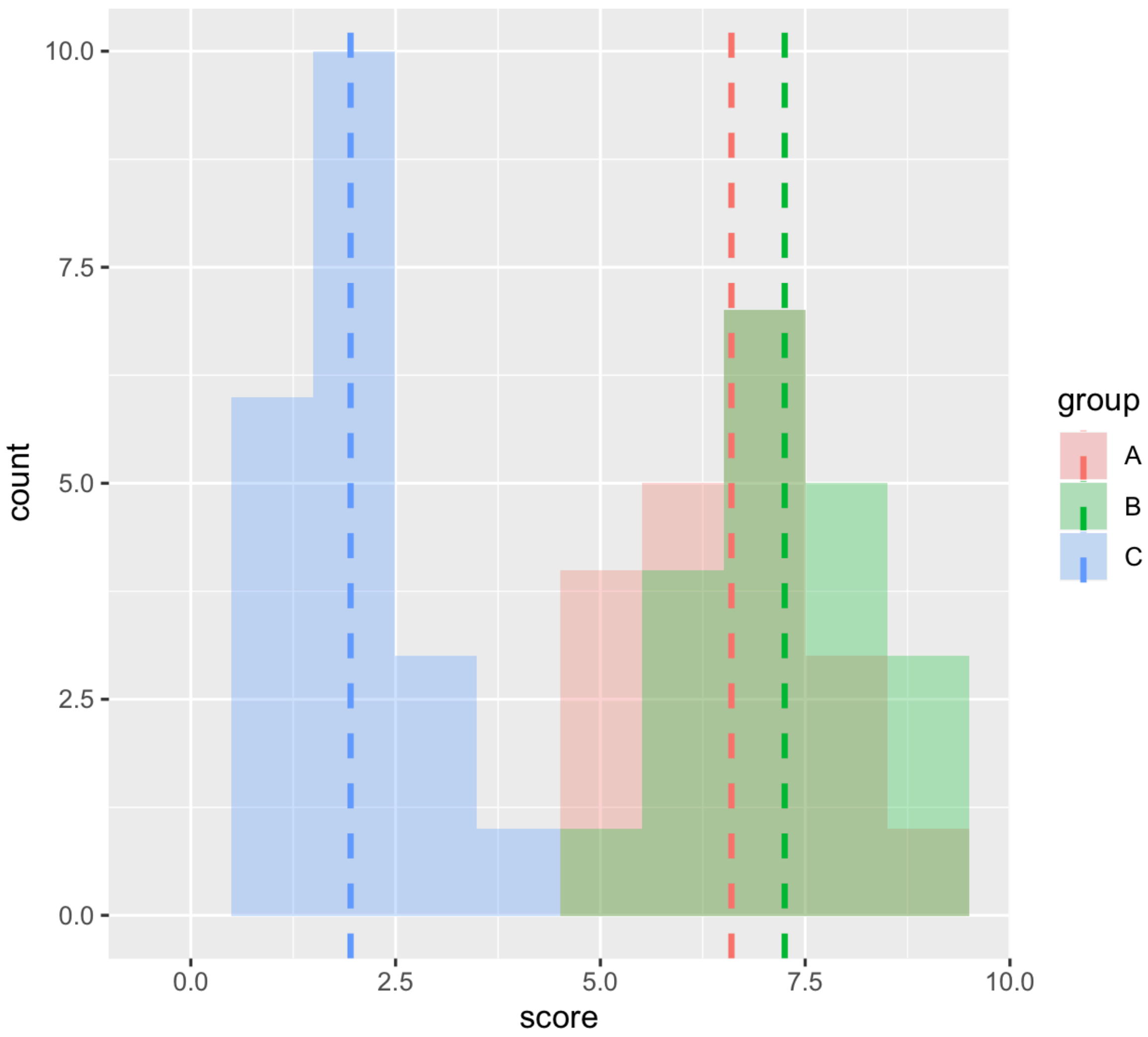


paired t-test :: divide by $\sqrt{n}$ because data point paired

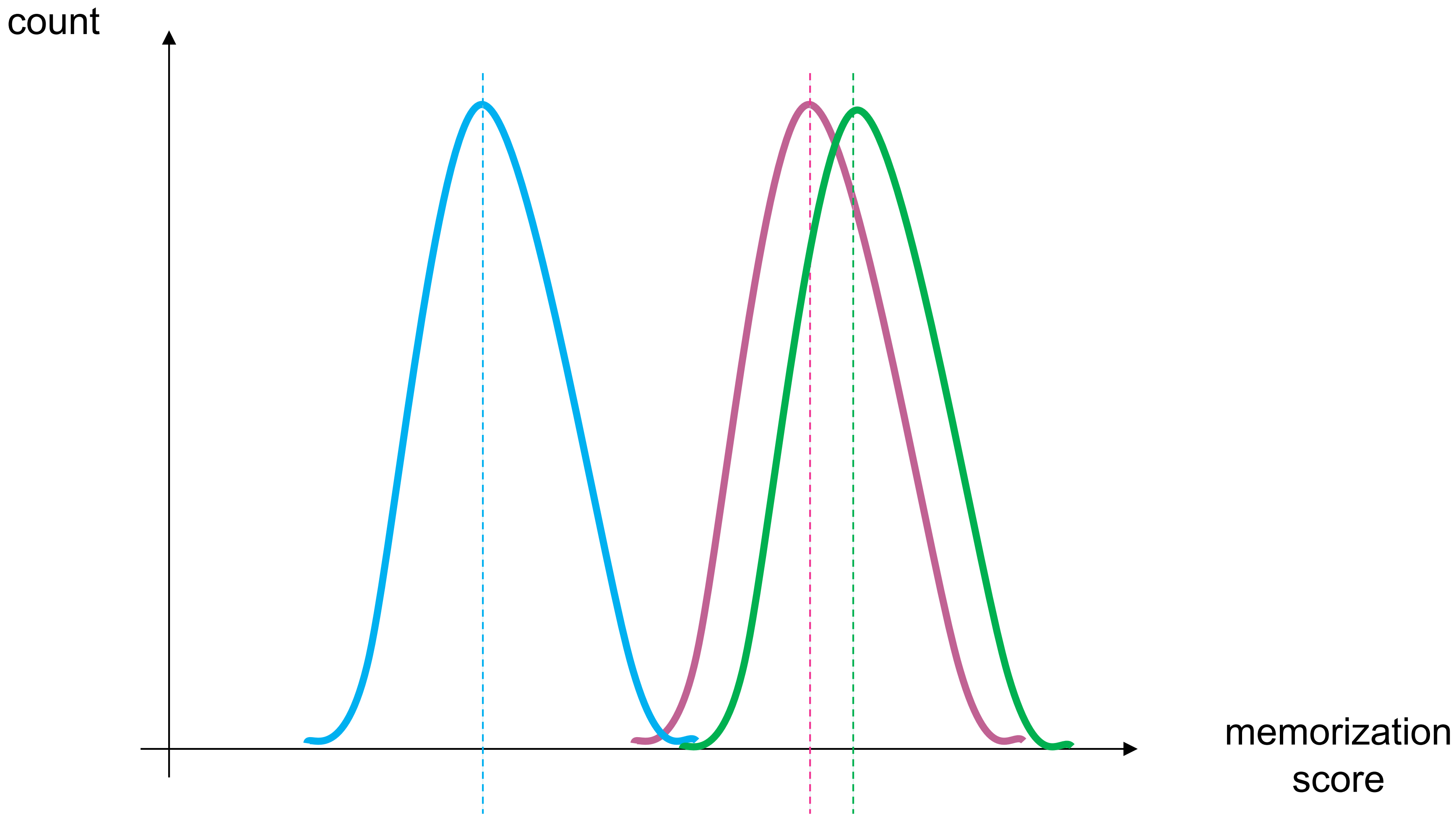
unpaired t-test :: by $\sqrt{\left(\frac{\mathbf{1}}{\boldsymbol{n1}}+\frac{\mathbf{1}}{\boldsymbol{n2}}\right)}$

you can do the math: unpaired t-test the denominator (noise) is larger because we add n1 + n2

