Hypotheses and Hypothesis Testing Intro

February 20, 2020
Data Science CSCI 1951A
Brown University
Instructor: Ellie Pavlick
HTAs: Josh Levin, Diane Mutako, Sol Zitter

Announcements

- Final Project Pitches and Feedback
- Data Deliverable—***not*** a ceremonial checkin, please start soon!
- Grades, regrades—read policy on Piazza
- Map Reduce out later today—follow announcements about the cluster
- Map Reduce lab released today, no sections until next week. One optional lab.

Today

- What is a hypothesis?
- Some definitions/notation
- Intuition behind modeling/hypothesis testing

Today

Not to be dramatic, but...

THIS IS PROBABLY THE MOST IMPORTANT LECTURE IN THE WHOLE COURSE!!!!!

- What is a hypothesis?
- Some definitions/notation
- Intuition behind modeling/hypothesis testing

What is a hypothesis?

- #1 most important thing: *falsifiable*

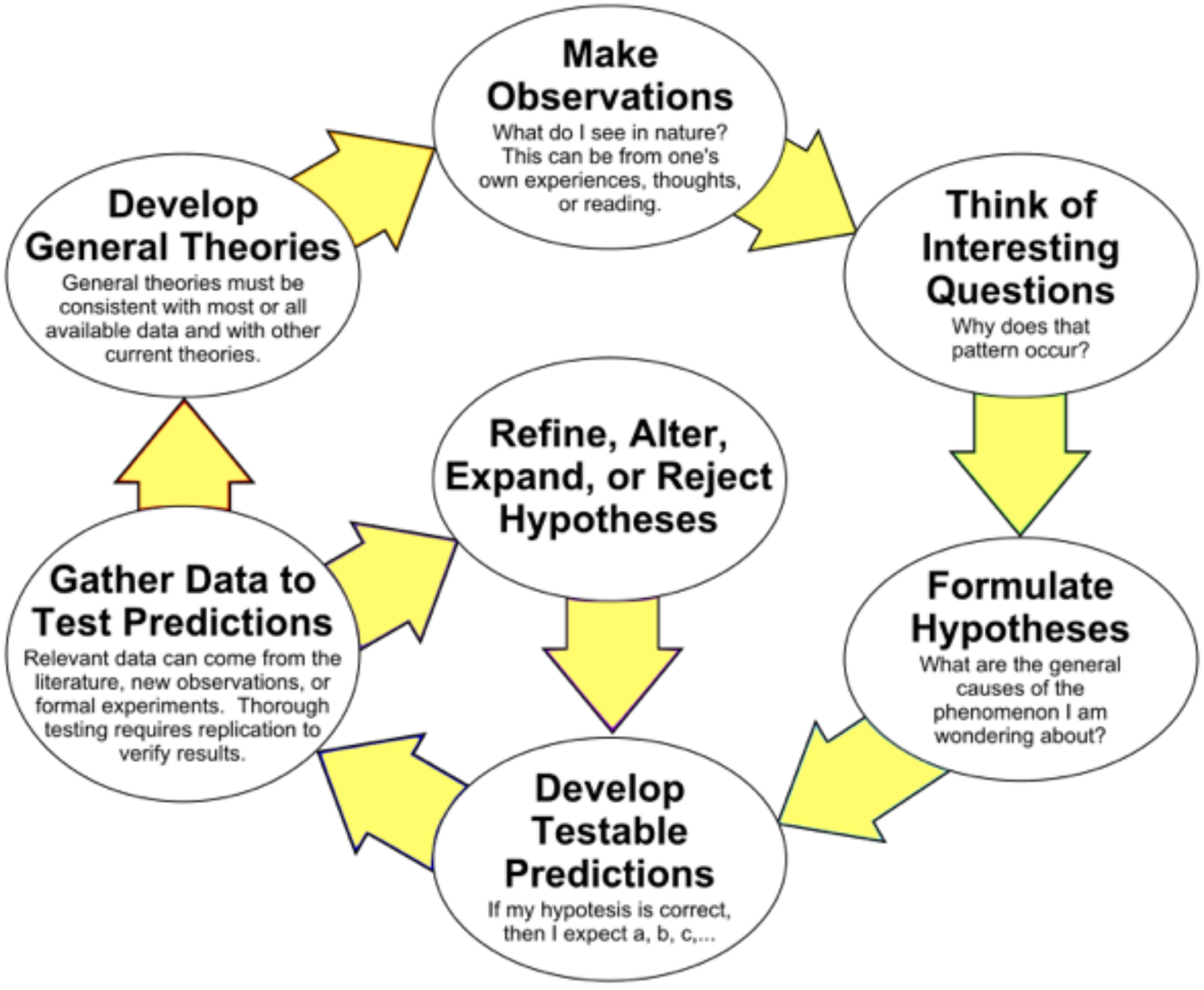
What is a hypothesis?

- #1 most important thing: *falsifiable*
- But also, should be:
    - Disputed (at least a little bit). I.e. it should be tied to a question people are actually asking

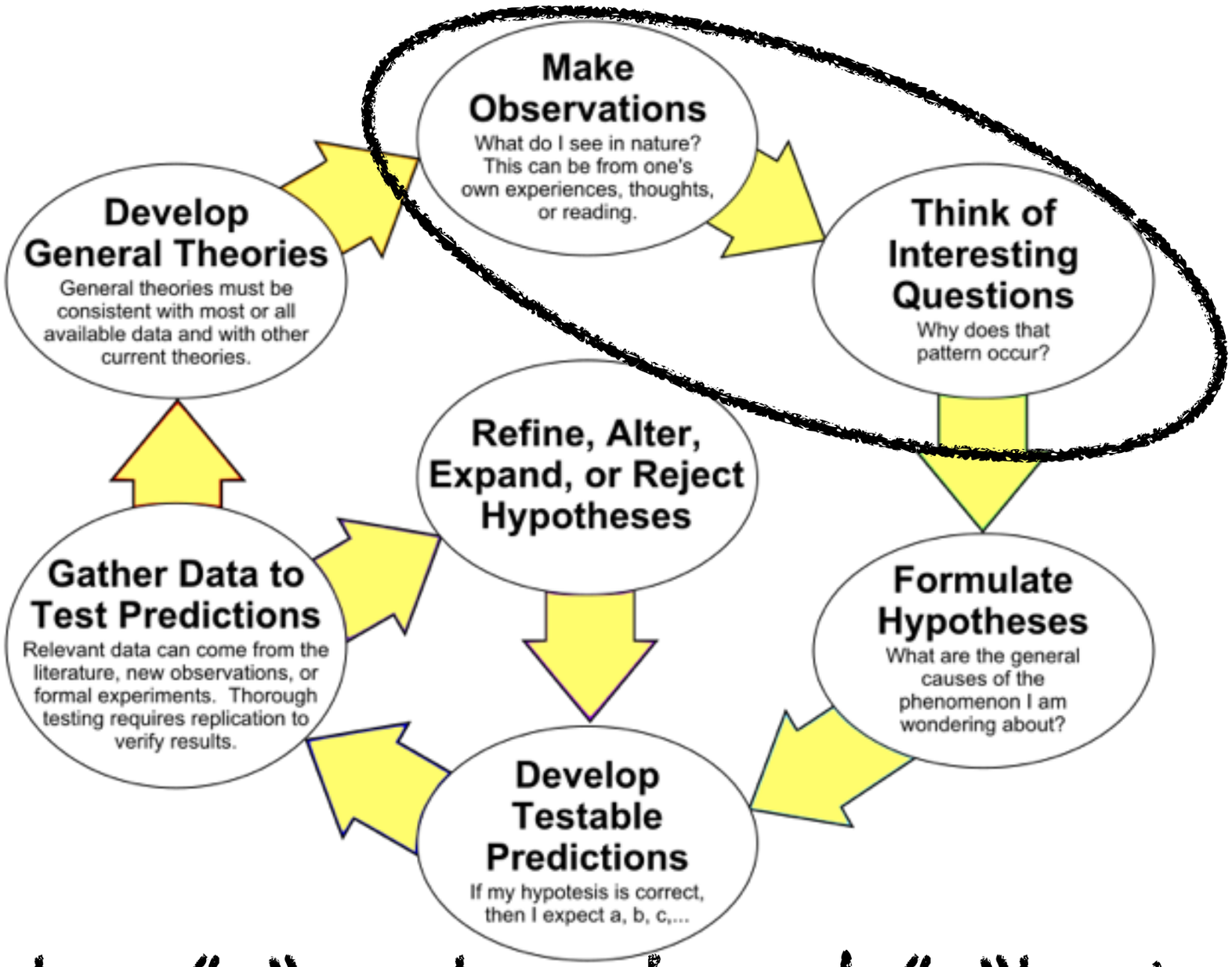
What is a hypothesis?

- #1 most important thing: *falsifiable*
- But also, should be:
  - Disputed (at least a little bit). I.e. it should be tied to a question people are actually asking
  - Specific. Avoid subjective terms like “better than”

What is a hypothesis?

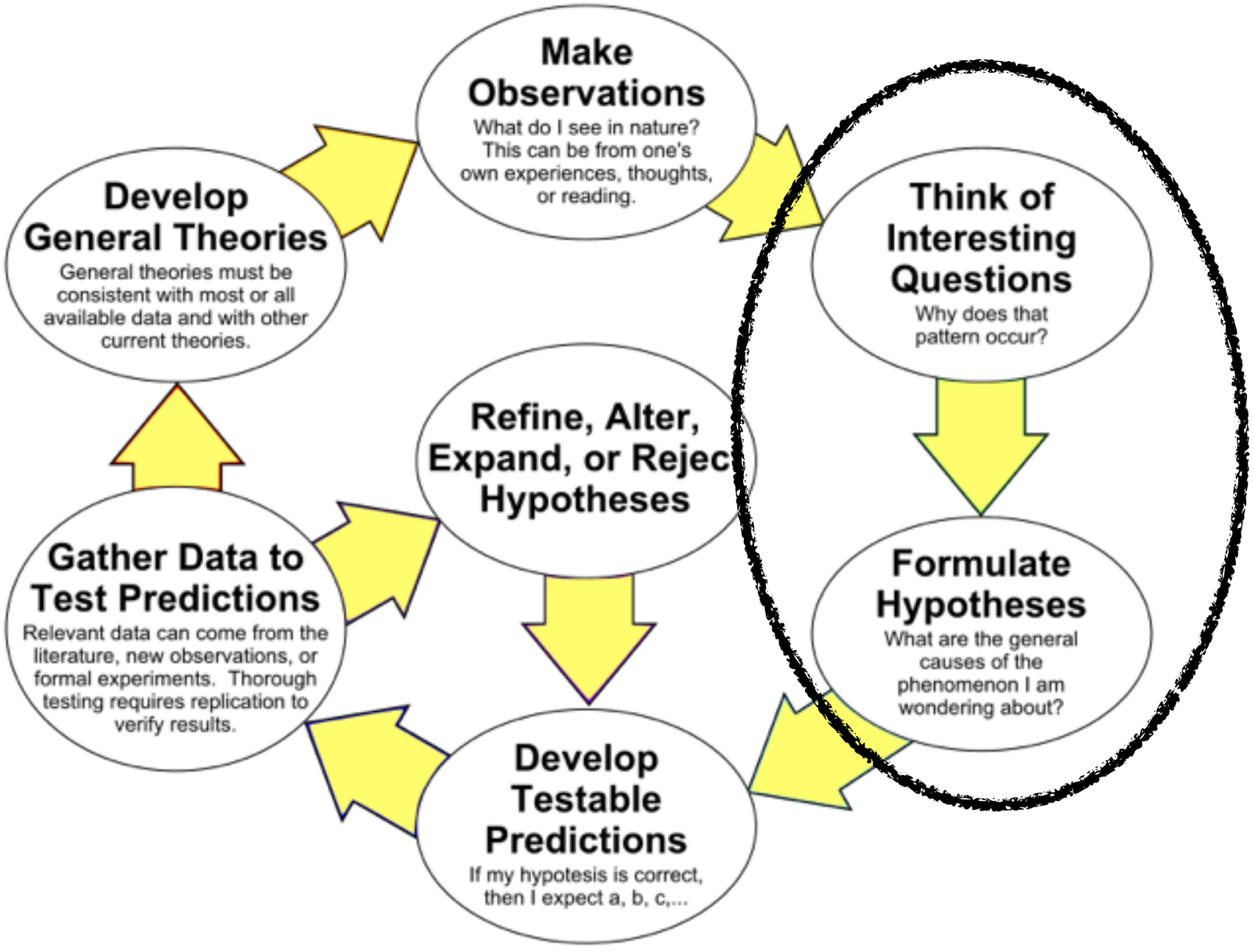


What is a hypothesis?



"explore", "analyze trends", "look for patterns", "visualize"

What is a hypothesis?



Literally the hardest part!

Clicker Question!

Clicker Question!

"Look for differences in political affiliations between universities"

Is this a valid hypothesis?

a) Yes
b) No

Clicker Question!

"Look for differences in political affiliations between universities"

Is this a valid hypothesis?

a) Yes
b) No

Clicker Question!

"Look for differences in political affiliations between universities"

"There are differences in political affiliations between universities"

Is this a valid hypothesis?

a) Yes
b) No

Clicker Question!

"Look for differences in political affiliations between universities"

"There are differences in political affiliations between universities"

Is this a valid hypothesis?

a) Yes
b) No

Clicker Question!

"Look for differen

"There are differer

yes, testable.

no, probably not really in question. (though could be)

Is this a valid hypothesis?

a) Yes
b) No

Clicker Question!

"Look for differences in political affiliations between universities"

"There are differences in political affiliations between universities"

"Coastal universities are more liberal than universities in the heartland"

Is this a valid hypothesis?

a) Yes
b) No

Clicker Question!

"Look for differences in political affiliations between universities"

"There are differences in political affiliations between universities"

"Coastal universities are more liberal than universities in the heartland"

Is this a valid hypothesis?

a) Yes
b) No

Clicker Question!

"Look for differences in political affiliations between universities"

"There are differences in political affiliations between universities"

"**Coastal** universities are more liberal than universities in the **heartland**"

need to define these

Clicker Question!

"Look for differences in political affiliations between universities"

"There are differences in political affiliations between universities"

"Coastal universities are more **liberal** than universities in the heartland"

and this: party affiliation? voting record? general opinions?

Clicker Question!

"Look for differences in political affiliations between universities"

scary truth #1: no single right way to do this

"Coastal universities are more **liberal** than universities in the heartland"

and this: party affiliation? voting record? general opinions?

Clicker Question!

"Look for differences in political affiliations between

scary truth #2: you are biased, you have to work hard to not let yourself just - "find what you want to find"

"Coastal universities are more **liberal** than universities in the heartland"

and this: party affiliation? voting record? general opinions?

Clicker Question!

"Look for differences in political affiliations between

scary truth #2: you are biased, you have to work hard to not let yourself just - "find what you want to find"

"Coastal universities ar more **liberal** than universities in the

to be continued... (in a few lectures)

and this: par g record? general opinions!

Today

- ~~What is a hypothesis?~~
- Some definitions/notation
- Intuition behind modeling/hypothesis testing

Statistics vs. Prob. Theory

Statistics vs. Prob. Theory

- Probability theory: mathematical theory that describes uncertainty.

Statistics vs. Prob. Theory

- Probability theory: mathematical theory that describes uncertainty.
- Distributions/parameters are known

Statistics vs. Prob. Theory

- Probability theory: mathematical theory that describes uncertainty.
  - Distributions/parameters are known

- Statistics: techniques for extracting useful information from data.

