Hypothesis for correlation and theories of hypothesis testing

Overview

Hypothesis tests for correlation

Hypothesis test for correlation in R

Theories of hypothesis testing

Parametric probability distributions

Hypothesis tests for correlation

Hypothesis tests for correlation

Is there a positive correlation between the number of carbohydrates in a cereal and the number calories?

What is the population parameter and the statistic of interest?

- Parameter: ρ
- Statistic: r

Significance tests for correlation

Suppose we had some data from 30 randomly selected cereals

	Calories	Carbohydrates
AppleJacks	117	27
Boo Berry	118	27
Cap'n Crunch	144	31
Cinnamon Toast Crunch	169	32

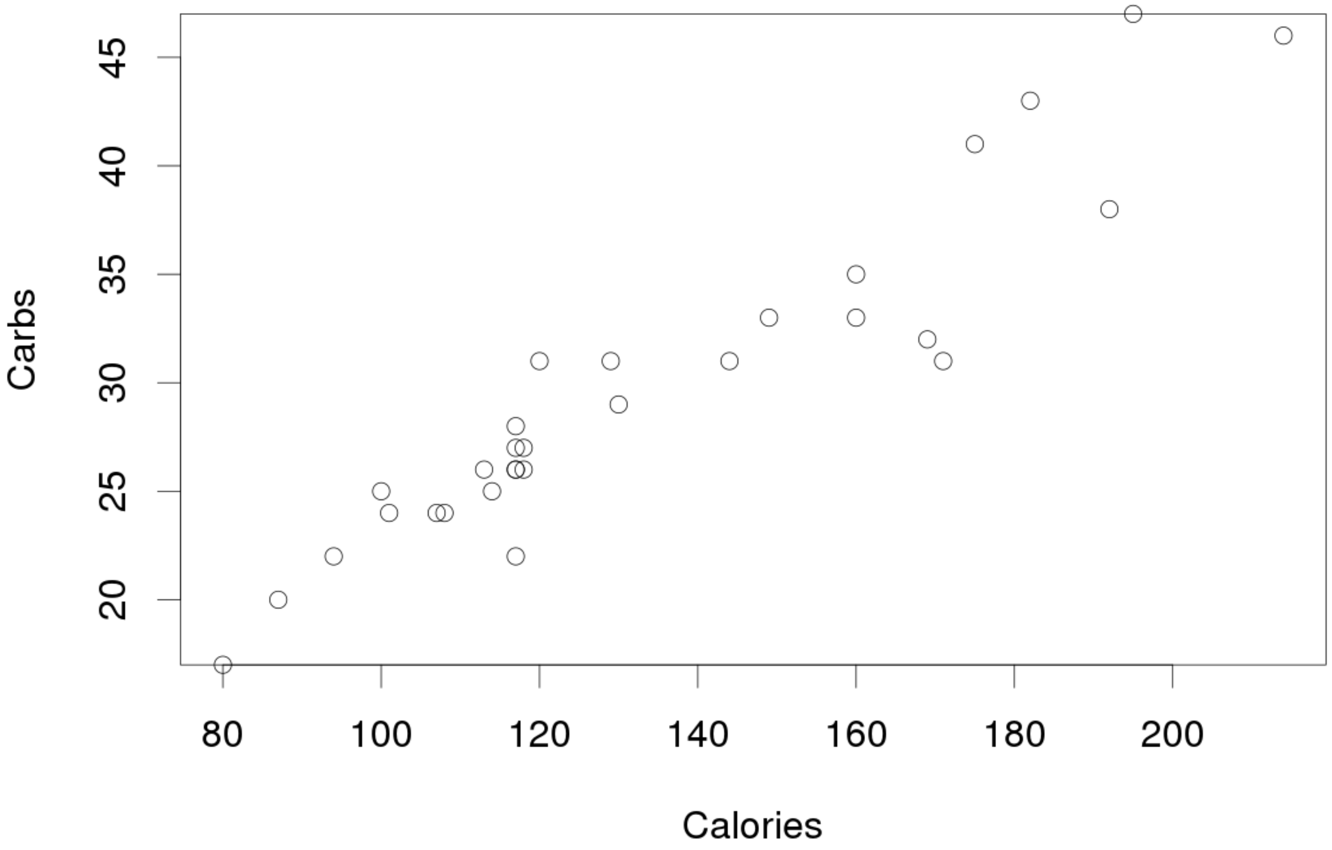
What is the first step we should do for running a hypothesis test?

Hypothesis testing for correlation

1. $H_0: \rho = 0$ There is no correlation between calories and carbs

   $H_A: \rho > 0$ There is a positive correlation between calories and carbs

Correlation between Carbohydrates and Calories



Guesses on the observed correlation r?

r = 0.94

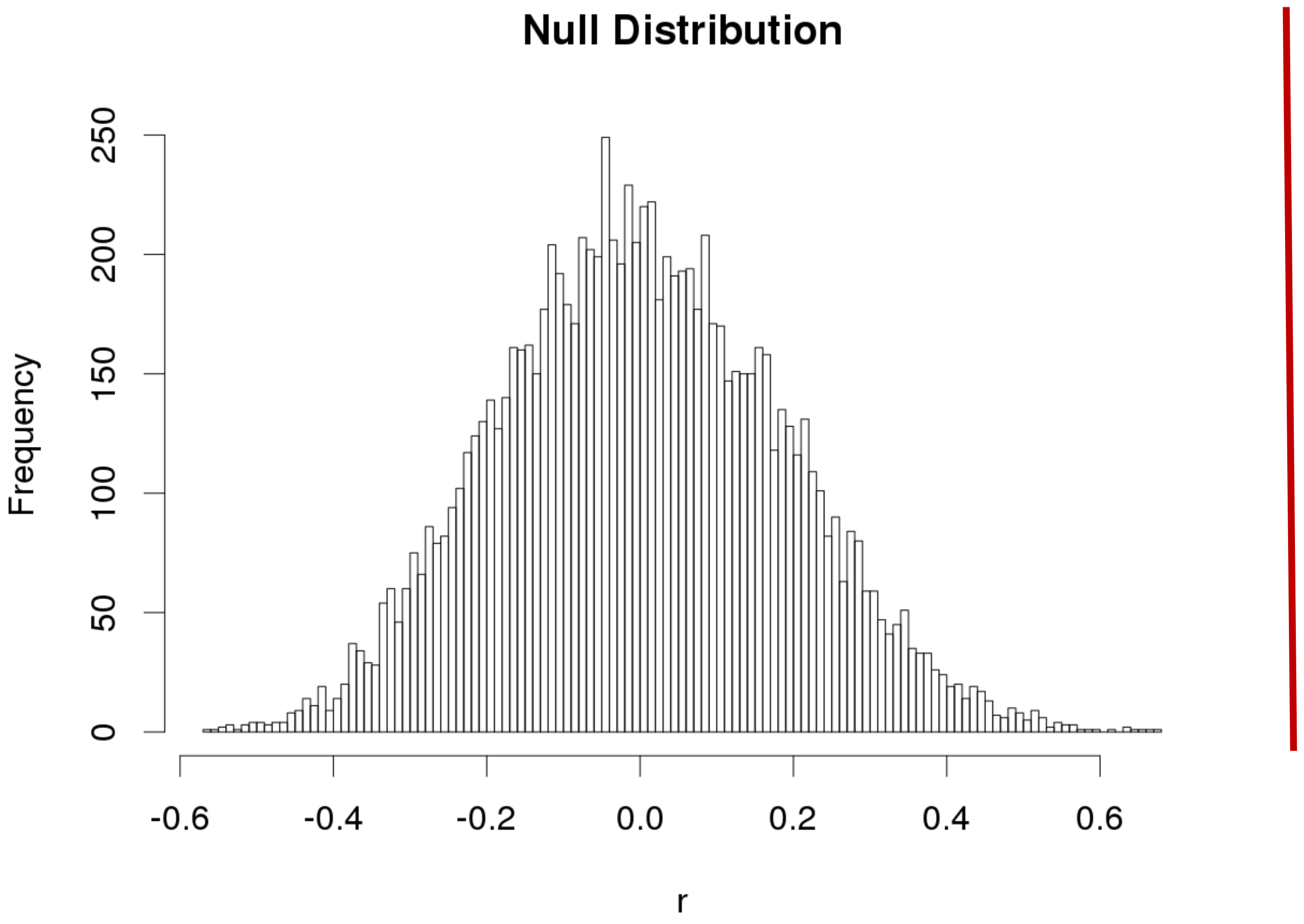
Creating the null distribution

	Calories	Carbohydrates
AppleJacks	117	27
Boo Berry	118	27
Cap'n Crunch	144	31
Cinnamon Toast Crunch	169	32

How can we create one point in the null distribution?

If there is no relationship between calories and carbohydrates (H_0 is true), then…

we can shuffle the order of the calories vector and get an equally valid r



Where is the observed statistic on this distribution?

What is the p-value?

NHST for correlation

What conclusion would you draw?

Let’s try it in R!

Homework 7 – part 3: 1969 Vietnam Draft

date	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	305	86	108	32	330	249	93	111	225	359	19	129
2	159	144	29	271	298	228	350	45	161	125	34	328
3	251	297	267	83	40	301	115	261	49	244	348	157
4	215	210	275	81	276	20	279	145	232	202	266	165
5	101	214	293	269	364	28	188	54	82	24	310	56
6	224	347	139	253	155	110	327	114	6	87	76	10
7	306	91	122	147	35	85	50	168	8	234	51	12
8	199	181	213	312	321	366	13	48	184	283	97	105
9	194	338	317	219	197	335	277	106	263	342	80	43
10	325	216	323	218	65	206	284	21	71	220	282	41
11	329	150	136	14	37	134	248	324	158	237	46	39
12	221	68	300	346	133	272	15	142	242	72	66	314
13	318	152	259	124	295	69	42	307	175	138	126	163
14	238	4	354	231	178	356	331	198	1	294	127	26
15	17	89	169	273	130	180	322	102	113	171	131	320
16	121	212	166	148	55	274	120	44	207	254	107	96
17	235	189	33	260	112	73	98	154	255	288	143	304
18	140	292	332	90	278	341	190	141	246	5	146	128
19	58	25	200	336	75	104	227	311	177	241	203	240
20	280	302	239	345	183	360	187	344	63	192	185	135
21	186	363	334	62	250	60	27	291	204	243	156	70
22	337	290	265	316	326	247	153	339	160	117	9	53
23	118	57	256	252	319	109	172	116	119	201	182	162
24	59	236	258	2	31	358	23	36	195	196	230	95
25	52	179	343	351	361	137	67	286	149	176	132	84
26	92	365	170	340	357	22	303	245	18	7	309	173
27	355	205	268	74	296	64	289	352	233	264	47	78
28	77	299	223	262	308	222	88	167	257	94	281	123
29	349	285	362	191	226	353	270	61	151	229	99	16
30	164		217	208	103	209	287	333	315	38	174	3
31	211		30		313		193	11		79		100

