Urban Informatics

Fall 2015

dr. federica bianco fb55@nyu.edu

@fedhere

CUSP

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

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
- Statistical and Systematic errors
- Correlation vs Causation

hypothesis testing

did we detect a phenomenon?

null hypothesis: 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?

A simple (too simple?) answer

did we detect a phenomenon?

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 statistcs for *a sample* to be compared to the *parameter of a population*

OR

Measure a statistcs 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 statistcs for *a sample* to be compared to the *parameter of a population*

OR

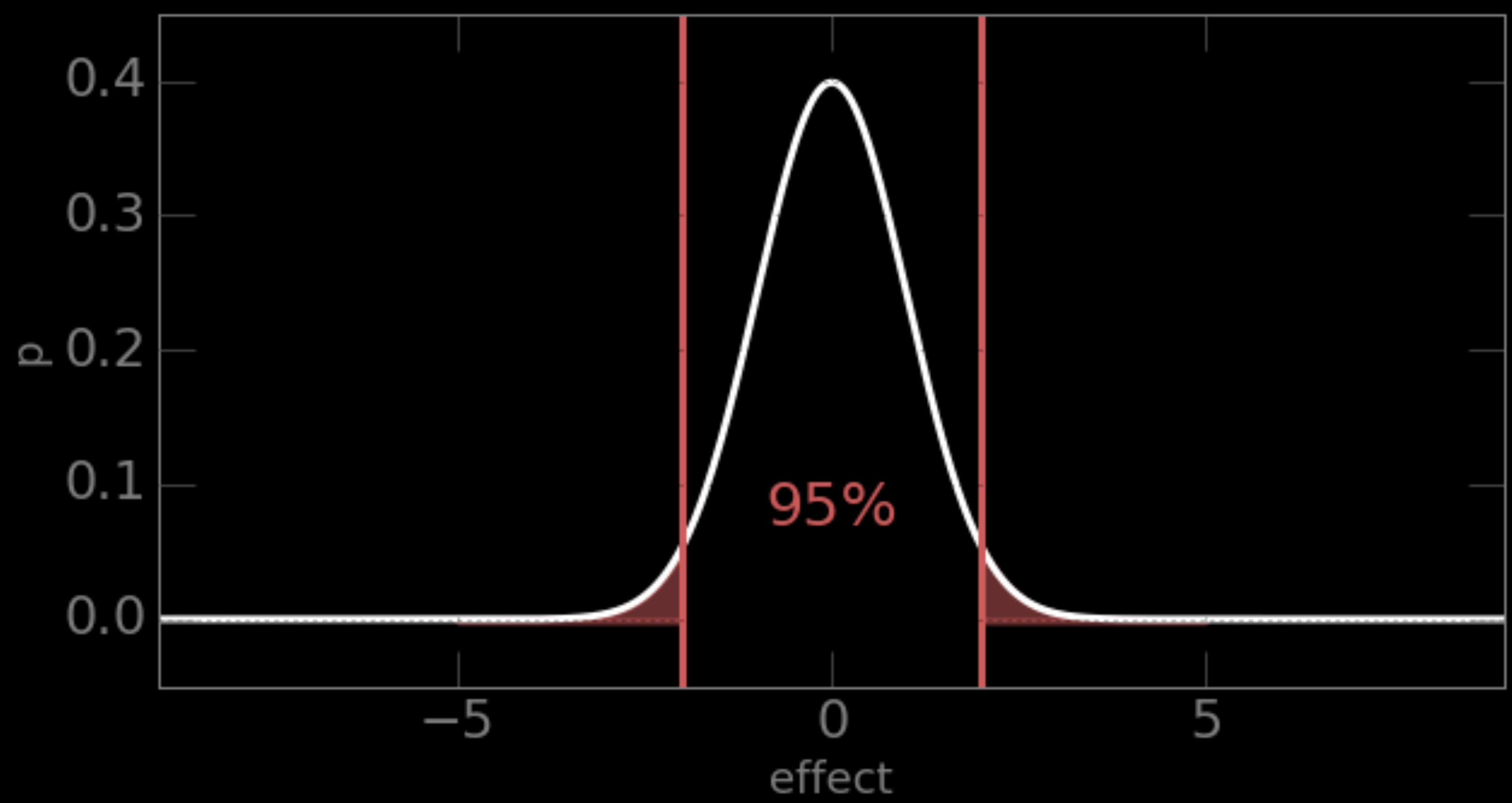
Measure a statistcs for *two or more samples* to be compared to *each other*

