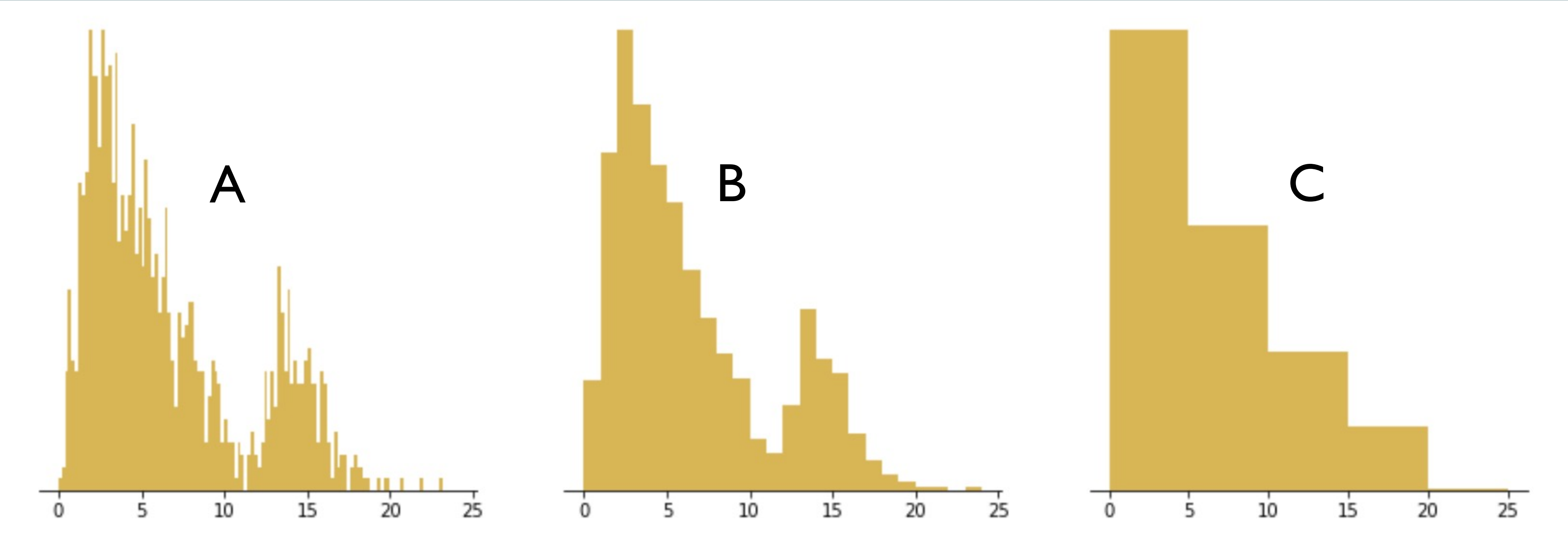
EXAM 1

Review

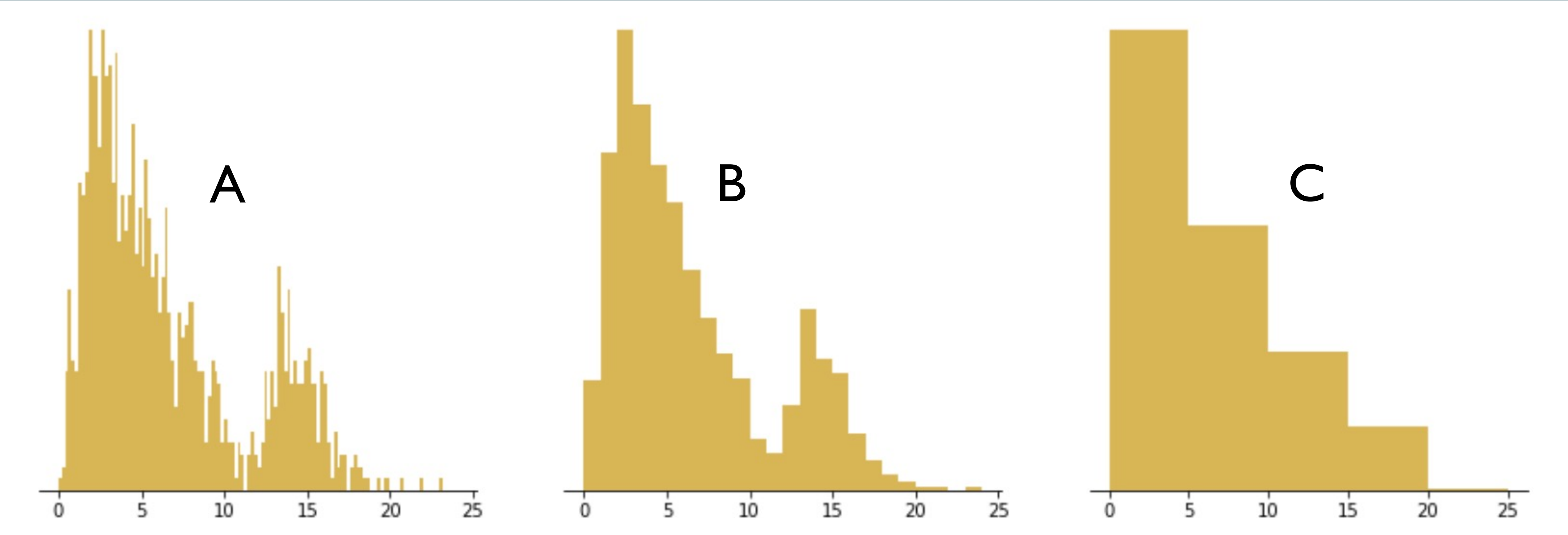
Histogram B has a bin size of 1.