Statistics vs. Prob. Theory

- Probability theory: mathematical theory that describes uncertainty.
    - Distributions/parameters are known
- Statistics: techniques for extracting useful information from data.
    - Distributions/parameters are generally unknown and need to be estimated from the data

The bigger picture

- Start with real world phenomenon/observations
- Make assumptions about the underlying model
- Fit the parameters of the model based on data

The bigger picture

Whether a coin is heads or tails

What a person will say next

Whether someone will click on an ad

- Start with real world phenomenon/observations
- Make assumptions about the underlying model
- Fit the parameters of the model based on data

The bigger picture

You *always* make assumptions about the structure of the process that is generating the data

- Start with real world phenomenon/observations
- Make assumptions about the underlying model
- Fit the parameters of the model based on data

"All models are bad, but some are useful"

The bigger picture

Goal is always to "explain the data"

- Start with real world phenomenon/observations
- Make assumptions about the underlying model
- Fit the parameters of the model based on data

Two typical (different) use cases:

1) understand the underlying model better
2) make predictions

Probability Space

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

Probability Space

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

the set of all possible outcomes of the random process modeled

Probability Space

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

A family of sets F representing the allowable events, where each set in F is a subset of the sample space Ω

$$F = \{E_i \subseteq \Omega\}_i$$

Probability Space

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

A family of sets F representing the allowable events, where each set in F is a subset of the sample space Ω

$$F = \{E_i \subseteq \Omega\}_i$$

$$F = 2^{\Omega}$$

Probability Space

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

Probability function which assigns a real number to each event in F

$$P : F \rightarrow \mathbb{R}$$

Probability Space

Valid Probability Function:

$$0 \leq P(E) \leq 1 \; \forall E \in F$$

$$P(\Omega) = 1$$

$$P(\bigcup_i E_i) = \sum_i P(E_i)$$

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

robability function which
gns a real number to each
event in F

$$P : F \rightarrow \mathbb{R}$$

Over-used Example

Tossing a fair coin once

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

Over-used Example

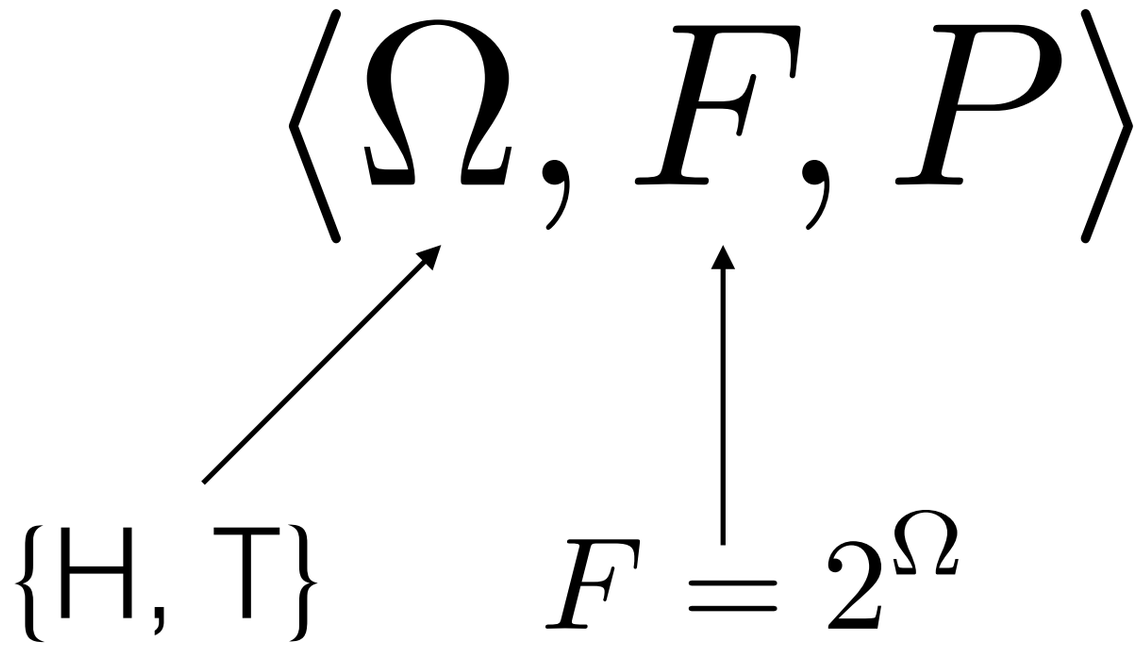
Tossing a fair coin once

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

{H, T}

Over-used Example

Tossing a fair coin once



Over-used Example

Tossing a fair coin once

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

{H, T}

$F = 2^{\Omega}$

{}, {H}, {T}, {H, T}

{} -> 0
{H} -> 0.5
{T} -> 0.5
{H, T} -> ???

sed Example

ng a fair coin once

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

{H, T}

$F = 2^{\Omega}$

{}, {H}, {T}, {H, T}

{} -> 0
{H} -> 0.5
{T} -> 0.5
{H, T} -> ???

sed Example

ng a fair coin once

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

{H, T} $F =$

{}, {H}, {T

Valid Probability Function:

$$0 \leq P(E) \leq 1 \; \forall E \in F$$

$$P(\Omega) = 1$$

$$P(\bigcup_i E_i) = \sum_i P(E_i)$$

{} -> 0

{H} -> 0.5

{T} -> 0.5

{H, T} -> P({H} ∪{T}) = P({H}) + P({T}) = 1

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

{H, T} $F =$

{}, {H}, {T

Valid Probability Function:

$$0 \leq P(E) \leq 1 \; \forall E \in F$$

$$P(\Omega) = 1$$

$$P(\bigcup_i E_i) = \sum_i P(E_i)$$

Random Variables

Random Variables

- Random variable X assigns a number to each outcome: $X : \Omega \to \mathbb{R}$

Random Variables

- Random variable X assigns a number to each outcome: $X : \Omega \to \mathbb{R}$
- Use $X = a$ to mean the event $\{\omega | X(\omega) = a\}$

Random Variables

- Random variable X assigns a number to each outcome: $X : \Omega \rightarrow \mathbb{R}$

- Use $X = a$ to mean the event $\{\omega | X(\omega) = a\}$

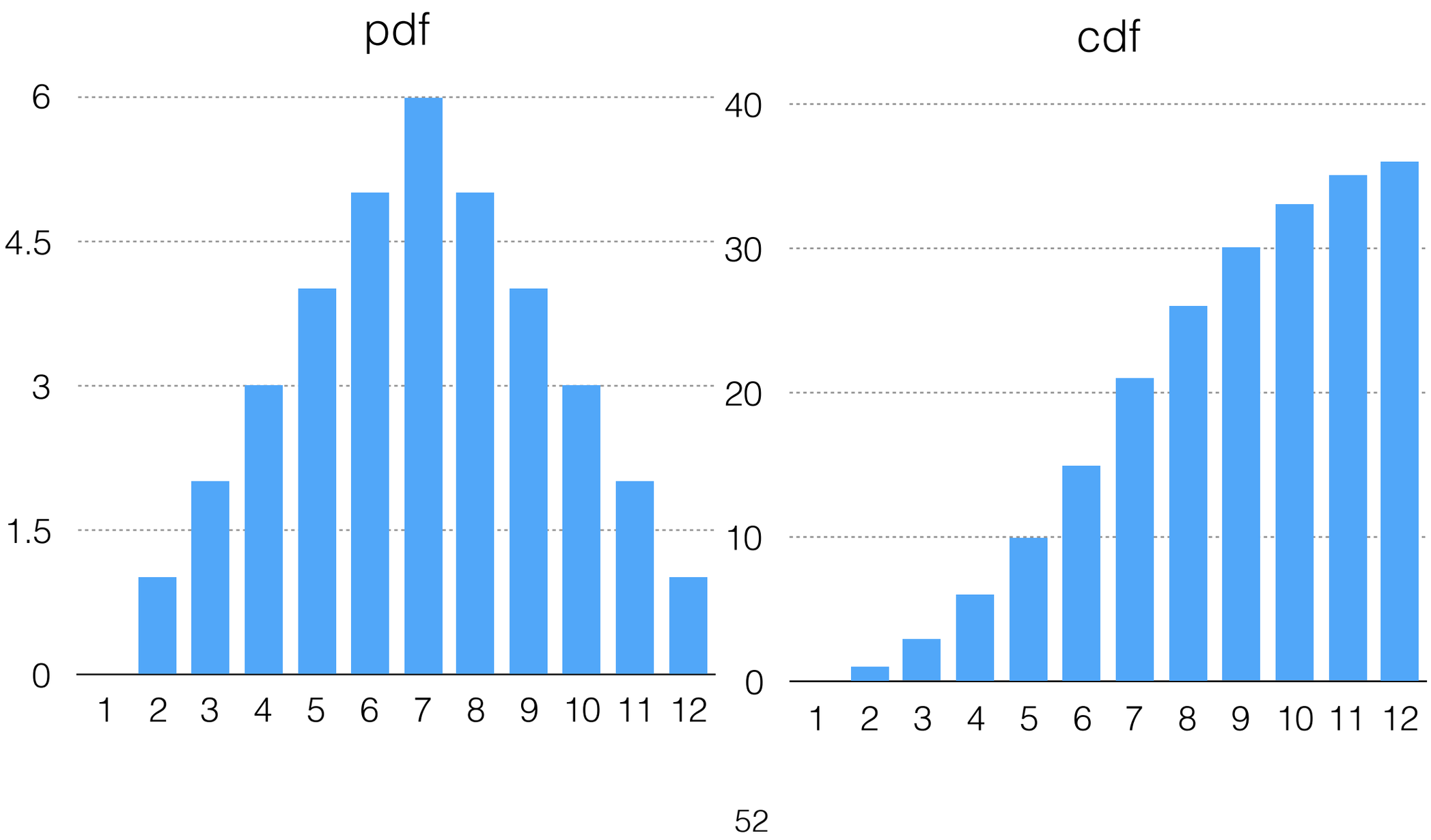
- Probability mass function (pmf) gives probability that X takes the value a: $p(a) = Pr(X = a)$

Random Variables

- Random variable X assigns a number to each outcome: $X : \Omega \rightarrow \mathbb{R}$
- Use $X = a$ to mean the event $\{\omega | X(\omega) = a\}$
- Probability mass function (pmf) gives probability that X takes the value a: $p(a) = Pr(X = a)$
- Cumulative distribution function (cdf) gives probability that X takes any value up to a: $F(a) = Pr(X \leq a)$

Random Variables

X = sum of two dice



Clicker Question!

Clicker Question!

X is a random variable with the below cdf.

X	1	2	3	4
cdf F(a)	0.5	0.75	0.9	1

What is P(X≤3)?

(a) 0 (b) 0.15 (c) 0.9 (d) 1

Clicker Question!

X is a random variable with the below cdf.

X	1	2	3	4
cdf F(a)	0.5	0.75	0.9	1

What is P(X≤3)?

(a) 0 (b) 0.15 (c) 0.9 (d) 1

Clicker Question!

X is a random variable with the below cdf.

X	1	2	3	4
cdf F(a)	0.5	0.75	0.9	1

What is P(X=3)?

(a) 0 (b) 0.15 (c) 0.9 (d) 1

Clicker Question!

X is a random variable with the below cdf.

X	1	2	3	4
cdf F(a)	0.5	0.75	0.9	1

What is P(X=3)?

(a) 0 (b) 0.15 (c) 0.9 (d) 1

Expected Value

$$E(X) = \sum_i x_i Pr(x_i)$$

Expected Value

$$E(X) = \sum_i x_i Pr(x_i)$$

$$\Downarrow\Uparrow$$

$$E(X) = \int_i x_i Pr(x_i)$$

Expected Value

$$E(X) = \sum_i x_i Pr(x_i)$$

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1

Expected Value

$$E(X) = \sum_i x_i Pr(x_i)$$

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1

0.5

Expected Value

$$E(X) = \sum_i x_i Pr(x_i)$$

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1

0.5 + 0.5

Expected Value

$$E(X) = \sum_i x_i Pr(x_i)$$

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1

0.5 + 0.5 + 0.45

Expected Value

$$E(X) = \sum_i x_i Pr(x_i)$$

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1

0.5 + 0.5 + 0.45 + 0.4

Expected Value

$$E(X) = \sum_i x_i Pr(x_i)$$

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1

$$0.5 + 0.5 + 0.45 + 0.4 = 1.85$$

Variance

$$Var(X) = E((X - E(X))^2)$$

Variance

$$Var(X) = E((X - E(X))^2)$$

E(X) = 1.85

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1

Variance

$$Var(X) = E((X - E(X))^2)$$

E(X) = 1.85

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1
X - E(X)	-0.85	0.15	1.15	2.15

Variance

$$Var(X) = E((X - E(X))^2)$$

E(X) = 1.85

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1
X - E(X)	-0.85	0.15	1.15	2.15
$(X - E(X))^2$	0.722	0.023	1.32	4.62

Variance

$$Var(X) = E((X - E(X))^2)$$

E(X) = 1.85

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1
X - E(X)	-0.85	0.15	1.15	2.15
$(X - E(X))^2$	0.722	0.023	1.32	4.62

0.361

Variance

$$Var(X) = E((X - E(X))^2)$$

E(X) = 1.85

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1
X - E(X)	-0.85	0.15	1.15	2.15
$(X - E(X))^2$	0.722	0.023	1.32	4.62

0.361 + .006

Variance

$$Var(X) = E((X - E(X))^2)$$

E(X) = 1.85

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1
X - E(X)	-0.85	0.15	1.15	2.15
$(X - E(X))^2$	0.722	0.023	1.32	4.62

0.361 + .006 + 0.198

Variance

$$Var(X) = E((X - E(X))^2)$$

E(X) = 1.85

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1
X - E(X)	-0.85	0.15	1.15	2.15
$(X - E(X))^2$	0.722	0.023	1.32	4.62

0.361 + .006 + 0.198 + 0.462

Variance

$$Var(X) = E((X - E(X))^2)$$

E(X) = 1.85

X	1	2	3	4
pdf	0.5	0.25	0.15	0.1
X - E(X)	-0.85	0.15	1.15	2.15
$(X - E(X))^2$	0.722	0.023	1.32	4.62

0.361 + .006 + 0.198 + 0.462 = 1.027

Interpreting Expectation

Would you accept a gamble that offers a 10% chance to win $95 and a 90% chance of losing $5?

Interpreting Expectation

Would you accept a gamble that offers a 10% chance to win $95 and a 90% chance of losing $5?

$$E(\text{Payoff}) = (95 \times 0.10) - (5 \times 0.9)$$

Interpreting Expectation

Would you accept a gamble that offers a 10% chance to win $95 and a 90% chance of losing $5?

$$E(\text{Payoff}) = (95 \times 0.10) - (5 \times 0.9)$$

$$E(\text{Payoff}) = (9.5) - (4.5)$$

Interpreting Expectation

Would you accept a gamble that offers a 10% chance to win $95 and a 90% chance of losing $5?

$$E(\text{Payoff}) = (95 \times 0.10) - (5 \times 0.9)$$

$$E(\text{Payoff}) = (9.5) - (4.5)$$

$$E(\text{Payoff}) = 5$$

Clicker Question!

Clicker Question!

How much would you pay for a lottery ticket that offers a 10% percent chance of winning $100 and a 90% chance of winning nothing?

(a) $0
(b) no more than $2
(c) no more than $5
(d) no more than $10

Clicker Question!

How much would you pay for a lottery ticket that offers a 10% percent chance of winning $100 and a 90% chance of winning nothing?

(a) $0
(b) no more than $2
(c) no more than $5
(d) no more than $10

Clicker Question!

How much would you pay for a lottery ticket that offers a 10% percent chance of winning $100 and a 90% chance of winning nothing?