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…

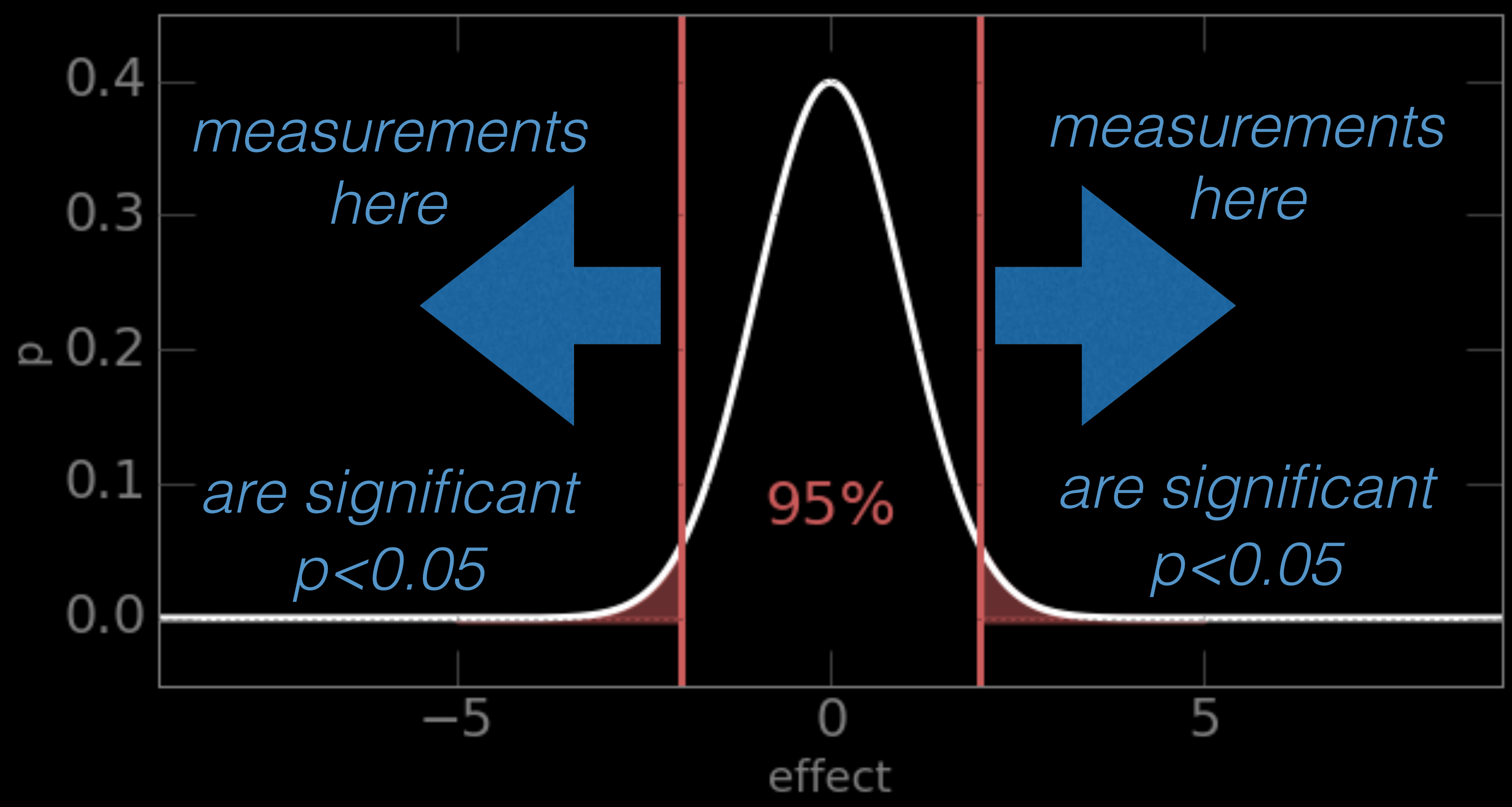
$\alpha = 0.05$

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



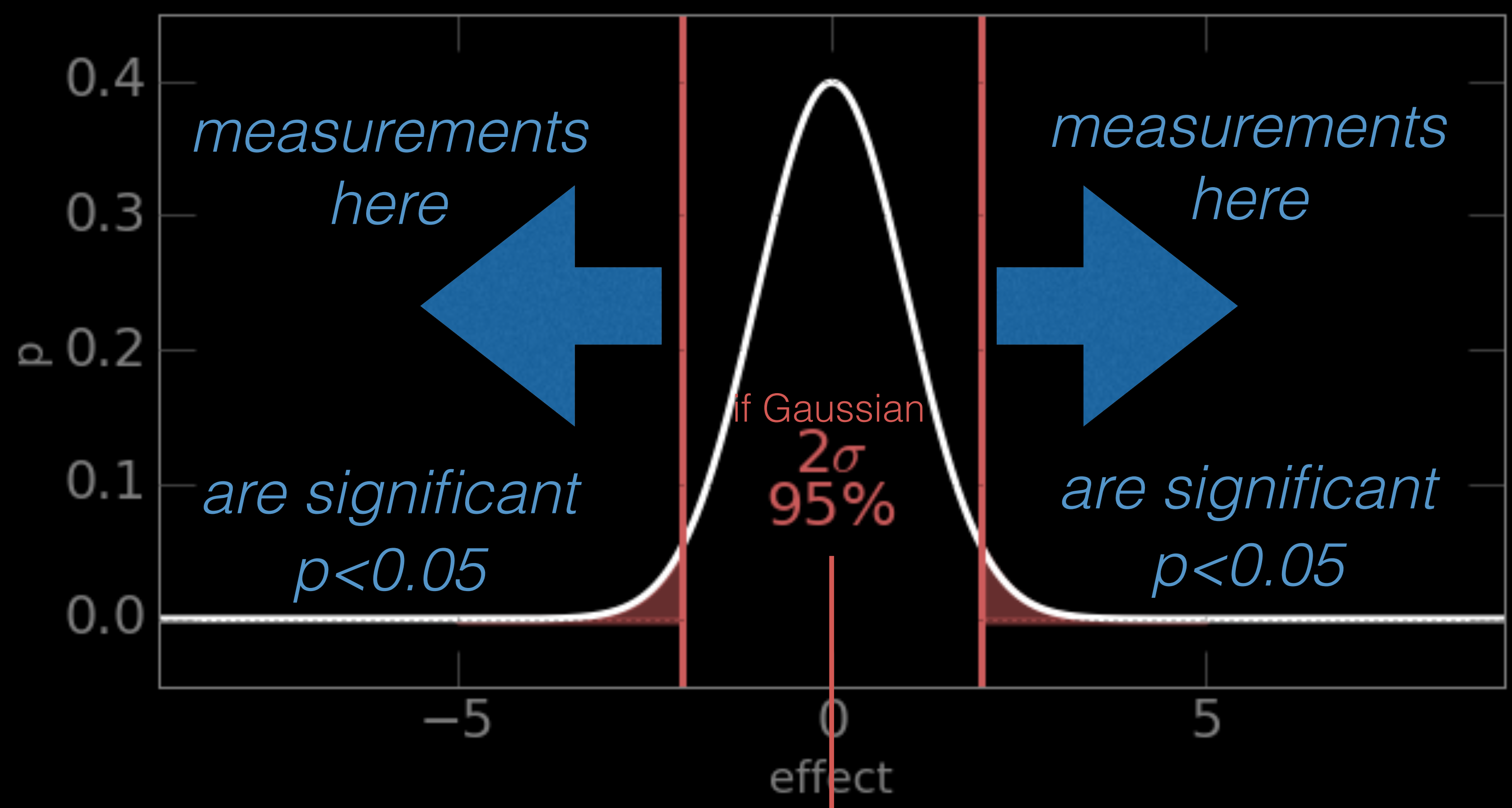
$\alpha = 0.05$

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



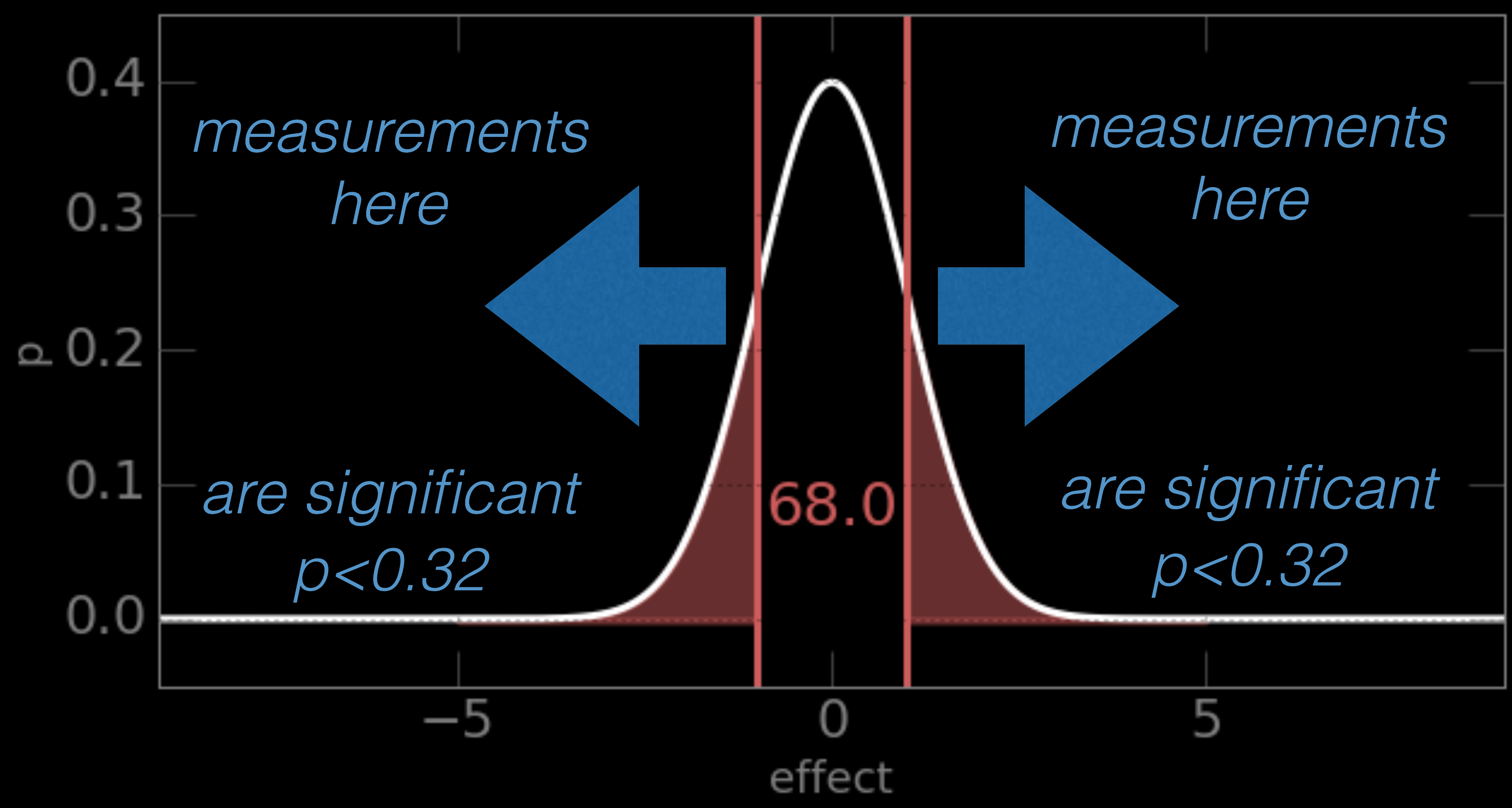
$\alpha = 0.05$

$1-0.05 = 0.95 \quad => 95\%$



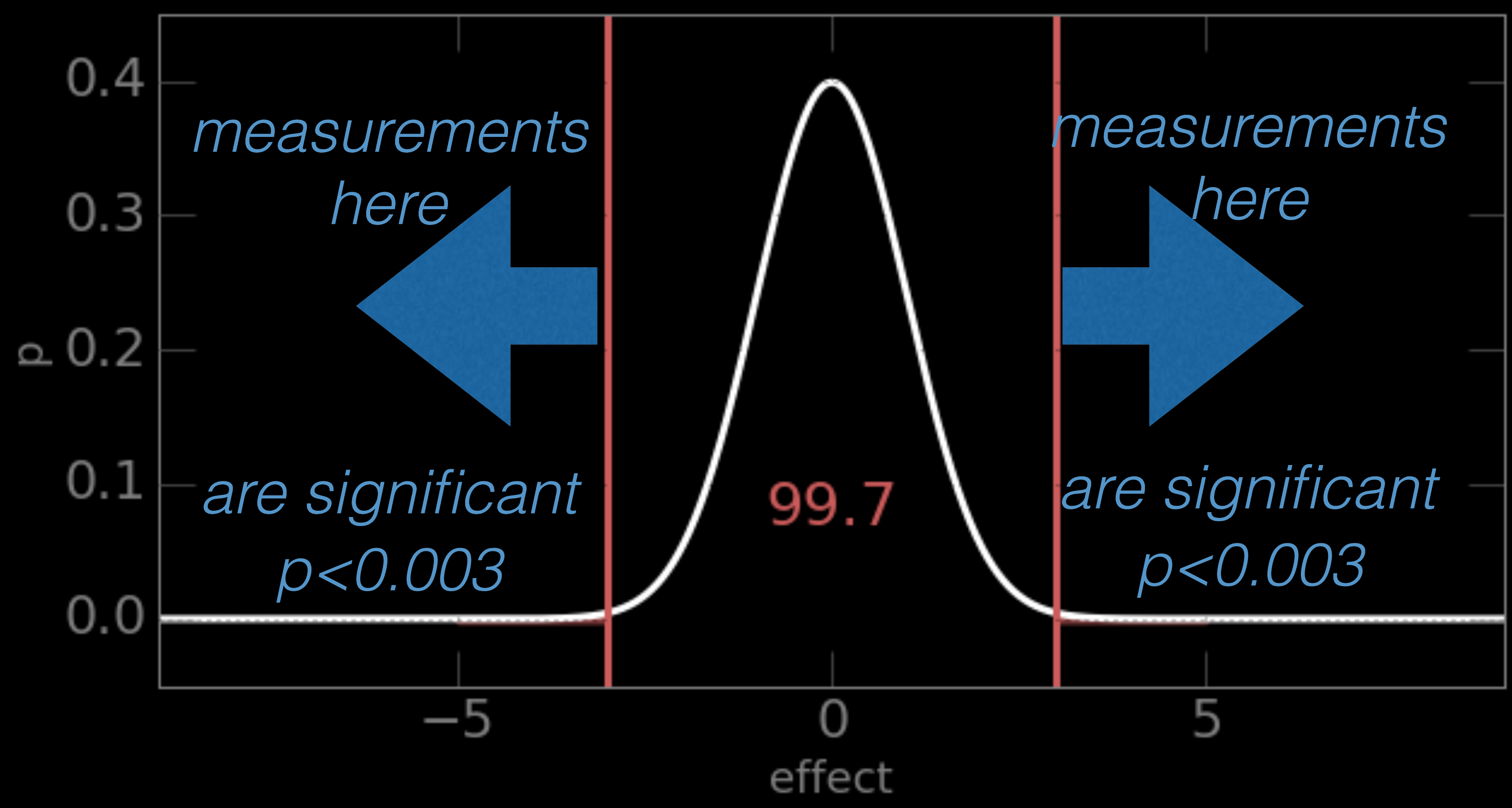
Z test

$\alpha = 0.32$



0.4
0.3
0.2
0.1
0.0
p
measurements here
measurements here
are significant p<0.32
are significant p<0.32
68.0
−5
0
5
effect

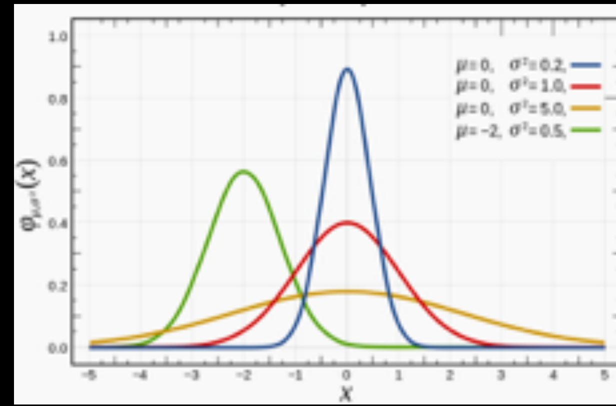
$\alpha = 0.003$



0.4
0.3
0.2
0.1
0.0
p
measurements here
are significant p<0.003
99.7
measurements here
are significant p<0.003
−5
0
5
effect

Z statistics Gaussian

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

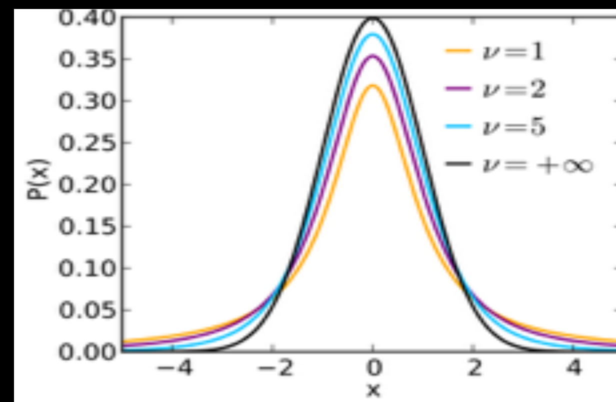


Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbf{R}$ — mean (location) $\sigma^2 > 0$ — variance (squared scale)
Support	$x \in \mathbf{R}$
PDF	$\frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$

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

Student's *t*

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

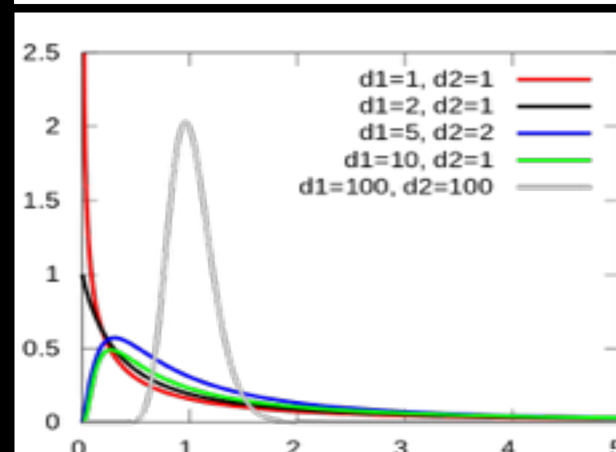


Parameters	$\nu > 0$ degrees of freedom (real)
Support	$x \in (-\infty; +\infty)$
PDF	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$
CDF	$\frac{1}{2} + x\Gamma\left(\frac{\nu+1}{2}\right) \times \frac{{}_2F_1\left(\frac{1}{2},\frac{\nu+1}{2};\frac{3}{2};-\frac{x^2}{\nu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

F statistics

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

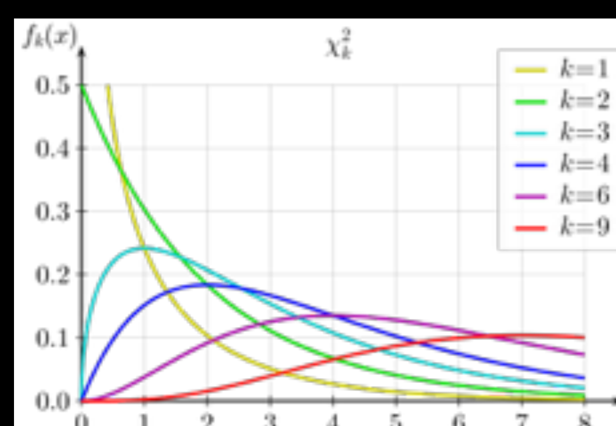


Parameters	$d_1, d_2 > 0$ deg. of freedom
Support	$x \in [0, +\infty)$
PDF	$\frac{\sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1+d_2}}}}{x\,\mathrm{B}\left(\frac{d_1}{2},\frac{d_2}{2}\right)}$
CDF	$I_{\frac{d_1 x}{d_1 x + d_2}}\left(\frac{d_1}{2},\frac{d_2}{2}\right)$

Mean	$\frac{d_2}{d_2-2}$ for $d_2 > 2$
Mode	$\frac{d_1-2}{d_1}\,\frac{d_2}{d_2+2}$ 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

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



Notation	$\chi^2(k)$ or χ^2_k
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	$\sqrt{8/k}$

CUSP

goodness of fit χ^2 $\chi^2_F = \sum_i \frac{(m_i - x_i)^2}{e_i}$ -

see
Statistics in a Nutshell
IV: Statistical analysis

Steps in Hypothesis Testing

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

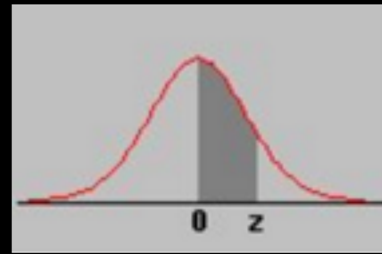
2. Measure a statistcs for *a sample* to be compared to the *parameter of a population*

OR

Measure a statistcs 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

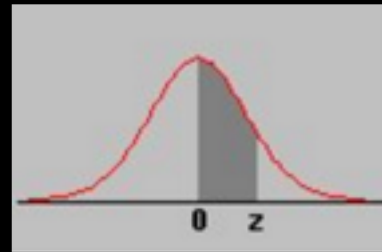


$$Z = \frac{\mu_{\text{pop}} - \mu_{\text{sample}}}{\sigma / \sqrt{N}} = 2.55$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
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	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

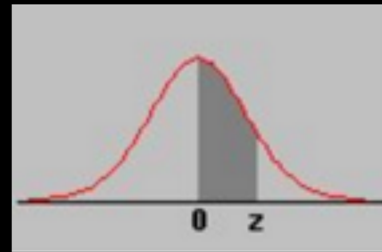
https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md



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

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

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



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

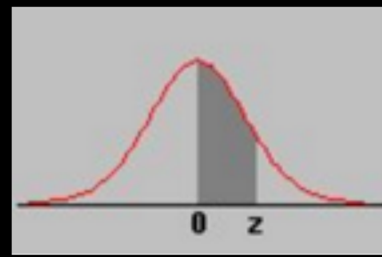
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293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

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

2 sided test $1 - (0.4946 * 2) = 0.0108$

$p < 0.05$

H_0 IS REJECTED ($p<0.05$)



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

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319

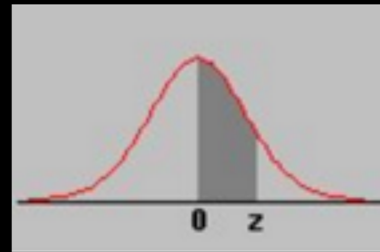
	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
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	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

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

1 sided test   1 - 0.4946 - 0.5 = 0.0054

$p < 0.05$

H_0 IS REJECTED ($p<0.05$)



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

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	**0.4418**	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
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	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	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

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

$\alpha = 0.05$

4 observations - 1 independent variable = 3 degreed of freedom

Accept Hypothesis

Reject Hypothesis

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^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

$p < 0.05$

H_0 REJECTED

$$\chi_F^2 = \sum_{i=1}^{4} \frac{(m_i - x_i)^2}{e_i} = 7.11$$

$\alpha = 0.05$

4 observations - 1 independent variable =
3 degreed of freedom

Accept Hypothesis ← → Reject Hypothesis

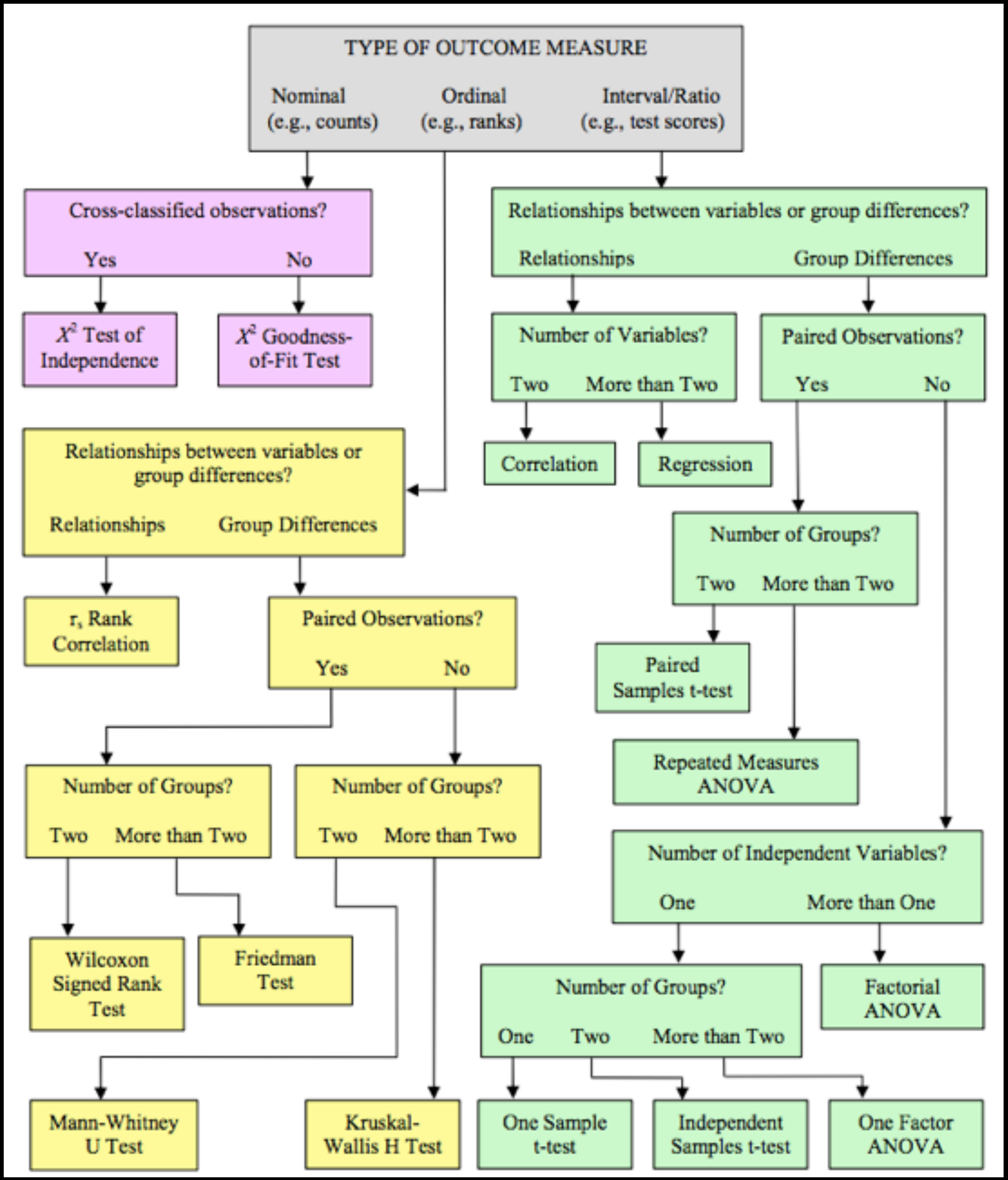
Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^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