The first date picked was Sept 14 (sequential number 258)

date	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	305	86	108	32	330	249	93	111	225	359	19	129
2	159	144	29	271	298	228	350	45	161	125	34	328
3	251	297	267	83	40	301	115	261	49	244	348	157
4	215	210	275	81	276	20	279	145	232	202	266	165
5	101	214	293	269	364	28	188	54	82	24	310	56
6	224	347	139	253	155	110	327	114	6	87	76	10
7	306	91	122	147	35	85	50	168	8	234	51	12
8	199	181	213	312	321	366	13	48	184	283	97	105
9	194	338	317	219	197	335	277	106	263	342	80	43
10	325	216	323	218	65	206	284	21	71	220	282	41
11	329	150	136	14	37	134	248	324	158	237	46	39
12									242	72	66	314
13									175	138	126	163
14	238	4	354	231	178	356	331	198	1	294	127	26
15	17	89	169	273	130	180	322	102	113	171	131	320
16	121	212	166	148	55	274	120	44	207	254	107	96
17	235	189	33	260	112	73	98	154	255	288	143	304
18	140	292	332	90	278	341	190	141	246	5	146	128
19	58	25	200	336	75	104	227	311	177	241	203	240
20	280	302	239	345	183	360	187	344	63	192	185	135
21	186	363	334	62	250	60	27	291	204	243	156	70
22	337	290	265	316	326	247	153	339	160	117	9	53
23	118	57	256	252	319	109	172	116	119	201	182	162
24	59	236	258	2	31	358	23	36	195	196	230	95
25	52	179	343	351	361	137	67	286	149	176	132	84
26	92	365	170	340	357	22	303	245	18	7	309	173
27	355	205	268	74	296	64	289	352	233	264	47	78
28	77	299	223	262	308	222	88	167	257	94	281	123
29	349	285	362	191	226	353	270	61	151	229	99	16
30	164		217	208	103	209	287	333	315	38	174	3
31	211		30		313		193	11		79		100

date	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	305	86	108	32	330	249	93	111	225	359	19	129
2	159	144	29	271	298	228	350	45	161	125	34	328
3	251	297	267	83	40	301	115	261	49	244	348	157
4	215	210	275	81	276	20	279	145	232	202	266	165
5	101	214	293	269	364	28	188	54	82	24	310	56
6	224	347	139	253	155	110	327	114	6	87	76	10
7	306	91	122	147	35	85	50	168	8	234	51	12
8	199	181	213	312	321	366	13	48	184	283	97	105
9	194	338	317	219	197	335	277	106	263	342	80	43
10	325	216	323	218	65	206	284	21	71	220	282	41
11	329	150	136	14	37	134	248	324	158	237	46	39
12	221	68	300	346	133	272	15	142	242	72	66	314
13	318	152	259	124	295	69	42	307	175	138	126	163
14	238	4	354	231	178	356	331	198	1	294	127	26
15	17	89	169	273	130	180	322	102	113	171	131	320
16	121	212	166	148	55	274	120	44	207	254	107	96
17	235	189	33	260	112	73	98	154	255	288	143	304
18	140	292	332	90	278	341	190	141	246	5	146	128
19	58	25	200	336	75	104	227	311	177	241	203	240
20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	135
21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	70
22	337	290	265	316	326	247	153	339	160	117	9	53
23	118	57	256	252	319	109	172	116	119	201	182	162
24	59	236	258	2	31	358	23	36	195	196	230	95
25	52	179	343	351	361	137	67	286	149	176	132	84
26	92	365	170	340	357	22	303	245	18	7	309	173
27	355	205	268	74	296	64	289	352	233	264	47	78
28	77	299	223	262	308	222	88	167	257	94	281	123
29	349	285	362	191	226	353	270	61	151	229	99	16
30	164		217	208	103	209	287	333	315	38	174	3
31	211		30		313		193	11		79		100

The second date picked was April 24th (sequential number 115)

What is your Draft number?

date	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	305	86	108	32	330	249	93	111	225	359	19	129
2	159	144	29	271	298	228	350	45	161	125	34	328
3	251	297	267	83	40	301	115	261	49	244	348	157
4	215	210	275	81	276	20	279	145	232	202	266	165
5	101	214	293	269	364	28	188	54	82	24	310	56
6	224	347	139	253	155	110	327	114	6	87	76	10
7	306	91	122	147	35	85	50	168	8	234	51	12
8	199	181	213	312	321	366	13	48	184	283	97	105
9	194	338	317	219	197	335	277	106	263	342	80	43
10	325	216	323	218	65	206	284	21	71	220	282	41
11	329	150	136	14	37	134	248	324	158	237	46	39
12	221	68	300	346	133	272	15	142	242	72	66	314
13	318	152	259	124	295	69	42	307	175	138	126	163
14	238	4	354	231	178	356	331	198	1	294	127	26
15	17	89	169	273	130	180	322	102	113	171	131	320
16	121	212	166	148	55	274	120	44	207	254	107	96
17	235	189	33	260	112	73	98	154	255	288	143	304
18	140	292	332	90	278	341	190	141	246	5	146	128
19	58	25	200	336	75	104	227	311	177	241	203	240
20	280	302	239	345	183	360	187	344	63	192	185	135
21	186	363	334	62	250	60	27	291	204	243	156	70
22	337	290	265	316	326	247	153	339	160	117	9	53
23	118	57	256	252	319	109	172	116	119	201	182	162
24	59	236	258	2	31	358	23	36	195	196	230	95
25	52	179	343	351	361	137	67	286	149	176	132	84
26	92	365	170	340	357	22	303	245	18	7	309	173
27	355	205	268	74	296	64	289	352	233	264	47	78
28	77	299	223	262	308	222	88	167	257	94	281	123
29	349	285	362	191	226	353	270	61	151	229	99	16
30	164		217	208	103	209	287	333	315	38	174	3
31	211		30		313		193	11		79		100

1969 Vietnam Draft sorted by sequential date

Date	Sequential date	Draft number
Jan 1	1	305
Jan 2	2	159
Jan 3	3	251
Jan 4	4	215
Jan 5	5	101
Jan 6	6	224
Jan 7	7	306
Jan 8	8	199
Jan 9	9	194

1969 Vietnam Draft

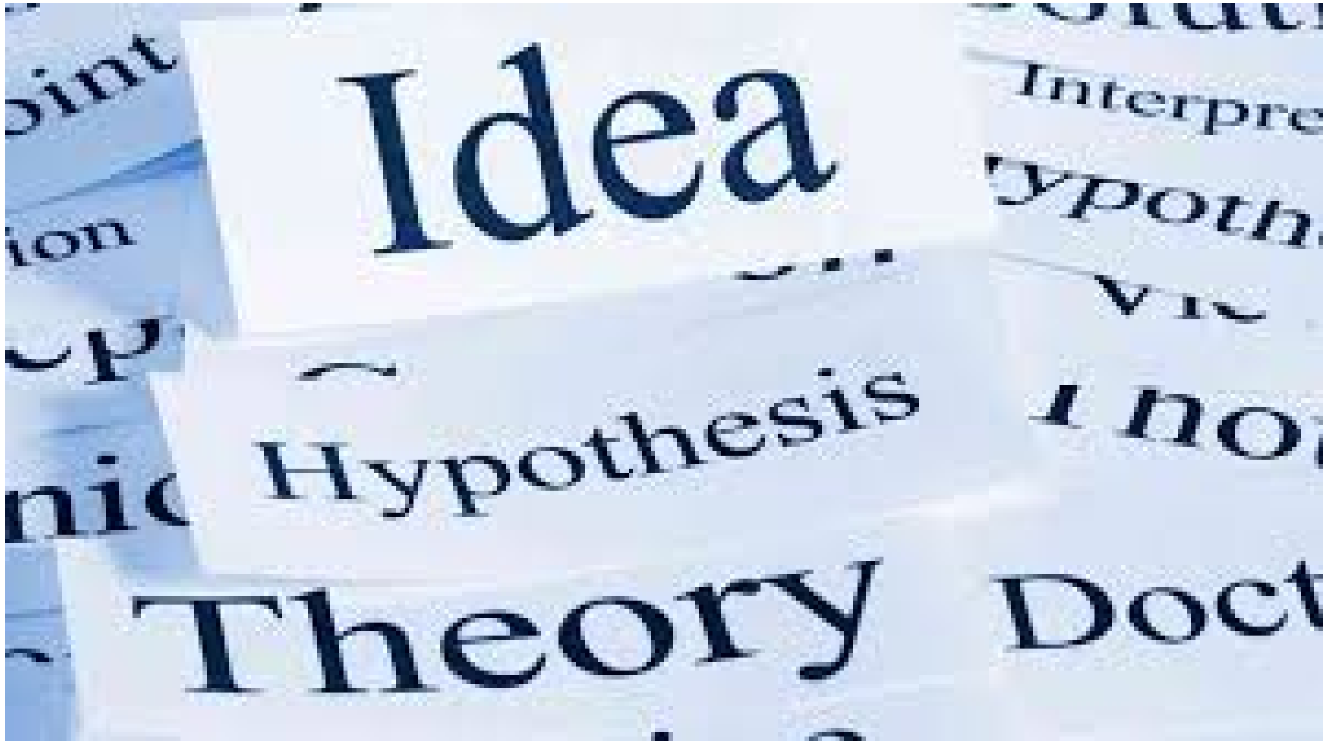
In a perfectly “fair”, random lottery, what should be the value of the correlation coefficient between **draft number** and **sequential date** of birthday?

Homework 7

Use hypothesis testing to assess whether there is a correlation between sequential date and draft number

- i.e., was the draft really random?

Theories of hypothesis tests



Two theories of hypothesis testing

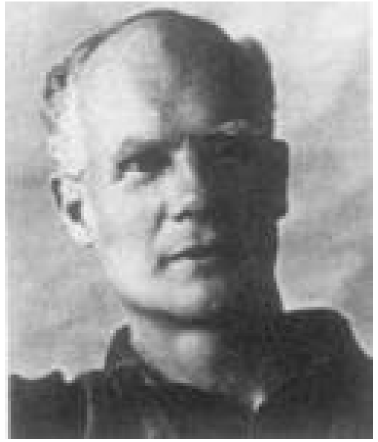
Null-hypothesis significance testing (NHST) is a hybrid of two theories:

1. Significance testing of Ronald Fisher

2. Hypothesis testing of Jezy Neyman and Egon Pearson

Fisher (1890-1962)

Neyman (1894-1981)

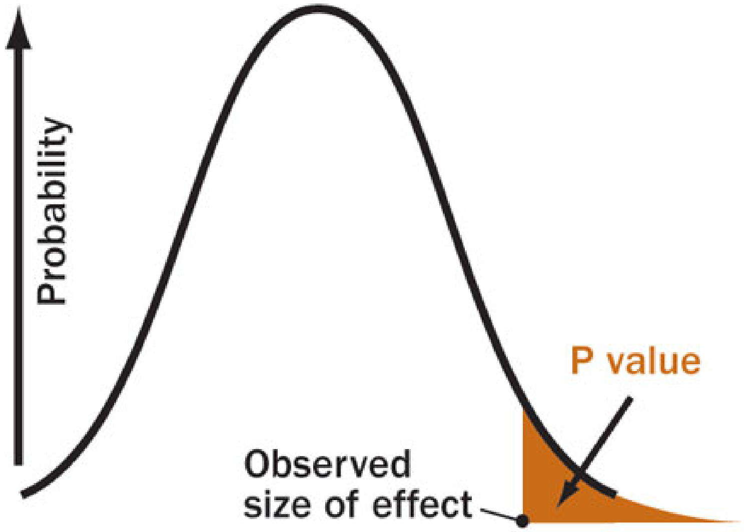


Pearson (1895-1980)

Ronald Fisher’s significance testing

Views the p-value as strength of evidence against the null hypothesis

- P-values part of an on-going scientific process: tells the experimenter “what results to ignore”



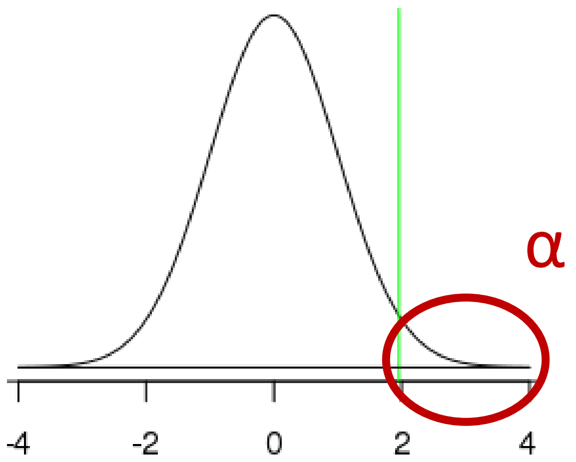
Neyman-Pearson null hypothesis testing

Makes ***a formal decision*** in statistical tests

Reject H_0: if the observed sample statistic is beyond a fixed value

- i.e., reject H_0 if the p-value is less than some predetermined **significance level** α

Null distribution



Do not reject H_0: if the observed sample statistic is not beyond a fixed value. This means the test is inconclusive.