(a) $0
(b) no more than $2
(c) no more than $5
(d) no more than $10

$0 = 0.1(100 - cost) - 0.9(cost)$

$0 = 10 - cost$

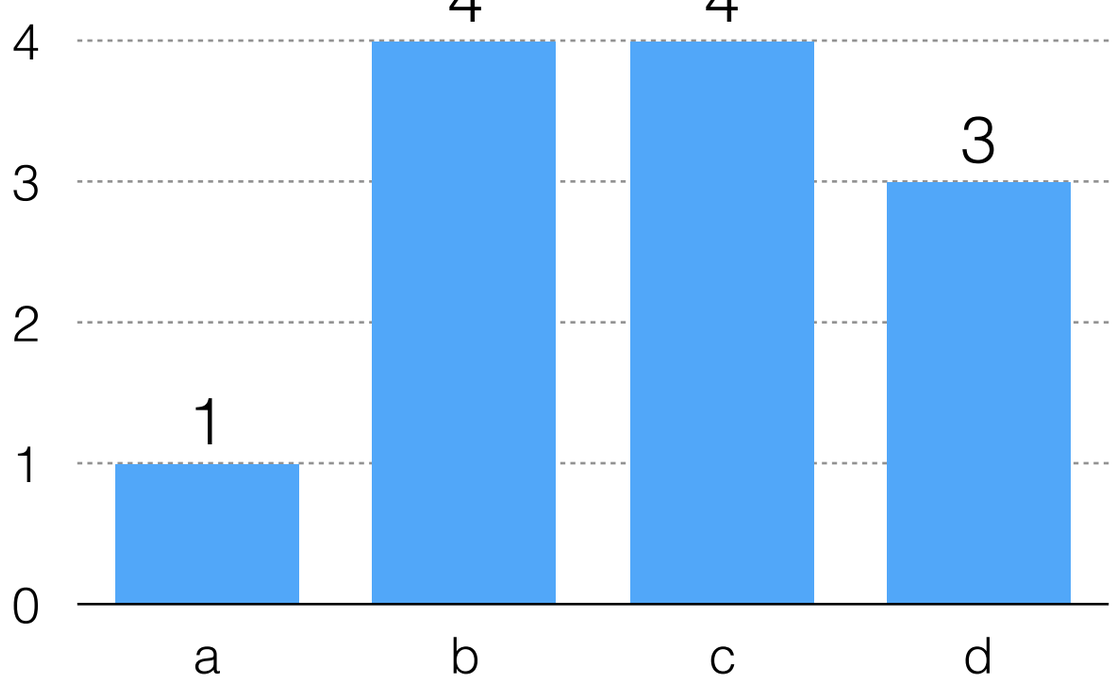
$cost = 10$

Today

- ~~What is a hypothesis?~~
- ~~Some definitions/notation~~
- Intuition behind modeling/hypothesis testing

Gaming Clicker Questions!

Are the answers to my clicker questions random?



Are the answers to my clicker questions random?

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

Are the answers to my clicker questions random?

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

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

Are the answers to my clicker questions random?

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

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

Probability of this?

Are the answers to my clicker questions random?

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

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

Probability of this?

The bigger picture

(Yes, I will belabor this point. Why do you ask?)

- Start with real world phenomenon/observations
- Make assumptions about the underlying model
- Fit the parameters of the model based on data

The bigger picture

- Start with real world phenomenon/observations
- Make assumptions about the underlying model
- Fit the parameters of the model based on data
    - Chose parameters of the model based on theories, do analysis to see if its a good fit (hypothesis testing!!)

AND/OR

    - Set parameters of the model based on data, try to make forecast for unseen/future data (prediction!!)

The bigger picture

- Start with real world phenomenon/observations
- Make assumptions about the underlying model
- Fit the parameters of the model based on data

- Chose parameters of the model based on theories, do analysis to see if its a good fit (hypothesis testing!!)

AND/OR

- Set parameters of the model based on data, try to make forecast for unseen/future data (prediction!!)

The bigger picture

- Start with real world phenomenon/observations
- Make assumptions about the underlying model
- Fit the parameters of the model based on data

Now

- Chose parameters of the model based on theories, do analysis to see if its a good fit (hypothesis testing!!)

AND/OR

- Set parameters of the model based on data, try to make forecast for unseen/future data (prediction!!)

Later

Are the answers to my clicker questions random?

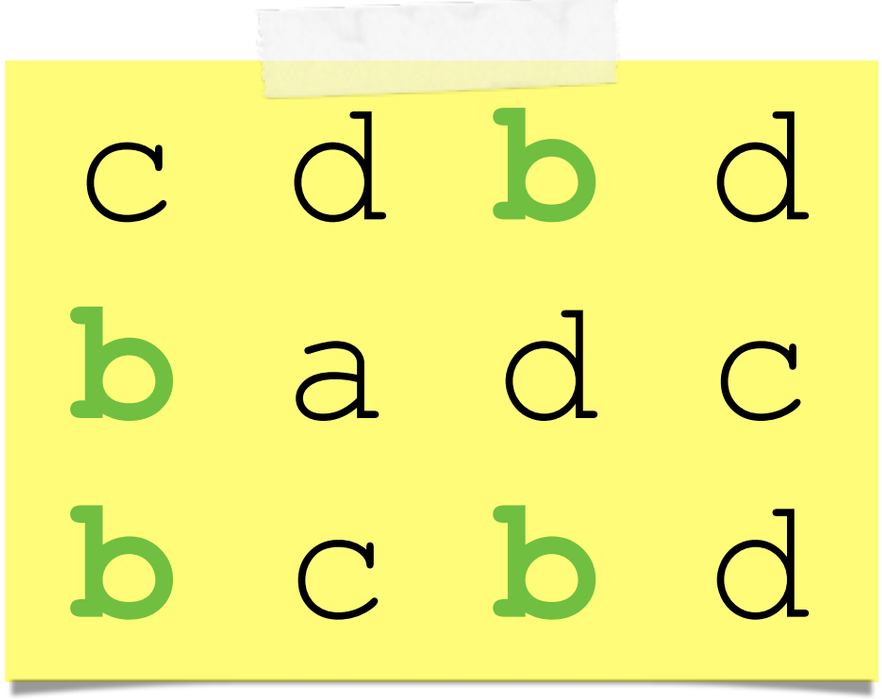
"I swear literally like 80% of the answers are just (b)"

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

Probability of this?

Are the answers to my clicker questions random?

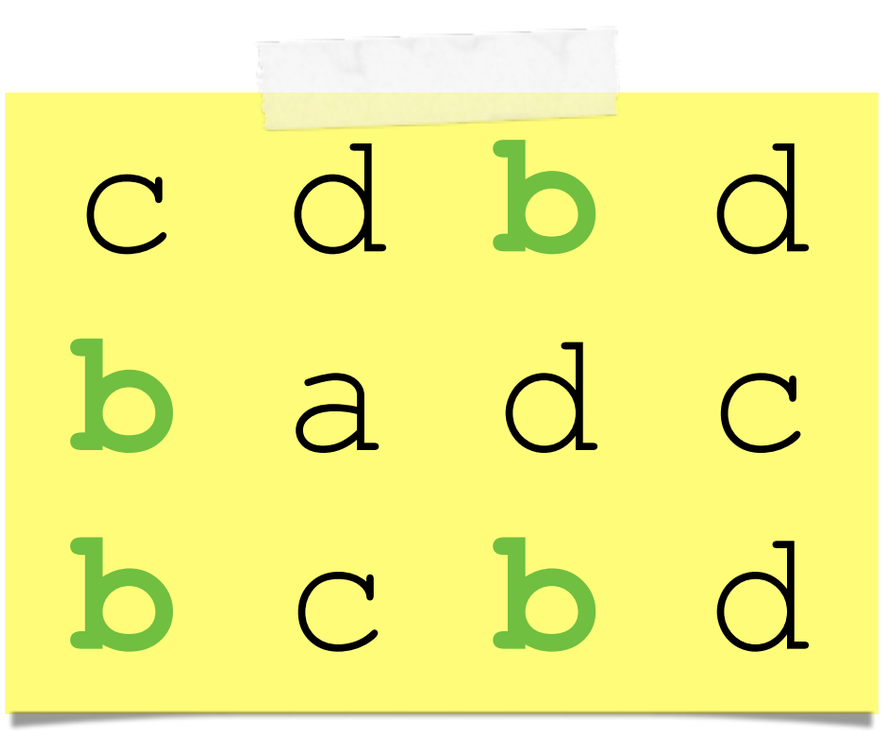
"I swear literally like 80% of the answers are just (b)"



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

Are the answers to my clicker questions random?

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

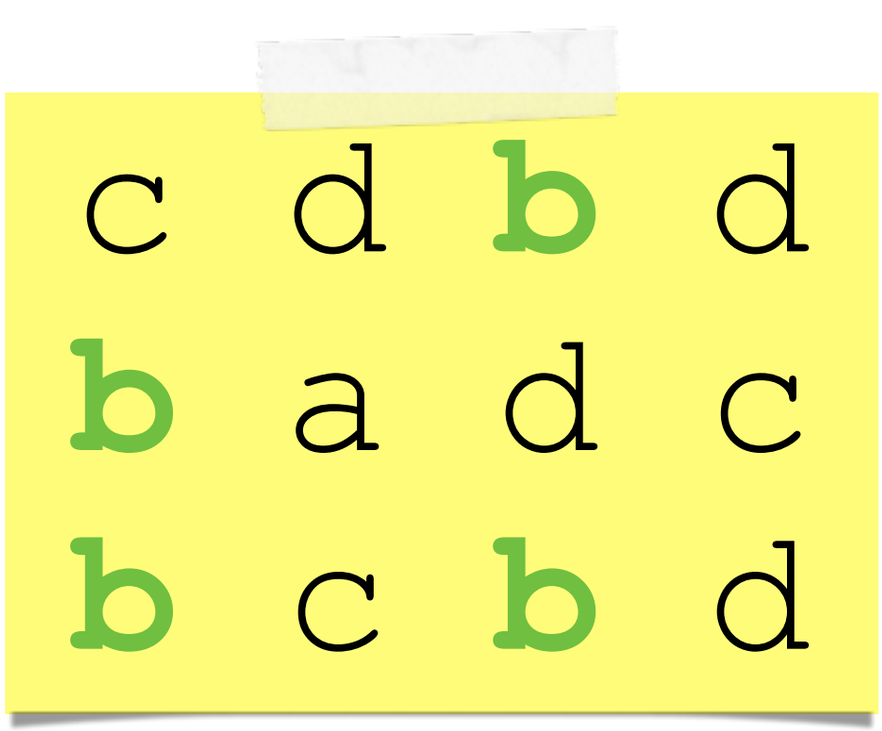


$$\langle \Omega, F, P \rangle$$

{b, not b}

Are the answers to my clicker questions random?

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



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

{b, not b}

p(B) = 0.8
p(not B) = 0.2

Are the answers to my clicker questions random?

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

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

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

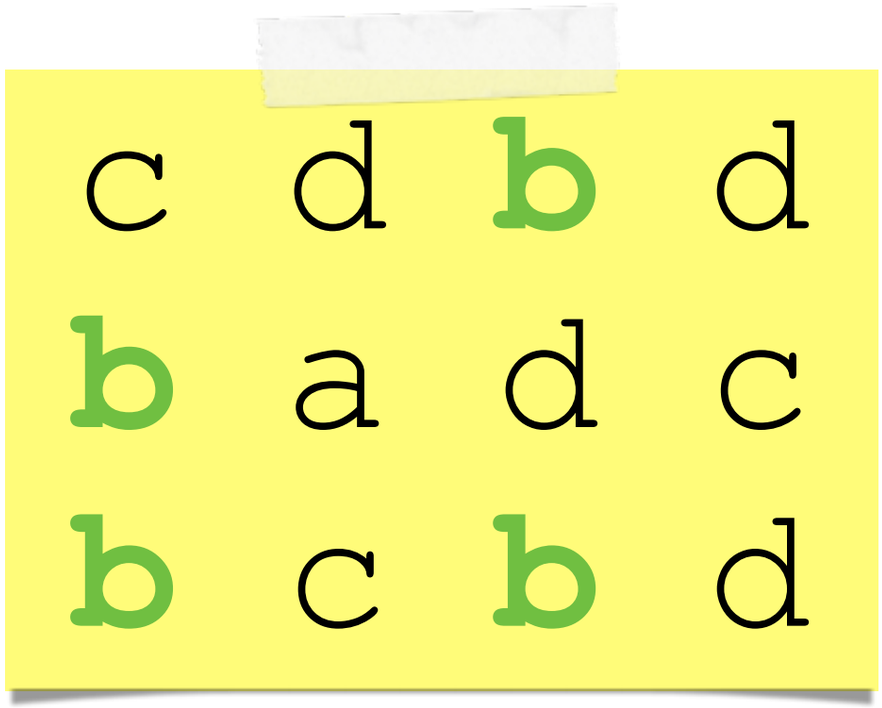
{b, not b}

p(B) = 0.8
p(not B) = 0.2

Assumption under investigation

Are the answers to my clicker questions random?

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



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

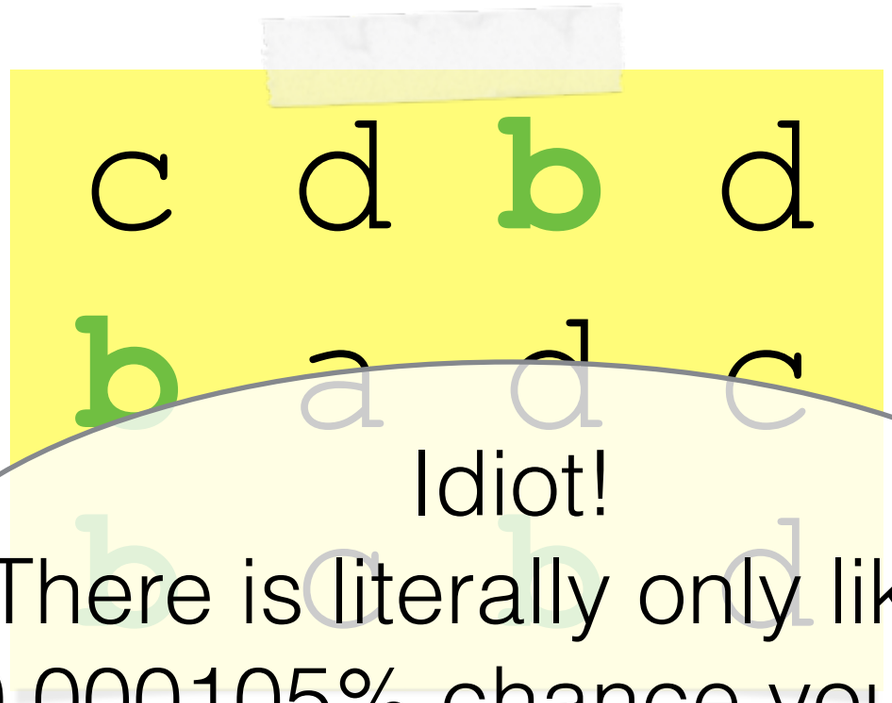
$$\langle \Omega, F, P \rangle$$

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

0.2 x 0.2 x 0.8 x 0.2 x
0.8 x 0.2 x 0.2 x 0.2 x
0.8 x 0.2 x 0.8 x 0.2 =
0.00000105

Are the answers to my clicker questions random?

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



Idiot!
There is literally only like a 0.000105% chance you are right!

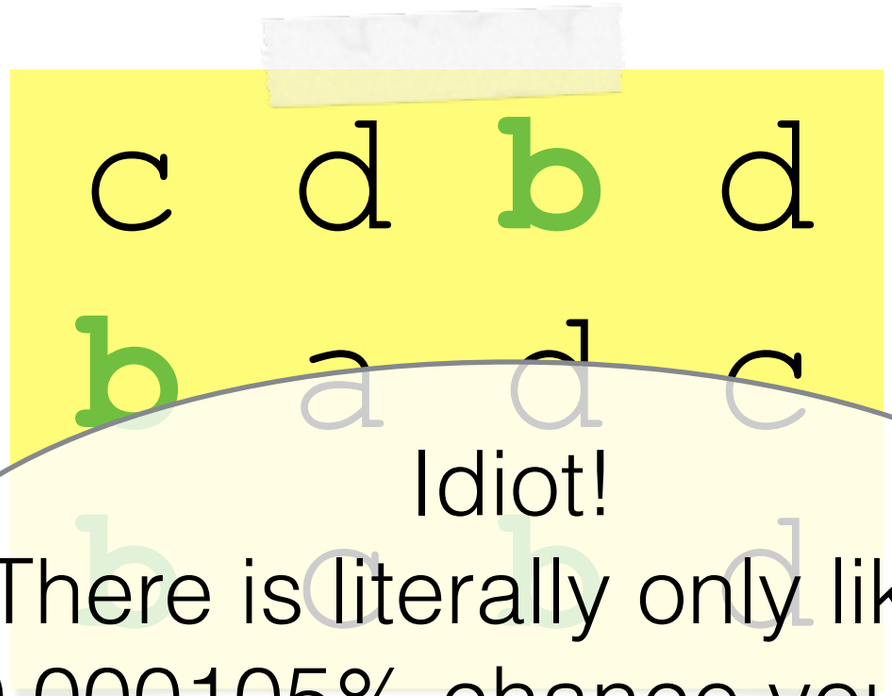
$$\langle \Omega, F, P \rangle$$

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

0.2 x 0.2 x 0.8 x 0.2 x 0.8 x 0.2 x 0.2 x 0.2 x 0.8 x 0.2 x 0.8 x 0.2 = 0.00000105

Are the answers to my clicker questions random?

