Hypothesis Testing

February 25, 2020

Data Science CSCI 1951A

Brown University

Instructor: Ellie Pavlick

HTAs: Josh Levin, Diane Mutako, Sol Zitter

Announcements

- ....?

Today

- **Two quick preliminaries: LoLN and CLT**
- Follow up from last time
- Common tests: chi-squared test, z-test/t-tests

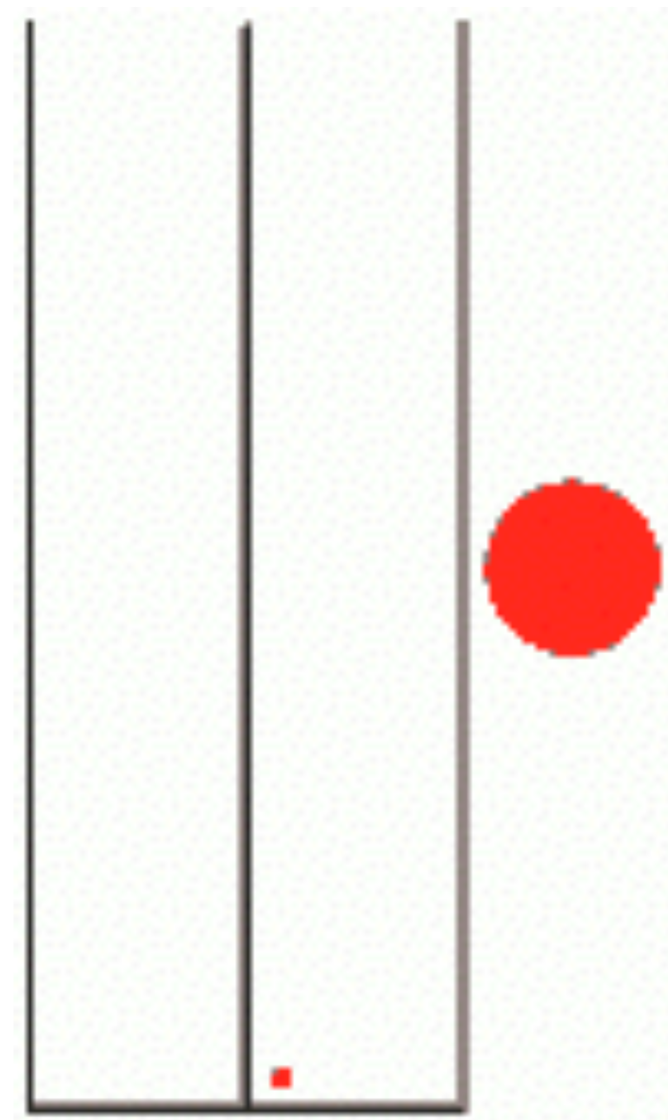
Law of Large Numbers

- If you perform the same experiment a large number times, the *average* will converge to the expected value

- Assumes that errors are "random" and uncorrelated, so will balance out over time

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

$$\bar{X}_n \to \mu \text{ as } n \to \infty$$

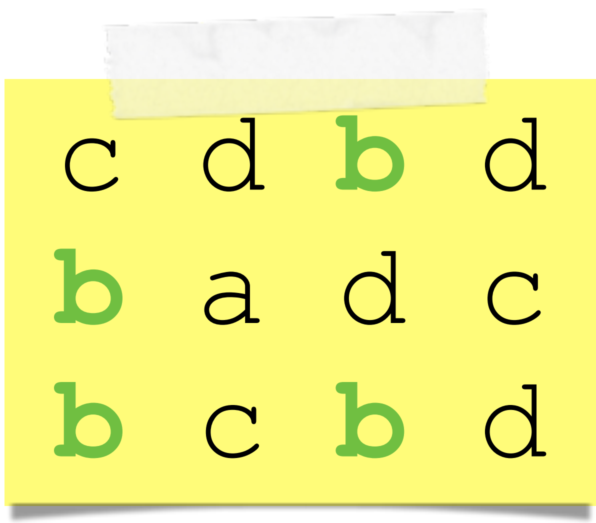


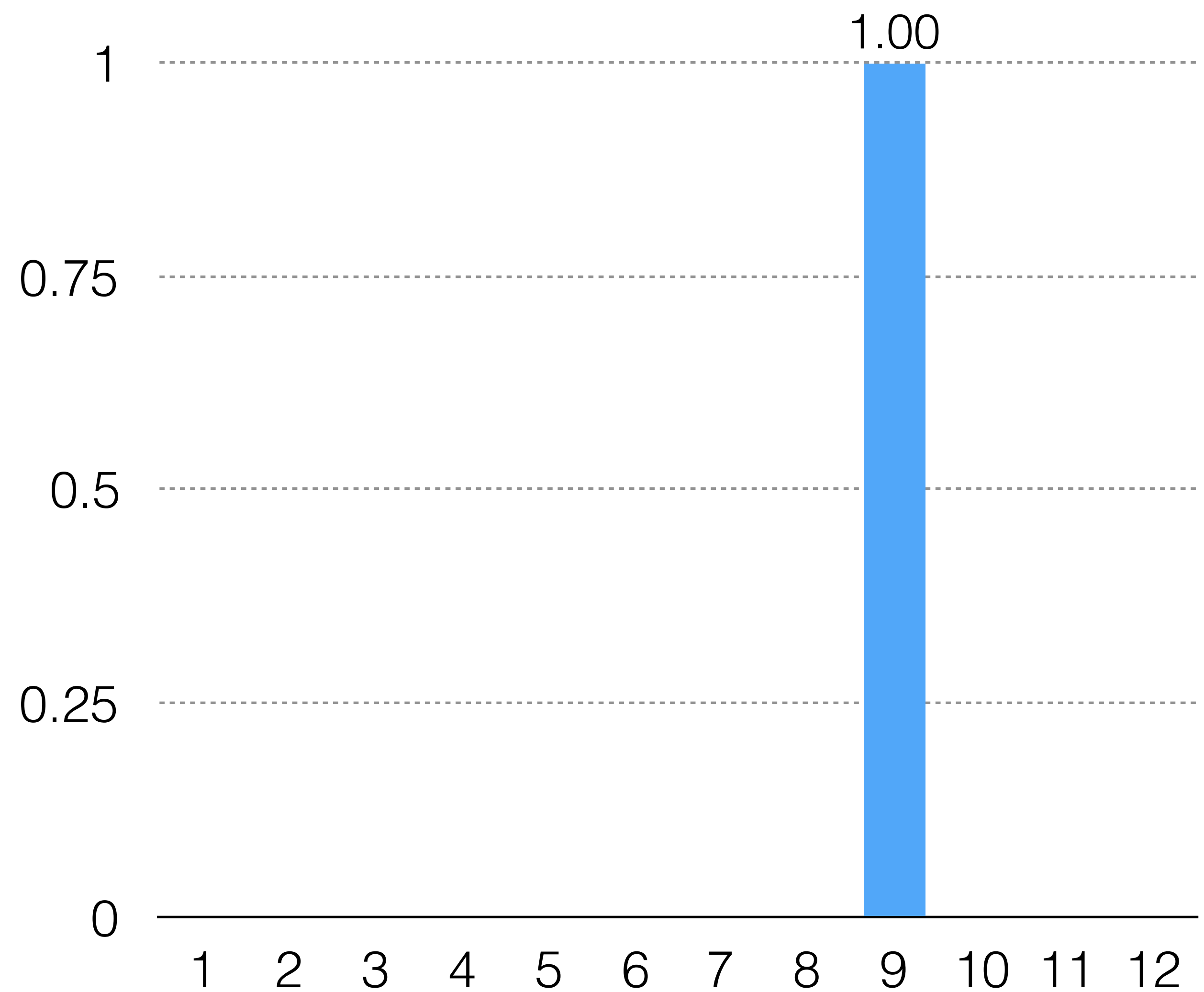
https://en.wikipedia.org/wiki/Law_of_large_numbers

Central Limit Theorem

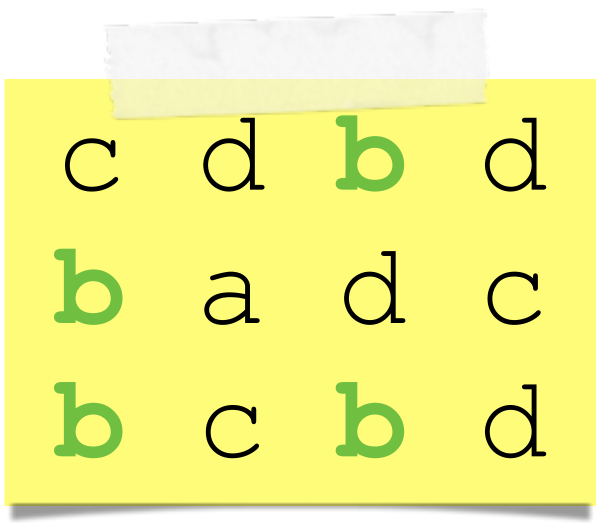
- Given $X_1 \ldots X_n$
- Not only does a $\bar{X}_n \rightarrow \mu \text{ as } n \rightarrow \infty$
- But the distribution approaches a normal distribution

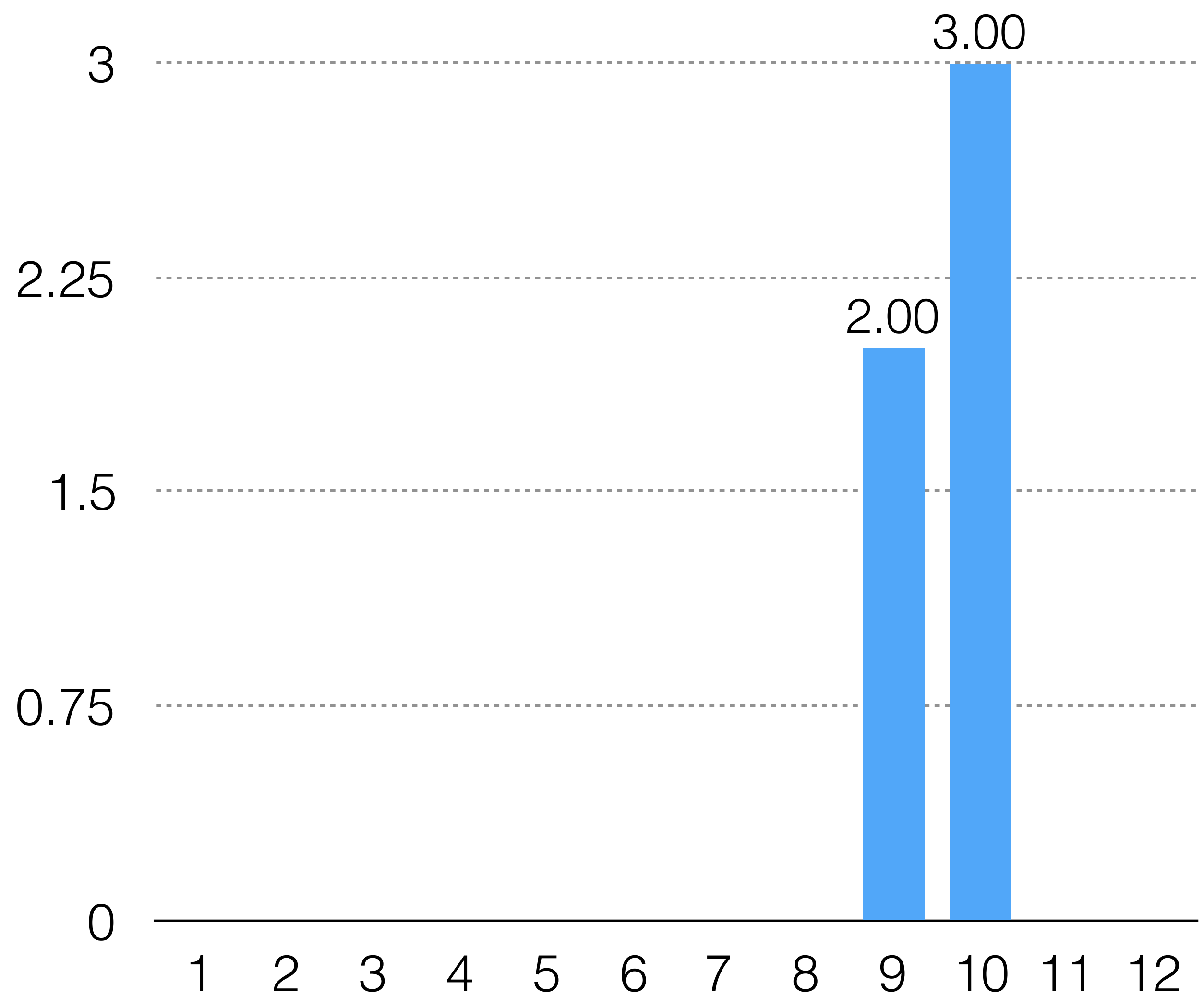
Central Limit Theorem



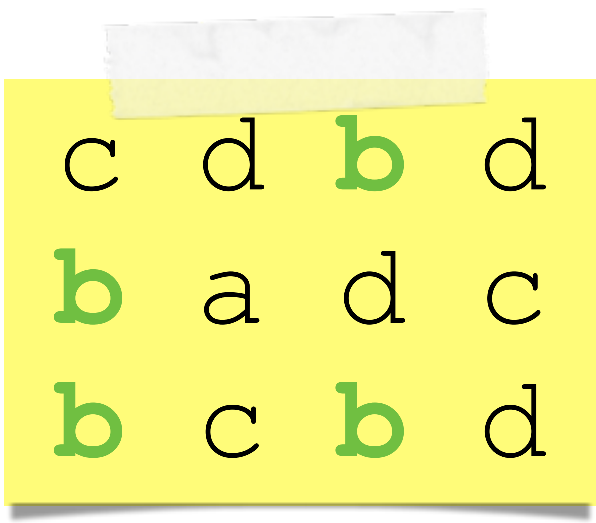


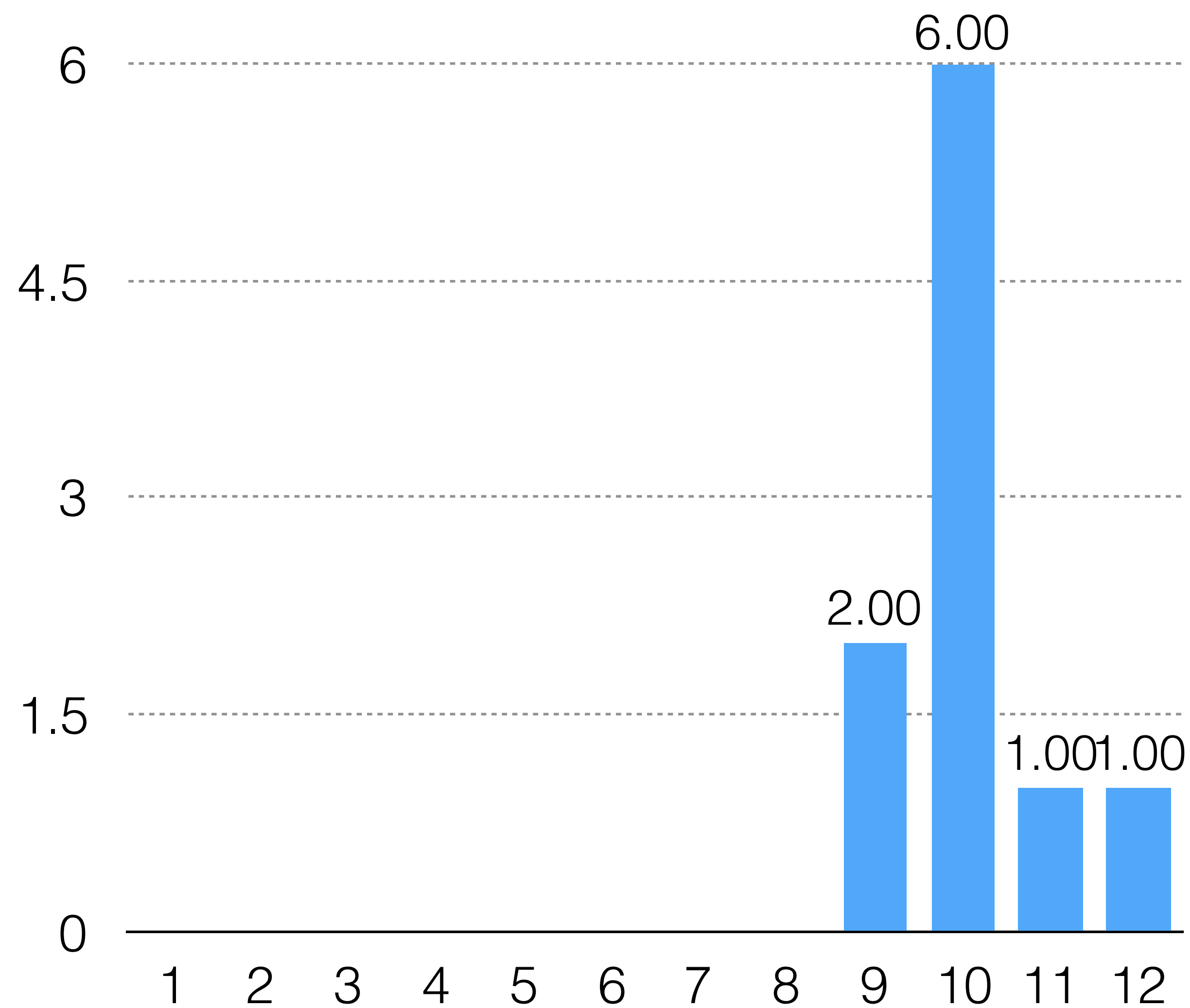
Central Limit Theorem



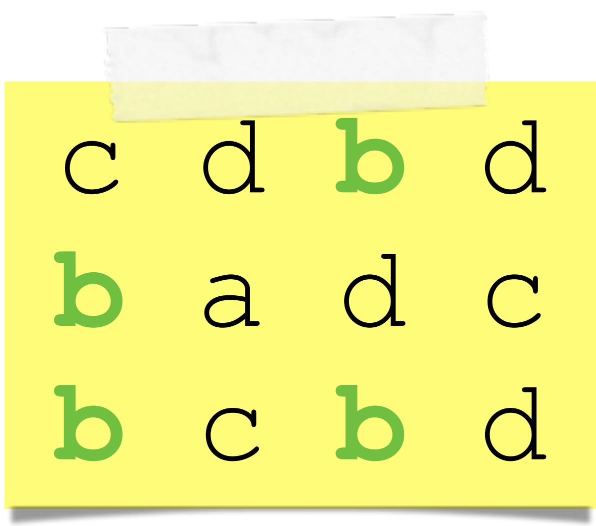


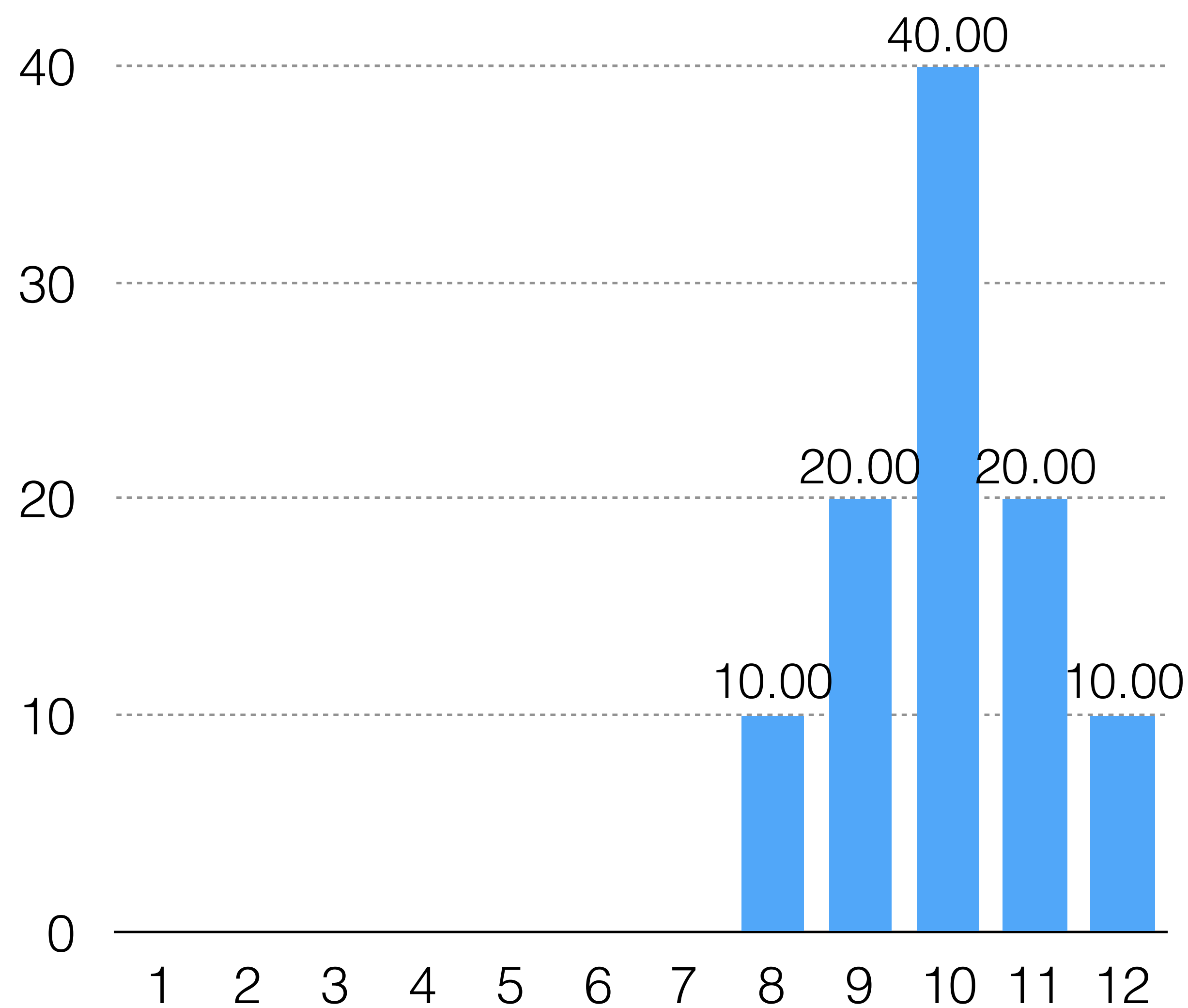
Central Limit Theorem





Central Limit Theorem

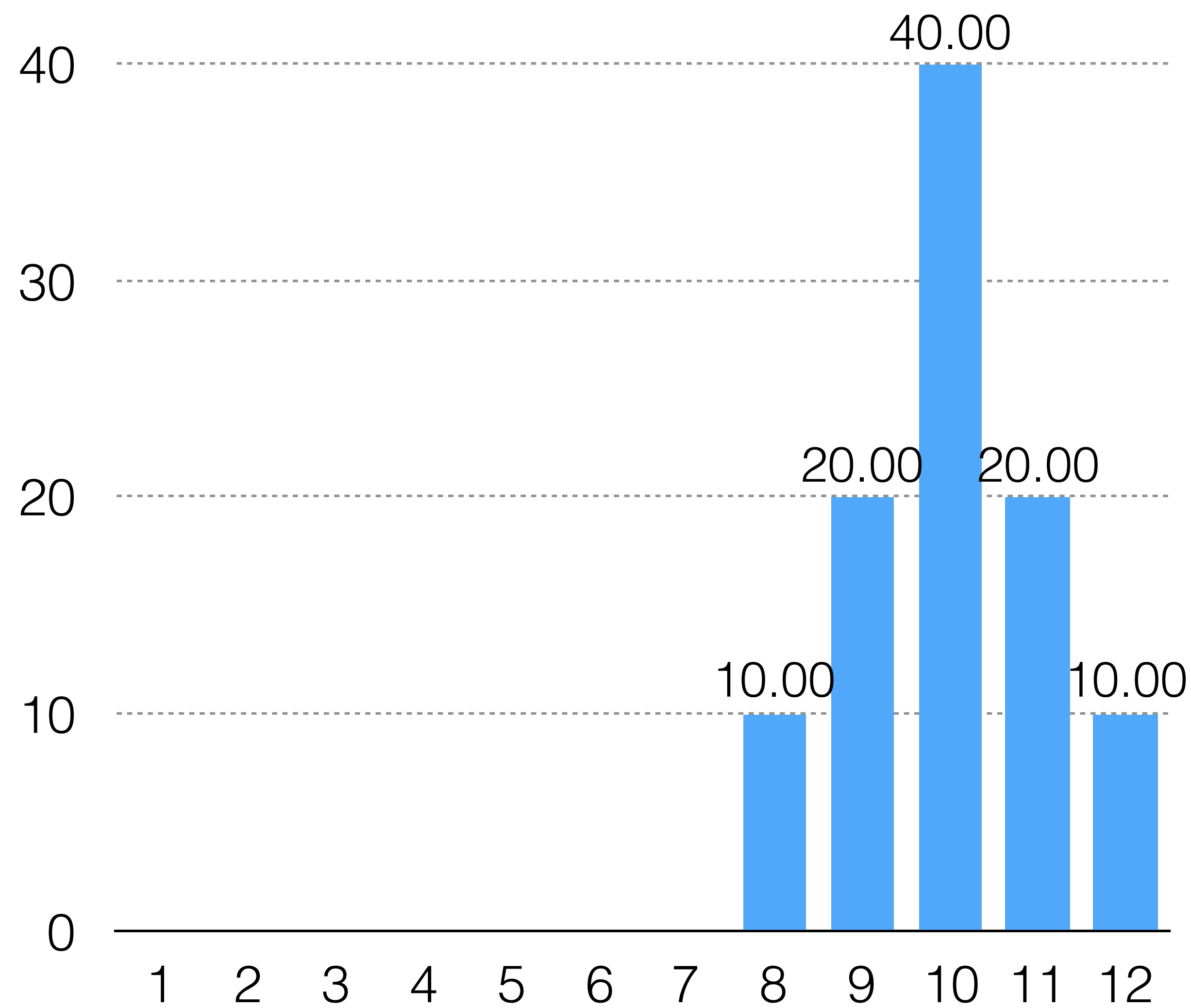




Central Limit Theorem

c	d	**b**	d
b	a	d	c
b	c	**b**	d

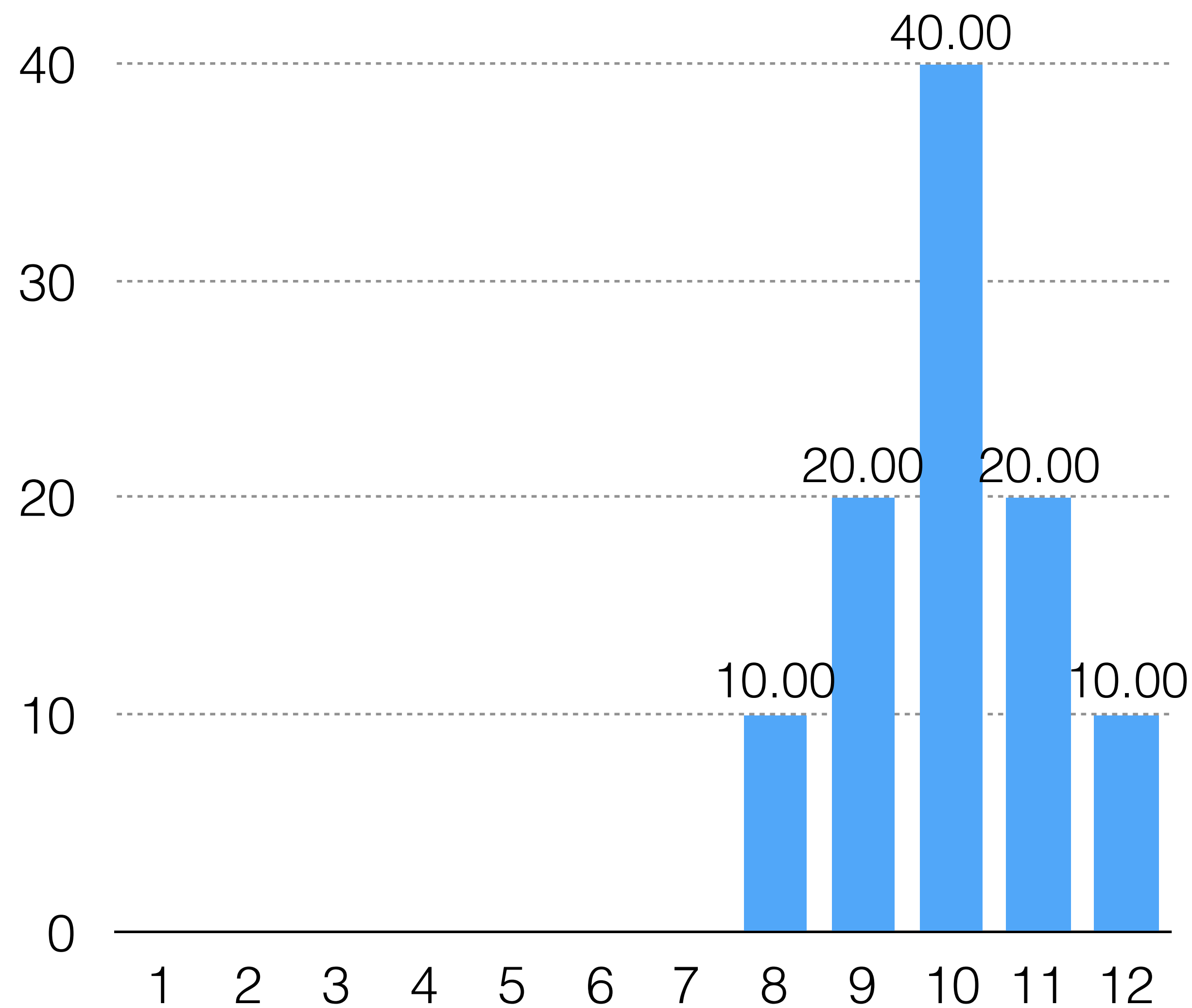
I.e. test statistics are often normally distributed...



Central Limit Theorem

c	d	**b**	d
b	a	d	c
b	c	**b**	d

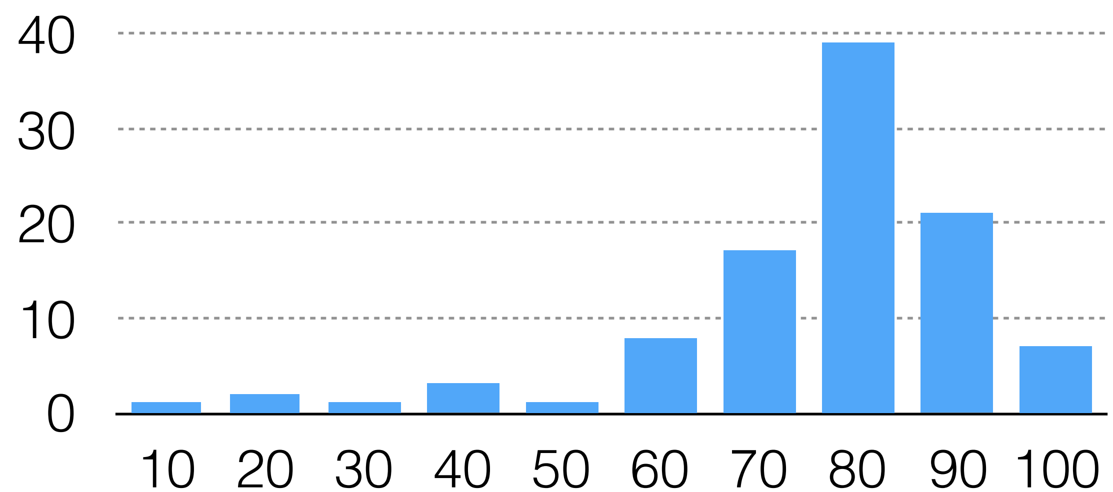
Can apply statistical methods designed for normal distributions even when underlying distribution is not normal



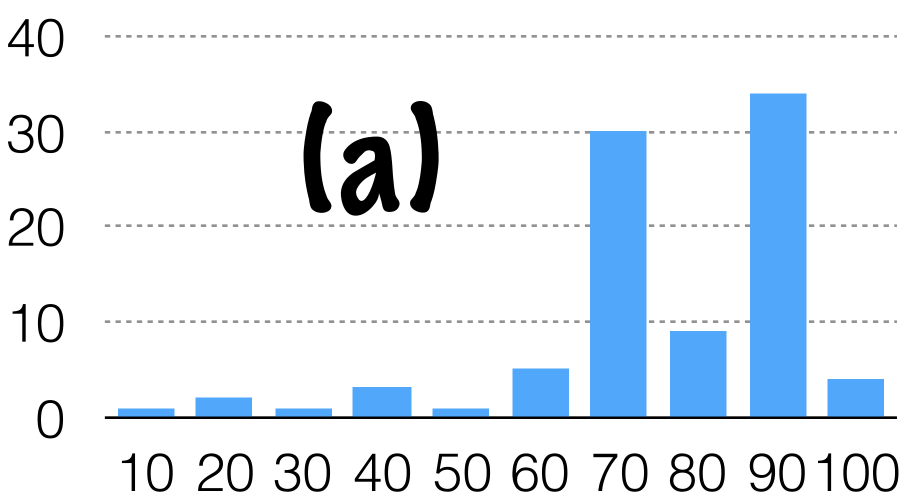
Clicker Question!