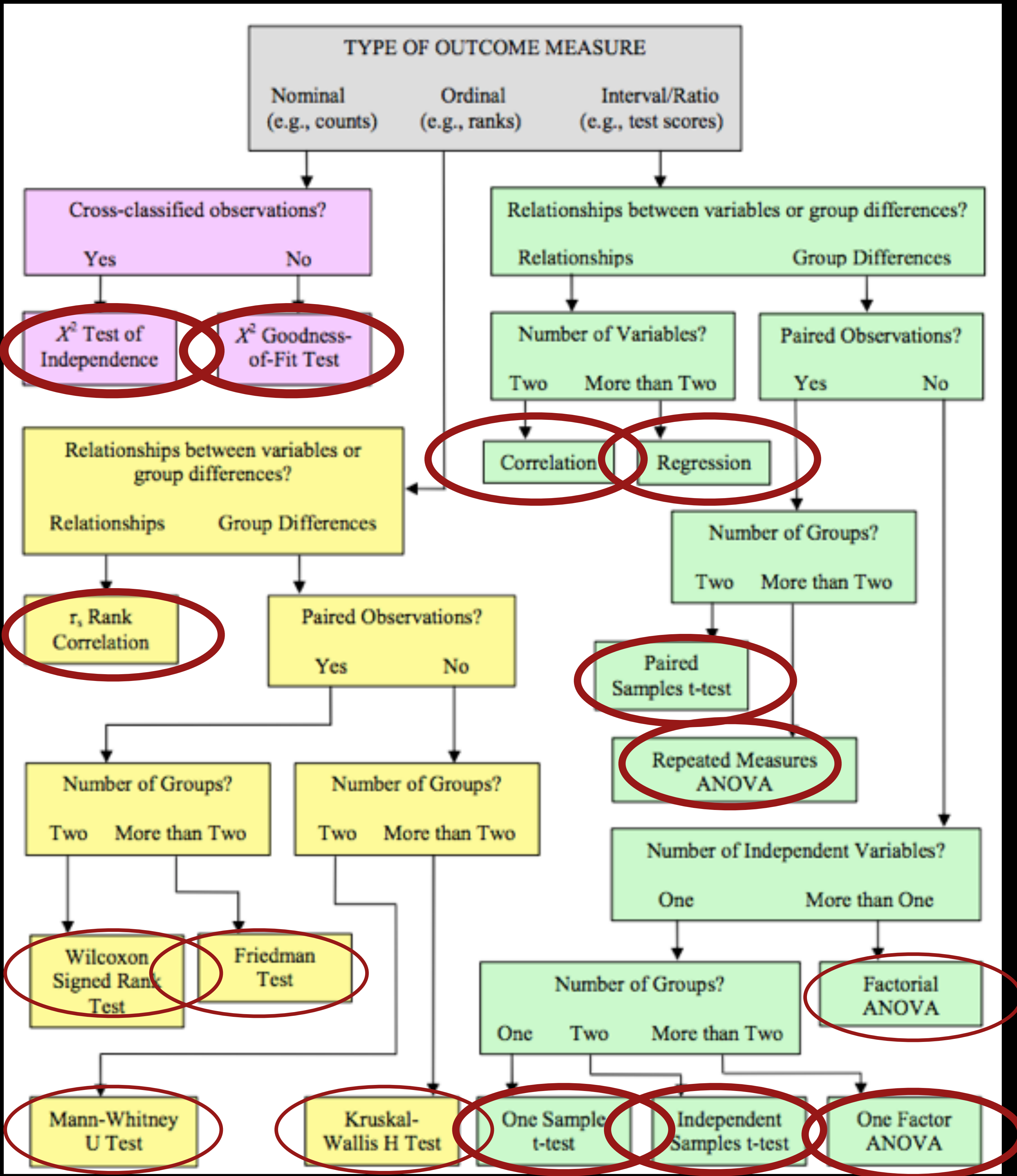
https://github.com/fedhere/UInotebooks/blob/master/HowToReadZandChisqTables.md

$p > 0.05$

H_0 CANNOT BE REJECTED



IV: Statistical analysis



IV: Statistical analysis

When to Use a Particular Statistical Test

Anne Marenco, CalState 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: When to Use a Particular Statistical Test

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

Anne Marenco, CalState http://www.csun.edu/~amarenco/Fcs%20682/When%20to%20use%20what%20test.pdf

choose 1

choose 1

Statistical Analyses	Independent Variables: # of IVs	Independent Variables: Data Type	Dependent Variables: # of DVs	Dependent Variables: Type of Data	Control Variables	Question Answered by the Statistic
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 2: When to Use a Particular Statistical Test

Example: ANCOVA

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

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

Comparing to population

Standard DEVIATION of Sample Estimates

Sample mean, $\bar{x}$	$Z = \dfrac{\mu_{pop} - \mu_{sample}}{\sigma / \sqrt{N}}$	$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$
Sample proportion, p	$z = \dfrac{p - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$	$\sigma_p = \sqrt{\dfrac{P(1-P)}{n}}$
Difference between proportions, p1 - p2	$z = \dfrac{(p_2 - p_1)}{\sqrt{p(1-p)(\frac{1}{n_2} + \frac{1}{n_1})}}, p = \dfrac{p_2 n_2 + p_1 n_1}{n_2 + n_1}$	$\sigma_{p_1 - p_2} = \sqrt{\dfrac{P_1(1-P_1)}{n_1} + \dfrac{P_2(1-P_2)}{n_2}}$

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	$SE_p = \frac{p(1-p)}{n}$
Difference between means, x1 - x2	$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
Difference between proportions, p1 - p2	$SE_{p_1 - p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

use if you DO NOT know the parameters of the population:

e.g. in the t-test

Beyond Significance Based Hypothesis Rejection

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" (PDF), Educational and Psychological Measurement, 20 (1): 37–46,

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 *d*	$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$ $\quad s = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$	for ***Z*** and ***t***	.20 .50 .80
Cohen's *w*	$w = \sqrt{\sum_{i=1}^{N} \frac{(p_{0i}-p_{1i})^2}{p_{0i}}}$	for χ^2	.10 .30 .50
Cohen's *h*	$h = 2(\arcsin\sqrt{p_1} - \arcsin\sqrt{p_2})$	for proportions	.20 .50 .80
f-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

http://core.ecu.edu/psyc/wuenschk/docs30/EffectSizeConventions.pdf

ERRORS

What if our conclusion is after all wrong??

we have not ruled out the null, just decided that is it unlikely

ERRORS

reality → *result of analysis* ↓	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

Errors and uncertainties.

Errors and uncertainties.

- Systematic error
- Statistical error (Stochastic & Random)

systematic errors

Errors and uncertainties.

- Systematic error

  <u>tendency to *systematically* underestimate/overestimate the average</u> difference between the *population* and the subset you test or *sample* because the sample is intrinsically different or the measurements are consistently off

Physical Activity and Transit Survey

Physical Activity and Transit Survey: Methodology

Target Population | Health Topics | Sampling | Limitations | Sample Size, Response and Cooperation Rates | Data Analysis

The Physical Activity and Transit (PAT) survey was conducted in 2010 and 2011 by the New York City Department of Health and Mental Hygiene. Data were collected to measure the level and context of physical activity in New York City and to improve understanding of what motivates individuals to be physical active including opportunities for activity, perceptions of safety and security, and other neighborhood factors. For more information on the PAT, please visit:

- PAT Overview
- PAT Device Methodology
- PAT Public Use Datasets

TARGET POPULATION
The target population of the PAT was adults aged 18 and older who were able to walk more than 10 feet and who lived in one of the five boroughs of New York City. Of the 3908 adults who completed the initial telephone screener, 3811 were mobile and completed the full survey.

HEALTH TOPICS
The PAT asked approximately 125 questions, covering the following: a modified version the Global Health and Physical Activity Questionnaire (GPAQ) designed by the World Health Organization on physical activity in the work, recreation and transportation domains. Also included were questions on chronic disease, diet, alcohol and tobacco, neighborhood conditions, and mental health. The survey also asked a multiple demographic variables to facilitate weighting and comparisons among different groups of New Yorkers.

SAMPLING
The PAT was conducted using a fully overlapping dual frame design, using randomly generated landline and cellular telephone samples. (Roughly 25% were completed on a cell phone.) To provide equal statistical power for borough-level comparisons, a similar number of participants were interviewed in each borough of New York (Bronx, Brooklyn, Manhattan, Queens and Staten Island). All data were then weighted to adjust for the probability of selection and differential nonresponse and sum to Census estimates of the number of people living in each borough.

Interviewing was done by Abt-SRBI, a survey research company based in New York City. Interviews were conducted in English, Spanish, Russian and Chinese (Cantonese and Mandarin). Data collection for wave 1 occurred in September – November of 2010; wave 2 was conducted in March – November of 2011. The average length of the survey was 35 minutes.

back to top

LIMITATIONS
The sampling methodology did not capture adults who could not be reached by either landline or cellular telephone. The PAT also excluded adults living in institutional group housing, such as incarcerated persons or those living in college dormitories.

back to top

http://www.nyc.gov/html/doh/html/data/pat-methods.shtml#4

Errors and uncertainties.

Errors and uncertainties.

Errors and uncertainties.

projection induces a *systematic underestimation*

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

UNBIASED survey:
average from *all* samples
equals *population* average

different surveys/experiments give different results
because they have different systematic errors or bias

CUSP

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

UNBIASED survey:
average from *all* samples
equals *population* average

Undercoverage bias

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

UNBIASED survey:
average from *all* samples
equals *population* average

Undercoverage bias

the surveyed segment of the population is lower in a sample than it is in the population. This can happen because the frame used to obtain the sample is incomplete or not representative of the population.

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

  UNBIASED survey:
  average from *all* samples
  equals *population* average

  Self selection bias

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

UNBIASED survey:
average from *all* samples
equals *population* average

<u>Self selection bias</u>

higher test scores observed among students who participate in a test preparation courses, but due to self-selection, people who choose to take the course may be more motivated, have more support…

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

UNBIASED survey:
average from *all* samples
equals *population* average

Self selection bias

higher test scores observed among students who participate in a test preparation courses, but due to self-selection, people who choose to take the course may be more motivated, have more support…
people willing to answer a survey about climate are likely more concerned and responsible citizens

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

UNBIASED survey:
average from *all* samples
equals *population* average

Social desirability bias

tendency of survey respondents to answer questions in a manner that will be viewed favorably: over-reporting "good behavior" or under-reporting undesirable behavior (e.g. drug+alcohol use).

Bias in measurements: know your data

Organizational Research Methods
00(0) 1-34

DOI: 10.1177/1094428113493120
orm.sagepub.com
SAGE

Random and Systematic Error Effects of Insomnia on Survey Behavior

Larissa K. Barber[1], Christopher M. Barnes[2], and Kevin D. Carlson[2]

Abstract

Insomnia is a prevalent experience among employees and survey respondents. Drawing from research on sleep and self-regulation, we examine both random (survey errors) and systematic (social desirability) effects of research participant insomnia on survey responses. With respect to random effects, we find that insomnia leads to increased survey errors, and that this effect is mediated by a lack of self-control and a lack of effort. However, insomnia also has a positive systematic effect, leading to lower levels of social desirability. This effect is also mediated by self-control depletion and a lack of

http://www.researchgate.net/publication/244478619_Random_and_Systematic_Error_Effects_of_Insomnia_on_Survey_Behavior

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

UNBIASED survey:
average from *all* samples
equals *population* average

Undercoverage bias
Self selection bias
Social desirability bias

Small number statistic
Publication Bias
Data Dredging

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | F

News & Comment > News > 2015 > August > Article

NATURE | NEWS

Social sciences suffer from severe publication bias

Survey finds that 'null results' rarely see the light of the day.

Mark Peplow

28 August 2014

Publication Bias

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

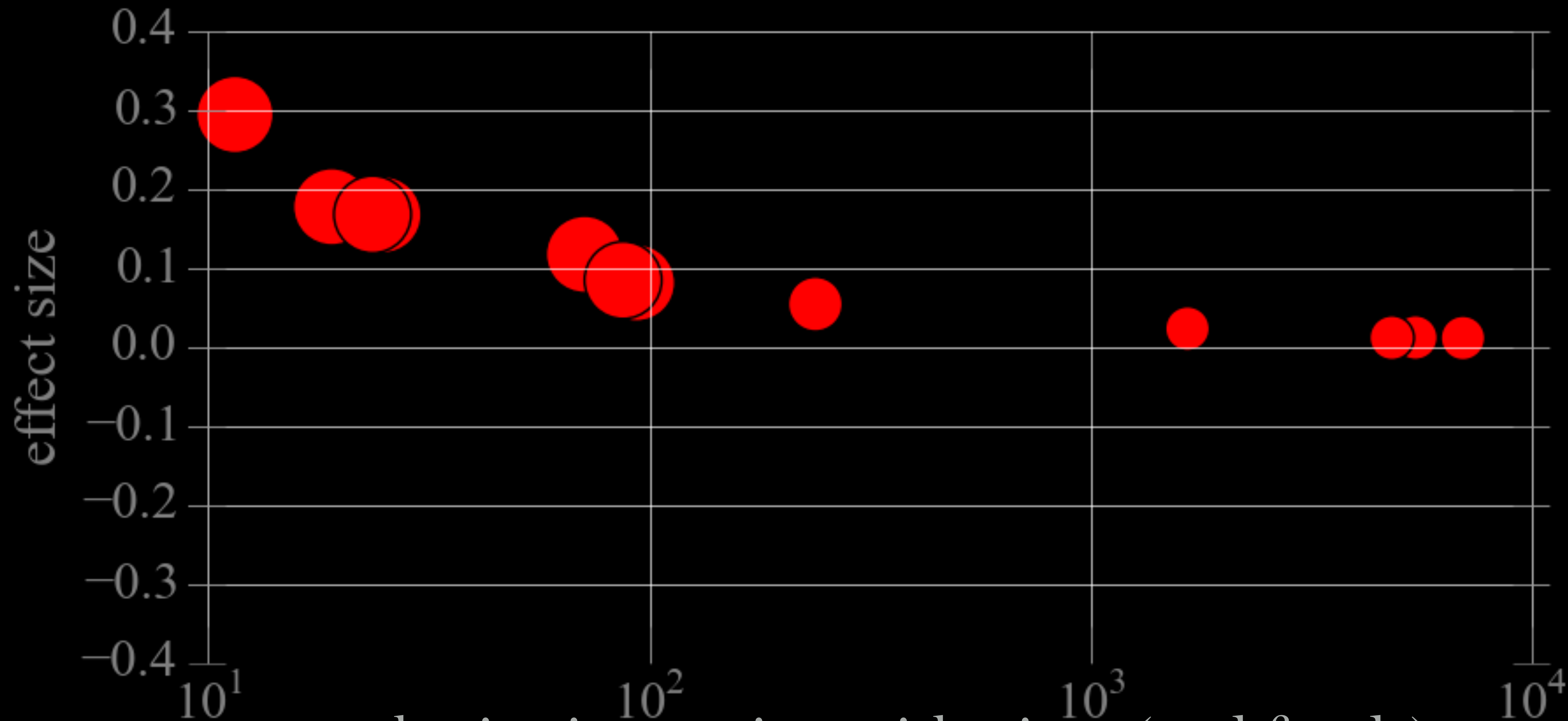
His team investigated the fate of 221 sociological studies conducted between 2002 and 2012, which were recorded by Time-sharing Experiments for the Social Sciences (TESS), a US project that helps social scientists to carry out large-scale surveys of people's views.

Only 48% of the completed studies had been published. So the team contacted the remaining authors to find out whether they had written up their results, or submitted them to a journal or conference. They also asked whether the results supported the researchers' original hypothesis.

Of all the null studies, just 20% had appeared in a journal, and 65% had not even been written up. By contrast, roughly 60% of studies with strong results had been published. Many of the researchers contacted by Malhotra's team said that they had not written up their null results because they thought that journals would not publish them, or that the findings were neither interesting nor important enough to warrant any further effort.

