

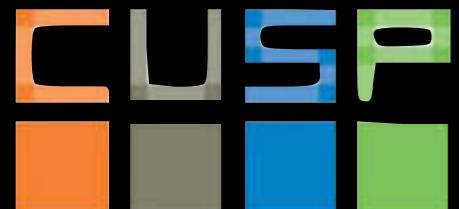
Urban Informatics

Fall 2017

dr. federica bianco fbianco@nyu.edu



@fedhere

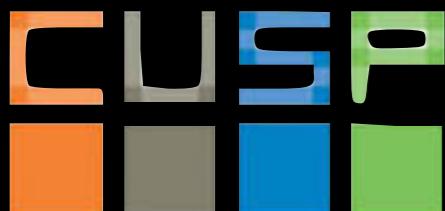


Recap:

- Good practices with data: falsifiability, reproducibility
- Basic data retrieving and munging: APIs, Data formats
- Basic statistics: distributions and their moments
- Hypothesis testing: p -value, statistical significance

Today:

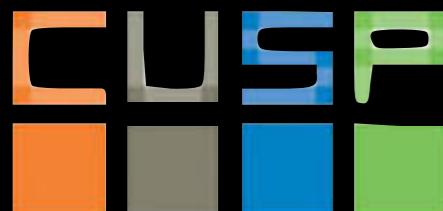
- How to choose the right statistical test
- Z, t, F test and tests for correlation
- Correlation vs Causation



hypothesis testing

null hypothesis 2 tailed: no relationship between two measured phenomena,
or no difference among groups
if you have a test control sample: test sample and control sample are the same - no effect

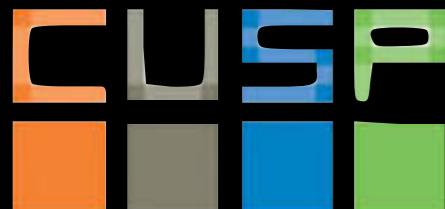
falsify the null hypothesis: do you see an effect?
do you see a difference b/w samples?



hypothesis testing

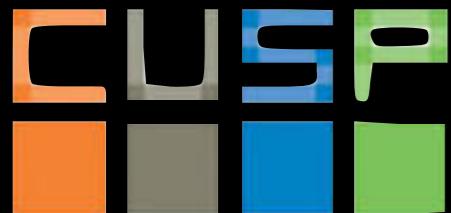
null hypothesis 1 tailed: measured phenomenon larger/smaller for one group than the other

falsify the null hypothesis: do you see an effect in a specific direction?
do you see a difference b/w samples with the correct sign?



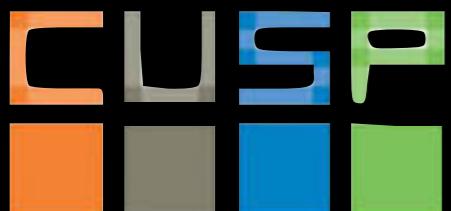
A simple (too simple?) answer

p-value a measure of the probability that the result you observed could have been observed by chance under the *Null hypothesis*



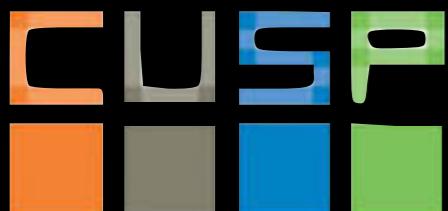
Steps in Hypothesis Testing

1. Formulate Null (and alternative) Hypothesis
2. Choose a significance level α
3. Measure a parameter for the population and compare it to the statistics of a sample
OR
Measure a statistics for *two or more samples* to be compared to *each other*
4. Assess if your statistics is significant or not. In practice: compare the statistics (Z, t, F, chisq) with a distribution table

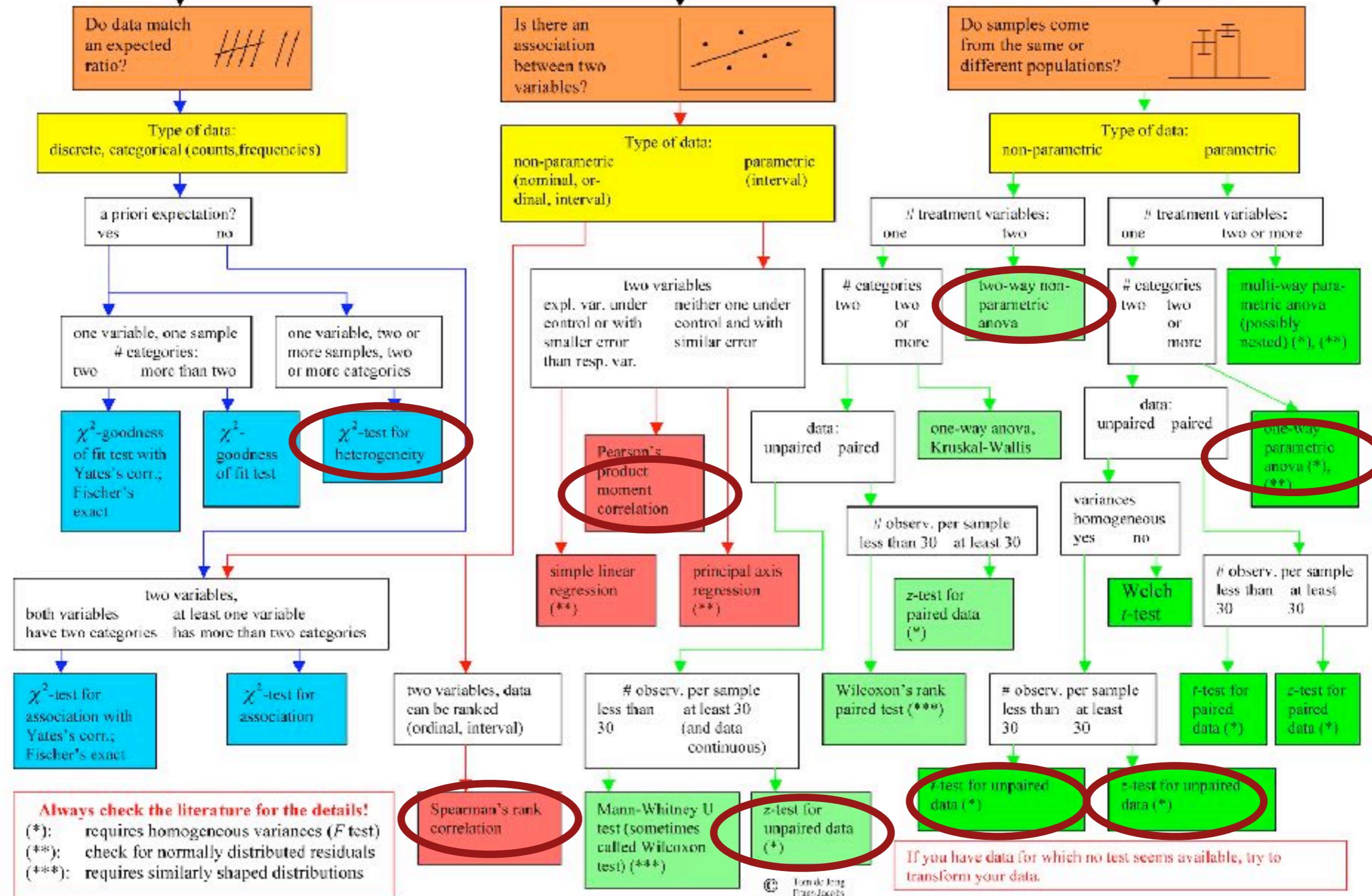


Steps in Hypothesis Testing

1. Formulate Null (and alternative) Hypothesis
2. Choose a significance level α
3. Measure a parameter for the population and compare it to the statistics of a sample
OR
Measure a statistics for *two or more samples* to be compared to *each other*
4. Assess if your statistics is significant or not. In practice: compare the statistics (Z, t, F, chisq) with a distribution table



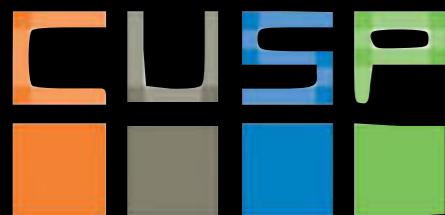
Type of question:



When to Use a Particular Statistical Test

[http://www.csun.edu/~amarenco/Fcs%20682/
When%20to%20use%20what%20test.pdf](http://www.csun.edu/~amarenco/Fcs%20682/When%20to%20use%20what%20test.pdf)

Statistical Analyses	Independent Variables		Dependent Variables		Control Variables	Question Answered by the Statistic
	# of IVs	Data Type	# of DVs	Type of Data		
Chi square	1	categorical	1	categorical	0	Do differences exist between groups?
t-Test	1	dichotomous	1	continuous	0	Do differences exist between 2 groups on one DV?
ANOVA	1 +	categorical	1	continuous	0	Do differences exist between 2 or more groups on one DV?
ANCOVA	1 +	categorical	1	continuous	1 +	Do differences exist between 2 or more groups after controlling for CVs on one DV?
MANOVA	1 +	categorical	2 +	continuous	0	Do differences exist between 2 or more groups on multiple DVs?
MANCOVA	1 +	categorical	2 +	continuous	1 +	Do differences exist between 2 or more groups after controlling for CVs on multiple DVs?
Correlation	1	dichotomous or continuous	1	continuous	0	How strongly and in what direction (i.e., +, -) are the IV and DV related?
Multiple regression	2 +	dichotomous or continuous	1	continuous	0	How much variance in the DV is accounted for by linear combination of the IVs? Also, how strongly related to the DV is the beta coefficient for each IV?
Path analysis	2 +	continuous	1 +	continuous	0	What are the direct and indirect effects of predictor variables on the DV?
Logistic Regression	1 +	categorical or continuous	1	dichotomous	0	What is the odds probability of the DV occurring as the values of the IVs change?



assignment 1:

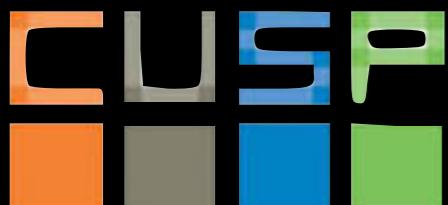
Brows Plos One for papers that uses each of the 4 tests

[http://www.csun.edu/~amarenco/Fcs%20682/
When%20to%20use%20what%20test.pdf](http://www.csun.edu/~amarenco/Fcs%20682/When%20to%20use%20what%20test.pdf)

choose 1 choose 1

choose 1 →

Statistical Analyses	Independent Variables		Dependent Variables		Control Variables	Question Answered by the Statistic
	# of IVs	Data Type	# of DVs	Type of Data		
Chi square	1	categorical	1	categorical	0	Do differences exist between groups?
t-Test	1	dichotomous	1	continuous	0	Do differences exist between 2 groups on one DV?
ANOVA	1 +	categorical	1	continuous	0	Do differences exist between 2 or more groups on one DV?
ANCOVA	1 +	categorical	1	continuous	1 +	Do differences exist between 2 or more groups after controlling for CVs on one DV?
MANOVA	1 +	categorical	2 +	continuous	0	Do differences exist between 2 or more groups on multiple DVs?
MANCOVA	1 +	categorical	2 +	continuous	1 +	Do differences exist between 2 or more groups after controlling for CVs on multiple Dvs?
Correlation	1	dichotomous or continuous	1	continuous	0	How strongly and in what direction (i.e., +, -) are the IV and DV related?
Multiple regression	2 +	dichotomous or continuous	1	continuous	0	How much variance in the DV is accounted for by linear combination of the IVs? Also, how strongly related to the DV is the beta coefficient for each IV?
Path analysis	2 +	continuous	1 +	continuous	0	What are the direct and indirect effects of predictor variables on the DV?
Logistic Regression	1 +	categorical or continuous	1	dichotomous	0	What is the odds probability of the DV occurring as the values of the IVs change?



IV: Statistical analysis

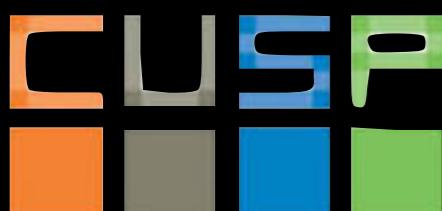
assignment 2:

Prepare a markdown file containing one table as below for each of the 3 tests you select describing the use of the test in a PlosOne publication of your choice

Example: ANCOVA

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0062593>

Statistical Analyses	Dependent Variables	Independent Variables	Control Variables	Question Answered
ANCOVA	Ratings about their values (ordinal)	Did Self Affirmation or not (category)	age	self-affirmation group rates the value significantly more than control group

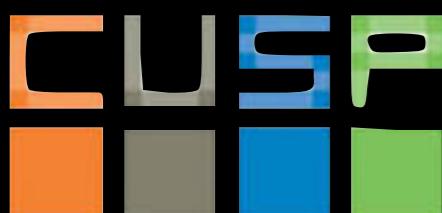


Comparing to population

Standard DEVIATION of Sample Estimates

Sample mean, \bar{x}	$Z = \frac{ \mu_{\text{pop}} - \mu_{\text{sample}} }{\sigma / \sqrt{N}}$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
Sample proportion, p	$z = \frac{ p - p_0 }{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$\sigma_p = \sqrt{\frac{P(1-P)}{n}}$
Difference between proportions, $p_1 - p_2$	$z = \frac{(p_2 - p_1)}{\sqrt{p(1-p)(\frac{1}{n_2} + \frac{1}{n_1})}}, p = \frac{p_2 n_2 + p_1 n_1}{n_2 + n_1}$	$\sigma_{p_1 - p_2} = \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$

use if you know the *parameters* of the *population*:
e.g. in the Z test



Comparing sample to population or between samples

Standard ERROR of Sample Estimates

Sample mean, \bar{x}

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Sample proportion, p

$$SE_p = \sqrt{\frac{p(1-p)}{n}}$$

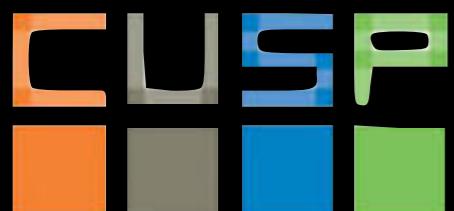
Difference between means,
 $x_1 - x_2$

$$SE_{x_1 - x_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Difference between proportions,
 $p_1 - p_2$

$$SE_{p_1 - p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

use if you DO NOT know the parameters of the population:
e.g. in the t-test



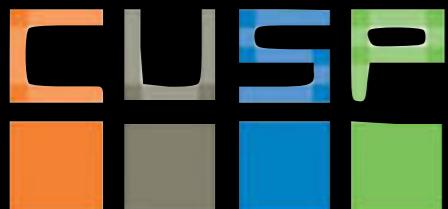
Steps in Hypothesis Testing

1. Formulate Null (and alternative) Hypothesis
2. Choose a significance level α

OR

Measure a statistics for *two or more samples* to be compared to *each other*

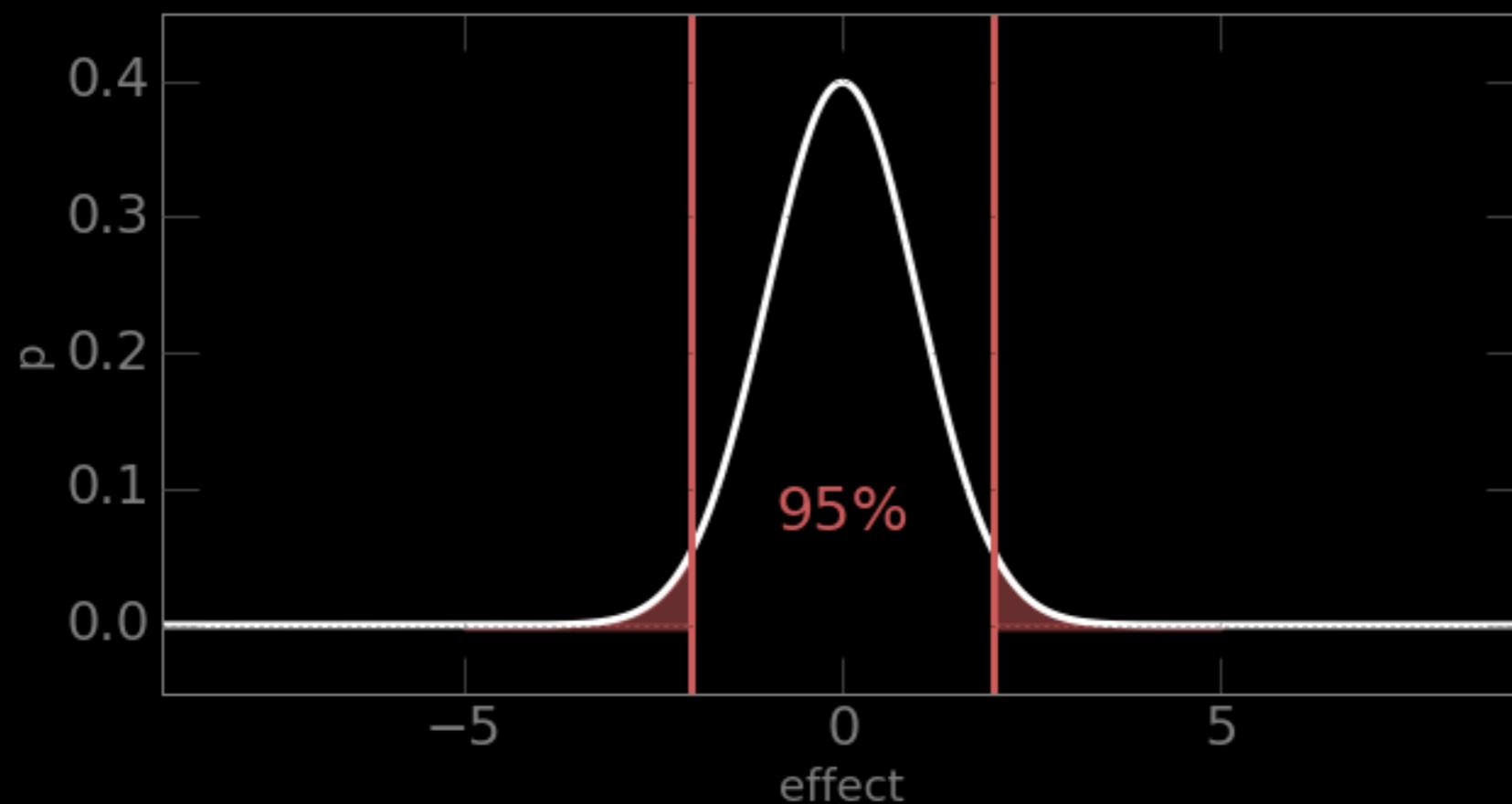
4. Assess if your statistics is significant or not. In practice: compare the statistics (Z, t, F, chisq) with a distribution table



if you knew how your statistics should be distributed...