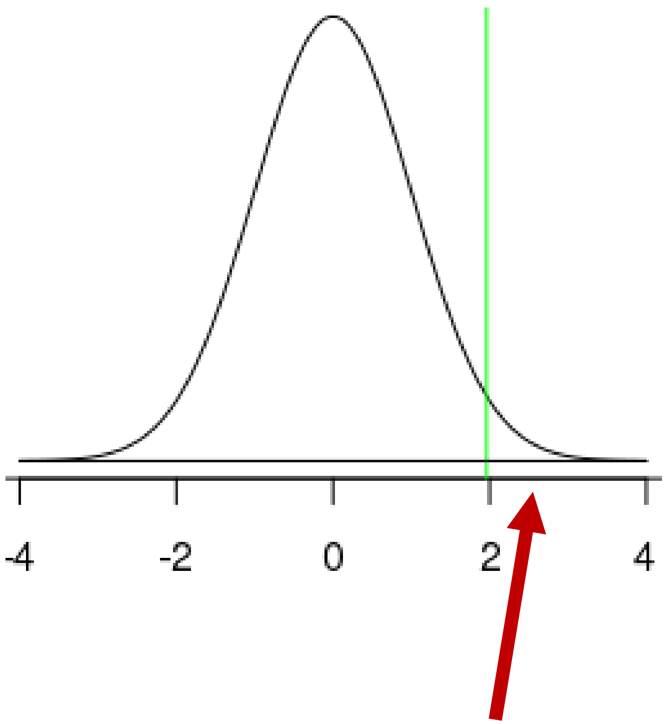
Neyman-Pearson frequentist logic

Type I error: incorrectly rejecting the null hypothesis when it is true

If Neyman-Pearson null hypothesis testing paradigm was followed perfectly, then only ~5% of all published research findings should be wrong (for $\alpha = 0.05$)

- i.e., we would only make type I errors 5% of the time

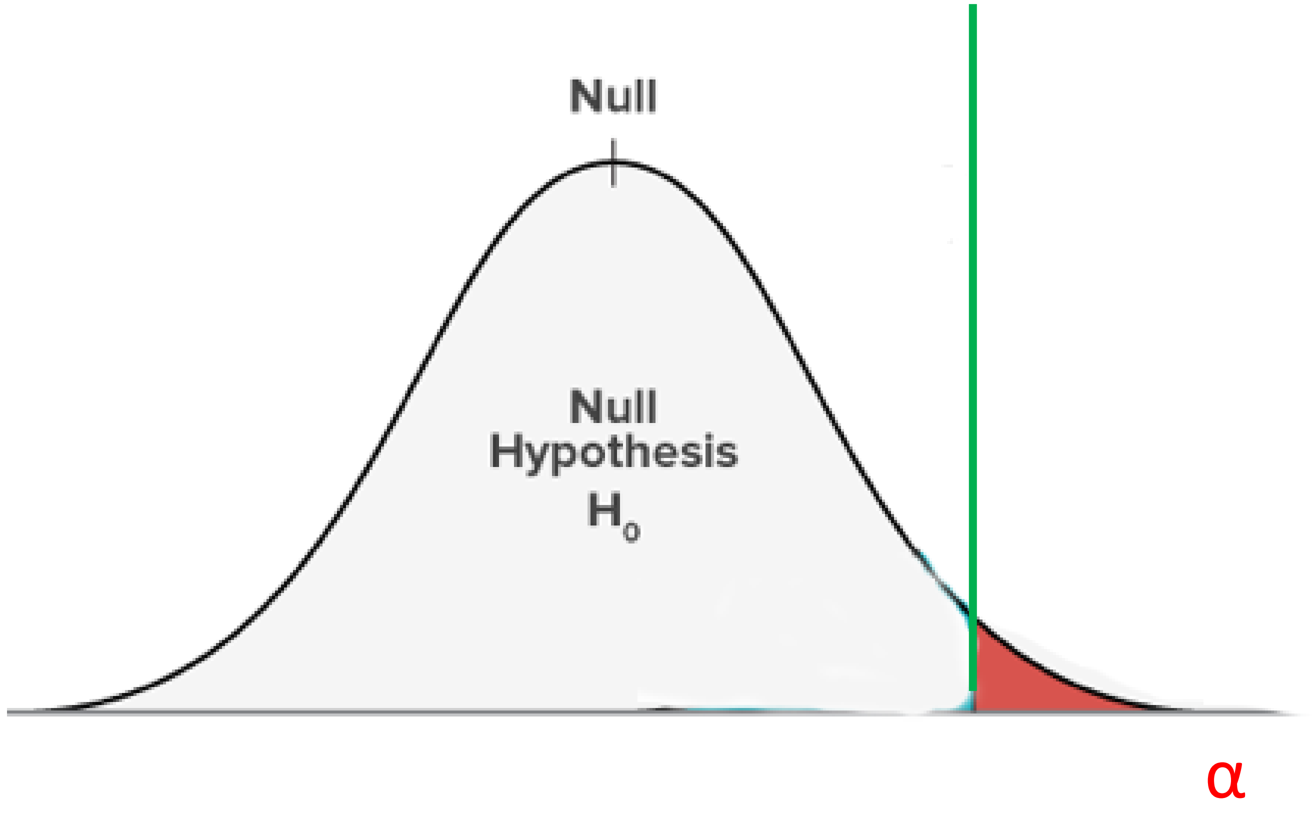
Null distribution



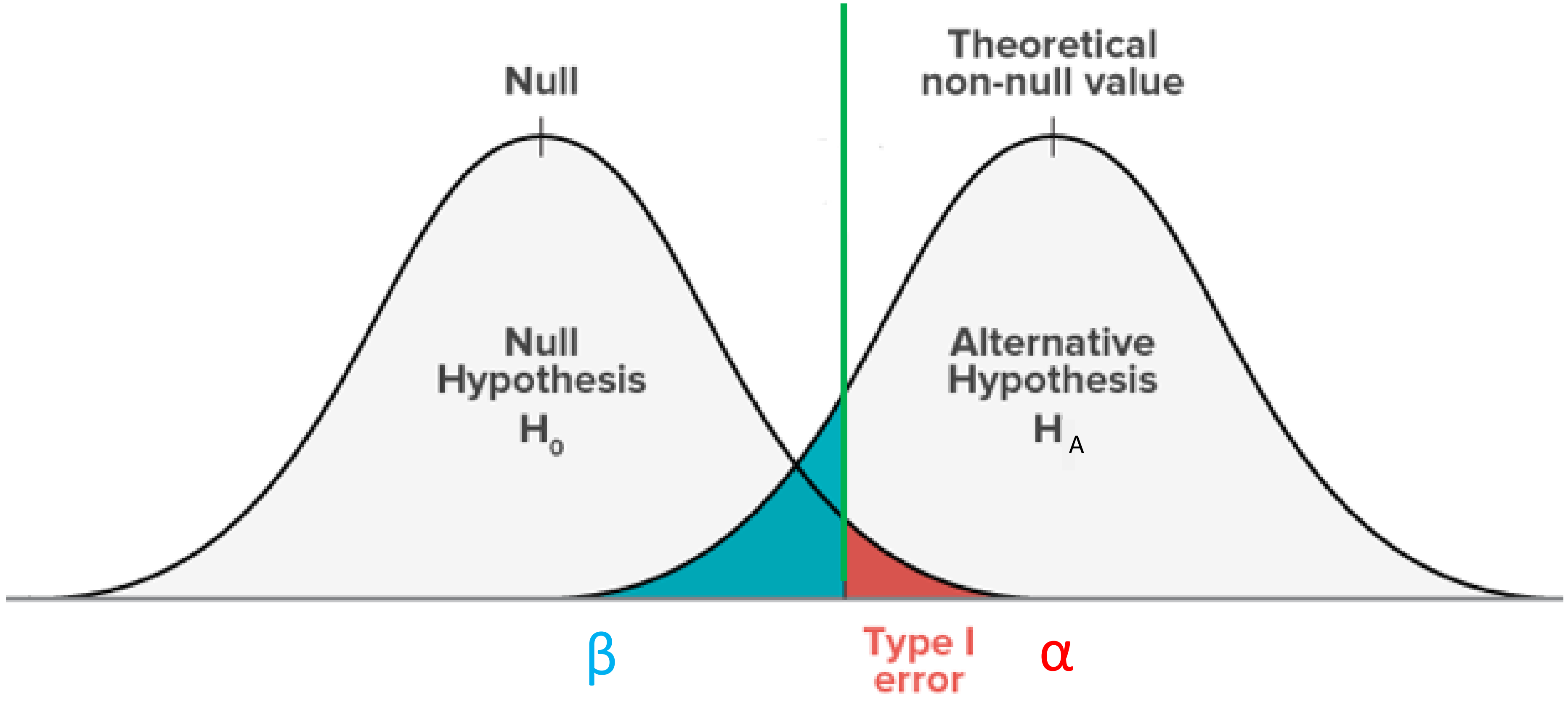
The null distribution is true but statistic landed here