<u>Publication Bias</u>

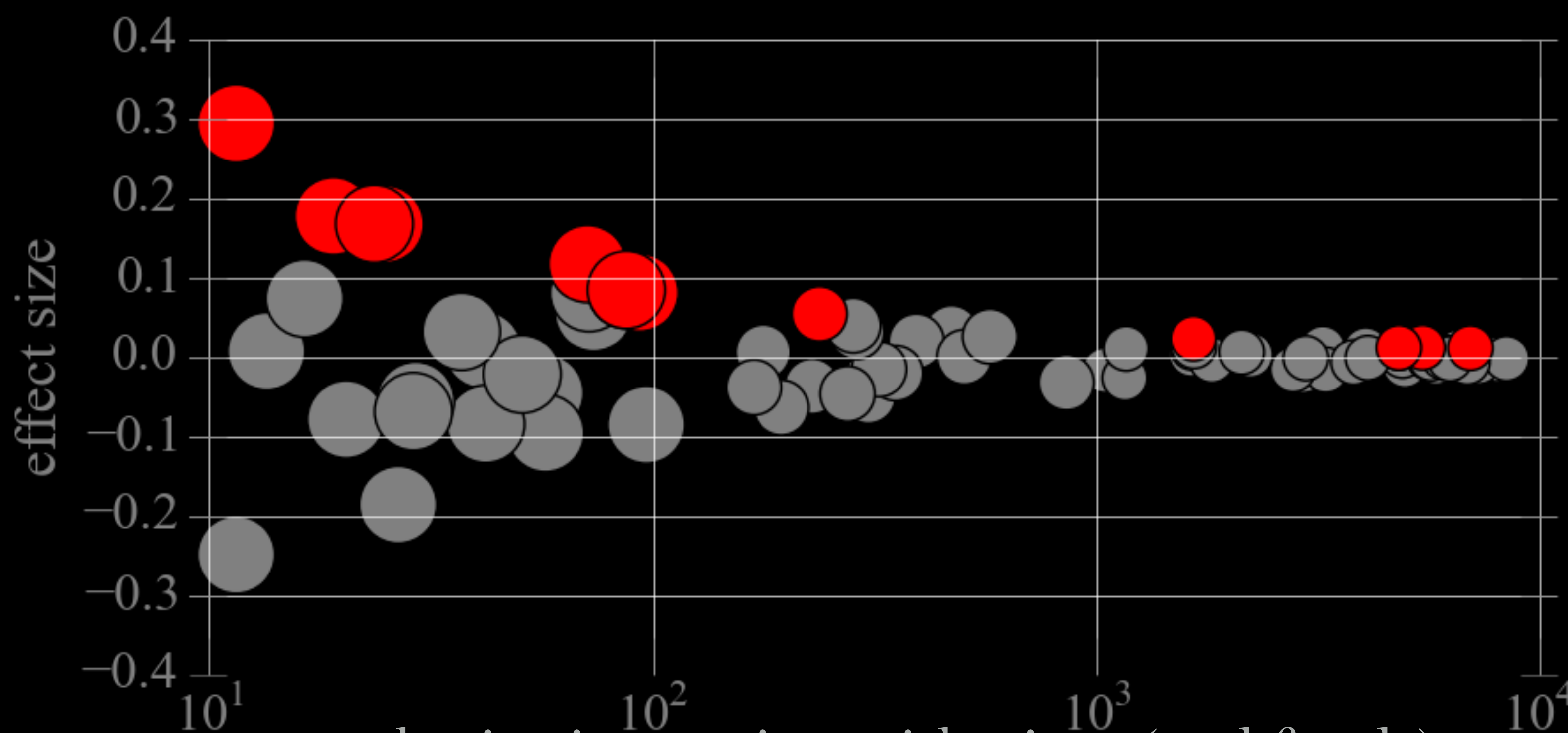
Bias in measurements: know your data



sample size increasing with time (and funds)

Publication Bias

Bias in measurements: know your data



sample size increasing with time (and funds)

Publication Bias

Bias in measurements: know your data

- Systematic error: SURVEY BIAS

EconPapers
Economics at your fingertips

EconPapers Home
About EconPapers

Working Papers
Journal Articles
Books and Chapters
Software Components

Authors

JEL codes
New Economics Papers

Advanced Search

Publication Bias in Measuring Climate Sensitivity

Share

Dominika Rečková and Zuzana Iršová (zuzana.irsova@ies-prague.org)

No 2015/14, Working Papers IES from Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies

Abstract: We present a meta-regression analysis of the relation between the concentration of carbon dioxide in the atmosphere and changes in global temperature. The relation is captured by "climate sensitivity", which measures the response to a doubling of carbon dioxide concentrations compared to pre-industrial levels. Estimates of climate sensitivity play a crucial role in evaluating the impacts of climate change and constitute one of the most important inputs into the computation of the social cost of carbon, which reflects the socially optimal value of a carbon tax. Climate sensitivity has been estimated by many researchers, but their results vary significantly. We collect 48 estimates from 16 studies and analyze the literature quantitatively. We find evidence for publication selection bias: researchers tend to report preferentially large estimates of climate sensitivity. Corrected for publication bias, the bulk of the literature is consistent with climate sensitivity lying between 1.4 and 2.3C.

Publication Bias

http://ies.fsv.cuni.cz/default/file/download/id/28421

Bias in measurements: know your data

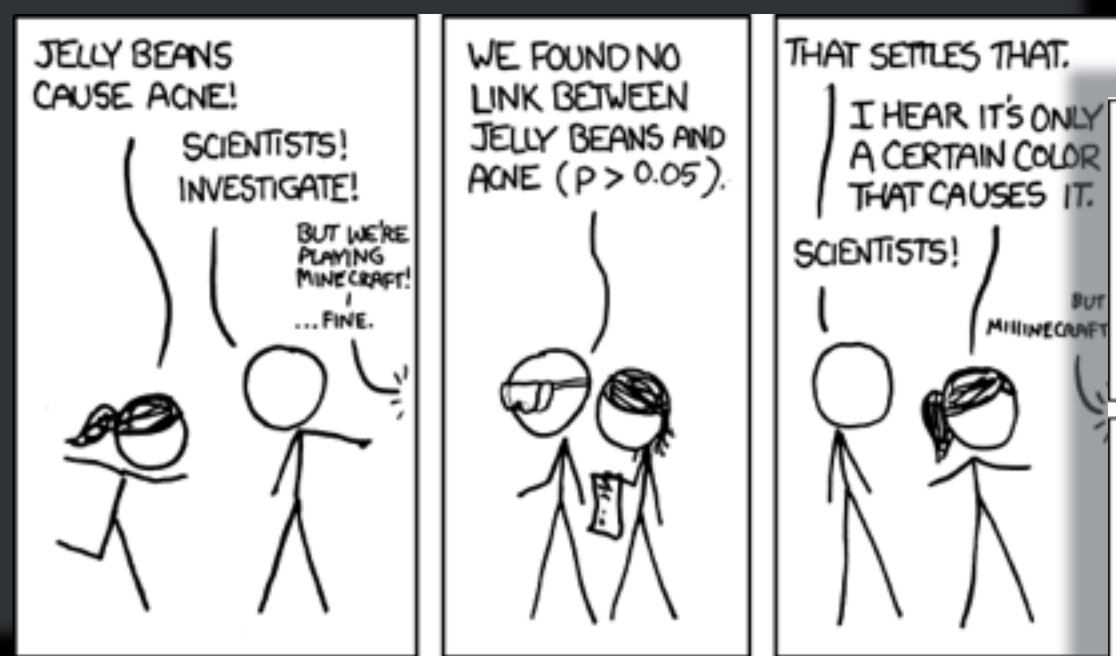
- Systematic error: SURVEY BIAS

UNBIASED survey:
average from *all* samples
equals *population* average

Undercoverage bias
Self selection bias
Social desirability bias

Small number statistic
Publication Bias
Data Dredging

Bias in measurements: know your data



https://xkcd.com/882/

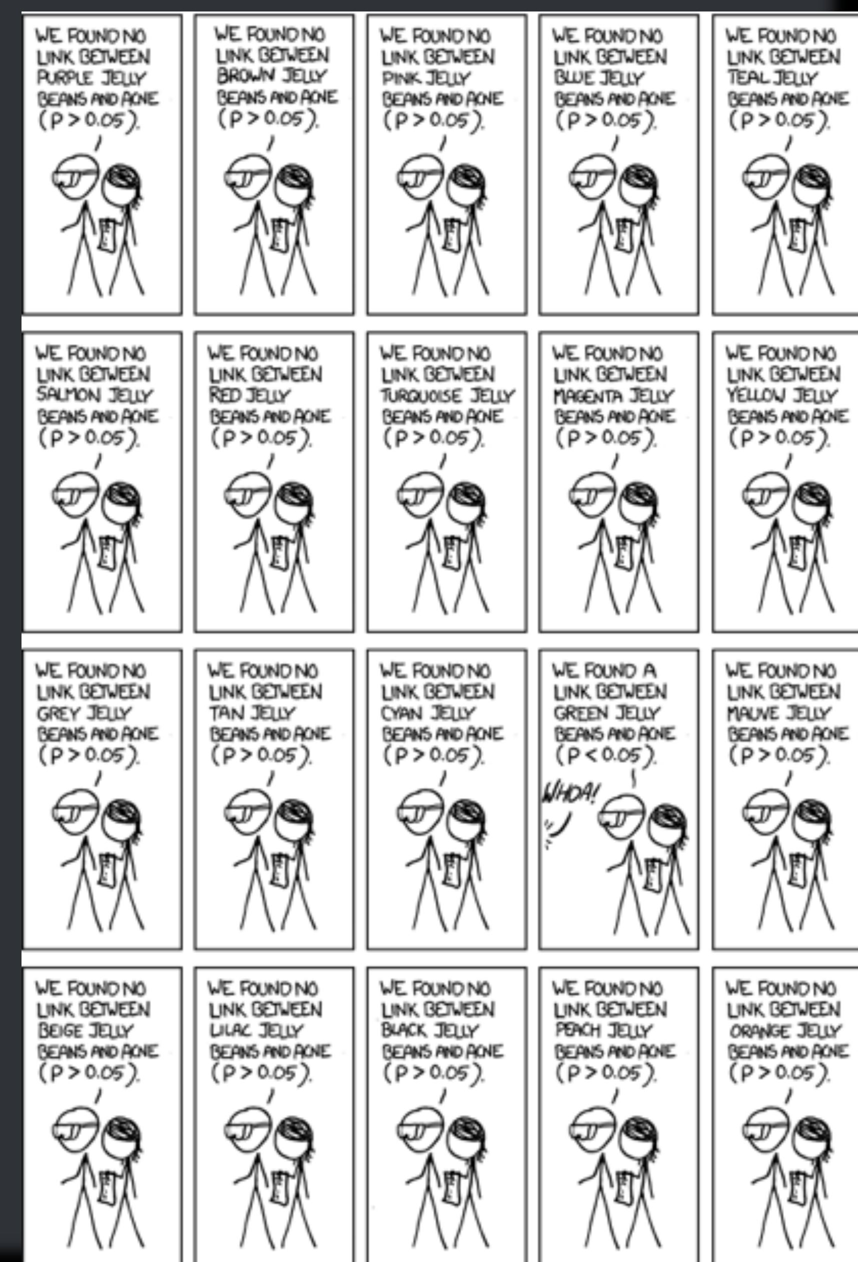
Data Dredging

IV: Statistical analysis

Bias in measurements: know your data

each test has a
probability p<=0.05
of Type I error
significance 95%

20 tests are preformed



Data Dredging

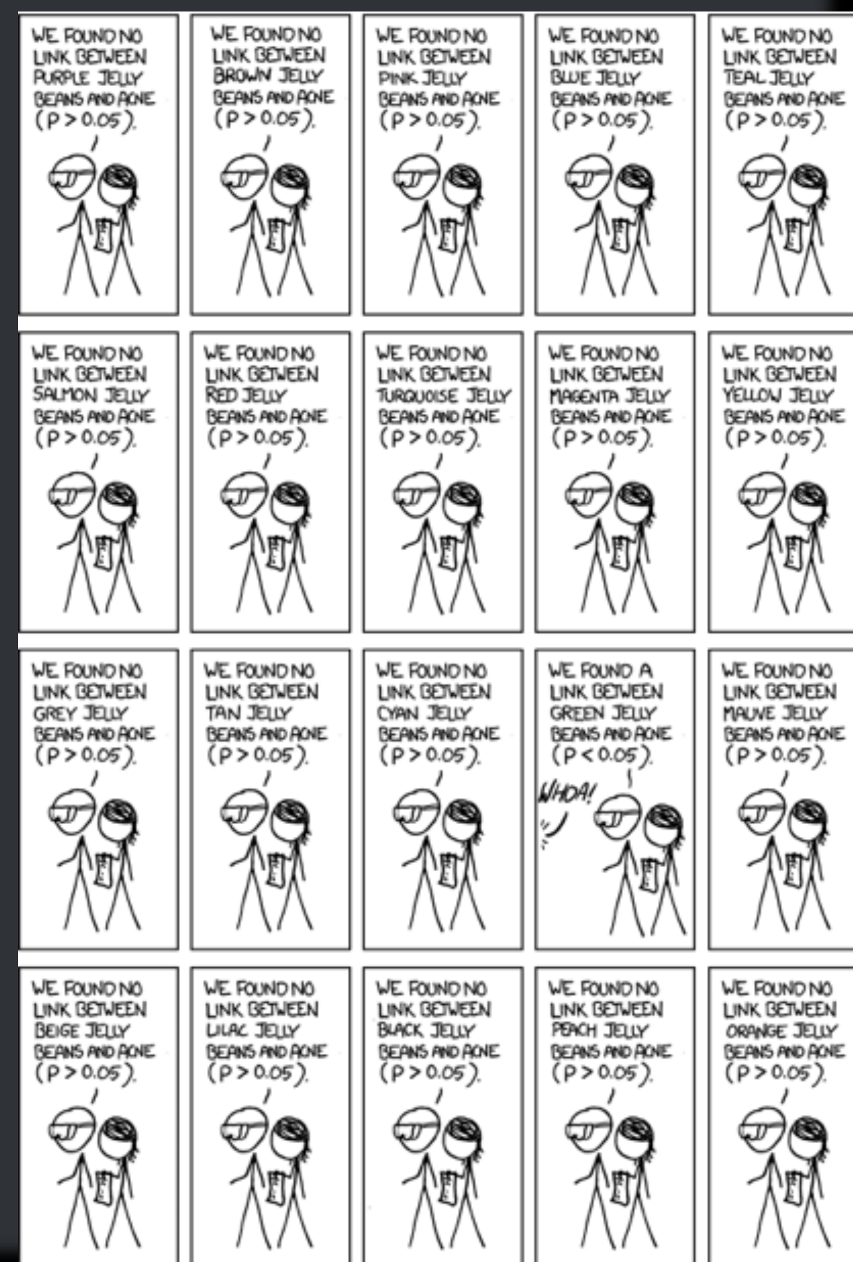
Bias in measurements: know your data

each test has a
probability $p<=0.05$
of Type I error
significance 95%

20 tests are preformed

assume independence:
if $p_i = 0.05$ for each i=1..20

total significance=
$1-(1-0.05)^{20}$
$p_{tot} = (1-0.05)^{20}$



Data Dredging

Bias in measurements: know your data

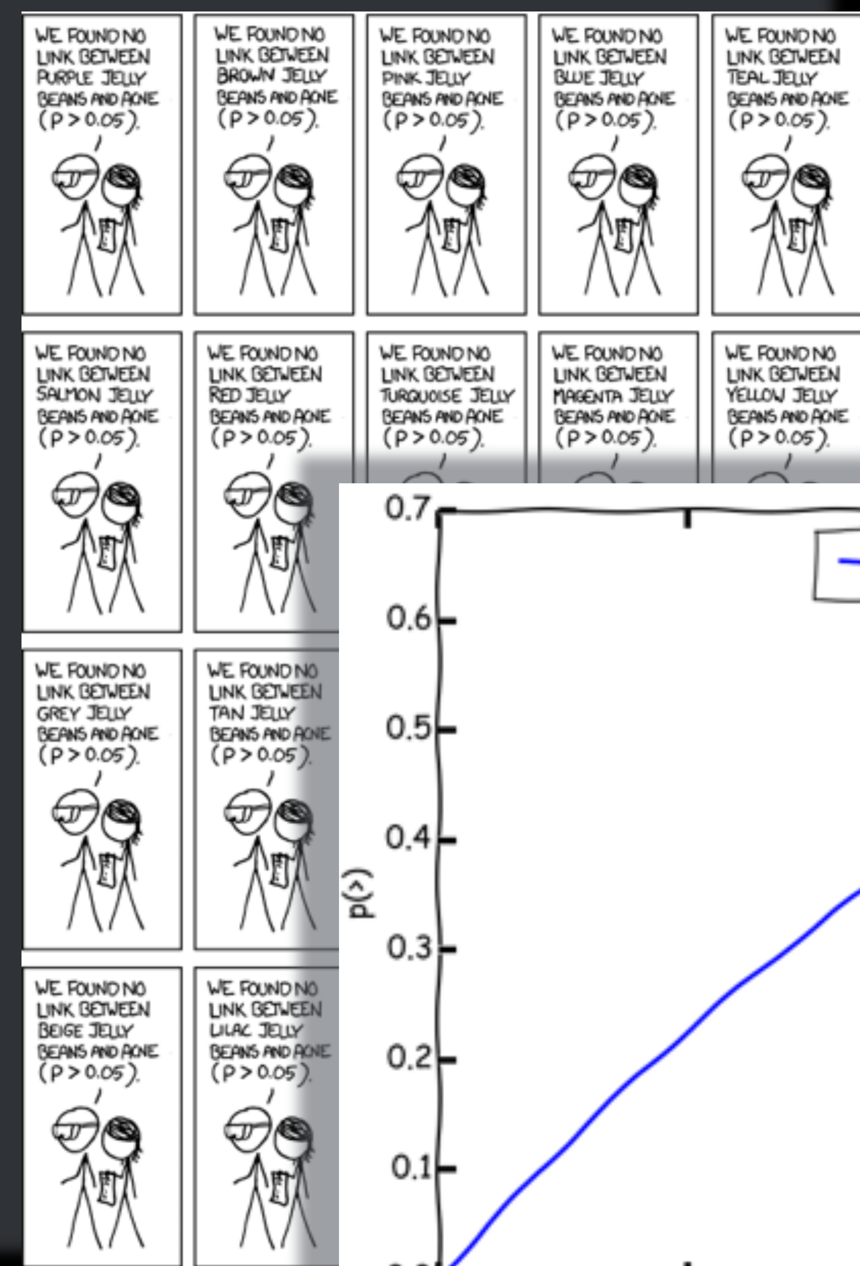
each test has a probability $p<=0.05$ of Type I error significance 95%