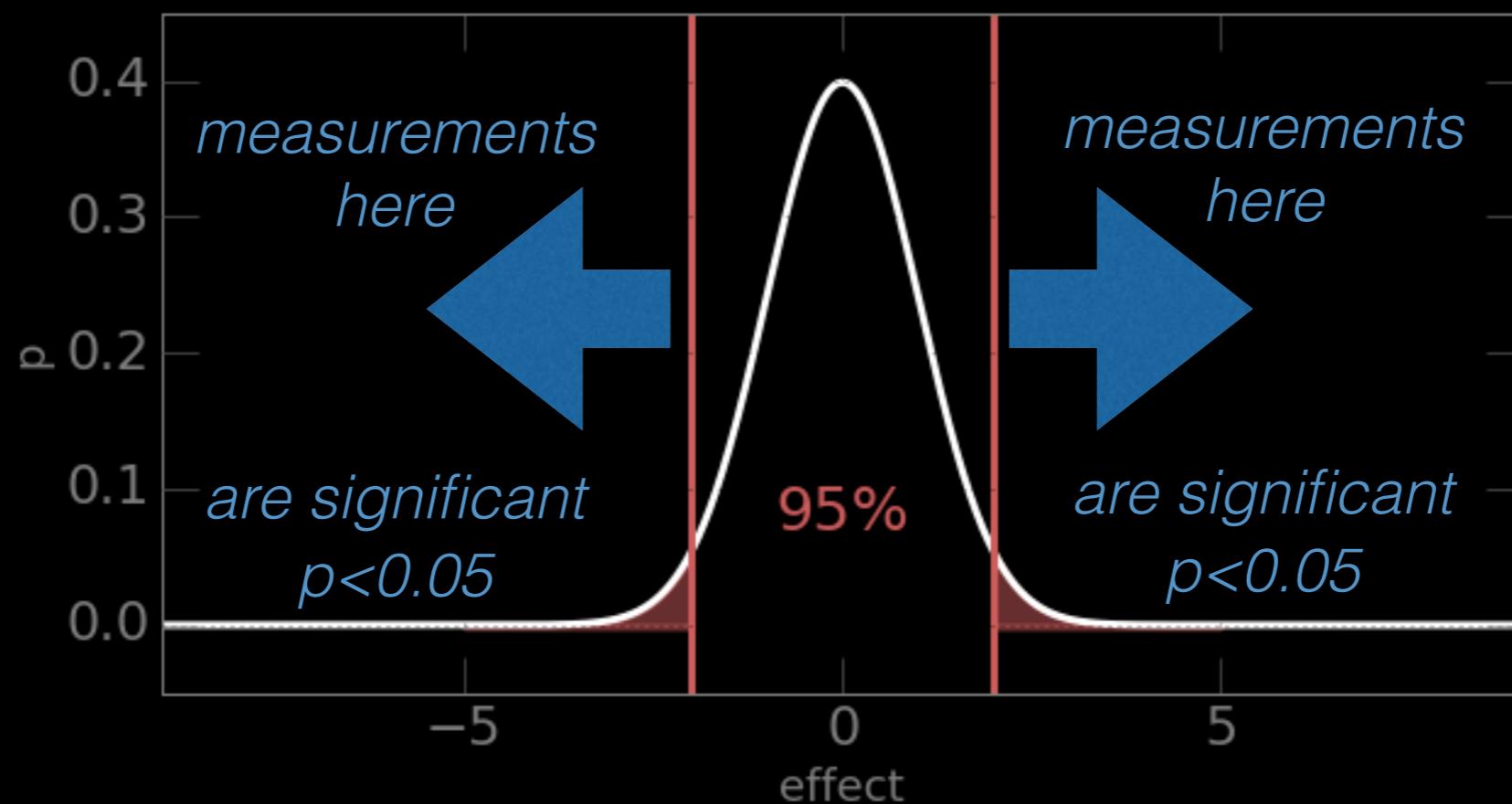
$$\alpha = 0.05$$

$$1 - 0.05 = 0.95 \Rightarrow 95\%$$



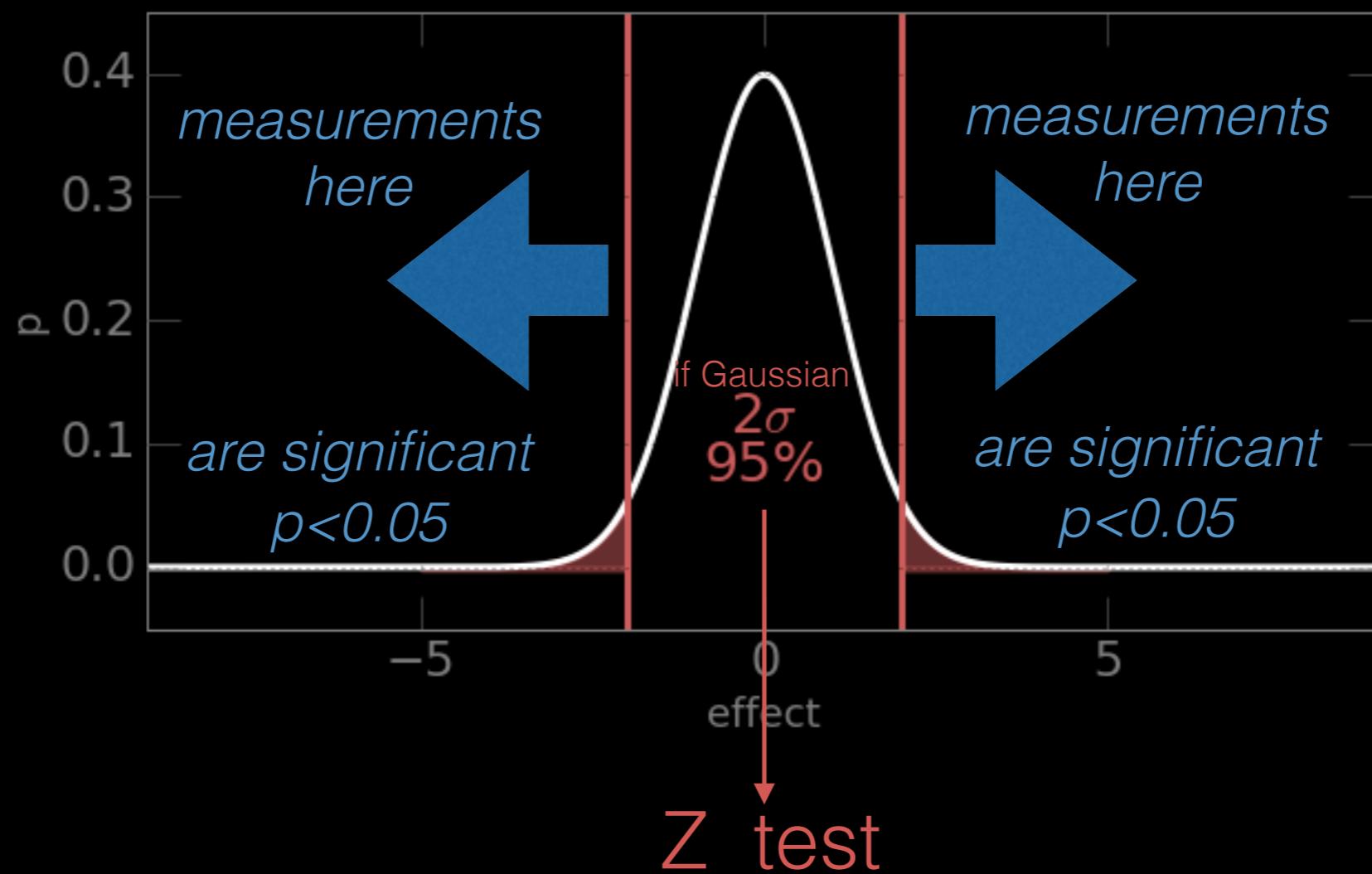
$$\alpha = 0.05$$

$$1 - 0.05 = 0.95 \Rightarrow 95\%$$

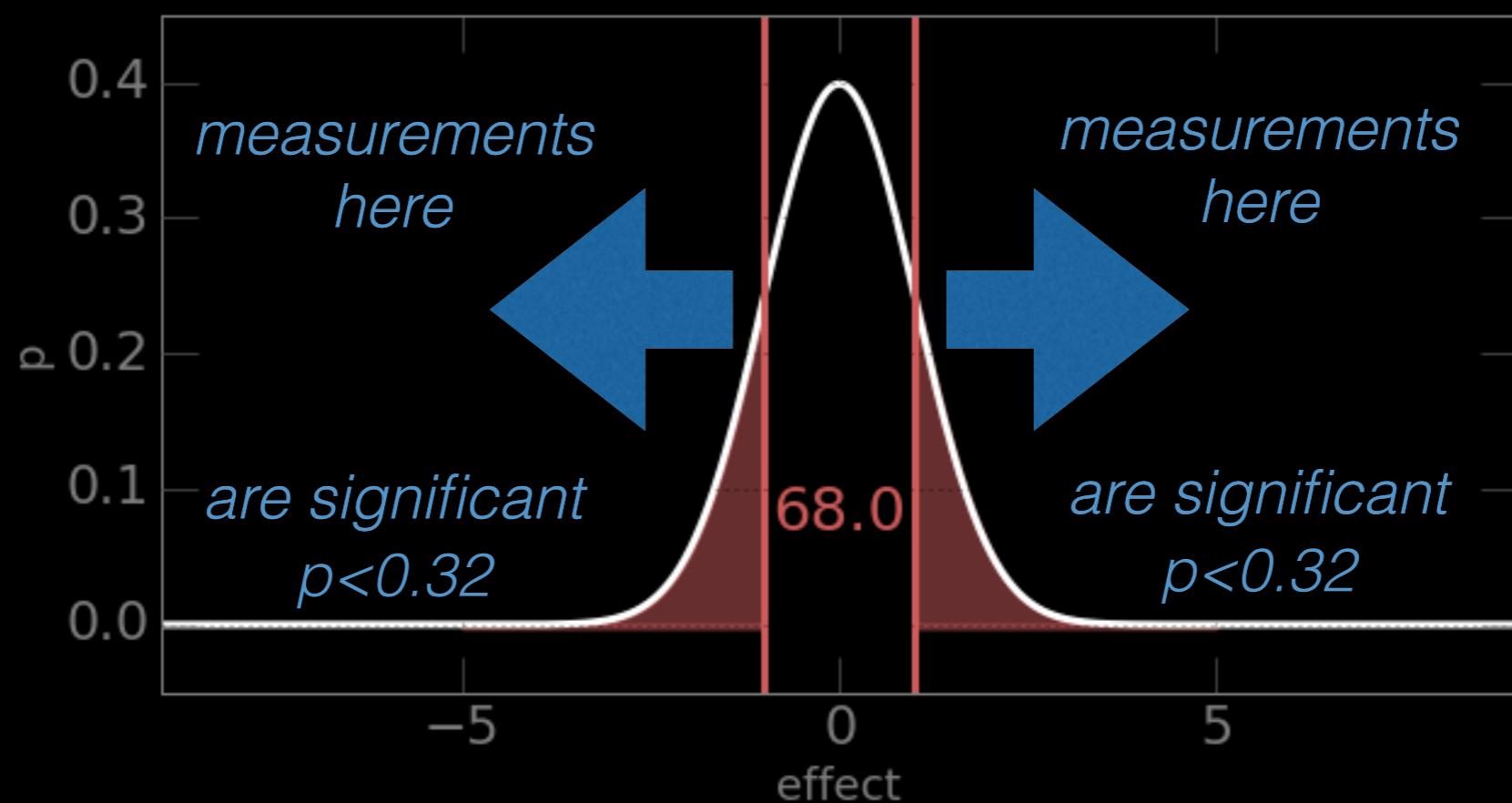


$$\alpha = 0.05$$

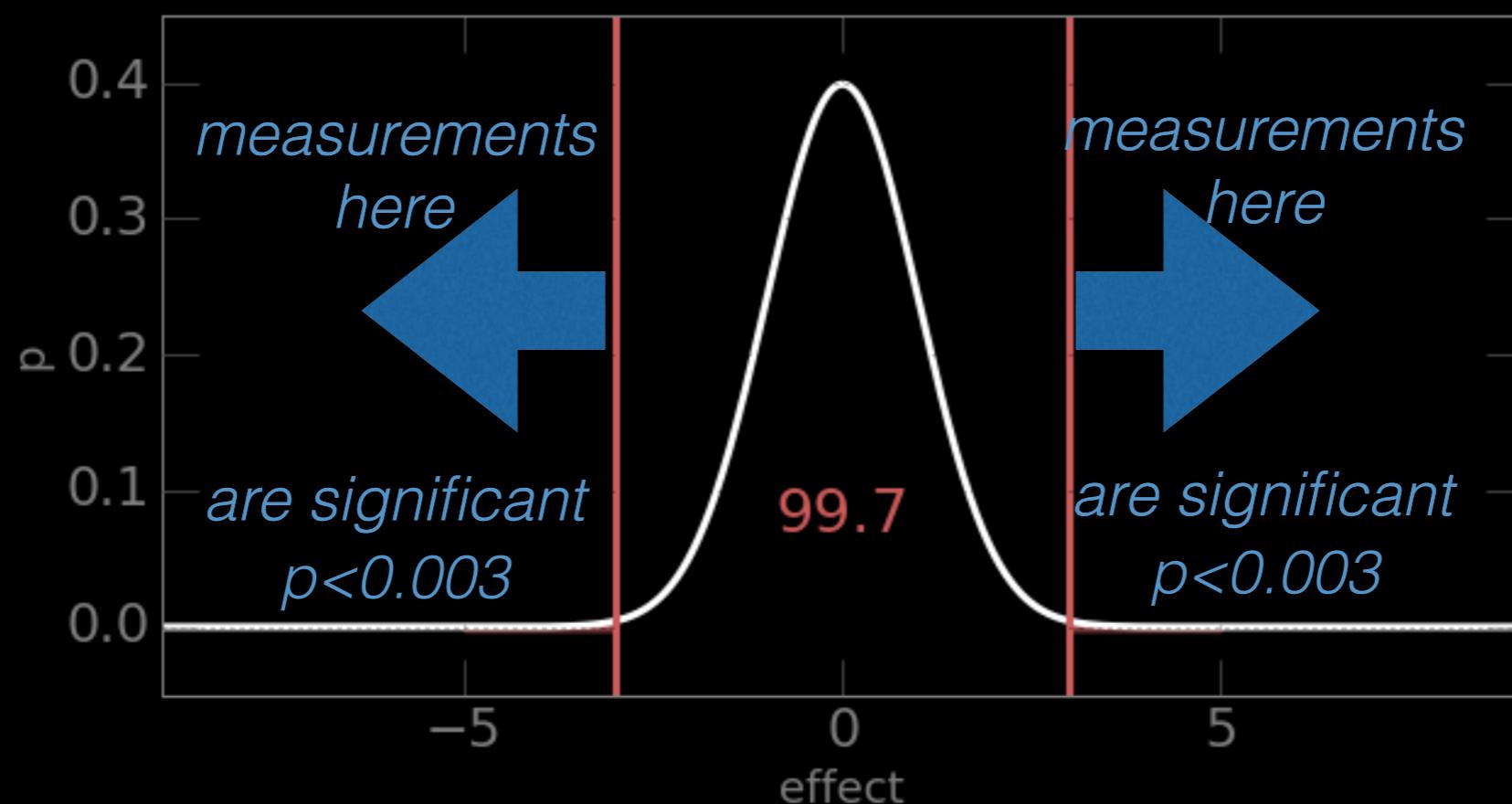
$$1 - \alpha = 0.95 \Rightarrow 95\%$$



$$a = 0.32$$



$$\alpha = 0.003$$



What is the distribution of a statistics?