HEY GIRL
I MADE A TYPE 1 ERROR, I SHOULDN'T HAVE REJECTED YOU
memegenerator.net

Neyman-Pearson Frequentist logic



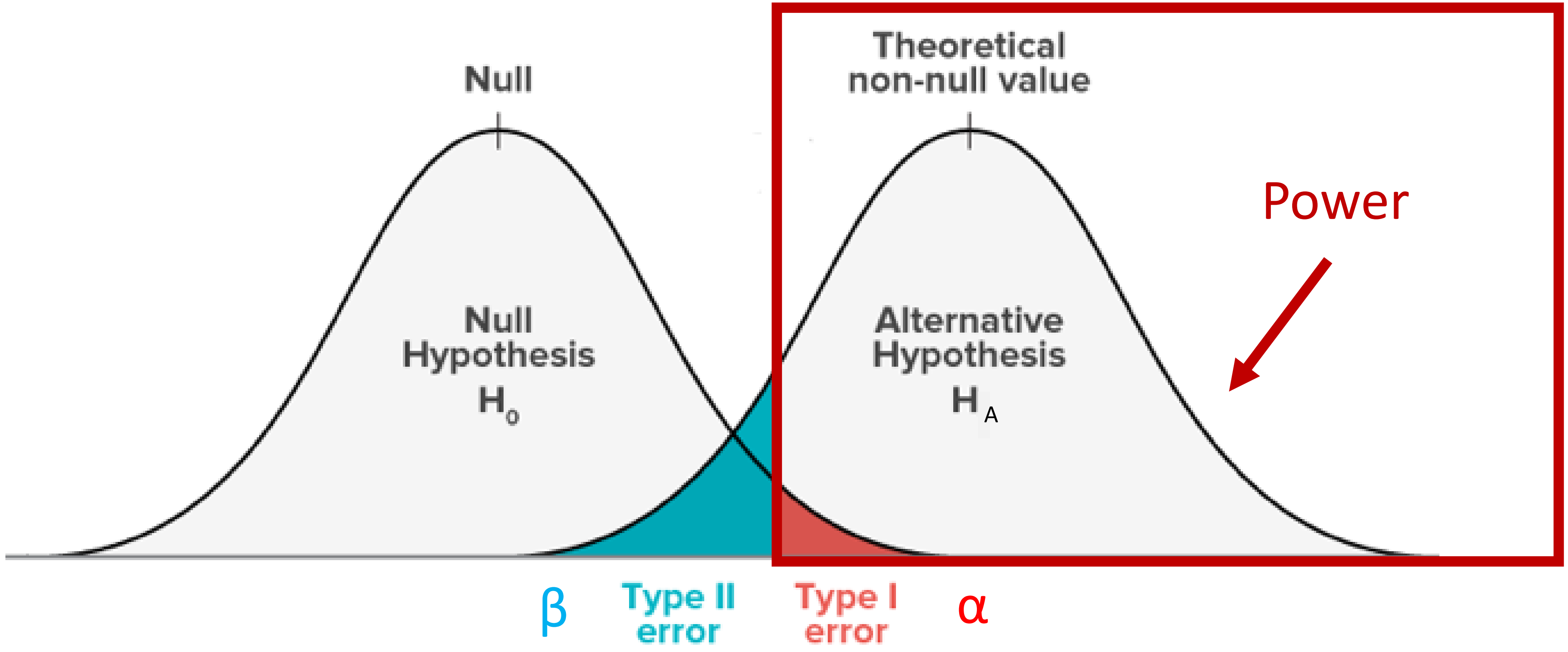
Neyman-Pearson Frequentist logic



Type 2 error: incorrectly rejecting failing to reject H_0 when it is false

- The rate at which we make type 2 errors is often denoted with the symbol β

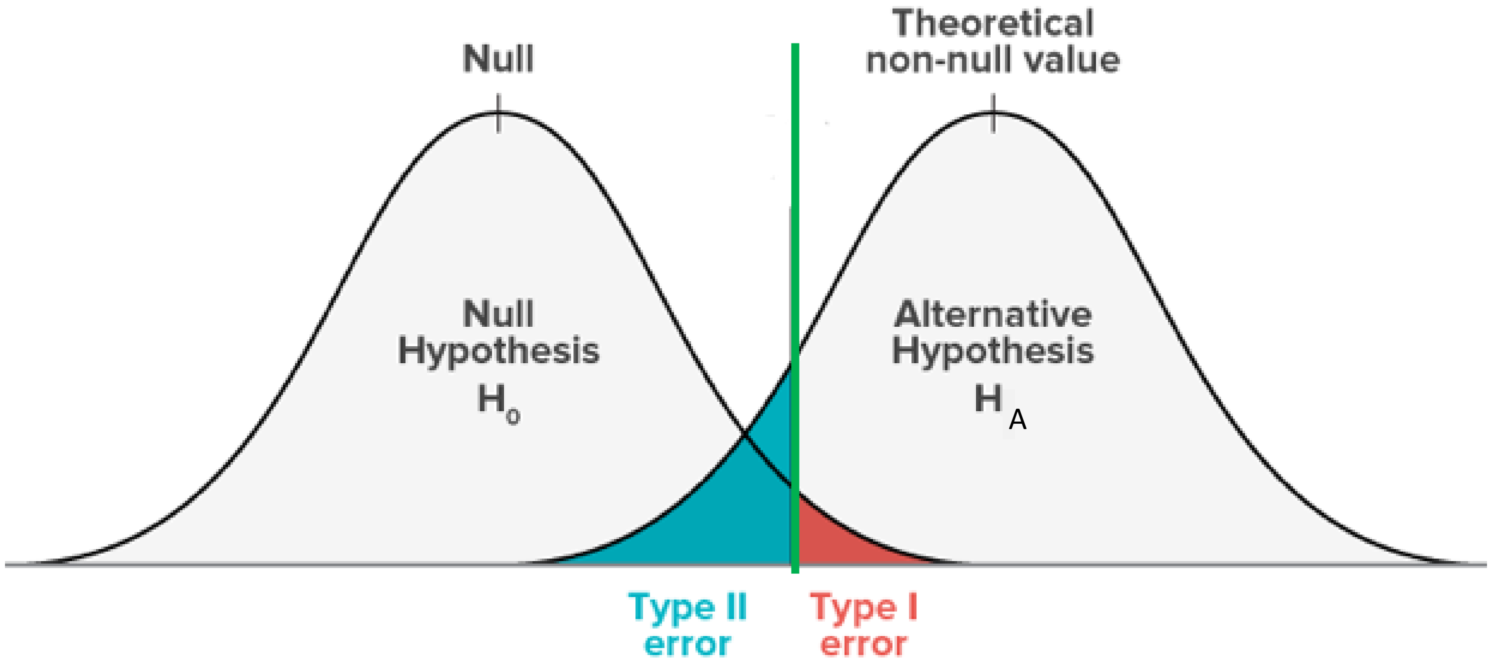
Neyman-Pearson Frequentist logic



The **power** of a test is the probability we reject the H_0 when it is **false**

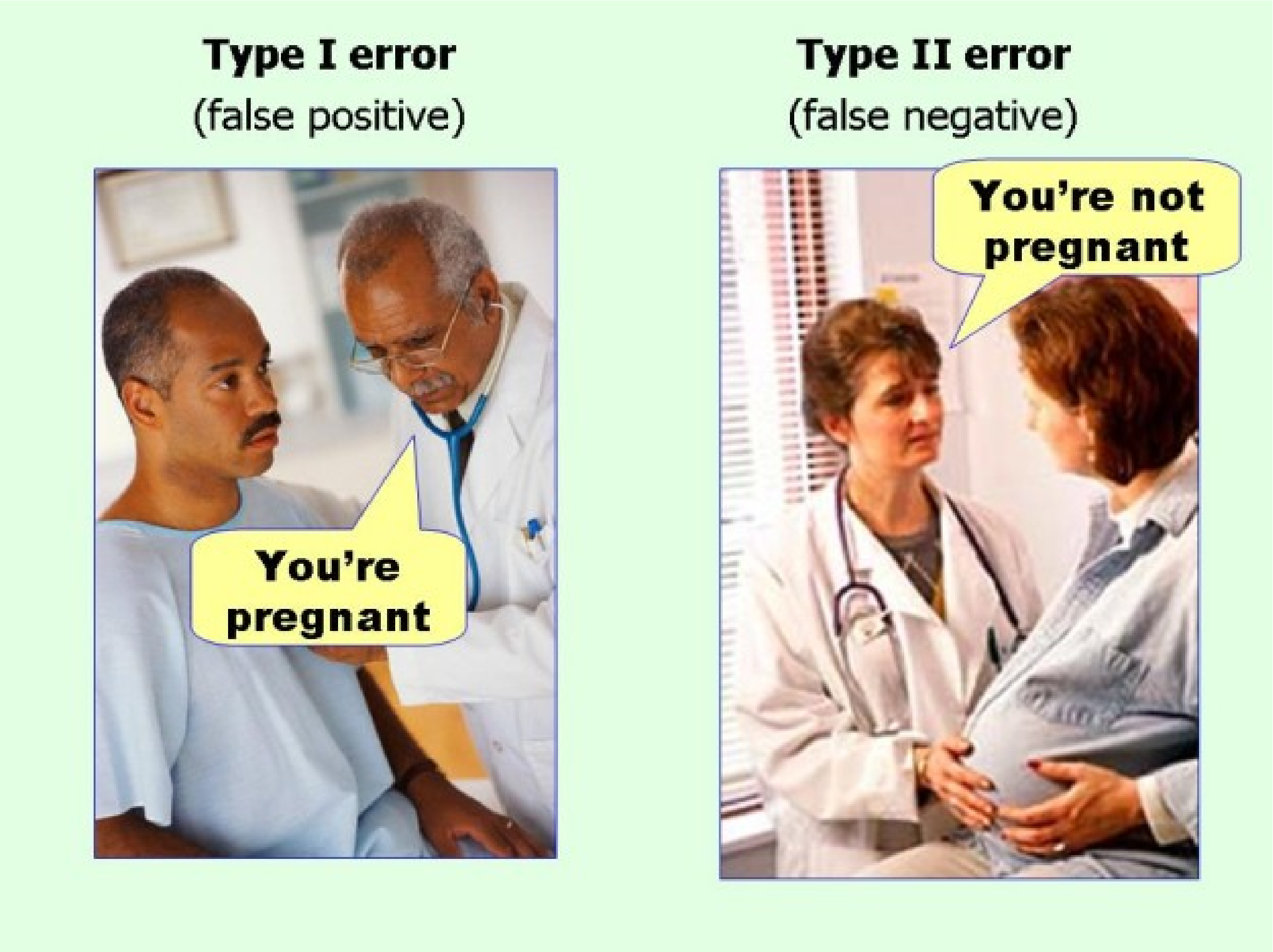
- 1 - β
- For a fixed α level, it would be best to use the most powerful test

Type I and Type II Errors



	Reject H_0	Do not reject H_0
H_0 is true	Type I error (α) (false positive)	No error

Type I and Type II Errors



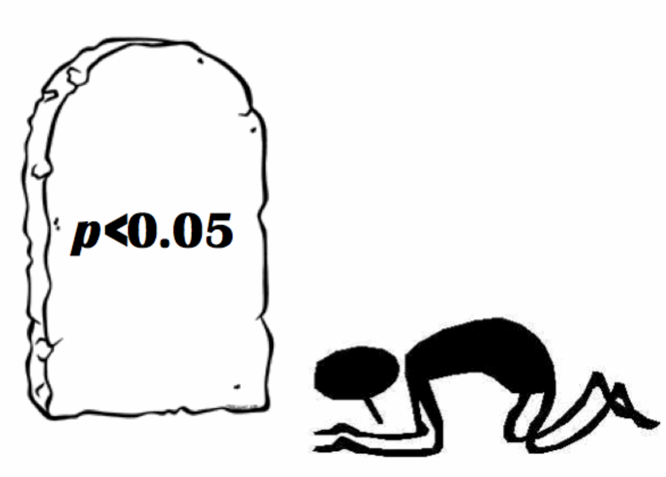
Problems with the NP hypothesis tests

Problem 1: we are interested in the results of a specific experiment, not whether we are right most of the time

- E.g., 95% of these statements are true:
  - Calcium is good for your heart, Paul is psychic, Buzz and Doris can communicate, …

Problem 2: Arbitrary thresholds for alpha levels

- P-value = 0.051, we don't reject H_0



Collectively Unconscious

News from the Frontiers of Science

ABOUT

NOVEMBER 3, 2012

New version SPSS will include 'celebratory fireworks' for significant results

An official press release has confirmed that the newest release of SPSS will be equipped with 'performance-rewarding features'. The new installment of the popular data-analysis package will light up with song, dance and fireworks whenever a statistical test is significant. 'We want to provide a package that is in line with the day-to-day experiences of researchers. We understand the pressure the publish, and the relief that is felt by many when those Stars of Significance appear in the results table. '

The level of significance will determine the abundance of the celebrations. If the *p*-value is below 0.05, researchers will automatically hear what is described as 'a cheerful tone', according to a company spokesman. "But if your *p*-value is below 0.01, the software package will play a series of congratulatory videos, complimenting your

SUBTITLE

RECENT POSTS

- Scientists may have 'sixth sense' for poor PSI research
- Matrix dimensions reach agreement at peace summit
- Controversial trial will provide free polymerase to junk DNA
- Animal rights activists outraged by infinite monkey experiment
- Scientists receive 12.6 million dollar grant to format references correctly

ARCHIVES

Problems with the NP hypothesis tests

<u>Problem 1</u>: we are interested in the results of a specific experiment, not whether we are right most of the time

- E.g., 95% of these statements are true:
  - Calcium is good for your heart, Paul is psychic, Buzz and Doris can communicate, …

<u>Problem 2</u>: Arbitrary thresholds for alpha levels

- P-value = 0.051, we don't reject H_0?