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



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

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

0.2 x 0.2 x 0.8 x 0.2 x 0.8 x 0.2 x 0.2 x 0.2 x 0.8 x 0.2 x 0.8 x 0.2 = 0.00000105

???

Idiot!
There is literally only like a 0.000105% chance you are right!

Clicker Question!

Clicker Question!

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

What is the probability of this event?

b b b d

b a b b

b b b b

a) 1.0
b) 0.4
c) 0.04
d) 0.004

Clicker Question!

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

What is the probability of this event?

b b b d

b a b b

b b b b

a) 1.0
b) 0.4
c) 0.04
d) 0.004

Clicker Question!

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

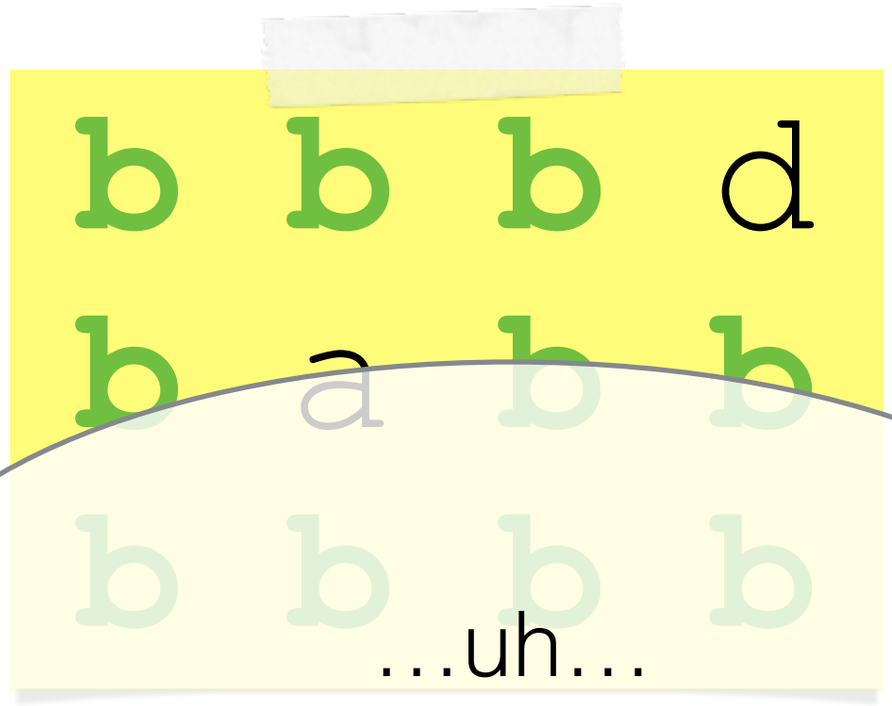
What is the probability of this event?

b	b	b	d
b	a	b	b
b	b	b	b

0.8 x 0.8 x 0.8 x 0.2 x
0.8 x 0.2 x 0.8 x 0.8 x
0.8 x 0.8 x 0.8 x 0.8 =
0.004

Are the answers to my clicker questions random?

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



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

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

0.8 x 0.8 x 0.8 x 0.2 x 0.8 x 0.2 x 0.8 x 0.8 x 0.8 x 0.8 x 0.8 x 0.8 = 0.004

Are the answers to my clicker questions random?

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

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

Probability of this?

Are the answers to my clicker questions random?

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

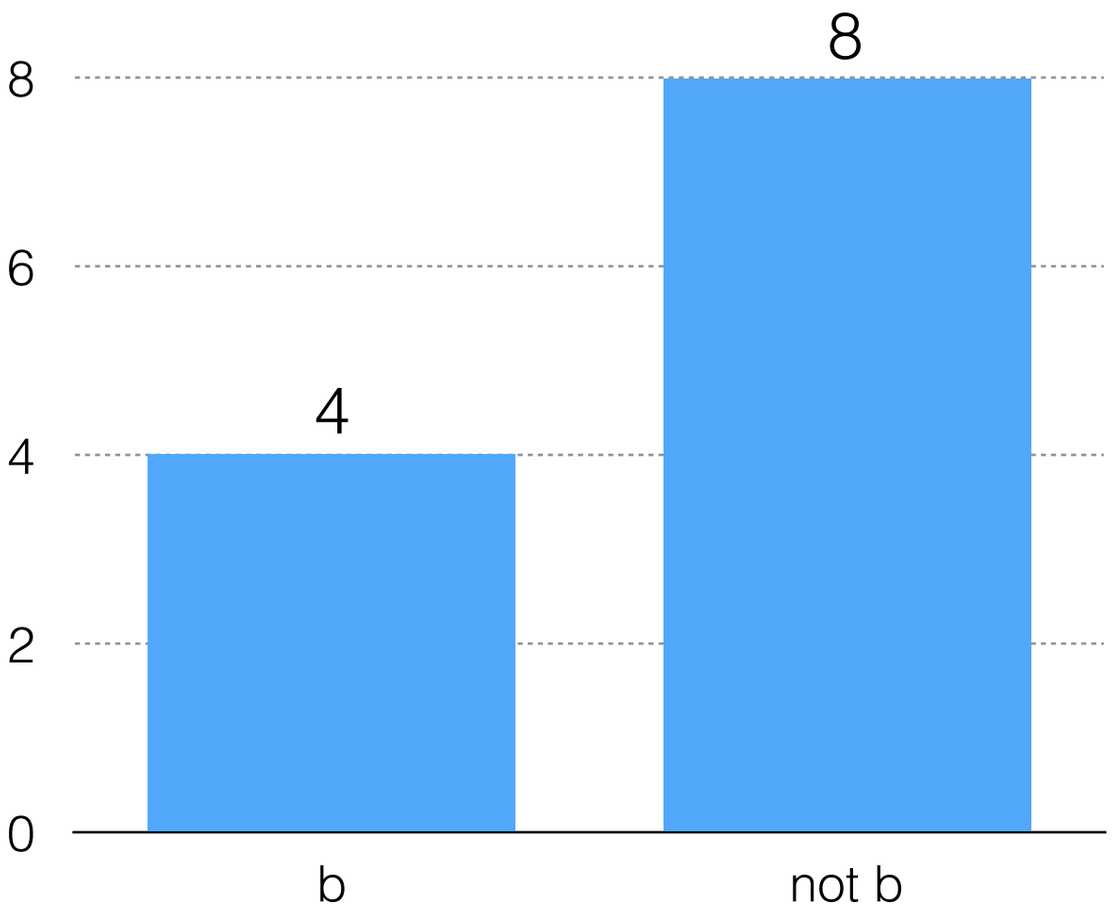
c	d	**b**	d
b	a	d	c
b	c	**b**	d

~~Probability of this?~~

Probability of anything as surprising as this

Are the answers to my clicker questions random?

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



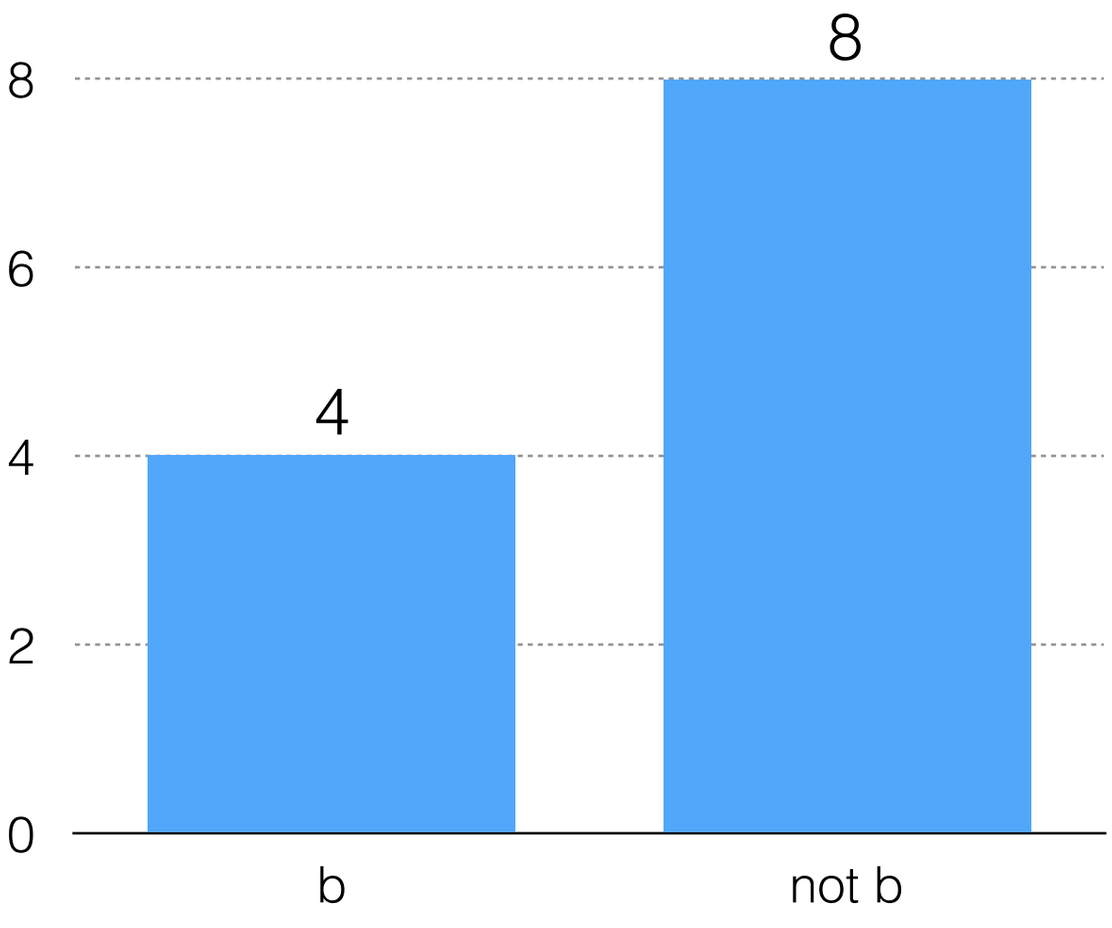
$$\langle \Omega, F, P \rangle$$

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

X = number of (b)s

Are the answers to my clicker questions random?

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

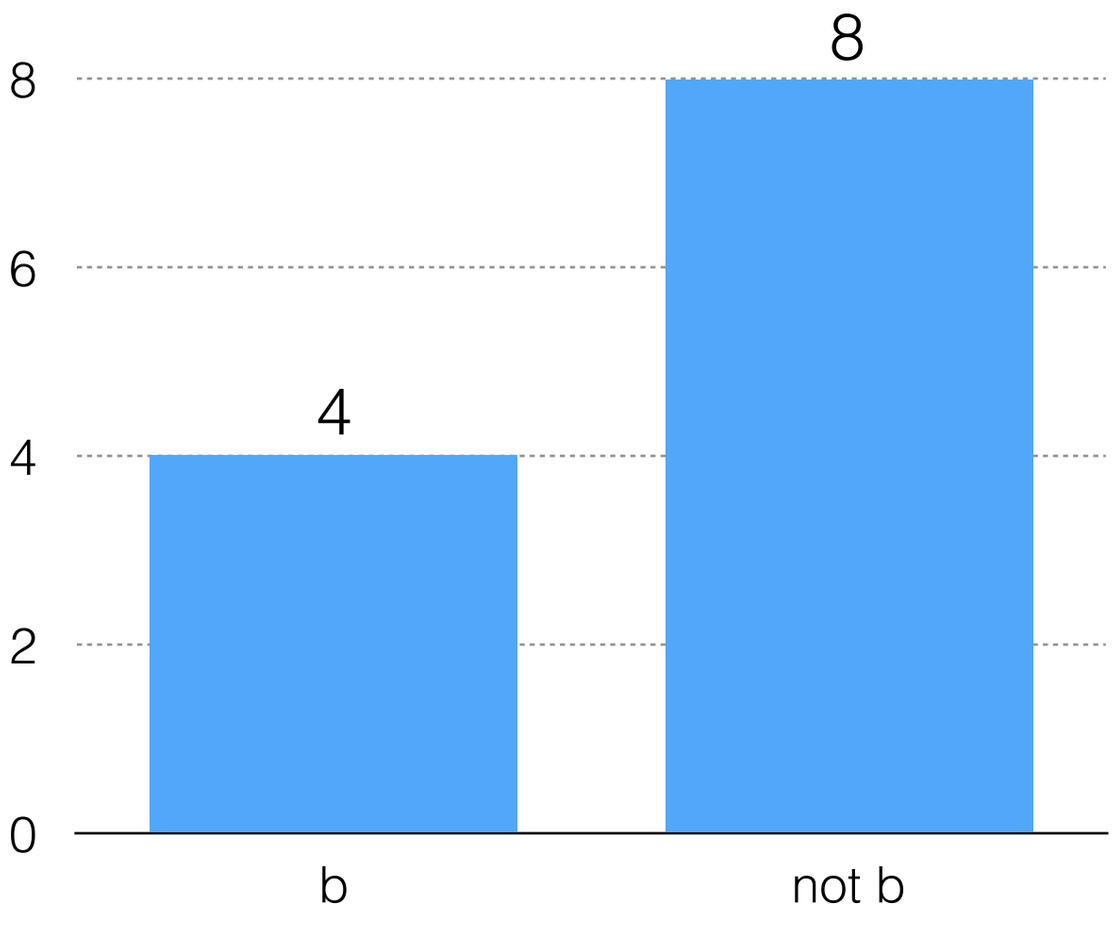


X = number of (b)s

How can we define the pdf for this?

Are the answers to my clicker questions random?

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



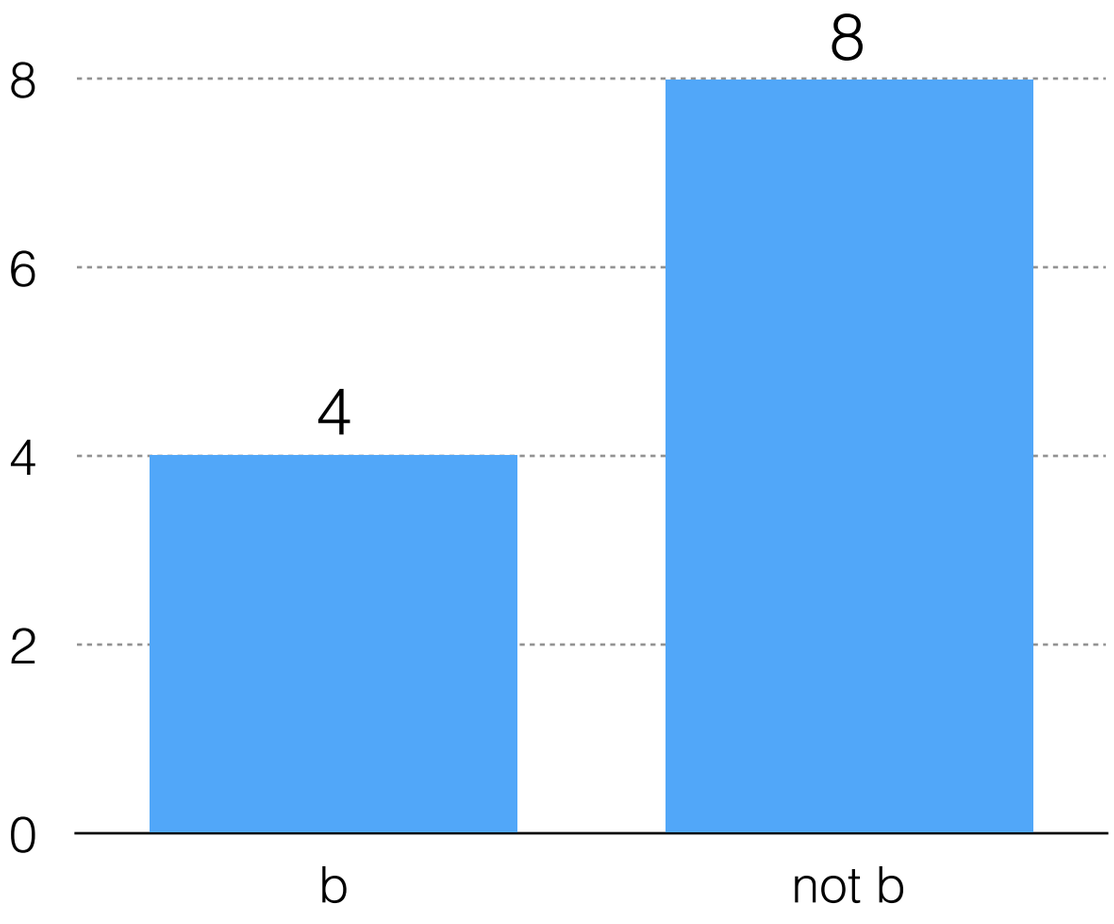
X = number of (b)s

How can we define the pdf for this?

i.e. how do we model P(# bs = k) for any value of k?