To measure the probability of the measured value of a statistic I need to know *how the statistics is distributed* (under the alternative hypothesis)
Each statistics follows some distribution, which though?

Steps in Hypothesis Testing

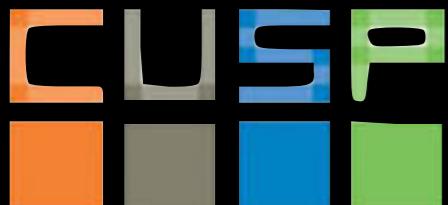
1. Formulate Null (and alternative) Hypothesis
2. Choose a significance level α

OR

Measure a statistics for *two or more samples* to be compared to *each other*

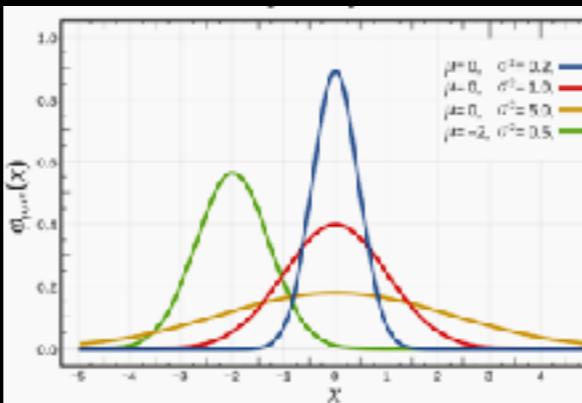
4. Assess if your statistics is significant or not. In practice: compare the statistics (Z, t, F, chisq) with a distribution table

<https://documents.software.dell.com/statistics/textbook/distribution-tables>



Z statistics Gaussian

$$Z = \frac{\mu - \bar{x}}{\sigma / \sqrt{n}}$$

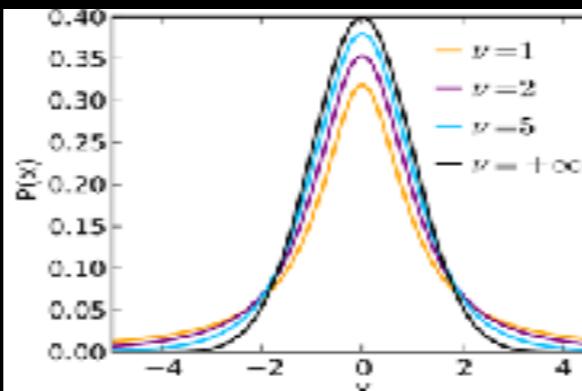


Notation	$N(\mu, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ — mean (location) $\sigma^2 > 0$ — variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}} \right) \right]$
Quantile	$\mu + \sigma\sqrt{2} \operatorname{erf}^{-1}(2F - 1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2

Quantile	$\mu + \sigma\sqrt{2} \operatorname{erf}^{-1}(2F - 1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2

Student's t

$$t = \frac{\mu - \bar{x}}{s / \sqrt{n}}$$

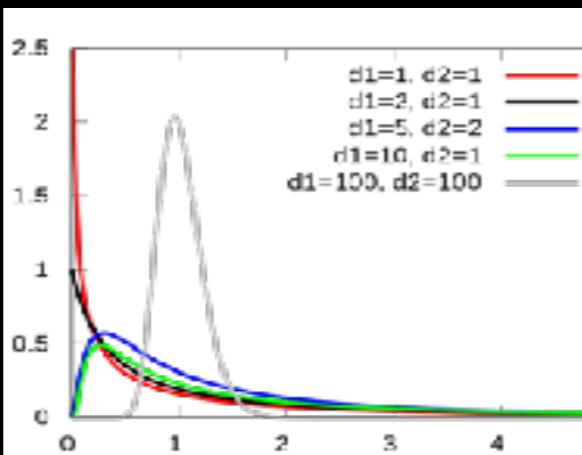


Parameters	$\nu > 0$ degrees of freedom (real)
Support	$x \in (-\infty; +\infty)$
PDF	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
CDF	$\frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right) \times \\ \frac{{}_2F_1\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{x^2}{\nu}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}$
where ${}_2F_1$ is the hypergeometric function	

Mean	0 for $\nu > 1$, otherwise undefined
Median	0
Mode	0
Variance	$\frac{\nu}{\nu-2}$ for $\nu > 2$, = for $1 < \nu \leq 2$, otherwise undefined

F statistics

$$F = \frac{\sum_i n_i (\bar{x}_i - \bar{\bar{x}})^2 / (K-1)}{\sum_{ij} (x_{ij} - \bar{x}_i)^2 / (N-K)}$$

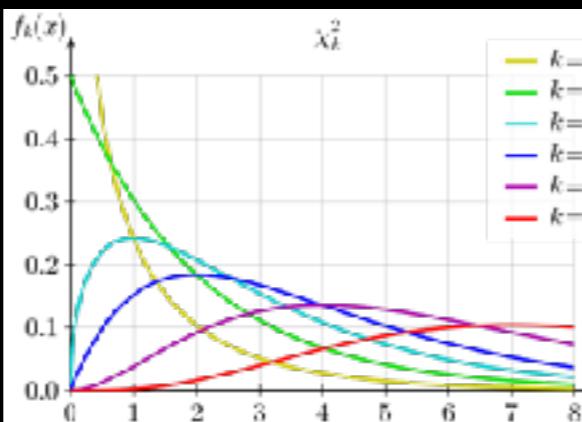


Parameters	$d_1, d_2 > 0$ deg. of freedom
Support	$x \in [0, +\infty)$
PDF	$\frac{\sqrt{(d_1 x)^{d_1} d_2^{d_2}}}{\sqrt{(d_1 + d_2)^{d_1+d_2}}} x \operatorname{B}\left(\frac{d_1}{2}, \frac{d_2}{2}\right)$
CDF	$I \frac{d_1 x}{d_1 x + d_2} \left(\frac{d_1}{2}, \frac{d_2}{2}\right)$

Mean	$\frac{d_2}{d_2 - 2}$ for $d_2 > 2$
Mode	$\frac{d_1 - 2}{d_1 - 4}$ $\frac{d_2}{d_2 + 2}$ for $d_1 > 2$
Variance	$\frac{2 d_2^2 (d_1 + d_2 - 2)}{d_1 (d_2 - 2)^2 (d_2 - 4)}$ for $d_2 > 4$
Skewness	$\frac{(2d_1 + d_2 - 2)\sqrt{8(d_2 - 4)}}{(d_2 - 6)\sqrt{d_1(d_1 + d_2 - 2)}}$ for $d_2 > 6$

Pearson's χ^2

$$\chi_P^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

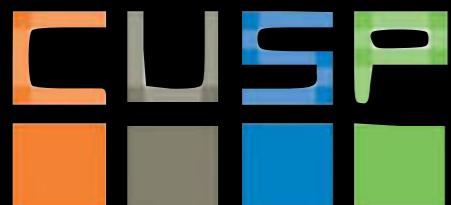


Notation	$\chi^2(k)$ or χ_k^2
Parameters	$k \in \mathbb{N}_{>0}$ (known as "degrees of freedom")
Support	$x \in [0, +\infty)$
PDF	$\frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$
CDF	$\frac{1}{\Gamma\left(\frac{k}{2}\right)} \gamma\left(\frac{k}{2}, \frac{x}{2}\right)$

Mean	k
Median	$\approx k \left(1 - \frac{2}{9k}\right)^3$
Mode	$\max\{k-2, 0\}$
Variance	$2k$
Skewness	$\sqrt{\frac{8}{k}}$

see

goodness of fit χ^2 $\chi_F^2 = \sum_i \frac{(m_i - x_i)^2}{e_i}$ - Statistics in a Nutshell
IV: Statistical analysis



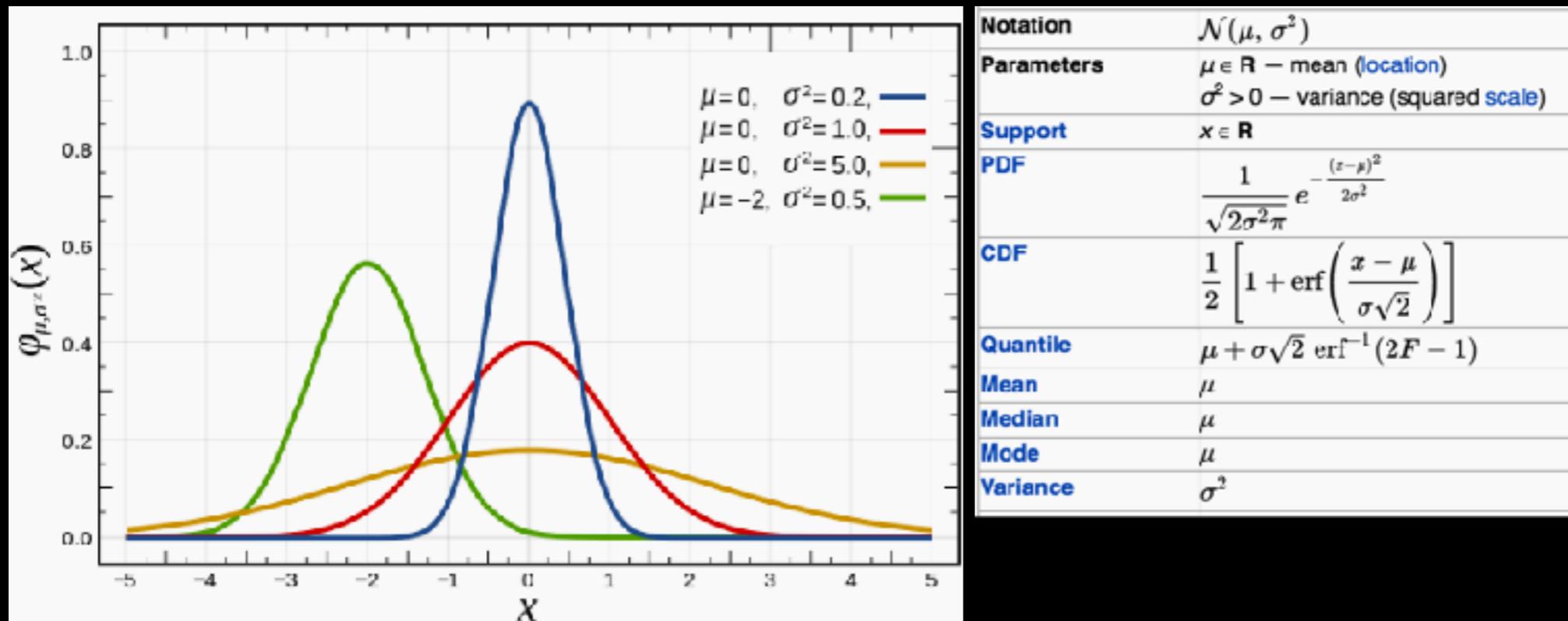
Is there a difference between means or population and sample,
difference between proportion in 2 samples?

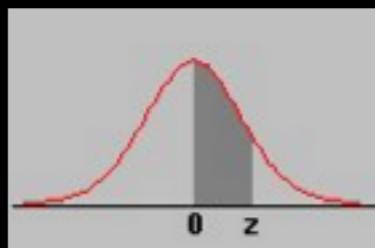
Z statistics

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

In absence of effect (i.e. under the Null)

Z is distributed according to a Gaussian $\mathcal{N}(\mu=0, \sigma=1)$

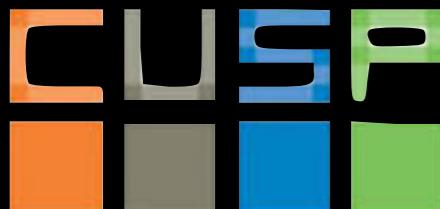


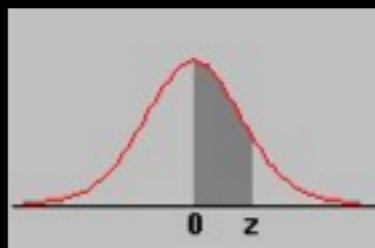


$$Z = \frac{\mu_{\text{pop}} - \mu_{\text{sample}}}{\sigma / \sqrt{N}} = 2.55$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879	1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852	2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015	2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177	2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319	2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
											3.0	0.4987	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>

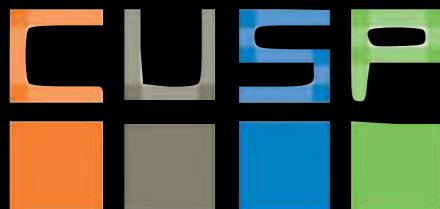


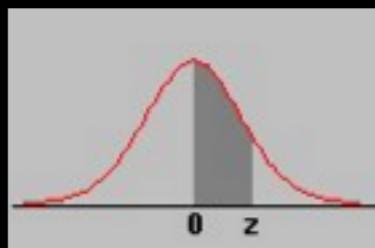


$$Z = \frac{\mu_{\text{pop}} - \mu_{\text{sample}}}{\sigma / \sqrt{N}} = 2.56$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879	1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852	2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015	2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177	2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319	2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
											3.0	0.4987	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>

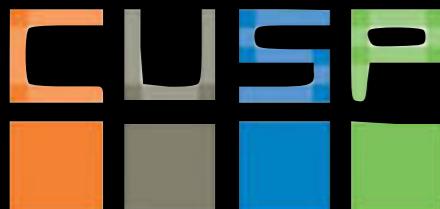


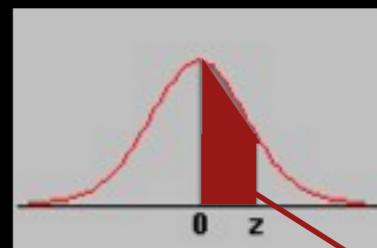


$$Z = \frac{\mu_{\text{pop}} - \mu_{\text{sample}}}{\sigma / \sqrt{N}} = 2.56$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879	1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852	2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015	2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177	2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319	2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
											3.0	0.4987	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>



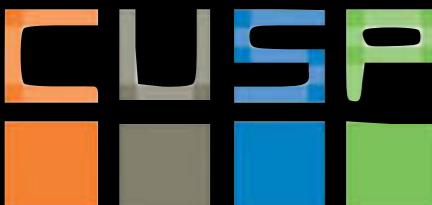


$$Z = \frac{\mu_{\text{pop}} - \mu_{\text{sample}}}{\sigma / \sqrt{N}} = 2.55$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359		1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753		1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141		1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517		1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879		1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224		2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549		2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852		2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133		2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389		2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621		2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830		2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4951	0.4962	0.4963	0.4964
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015		2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177		2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319		2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
												3.0	0.4987	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

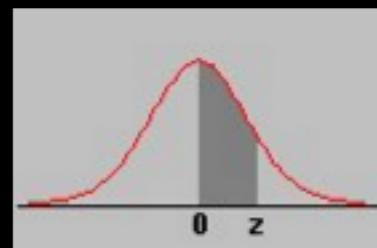
<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>

2 sided test $1 - (0.4948 * 2) = 0.0106$
 $p < 0.05$



H_0 IS REJECTED ($p < 0.05$)

IV: Statistical analysis

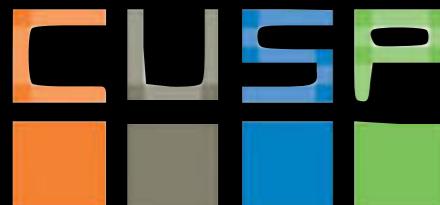


$$Z = \frac{\mu_{\text{pop}} - \mu_{\text{sample}}}{\sigma / \sqrt{N}} = 2.55$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879	1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852	2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4951	0.4962	0.4963	0.4964
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015	2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177	2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319	2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
											3.0	0.4987	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

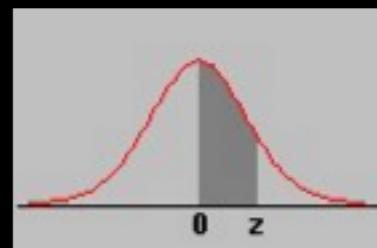
<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>