<u>Problem 3</u>: running many tests can give rise to a high number of type 1 errors

Genes and leukemia example

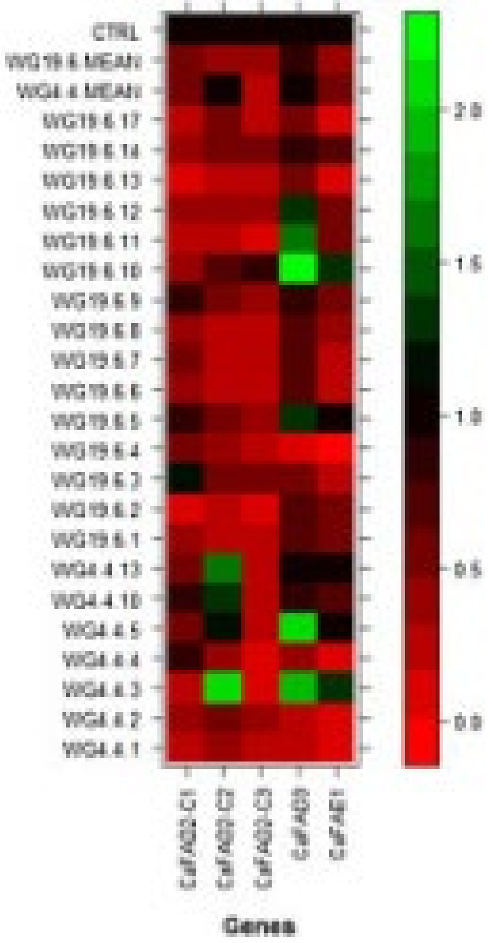
Scientists collected 7129 gene expression levels from 38 patients to find genetic differences between two types leukemia (L1 and L2)

Suppose there was no genetic differences between the types of leukemia

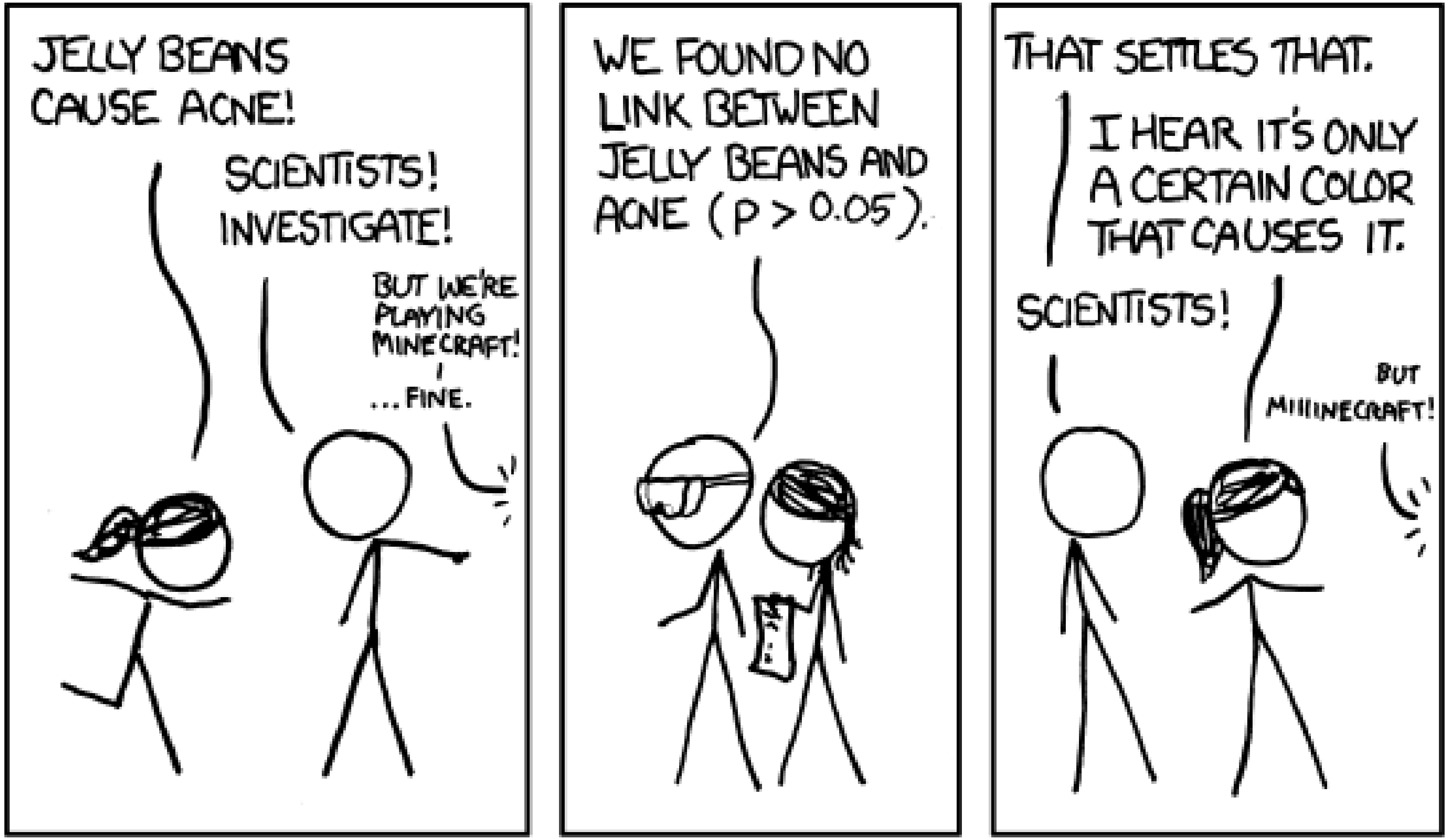
- H_0: $\mu_{L1} = \mu_{L2}$ is true for all genes

Q: If each gene was tested separately using a significance level of $\alpha = 0.05$, approximately how many type 1 errors would be expected?