… thus why harder to reach low pvalue with unpaired t-test anovas



10.0
7.5
5.0
2.5
0.0
count
0.0
2.5
5.0
7.5
10.0
score
group
A
B
C

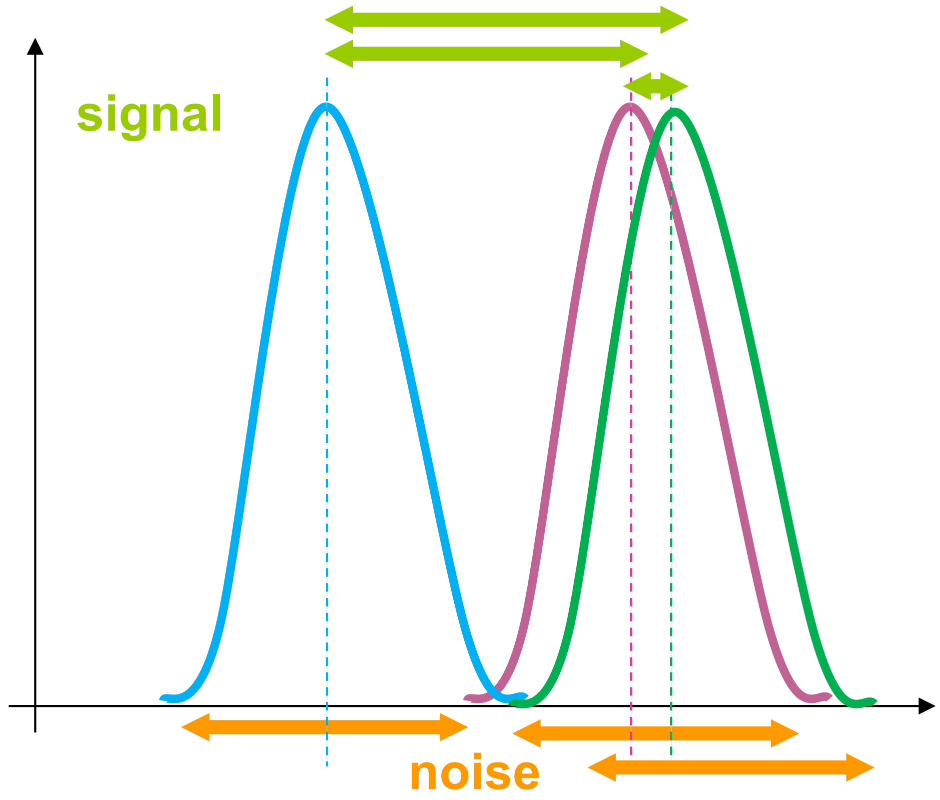


(let’s assume again these are normally distributed)

any statistical tests ::

$$\frac{\text{signal}}{\text{noise}}$$

ANOVA ::



$$\frac{\text{difference between group means}}{\text{variability of groups}}$$

ANOVA ::

$$F = \frac{MS_{between}}{MS_{within}}$$

$$MS_{between} \quad \frac{SS_{between}}{df_{between}}$$

$$MS_{within} \quad \frac{SS_{within}}{df_{within}}$$

$$SS_{between} = \sum_{j=1}^{p} n_j(\overline{x}_j - \overline{x})^2$$

$$SS_{within} = \sum_{j=1}^{p}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_j)^2$$

don’t be afraid by these equations,
these are just simple math

if you like that kind of calculation you can even do it by hand …

	Group 1	Group 2	Group 3
n (sample size)	70	70	70
M (mean)	4	3.7	3.4
s^2 (variance)	4.4	5.2	6.1

Anova by hand (step by step)		
1. compute th combined sample size N	210	
2. compute the degress of freedom between (dfbetween)	2	(number of groups - 1)
3. compute the degress of freedom within (dfwithin)	207	(n1-1)+(n2-1)+(n3-1)
the nominator		
4. compute the average mean	3.7	
5. compute the SSbetween	12.6	
6. compute the MSbetween (divide by dfbetween)	6.3	
the denominator		
7. compute the SSwithin	1083.3	(I multiply by ni here as the variance formula has a dividor which we don't need here)
8. Compute Mswithing (divide by dfwithin)	5.233333	
9. compute F	1.203822	
10. find p_value	0.302137	(p value NOT < alpha so DO NOT reject Ho)

Sample Variance = $s^2 = \frac{\Sigma(X-\overline{X})^2}{n-1}$

(excel file in the git hub repository)

A	B	C	D	E	F	G
	IDs	Before	After	1. difference	**Paired T-test example**	
	1	312	300	12	(In green the initial data, in blue the computation)	
	2	242	201	41		
	3	340	232	108	steps by steps	
	4	388	312	76		
	5	296	220	76	1. add new colum to compute the differences between conditions for each participants	
	6	254	256	-2	2. compute the mean of the differences (use excel formula =AVERAGE(new column))	56.1111111
	7	391	328	63	3. compute the standard deviation of the differences (use formula =STDEV(new column))	34.173983
	8	402	330	72	4. compute de standard error of the mean difference (difive 3. by SQRT(n))	11.3913277
	9	290	231	59	5. compute t_value, i.e. step 2. divided by step. 4	4.92577449