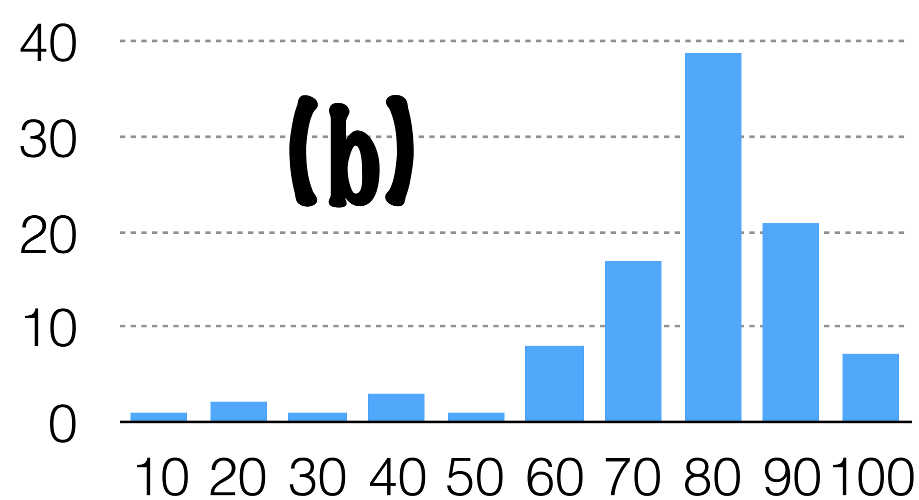
Clicker Question!

Every year, I compute the mean grade in my class. I never change the material or my methods for evaluating because, lazy. Over the 439 years that I have been teaching this class, this has resulted in the below distribution.



Which of these is mostly like the typical distribution on any given year?

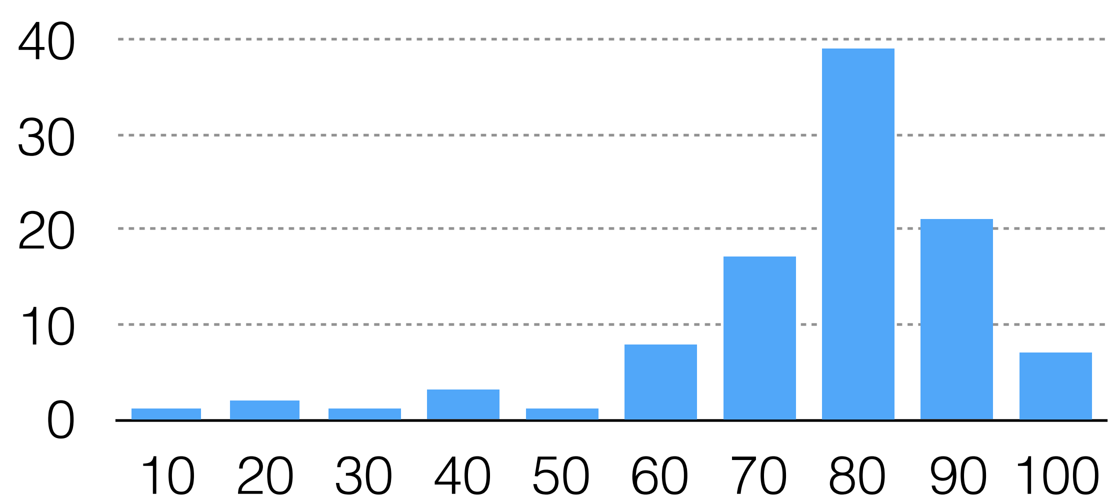




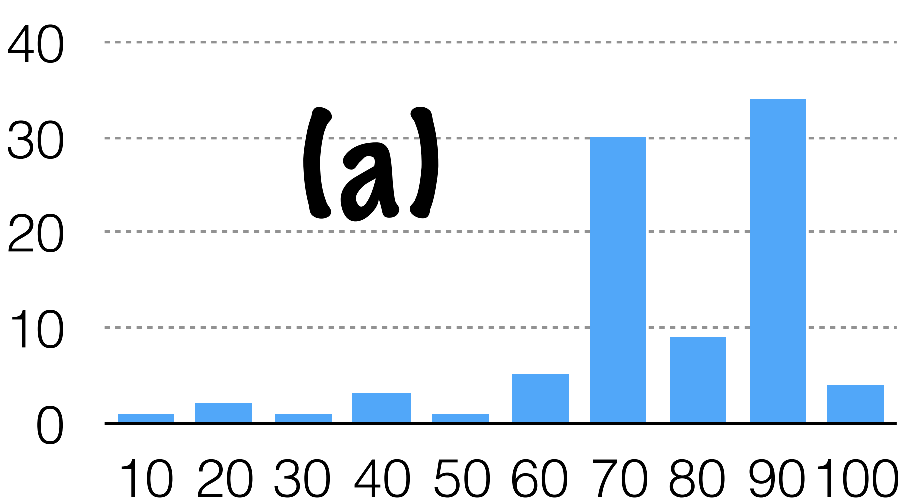
(c) can't say, could be either

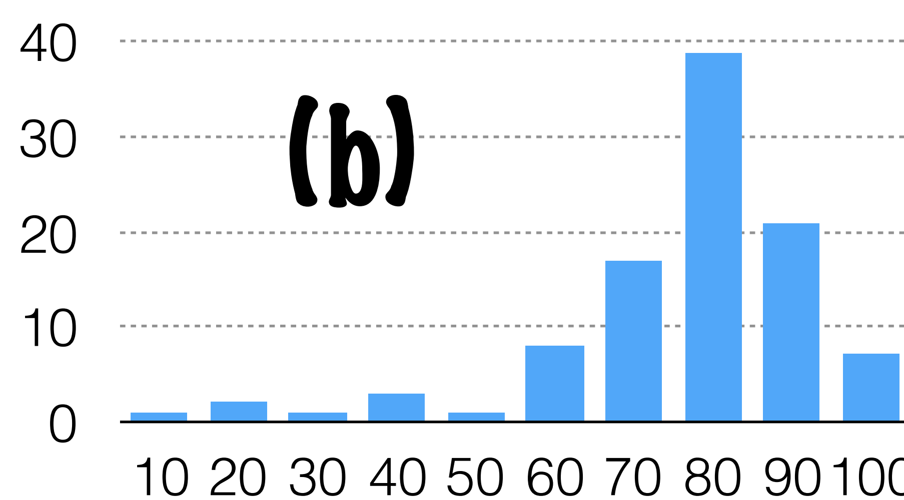
Clicker Question!

Every year, I compute the mean grade in my class. I never change the material or my methods for evaluating because, lazy. Over the 439 years that I have been teaching this class, this has resulted in the below distribution.



Which of these is mostly like the typical distribution on any given year?





(c) can't say, could be either

Clicker Question!

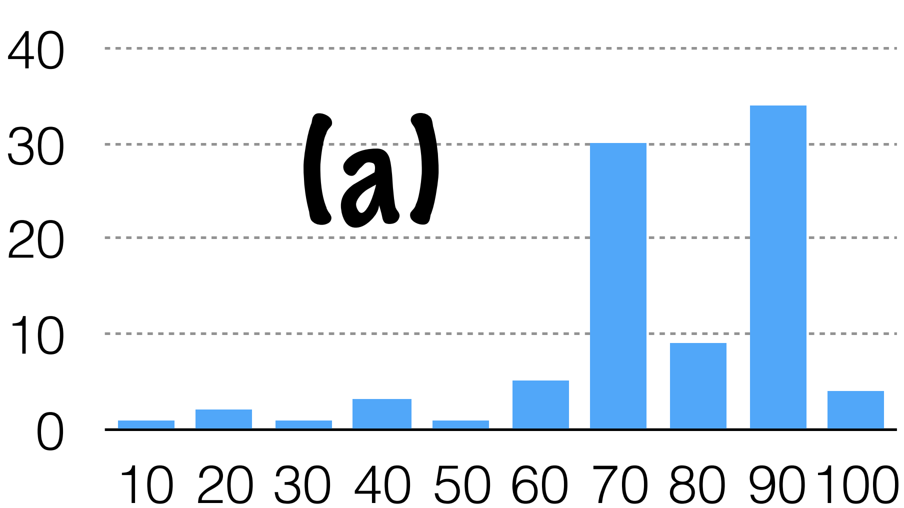
Every year, I compute the mean grade in my class. I never change the

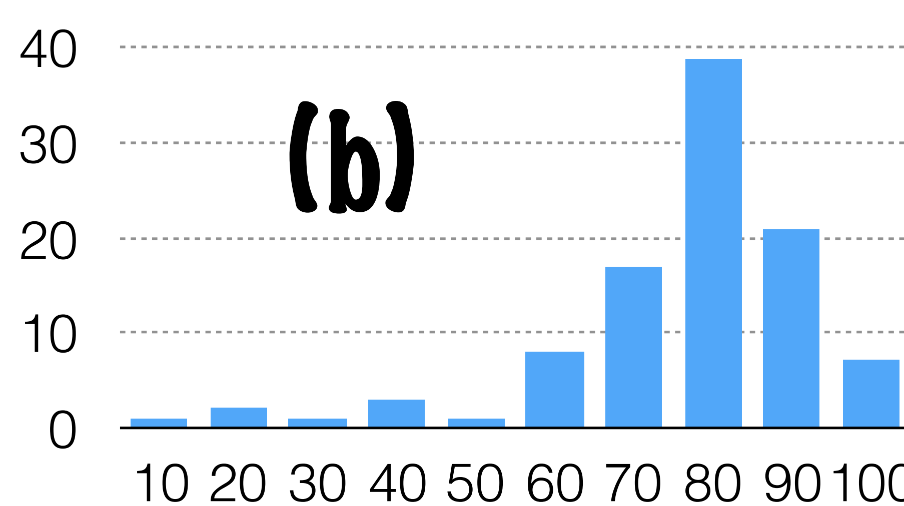
mater ... that I

hav ... on.

Central Limit Theorem: repeated measures of mean will be normally distributed, doesnt assume the population over which you are taking the mean is normally distributed.

Which of these is mostly like the typical distribution on any given year?





(c) can't say, could be either

Today

- Two quick preliminaries: LoLN and CLT
- **Follow up from last time**
- Common tests: chi-squared test, z-test/t-tests

Hypothesis Testing in General

Hypothesis Testing in General

- Null hypothesis (H_0) — the “nothing to see here” assumption

Hypothesis Testing in General

- Null hypothesis (H_0) — the “nothing to see here” assumption

- Alternative hypothesis (H_a)—the thing you know will lead to an explosive headline and are really hoping is true but you are a good scientist, so you will look to the data to confirm

Hypothesis Testing in General

Thing you can model

- Null hypothesis (H_0) — the "nothing to see here" assumption

- Alternative hypothesis (H_a)—the thing you know will lead to an explosive headline and are really hoping is true but you are a good scientist, so you will look to the data to confirm

Hypothesis Testing in General

Thing you can model

- Null hypothesis (H_0) — the "nothing to see here" assumption
- Alternative hypothesis (H_a)—the thing you know will lead to an explosive headline and are really hoping is true but you are a good scientist, so you will look to the data to confirm

Matters for how you compute p-values...more soon

Hypothesis Testing in General

- Assume the null hypothesis is true—i.e., don't deviate from status quo without good reason :)
- If there is enough evidence to suggest that H_0 is highly unlikely, then we can say we "reject the null hypothesis"
- If there is not enough evidence, we "fail to reject it"
- We don't "accept" or "prove" H_0 or H_a

Hypothesis Testing in General

- Determine an appropriate test statistic, given your null hypothesis

- Come up with a theoretical distribution of that test statistic (often, this work has been done for you)

- Compute the likelihood of your test statistic, given that theoretical distribution

Hypothesis Testing

"I swear literally like 80% of the answers are just (b)"

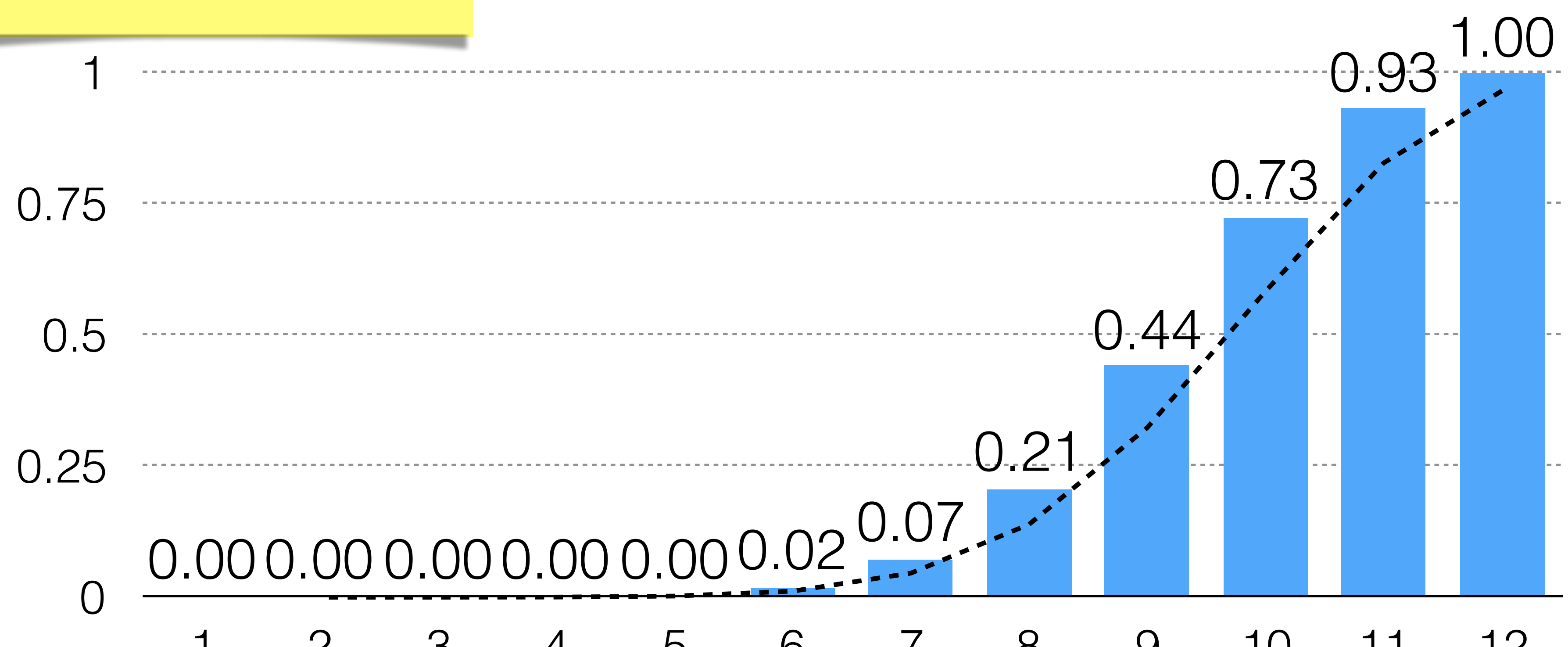
c	d	**b**	d
b	a	d	c
b	c	**b**	d

→ 4

$$\langle \Omega, F, P \rangle$$

{b, not b} p(B) = 0.8

cdf



Hypothesis Testing

Hypothesis

"I swear literally like 80% of the answers are just (b)"

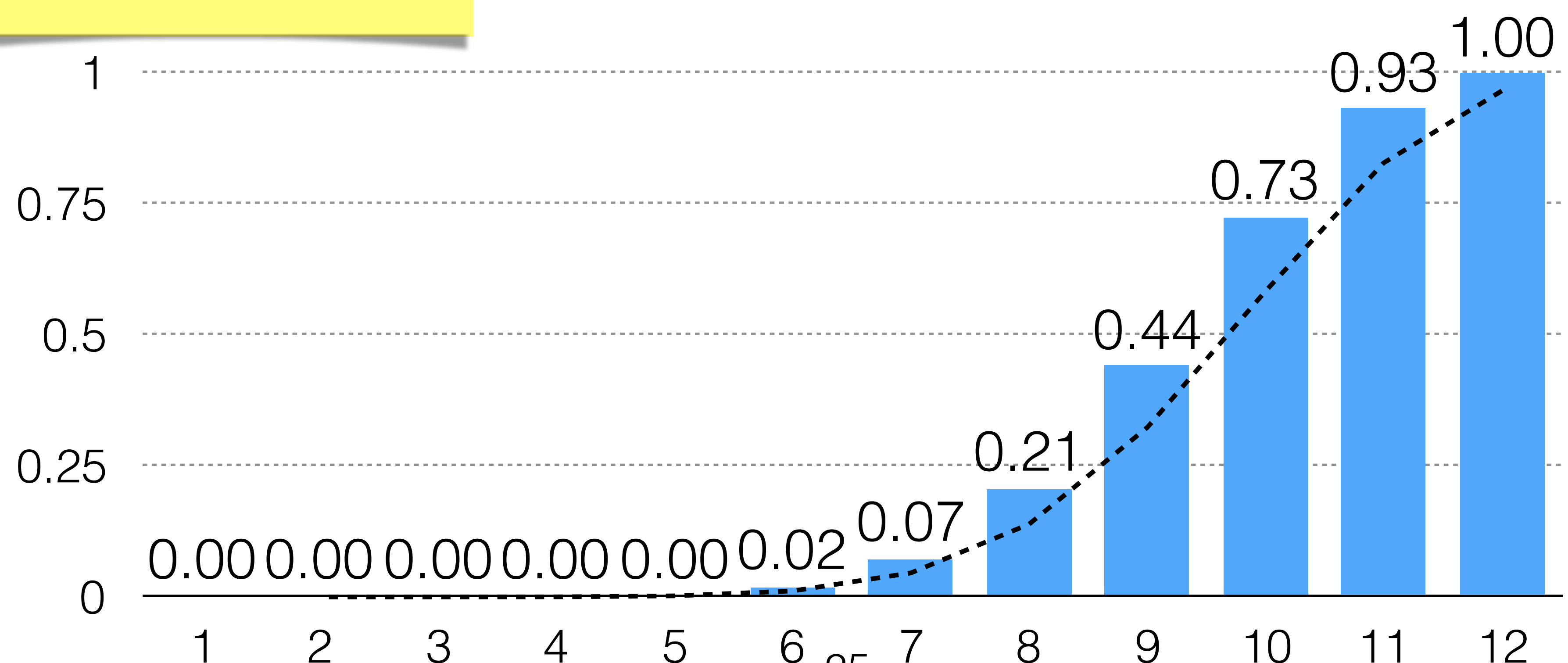
c	d	**b**	d
b	a	d	c
b	c	**b**	d

→ 4

$\langle \Omega, F, P \rangle$

{b, not b} p(B) = 0.8

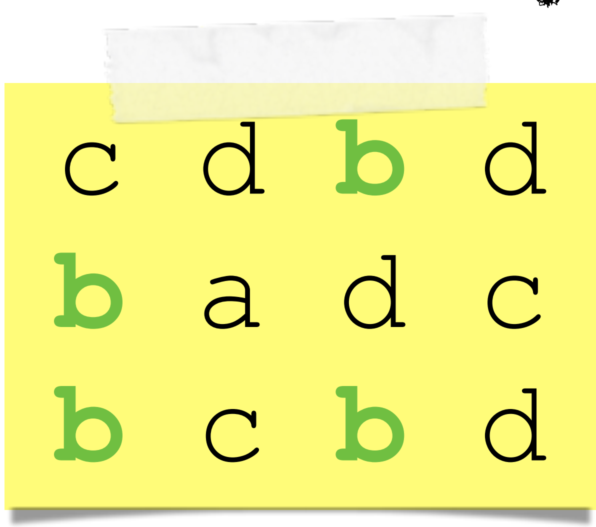
cdf



Hypothesis Testing

Observation/Sample

"I swear literally like 80% of the answers are just (b)"

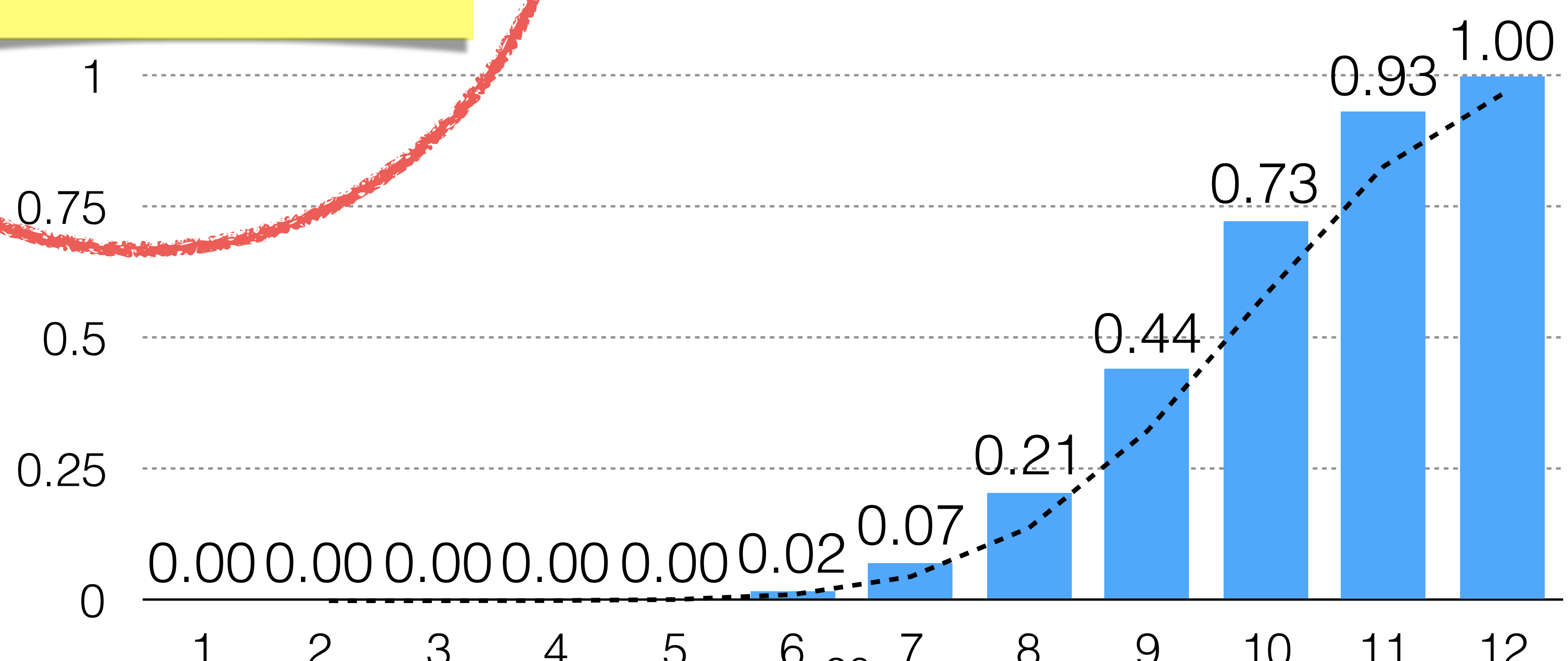


→ 4

$$\langle \Omega, F, P \rangle$$

{b, not b} p(B) = 0.8

cdf



Hypothesis Testing

Test Statistic

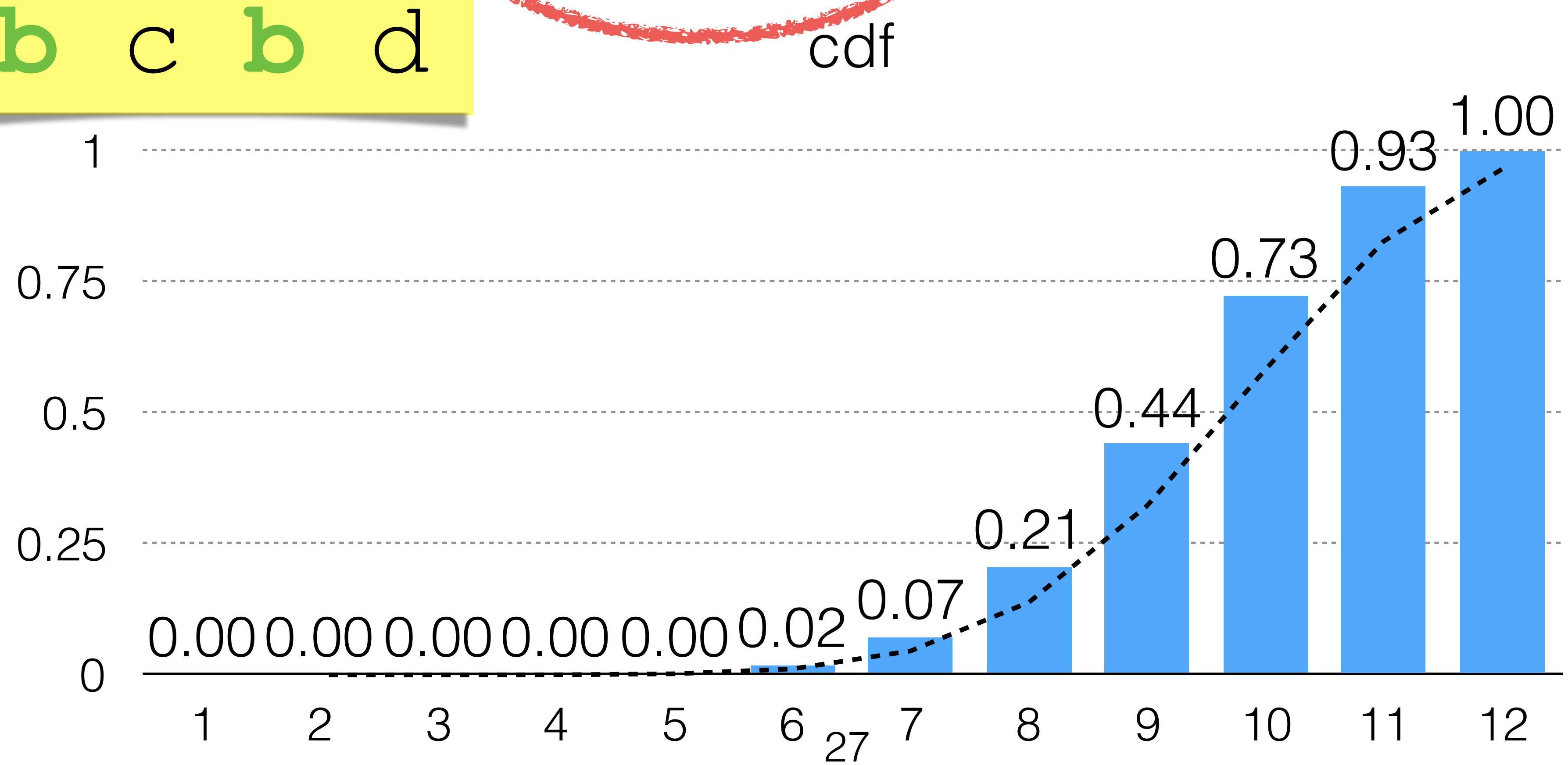
"I swear literally like 80% of the answers are just (b)"

c	d	**b**	d
b	a	d	c
b	c	**b**	d

→ 4

$$\langle \Omega, F, P \rangle$$

{b, not b} p(B) = 0.8



Hypothesis Testing

Theoretical Distribution

"I swear literally like 80% of the answers are just (b)"

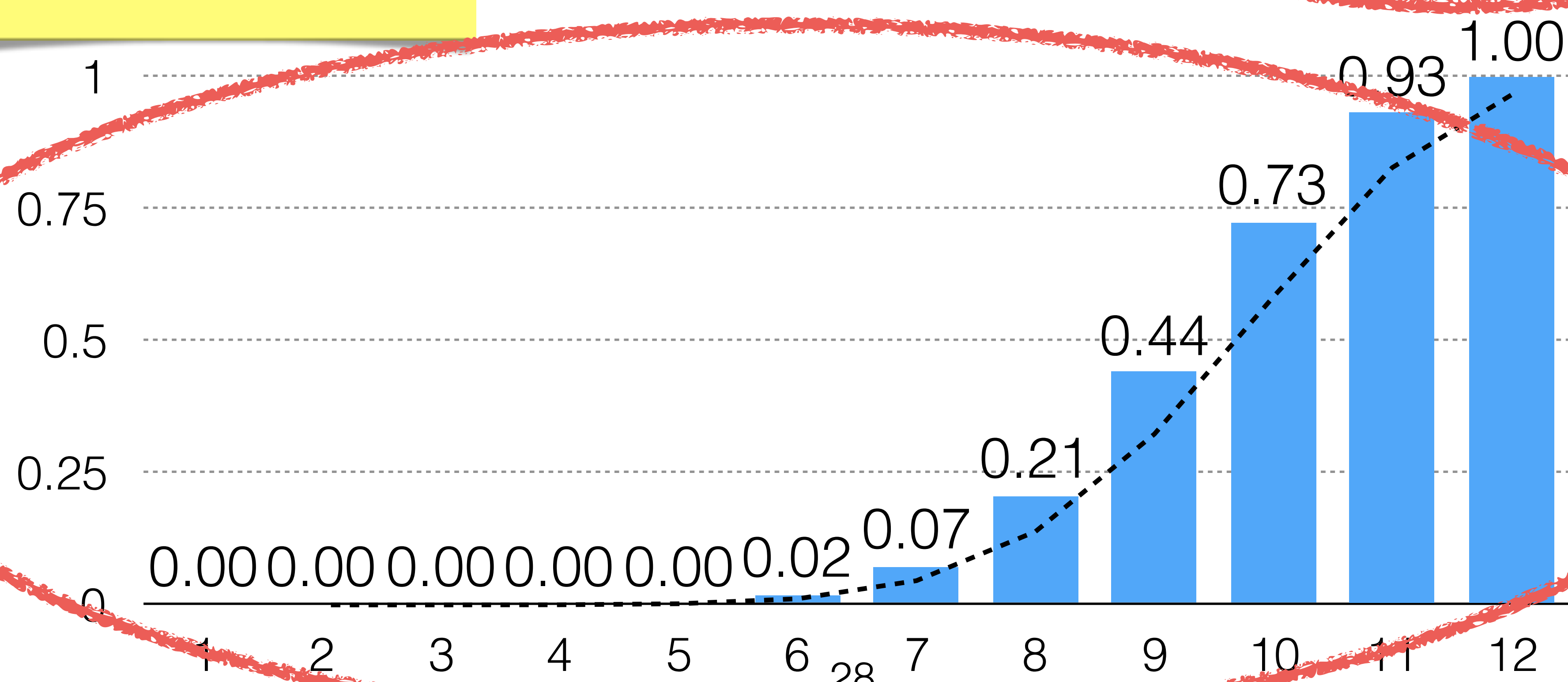
c	d	**b**	d
b	a	d	c
b	c	**b**	d

→ 4

$$\langle \Omega, F, P \rangle$$

{b, not b} p(B) = 0.8

cdf



Hypothesis Testing

"I swear literally like 80% of the answers are just (b)"