Are the answers to my clicker questions random?

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



X = number of (b)s

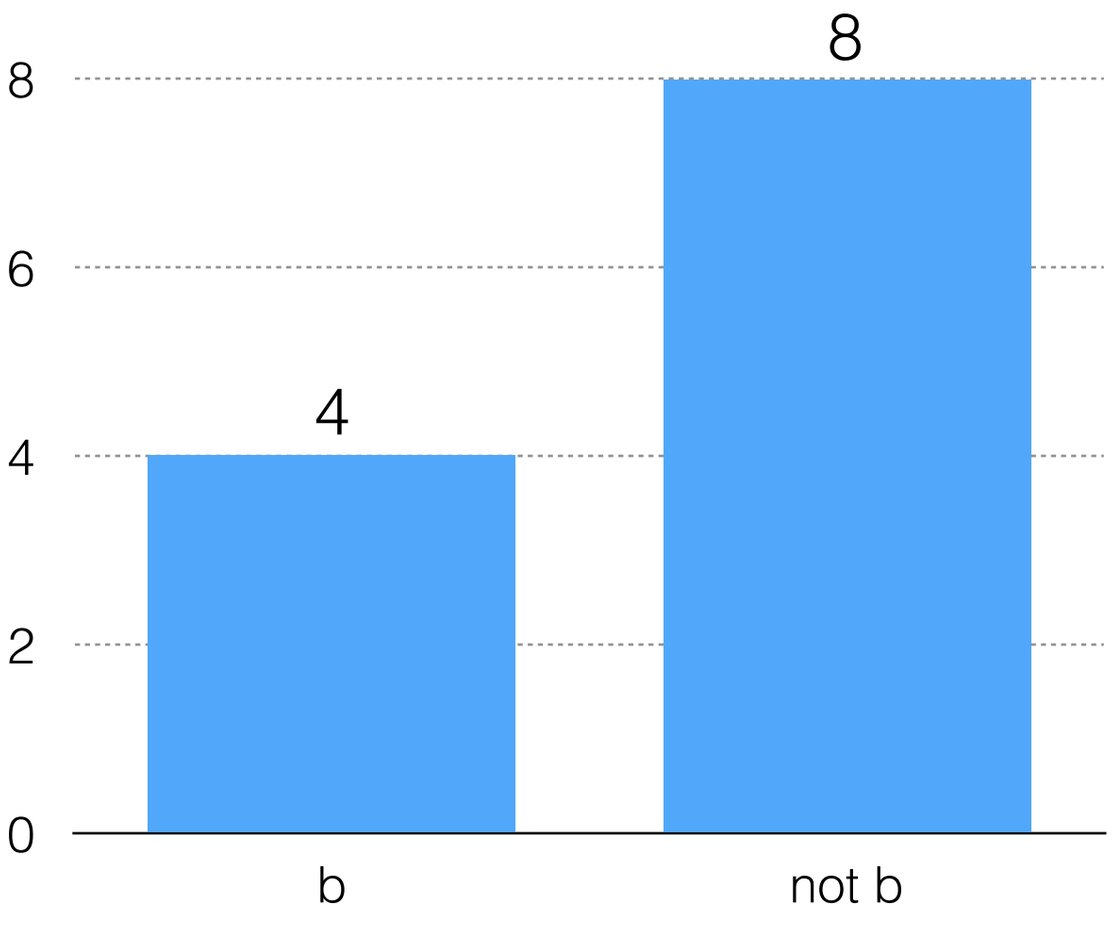
$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

binomial distribution!

("biased coin" distribution)

Are the answers to my clicker questions random?

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



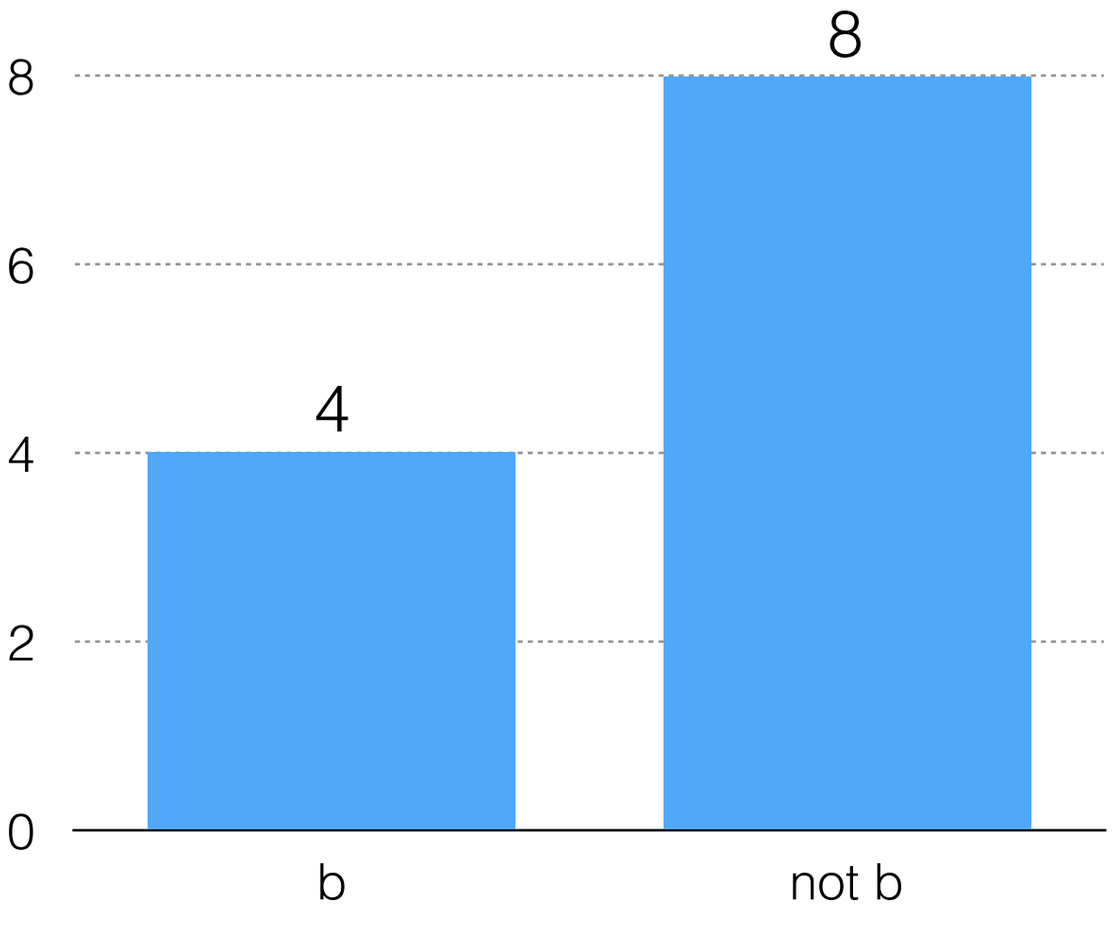
X = number of (b)s

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

P(#(b)s = k)

Are the answers to my clicker questions random?

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



X = number of (b)s

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

account for all the positions the k bs could occur in

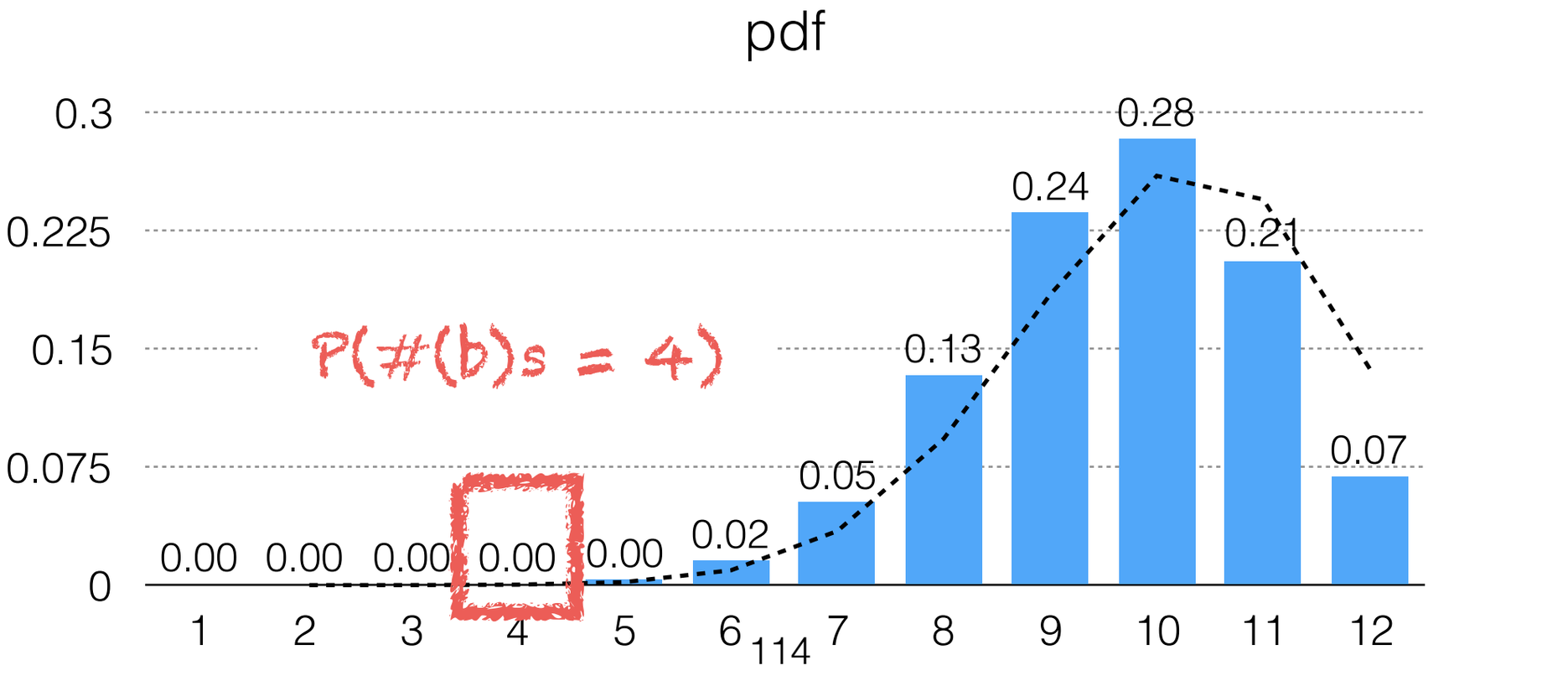
Are the answers to my clicker questions random?

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

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

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

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

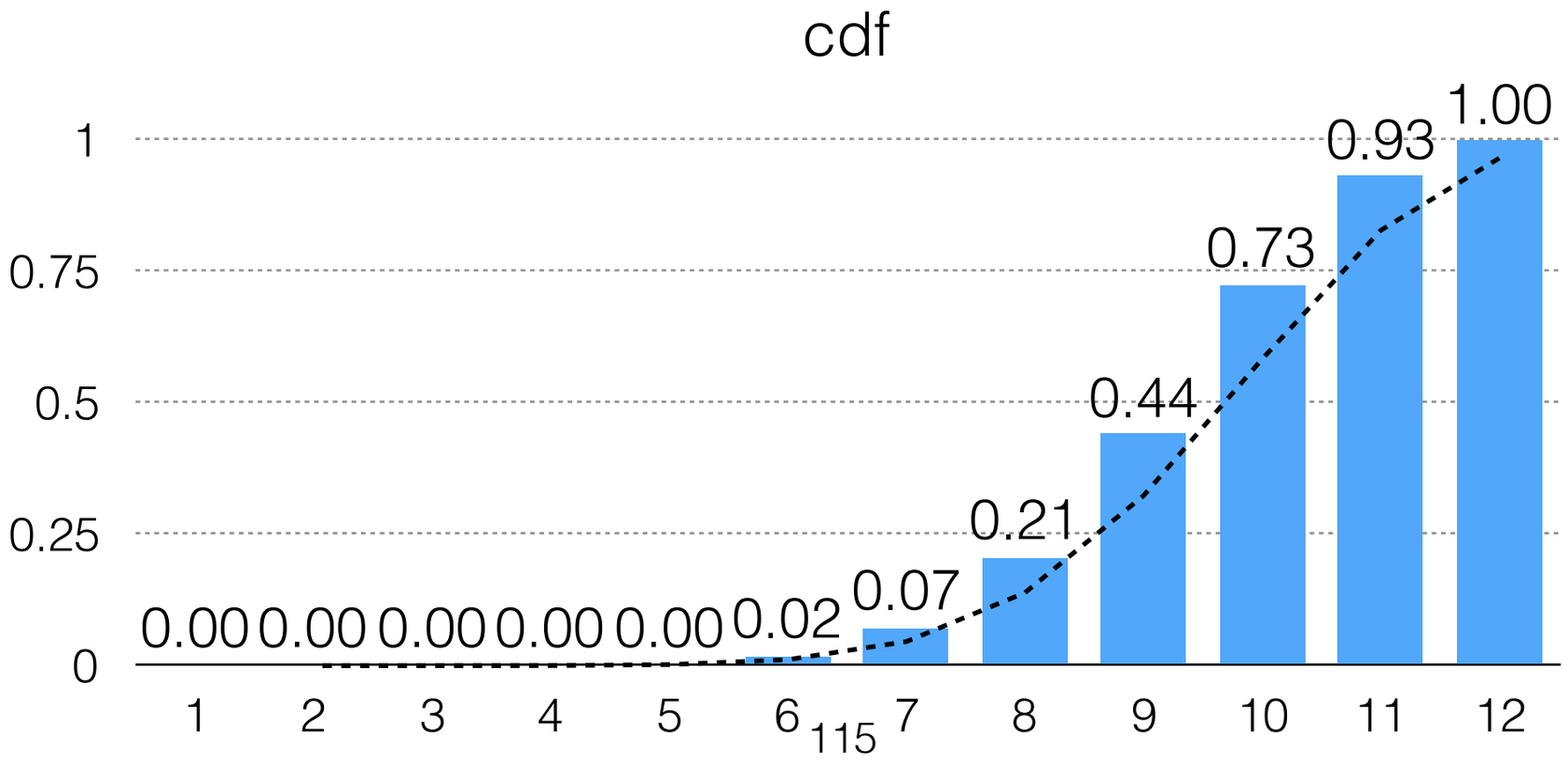


Are the answers to my clicker questions random?