(excel file in the git hub repository)

ANOVA ::

$$\boxed{F} = \frac{MS_{between}}{MS_{within}}$$

$$MS_{between} \quad \frac{SS_{between}}{df_{between}}$$

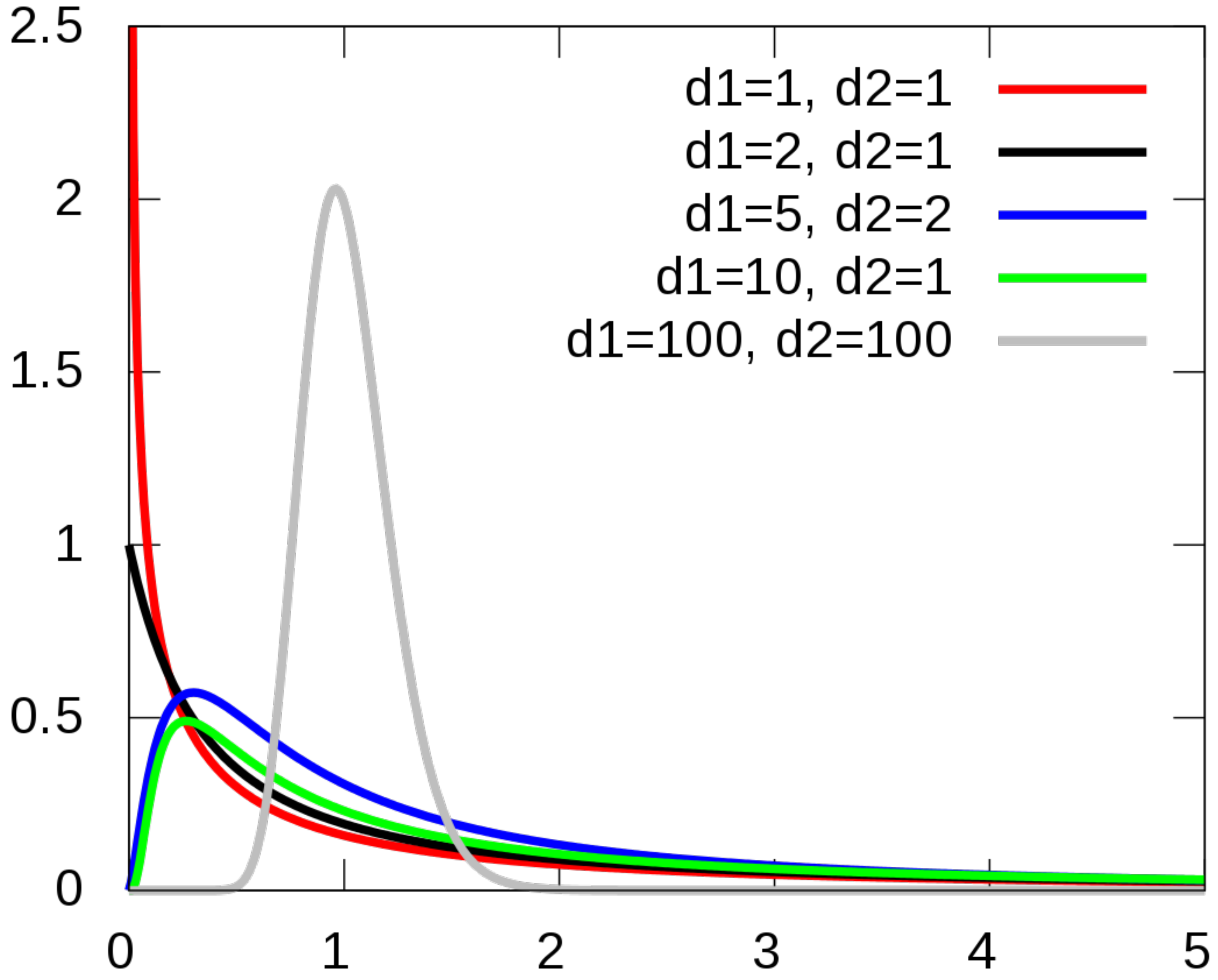
$$MS_{within} \quad \frac{SS_{within}}{df_{within}}$$

$$SS_{between} = \sum_{j=1}^{p} n_j(\overline{x}_j - \overline{x})^2$$

$$SS_{within} = \sum_{j=1}^{p}\sum_{i=1}^{n_j}(x_{ij} - \overline{x}_j)^2$$

how do we know if the Fvalue is anygood?