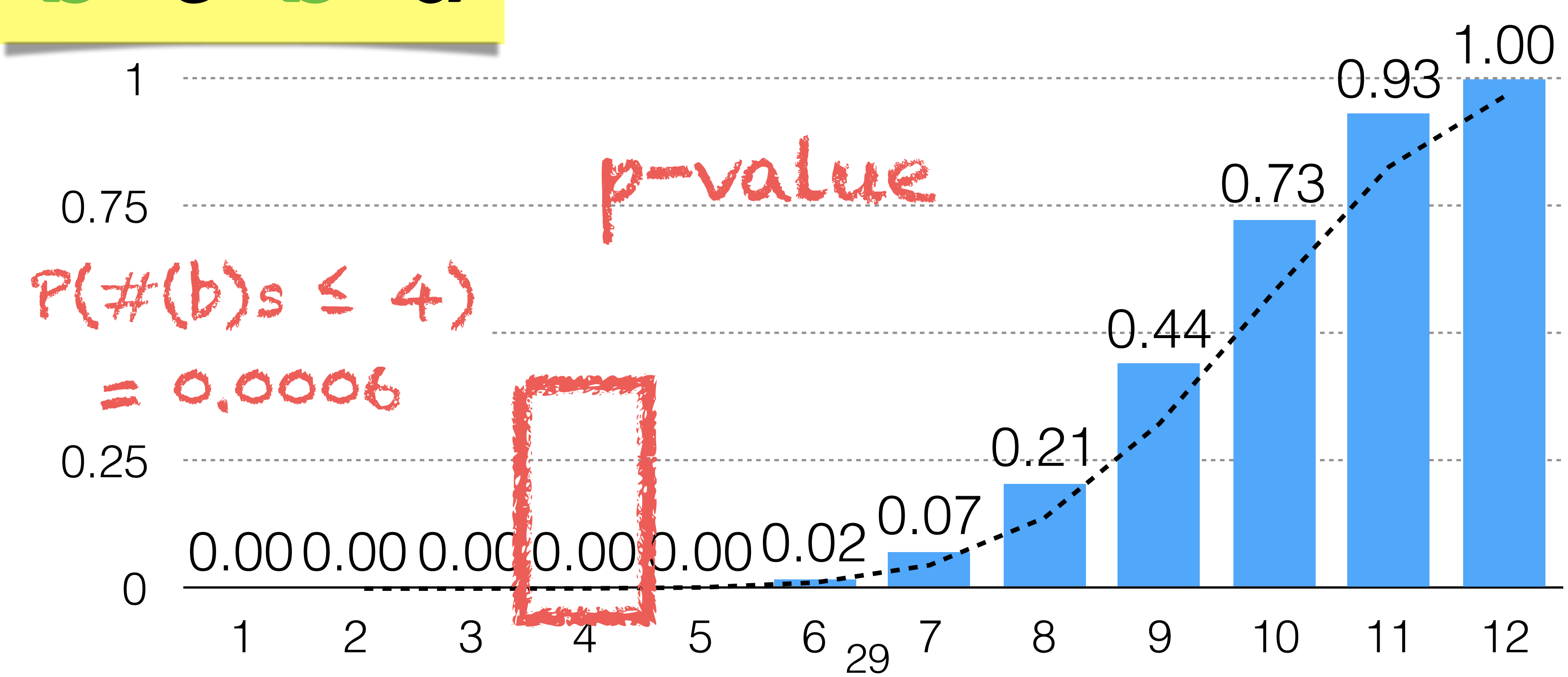
c d **b** d
b a d c
b c **b** d

→ 4

$\langle \Omega, F, P \rangle$

{b, not b} p(B) = 0.8

cdf



Today

- Two quick preliminaries: LoLN and CLT
- Follow up from last time
- **Common tests: chi-squared test, z-test/t-tests**

Some tests you are likely to use

Some tests you are likely to use

- **t-test**: difference of means; is the average value of some feature different between two populations
  - e.g. are men taller than women, are blue states more populated than red states, do CS work harder than other majors (::rolling_eyes::)

Some tests you are likely to use

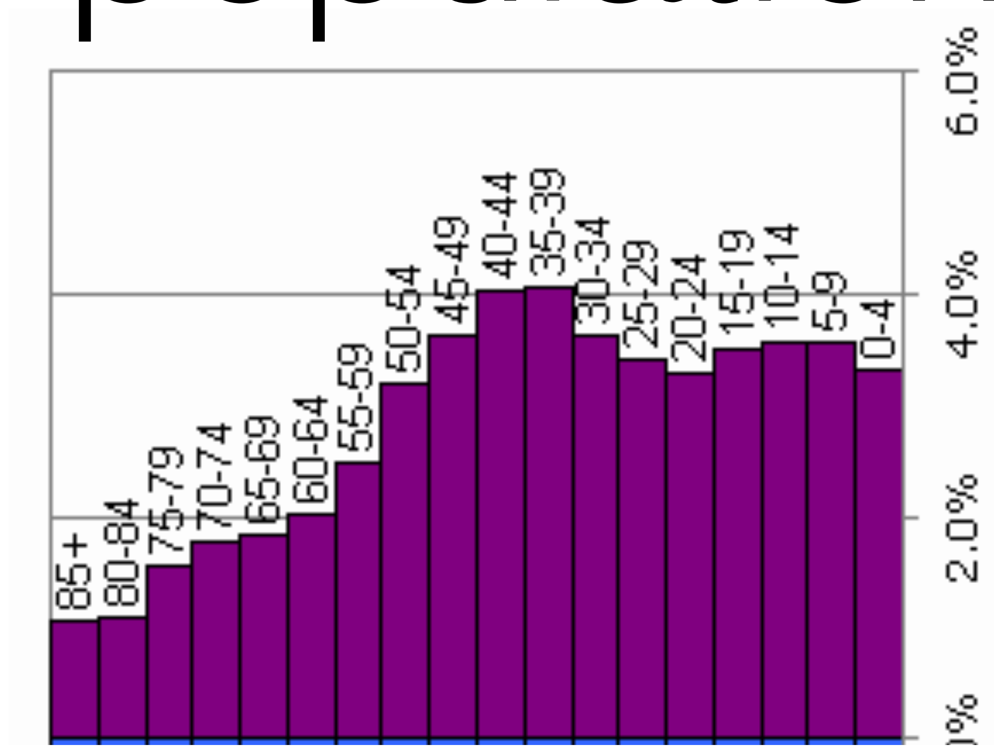
- **t-test**: difference of means; is the average value of some feature different between two populations
    - e.g. are men taller than women, are blue states more populated than red states, do CS work harder than other majors (::rolling_eyes::)
- **chi-squared test**: difference in frequencies of a categorical variable; is the distribution of some feature uniform across groups
    - e.g. do neighborhoods differ in terms of music preferences features; do college majors differ in terms of sociodemographic features

Some tests you are likely to use

- **t-test: difference of means; is the average value of some feature different between two populations**
  - e.g. are men taller than women, are blue states more populated than red states, do CS work harder than other majors (::rolling_eyes::)
- **chi-squared test**: difference in frequencies of a categorical variable; is the distribution of some feature uniform across groups
  - e.g. do neighborhoods differ in terms of music preferences features; do college majors differ in terms of sociodemographic features

Test for population means

Test for population means

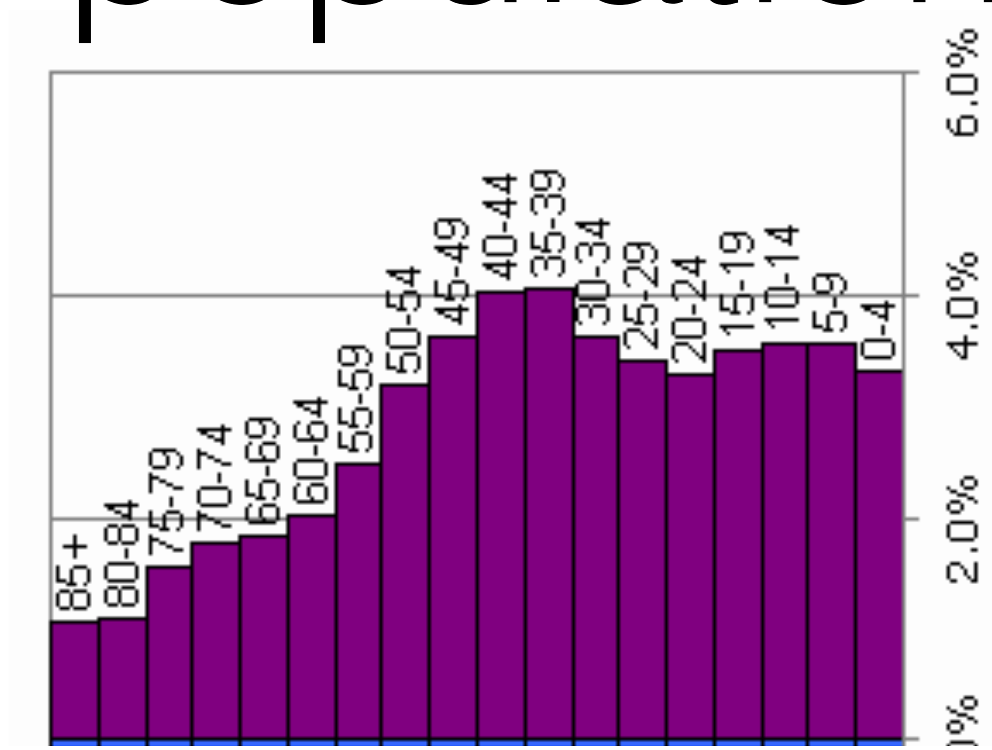


Distribution of ages in the US

Hypothesis: Mean age is 35.

http://www.censusscope.org/us/chart_age.html

Test for population means



Distribution of ages in the US

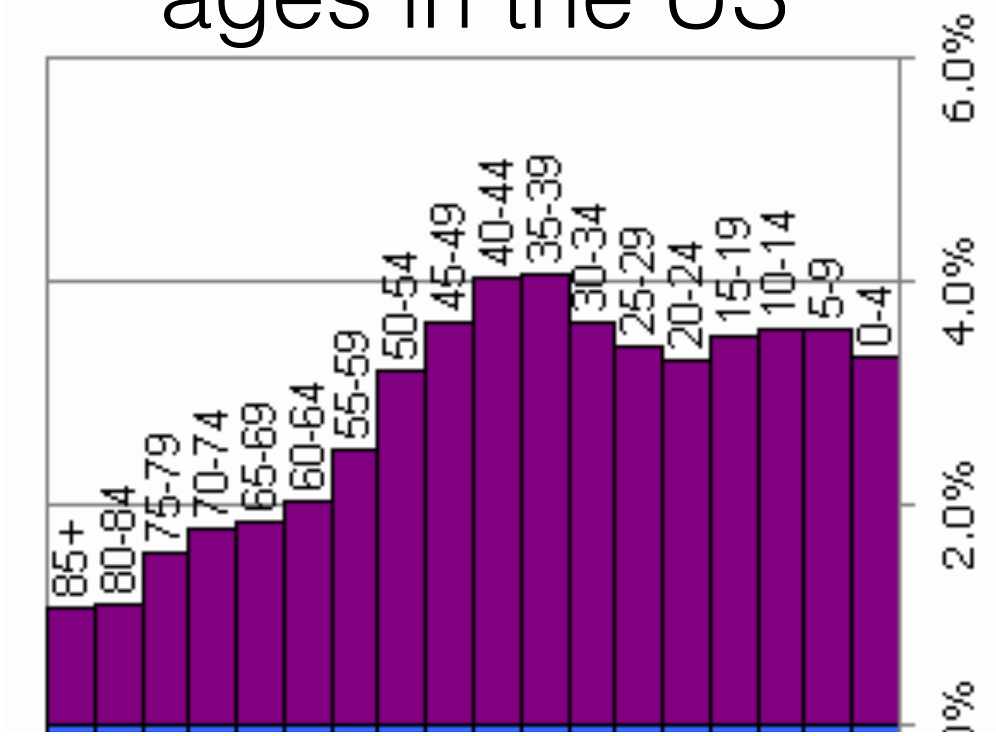
Hypothesis: Mean age is 35.

Thoughts about test statistics?

http://www.censusscope.org/us/chart_age.html

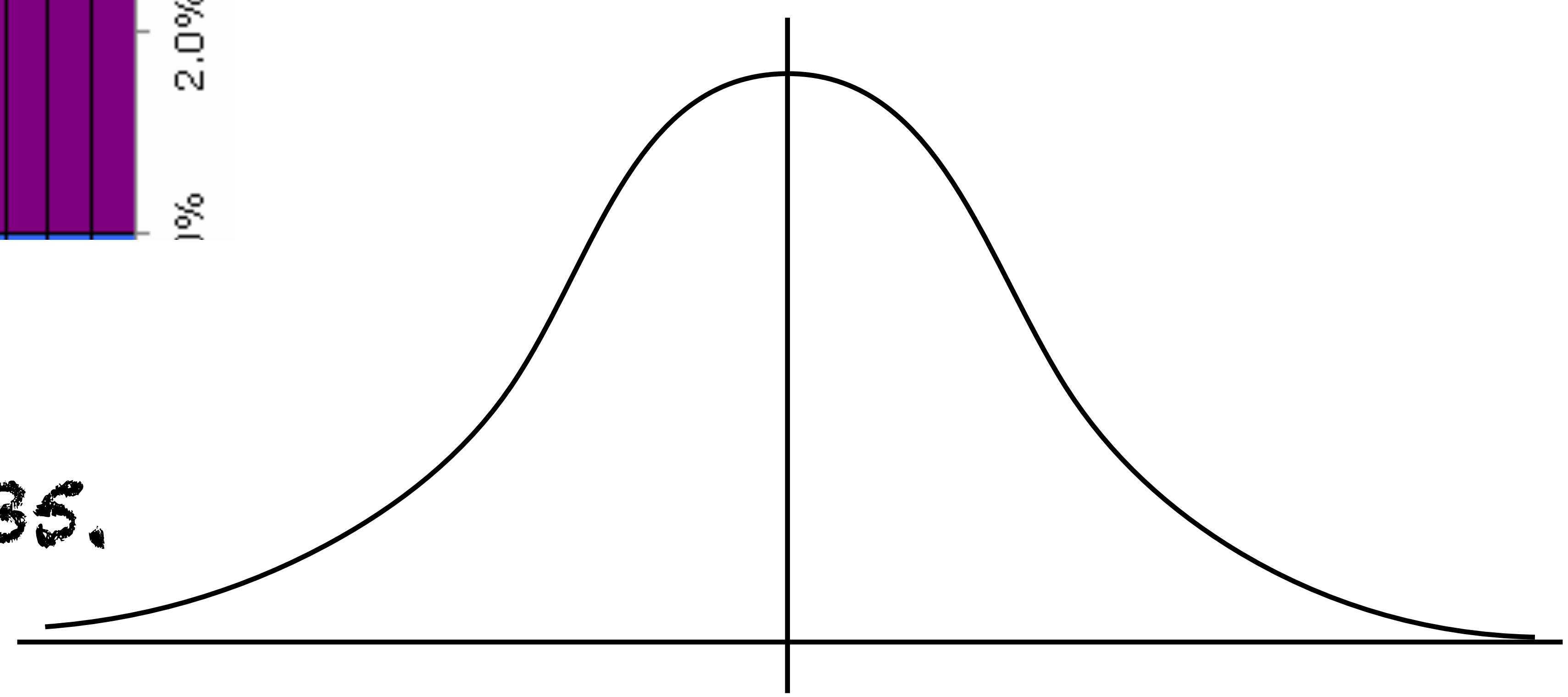
Test for population means

Distribution of ages in the US



$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

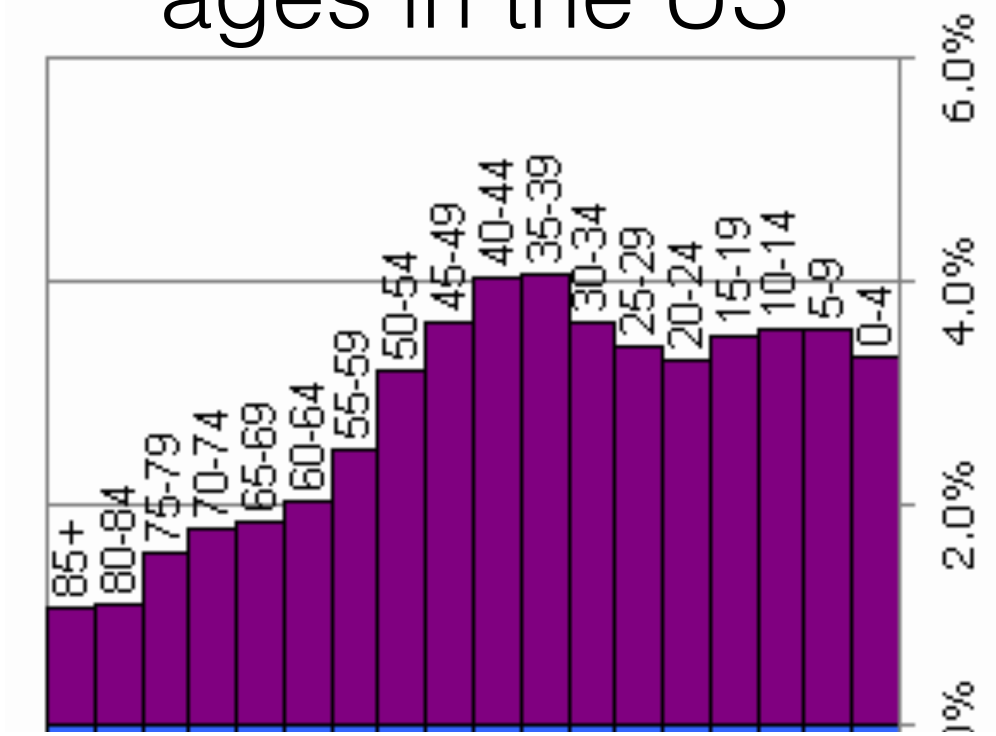
Hypothesis:
Mean age is 35.



http://www.censusscope.org/us/chart_age.html

Test for population means

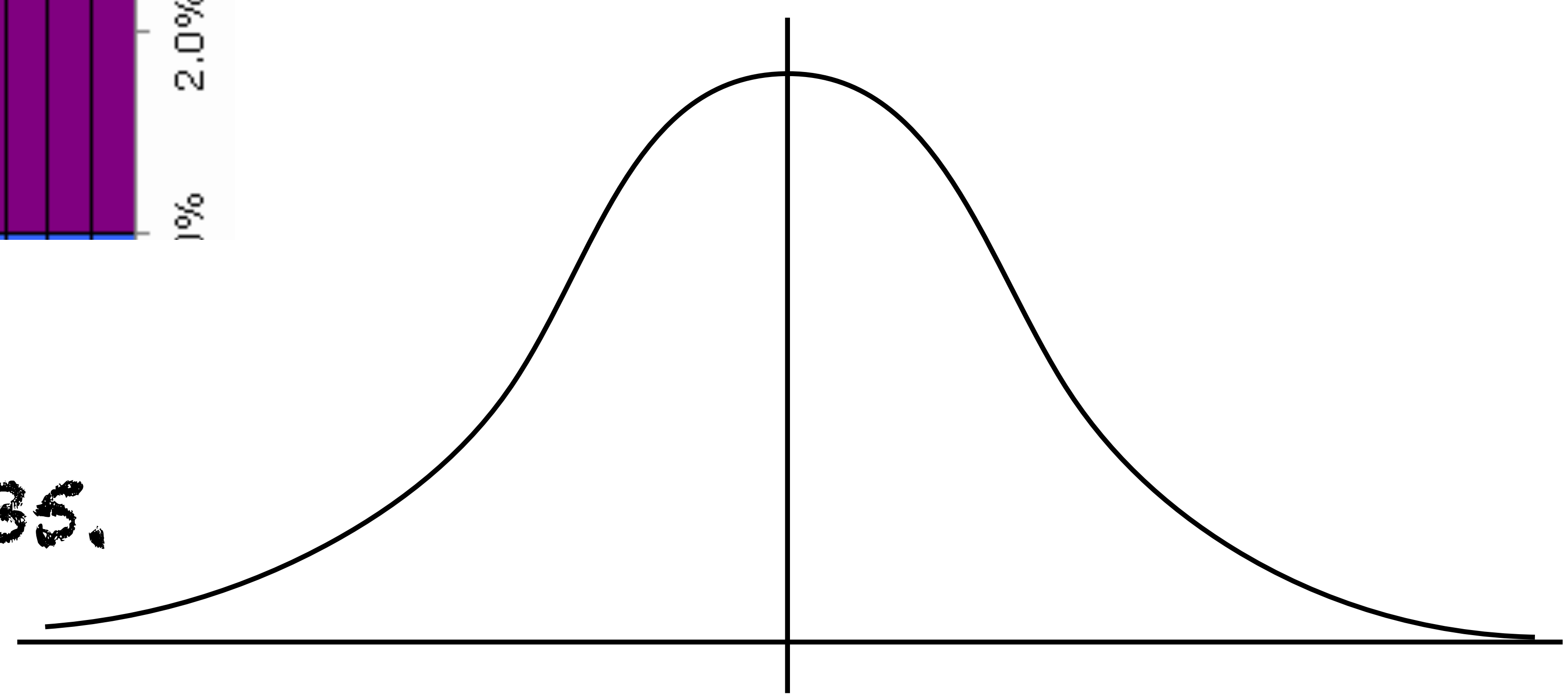
Distribution of ages in the US



Hypothesis:
Mean age is 35.

observed

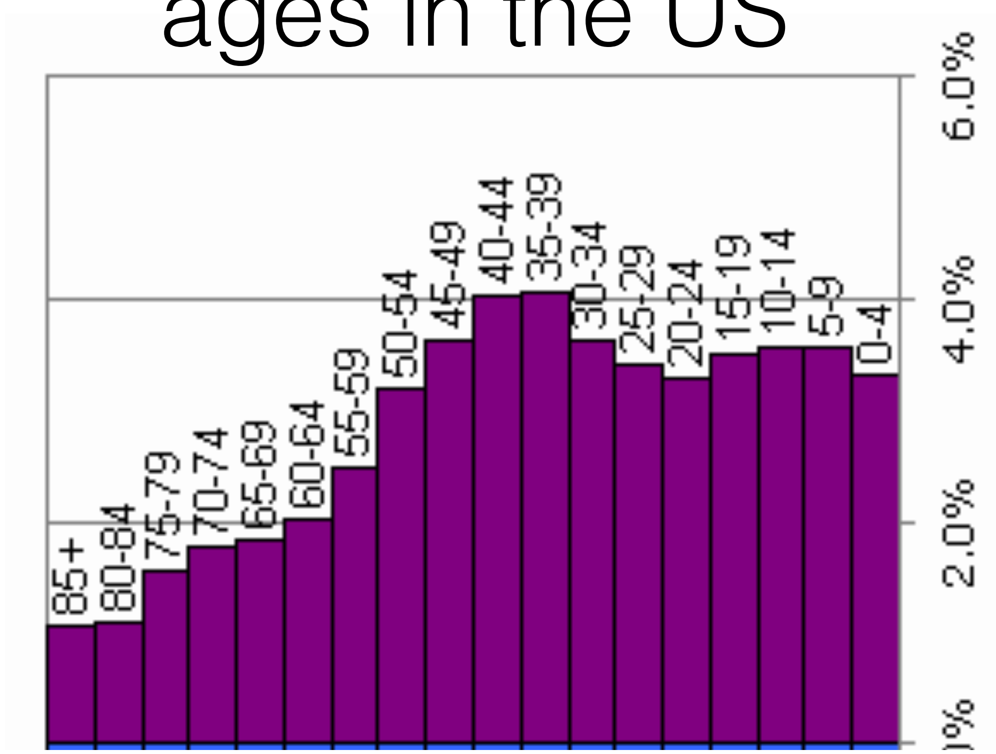
$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$



http://www.censusscope.org/us/chart_age.html

Test for population means

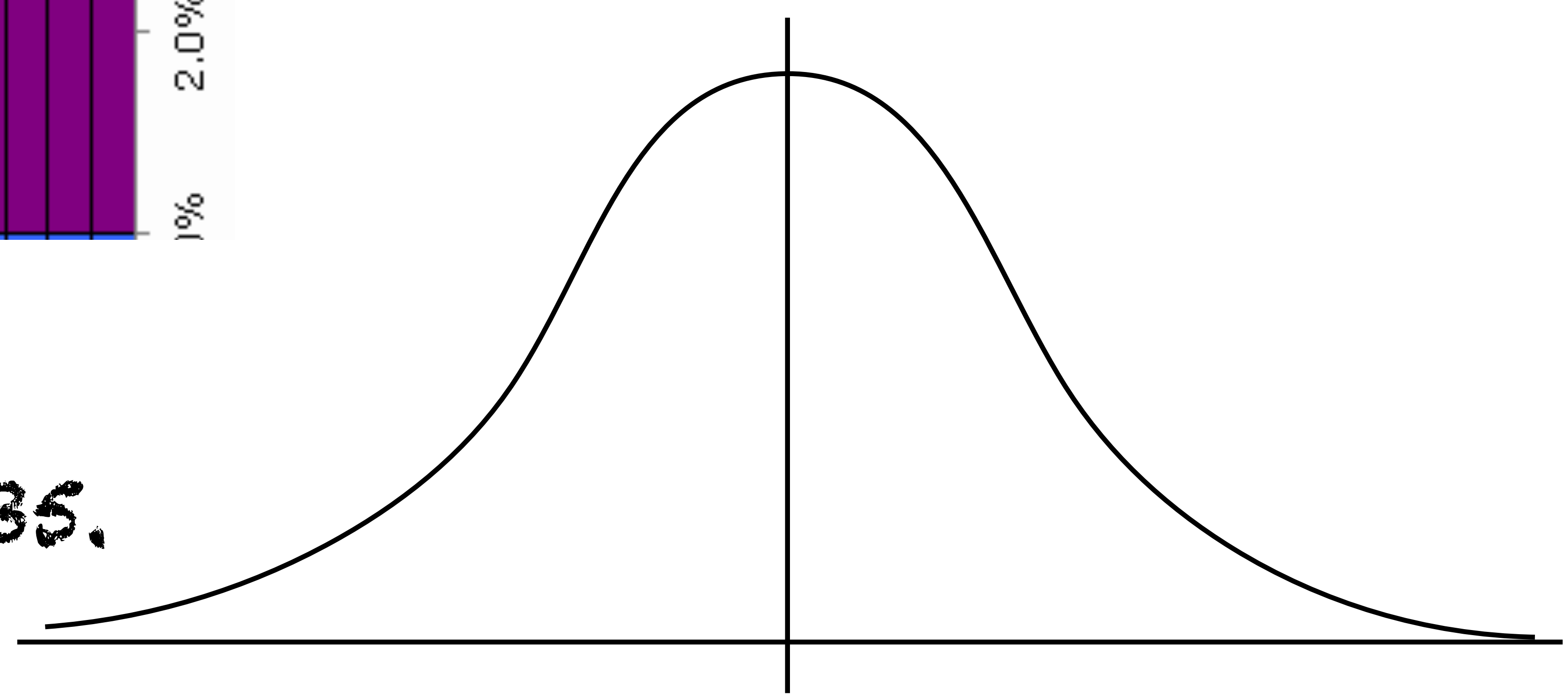
Distribution of ages in the US



$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

expected

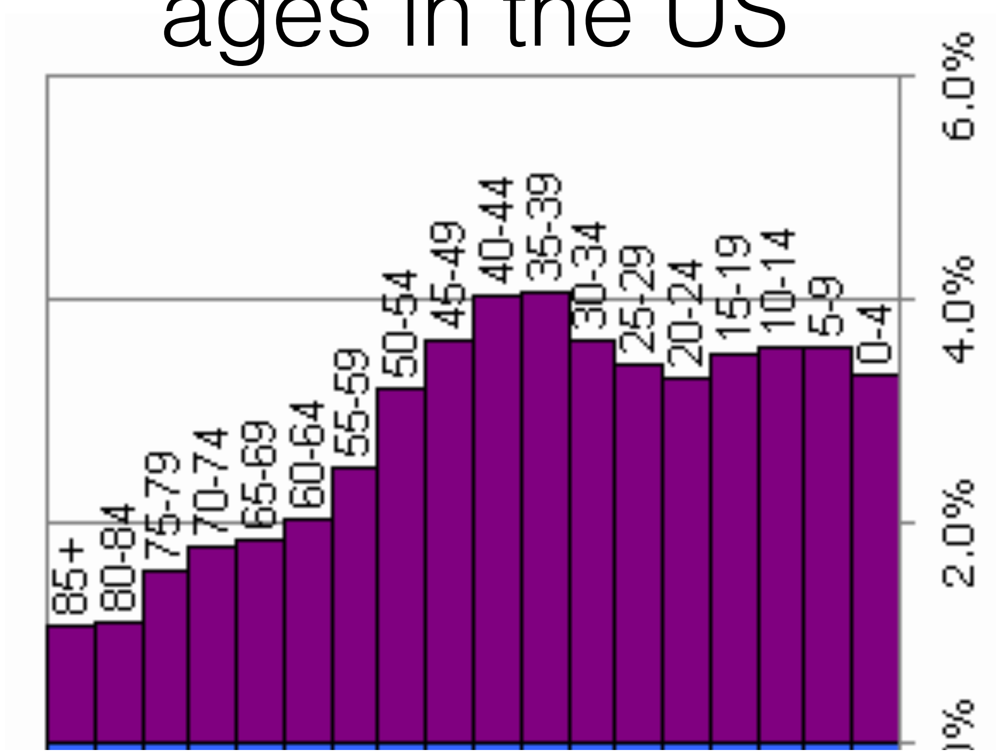
Hypothesis:
Mean age is 35.



http://www.censusscope.org/us/chart_age.html

Test for population means

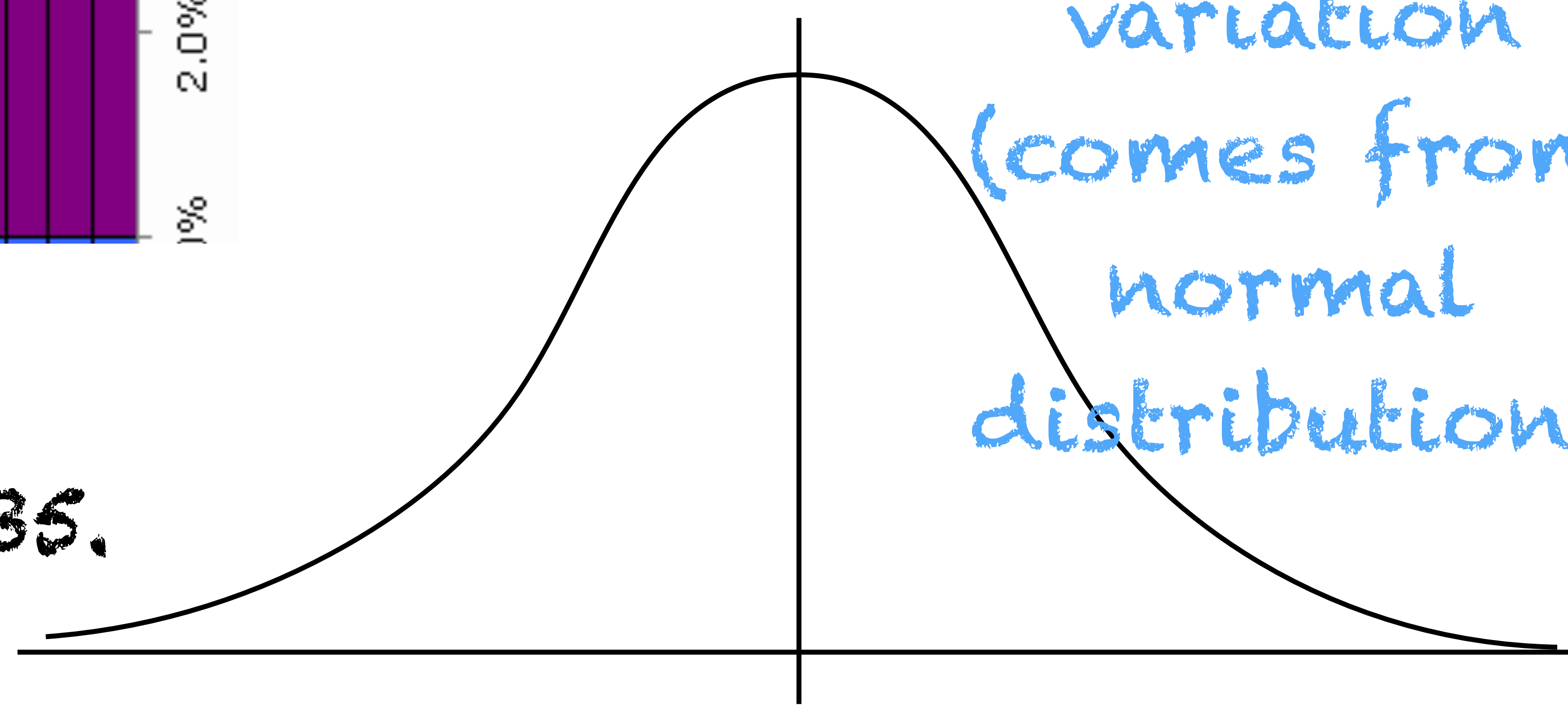
Distribution of ages in the US



$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

variation (comes from normal distribution)

Hypothesis:
Mean age is 35.