Which histogram would you say has a bin size of $\frac{1}{5}$?



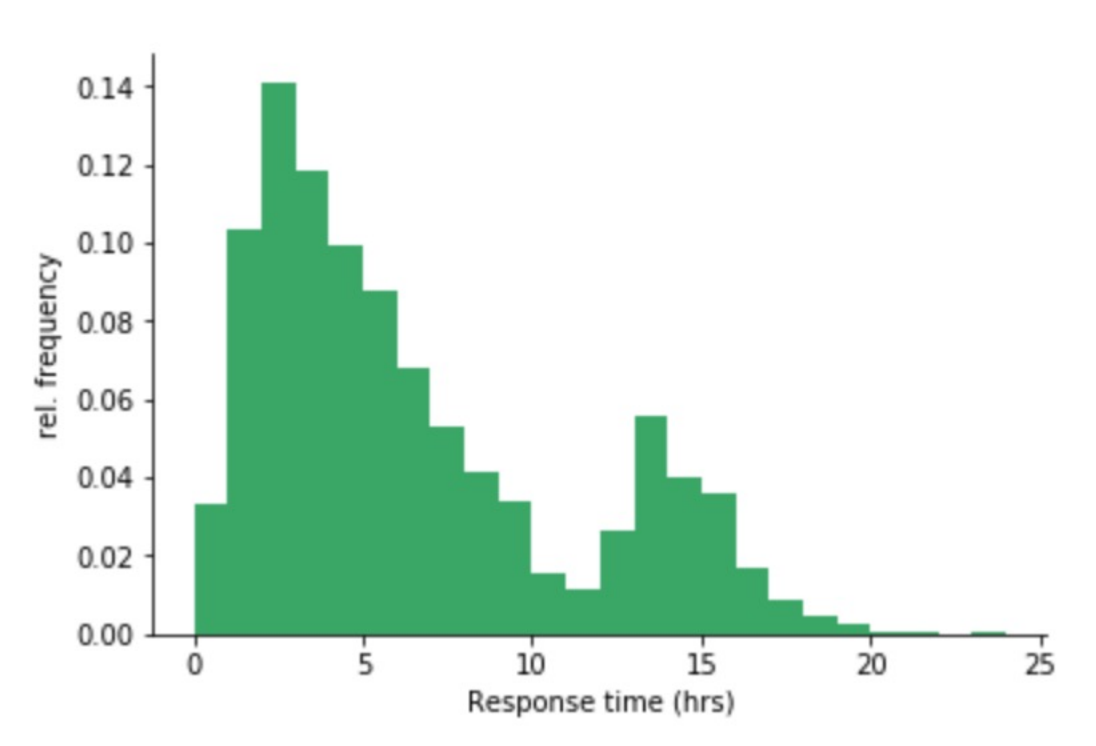
Histogram B has a bin size of 1.

Which histogram would you say has a bin size of $\frac{1}{5}$?