… this is where **F-distributions** come in



they take **two degrees of freedom** rather than just one like with t-distributions

F Distribution critical values for P=0.05

Denominator (the within df – also called the error)

Numerator DF (the between df)

DF	1	2	3	4	5	7	10	15	20	30	60	120	500	1000
1	161.45	199.50	215.71	224.58	230.16	236.77	241.88	245.95	248.01	250.10	252.20	253.25	254.06	254.19
2	18.513	19.000	19.164	19.247	19.296	19.353	19.396	19.429	19.446	19.462	19.479	19.487	19.494	19.495
3	10.128	9.5522	9.2766	9.1172	9.0135	8.8867	8.7855	8.7028	8.6602	8.6165	8.5720	8.5493	8.5320	8.5292
4	7.7086	6.9443	6.5915	6.3882	6.2560	6.0942	5.9644	5.8579	5.8026	5.7458	5.6877	5.6580	5.6352	5.6317
5	6.6078	5.7862	5.4095	5.1922	5.0504	4.8759	4.7351	4.6187	4.5582	4.4958	4.4314	4.3985	4.3731	4.3691
7	5.5914	4.7375	4.3469	4.1202	3.9715	3.7871	3.6366	3.5108	3.4445	3.3758	3.3043	3.2675	3.2388	3.2344
10	4.9645	4.1028	3.7082	3.4780	3.3259	3.1354	2.9782	2.8450	2.7741	2.6996	2.6210	2.5801	2.5482	2.5430
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7066	2.5437	2.4035	2.3275	2.2467	2.1601	2.1141	2.0776	2.0718
20	4.3512	3.4928	3.0983	2.8660	2.7109	2.5140	2.3479	2.2032	2.1241	2.0391	1.9463	1.8962	1.8563	1.8498
30	4.1709	3.3159	2.9223	2.6896	2.5336	2.3343	2.1646	2.0149	1.9317	1.8408	1.7396	1.6835	1.6376	1.6300
60	4.0012	3.1505	2.7581	2.5252	2.3683	2.1666	1.9927	1.8365	1.7480	1.6492	1.5343	1.4672	1.4093	1.3994
120	3.9201	3.0718	2.6802	2.4473	2.2898	2.0868	1.9104	1.7505	1.6587	1.5544	1.4289	1.3519	1.2804	1.2674
500	3.8601	3.0137	2.6227	2.3898	2.2320	2.0278	1.8496	1.6864	1.5917	1.4820	1.3455	1.2552	1.1586	1.1378
1000	3.8508	3.0047	2.6137	2.3808	2.2230	2.0187	1.8402	1.6765	1.5811	1.4705	1.3318	1.2385	1.1342	1.1096

Example: F for df = 2,207 is 3.0718

also in form of table (here for the dfwithin and dfbetween of our excel example)

```
# first we run the one-way anova
library(ez)
ezANOVA(dat,id,between=group,dv=score)
```

```
  Effect DFn DFd        F            p p<.05       ges
1  group   2  57 154.8886 9.056612e-24     * 0.8445923
```

ok something is going to be interesting here

```
# second, run the pairwise comparison

pairwise.t.test(dat$score,dat$group, paired=FALSE,
p.adjust.method="bonferroni")
```

```
  A      B
B 0.16   -
C <2e-16 <2e-16
```

here are significant differences

and we don't need to do the Bonferroni correction (already included)

this was from last week, R gives us all this numbers

degrees of
freedom

degrees of freedom::

the number of values in the final calculation of a **statistic** that are free to vary