[4] http://www.censusscope.org/us/chart_age.html

Clicker Question!

Clicker Question!

Why can we use a normally-distributed test statistic to evaluate mean age of a population?

a) Because ages are normally distributed
b) Because the test statistic is a random variable
c) Because of the law of large numbers
d) Because of the central limit theorem
e) The limit does not exist!

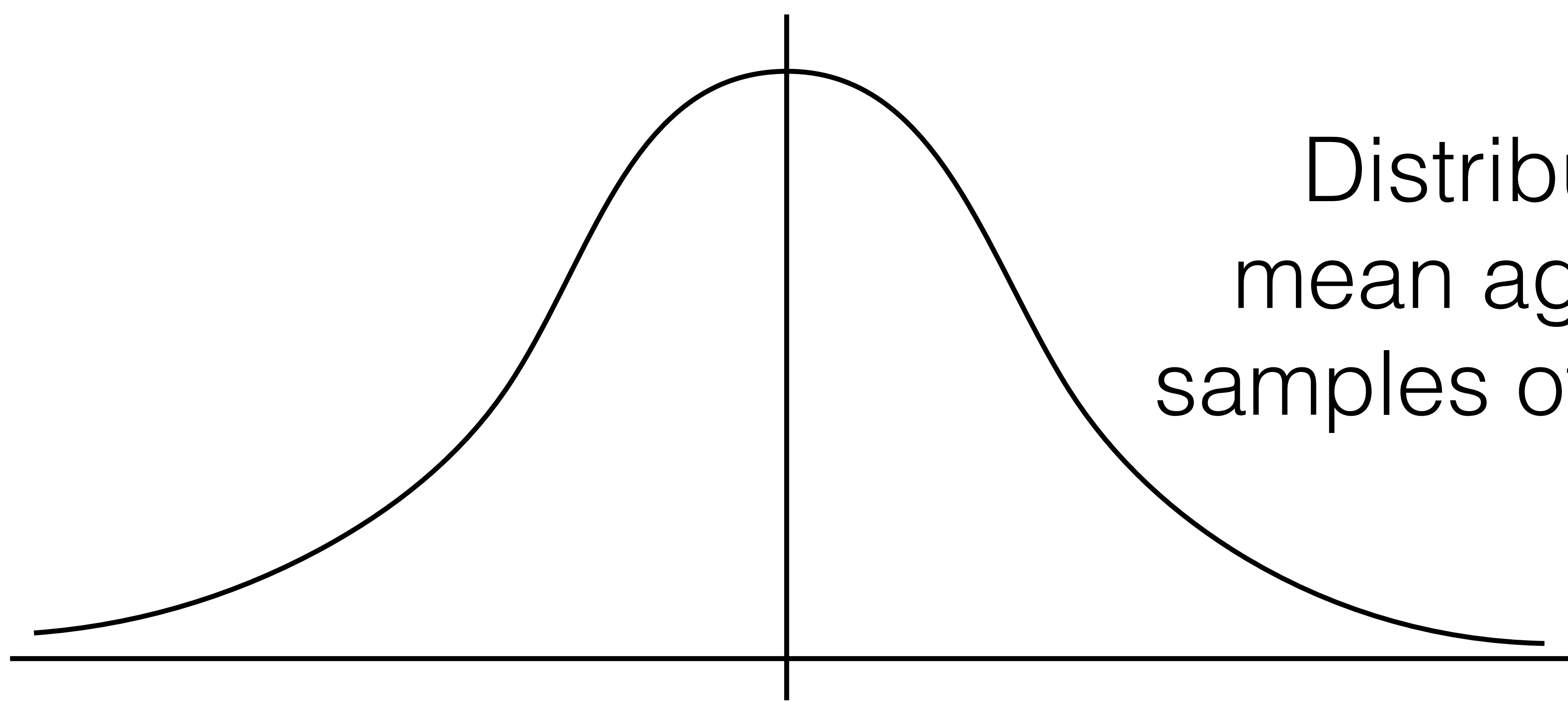
Clicker Question!

Why can we use a normally-distributed test statistic to evaluate mean age of a population?

a) Because ages are normally distributed
b) Because the test statistic is a random variable
c) Because of the law of large numbers
d) Because of the central limit theorem
e) The limit does not exist!

Test for population means

$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

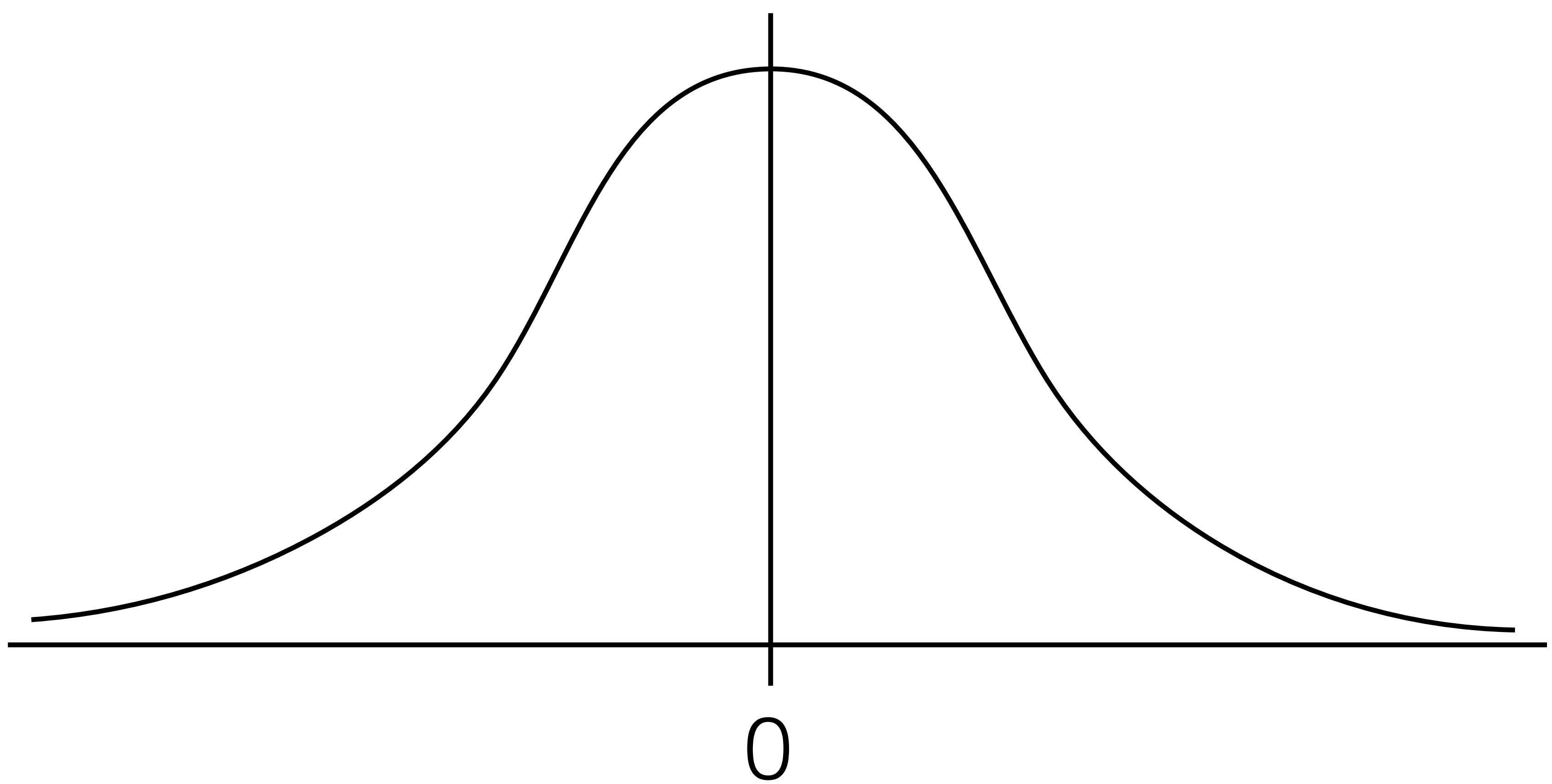


Distribution of mean ages given samples of US adults

http://www.censusscope.org/us/chart_age.html

Standard Normal Distr.

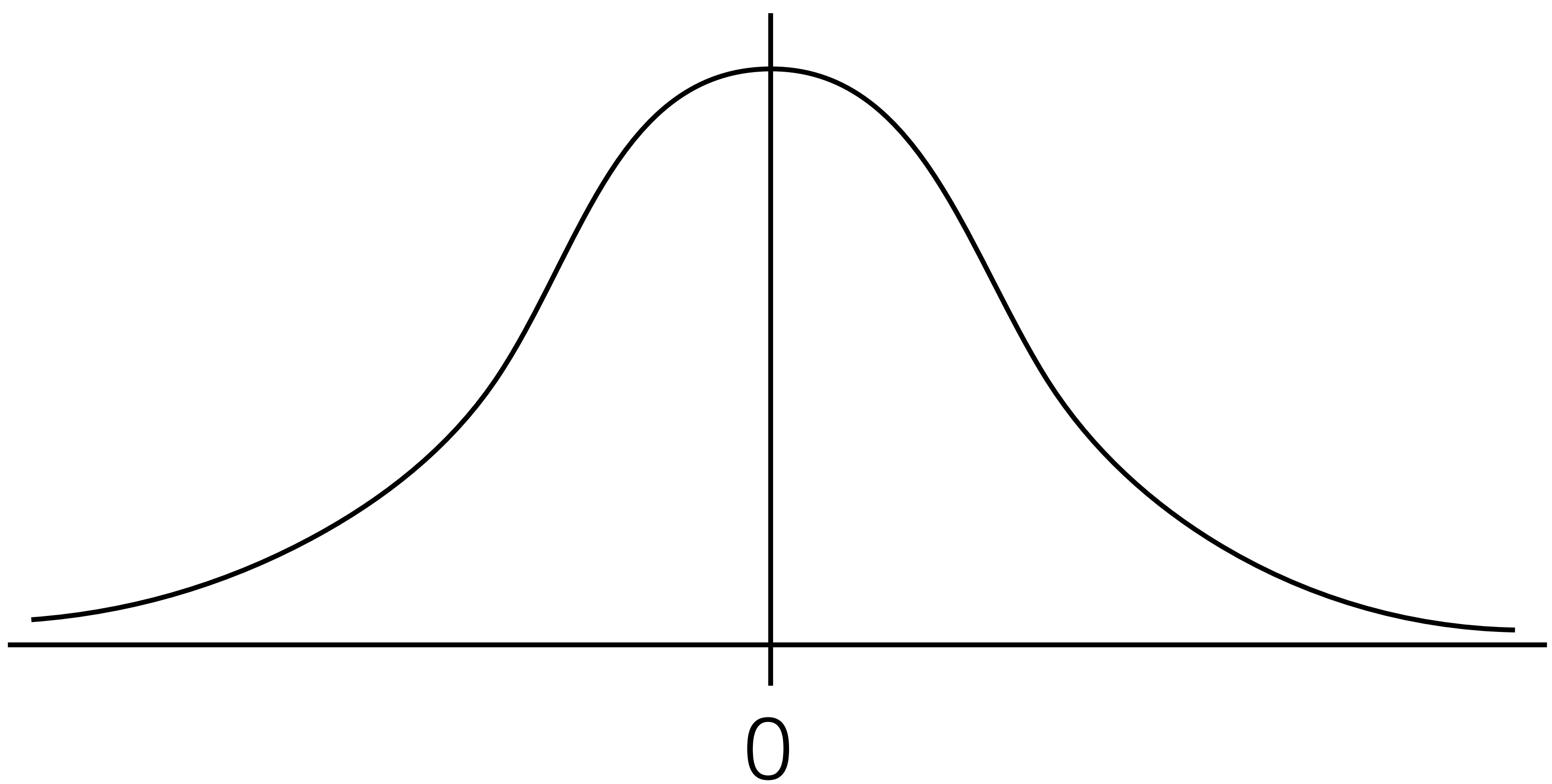
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$



z = distance from mean in std units

Standard Normal Distr.

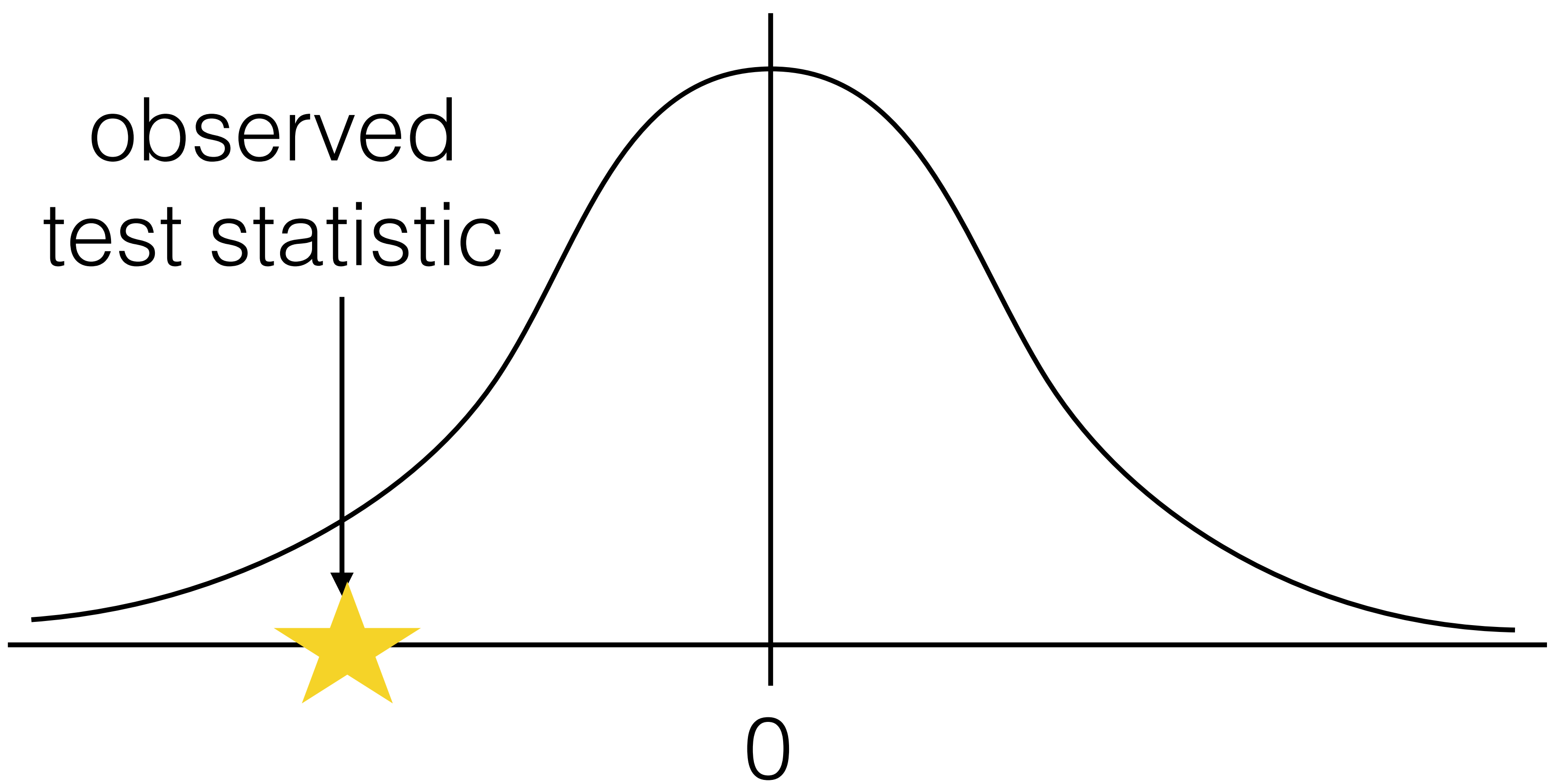
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$



z = (observed - expected) / standard deviation

Standard Normal Distr.

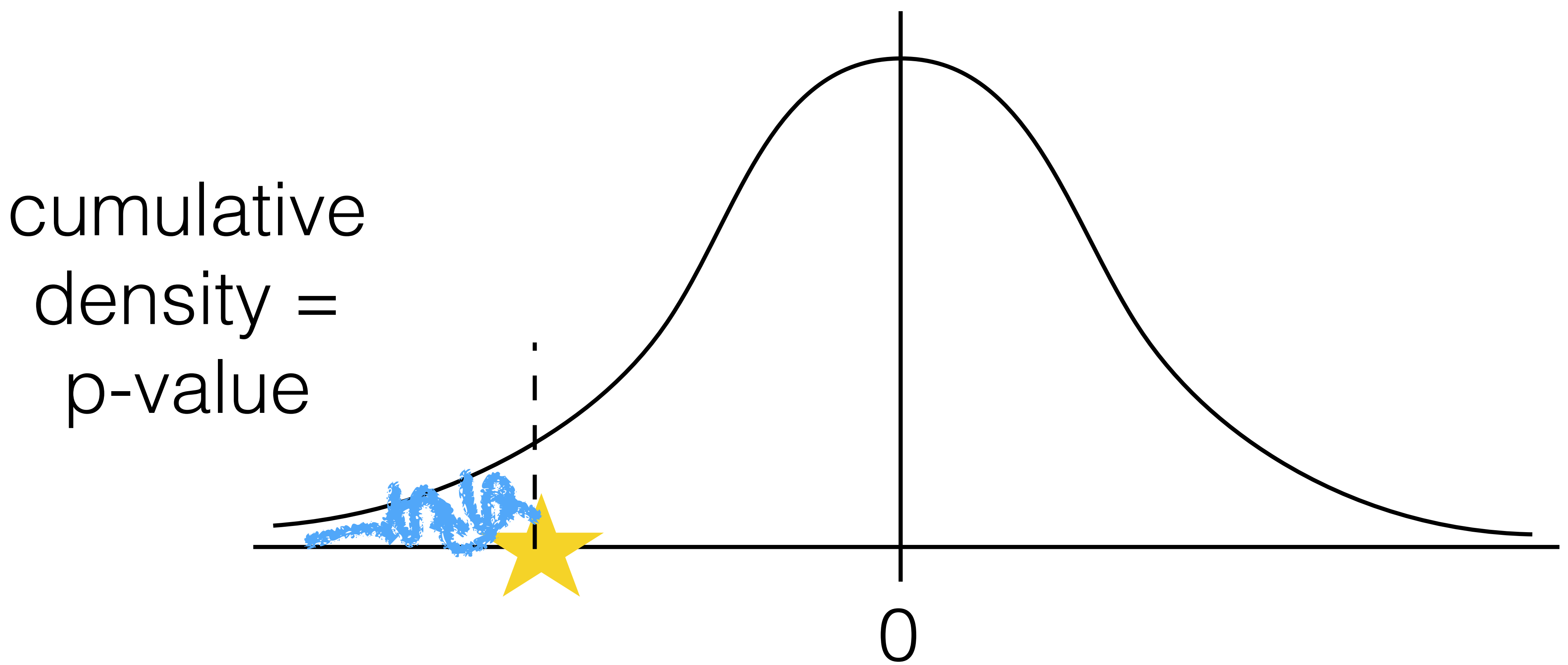
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$



z = (observed - expected) / standard deviation

Standard Normal Distr.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$



z = (observed - expected) / standard deviation

Detour! **α** and determining "significance"

Detour! α and determining "significance"

"I swear literally like 80% of the answers are just (b)"

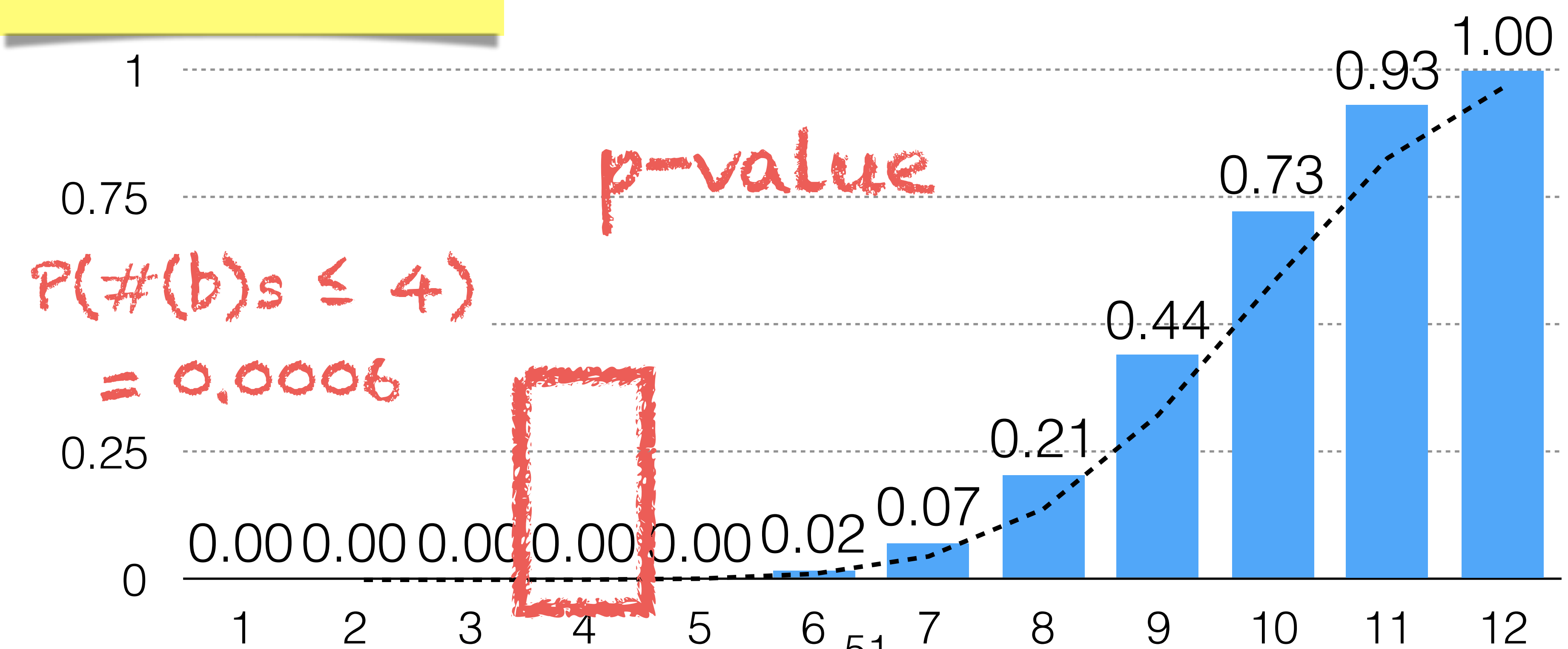
c	d	**b**	d
b	a	d	c
b	c	**b**	d

→ 4

$$\langle \Omega, F, P \rangle$$

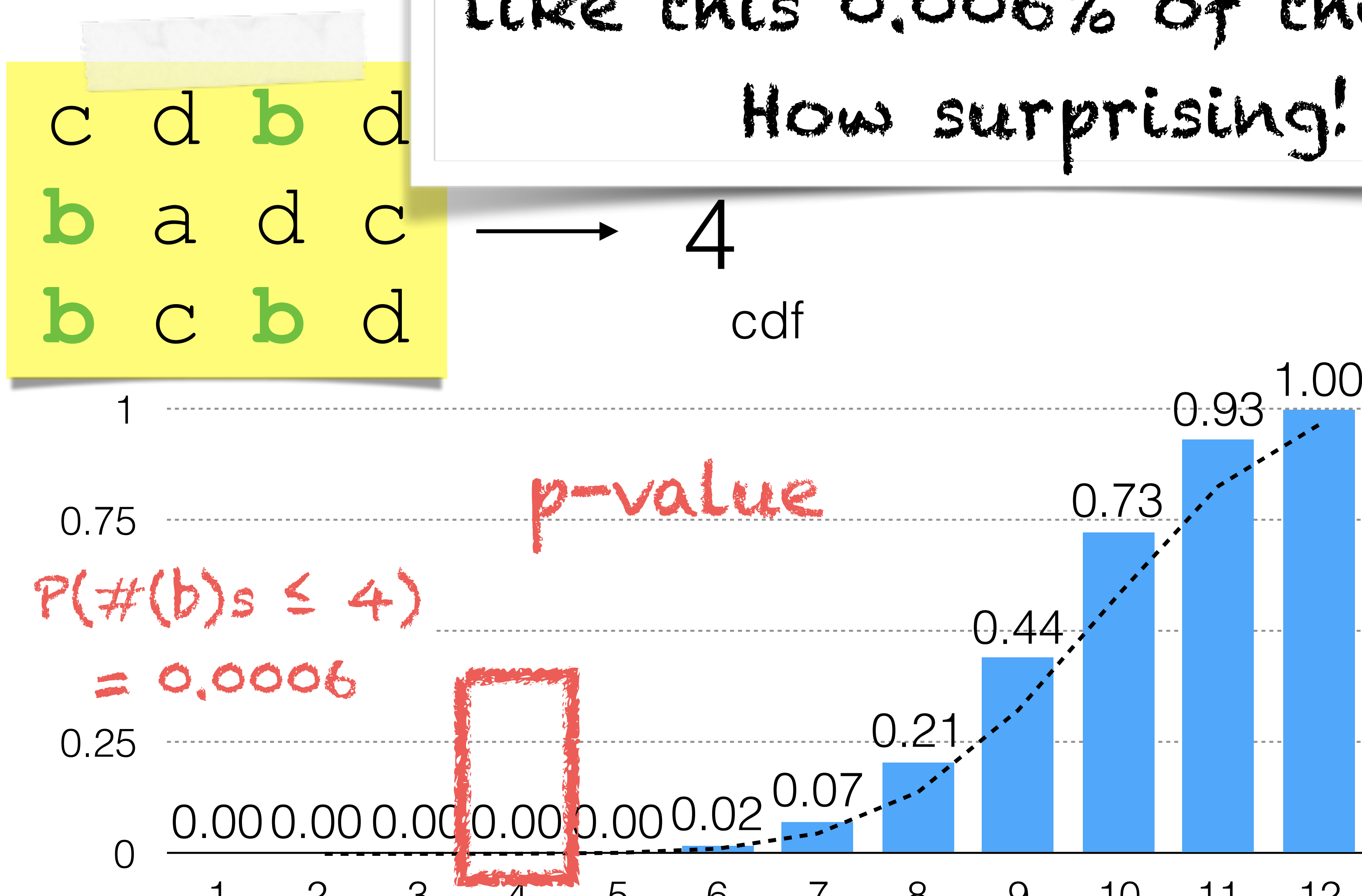
{b, not b} p(B) = 0.8

cdf



Detour

"assuming the null hypothesis were true, I would only expect to see a test statistic like this 0.006% of the time. How surprising!



Detour! ~ and determining

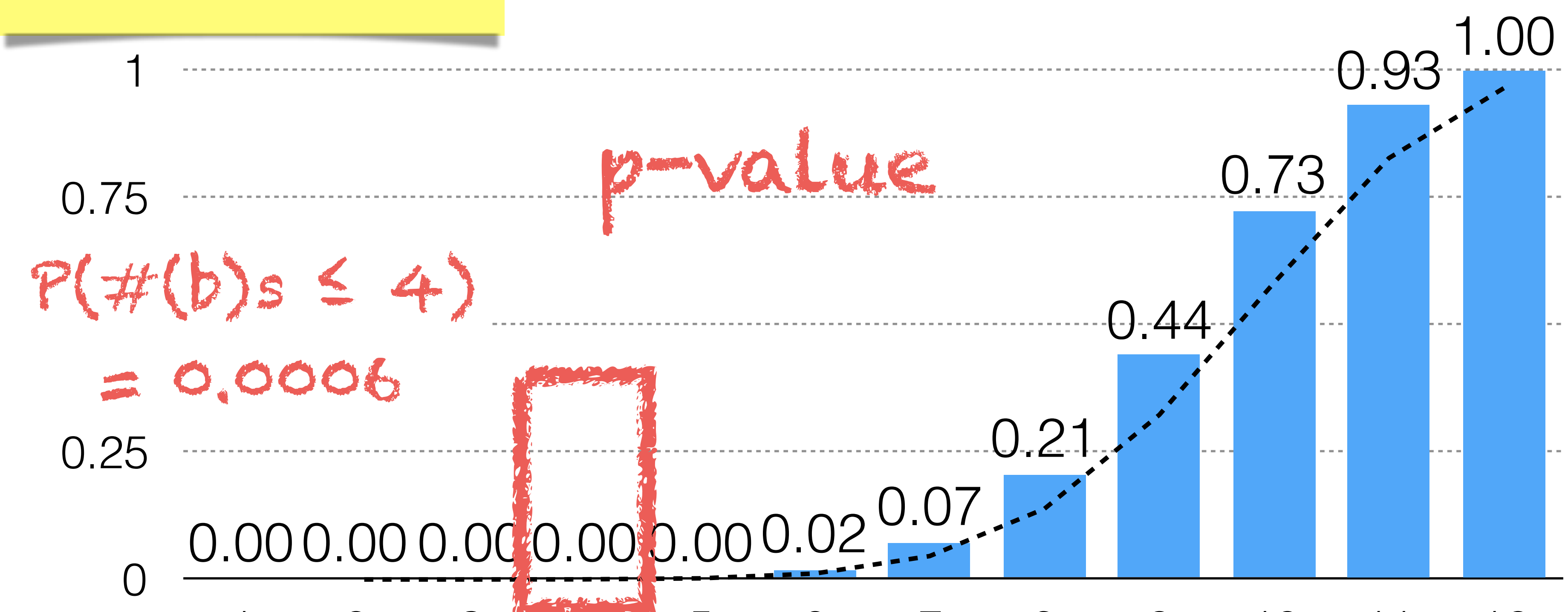
"assuming the null hypothesis were true, I would only expect to see a test statistic like this 0.006% of the time.