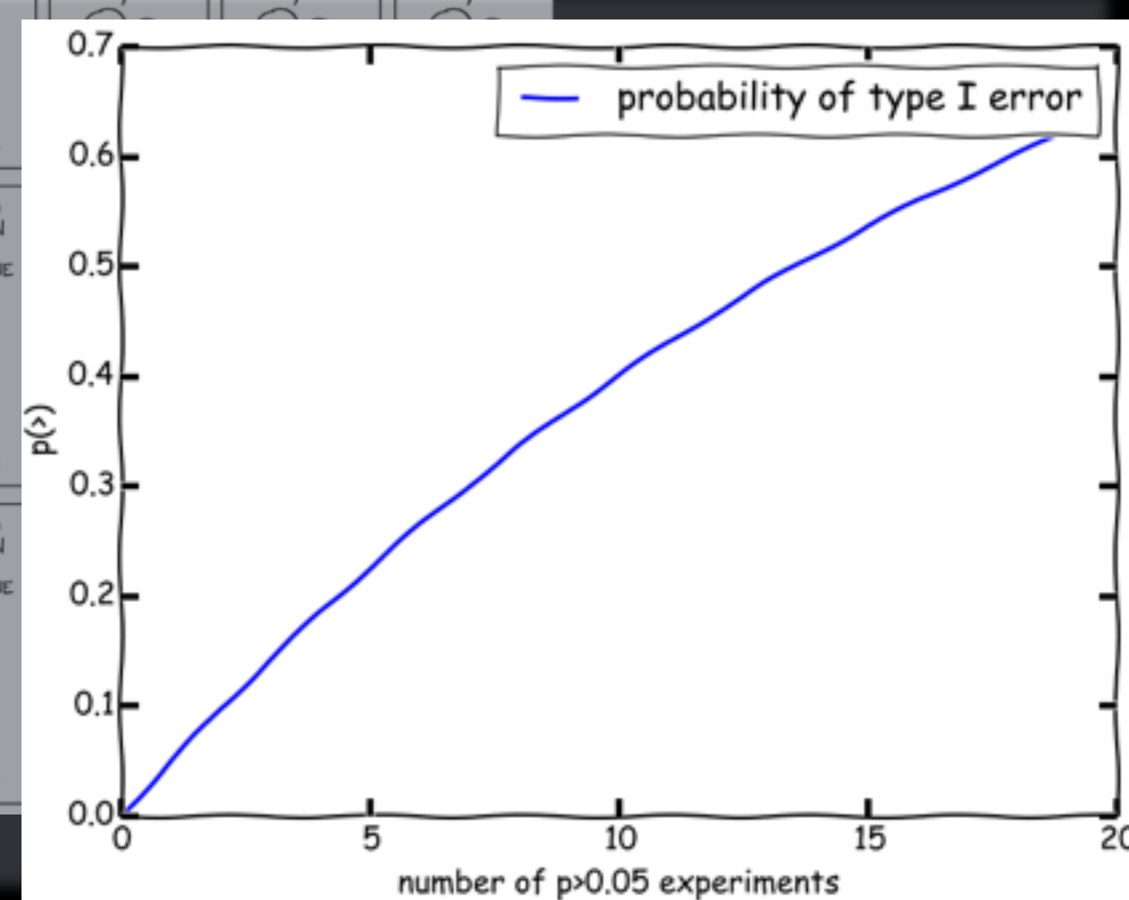
20 tests are preformed

assume independence:
if $p_i = 0.05$ for each i=1..20

total significance= $1-(1-0.05)^{20}$

$p_{tot} = (1-0.05)^{20}$





Data Dredging

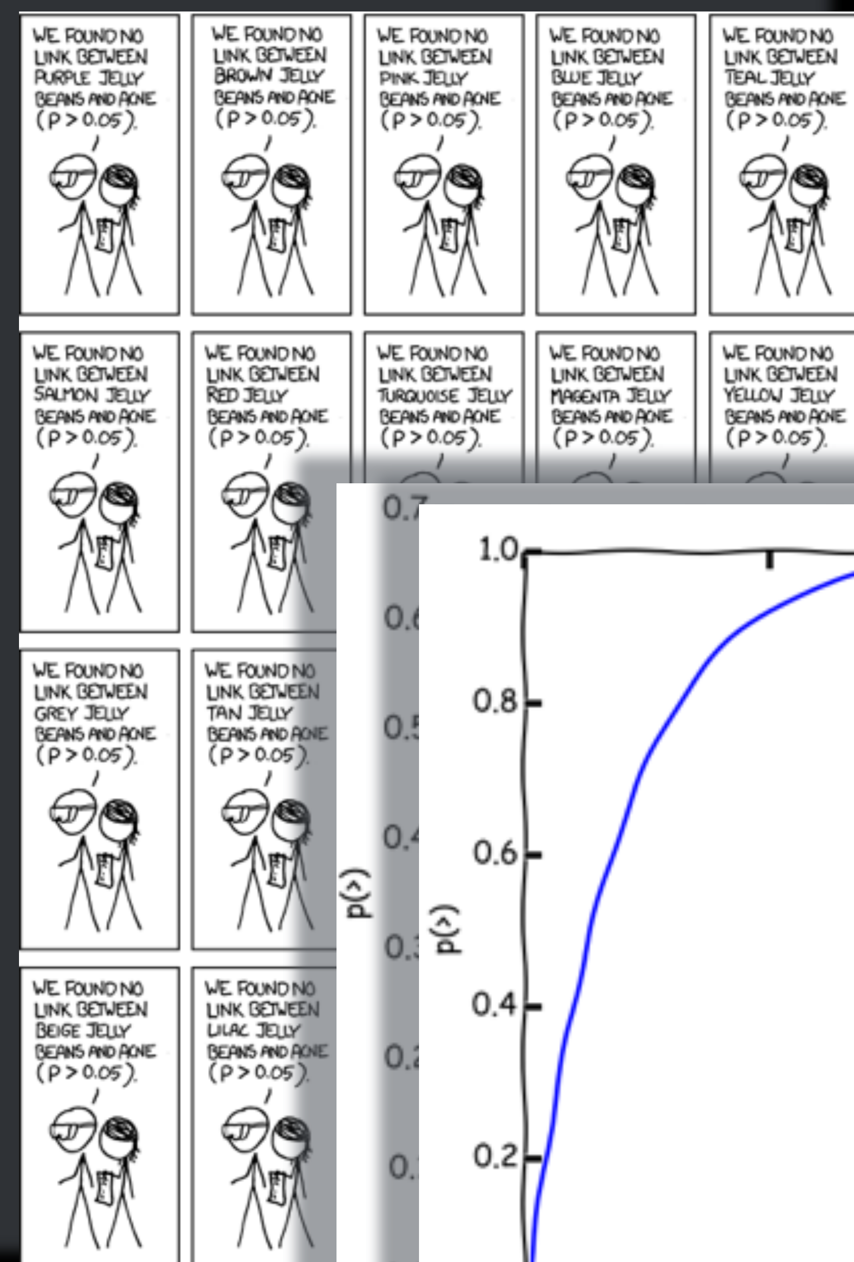
Bias in measurements: know your data

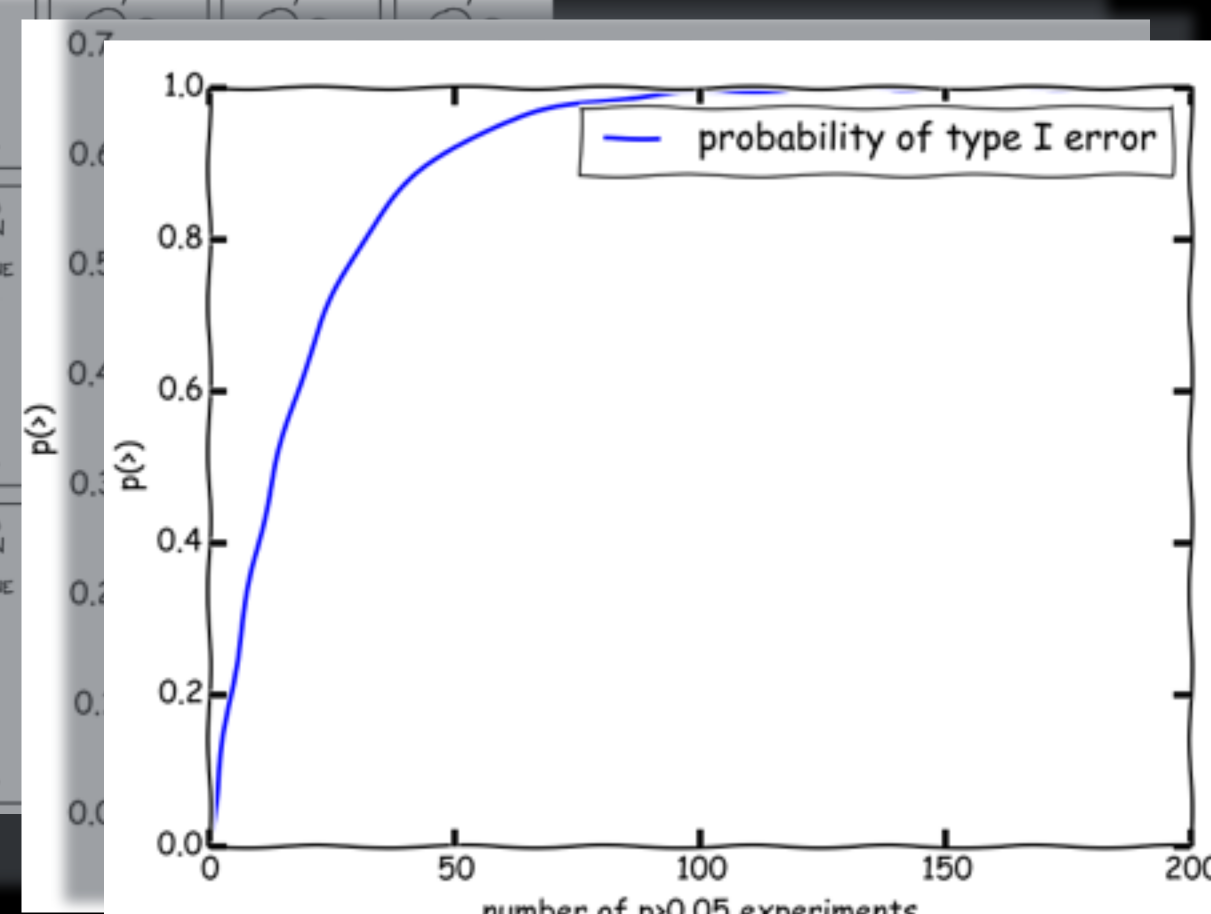
each test has a
probability $p<=0.05$
of Type I error
significance 95%

20 tests are preformed

assume independence:
if $p_i = 0.05$ for each i=1..20

total significance=
$1-(1-0.05)^{20}$
$p_{tot} = (1-0.05)^{20}$





Data Dredging

Systematic errors, biases, and reproducibility

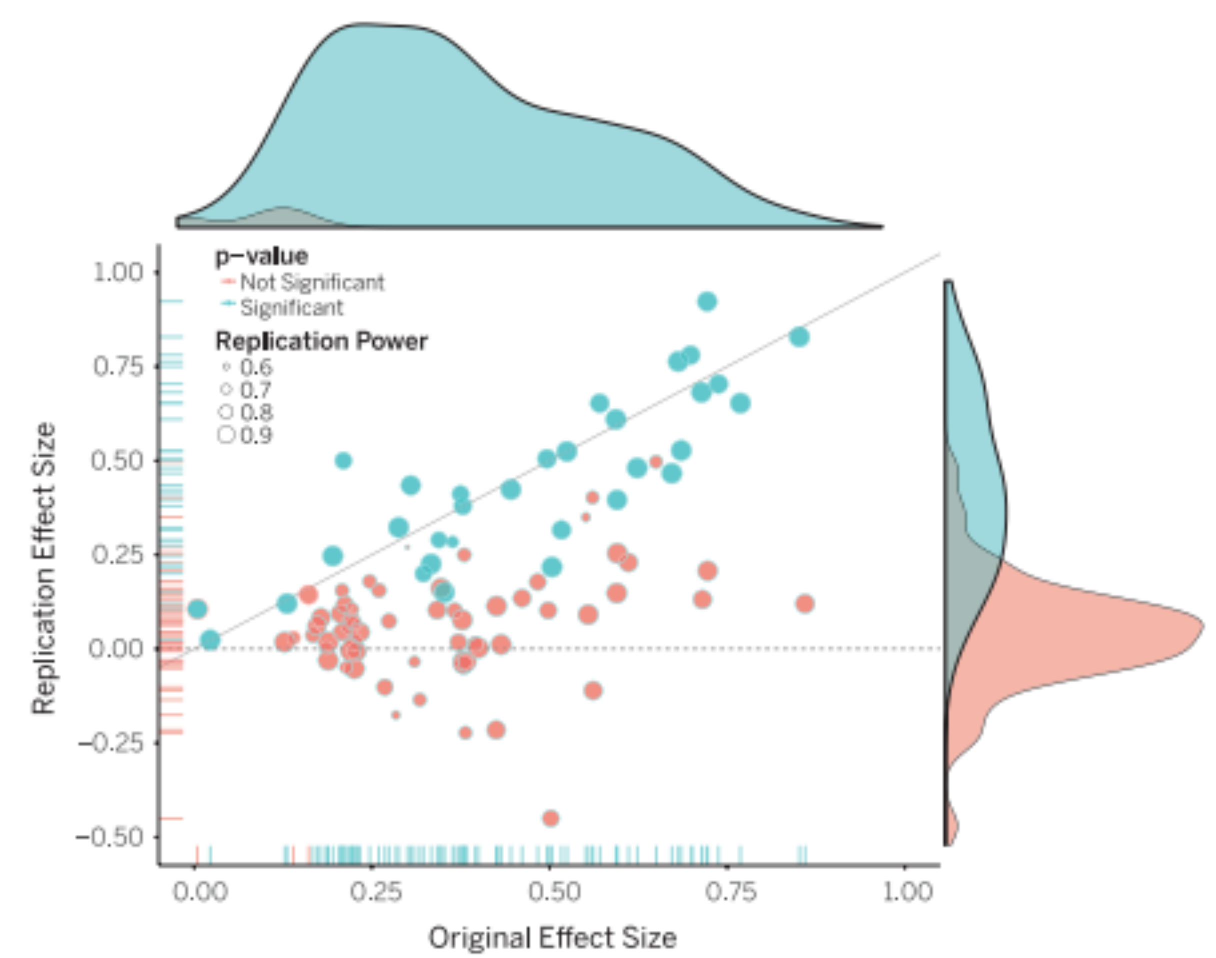
RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*