1 sided test $1 - 0.4948 - 0.5 = 0.0054$
 $p < 0.05$



H_0 IS REJECTED ($p < 0.05$)

IV: Statistical analysis

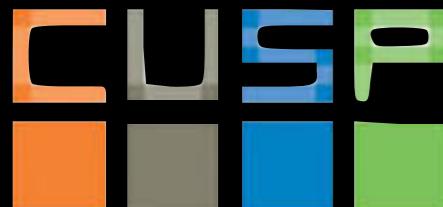


$$Z = \frac{\mu_{\text{pop}} - \mu_{\text{sample}}}{\sigma / \sqrt{N}} = 1.57$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879	1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852	2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015	2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177	2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319	2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
											3.0	0.4987	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

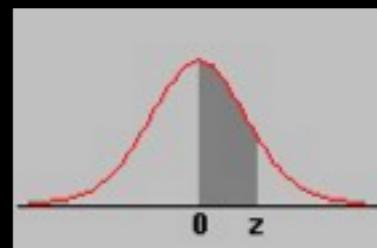
<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>

2 sided test $1 - 0.4418^2 = 0.1164$
 $p > 0.05$



H_0 CANNOT BE REJECTED

IV: Statistical analysis

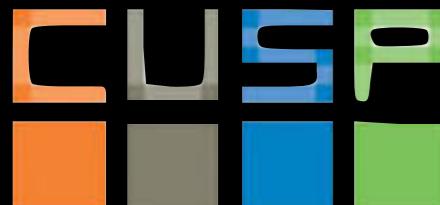


$$Z = \frac{\mu_{\text{pop}} - \mu_{\text{sample}}}{\sigma / \sqrt{N}} = 1.57$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879	1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852	2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015	2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177	2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319	2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
											3.0	0.4987	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

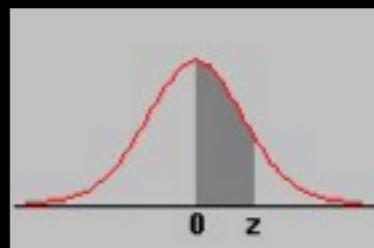
<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>

2 sided test $1 - 0.4418^2 = 0.1164$
 $p > 0.05$



H_0 CANNOT BE REJECTED

IV: Statistical analysis



$$Z = \frac{\mu_{\text{pop}} - \mu_{\text{sample}}}{\sigma / \sqrt{N}} = 1.96$$

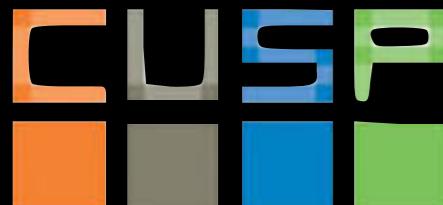
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359	1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4750	0.4418	0.4429	0.4441
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753	1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141	1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517	1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1871	1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224	2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549	2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852	2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133	2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389	2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621	2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830	2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015	2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177	2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319	2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
											3.0	0.4987	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990

<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>

2 sided test

$$1 - 0.4750^2 = 0.05$$

$$p = 0.05$$



H_0 CANNOT BE REJECTED

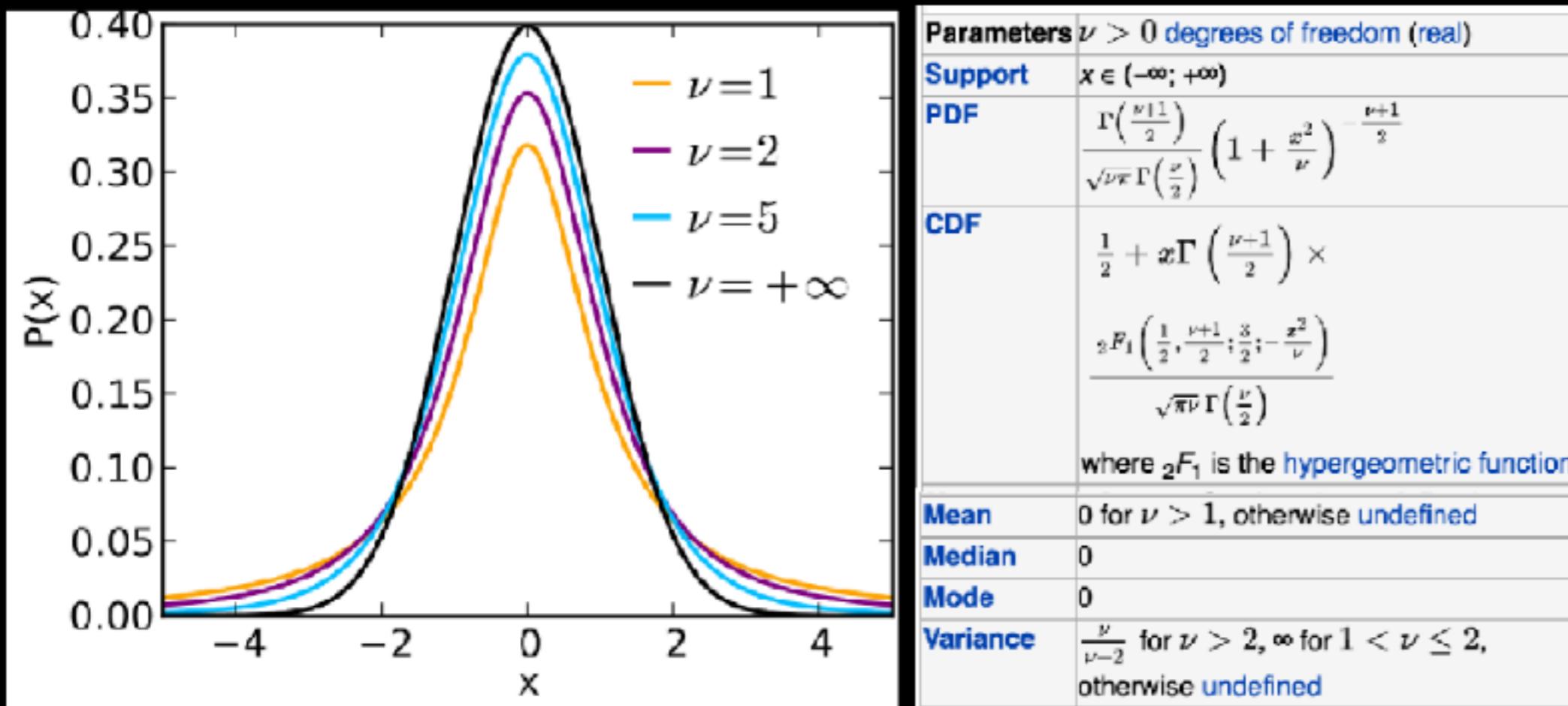
IV: Statistical analysis

Is there a difference between means of 2 sample?

Student's t

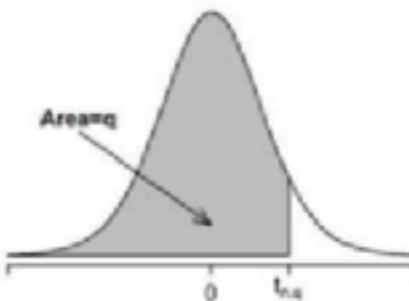
$$t = \frac{\mu - \bar{x}}{s/\sqrt{n}}$$

In absence of effect (i.e. under the Null)
 t is distributed according to a *student's t* distribution
 with ν number of degrees of freedom



$$t = \frac{\mu - \bar{x}}{s/\sqrt{n}} = 1.75$$

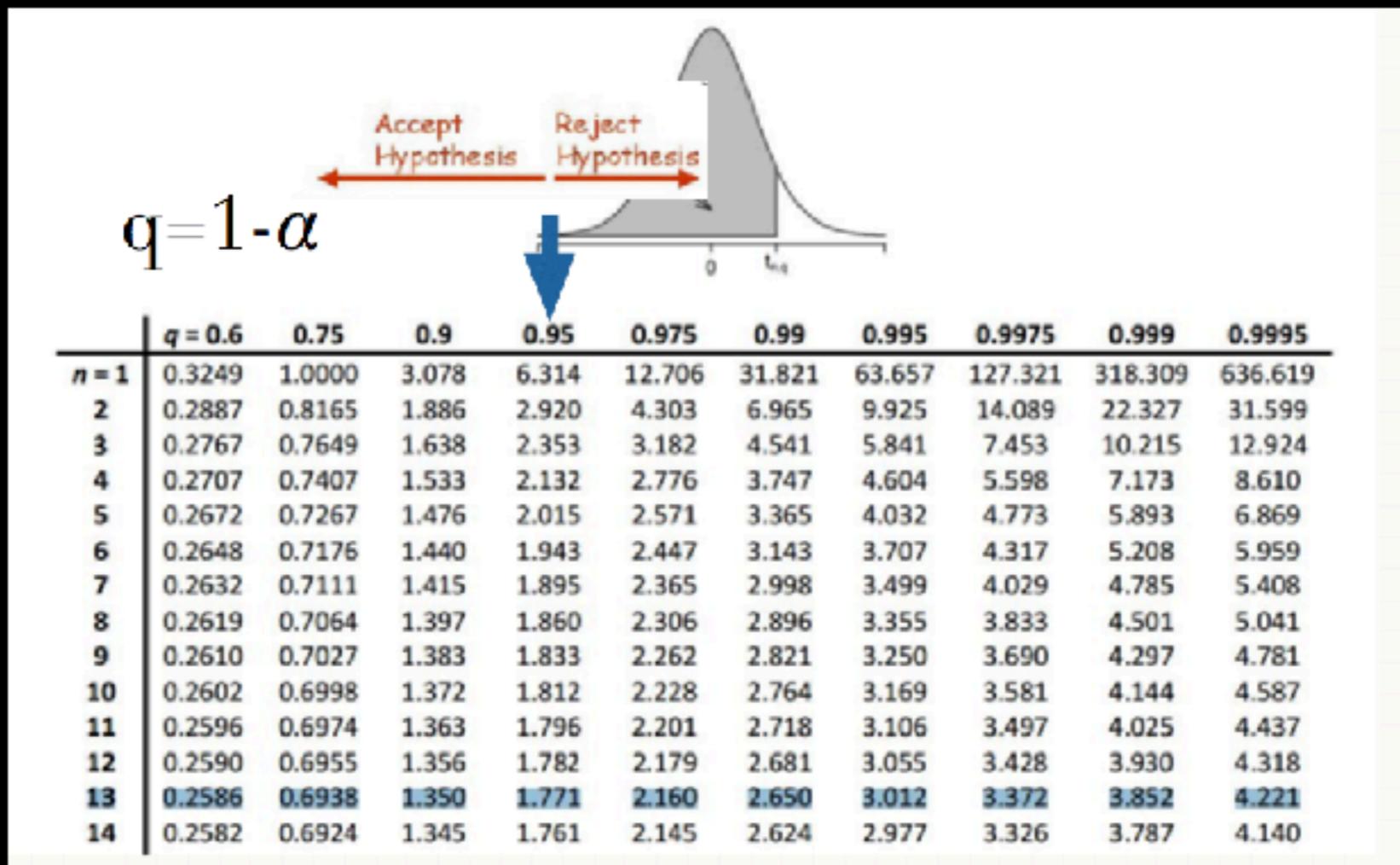
$n=13$,
 $q=.95$



	$q = 0.6$	0.75	0.9	0.95	0.975	0.99	0.995	0.9975	0.999	0.9995
$n = 1$	0.3249	1.0000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.2887	0.8165	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.2767	0.7649	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.2707	0.7407	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.2672	0.7267	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.2648	0.7176	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.2632	0.7111	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.2619	0.7064	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.2610	0.7027	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.2602	0.6998	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.2596	0.6974	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.2590	0.6955	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.2586	0.6938	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.2582	0.6924	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140

$$t = \frac{\mu - \bar{x}}{s/\sqrt{n}} = 1.75$$

$n=13$, $\alpha=.95$



Is $1.75 > 1.771$? NO:

Cannot reject the Null Hypothesis

Is there a difference between means or population and sample,
difference between proportion in 2 samples?

χ^2 statistics

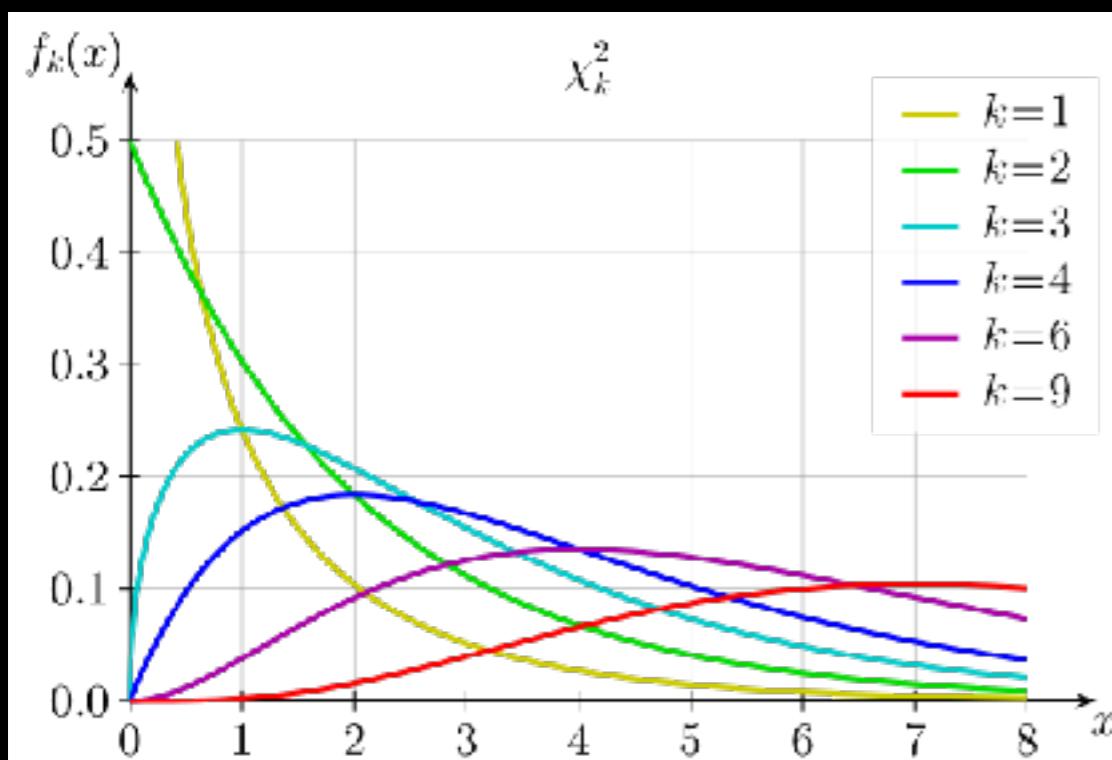
$$\chi_P^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}$$

In absence of effect (i.e. under the Null)
 χ^2 is distributed according to a χ^2 distribution
 with $k = \text{number of degrees of freedom}$

O : observed

E : expected (also model prediction)

N observations



Notation	$\chi^2(k)$ or χ_k^2
Parameters	$k \in \mathbb{N}_{>0}$ (known as "degrees of freedom")
Support	$x \in [0, +\infty)$
PDF	$\frac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$
CDF	$\frac{1}{\Gamma\left(\frac{k}{2}\right)} \gamma\left(\frac{k}{2}, \frac{x}{2}\right)$
Mean	k
Median	$\approx k \left(1 - \frac{2}{9k}\right)^3$
Mode	$\max\{k-2, 0\}$
Variance	$2k$
Skewness	$\sqrt{8/k}$

observed *expected*

$$\chi_P^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i} = 8.57$$

For test of proportion

4 observations - 1 independent variable =
1 degree of freedom $\alpha = 0.05$

Accept Hypothesis **Reject Hypothesis**

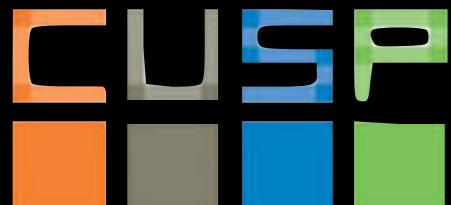
← →

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of χ^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
3	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
5	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09

<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>

$$8.57 > 3.84$$



H_0 REJECTED $p < 0.05$

IV: Statistical analysis

model prediction observations

$$\chi^2_F = \sum_{i=1}^4 \frac{(m_i - x_i)^2}{e_i} = 8.57$$

$e_i \rightarrow$ errors

For goodness of fit

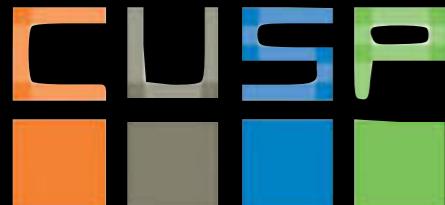
4 observations - 1 independent variable =
3 degrees of freedom $\alpha = 0.05$

The diagram shows a Chi-Square distribution table with a red arrow pointing to the 0.05 column and another red circle highlighting the 0.05 value.