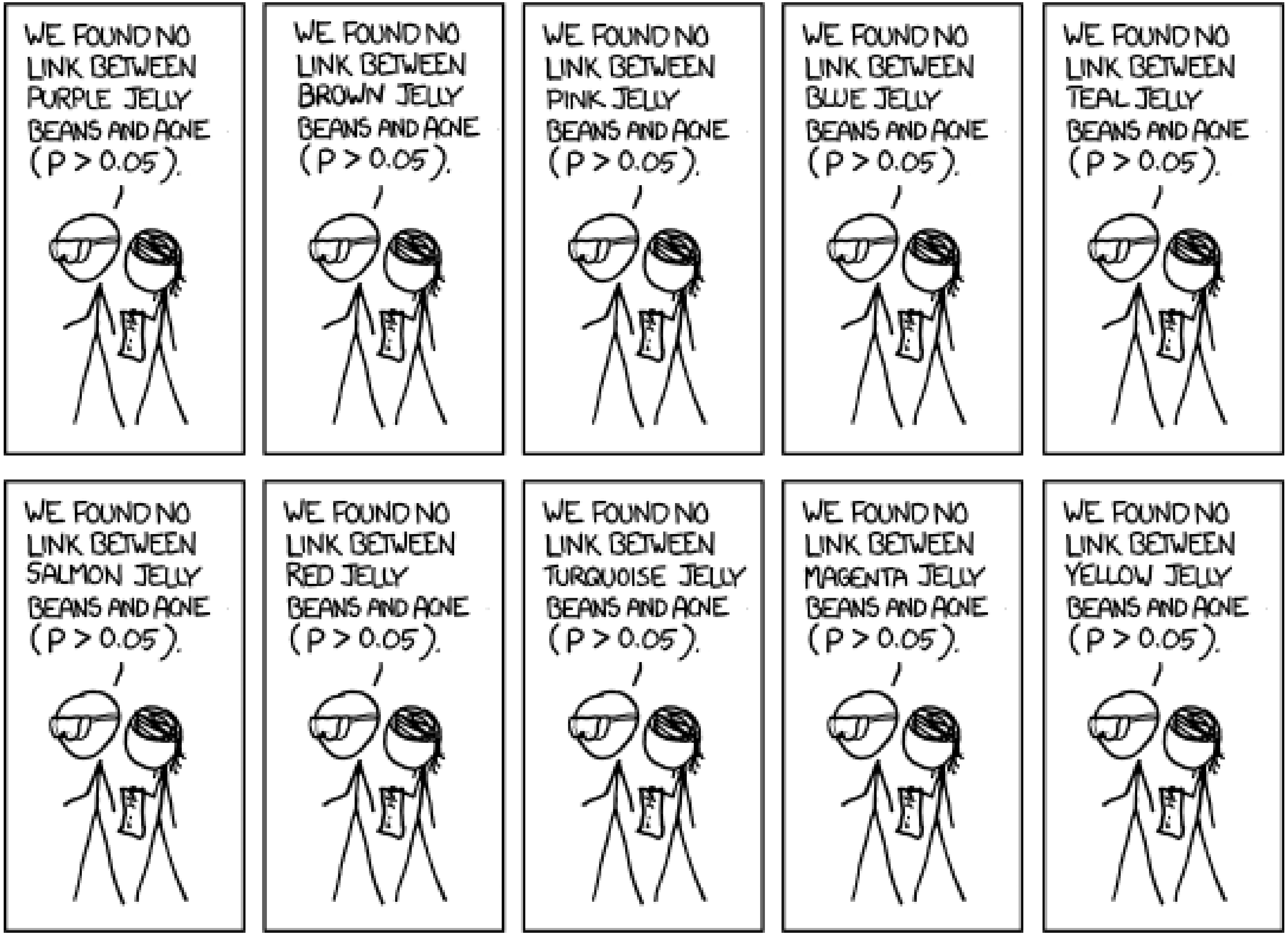
- A: 7129 x 0.05 = 356

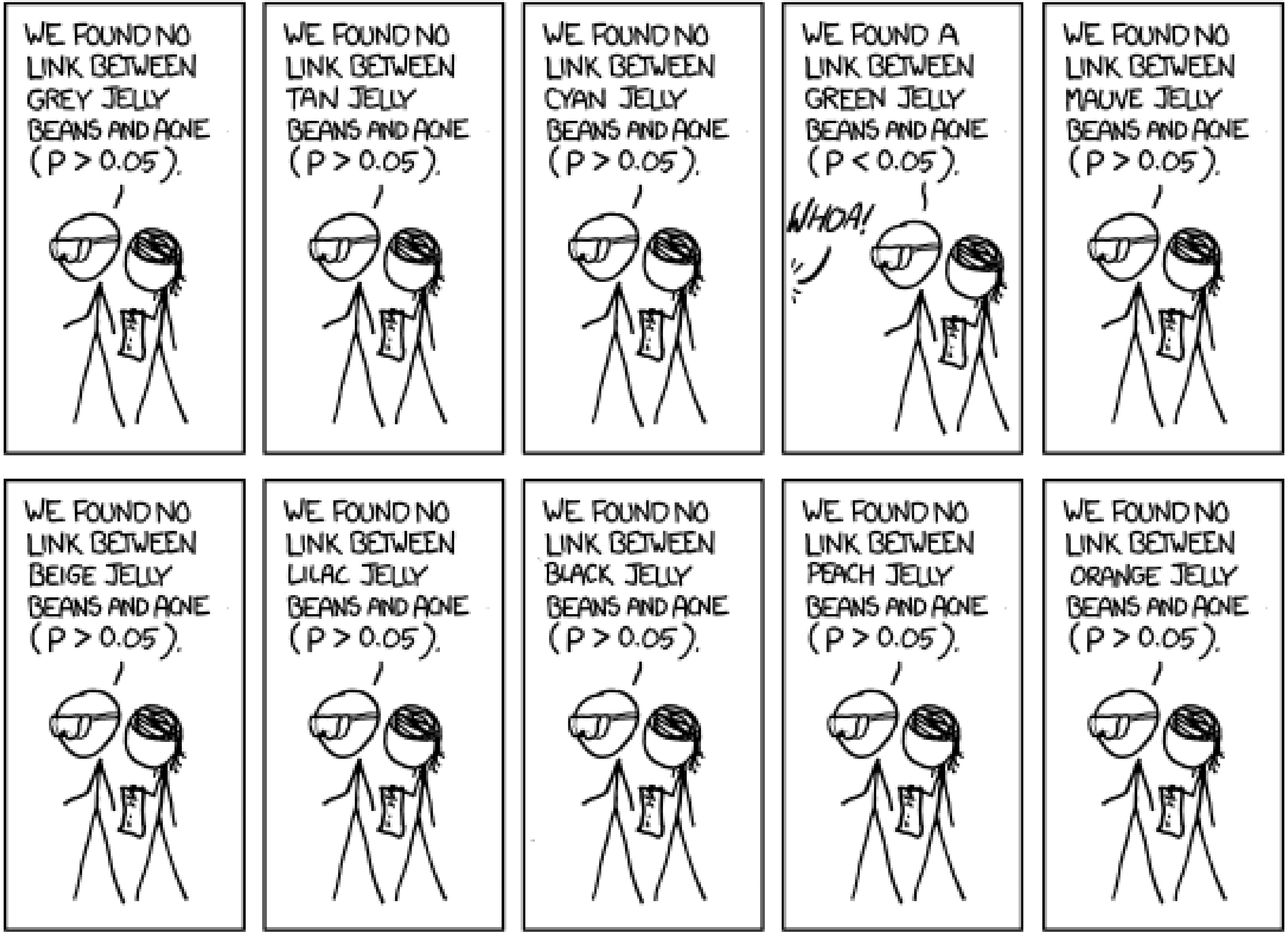


Multiple hypothesis tests



http://xkcd.com/882/


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).


WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).

Don’t ever do this!

http://xkcd.com/882/

Genes and leukemia example

There are methods that try to correct for running multiple hypothesis tests

The ***Bonferroni correction*** is one way that controls the probability of ***any*** hypothesis test giving a type 1 error

- i.e., controls the familywise error rate (no type 1 errors for any of the tests run)

It works by dividing the initial α level by the number of tests run

- E.g., $\alpha = 0.05/7129 = 0.000007$
- All p-values need to be below this level to be considered statistically significant
- This can lead to many type 2 errors (Type 2 error: failure to reject H_0 when it is false)

The problem of multiple testing

For $\alpha = 0.05$, ~5% of all published research findings should be wrong

Publication bias (file drawer effect): Generally positive results are more likely to be published, so if you read the literature, the number of incorrect results (type 1 errors) will be greater than 5%.

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

The Earth Is Round (p < .05)

Jacob Cohen

After 4 decades of severe criticism, the ritual of null hypothesis significance testing—mechanical dichotomous decisions around a sacred .05 criterion—still persists. This article reviews the problems with this practice, including

sure how to test H_0, chi-square with Yates's (1951) correction or the Fisher exact test, and wonders whether he has enough power. Would you believe it? And would you believe that if he tried to publish this result without a

[American Statistical Association's Statement on p-values](#)

Some thoughts…

Better to have hypothesis tests than none at all. Just need to think carefully and use your judgment.

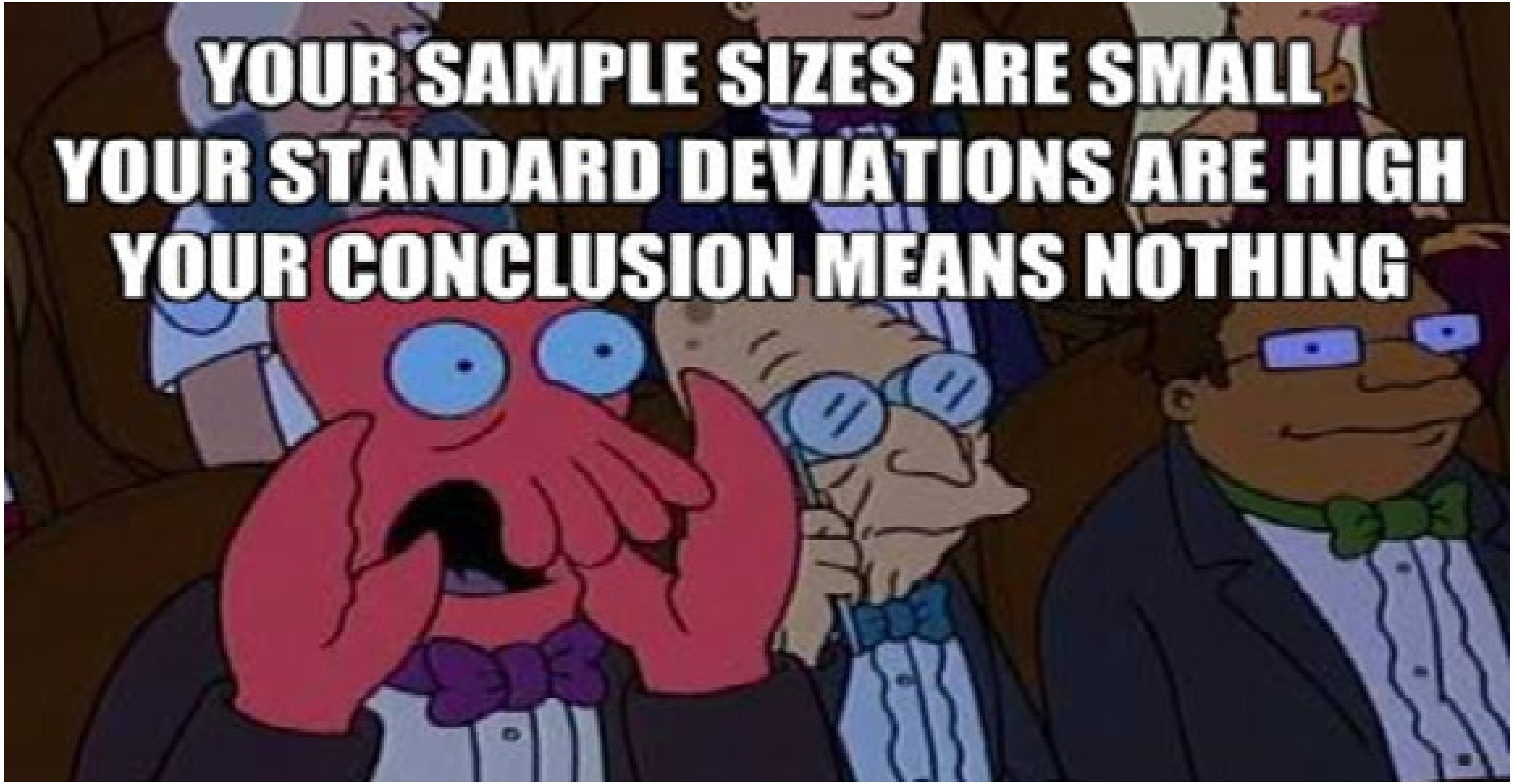
Report effect size in most cases – i.e., confidence intervals

Report the p-values rather than accept/reject H_0

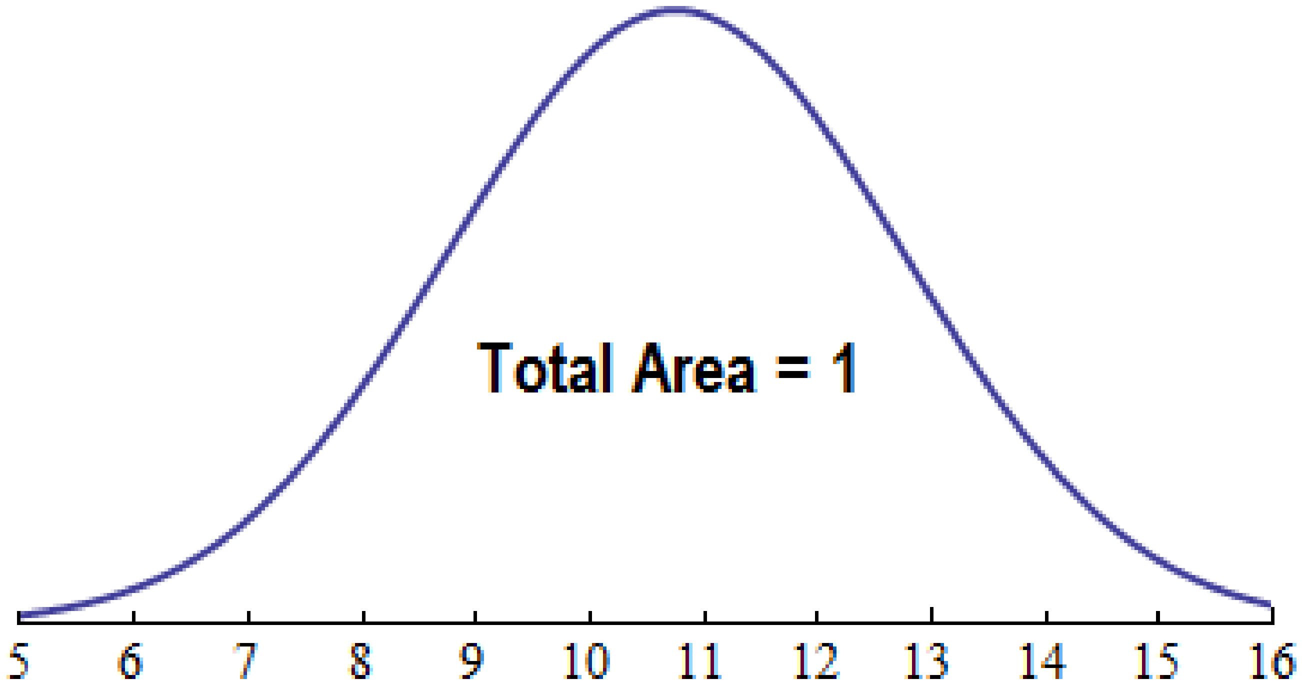
- i.e., report $p = 0.23$ not $p < 0.05$

Replicate findings (perhaps in different contexts) to make sure you get the same results

Be a good/honest scientists and try to get at the Truth!