How surprising!

c	d	**b**	d
b	a	d	c
b	c	**b**	d

→ 4

cdf



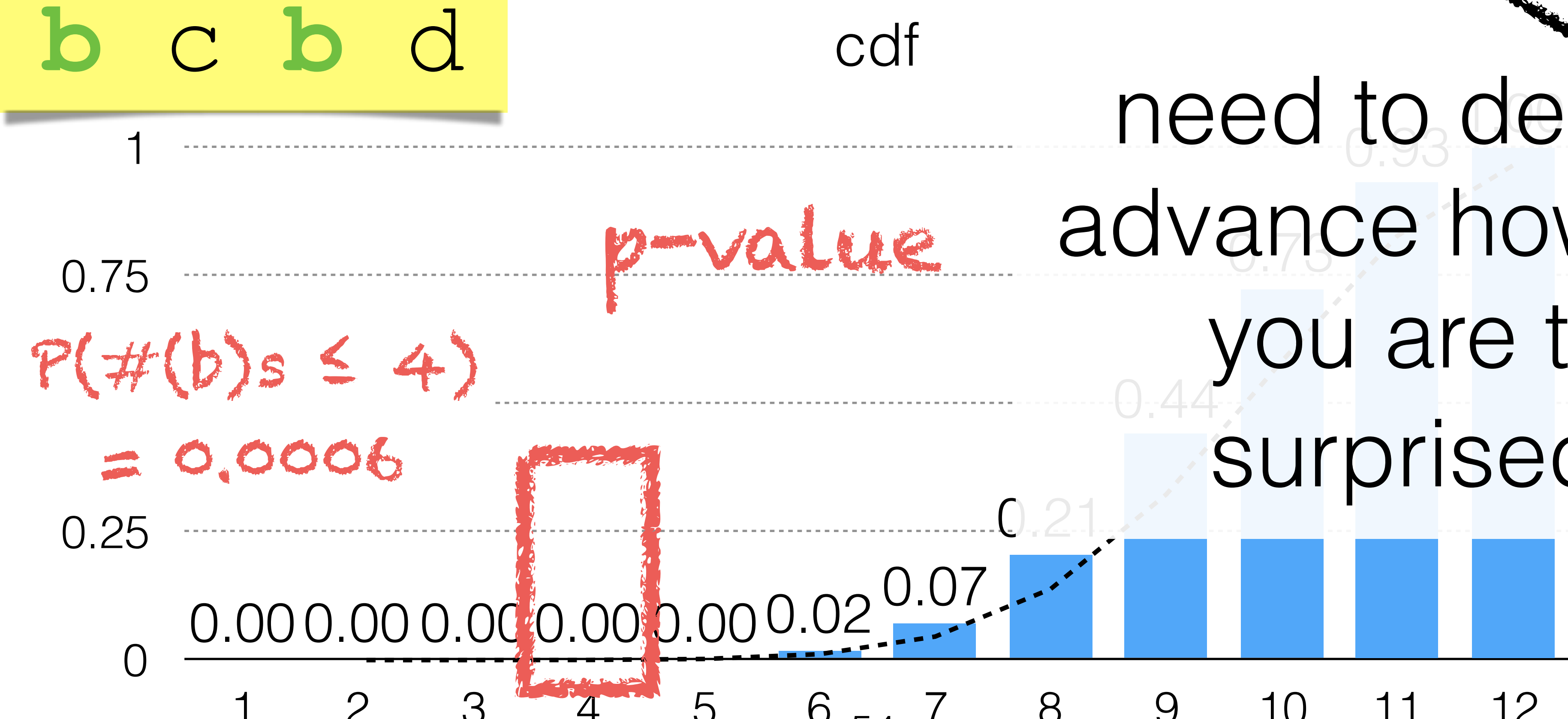
Detour: α and determining

"assuming the null hypothesis were true, I would only expect to see a test statistic like this 0.006% of the time.

How surprising!

c d b d
b a d c
b c b d

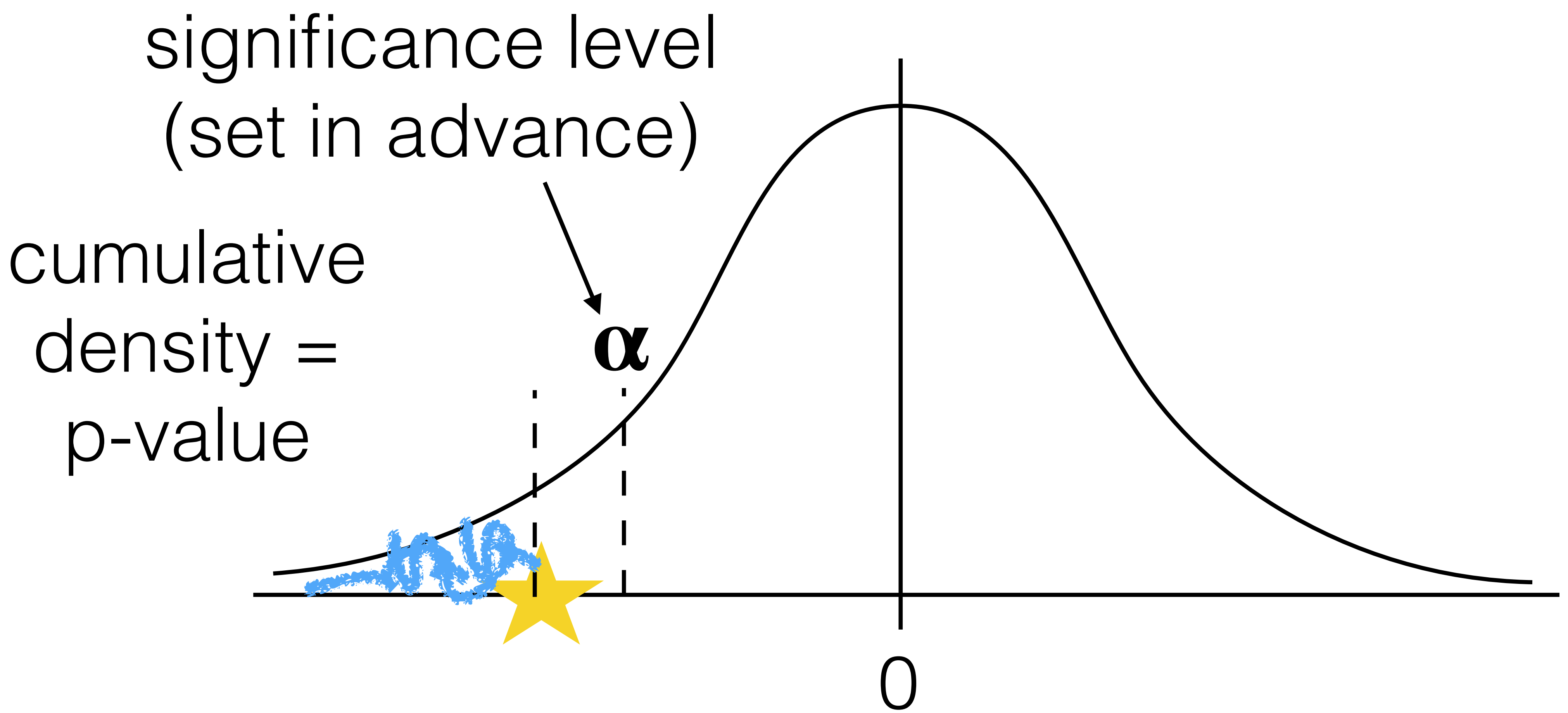
→ 4



need to decide in advance how willing you are to be surprised (**α**)

Standard Normal Distr.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$



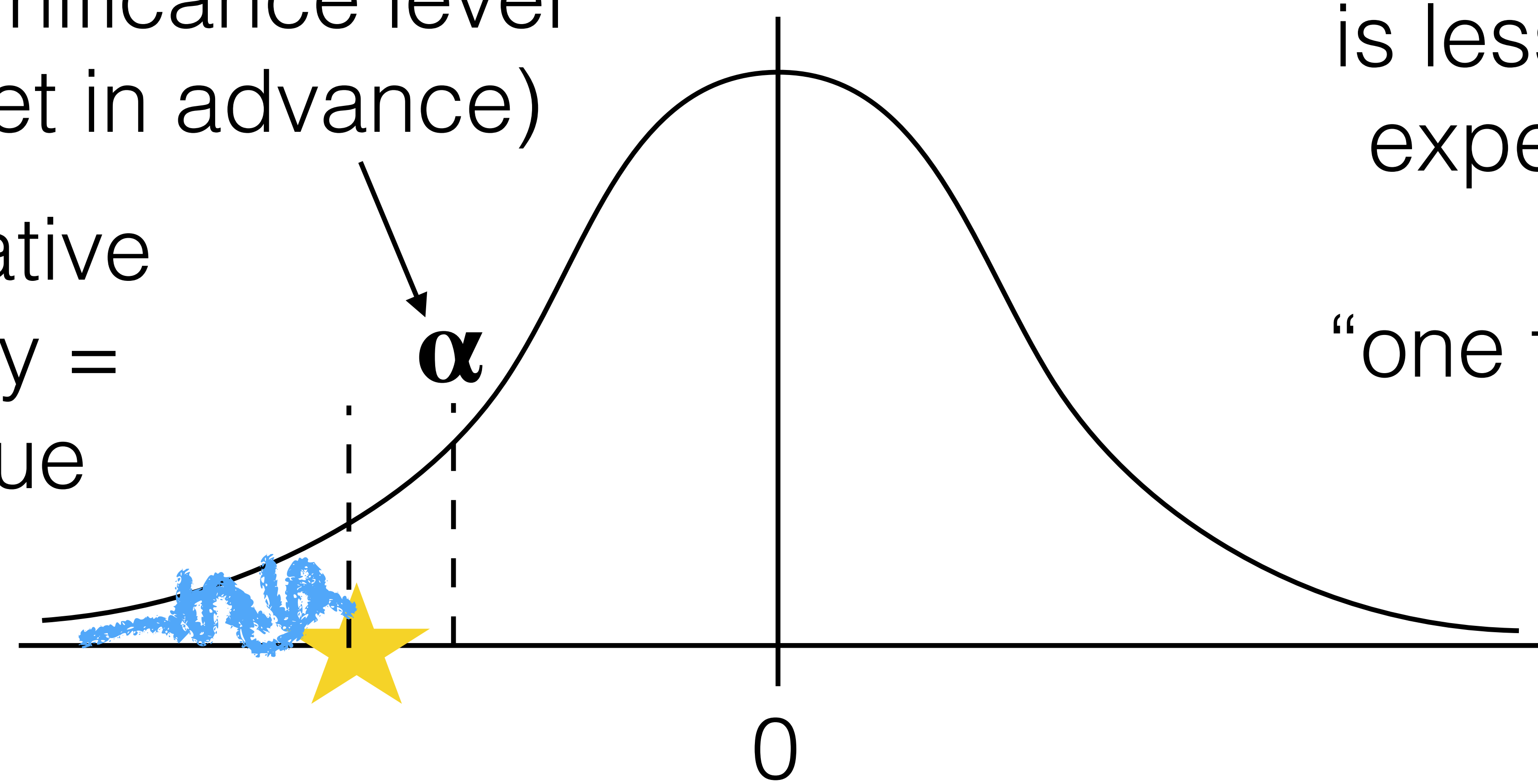
z = (observed - expected) / standard deviation

Standard Normal Distr.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

significance level
(set in advance)

cumulative
density =
p-value



Ha = test stat
is less than
expected

“one tailed”

z = (observed - expected) / standard deviation

assuming the null hypothesis is true, you will be still be "surprised" alpha % of the time

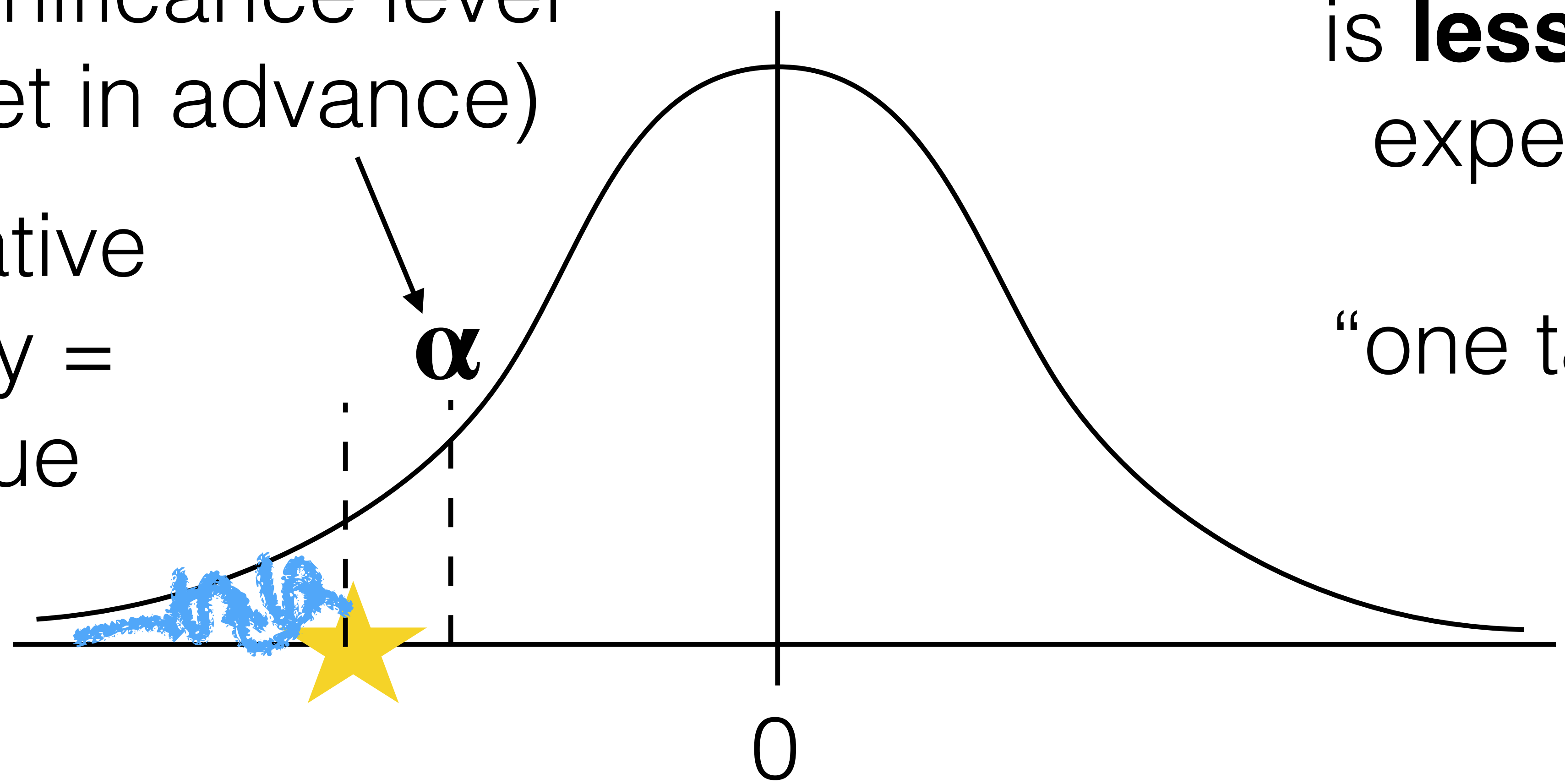
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Ha = test stat is **less than** expected

"one tailed"

significance level (set in advance)

cumulative density = p-value



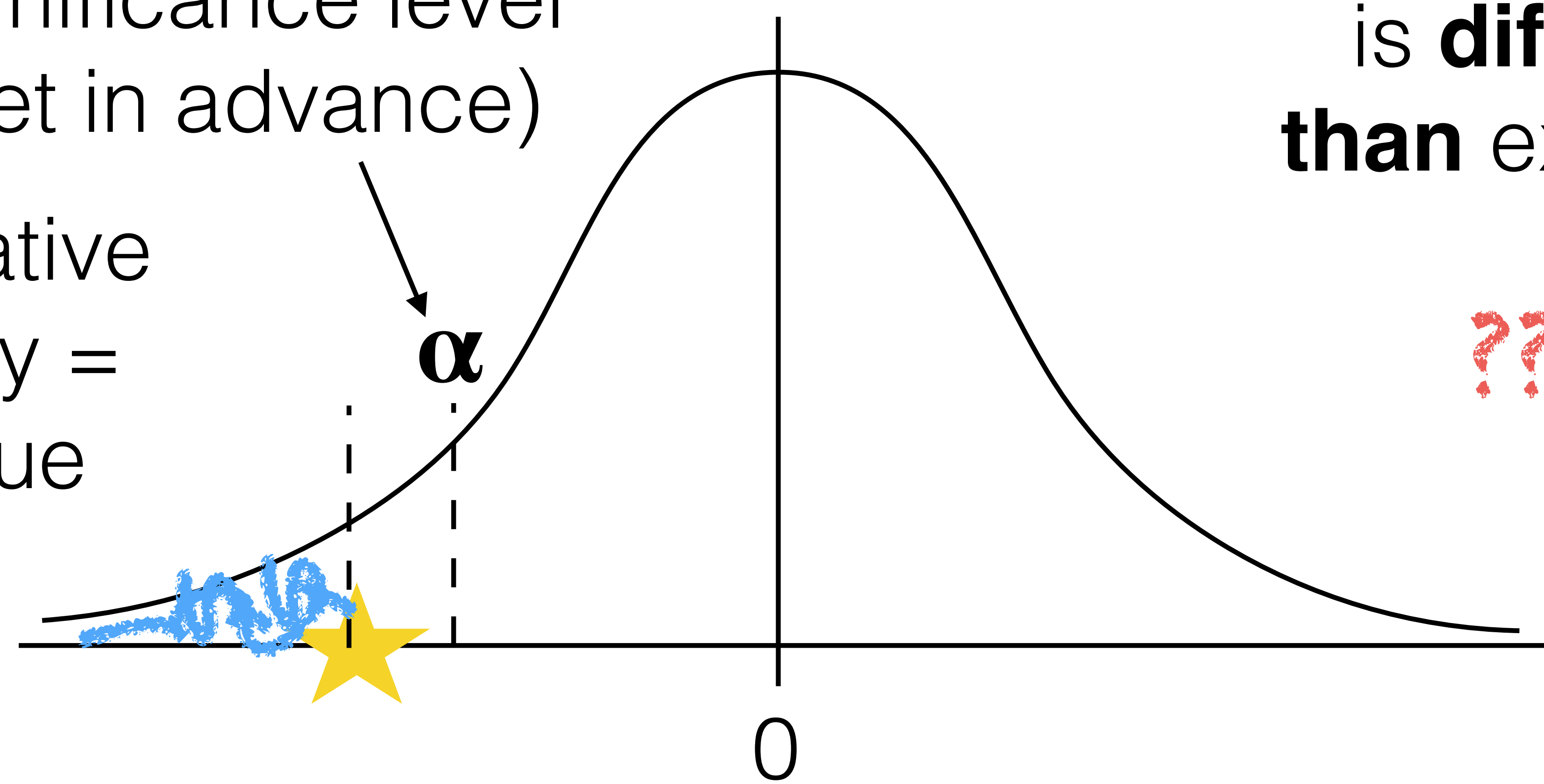
z = (observed - expected) / standard deviation

assuming the null hypothesis is true, you will be still be "surprised" alpha % of the time

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

significance level
(set in advance)

cumulative
density =
p-value



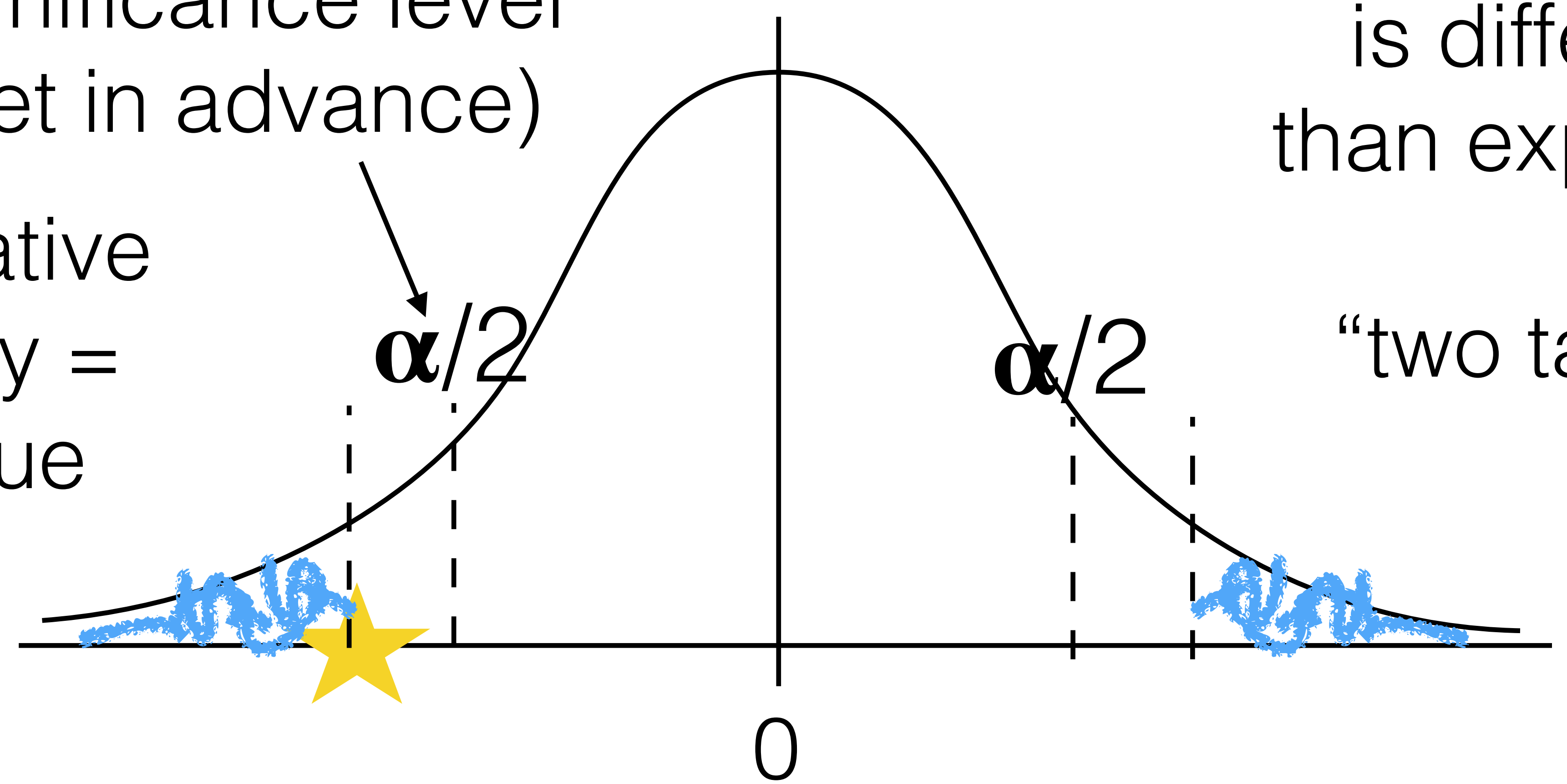
Ha = test stat
is **different**
than expected

???

z = (observed - expected) / standard deviation

assuming the null hypothesis is true, you will be still be "surprised" alpha % of the time

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$



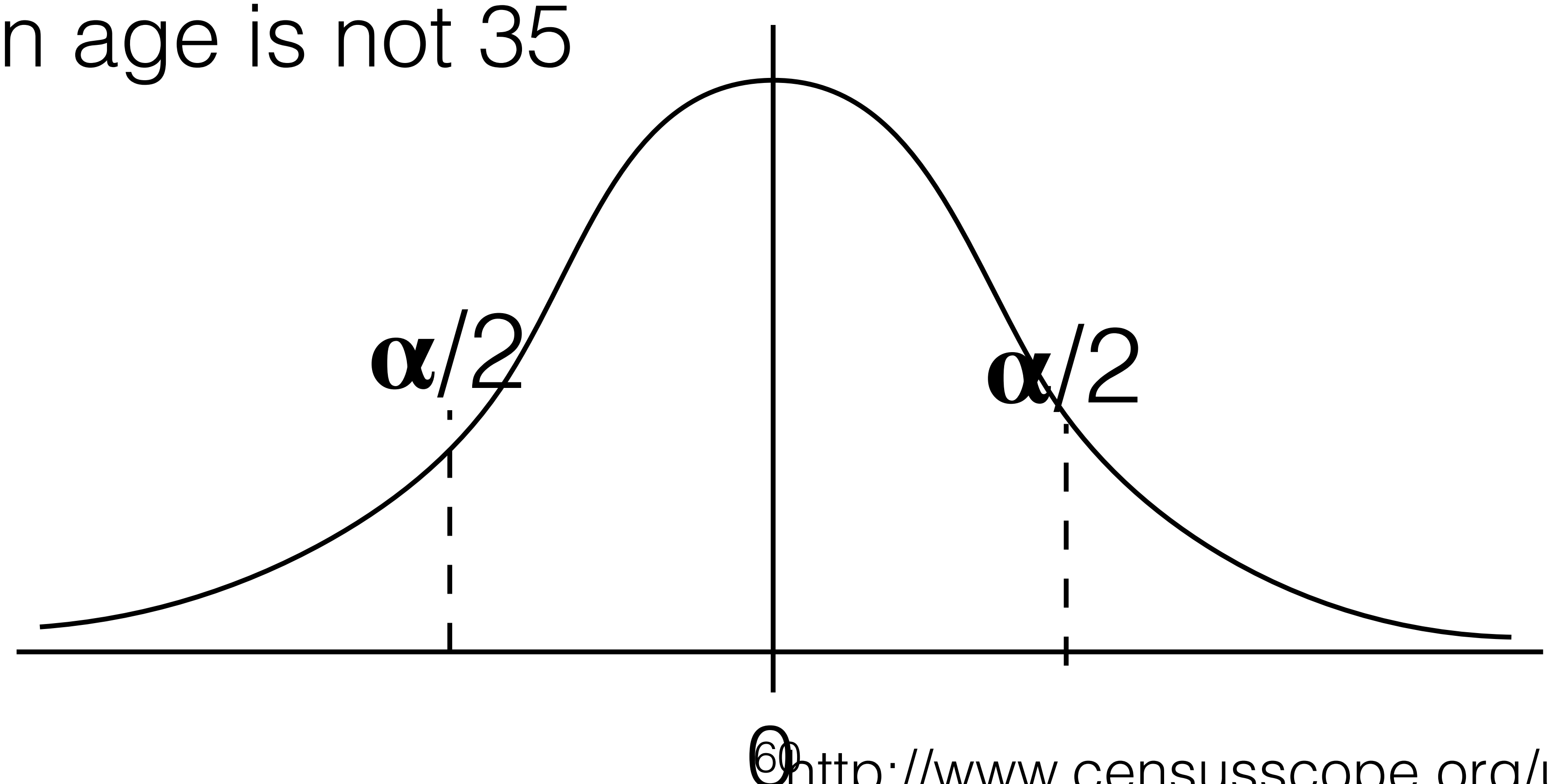
z = (observed - expected) / standard deviation

Test for population means

$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

H_0 = mean age is 35

H_a = mean age is not 35



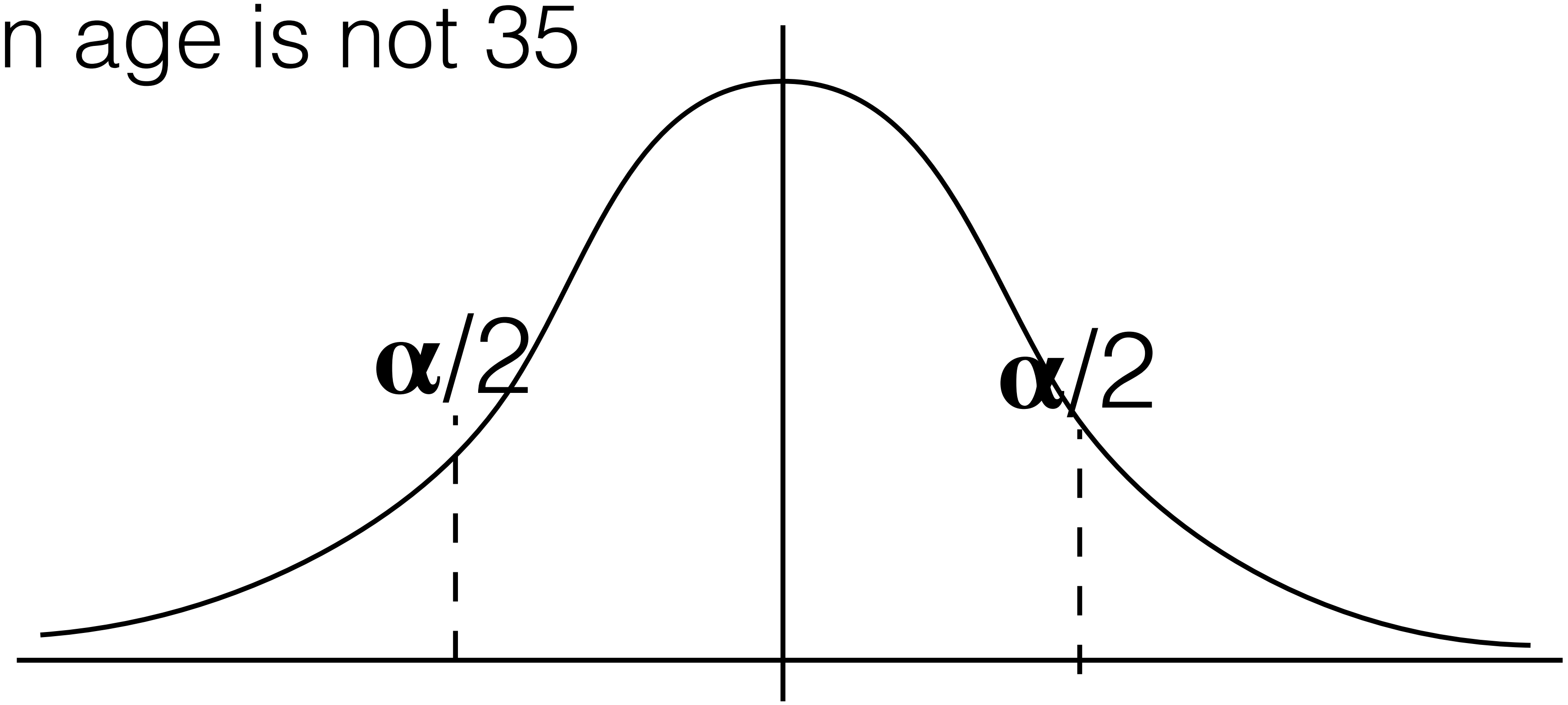
http://www.censusscope.org/us/chart_age.html