Science,
August 28, 2015



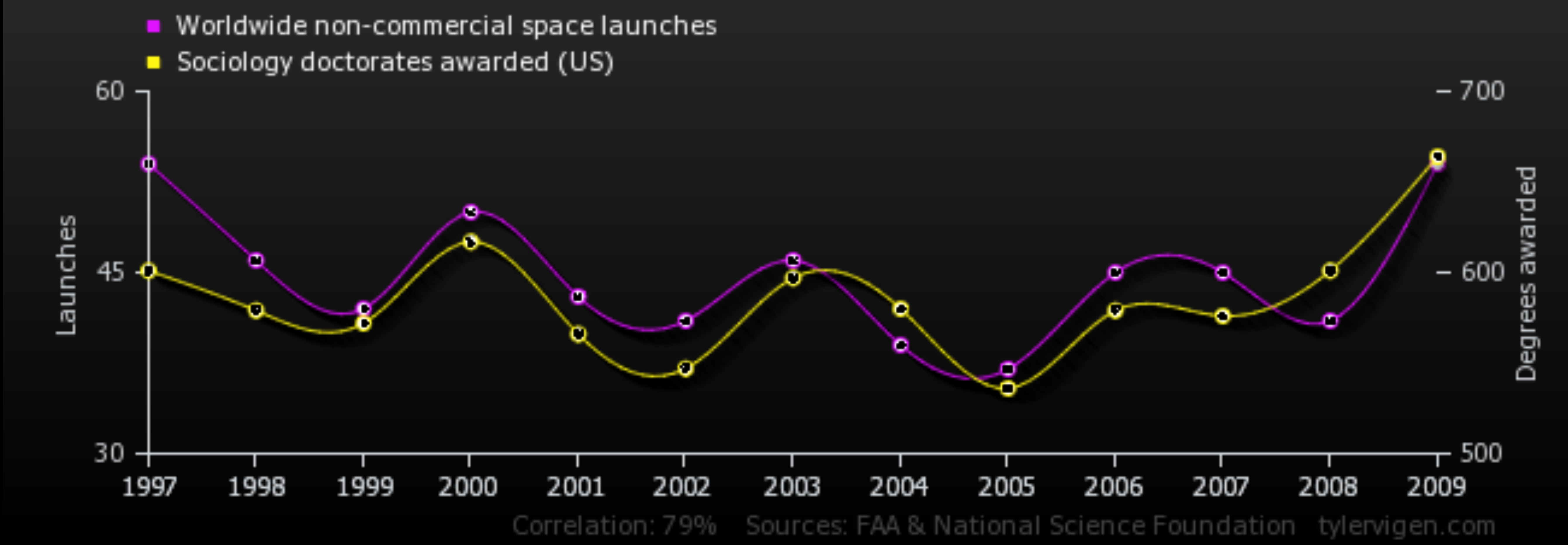
Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.

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

Errors and uncertainties.

- Systematic error
  tendency to *systematically* underestimate/overestimate the average difference between the *population* and the subset you test or *sample* because the sample is intrinsically different or the measurements are consistently off

  Solution: Good experimental design
  Calibration (to assess systematics induced by your measurements)
  Simulations (to assess the systematics induced by your analysis)

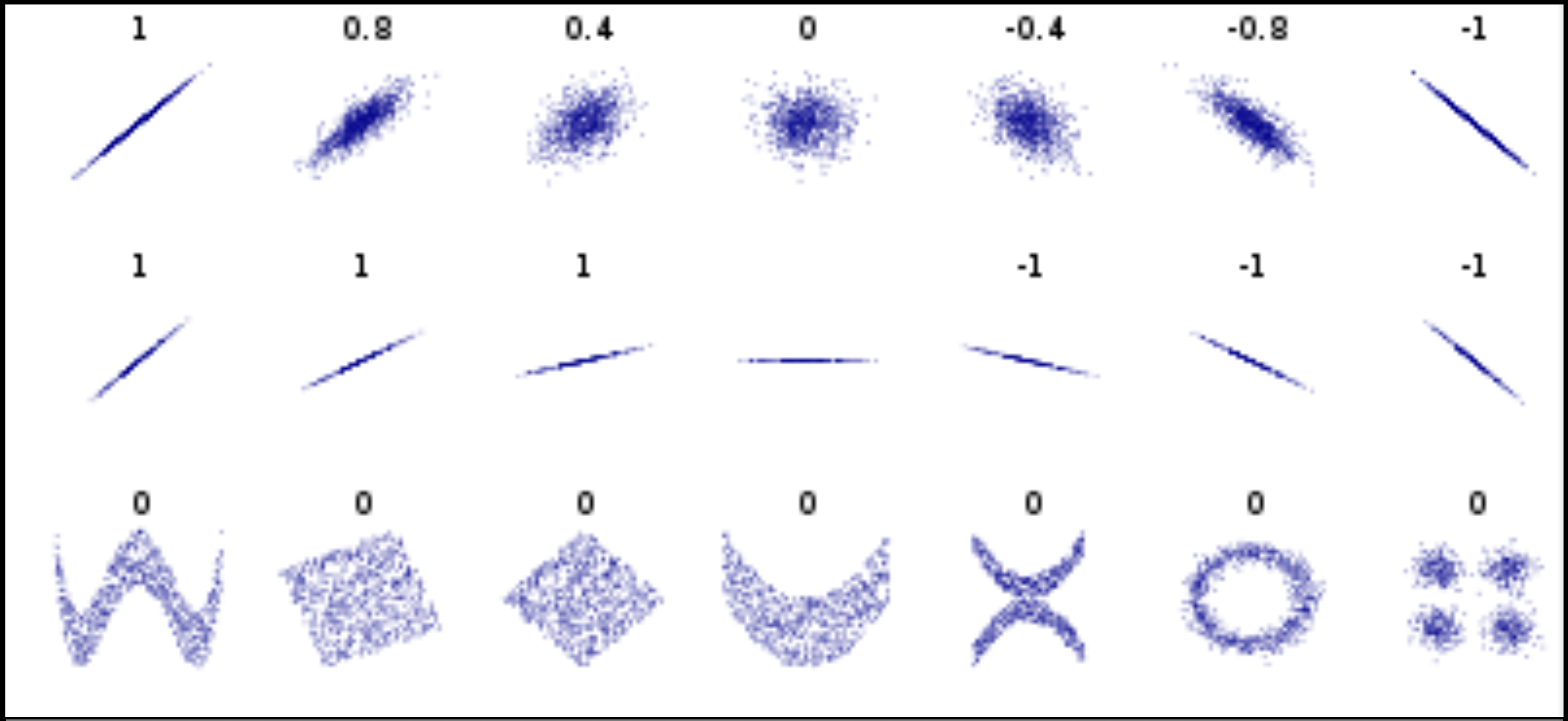


Correlation

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:

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

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.

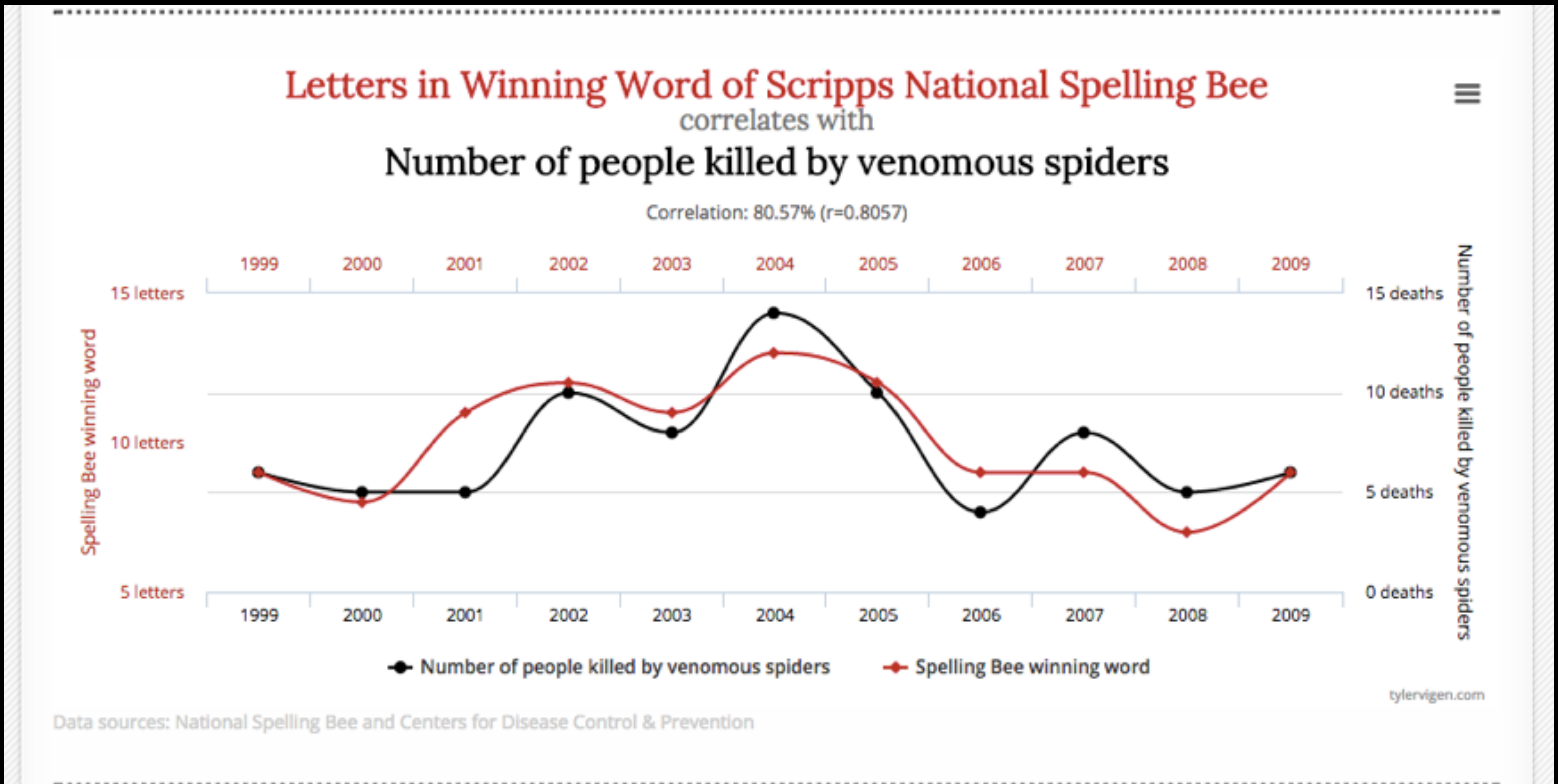
How many dichotomous+ (binary) variables?				
0	Both variables interval or ratio?			
	Y	Measures are linear? (No = monotonic*)		
		Y	Pearson correlation	
		N	Spearman correlation	
	N	Both variables are ordinal?		
		Y	Kendall correlation	
		N	Both variables can be ranked?	
			Y	Kendall correlation
			N	Convert to frequency data and use Chi-square test for independence
1	Biserial Correlation Coefficient			
2	2 x 2 table?			
	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

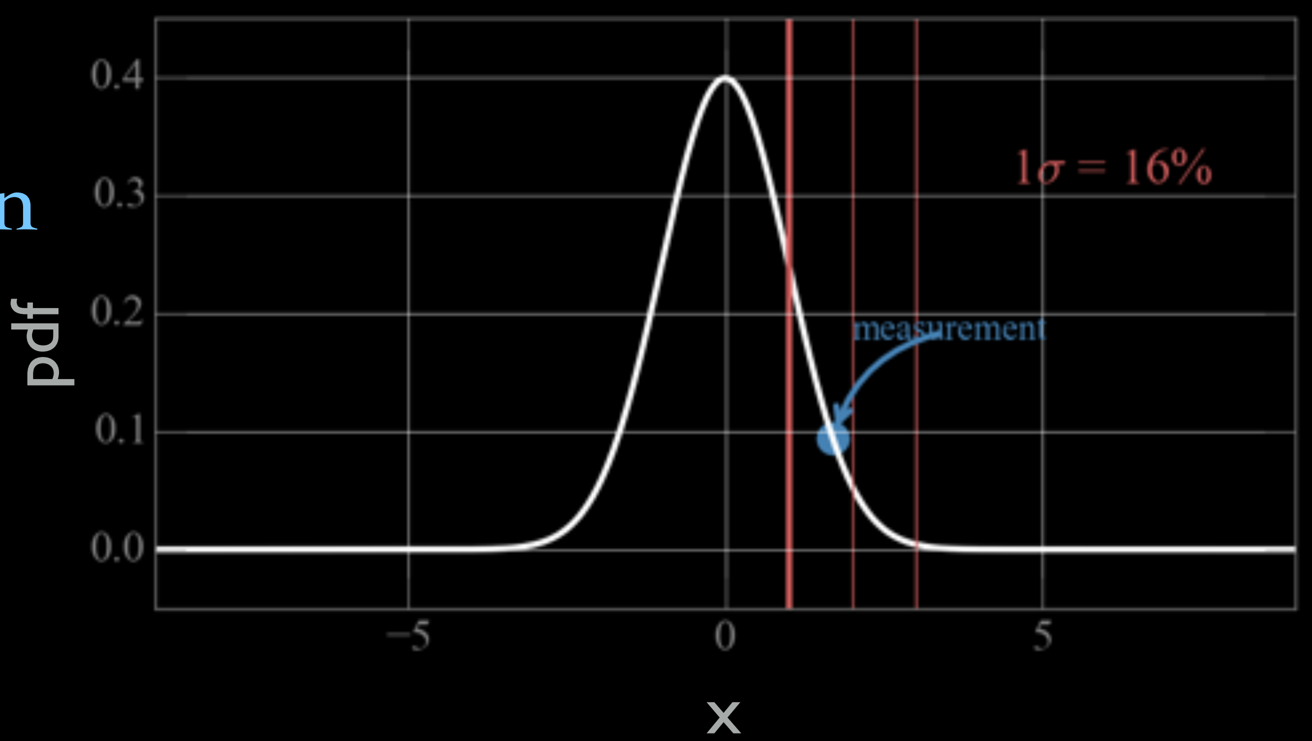
WARNING: Correlation is not causation!