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

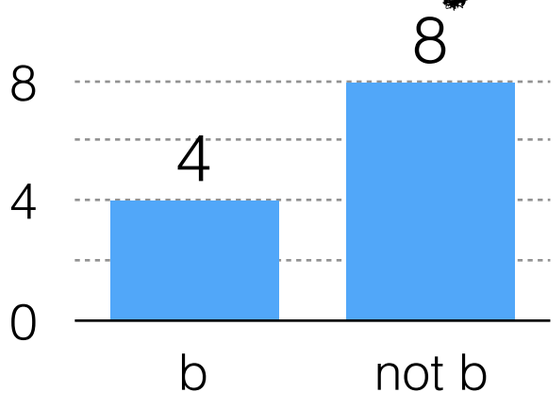
$$\langle \Omega, F, P \rangle$$

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

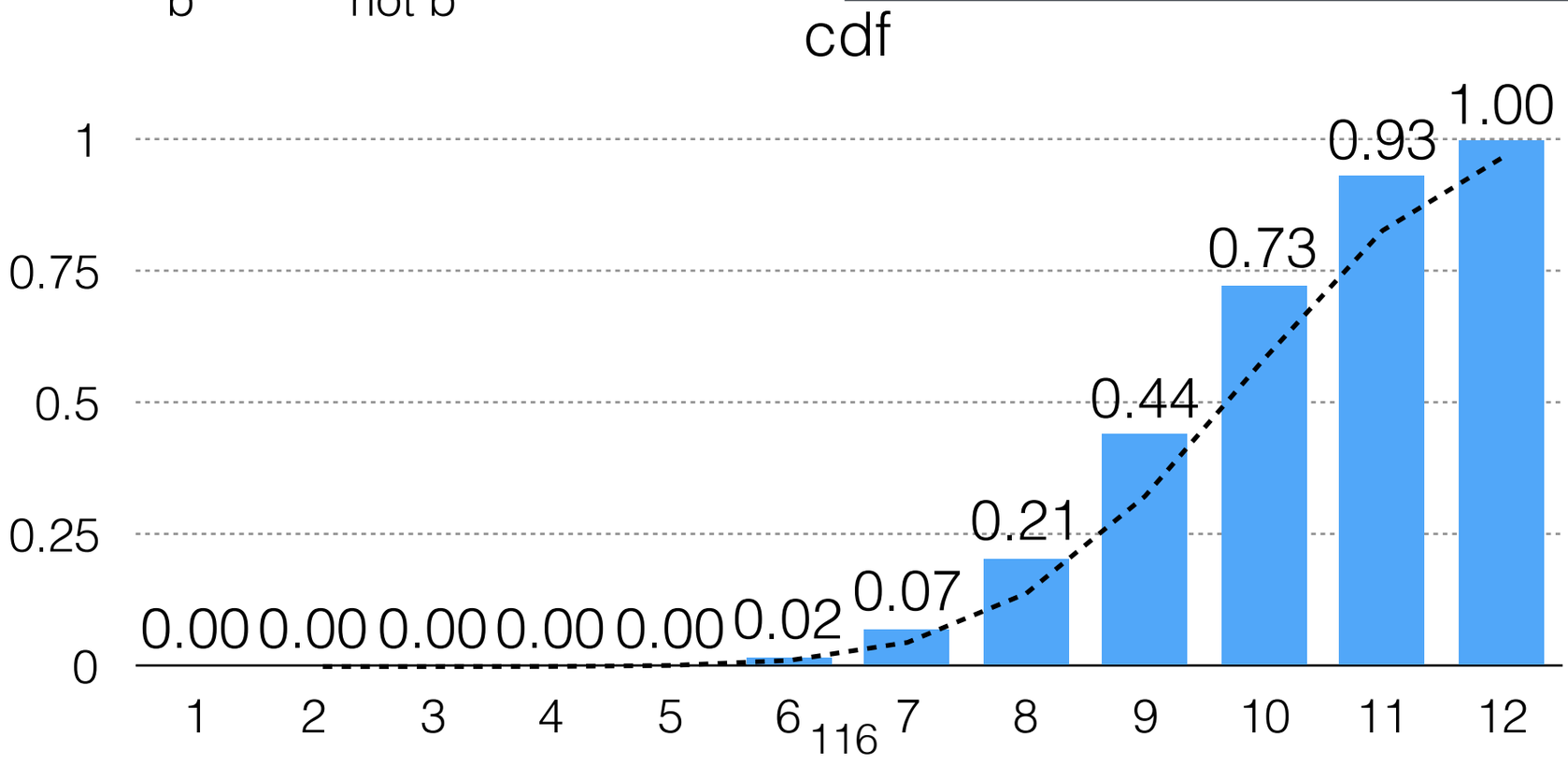


Are the answers to my clicker questions random?

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

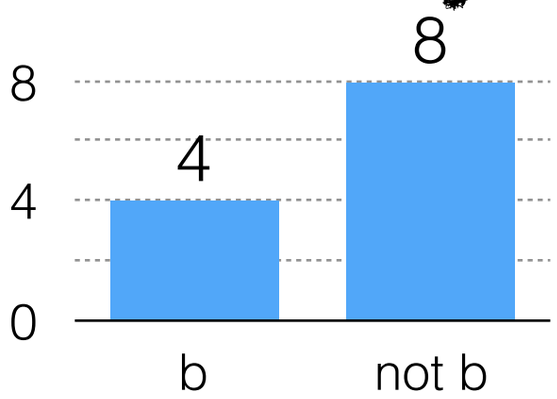


Is the 4 (b)s we observed significantly lower than what we would expect by chance, assuming that in fact 80% of answers are (b)?



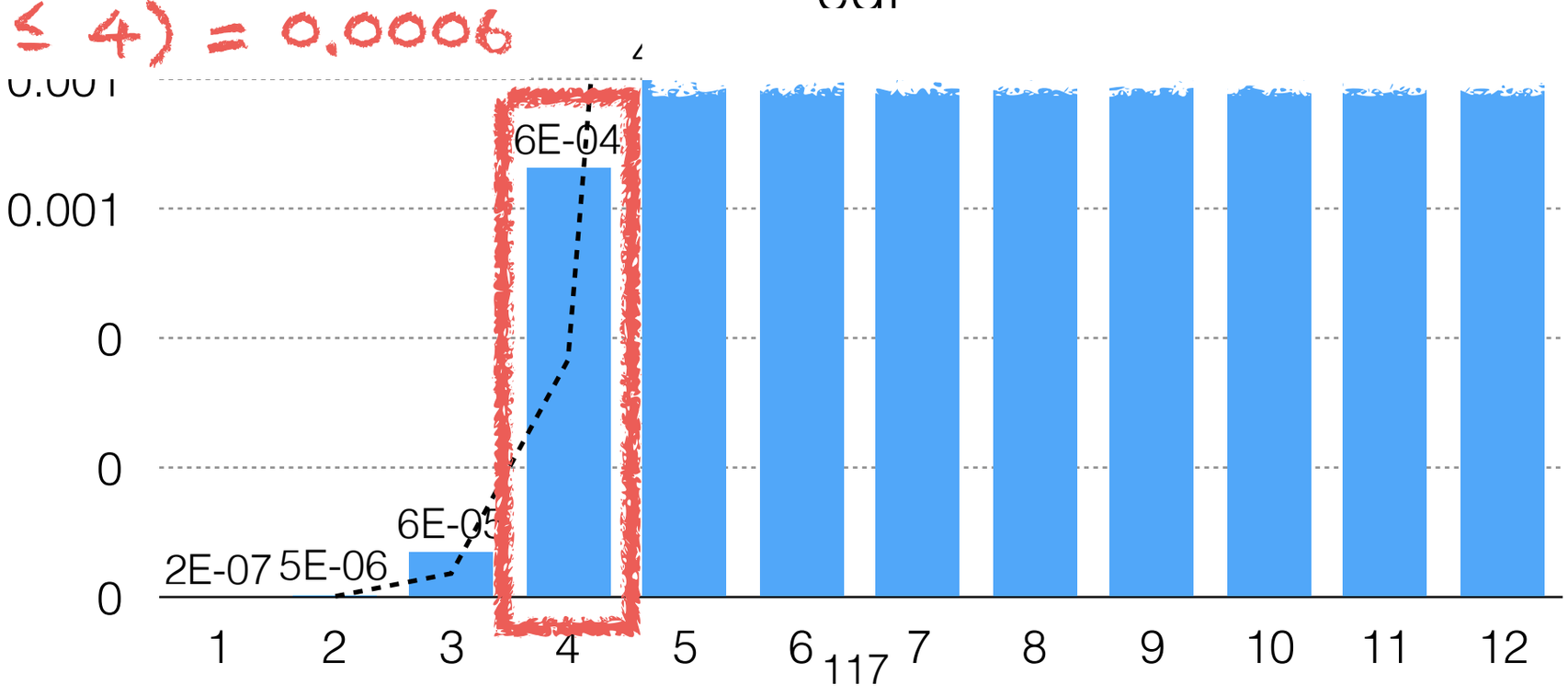
Are the answers to my clicker questions random?

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



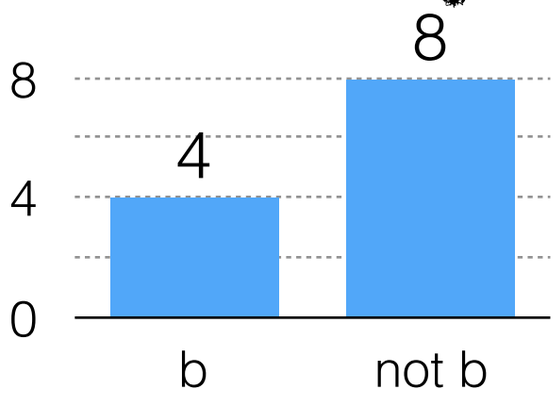
Is the 4 (b)s we observed significantly lower than what we would expect by chance, assuming that in fact 80% of answers are (b)?

P(#(b)s ≤ 4) = 0.0006



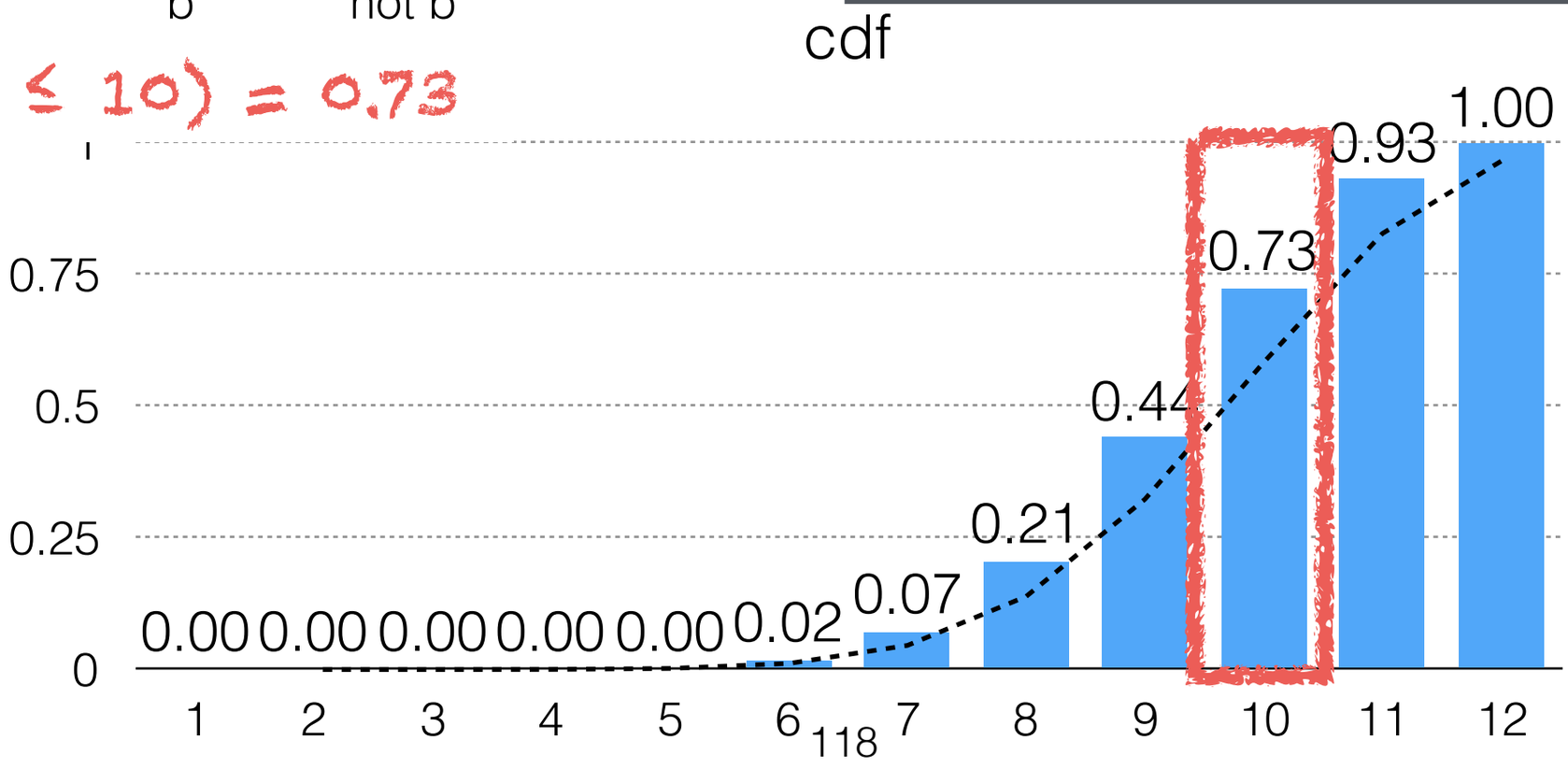
Are the answers to my clicker questions random?

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



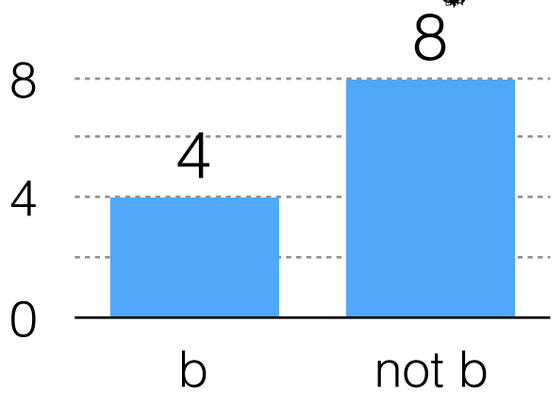
Is the 4 (b)s we observed significantly lower than what we would expect by chance, assuming that in fact 80% of answers are (b)?

P(#(b)s ≤ 10) = 0.73



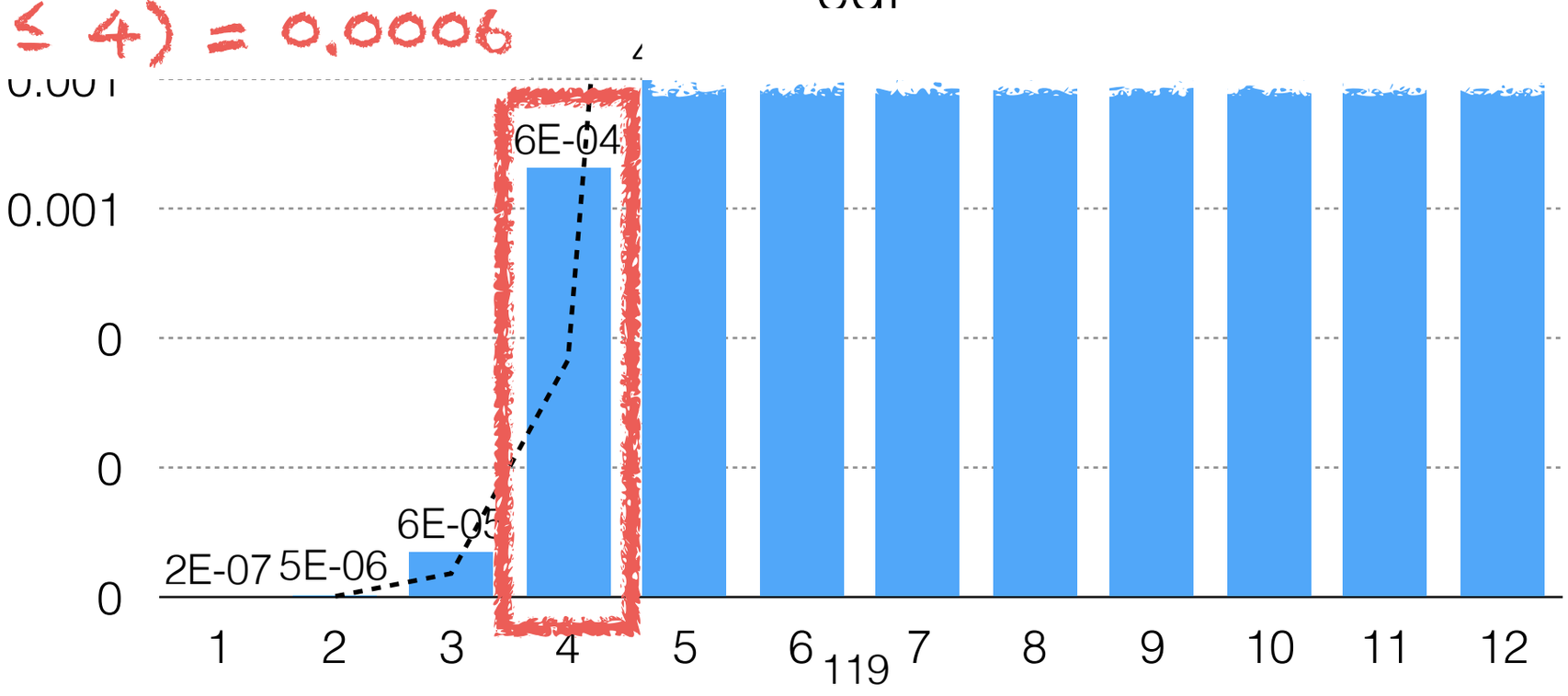
Are the answers to my clicker questions random?