Degrees of Freedom	Probability of a larger value of χ^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09

<https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md>

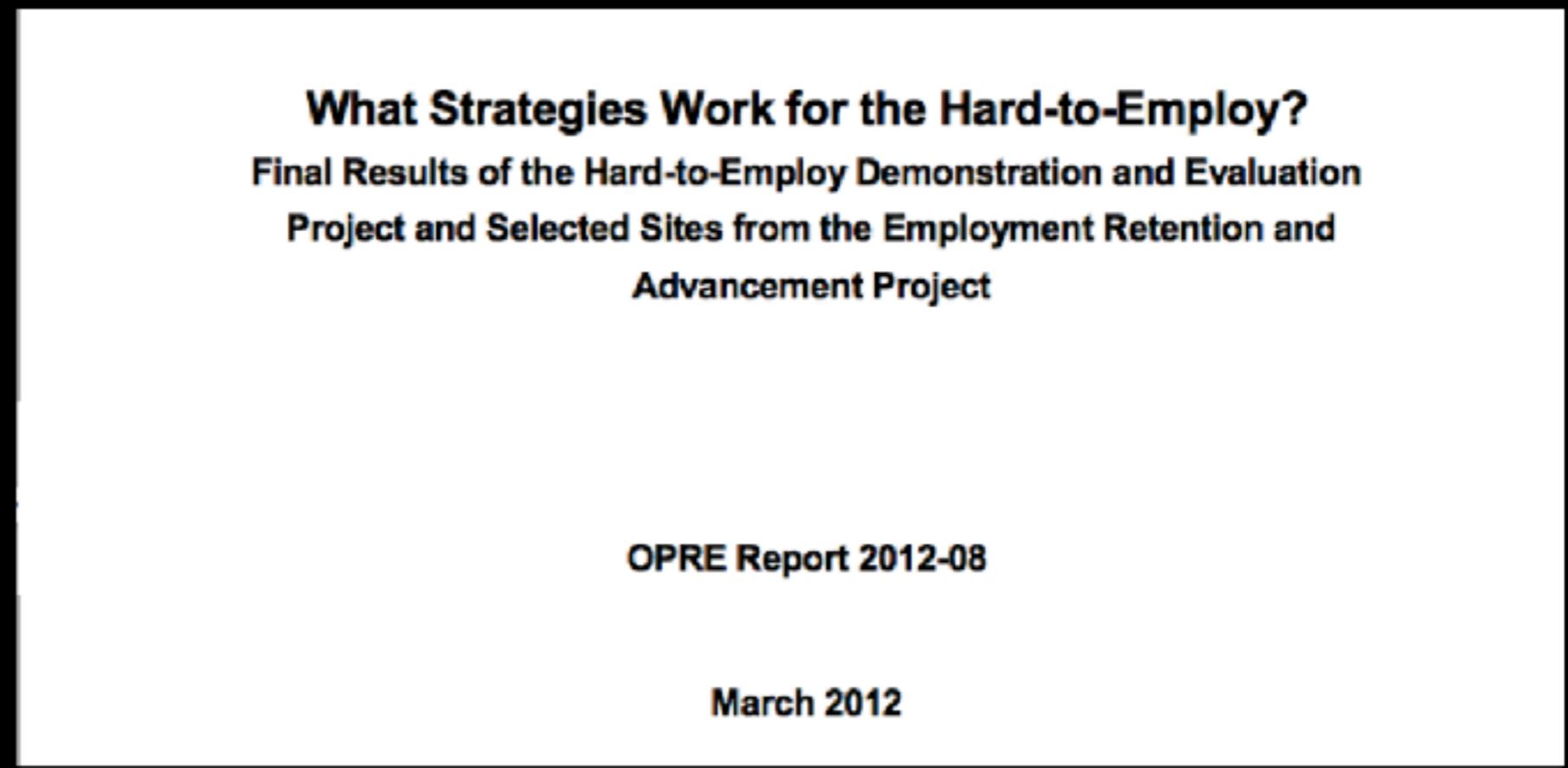
$$8.57 > 7.81$$



H_0 REJECTED $p < 0.05$

IV: Statistical analysis

Example: **NULL HYPOTHESIS:** the % of former prisoners employed 3 years after release is *the same or lower* for candidates who participated in the program as for the control group,
significance level p=0.05



<http://www.mdrc.org/sites/default/files/What%20Strategies%20Work%20for%20the%20Hard%20FR.pdf>

https://github.com/fedhere/PUI2017_fb55/blob/master/Lab4_fb55/effectiveness%20of%20NYC%20Post-Prison%20Employment%20Programs.ipynb

NULL HYPOTHESIS: the % of former prisoners employed 3 years after release is *the same or lower* for candidates who participated in the program as for the control group,
significance level p=0.05

The Enhanced Services for the Hard-to-Employ Demonstration and Evaluation Project

Table 2.1

Summary of Impacts, New York City Center for Employment Opportunities

Outcome	Program Group	Control Group	Difference (Impact)	P-Value
<u>Employment (Years 1-3) (%)</u>	P₁	P₀		
Ever employed	83.8	70.4	13.4 ***	0.000
Ever employed in a CEO transitional job ^a	70.1	3.5	66.6 ***	0.000
Ever employed in an unsubsidized job	63.7	69.0	-5.3 *	0.078
<u>Postprogram unsubsidized employment (Years 2-3)</u>				
Ever employed in an unsubsidized job (%)	53.3	52.1	1.2	0.713
Employed in an unsubsidized job, average per quarter (%)	28.2	27.2	1.1	0.618
Employed for six or more consecutive quarters (%)	14.7	11.9	2.8	0.195
Total UI-covered earnings ^b (\$)	10,435	9,846	589	0.658
Sample size (total = 973) ^c	564	409		

$$H_0: P_0 - P_1 \geq 0$$

$$H_a: P_0 - P_1 < 0$$

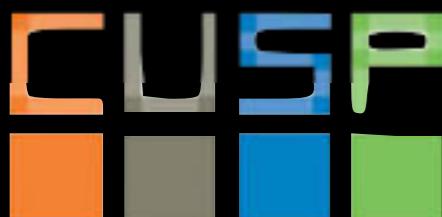
$$\alpha = 0.05$$

SOURCES: MDRC earnings calculations from the National Directory of New Hires (NDNH) database and employment calculations from the unemployment insurance (UI) wage records from New York State, MDRC calculations using data from the New York State Division of Criminal Justice Services (DCJS) and the New York City Department of Correction (DOC).

NOTES: Statistical significance levels are indicated as: *** = 1 percent; ** = 5 percent; * = 10 percent.

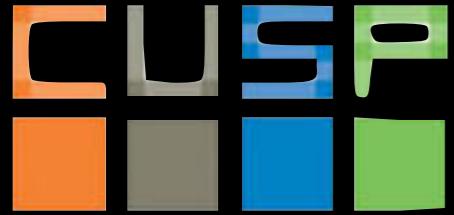
The p-value indicates the likelihood that the difference between the program and control groups arose by chance.

<http://www.mdrc.org/sites/default/files/What%20Strategies%20Work%20for%20the%20Hard%20FR.pdf>





https://github.com/fedhere/PUI2017_fb55/blob/master/Lab4_fb55/effectiveness%20of%20NYC%20Post-Prison%20Employment%20Programs.ipynb



IV: Statistical analysis

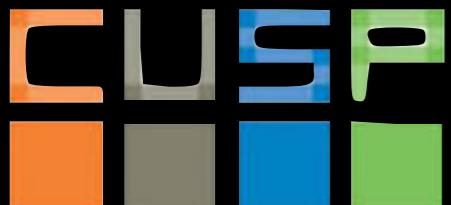
BEYOND SIGNIFICANCE TESTING

**REFORMING DATA ANALYSIS METHODS
IN BEHAVIORAL RESEARCH**

REX B. KLINE

РЕХ В. КЛІНЕ

[https://www.dropbox.com/s/0z74eh210fgucl5/
Screenshot%202017-10-05%2000.23.12.png?dl=0](https://www.dropbox.com/s/0z74eh210fgucl5/Screenshot%202017-10-05%2000.23.12.png?dl=0)



Recommendations

Specific suggestions are listed and then discussed afterward:

1. Only in very exploratory research where it is unknown whether effects exist may a primary role for NHST be appropriate.
2. If statistical tests are used, (a) information about power must be reported, and (b) the null hypothesis must be plausible.
3. In any kind of behavioral research, it is not acceptable anymore to describe results solely in terms of NHST outcomes.
4. Drop the word “significant” from our data analysis vocabulary. Use it only in its everyday sense to describe something actually noteworthy or important.
5. It is the researcher’s responsibility to report and interpret, whenever possible, effect size estimates and confidence intervals for primary results. This does not mean to report effect sizes only for H_0 rejections.
6. It is also the researcher’s responsibility to demonstrate the substantive (theoretical, clinical, or practical) significance of the results. Statistical tests are inadequate for this purpose.
7. Replication is the best way to deal with sampling error.
8. Education in statistical methods needs to be reformed, too. The role of NHST should be greatly deemphasized so that more time can be spent showing students how to determine whether a result has substantive significance and how to replicate it.
9. Researchers need more help from their statistical software to compute effect sizes and confidence intervals.

compute effect sizes and confidence intervals.
Researchers need more help from their statistical software to
compute effect sizes and confidence intervals.

whether a result has substantive significance and how to replicate it.
more time can be spent showing students how to determine whether a result has substantive significance and how to replicate it.

Arguments against NHST

1. $P(H_0 | D) = \frac{P(H_0) P(D | H_0)}{P(D)}$ $P(H_0 | D) \neq P(D | H_0)$
2. *Significance depends on sample size, e.g.:*

$$Z = \frac{|\mu_{\text{pop}} - \mu_{\text{sample}}|}{\sigma / \sqrt{N}}$$

If you measure the same effect in two experiments with 2 samples of different size, the significance of the effect will be different

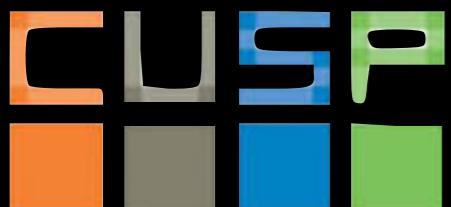
Effect size to measure the strength of a phenomenon

Estimate the strength of, for example, an apparent relationship, rather than assigning a significance level. The effect size does not directly determine the significance level, or vice versa.

Jacob Cohen (1960), "*A coefficient of agreement for nominal scales*", Educational and Psychological Measurement, 20 (1): 37–46,



[https://github.com/fedhere/PUI2016_fb55/blob/master/
HW3_fb55/citibikes_gender.ipynb](https://github.com/fedhere/PUI2016_fb55/blob/master/ HW3_fb55/citibikes_gender.ipynb)

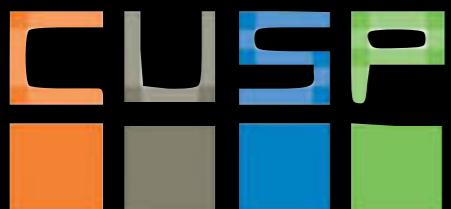


small,medium,large

Effect size to measure the strength of a phenomenon



Cohen's <i>d</i>	$d = \frac{ \bar{x}_1 - \bar{x}_2 }{s} \quad s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$	for Z and t	.20 .50 .80
Cohen's <i>w</i>	$w = \sqrt{\sum_{i=1}^N \frac{(p_{oi} - p_{1i})^2}{p_{oi}}}$	for X²	.10 .30 .50
Cohen's <i>h</i>	$h = 2(\arcsin \sqrt{p_1} - \arcsin \sqrt{p_2})$	for proportions	.20 .50 .80
<i>f</i> -squared	$\hat{f}_{effect} = \sqrt{(df_{effect}/N)(F_{effect}-1)}$	for F-test/ ANOVA	.10 .25 .40
R ² coefficient of determination		for regression	.02 .15 .35

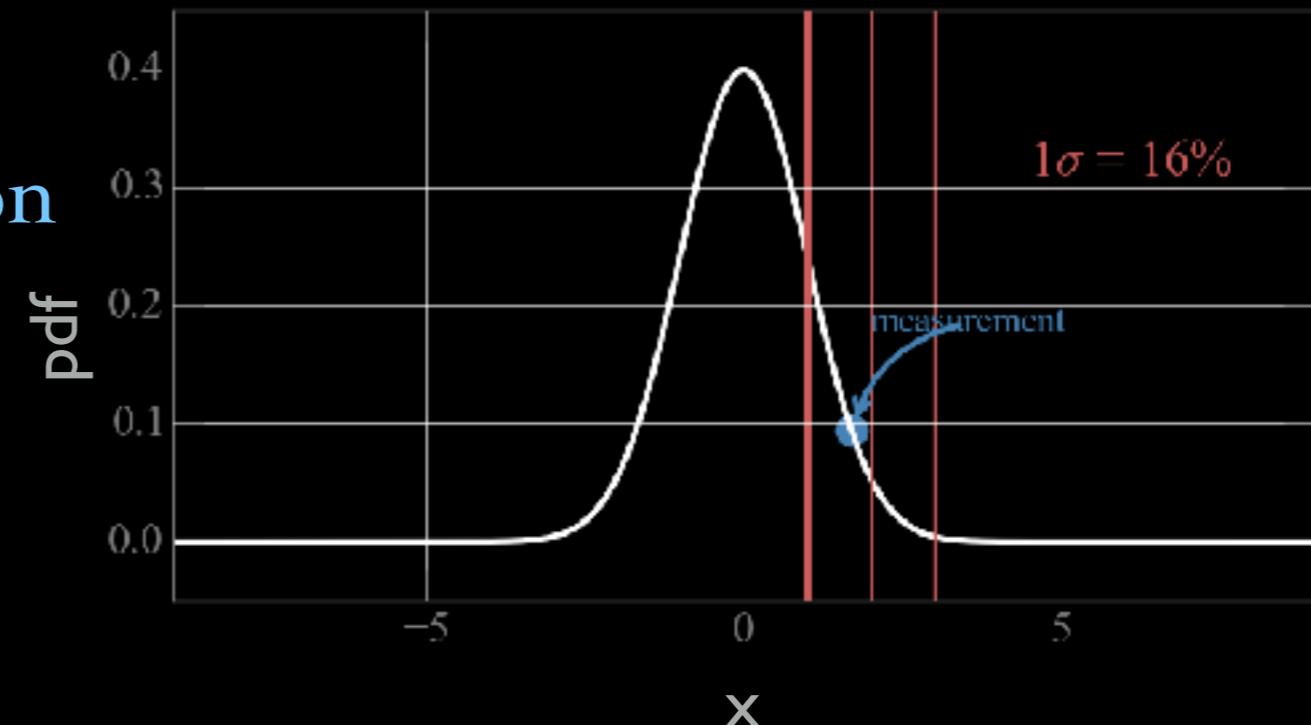


Tests for correlation and independence (continuous variables)