a complex notion but here is the intuition …

you have 7 hats, you want to wear one different every day for a week

Monday: 7 choices

Tuesday: 6 choices

Wednesday: 5 choices

Thursday: 4 choices

Friday: 3 choices

Saturday: 2 choices

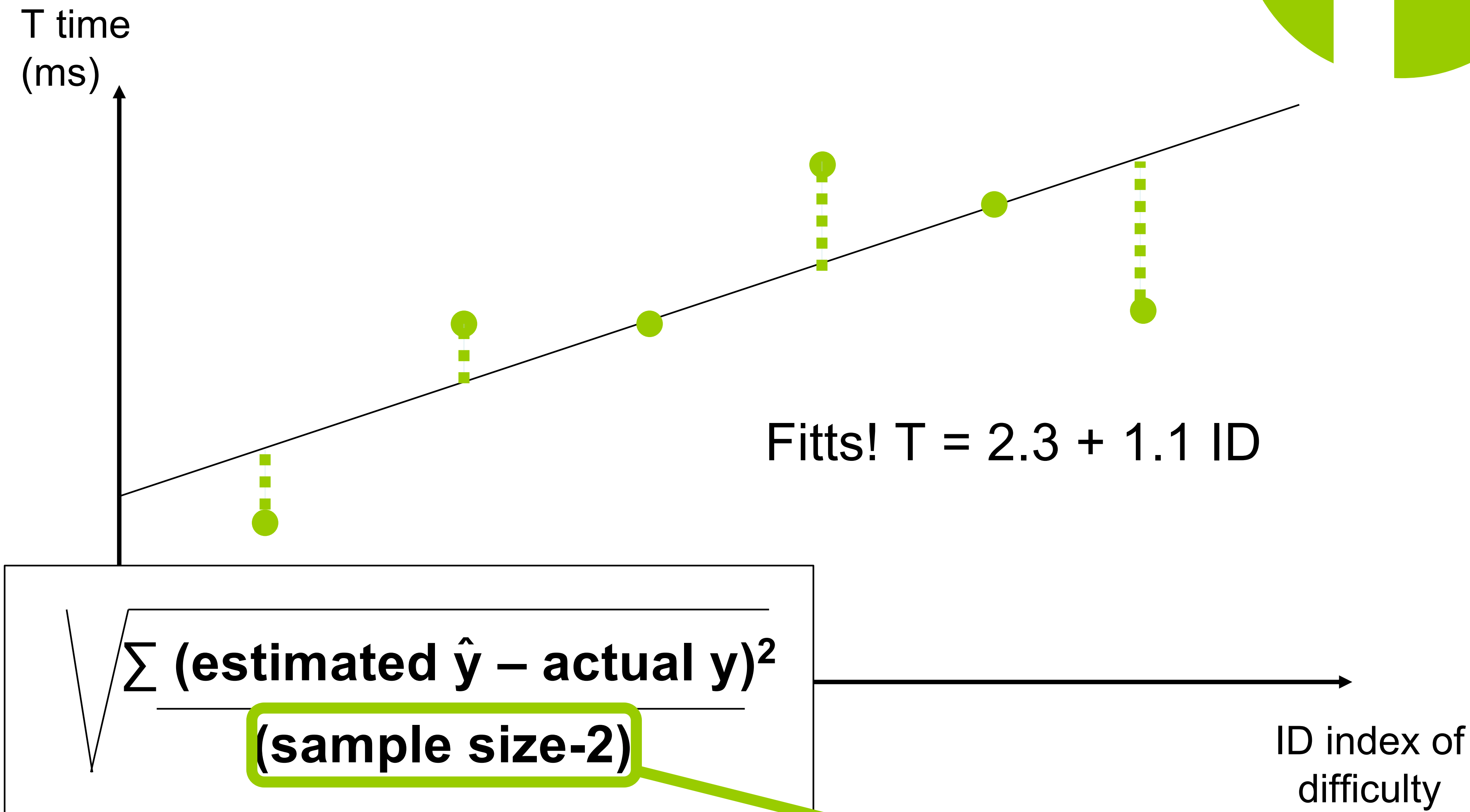
Sunday: **NO choice**

degrees of freedom is 7-1

if we have 7 observations and the mean of these observation (that you need to do t-test or Anova) your degree of freedom is 7-1

because if you know 6 observations you automatically know the 7^{th} one (thanks to the mean) remember the lecture on regression?

standard error of the estimate



$$\sqrt{\frac{\sum \text{(estimated } \hat{y} - \text{actual y)}^2}{\boxed{\text{(sample size-2)}}}}$$

also called degree of freedom

a simple (approximate) way to understand this is that we have two variables, the slope and the intercept of the regression line

that give us extra information, thus the minus 2

statistic tests on multiple ID variables

until now, we did statistical test on one independent variable (with multiple conditions or groups) and one dependant variable

e.g. effect of **chocolate, baseline, punishment (IV)** on **memorization score (DV)**

now it is possible to do tests for multiple IVs and multiple DVs however doing so decrease the power of your experiment (because you run more tests)

so it only works with powerful tests based on ANOVA (i.e. continuous variable and assumption of normality and homogeneity assumed)