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



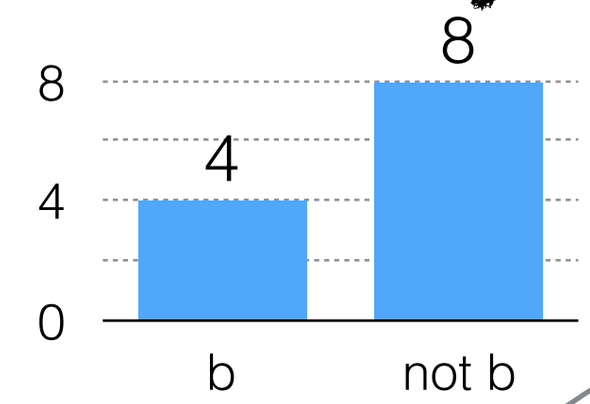
Is the 4 (b)s we observed significantly lower than what we would expect by chance, assuming that in fact 80% of answers are (b)?

P(#(b)s ≤ 4) = 0.0006

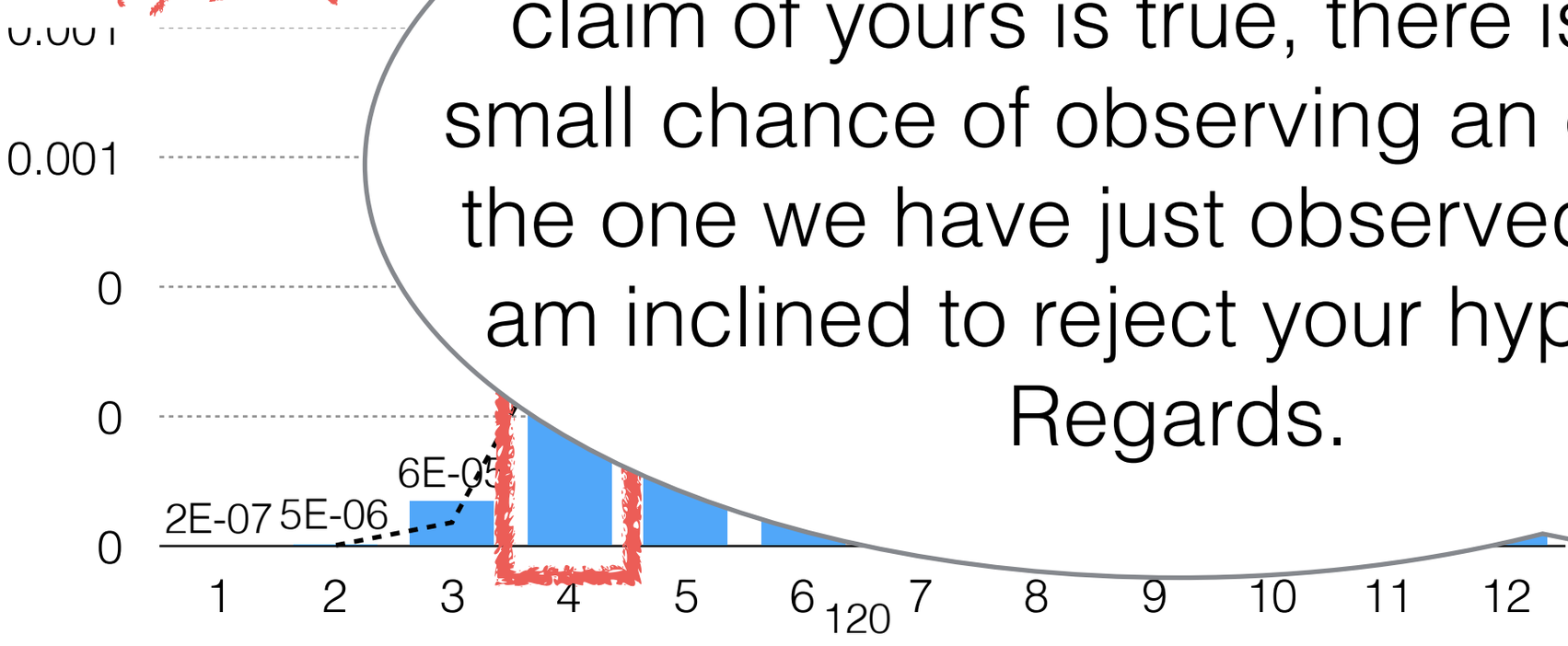


Are the answers to my clicker questions random?

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



P(#(b)s ≤ 4) = 0.000



Is the 4 (b)s we observed significantly lower than what we would [illegible] chance, assuming [illegible] (b)?

Dear friend, assuming this claim of yours is true, there is a very small chance of observing an event like the one we have just observed. Thus, I am inclined to reject your hypothesis. Regards.

Hypothesis Testing

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

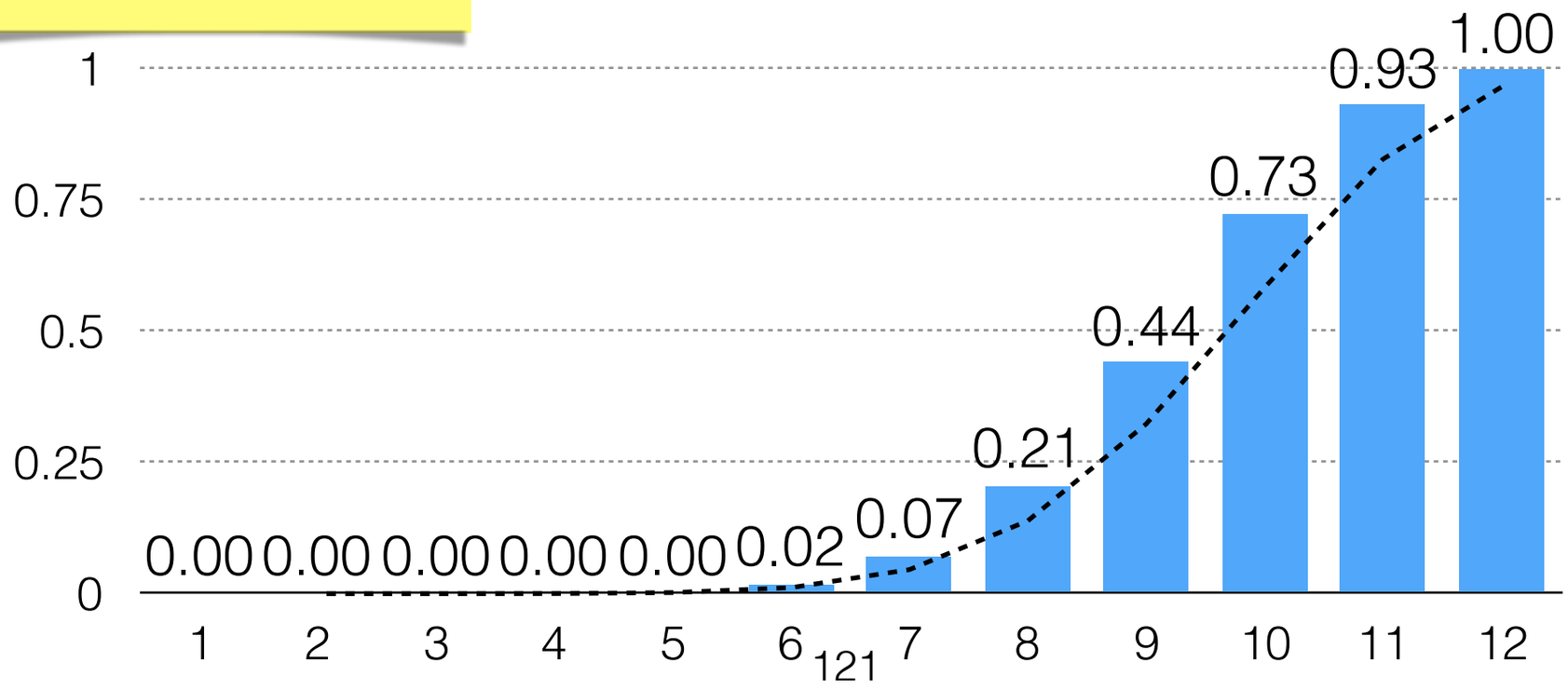
c	d	**b**	d
b	a	d	c
b	c	**b**	d

$\longrightarrow$ 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)"

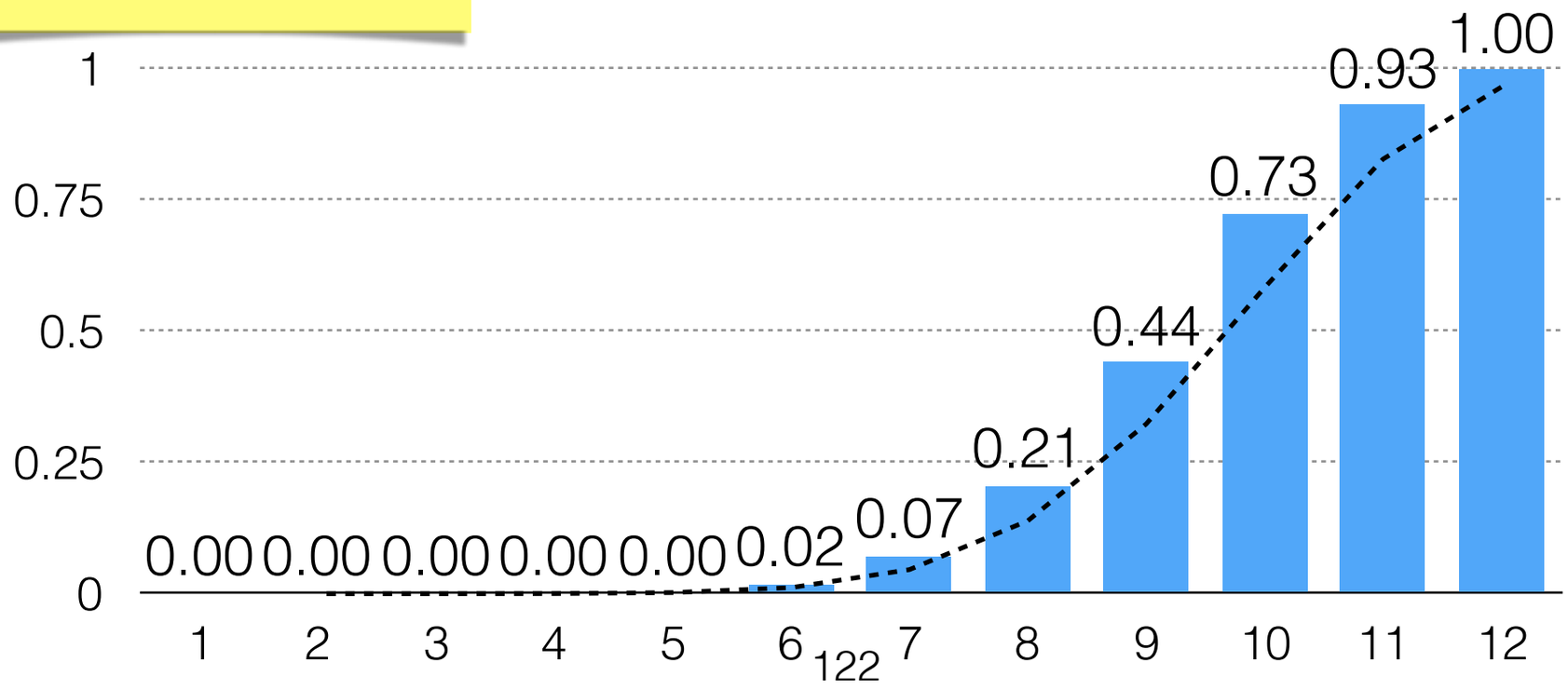
c	d	**b**	d
b	a	d	c
b	c	**b**	d

$\longrightarrow$ 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)"