YOUR SAMPLE SIZES ARE SMALL
YOUR STANDARD DEVIATIONS ARE HIGH
YOUR CONCLUSION MEANS NOTHING

Inference using parametric probability distributions



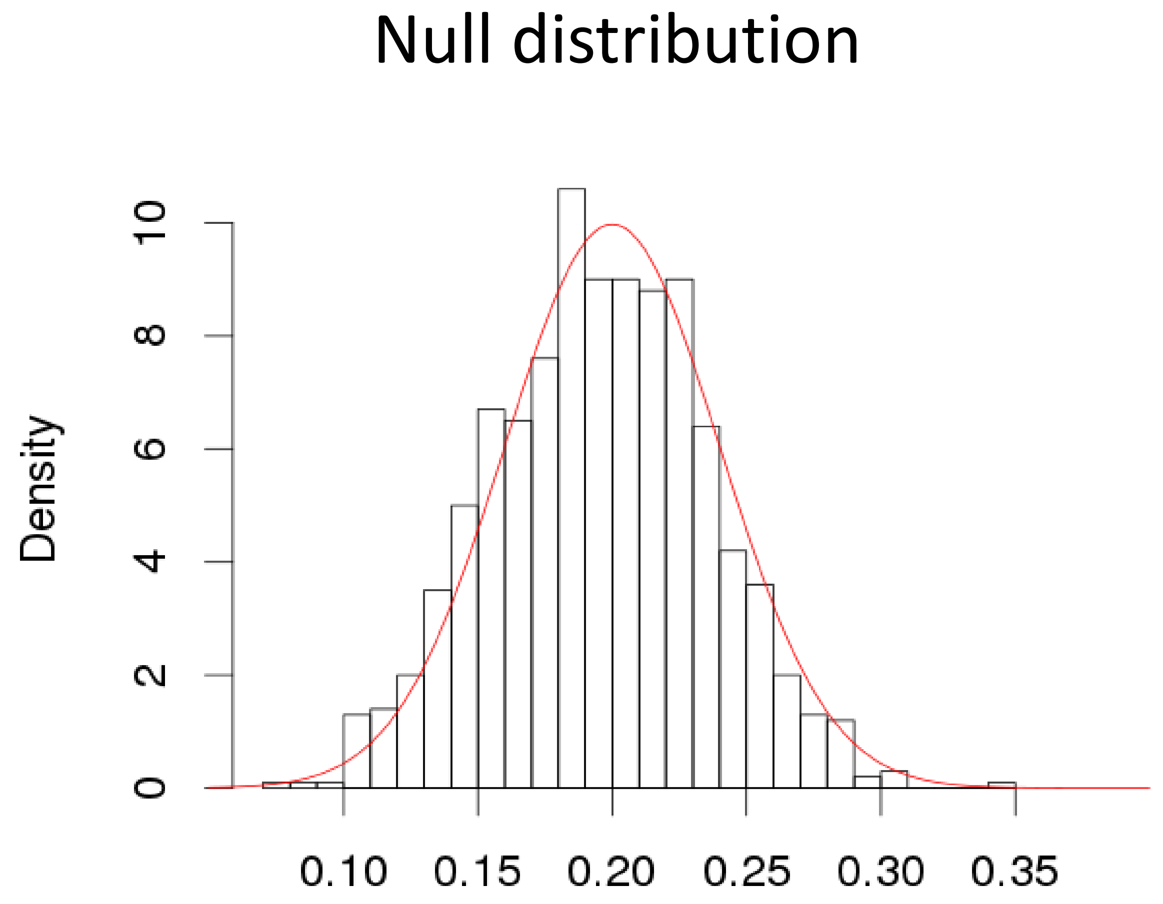
Inference using parametric probability distributions

In the past month we have learned how to use computer simulations to create confidence intervals and run hypothesis tests

Now we will use mathematical functions called **probability distributions** to do inference

- e.g. instead of running computer simulations to create null distributions we can just mathematical probability distributions

Comparing simulation based distribution and a probability distribution

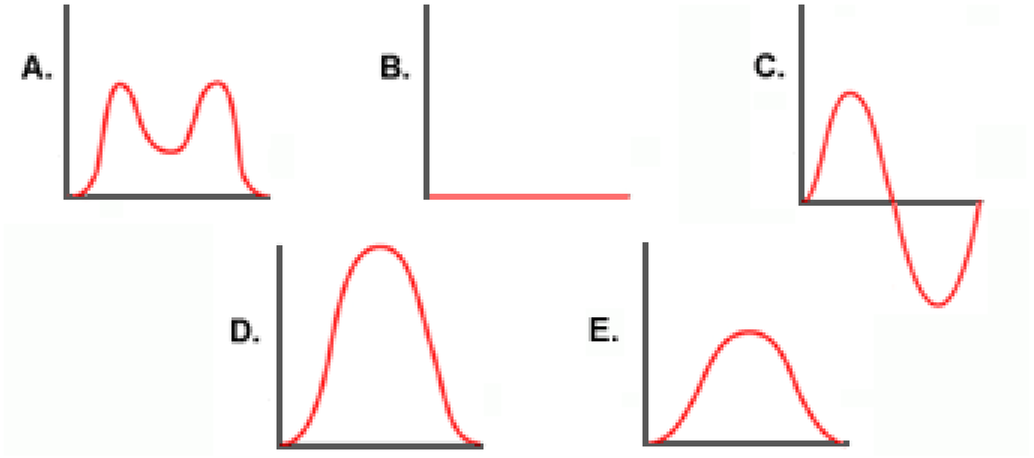


Density Curves

A **density curve** is a mathematical function f(x) that has two important properties:

1. The total area under the curve f(x) is equal to 1
2. The curve is always ≥ 0

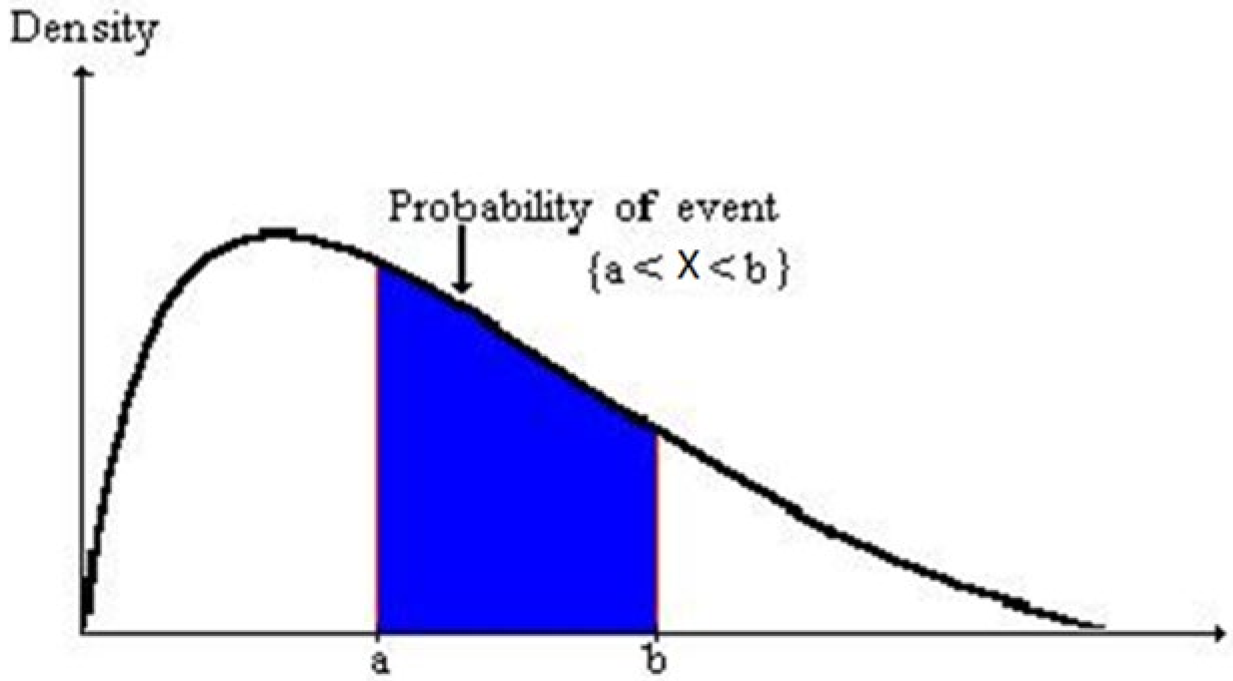
Which of these could ***<u>not</u>*** be a density curve?



Density Curves

The <u>area under the curve </u>in an interval [a, b] models the probability that a random number X will be in the interval

Pr(a < X < b) is the area under the curve from a to b



Example: Normal Density Curve

Normal distributions are a family of bell-shaped curves

There are two parameters that characterize normal curves, which are:

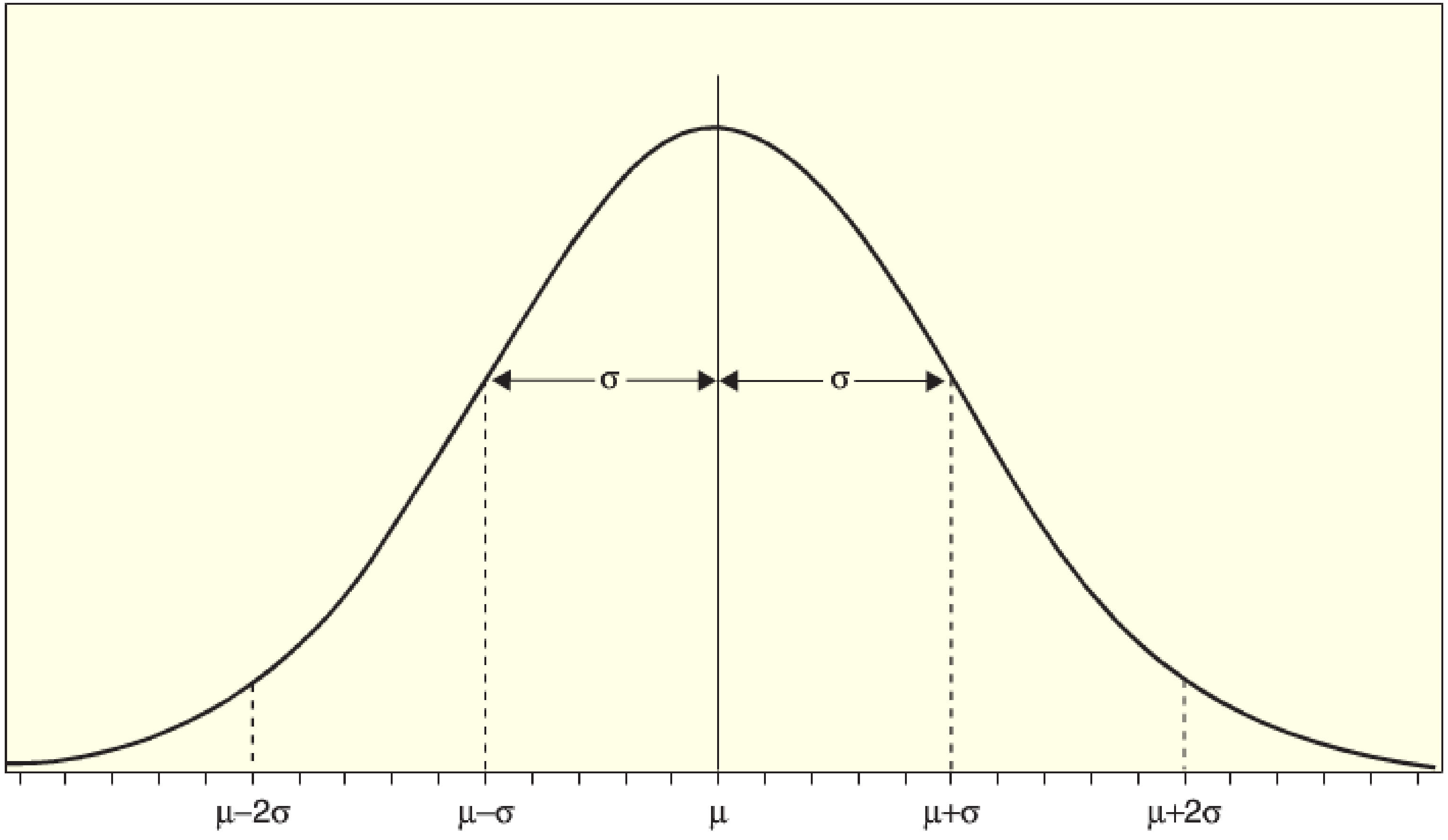
- The mean: μ
- The standard deviation: σ