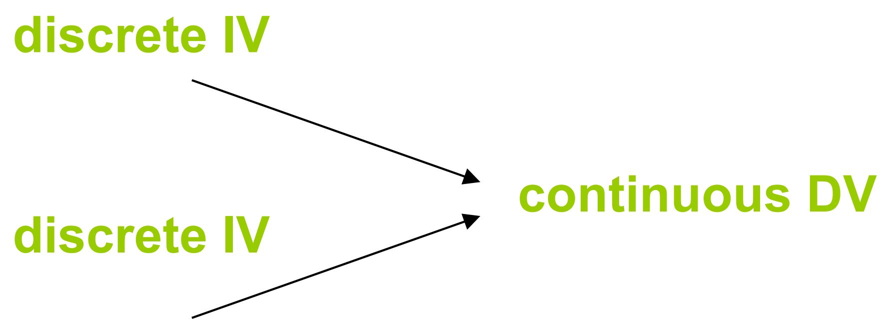
e.g. **two-ways ANOVA, MANOVA, ANVOVA**

discrete IV → continuous DV

one-way ANOVA::

compare the effect of **one discrete independent variables**, having 2 or more levels on **one dependant variable**

e.g. effect of alcohol consumption (none, 2-pints, 4-pints) on attractiveness ratings



two-way ANOVA::

compare the effect of **two discrete independent variables**, each of them having 2 or more levels on **one dependant variable**

e.g. effect of gender (female, male) and alcohol consumption (none, 2-pints, 4-pints) on attractiveness ratings continuous DV

discrete IV

continuous DV

one-way MANOVA::

(multivariate analysis of variance) compare the effect of **one independent variable**, having 2 or more levels on **two dependant variables**

e.g. effect of different memorization enhancing drugs (placebo, drug A, drug B) on memorization skills and emotional ratings (to find the sweet spot for a drug that enhance skills without depressing people!)

discrete IV

continuous covariate

continuous DV

one-way ANCOVA::

(analysis of **co**variance) compare the effect of **one independent variable**, having 2 or more levels and **one continuous covariate** on **one dependant variable**

e.g. effect of phone sizes (iphone 4, iphone 5, iphone 6, iphone 7) on the amplitude of phone movements made when texting **given the measure of the participants hand width (covariate)** discrete IV → continuous DV

continuous covariate → continuous DV

one-way MANCOVA::

(multivariate analysis of covariance) compare the effect of **one independent variable**, having 2 or more levels and **one continuous covariate** on **two dependant variable**

… you can even two a two-way MANCOVA (but your experience might not have much statistical power because there are too many tests to perform) although these tests exist, my advice is to keep the experimental design as simple as possible as you can, analysis will be easier and more powerful

… will look at power with guest lecturer: Luluah Al-Barrack

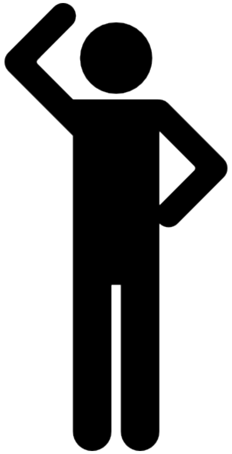
summary

1. Explain the general equation of a statistical test (signal/noise)
2. Explain how a t-test is computed but **I will not ask you to do it by** hand
3. Explain the 3 reasons why a t-value can be low (signal too low, noise too high, small sample)
4. Explain how an Anova is computed but **I will not ask you to do it by** hand
5. Explain what are t-distributions and f-distributions

take away

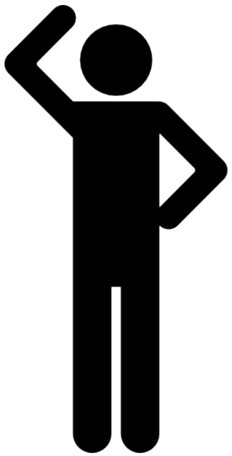
try this with a friend during reading weeks

https://tinyurl.com/statsBristol



tea
milk
water

don’t tell them how you made the cup



tea
water
milk

end

extra