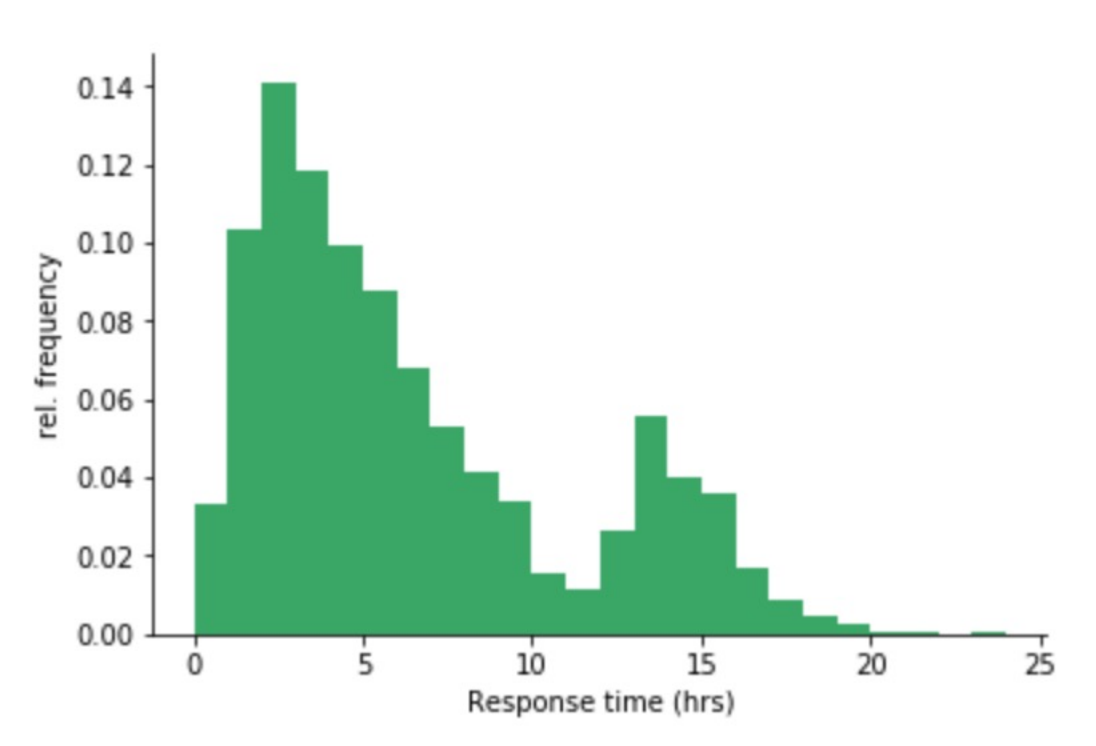


ANS: Histogram A

What is the mode of this distribution?

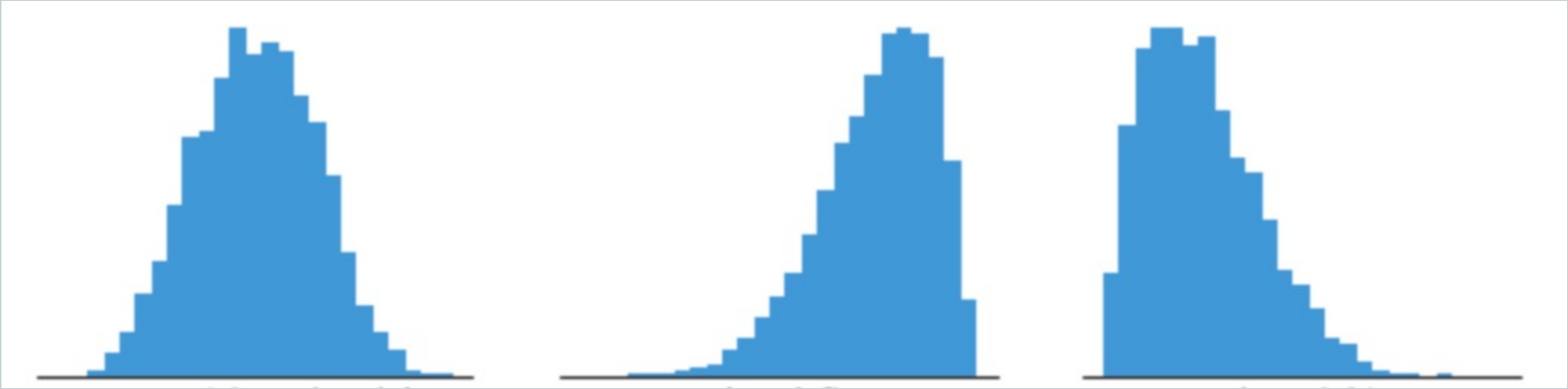


What is the mode of this distribution?



ANS: If the first height represents response time 0, then the mode is 2.
If the first height represents response time 1, then the mode is 3.