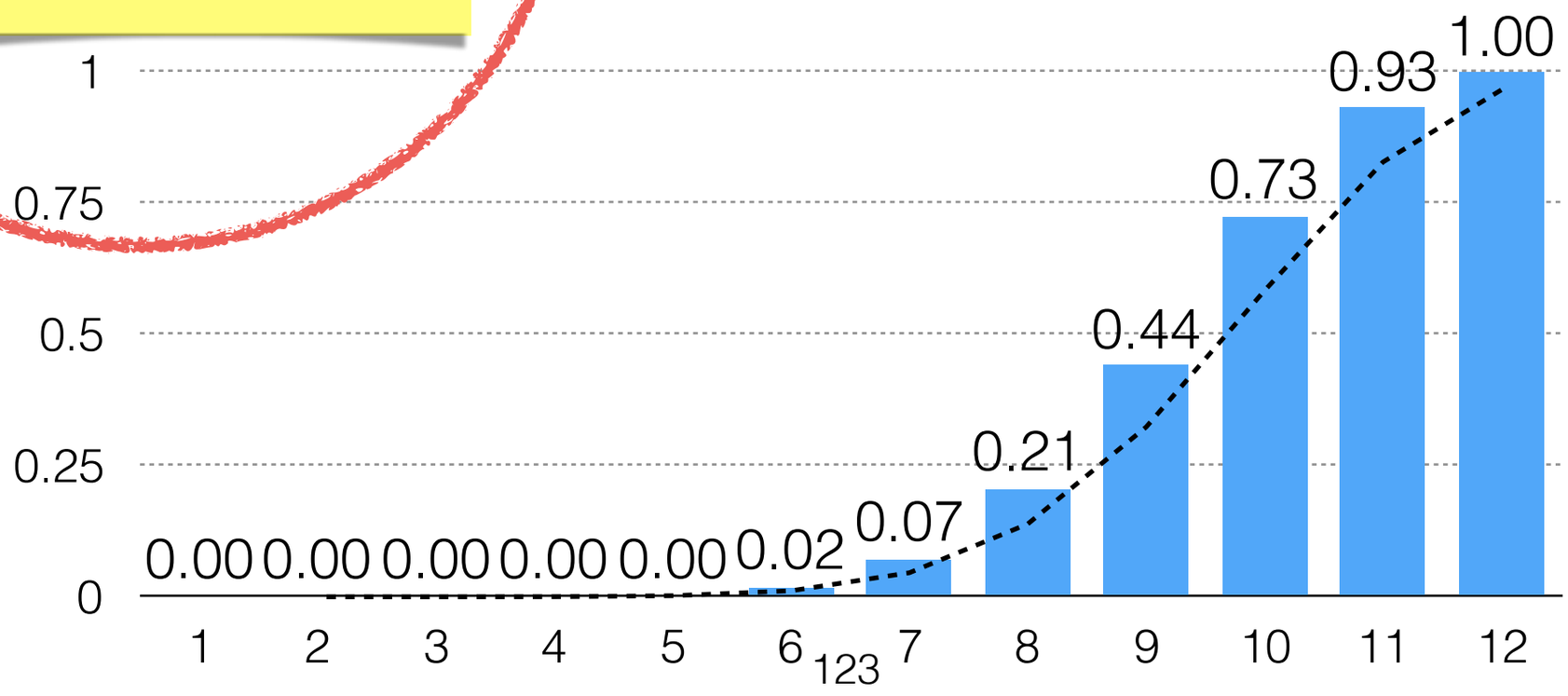
c	d	**b**	d
b	a	d	c
b	c	**b**	d

→ 4

cdf

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

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



Hypothesis Testing

Test Statistic

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

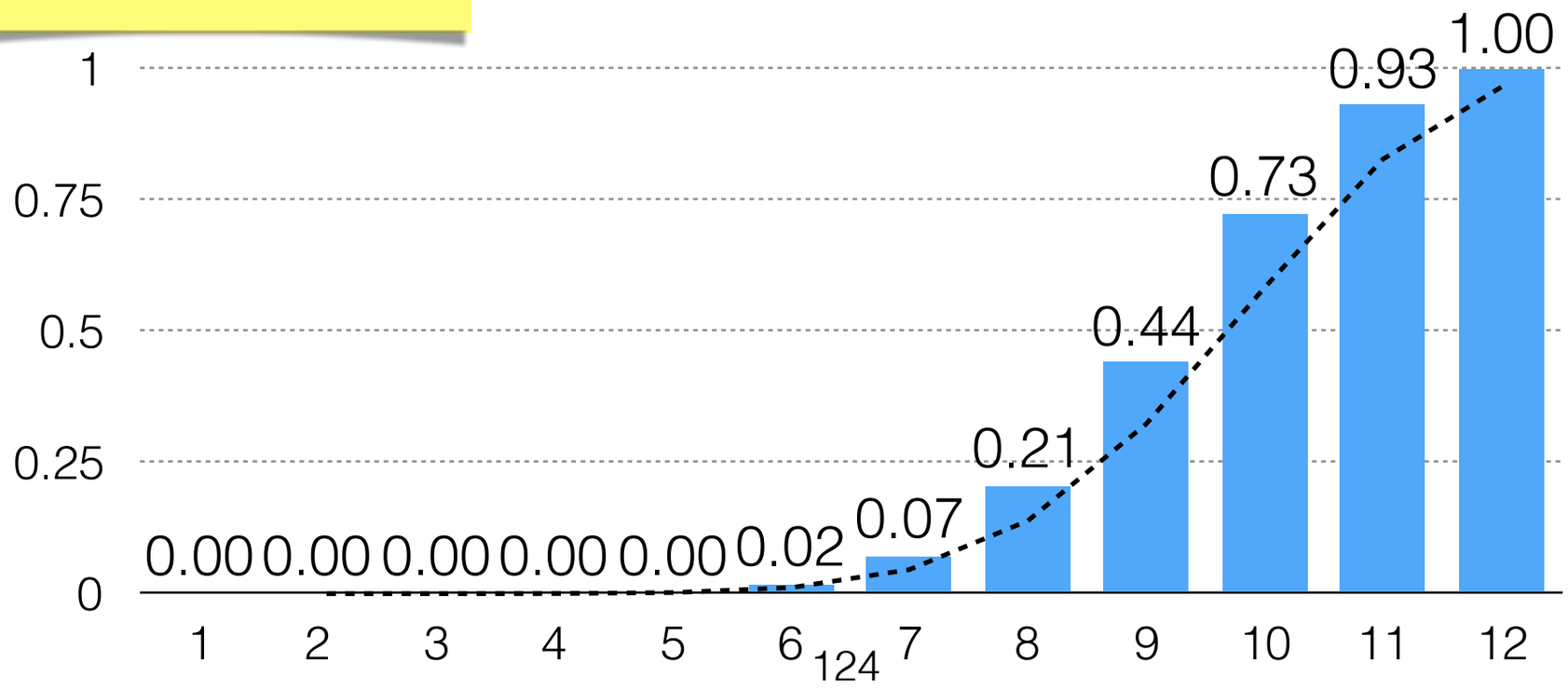
c	d	**b**	d
b	a	d	c
b	c	**b**	d

→ 4

cdf

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

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



Hypothesis Testing

Theoretical Distribution

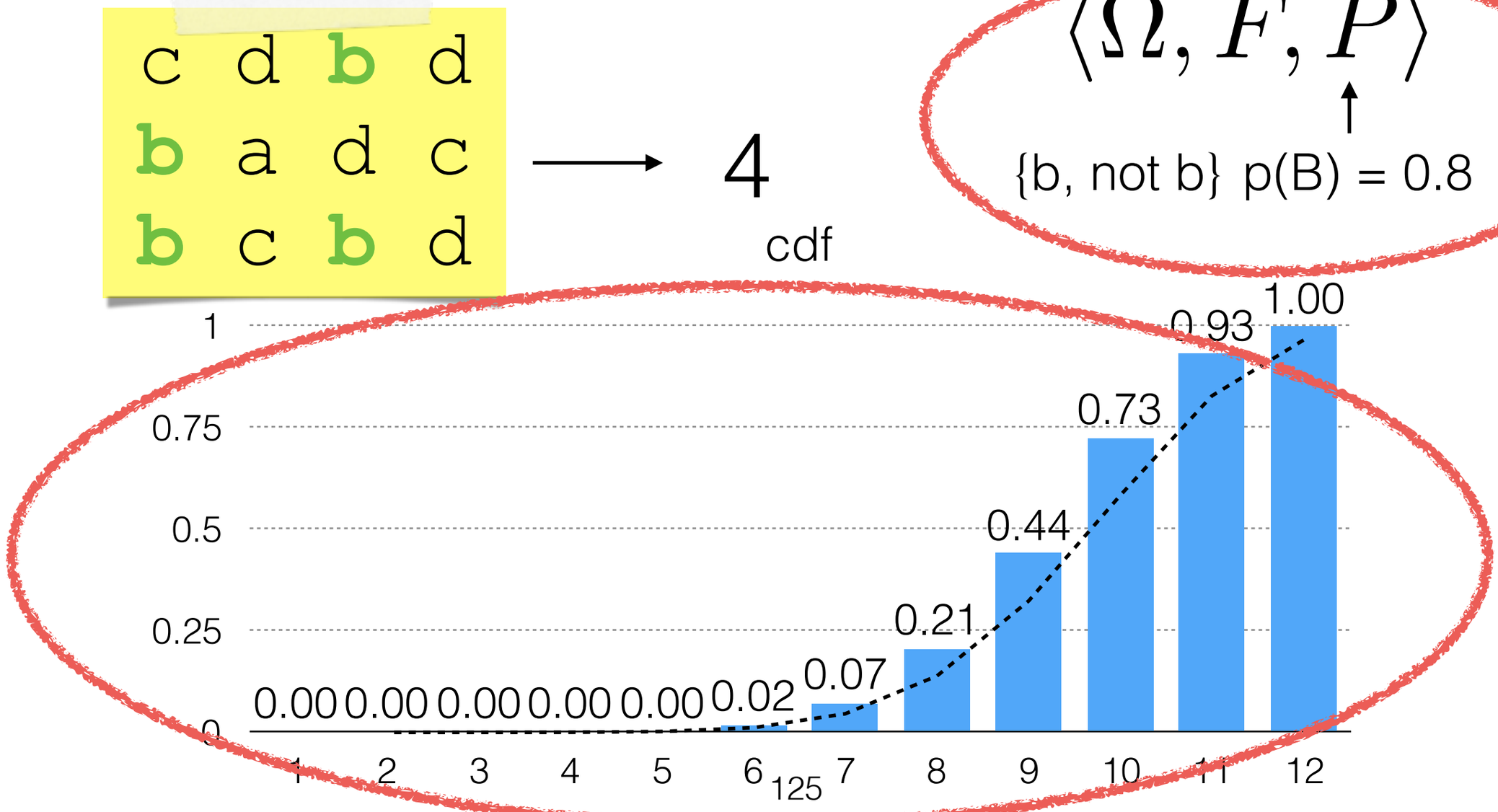
"I swear literally like 80% of the answers are just (b)"

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

$\longrightarrow$ 4

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

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



Hypothesis Testing

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

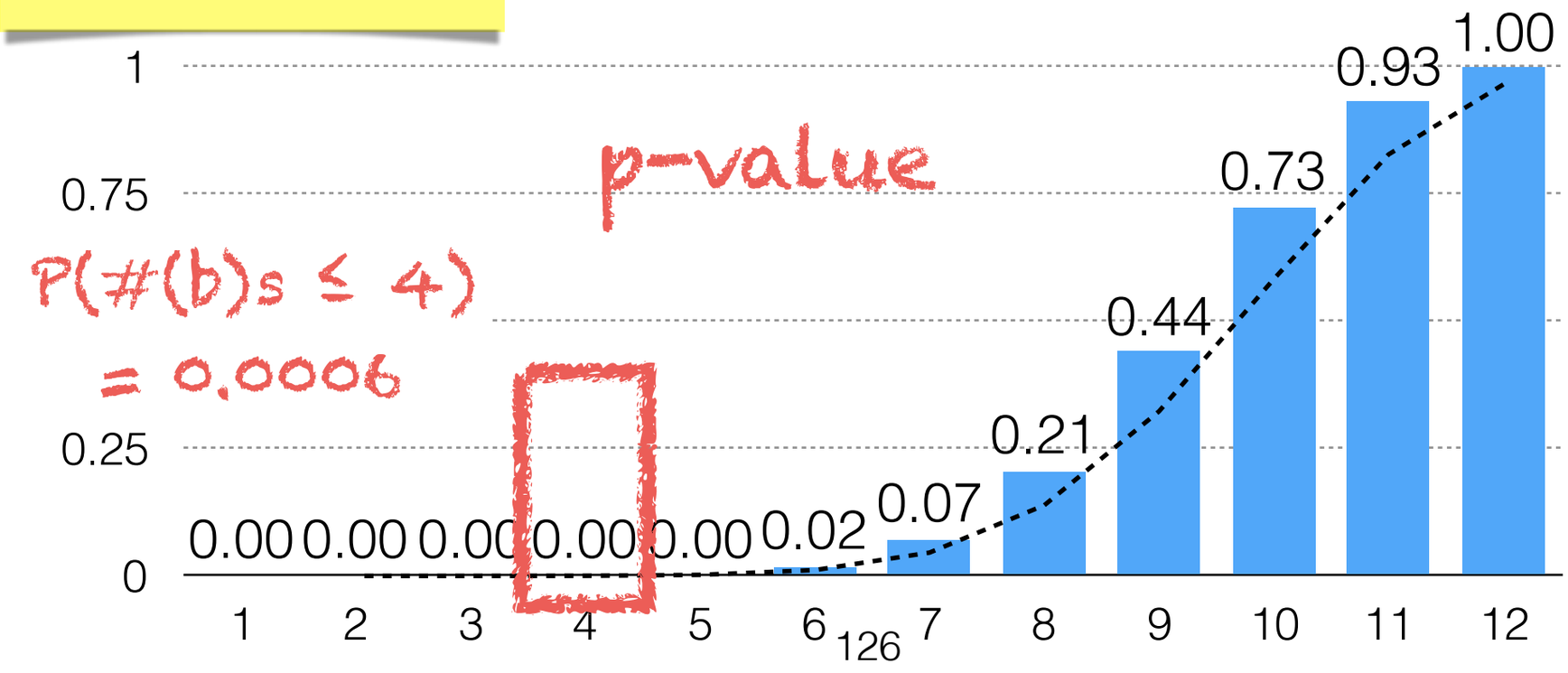
c	d	**b**	d
b	a	d	c
b	c	**b**	d

$\longrightarrow$ 4

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

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

cdf



Clicker Question!

Clicker Question!

Given all of this, is your friend wrong?

a) Yes!
b) No...

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

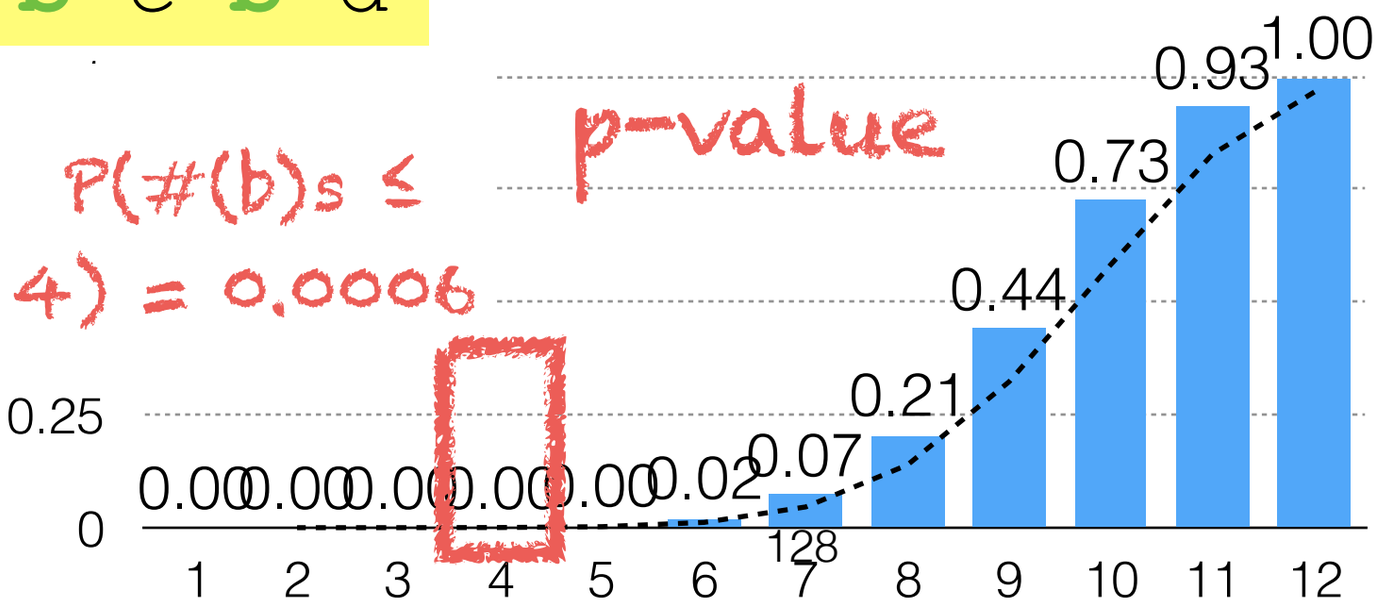
c	d	**b**	d
b	a	d	c
b	c	**b**	d

→ 4

cdf

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

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



Discussion Question!

Given all of this, is your friend wrong?

a) Yes!
b) No...

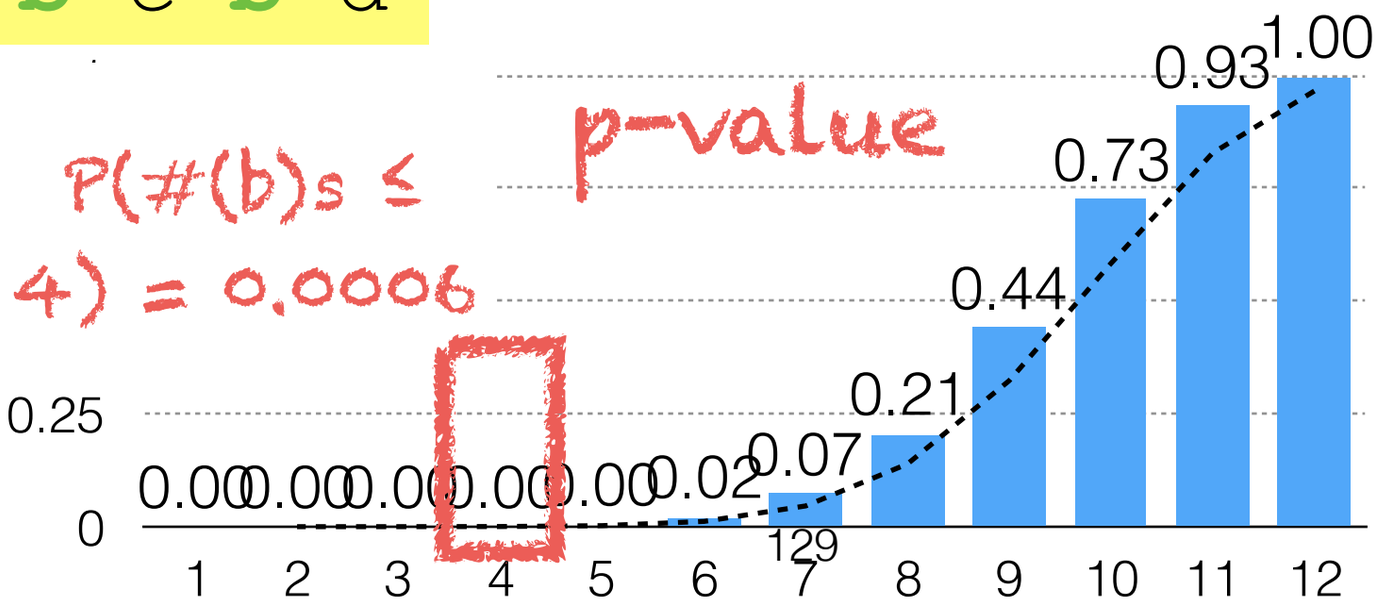
"I swear literally like 80% of the answers are just (b)"

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

→ 4 cdf

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

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



okie done now