Probability Distribution Function

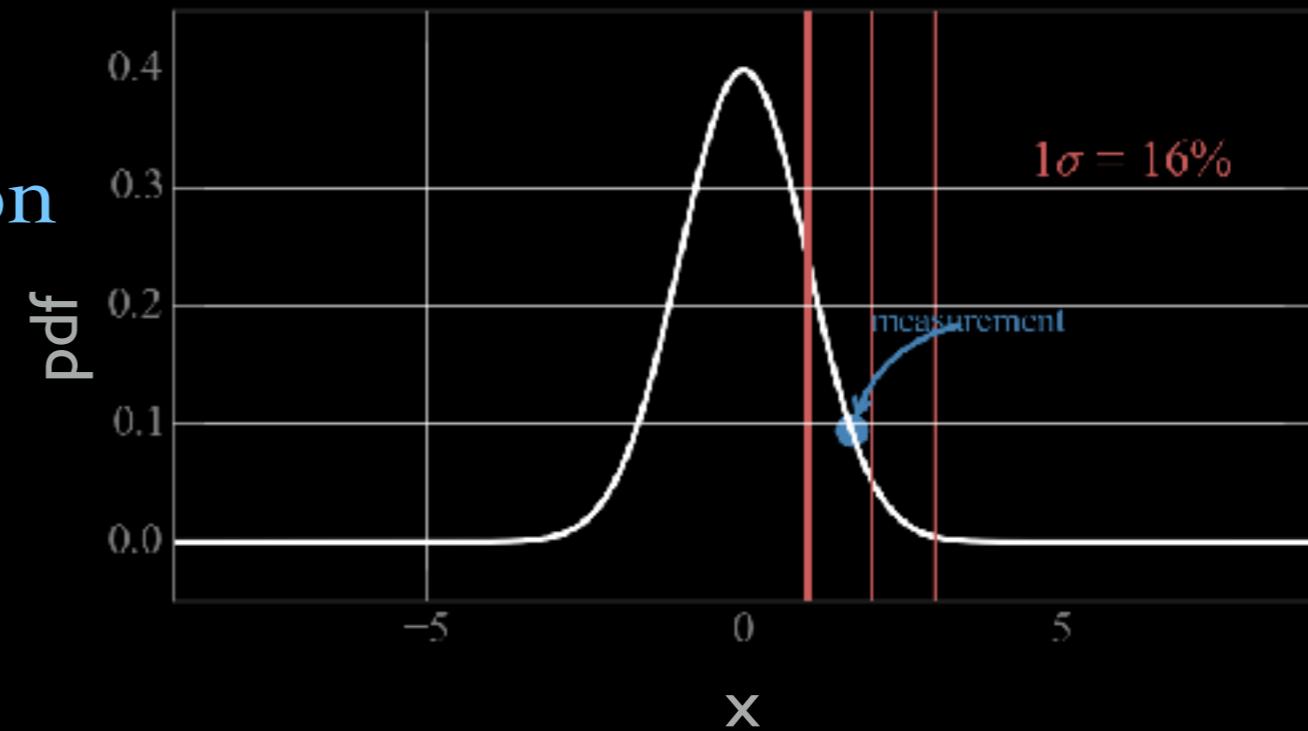
$$f_{x_0}(x) \sim p(x=x_0)$$

$$f_{x_0}(x) \sim p(x > x_0 - dx) \cap p(x < x_0 + dx)$$



Probability Distribution Function

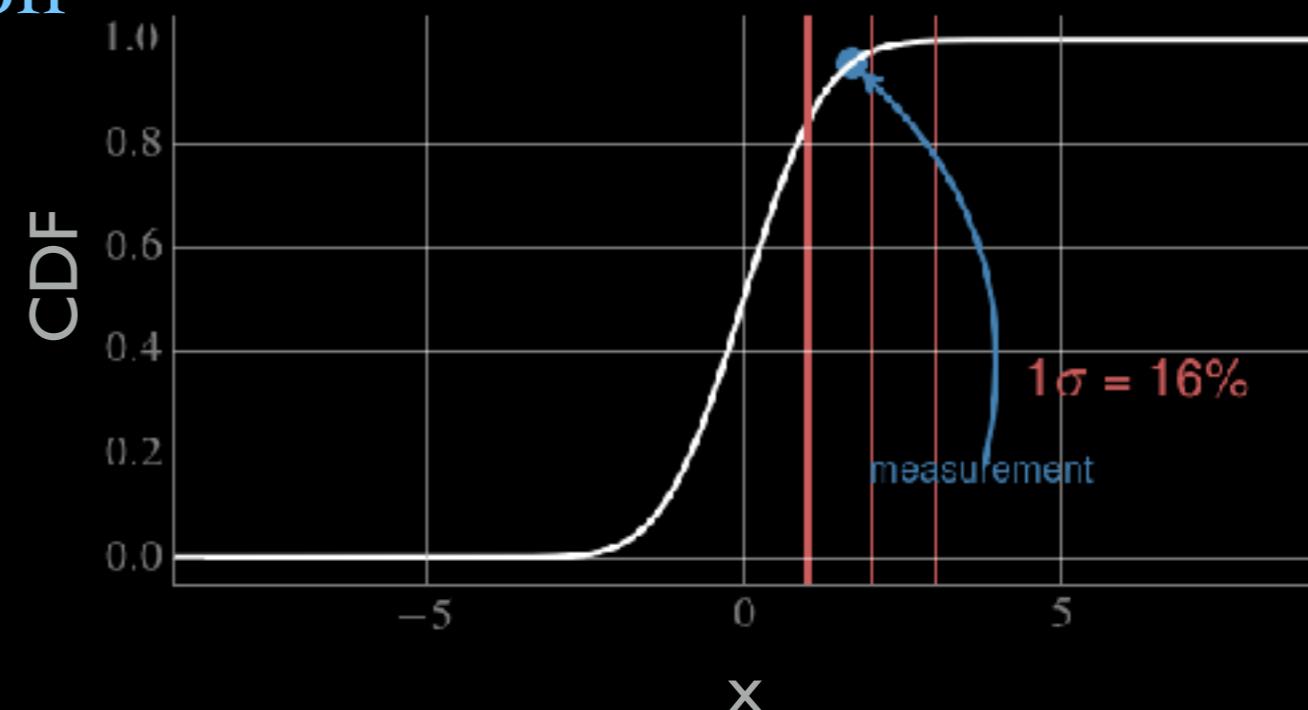
$$f_{x_0}(x) \sim p(x=x_0)$$

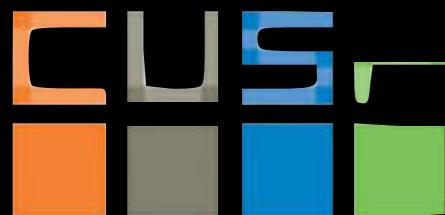
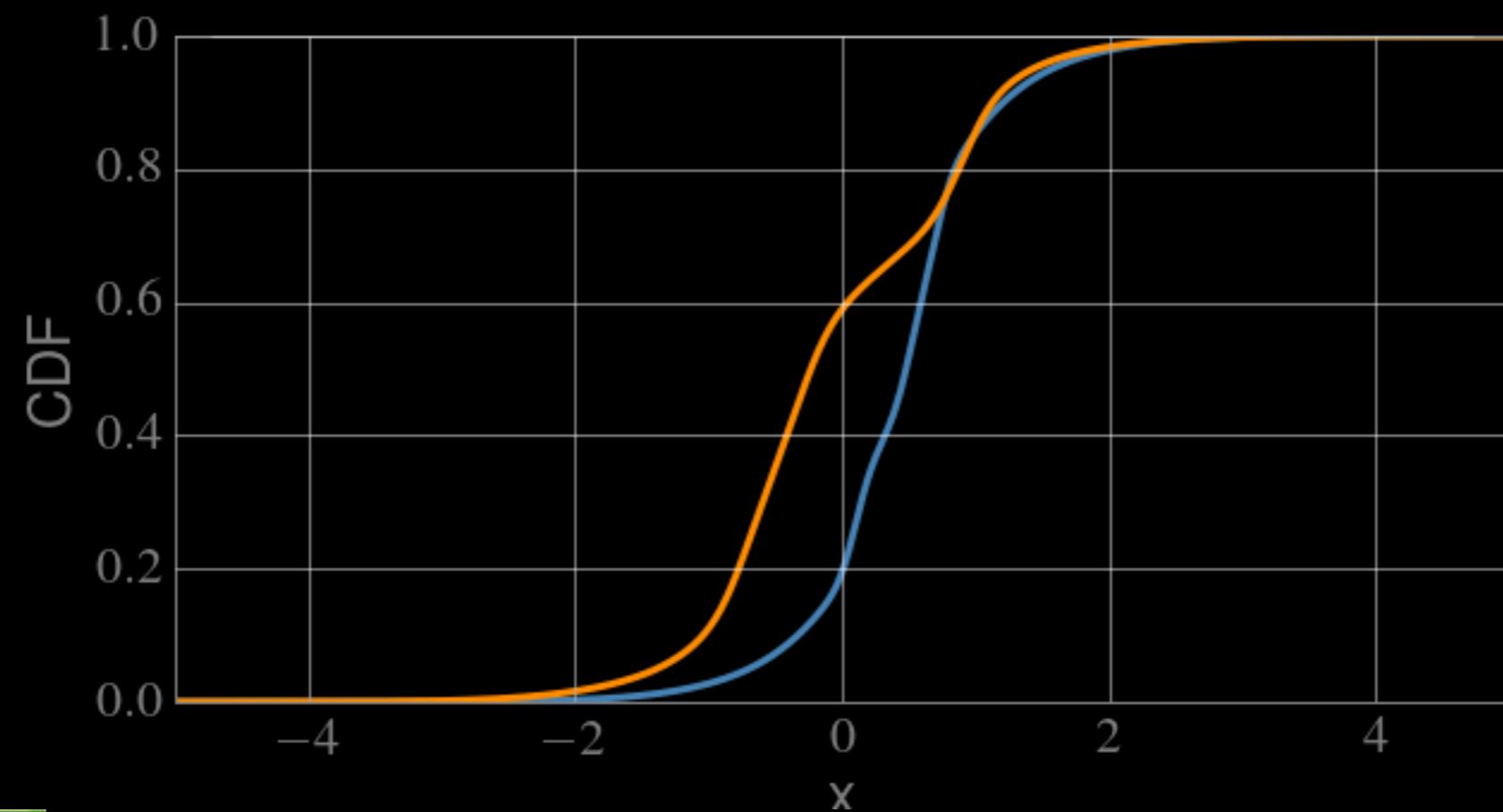
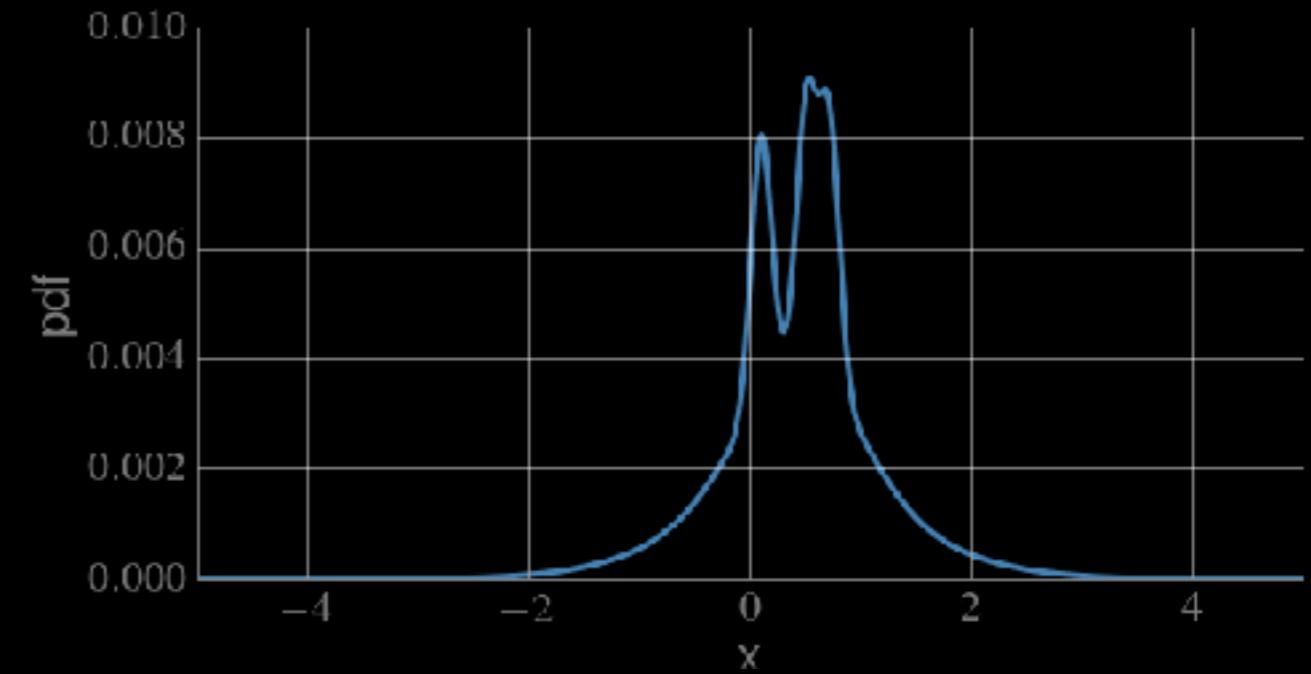
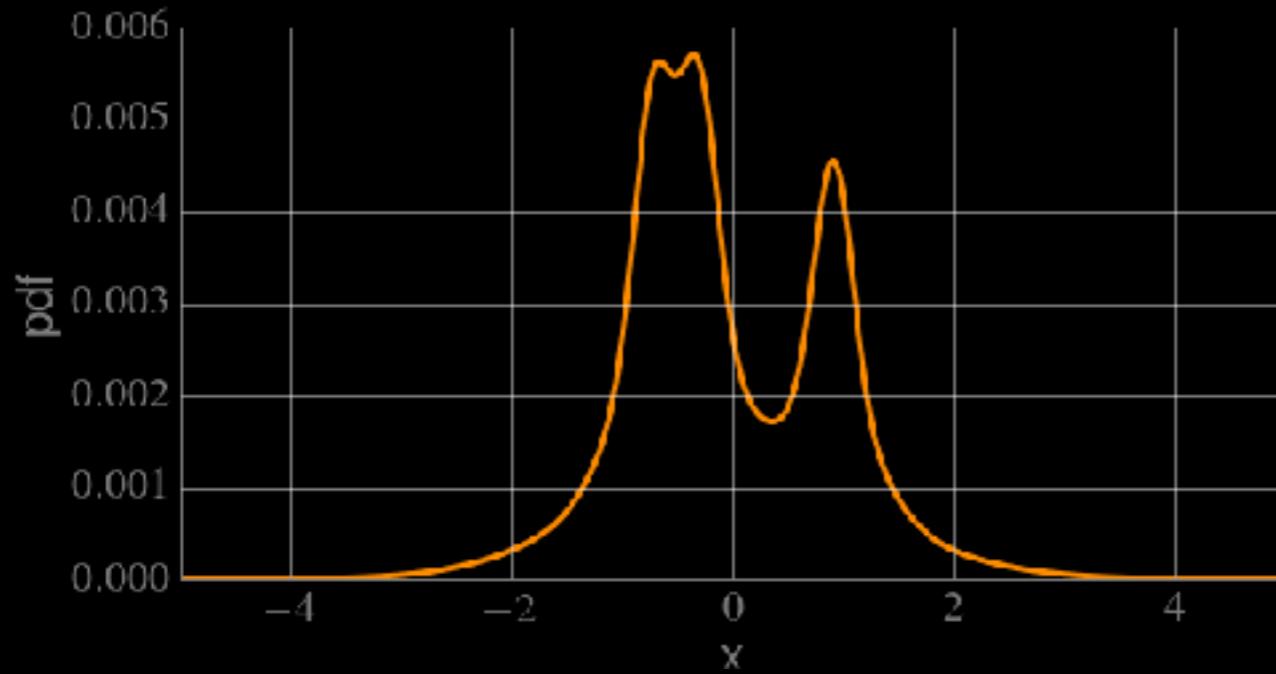