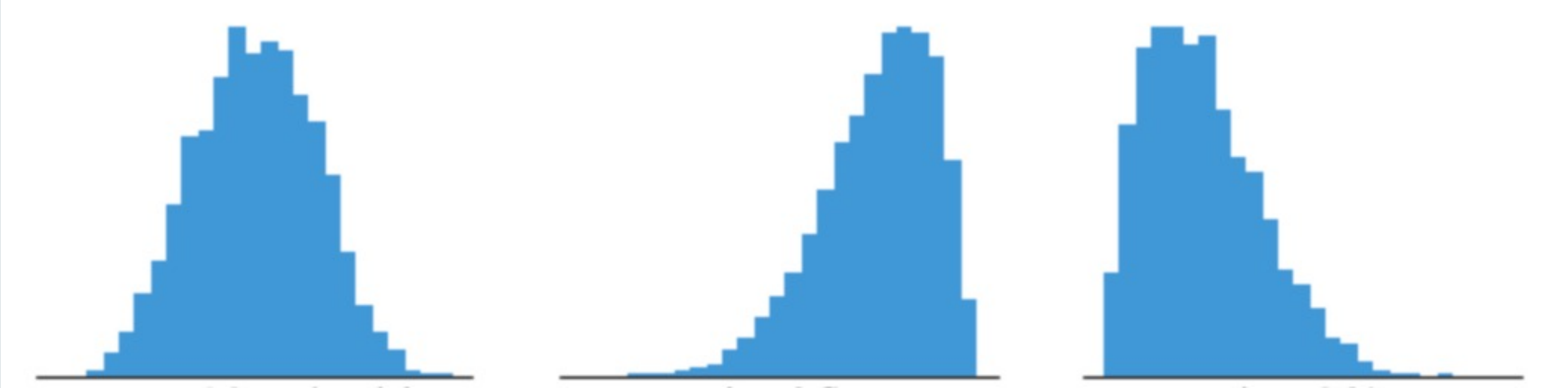
Can you give a shape descriptive name for these distributions?



A B C

D E F

Can you give a shape descriptive name for these distributions?