Test for population means

$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

H_0 = mean age is 35

H_a = mean age is not 35



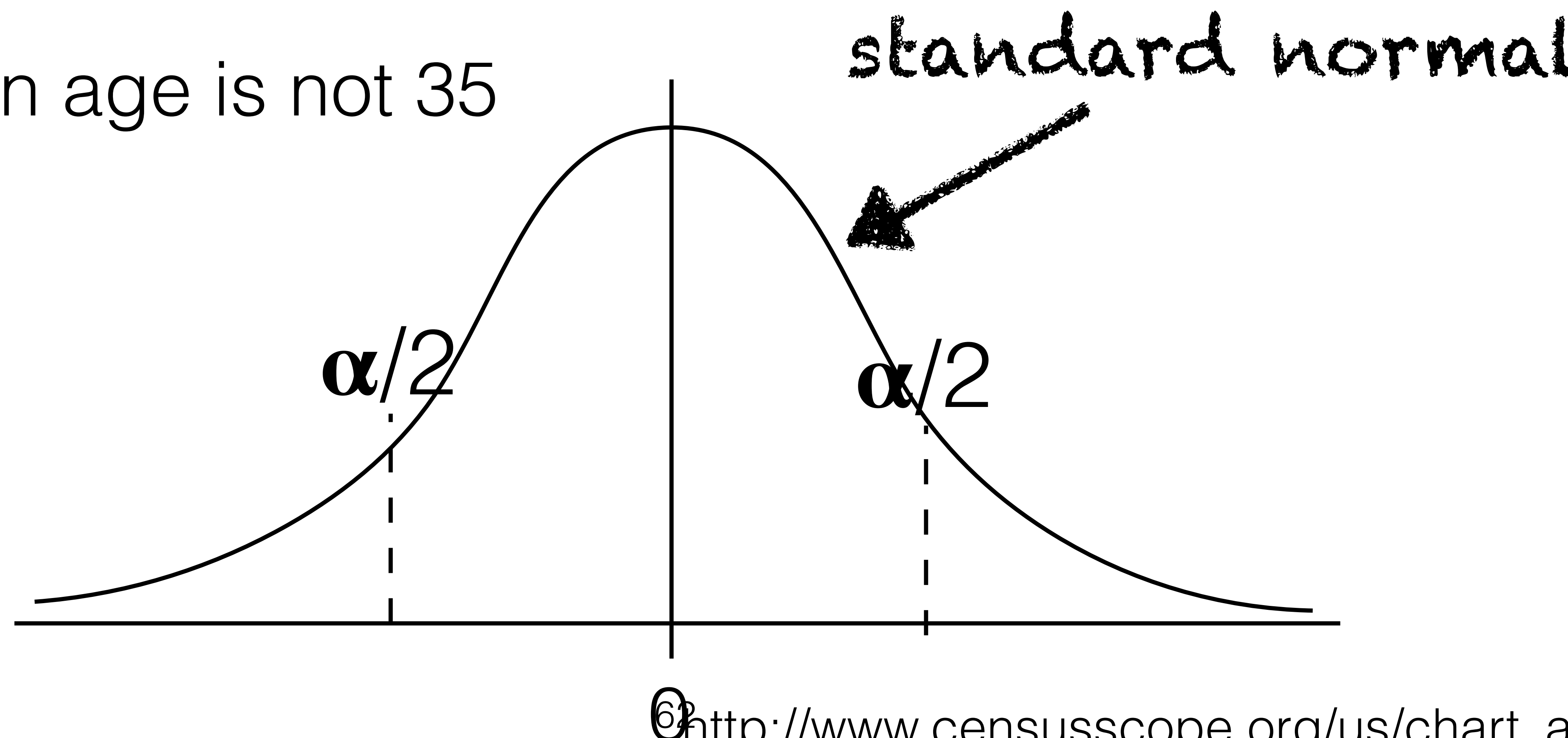
http://www.censusscope.org/us/chart_age.html

Test for population means

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

H_0 = mean age is 35

H_a = mean age is not 35



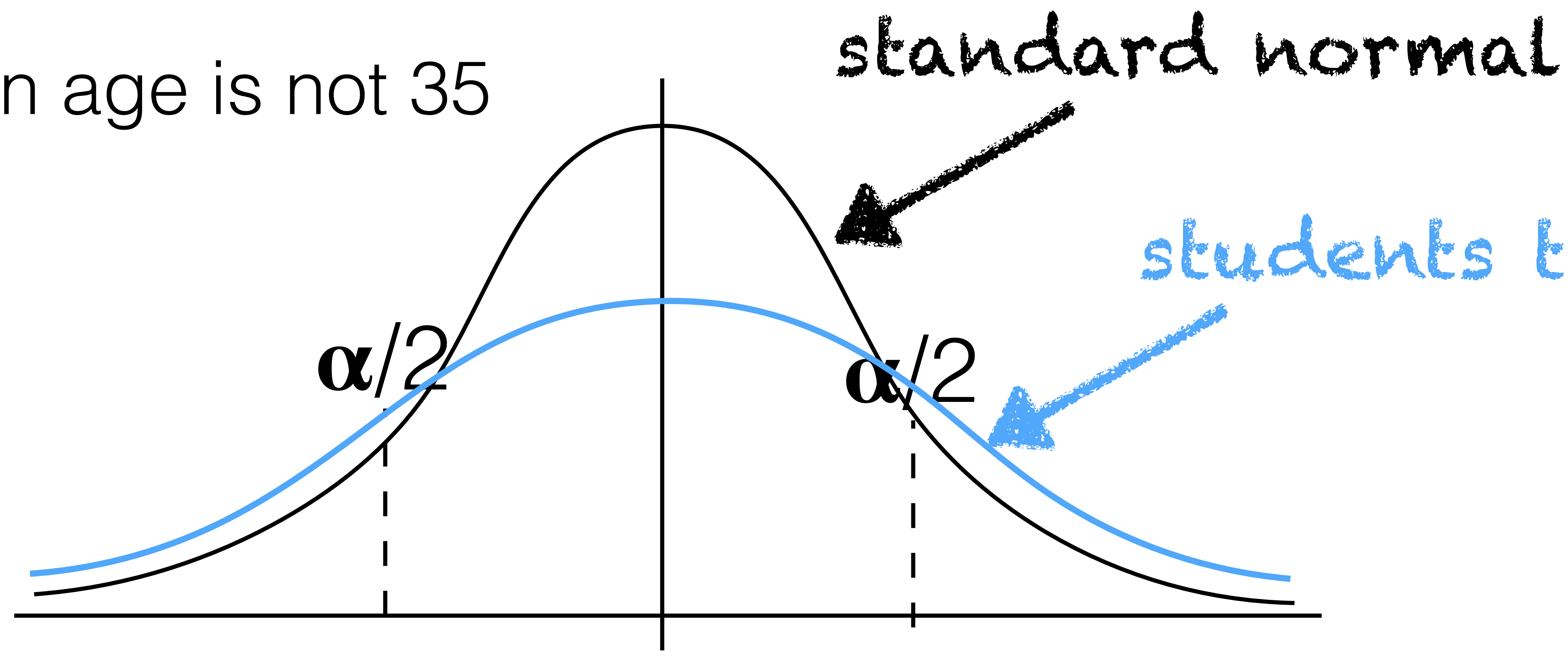
http://www.censusscope.org/us/chart_age.html

Test for population means

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

H_0 = mean age is 35

H_a = mean age is not 35



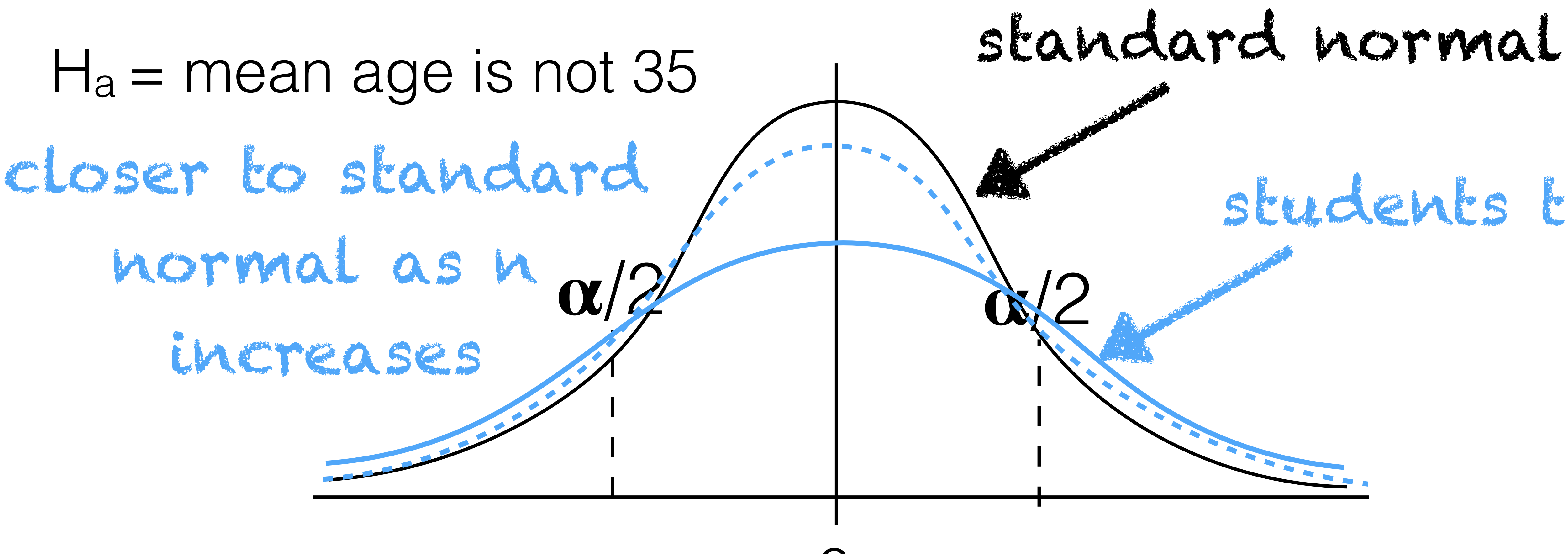
http://www.censusscope.org/us/chart_age.html

Test for population means

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

H_0 = mean age is 35

H_a = mean age is not 35



http://www.censusscope.org/us/chart_age.html

Is there evidence to reject the null hypothesis?

Null Hypothesis: The average grade is 85%.

Grades
90
92
80
87
98
78

Clicker Question!

Clicker Question!

What is the mean?

Null Hypothesis: The average grade is 85%.

Grades
90
92
80
87
98
78

a) 85
b) 87.5
c) 90
d) 92.5

Clicker Question!

What is the mean?

Null Hypothesis: The average grade is 85%.

Grades
90
92
80
87
98
78

a) 85
b) 87.5
c) 90
d) 92.5

Clicker Question!

What is the standard error?

Null Hypothesis: The average grade is 85%.

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2.$$

Grades
90
92
80
87
98
78

a) $\sqrt{47.25}$
b) $\sqrt{51.3}$
c) $\sqrt{56.7}$
d) $\sqrt{57.25}$

Clicker Question!

What is the standard error?

Null Hypothesis: The average grade is 85%.

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2.$$

Grades
90
92
80
87
98
78

a) $\sqrt{47.25}$

b) $\sqrt{51.3}$

c) $\sqrt{56.7}$

d) $\sqrt{57.25}$

Clicker Question!

What is the test statistic?

Null Hypothesis: The average grade is 85%.

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Grades
90
92
80
87
98
78

mean: 87.5
s: 7.5

a) 0.82
b) 0.85
c) 0.91
d) 0.95

Clicker Question!

What is the test statistic?

Null Hypothesis: The average grade is 85%.

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Grades
90
92
80
87
98
78

mean: 87.5
s: 7.5

a) 0.82
b) 0.85
c) 0.91
d) 0.95

Is there evidence to reject the null hypothesis?

Null Hypothesis: The average grade is 85%.

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

mean: 87.5
s: 7.5

Grades
90
92
80
87
98
78

a) 0.82
b) 0.85
c) 0.91
d) 0.95

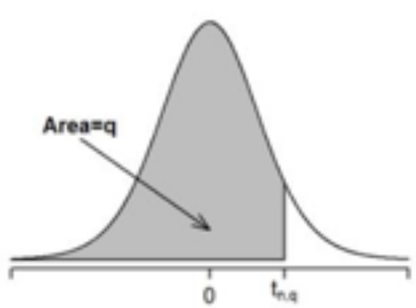
Is there evidence to reject the null hypothesis?

Null Hypothesis: The average grade is 85%.

$$\bar{x} - \mu_0$$

Quartiles of the t Distribution

The table gives the value if $t_{n;q}$ - the qth quantile of the t distribution for n degrees of freedom



	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

a) 0.82
b) 0.85
c) 0.91
d) 0.95

Is there evidence to reject the null hypothesis?

Null Hypothesis: The average grade is 85%.

$\bar{x} - \mu_0$

Quartiles of the t Distribution

The table gives the value if $t_{n;q}$ - the qth quantile of the t distribution for n degrees of freedom

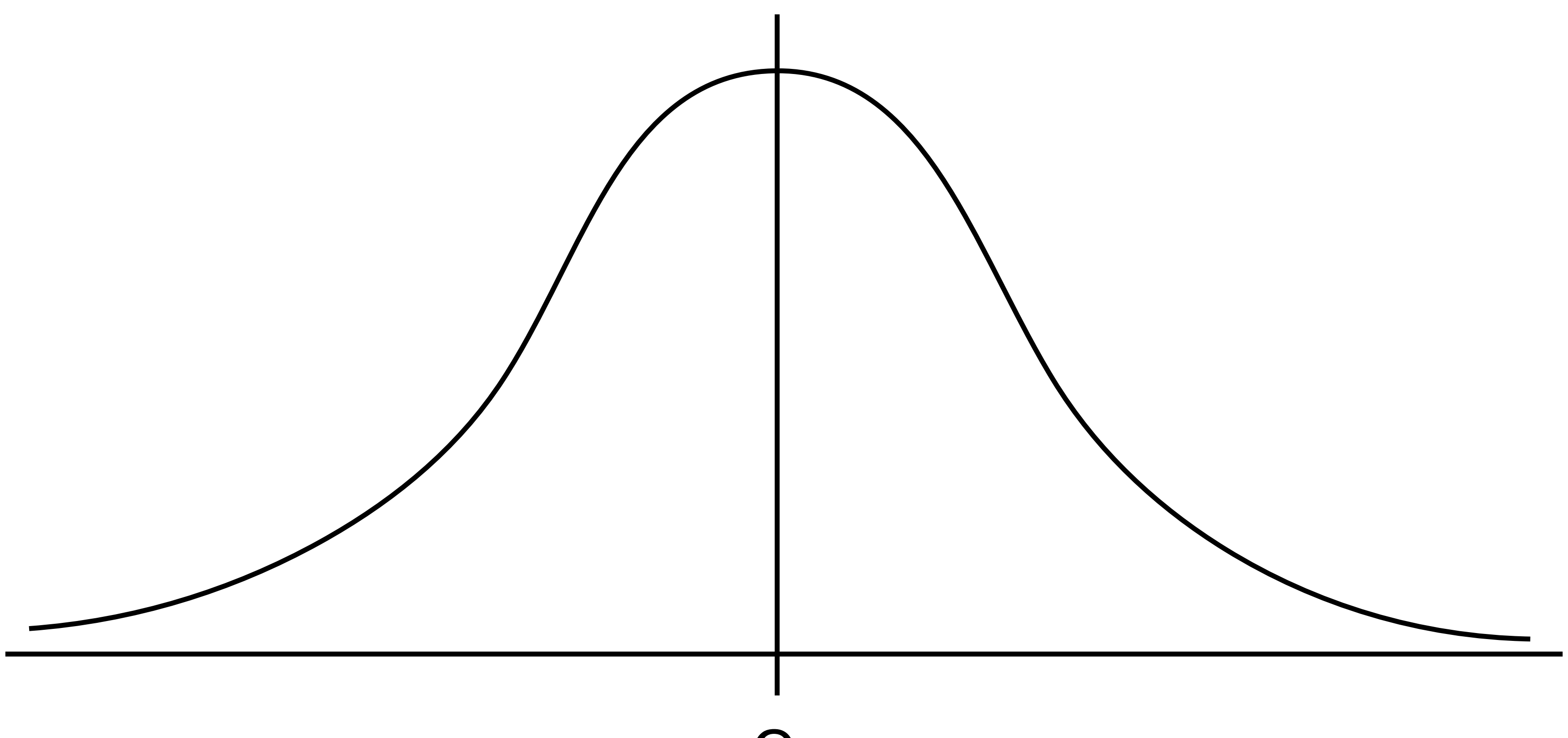
ttest_1samp

	q = 0.6	0.75	0.9	0.95	[illegible]	[illegible]	[illegible]	[illegible]	0.999	0.9995
n = 1	0.3249	1.0000	3.078	6.314	12.706	31.821	[illegible]	[illegible]	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

a) 0.82
b) 0.85
c) 0.91
d) 0.95

Test for difference in means

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

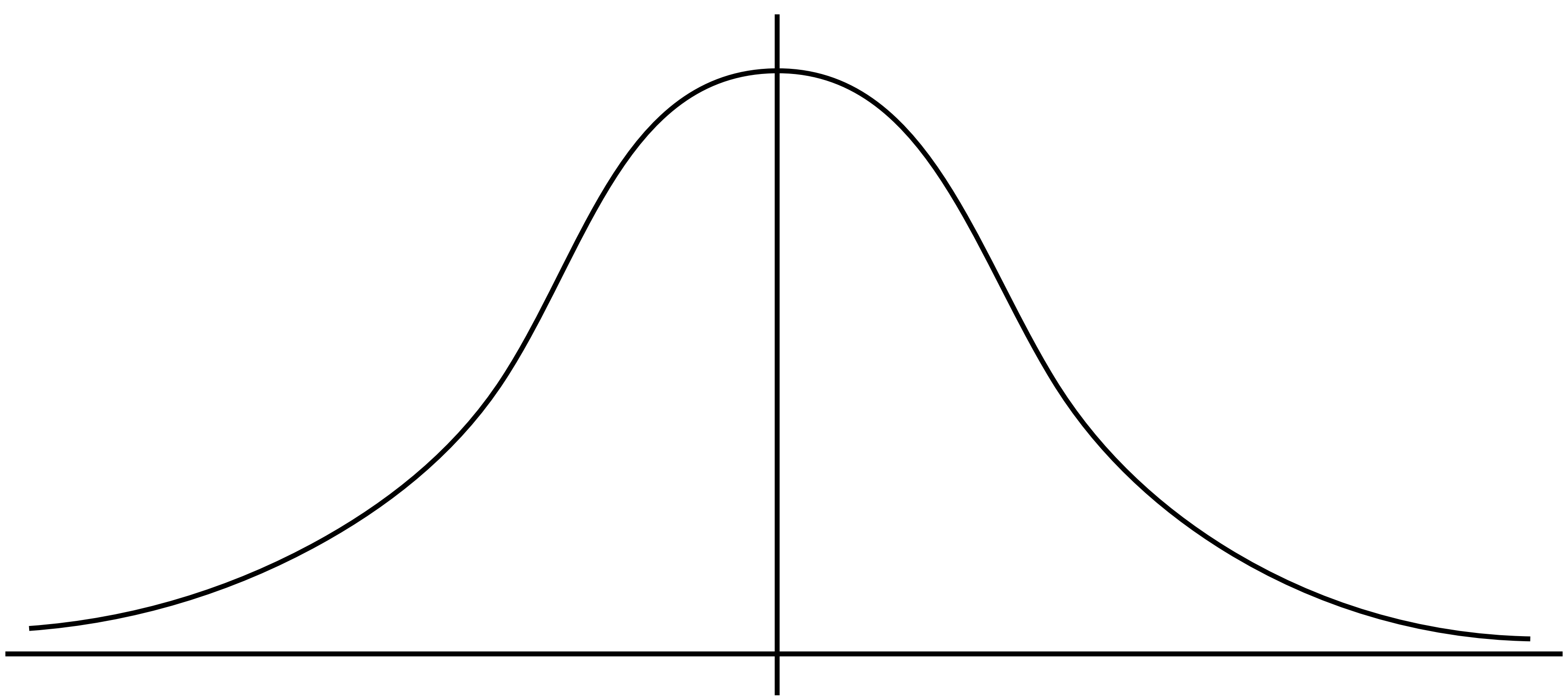


http://www.censusscope.org/us/chart_age.html

Test for difference in means

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

mean 1 - mean 2



0

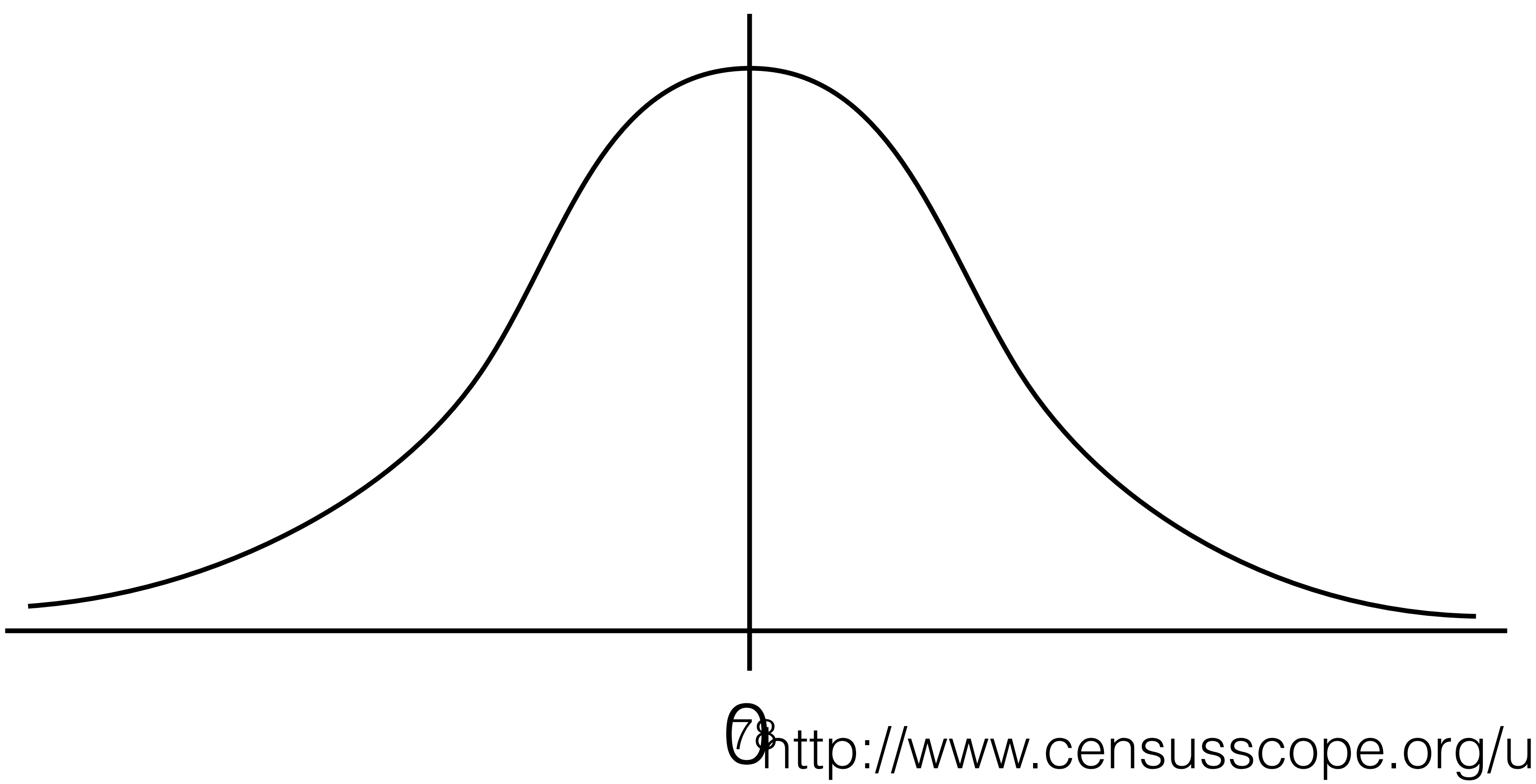
http://www.censusscope.org/us/chart_age.html

Test for difference in means

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

mean 1 - mean 2

zero



http://www.censusscope.org/us/chart_age.html

Is there evidence to reject the null hypothesis?

Null Hypothesis: The average grade is the same as last year.
Alt. Hypothesis: This year's grades are higher.

2019	2020
90	95
92	92
80	83
87	87
98	98
78	75

Is there evidence to reject the null hypothesis?

Null Hypothesis: The average grade is the same as last year.
Alt. Hypothesis: This year's grades are higher.

2019	2020
90	95
92	92
80	83
87	87
98	98
78	75

ttest_ind

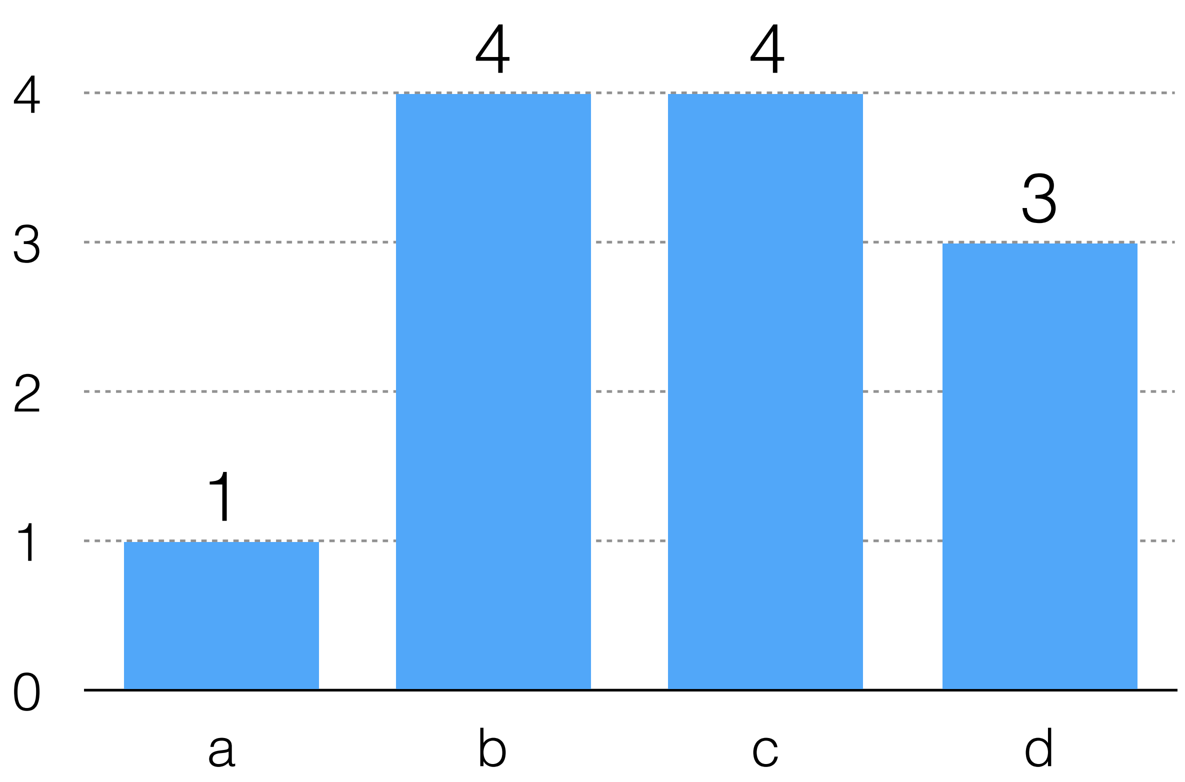
Some tests you are likely to use

- t-test: difference of means; is the average value of some feature different between two populations
  - e.g. are men taller than women, are blue states more populated than red states, do CS work harder than other majors (::rolling_eyes::)
- **chi-squared test: difference in frequencies of a categorical variable; is the distribution of some feature uniform across groups**
  - e.g. do neighborhoods differ in terms of music preferences features; do college majors differ in terms of sociodemographic features

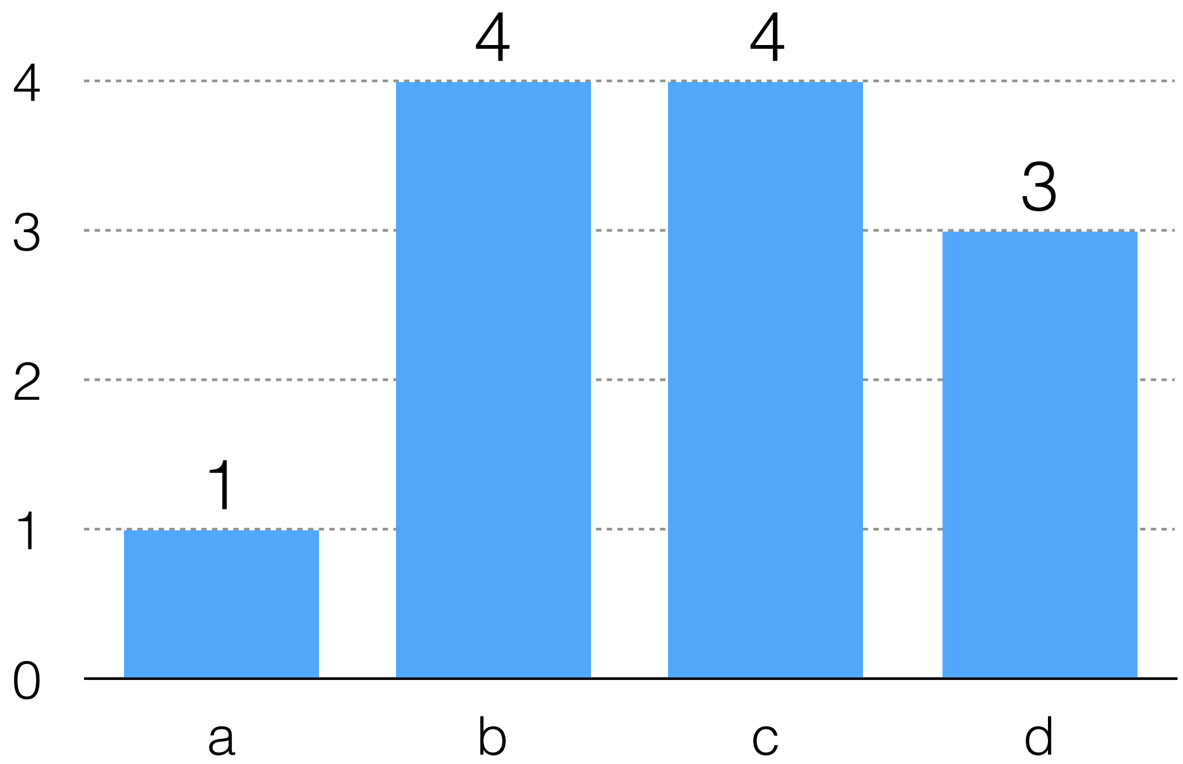
Are the answers to my clicker questions random?

c	d	b	d
b	a	d	c
b	c	b	d

Are the answers to my clicker questions random?