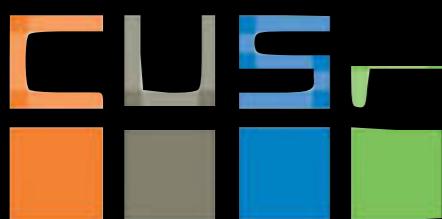
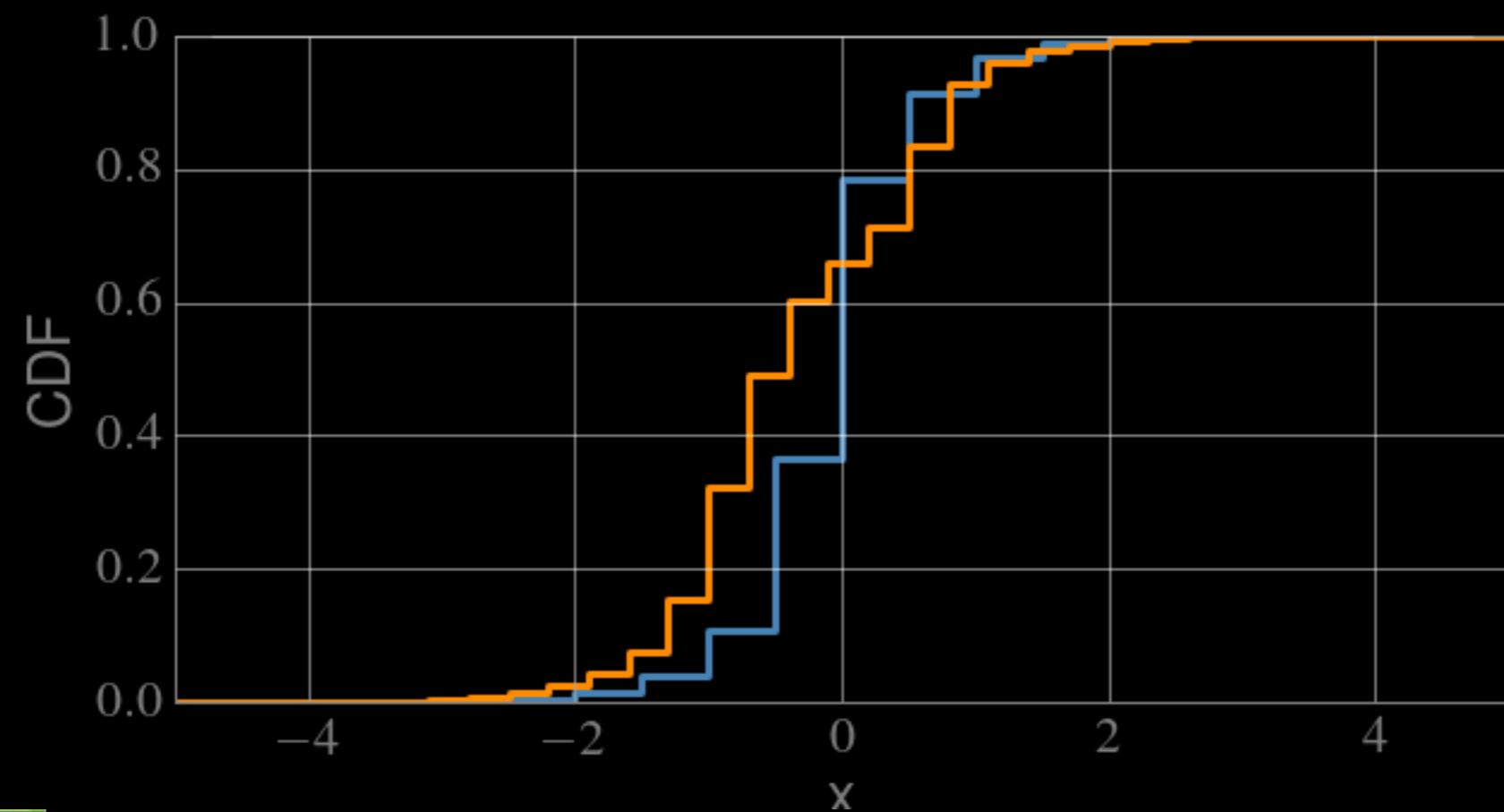
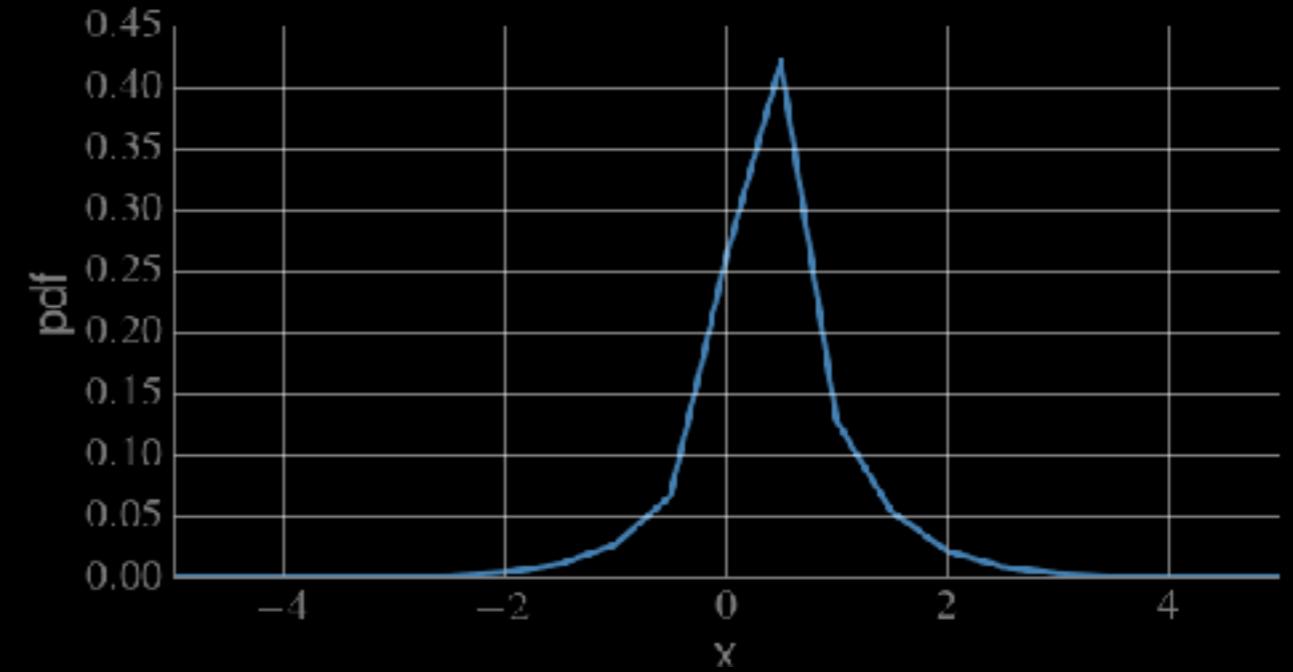
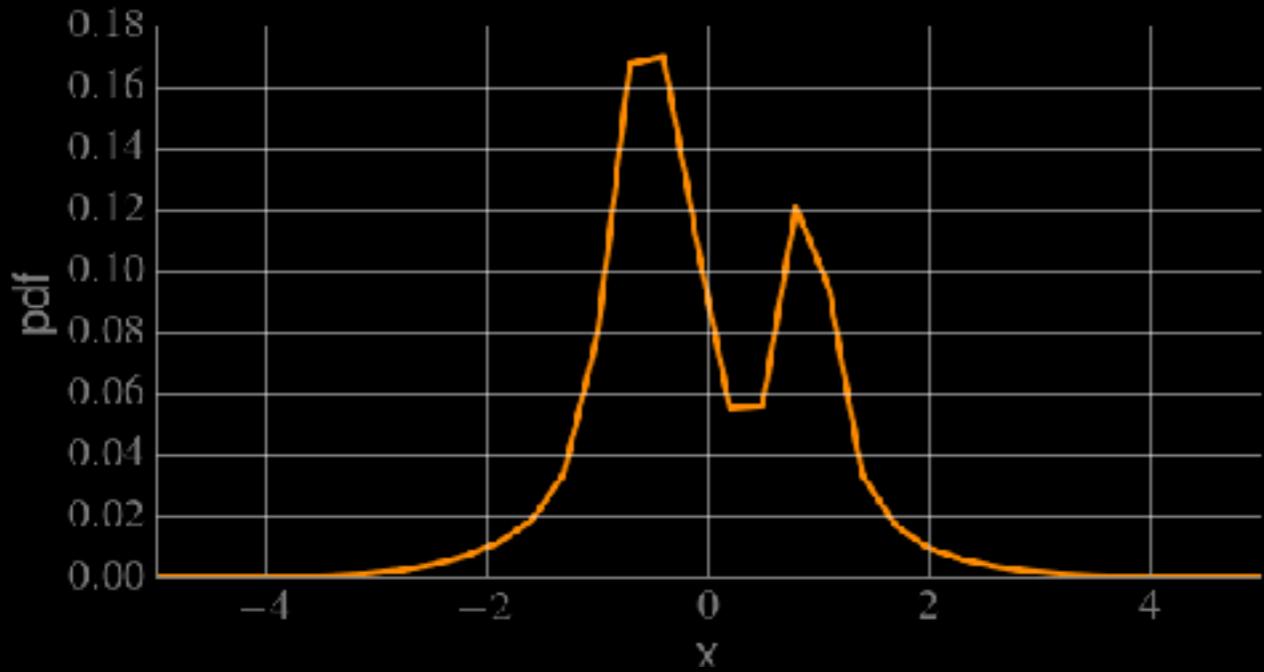
$$f_{x_0}(x) \sim p(x > x_0 - dx) \cap p(x < x_0 + dx)$$

Cumulative Distribution Function

$$F_{x_0}(x) = P(x < x_0)$$







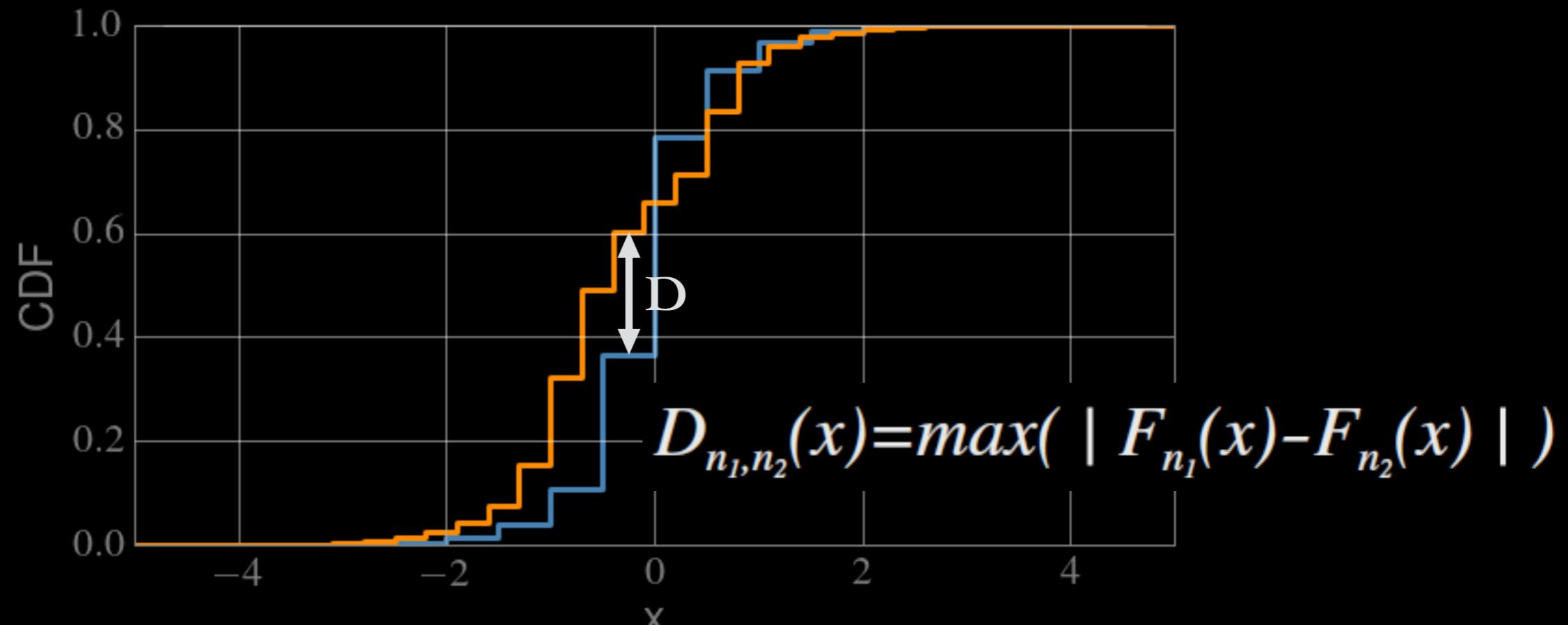
Two sample Kolmogorov Smirnoff test:

null hypothesis H_0 : the samples come from the same parent distribution

H_0 is rejected at level α if $D(n_1, n_2) > c(\alpha) \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$

with $c(\alpha)$ given by a table

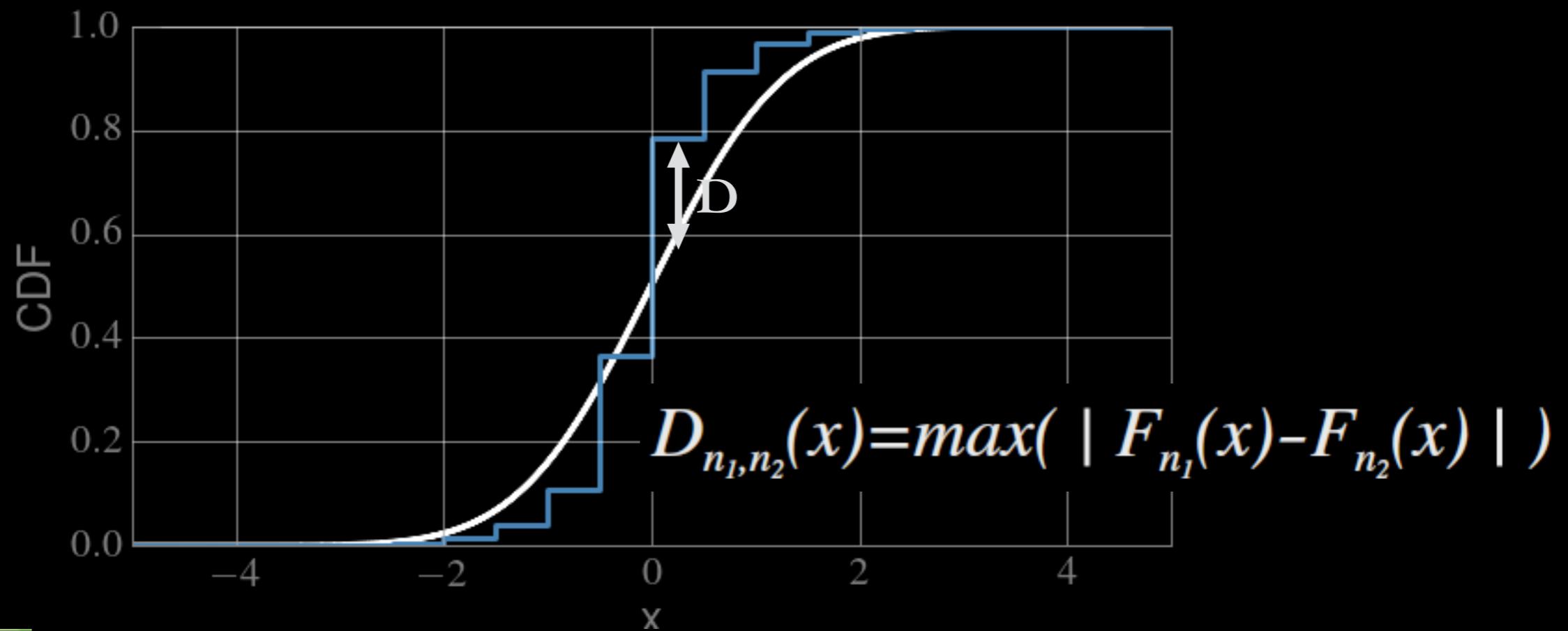
NOTE: it ONLY works in 2D where the Euclidian distance is uniquely defined!

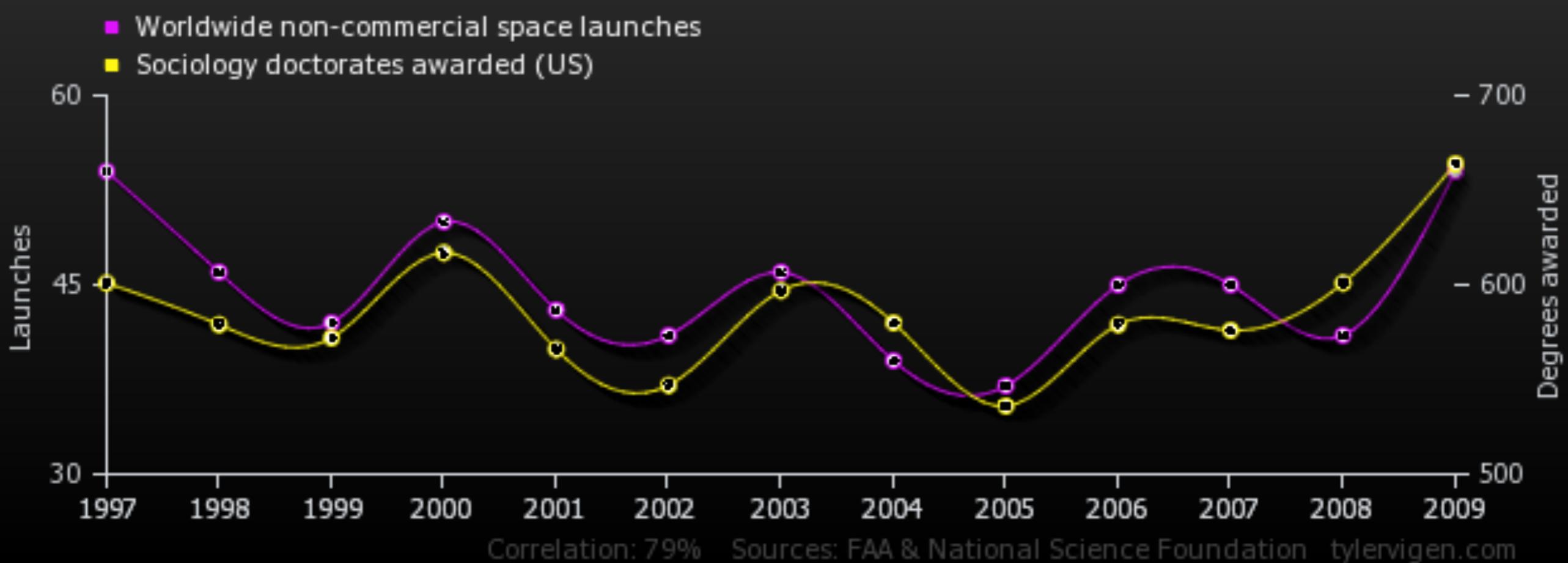


Goodness-of-fit Kolmogorov Smirnoff test:

null hypothesis H_0 : the sample does comes from the model distribution

H_0 is rejected at level α if $\sqrt{n} D_n > K_\alpha$ where $P(K \leq K_\alpha) = 1 - \alpha$