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

Tests for correlation and independence (continuous variables)

Probability Distribution Function

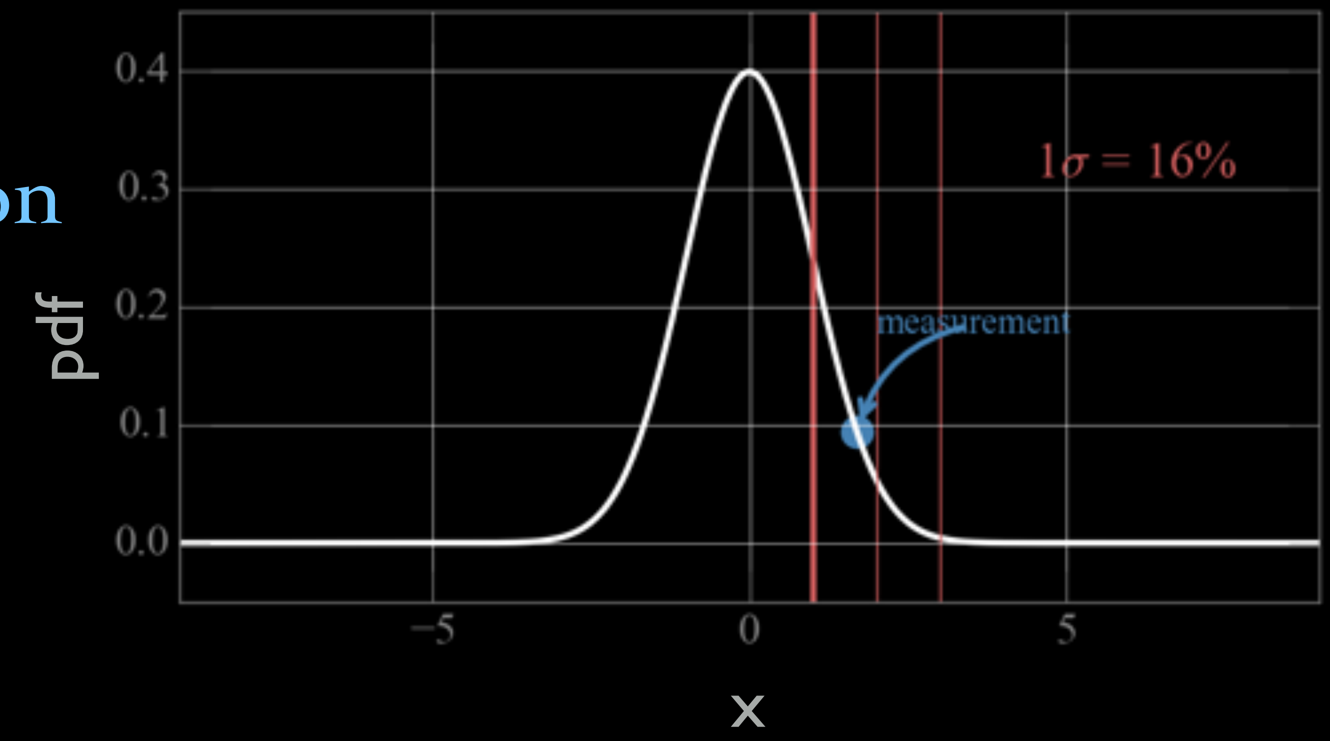


$$f_{x_0}(x) \sim \; p(x = x_0)$$

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

Probability Distribution Function

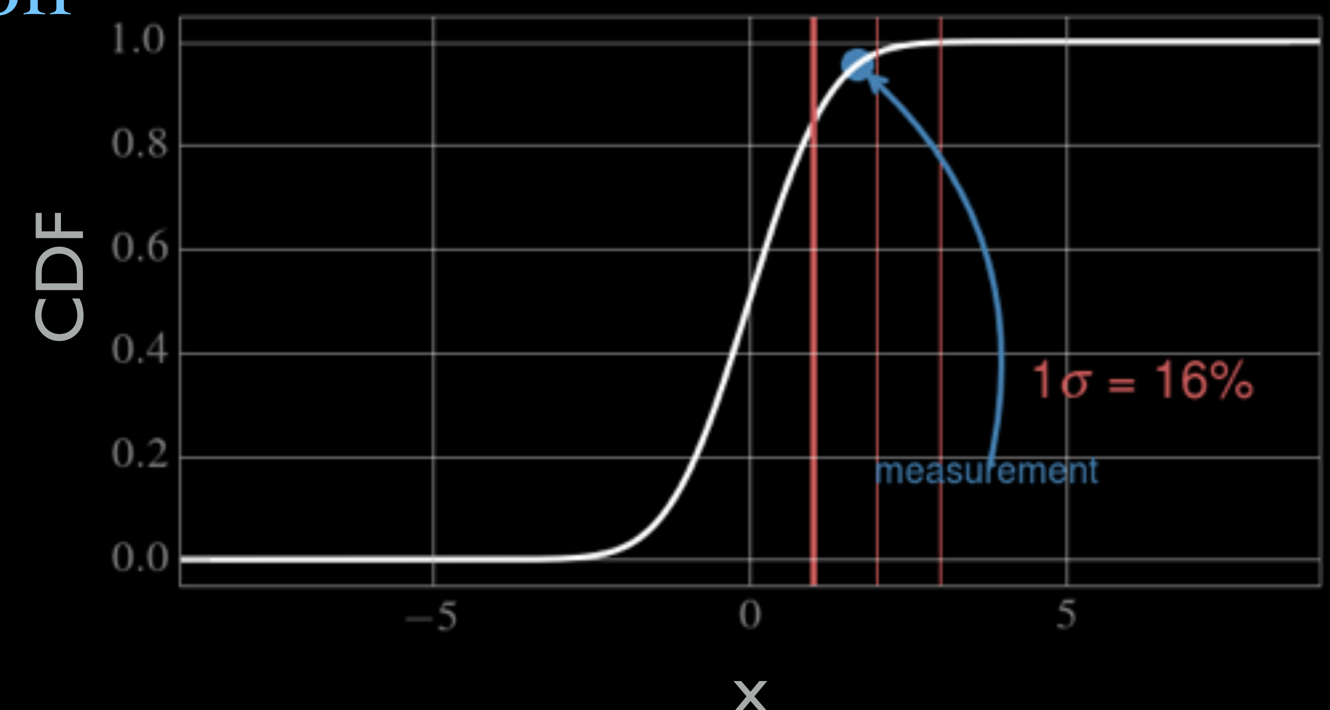
$$f_{x_0}(x) \sim p(x = x_0)$$

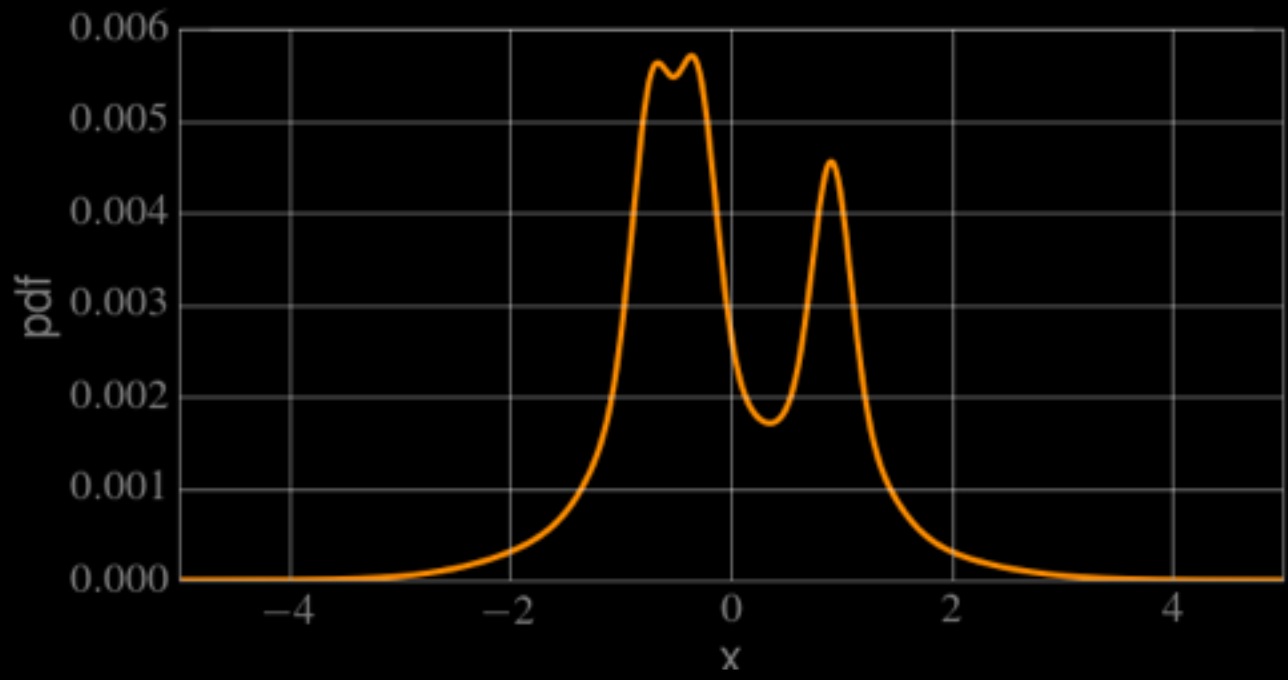


$$f_{x_0}(x) \sim p(x > x_0 - dx) \bigcap p(x < x_0 + dx)$$

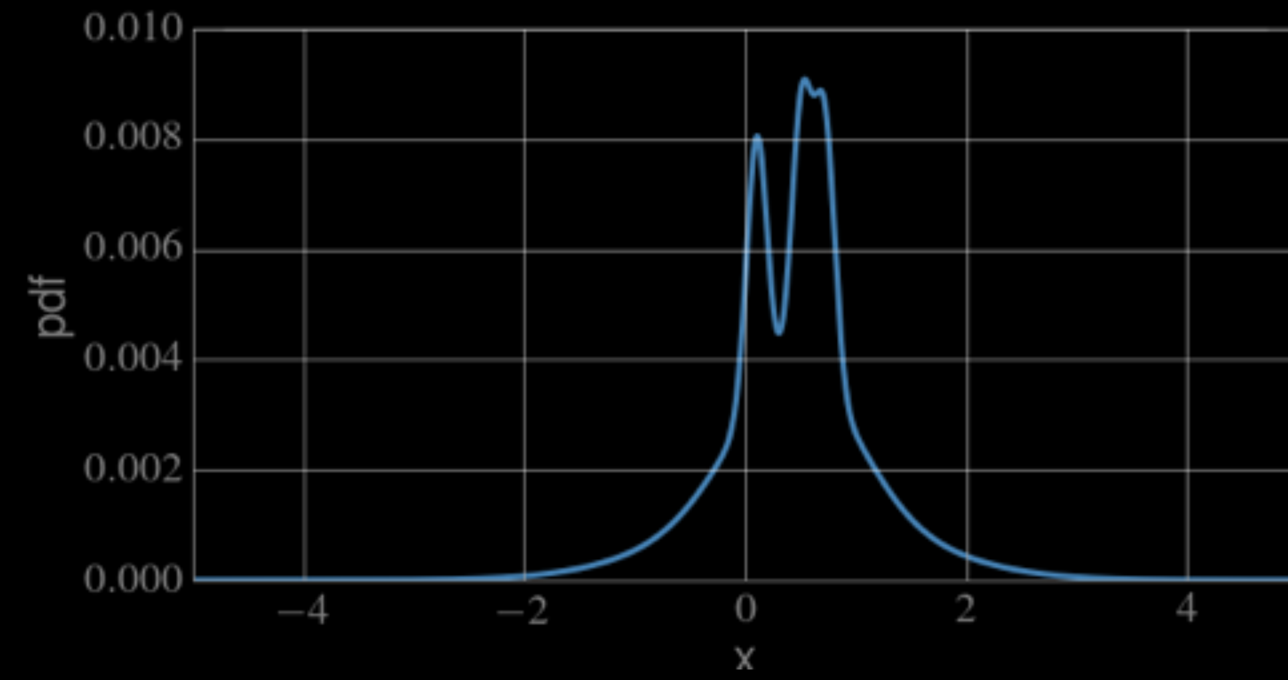
Cumulative Distribution Function

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

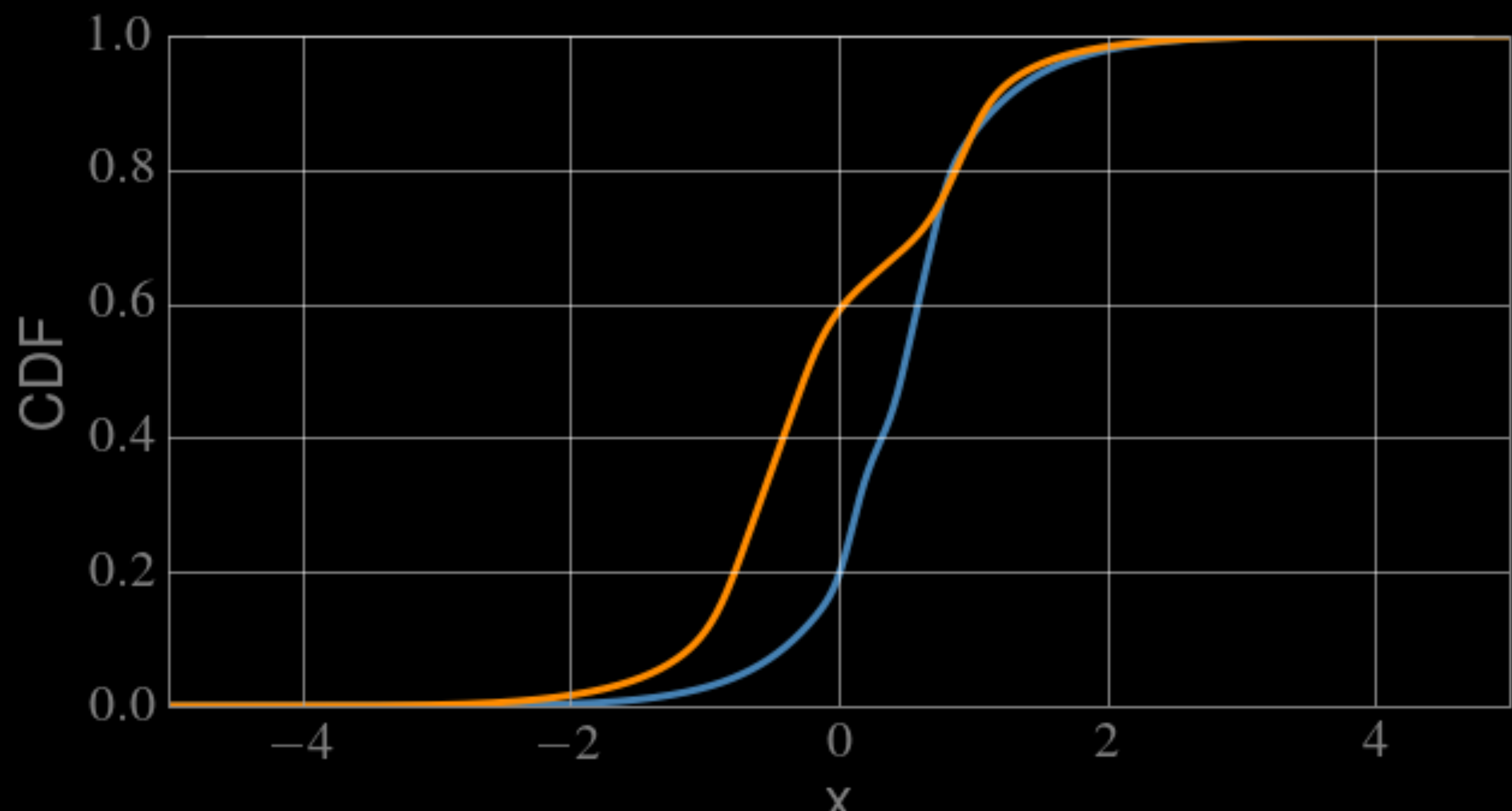




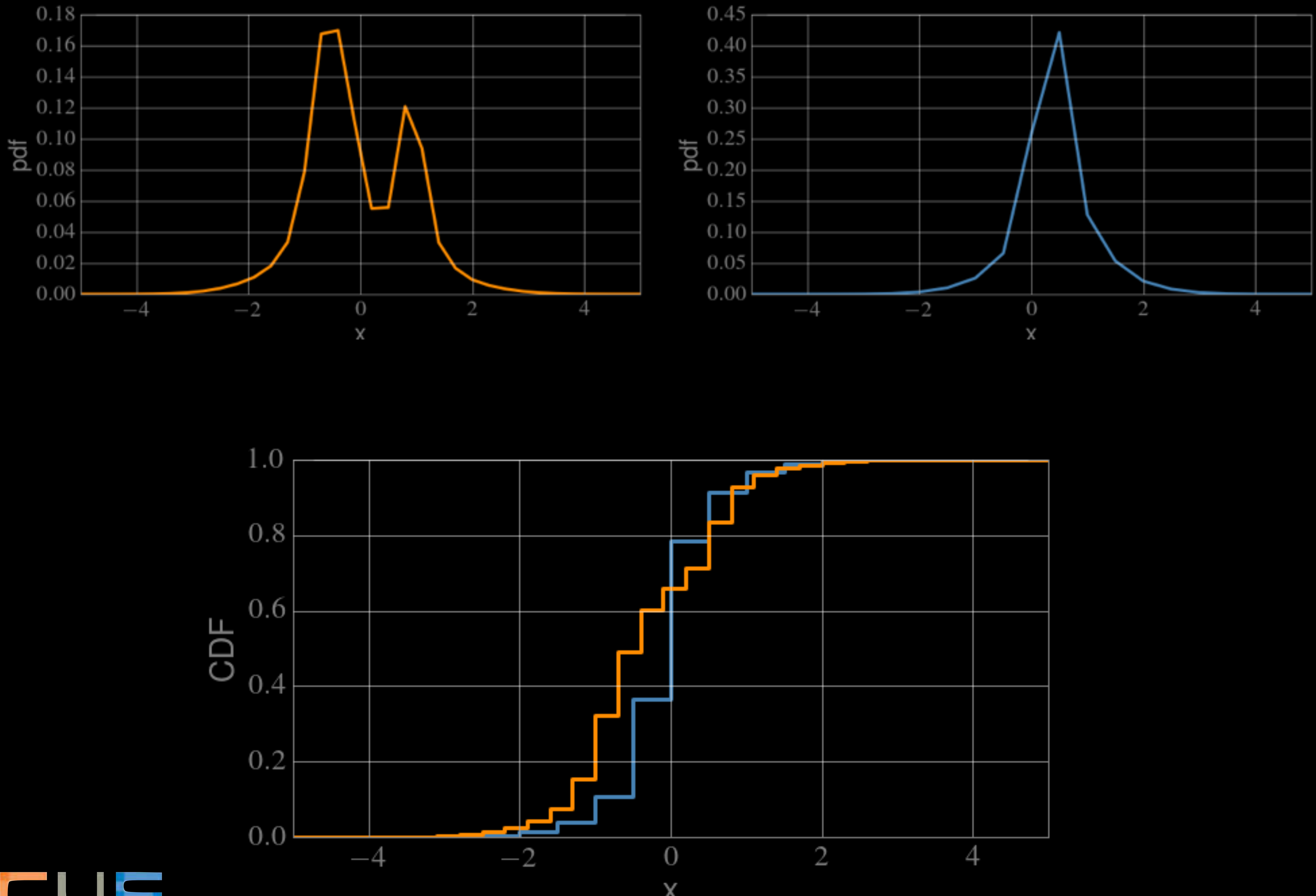
0.006
0.005
0.004
0.003
0.002
0.001
0.000
pdf
−4
−2
0
2
4
x