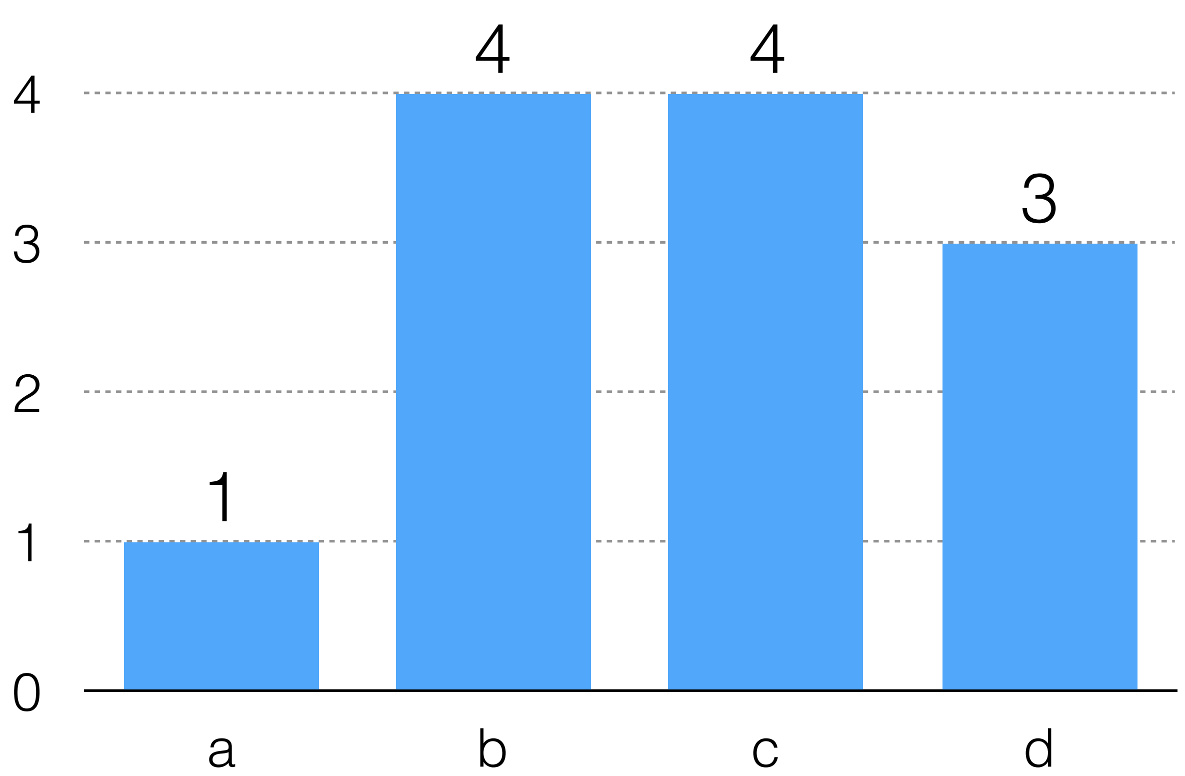
Are the answers to my clicker questions random?



Xi = count of answer i

p(a) = p(b) = p(c) = p(d) = 0.25

Are the answers to my clicker questions random?

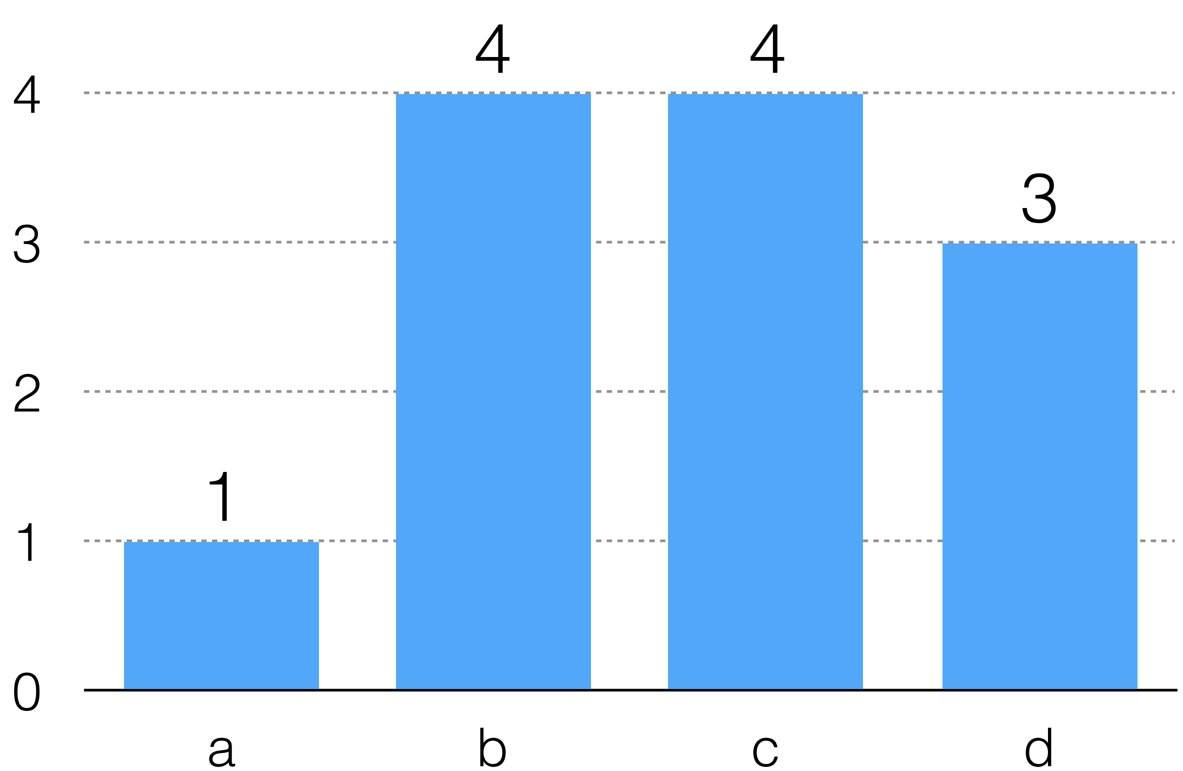


Is this distribution significantly different than what we would expect by chance, assuming that in fact all answers are equally likely?

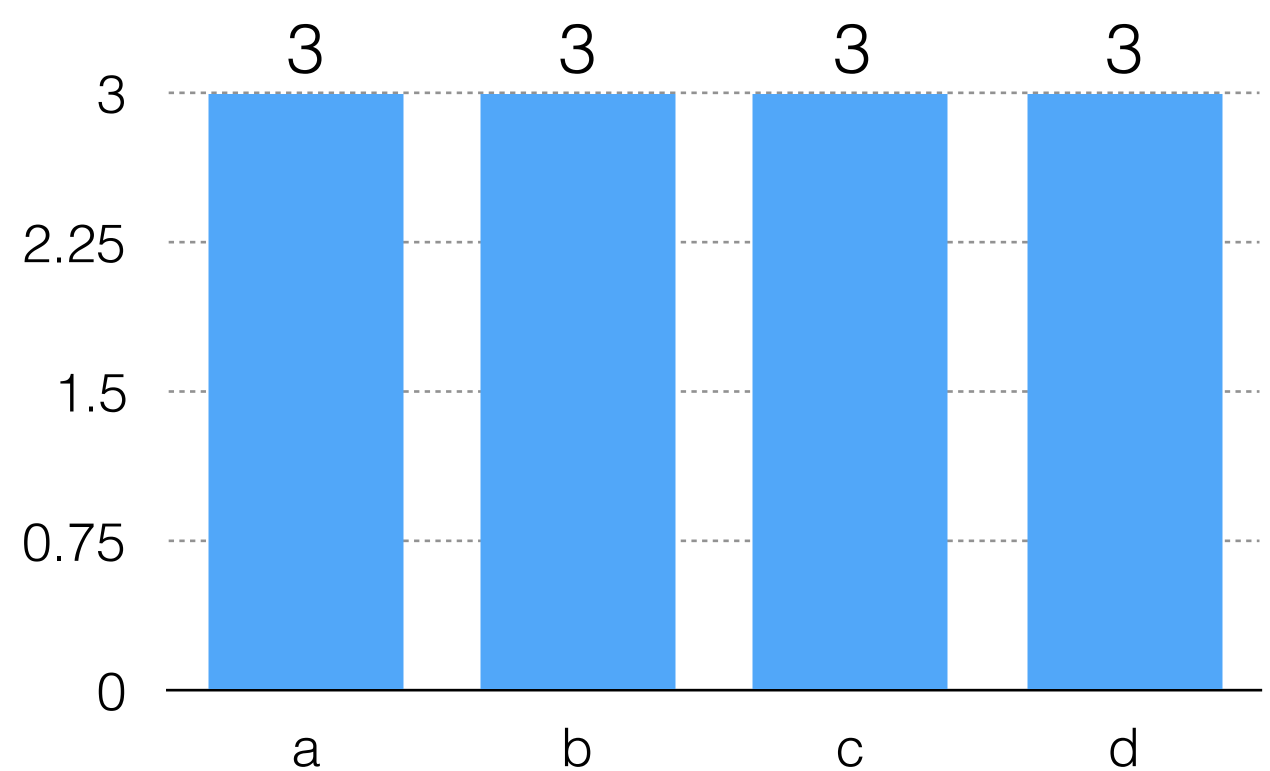
Xi = count of answer i

p(a) = p(b) = p(c) = p(d) = 0.25

Are the answers to my clicker questions random?

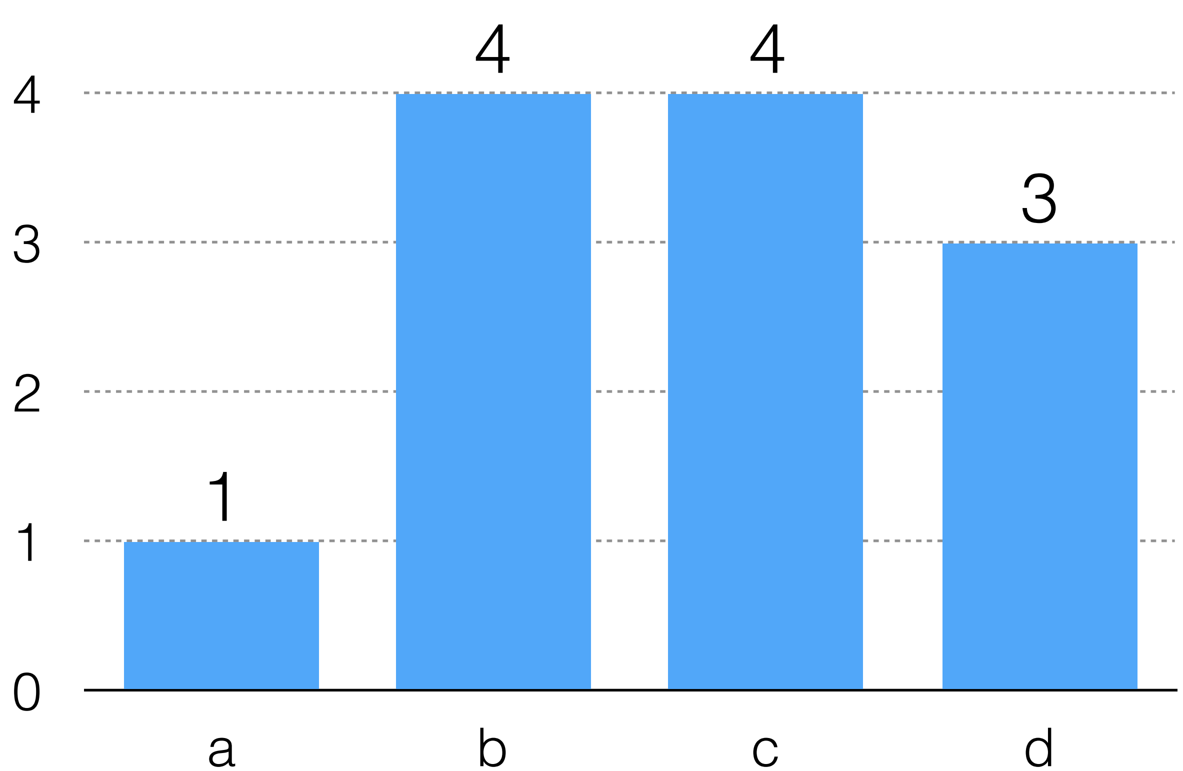


Observed

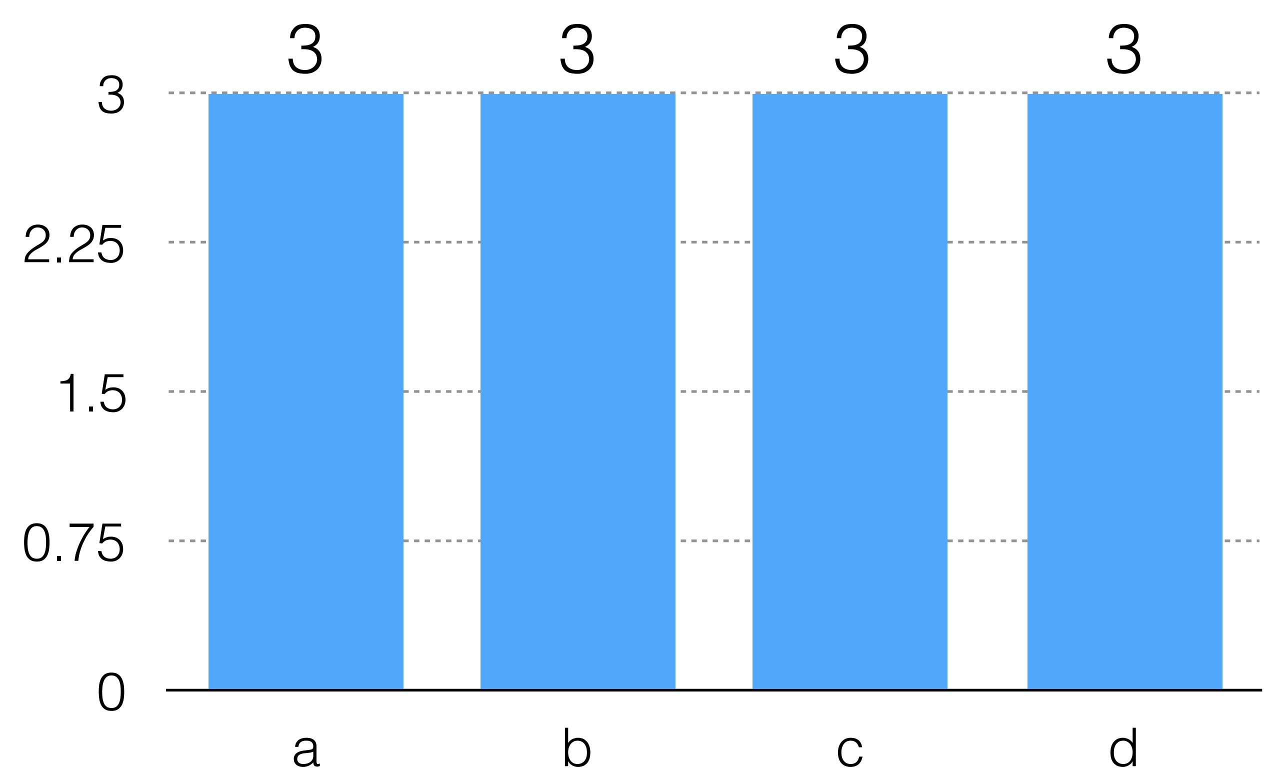


Expected

Are the answers to my clicker questions random?



Observed

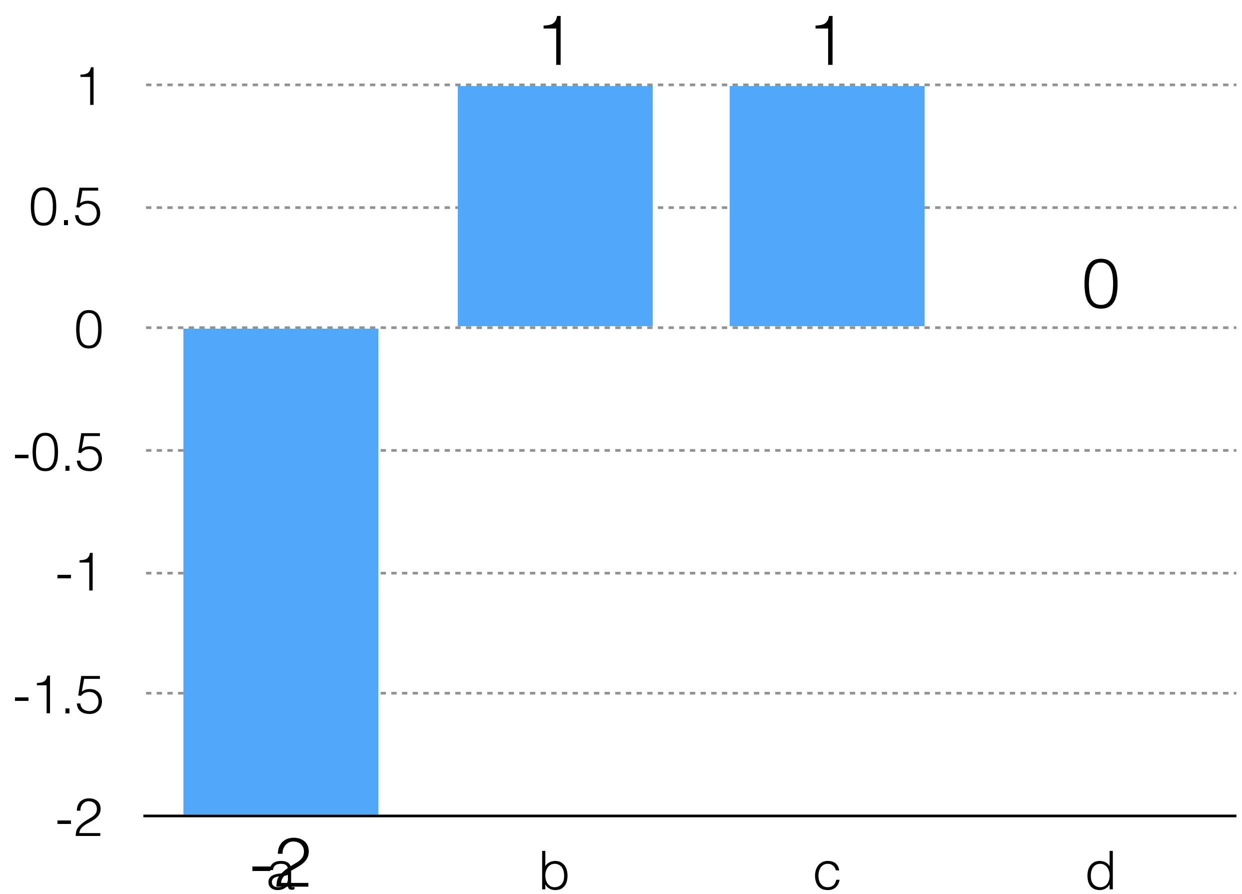


Expected

Want to model the difference between these

Are the answers to my clicker questions random?

$$X_i - E(X_i)$$



Clicker Question!

Clicker Question!

Should I use the total difference between observed and expected as my summary statistic? I.e.

$$\sum_i (X_i - E(X_i))$$

a) Yes! That sounds good.
b) No! I have qualms...

Clicker Question!

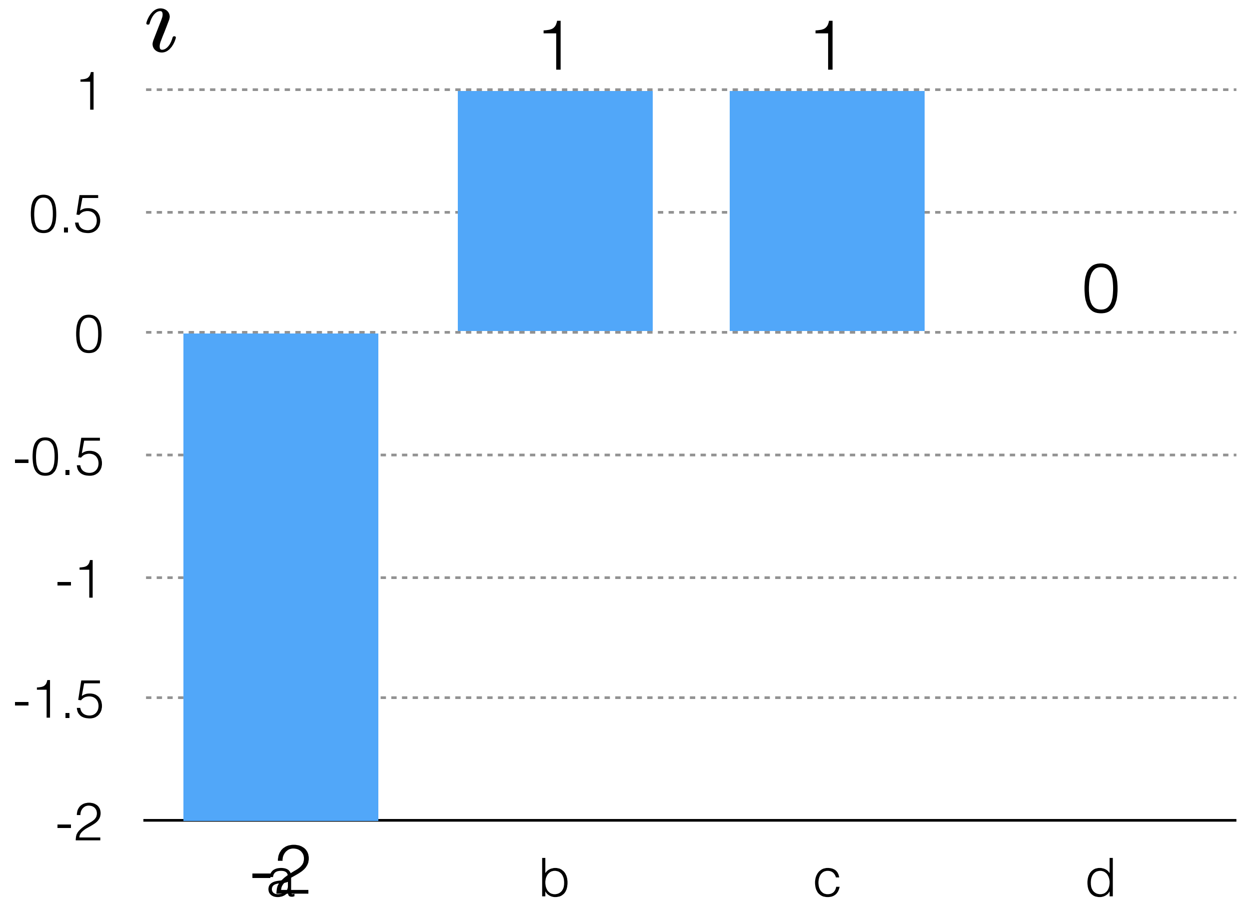
Should I use the total difference between observed and expected as my summary statistic? I.e.

$$\sum_i (X_i - E(X_i))$$

a) Yes! That sounds good.
b) No! I have qualms...

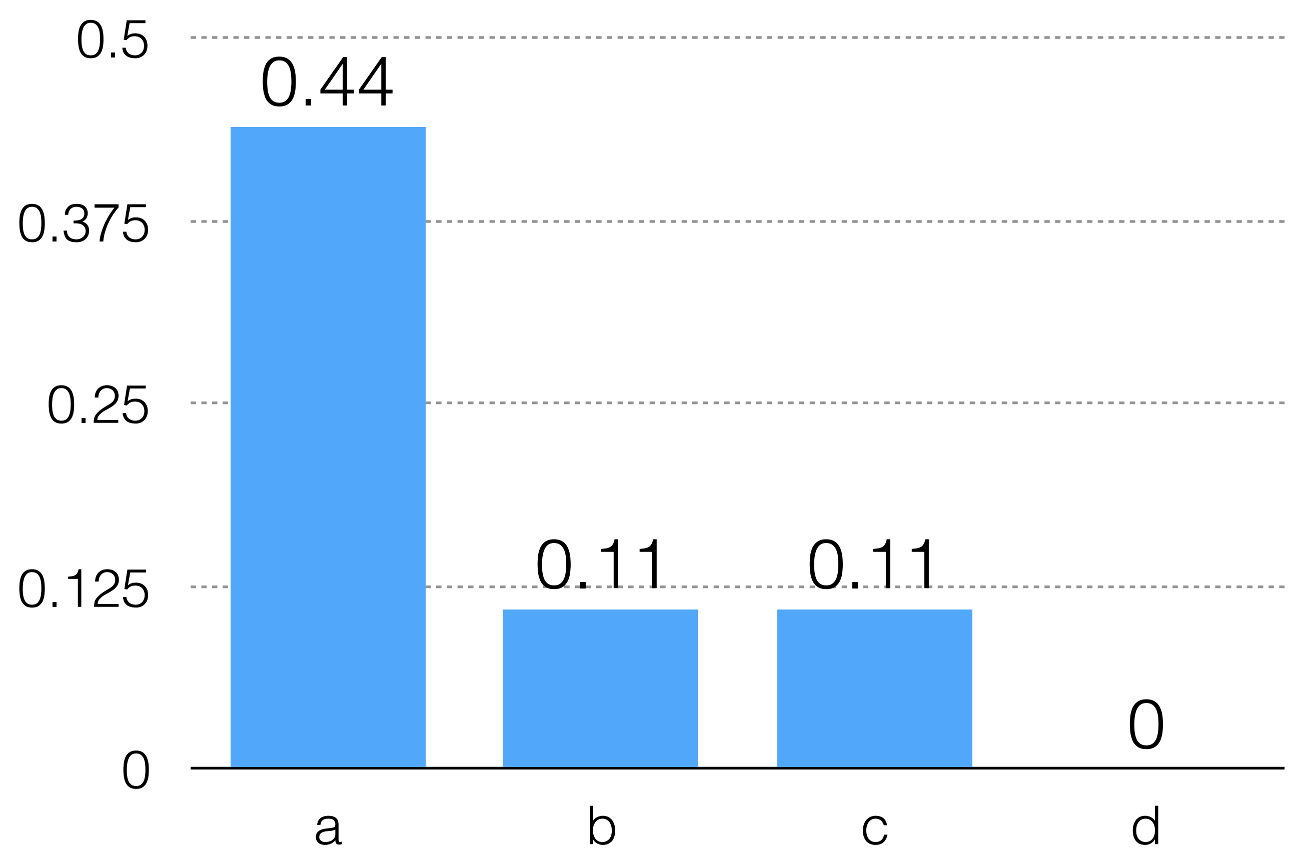
Are the answers to my clicker questions random?

$$\sum_i (X_i - E(X_i)) = 0$$



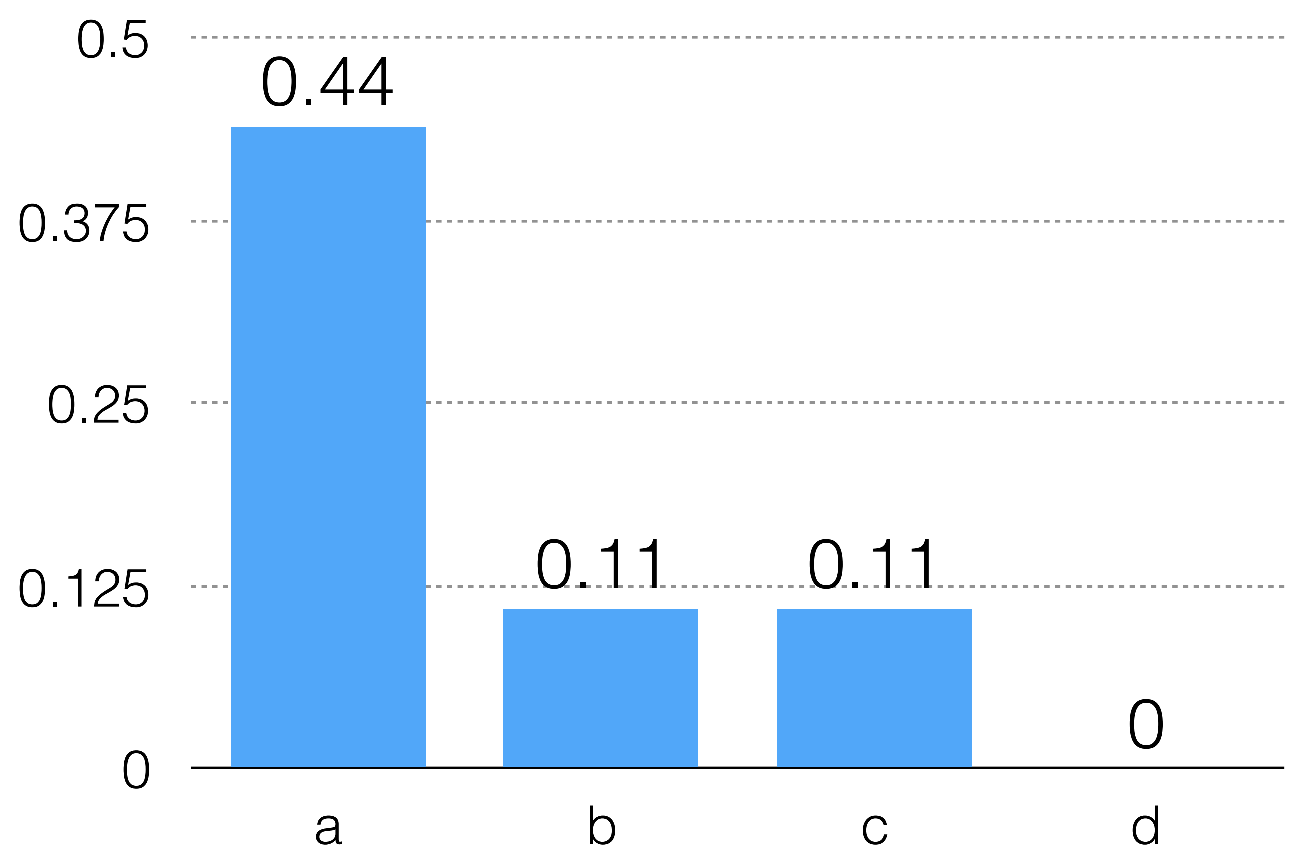
Are the answers to my clicker questions random?

$$\sum_i (X_i - E(X_i))^2$$



Are the answers to my clicker questions random?

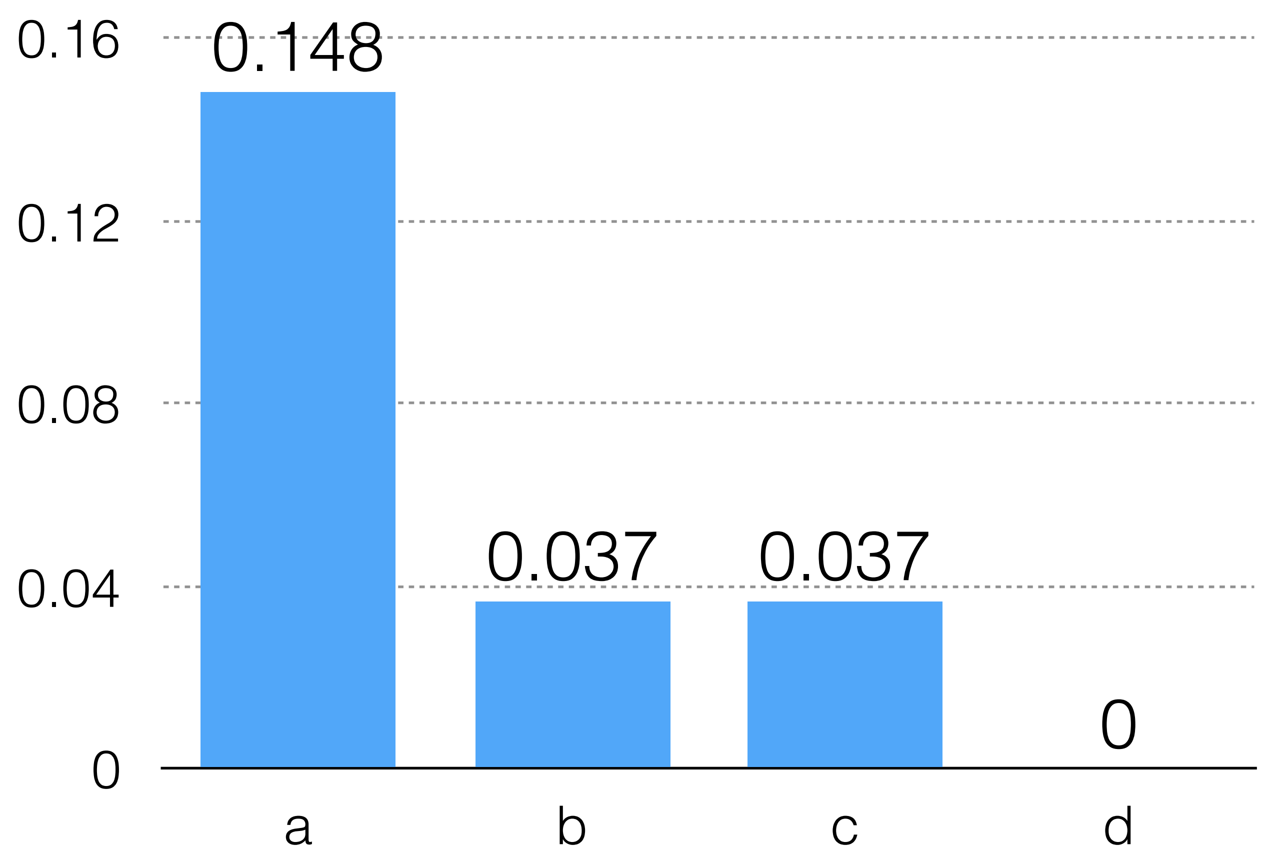
$$\sum_i (X_i - E(X_i))^2$$



Thoughts?