We use the symbols μ and σ because this is often a model for the population

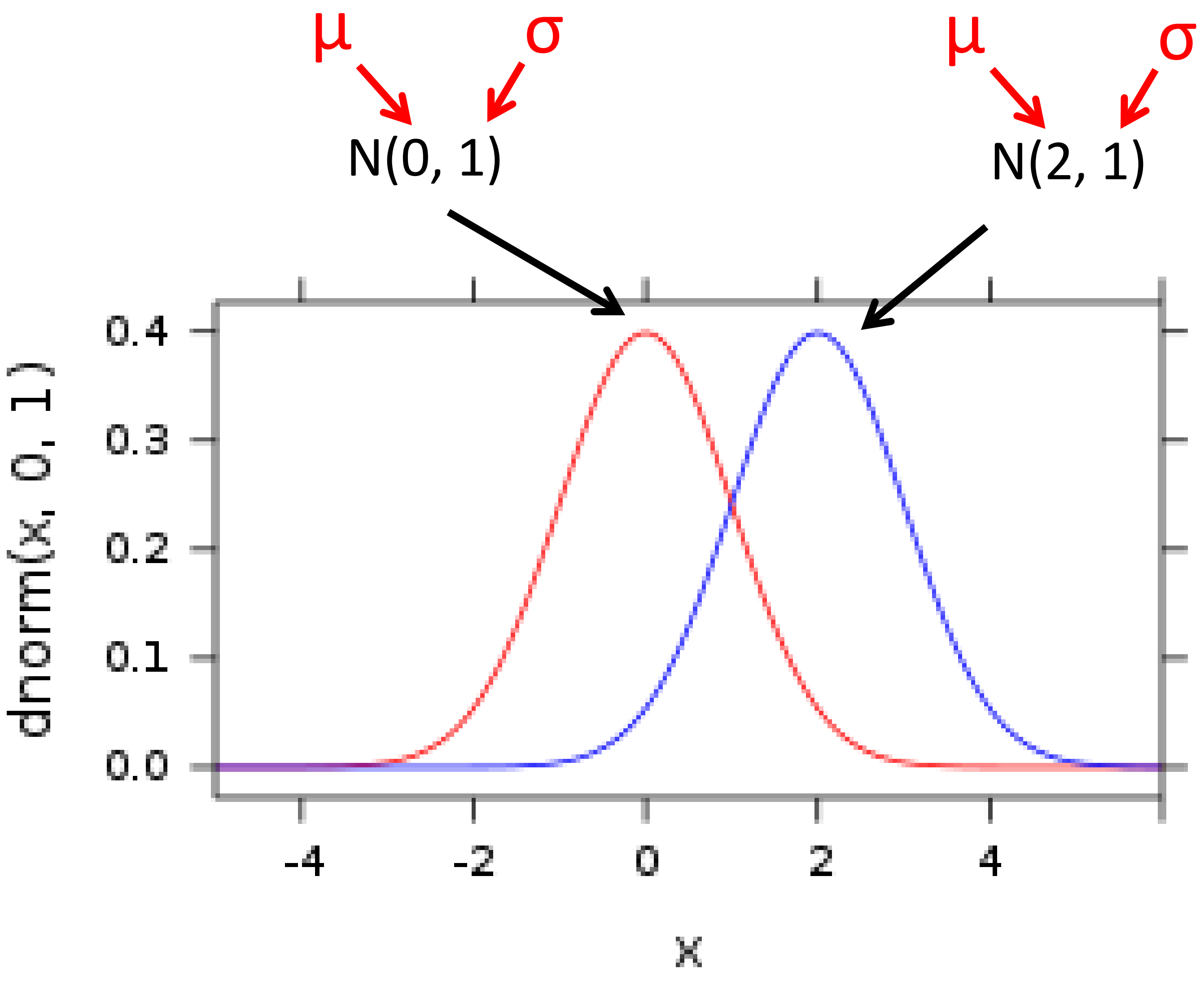
Notation: X ~ N(μ, σ)

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

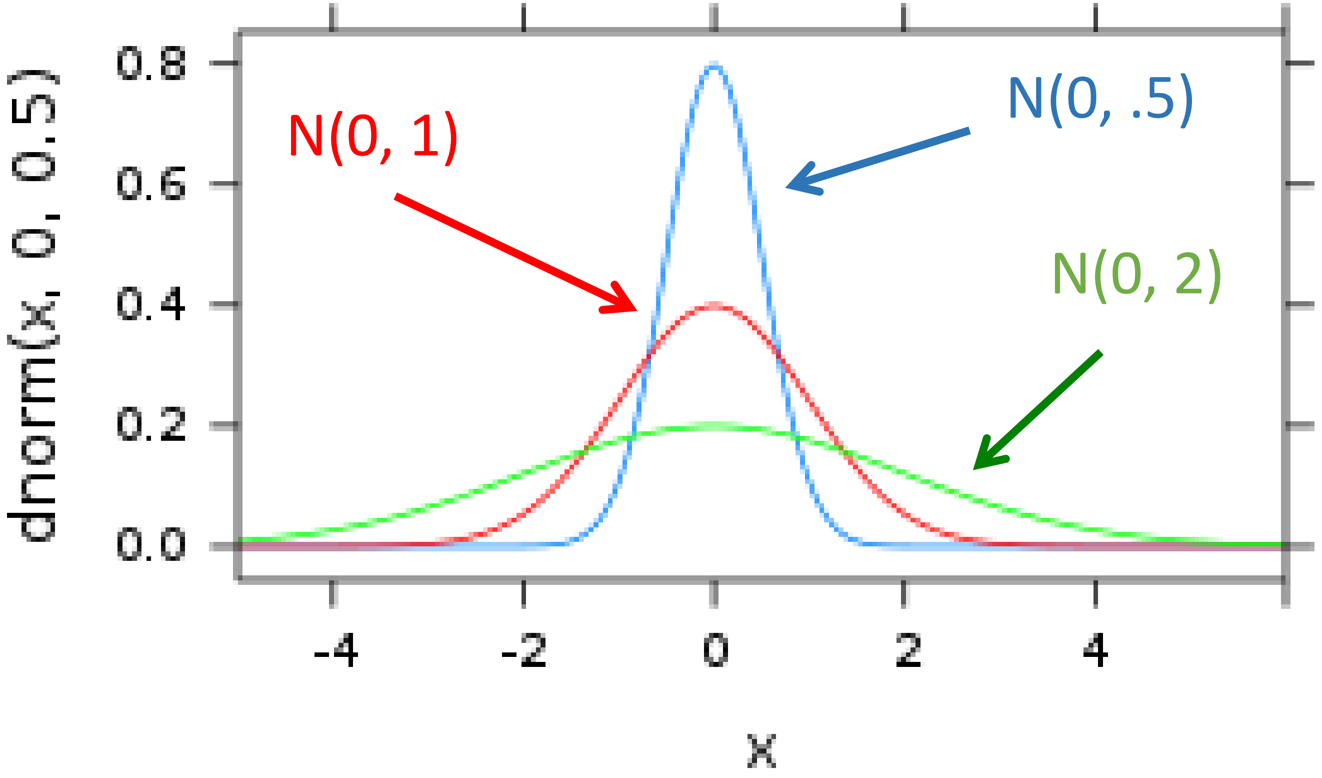
Graph of a Normal Density Curve



Normal curves with different means



Normal curves with different variances



Graphing Normal Curves

IQ scores are defined to have a mean of 100 and a standard deviation of 15

A) Draw this distribution by hand

B) Check your results in R

```
x <- 40:150
y <- dnorm(x, mu, sigma)
plot(x, y, type = "l")
```

Finding normal probabilities and percentiles

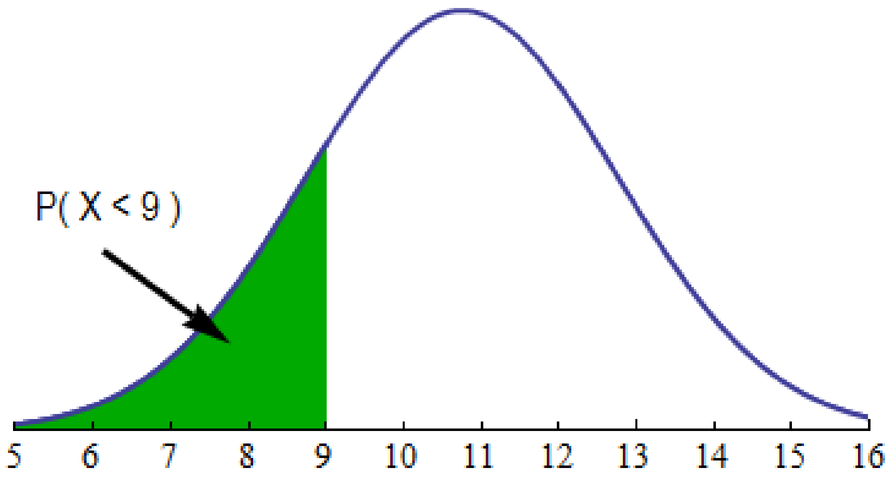
No simple formula exists for computing the areas under a normal curve

We can use R to find such areas by specifying:

- Mean and standard deviation
- The endpoints of the interval

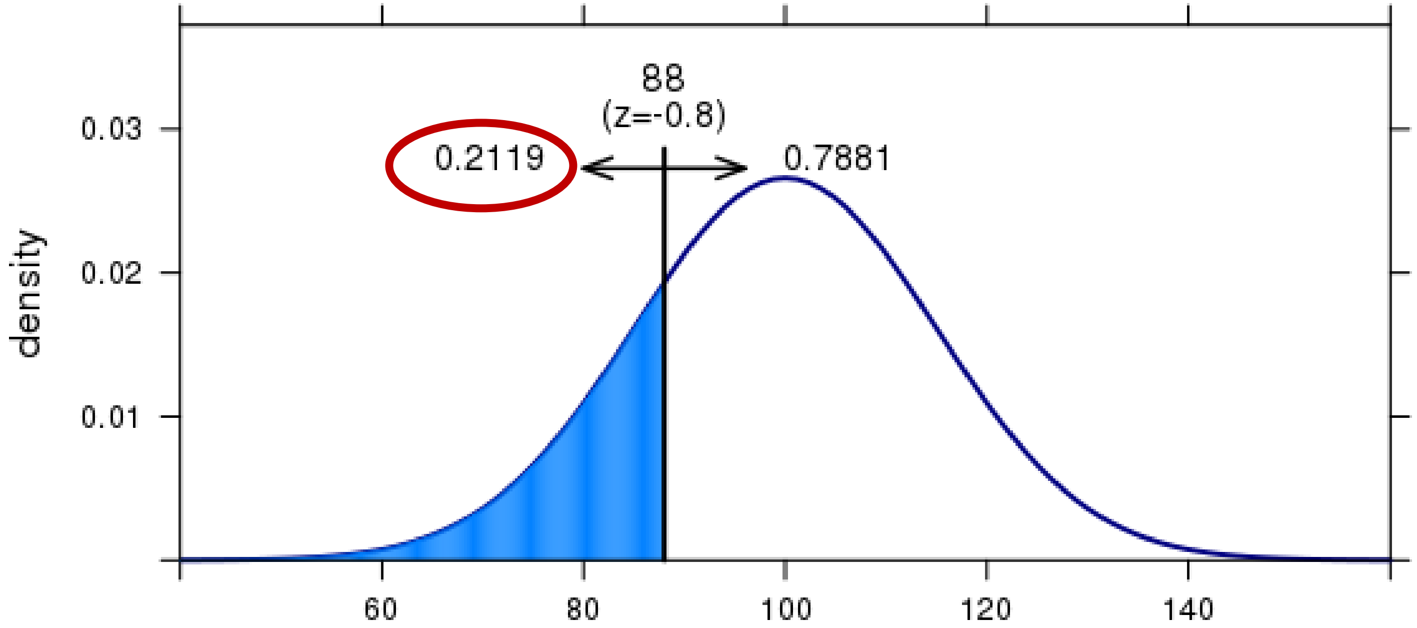
\# probability of getting a value ≤ x

```
pnorm(x, mu, sigma)
```



Calculate the probability a random person you meet has an IQ less than 88

```
pnorm(88, 100, 15)
```



[Normal area Pr(X ≤ x) app]   [Normal area Pr(a < X < b) app]

Probability practice questions

1. What is probability a randomly chosen person will have an IQ greater than 96?

2. What is the probability a randomly chosen person will have an IQ between 88 and 96?

Try this on your own, and we will discuss the answers next class…