0.010
0.008
0.006
0.004
0.002
0.000
pdf
−4
−2
0
2
4
x



1.0
0.8
0.6
0.4
0.2
0.0
CDF
−4
−2
0
2
4
x



pdf
x
CDF

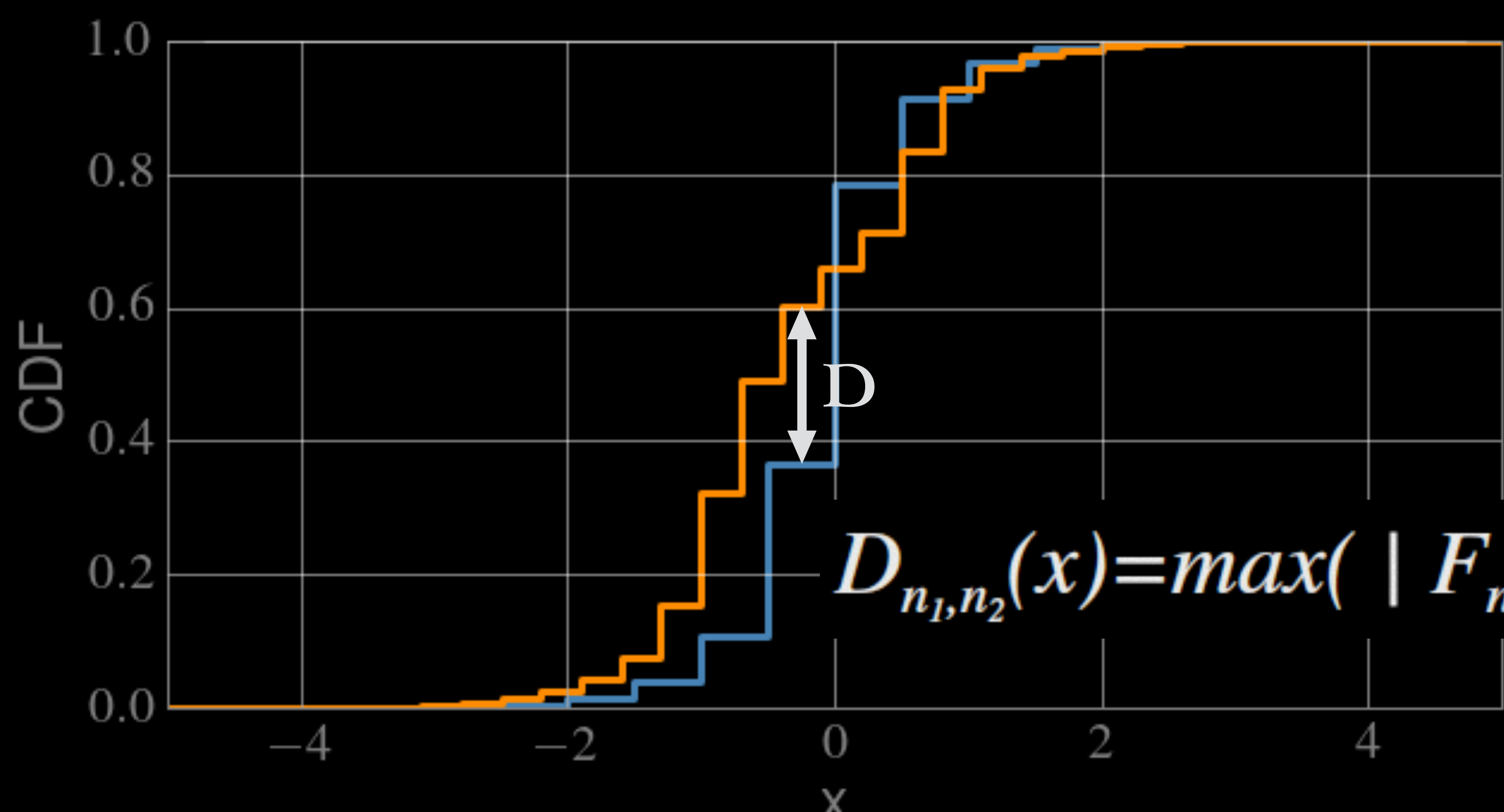
Two sample Kolmogorov Smirnoff test:

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(α) given by a table

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

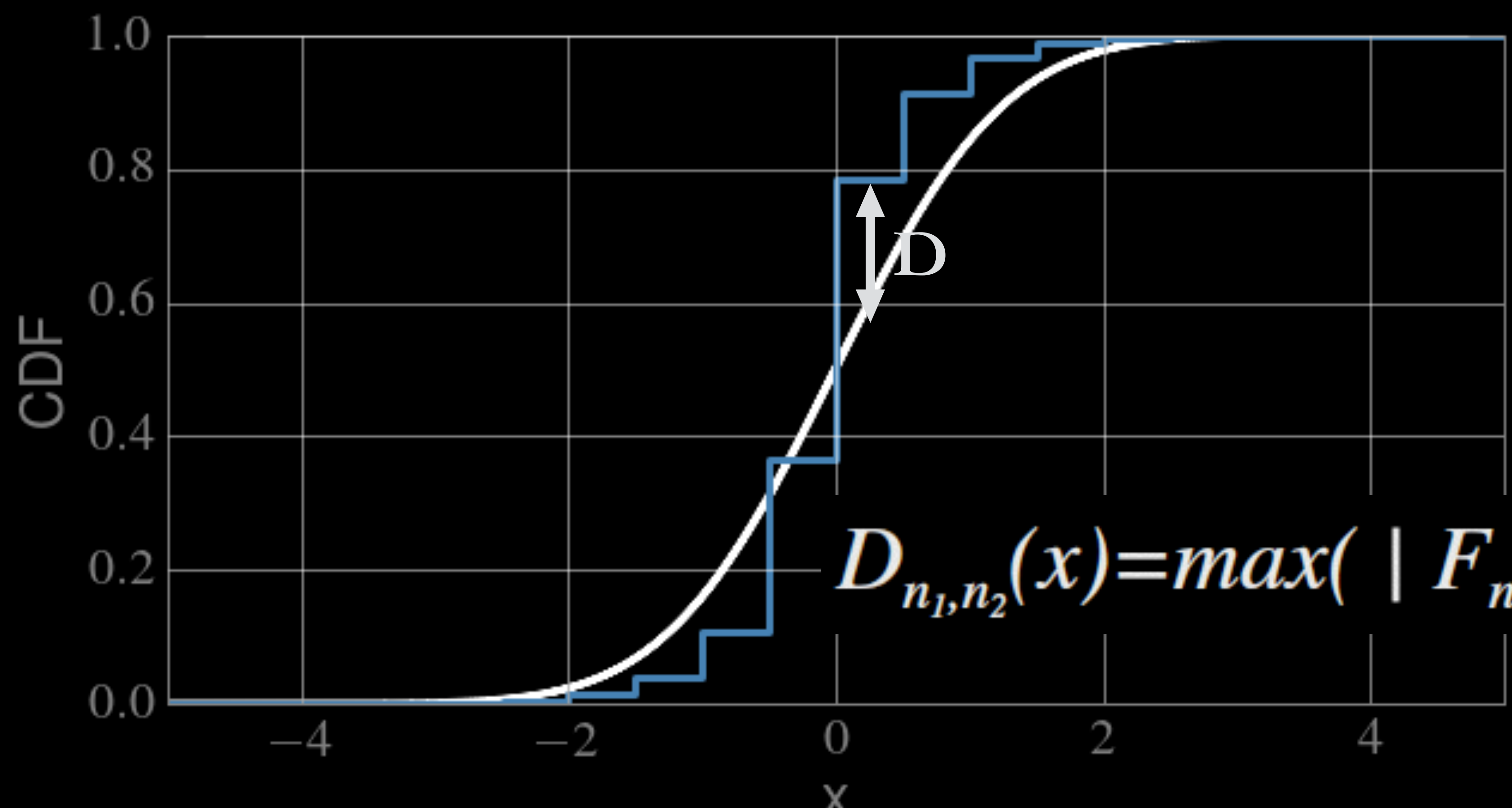


$$D_{n_1,n_2}(x)=max(\,|\,F_{n_1}(x)-F_{n_2}(x)\,|\,)$$

Goodness-of-fit Kolmogorov Smirnoff test:

null hypothesis H_0: the sample does comes from the <u>*model*</u> distribution

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



$$D_{n_1,n_2}(x) = max(\ |\ F_{n_1}(x) - F_{n_2}(x)\ |\ )$$

Tests Cheat Sheet:
2 (+) samples comparison

	metric (statistic)	compare to	
KS	$D_{n_1,n_2}(x)=max(\mid F_{n_1}(x)-F_{n_2}(x)\mid)$	$c(\alpha)\sqrt{\frac{n_1+n_2}{n_1 n_2}}$	*Non parametric 2 samples only*
K-sample Anderson-Darling	$ADK=\frac{n-1}{n^2(k-1)}\sum_{i=1}^{k}\frac{1}{n(i)}\left(\sum_{j=1}^{L}h_j\frac{(nF_{ij}-n_iH_j)^2}{H_j(n-H_j)-nh_j/4}\right)$	•AK table	*Non parametric, N samples*
Pearson's	$r_{xy}=\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{s_x}\right)\left(\frac{y_i-\overline{y}}{s_y}\right)$	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

assignment 3: Z-test and chi sq test

- Reproduce the analysis of the Hard to Employ program. Reproduce the results in cell 2 and 10. Follow the notebook in the HW directory (turn in the python notebook in the HW4_<netID> directory

assignment 4. 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. State your result in words in terms of the Null Hypothesis
Extra Credit: Use the age of bikers in day vs night and assess the correlation/independence of the 2 samples in each case..

CUSP

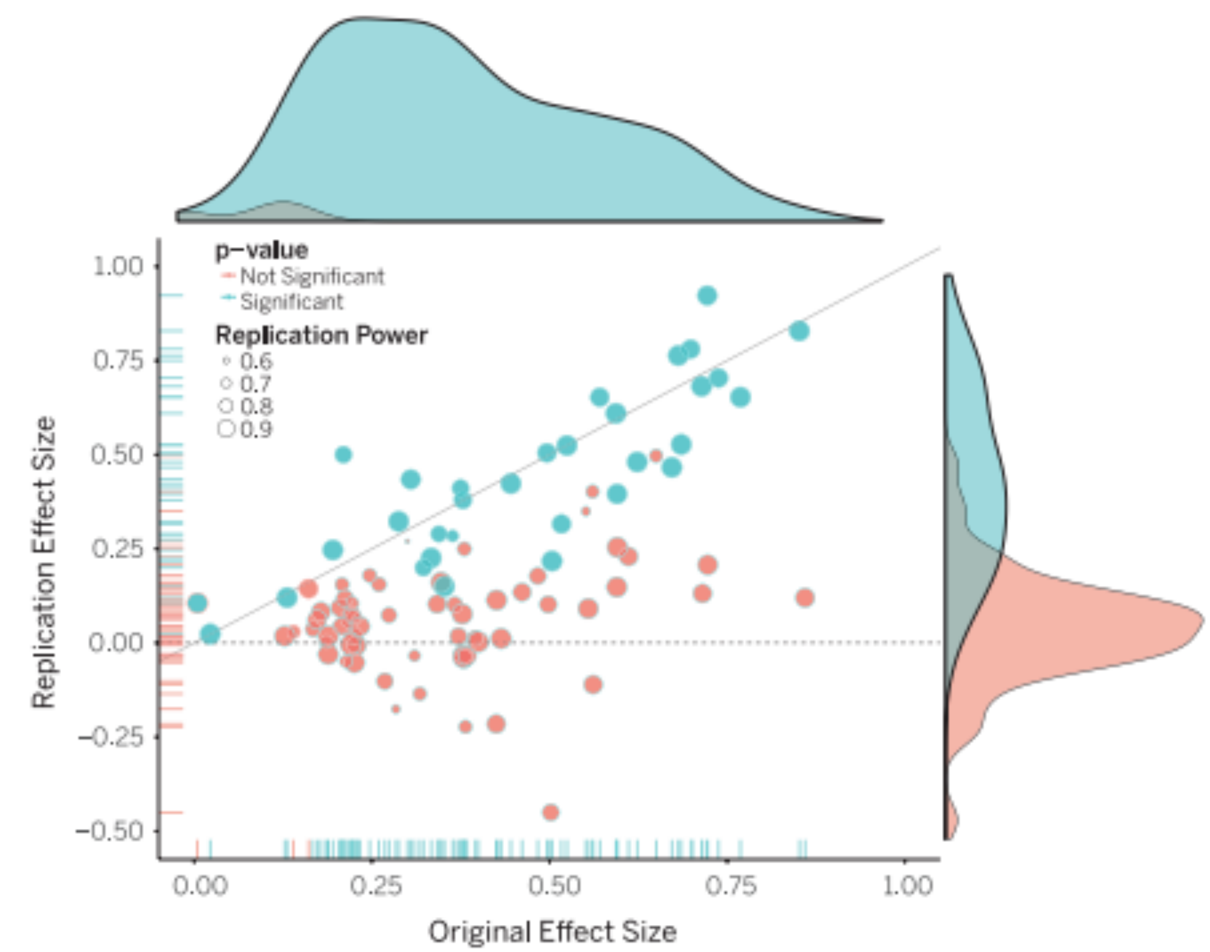
Homework: READING

RESEARCH ARTICLE SUMMARY

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*



Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.

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

MUST KNOWS:

- How to choose (and perform) a statistical test
- Statistical errors
- Systematic errors
- Undercoverage, SelfSelection, Social desirability publication bias, data dredging
- Precision vs accuracy
- PDF vs CDF
- correlation vs causation
- KS test for 2 samples, Pearson's, Spareman

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

Reckova & Irsova
Publication Bias in Measuring Climate Sensitivity
IES Working Paper: 14/2015
http://salserver.org.aalto.fi/vanhat_sivut/Opinnot/Mat-2.4108/pdf-files/emet03.pdf