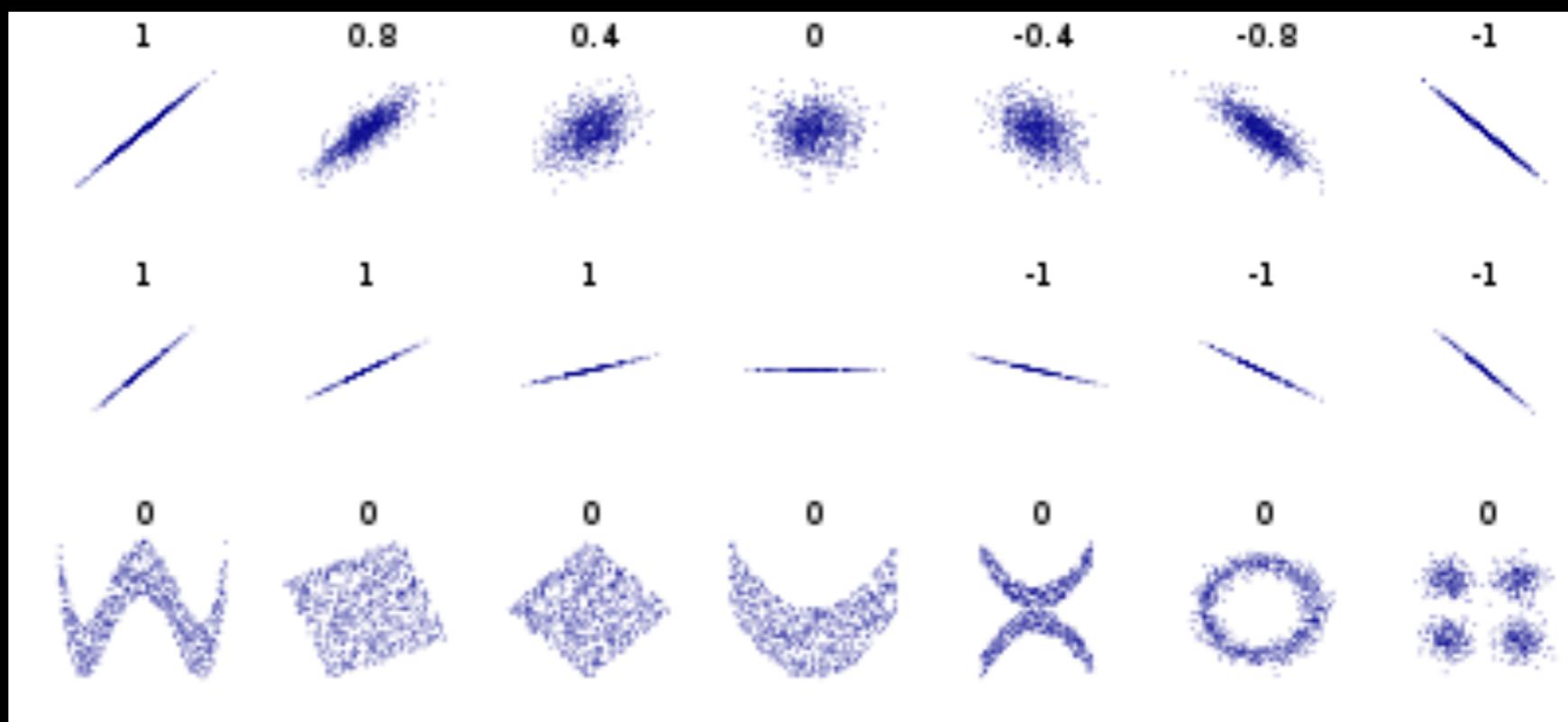




Correlation

Pearson's test:

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$
$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$



Pearson's test:

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

Spearman's test:
(Pearson's for ranks)

$$\rho = 1 - \frac{6 \sum (x_i - y_i)^2}{n(n^2 - 1)}$$

Choosing the test

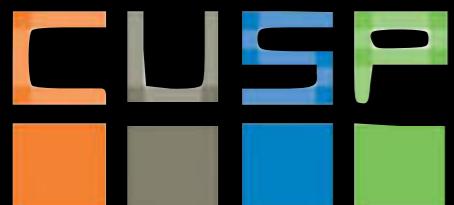
Use the table below to choose the test. See below for further details.

How many dichotomous* (binary) variables?				
Both variables interval or ratio?				
0	Y	Measures are linear? (No = monotonic*)		
		Y Pearson correlation		
0	N	N Spearman correlation		
		Both variables are ordinal?		
1	Y	Kendall correlation		
		Both variables can be ranked?		
1	N	Y Kendall correlation		
		N Convert to frequency data and use Chi-square test for independence		
1 serial Correlation Coefficient				
2 x 2 table?				
2	Y	Phi		
		N Cramer's V		
Data has frequency values for each category?				
Y	Chi-square test for independence			

*dichotomous = 'can have only two values' (eg. yes/no or 0/1).

†monotonic = constantly increasing or decreasing.

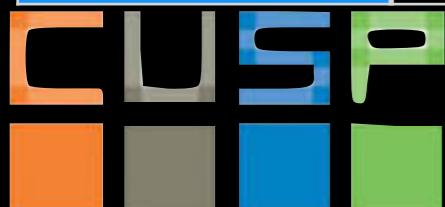
http://changingminds.org/explanations/research/analysis/choose_correlation.htm



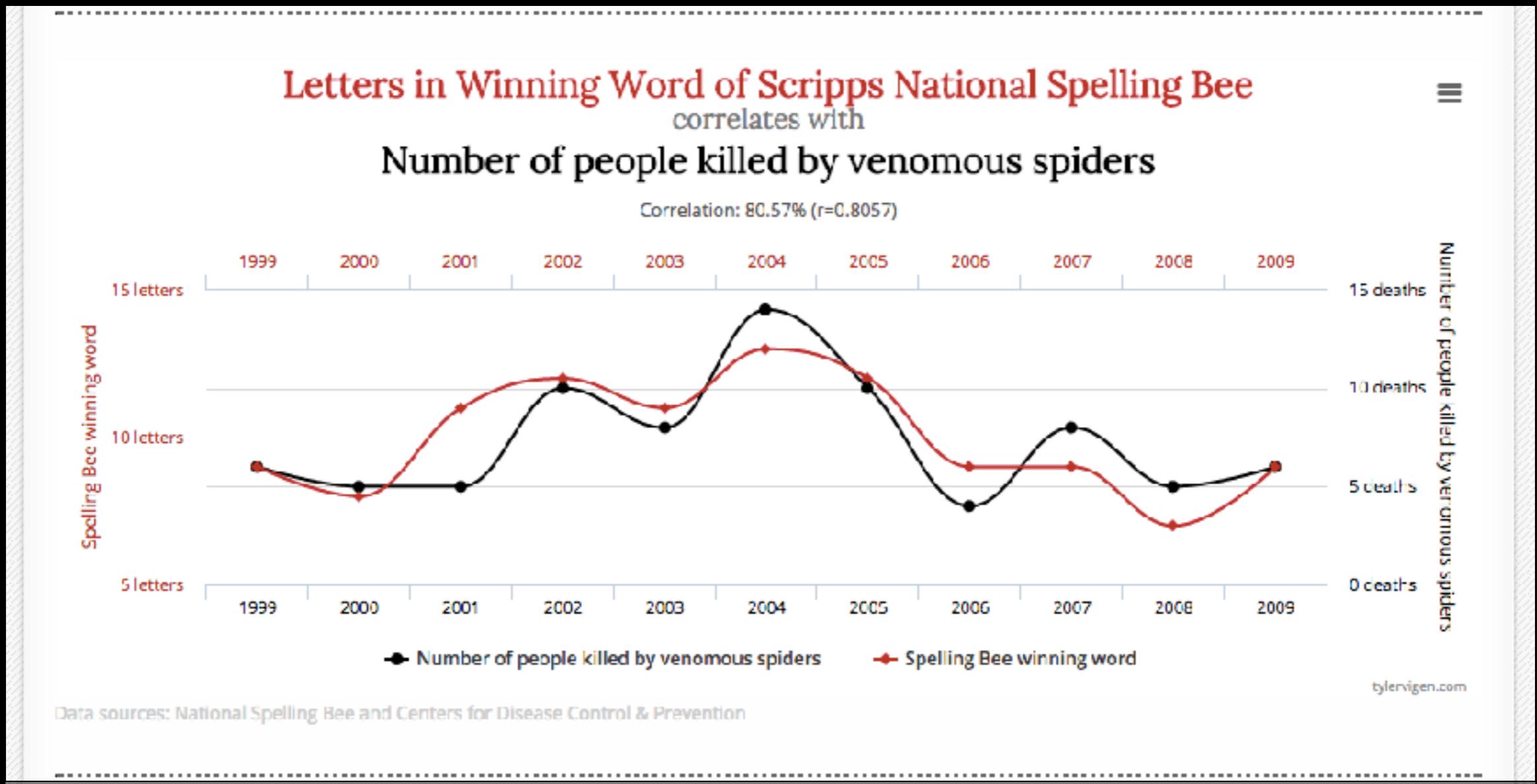
Tests Cheat Sheet:

2 (+) samples comparison

	metric (statistic)	compare to	
KS	$D_{n_1, n_2}(x) = \max(F_{n_1}(x) - F_{n_2}(x))$	$c(\alpha) \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$	Non parametric 2 samples only
K-sample Anderson-Darling	$ADK = \frac{n-1}{n^2(k-1)} \sum_{i=1}^k \frac{1}{n(i)} \left(\sum_{j=1}^L h_j \frac{(nF_{ij} - n_i H_j)^2}{H_j(n-H_j) - nh_j/4} \right)$	• AK table	Non parametric, N samples
Pearson's	$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$	The interpretation of a correlation coefficient depends on the context and purpose	-1 anticorrelated 0 uncorrelated 1 correlated .
Spearman's	$\rho = 1 - \frac{6 \sum (x_i - y_i)^2}{n(n^2 - 1)}$	t test $t = r \sqrt{\frac{n-2}{1-r^2}}$	ranked data only p-value from t-test, Fisher's transformation +z score, permutation test



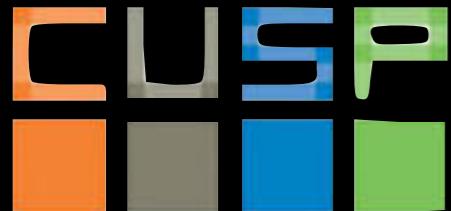
WARNING: Correlation is not causation!



<http://www.tylervigen.com/spurious-correlations>

ERRORS

What if our conclusion is after all wrong??
we have not ruled out the null, just decided that is it unlikely



ERRORS

<i>reality</i> → <i>result of analysis</i> ↓	H_0 is True	H_0 is False
H_0 is falsified	Type I error False Positive	True Positive
H_0 is not falsified	True Negative	Type II error False negative

ERRORS

	H_0 is True	H_0 is False
H_0 is falsified	Type I error False Positive important message gets spammed	True Positive
H_0 is not falsified	True Negative	Type II error False negative Spam in your Inbox

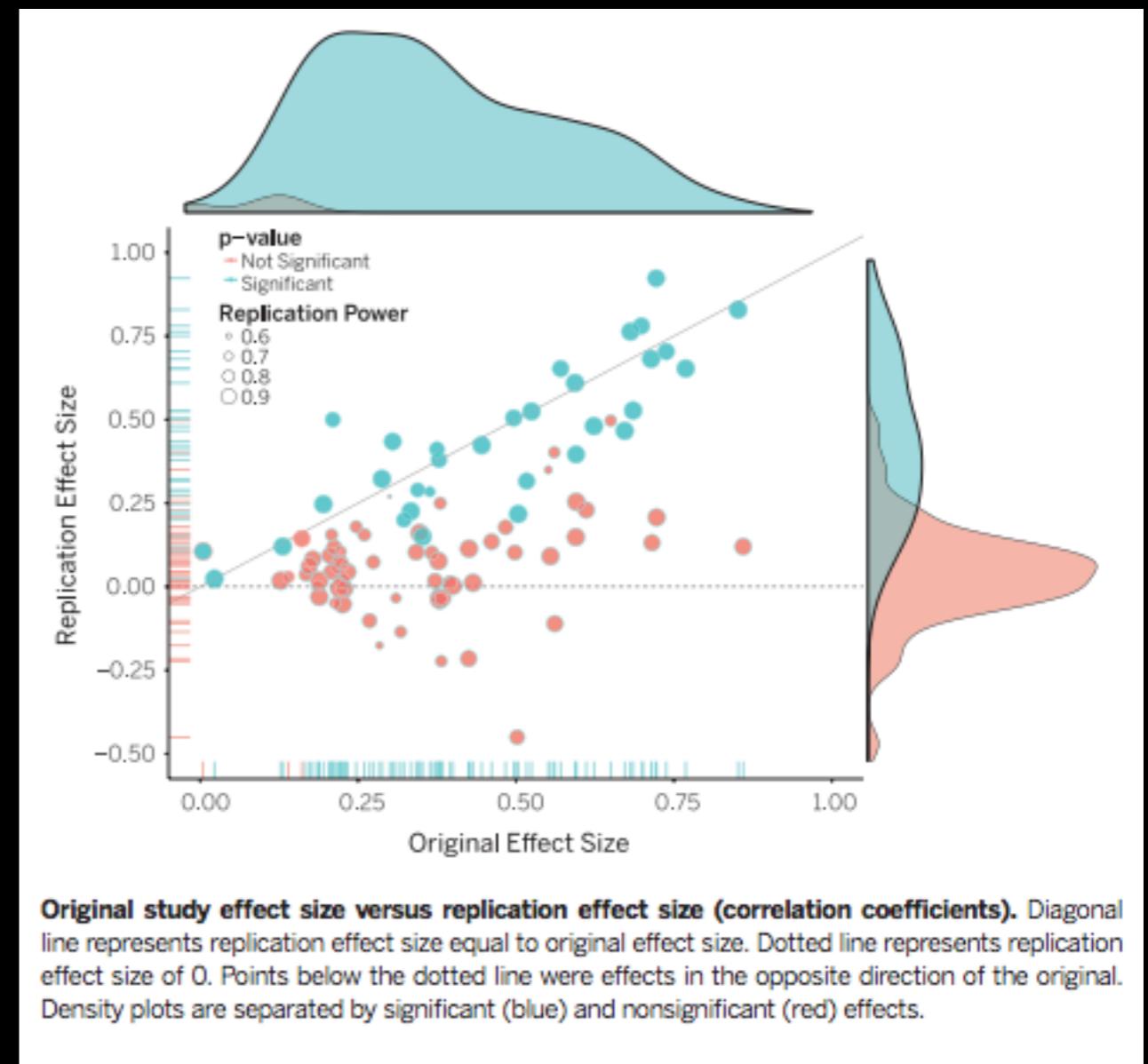
Homework: READING

RESEARCH ARTICLE SUMMARY

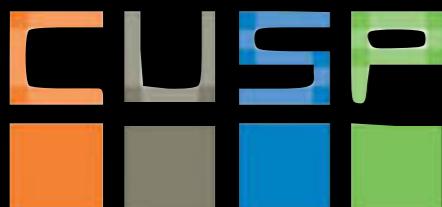
PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*



<http://www.sciencemag.org/content/349/6251/aac4716.full.pdf>



IV: Statistical analysis

Homework: 3. Compare Tests for Correlation

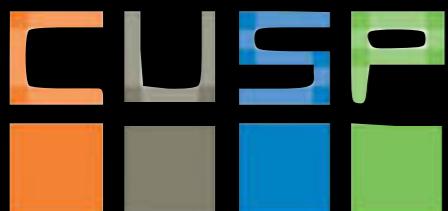
The following are 3 tests that assess correlation between 2 samples of citibike data:

- **Pearson's test** (answer: are the 2 samples correlated?)
- **Spearman's test** (answer: are the 2 samples correlated?)
- **K-S test** (answer: are the samples likely to come from the same parent distribution?)

Use: age of bikers for 2 genders, age of bikers in day vs night and assess the correlation/independence of the 2 samples in each case. State your result in words.

MUST KNOWS:

- p-values pros and cons
- How to chose a statistical test
- How to get from a statistics to a p-value
- Effect Size
- PDF vs CDF
- correlation vs causation
- KS test for 2 samples, Pearson's, Spearman's
- Type I & II errors and power of a test



Resources:

Sarah Boslaugh, Dr. Paul Andrew Watters, 2008

Statistics in a Nutshell (Chapters 3,4,5)

https://books.google.com/books/about/Statistics_in_a_Nutshell.html?id=ZnhgO65Pyl4C

David M. Lane et al.

Introduction to Statistics (XVIII)

http://onlinestatbook.com/Online_Statistics_Education.epub

<http://onlinestatbook.com/2/index.html>

Barun K. Nayak and Avijit Hazral

How to choose the right statistical test?

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3116565/>