Are the answers to my clicker questions random?

$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$

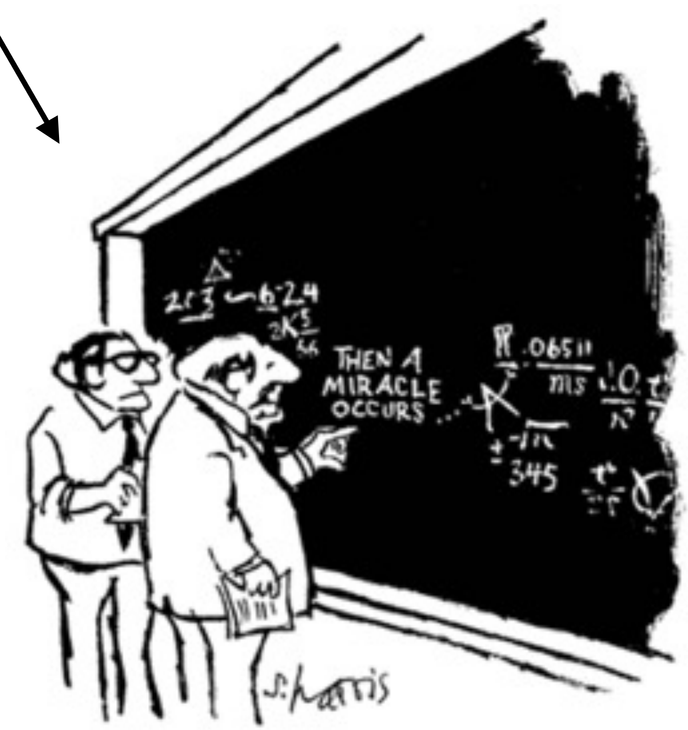


Are the answers to my clicker questions random?

$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$

Are the answers to my clicker questions random?

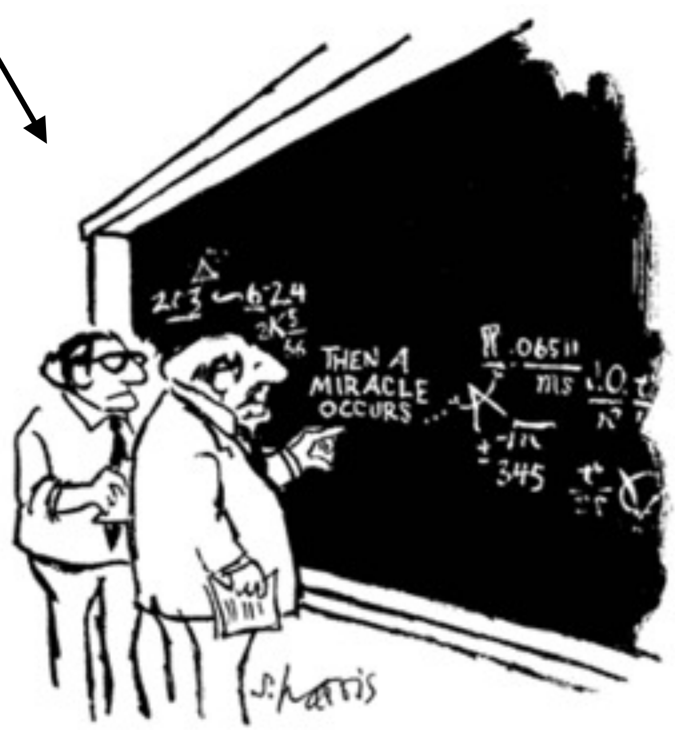
$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$



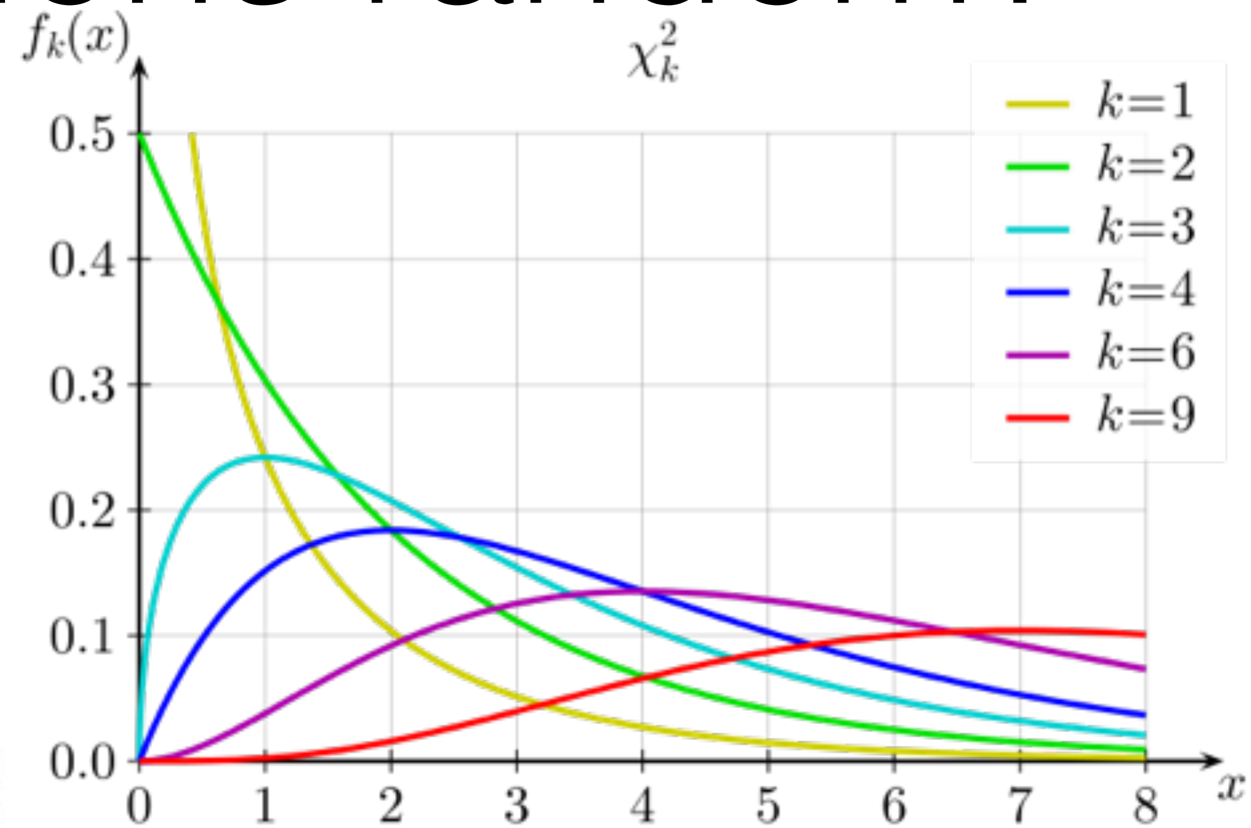
"I think you should be more explicit here in step two."

Are the answers to my clicker questions random?

$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$

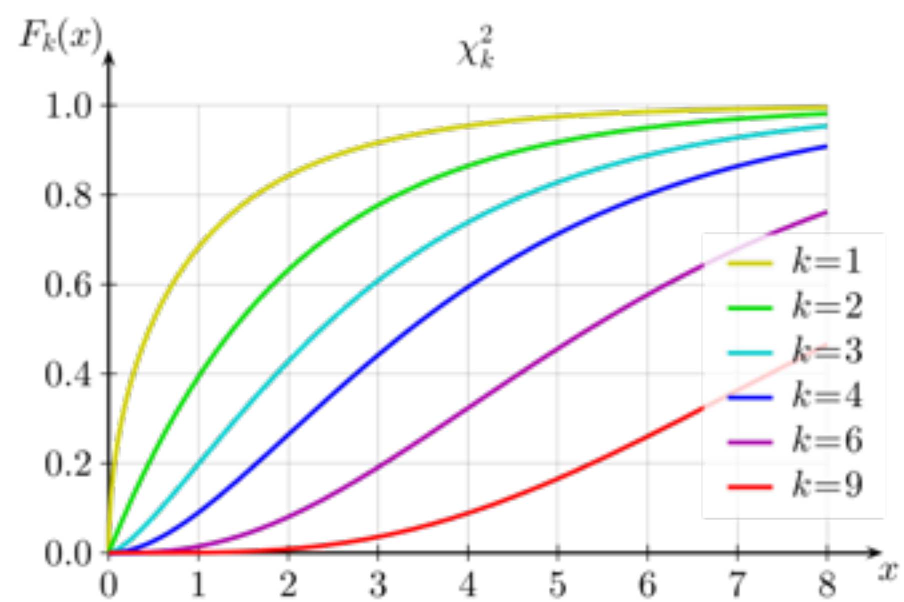


"I think you should be more explicit here in step two."



Chi Squared Test

$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$

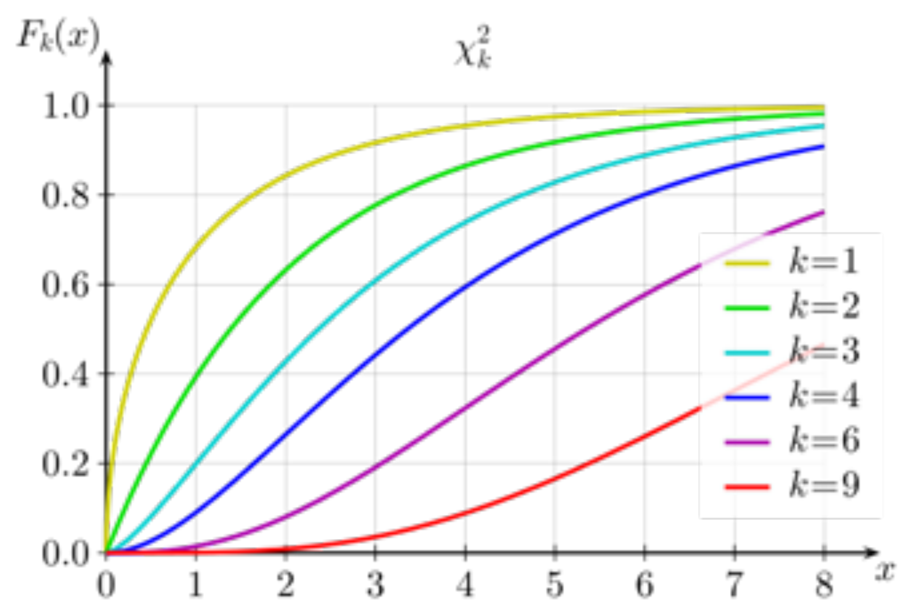


$$\frac{1}{\Gamma(k/2)}\gamma(\frac{k}{2}, \frac{x}{2})$$

Chi Squared Test

$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$

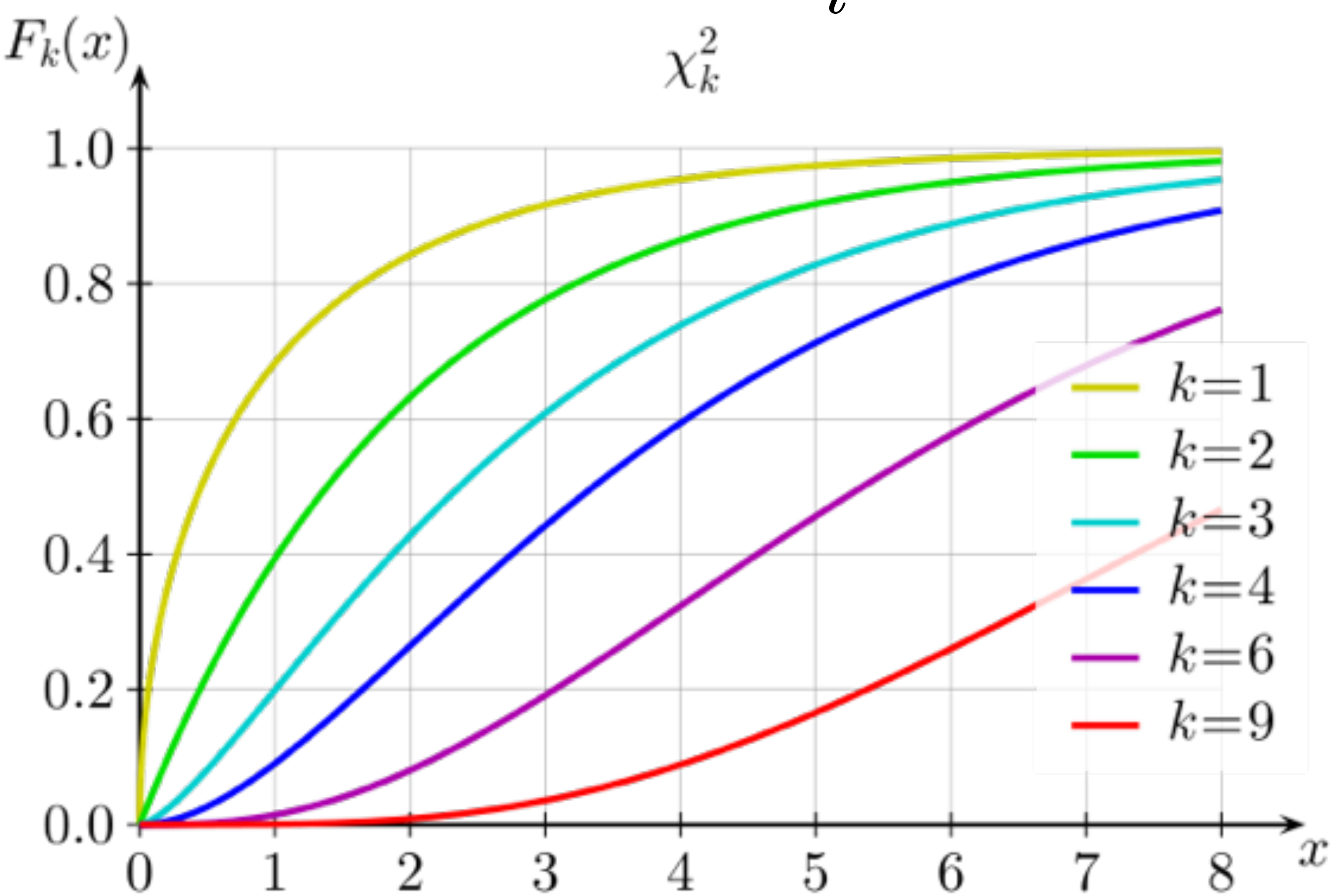
cdf that we can compute explicitly

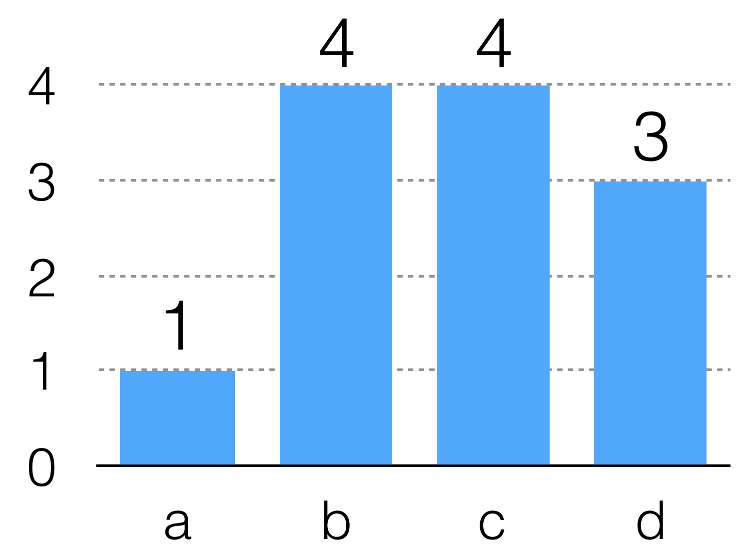


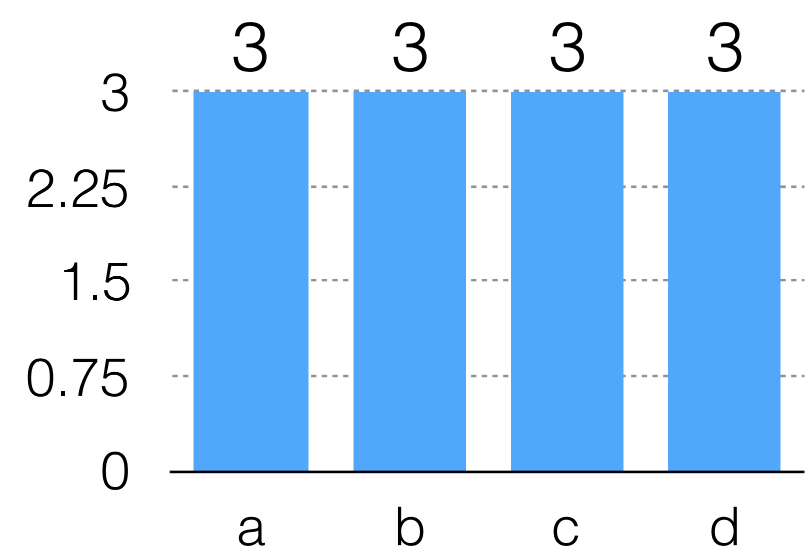
$$\frac{1}{\Gamma(k/2)}\gamma\left(\frac{k}{2}, \frac{x}{2}\right)$$

Chi Squared Test

$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$



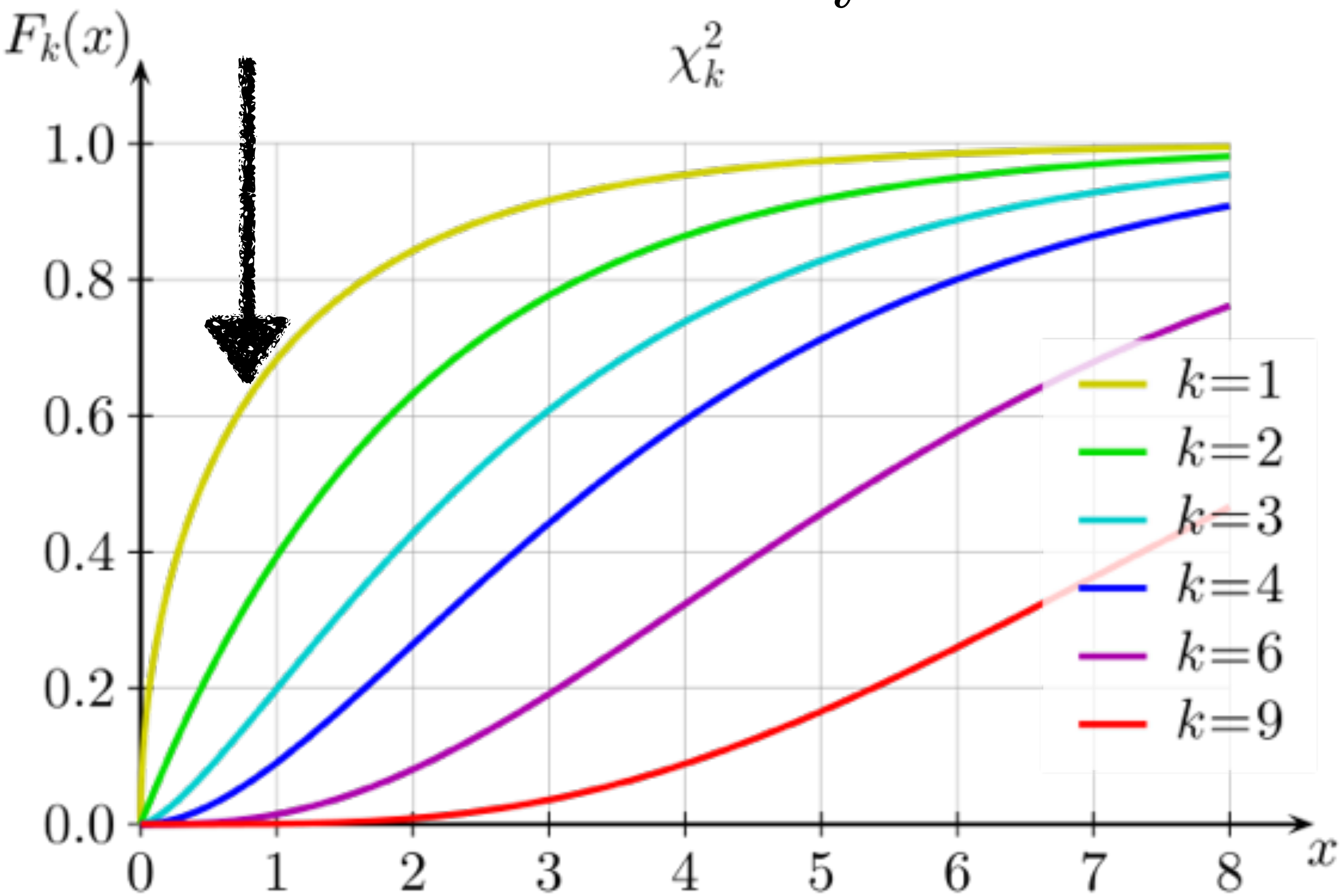


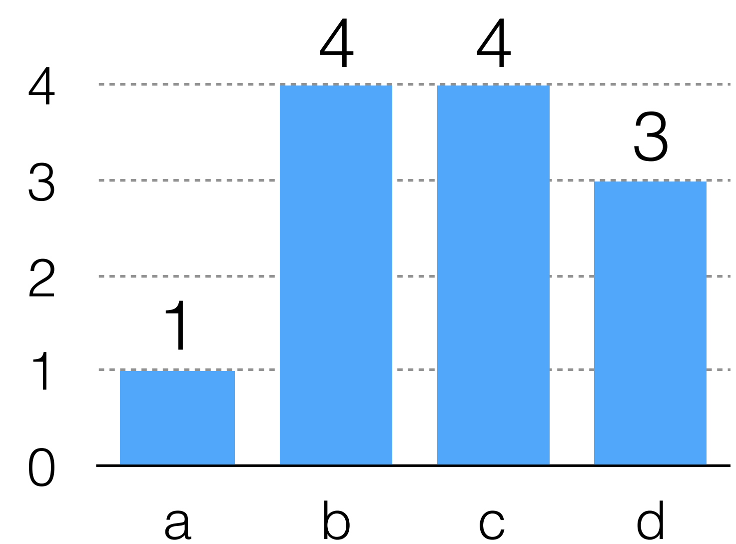


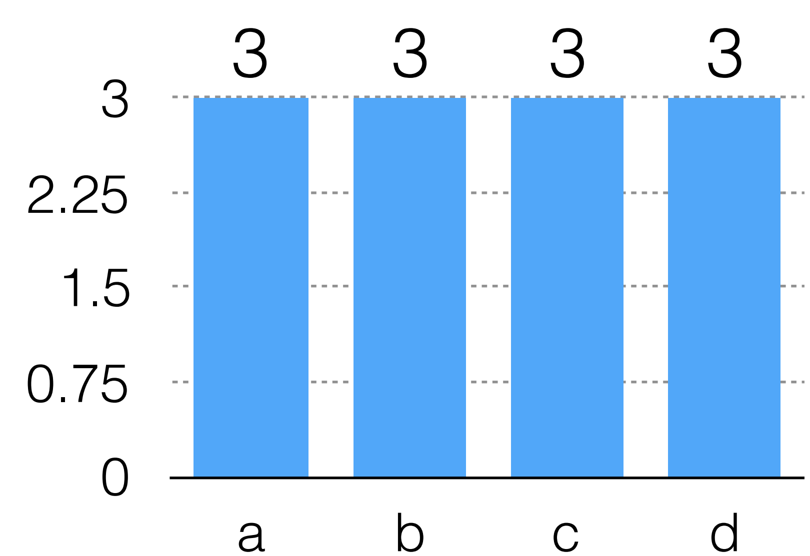
Chi Squared Test

not really remarkable

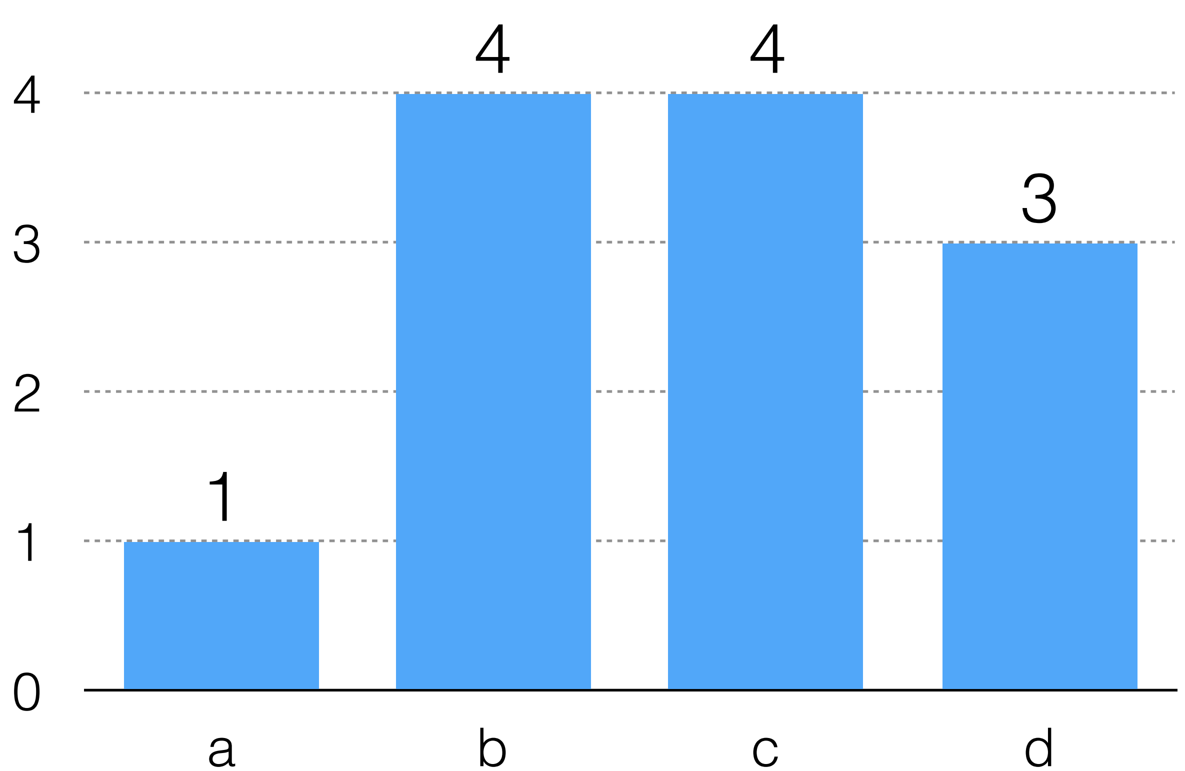
$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$



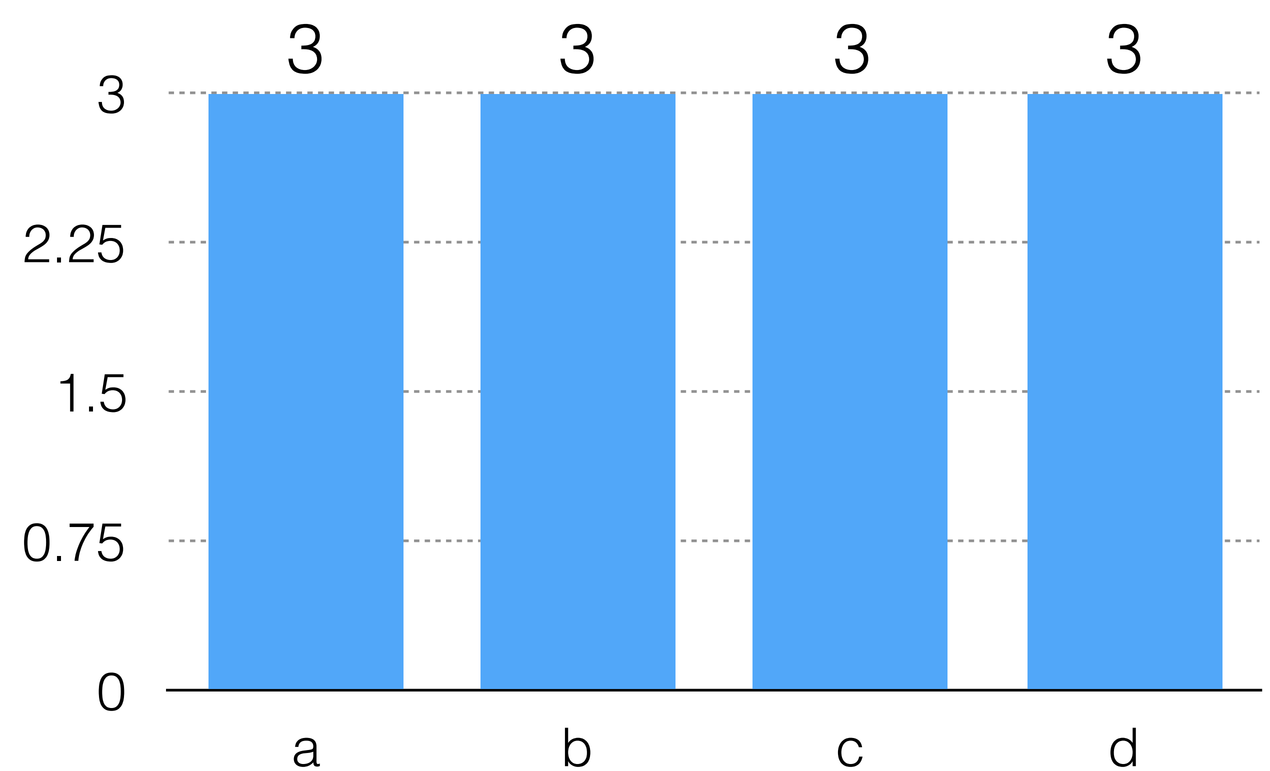




Chi Squared Test



Observed



Expected

Want to model the difference between these

Chi Squared Test

Overall goal: Understand how cities differ in terms of professions.

Null Hypothesis: No difference between Providence and Boston

	Art	Tech	Medicine
PVD	35	33	32
Boston	25	30	45

Chi Squared Test

Overall goal: Understand how cities differ in terms of professions.

Null Hypothesis: No difference between Providence and Boston

"expected"

	Art	Tech	Medicine
PVD	35	33	32
Boston	25	30	45

"observed"

Clicker Question!

Clicker Question!
Compute the test statistic

$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$

Null Hypothesis: No difference between Providence and Boston

"expected"	Art	Tech	Medicine
PVD	35	33	32
Boston	25	30	45

"observed"

a) 6.5
b) 7.3
c) 8.4

$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$

Clicker Question!
Compute the test statistic

Null Hypothesis: No difference between Providence and Boston

"expected"	Art	Tech	Medicine
PVD	35	33	32
Boston	25	30	45

"observed"

a) 6.5
b) 7.3
c) 8.4

$$\sum_i \frac{(X_i - E(X_i))^2}{E(X_i)}$$

Clicker Question!
Compute the test statistic

chisquare

Null Hypothesis: No difference betwe
Providence and Boston

"expected"	Art	Tech	Medicine
PVD	35	33	32
Boston	25	30	45

"observed"

a) 6.5
b) 7.3
c) 8.4

okie done now