A

Symmetric

B

Left Skewed

C

Right Skewed

Uniform

D

Bimodal

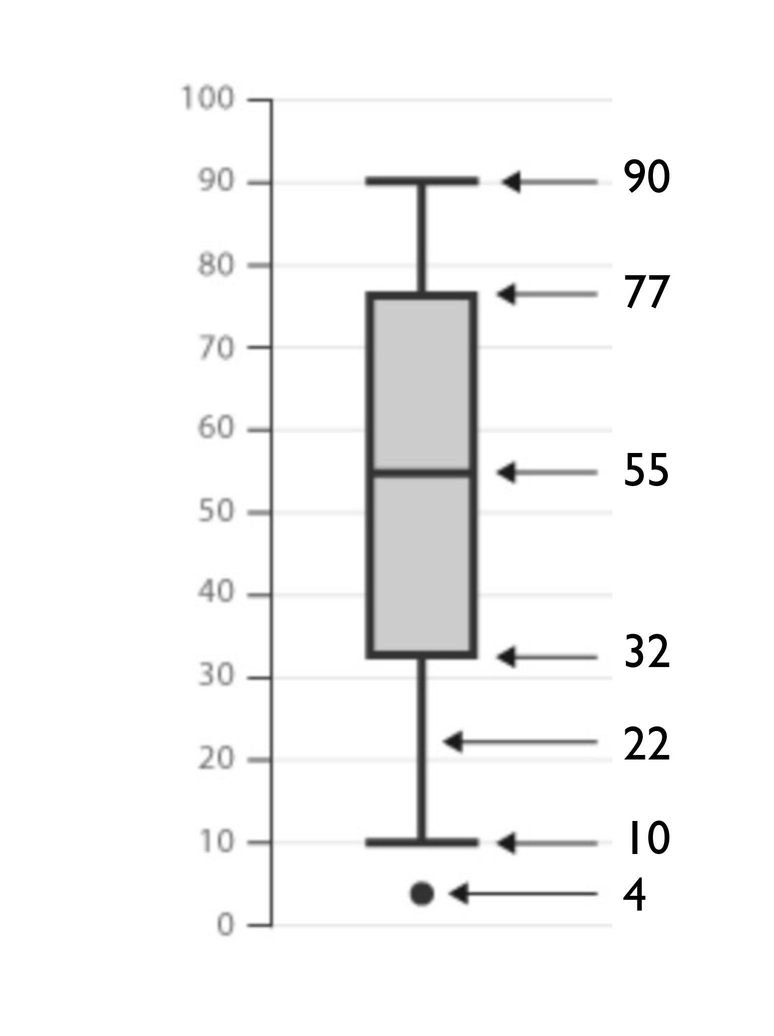
E

Multimodal

F

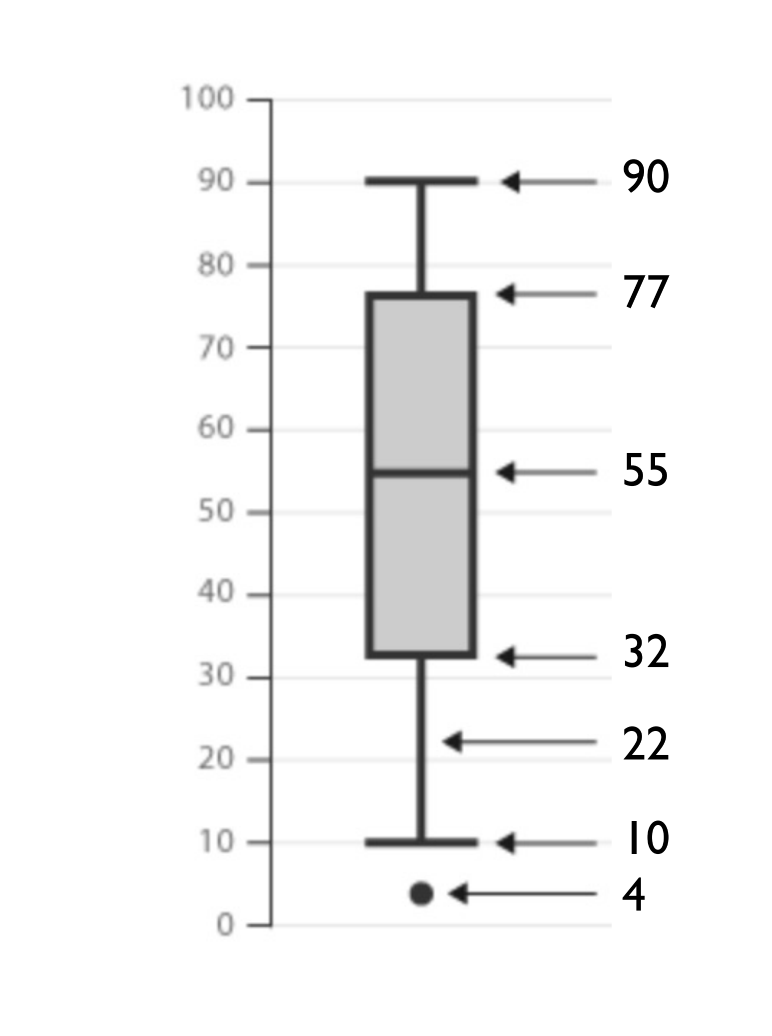
Consider the data set with the Box-and-Whisker plot seen at the right:

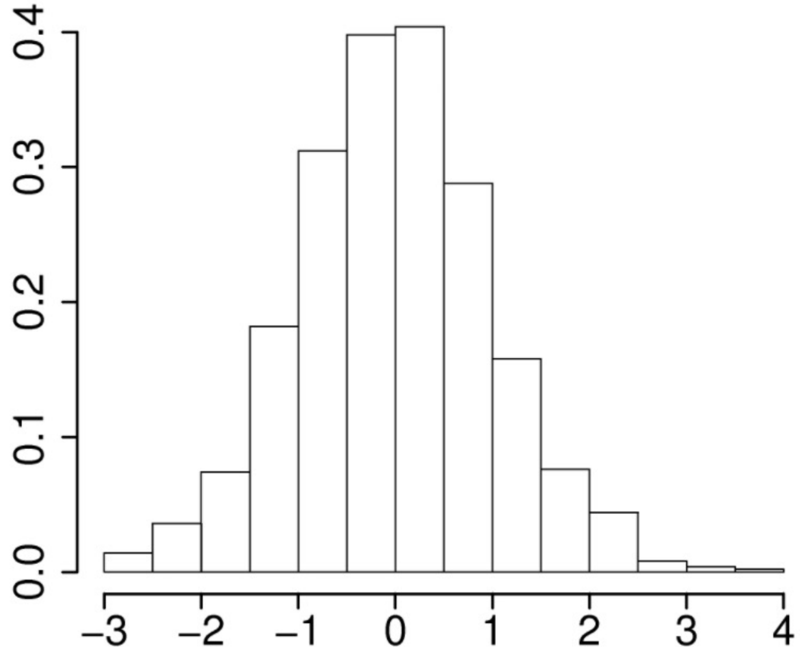
What is Q_1
What is Q_2
What is Q_3
What is the median
What is the minimum value
What is the IQR
Is there an outlier



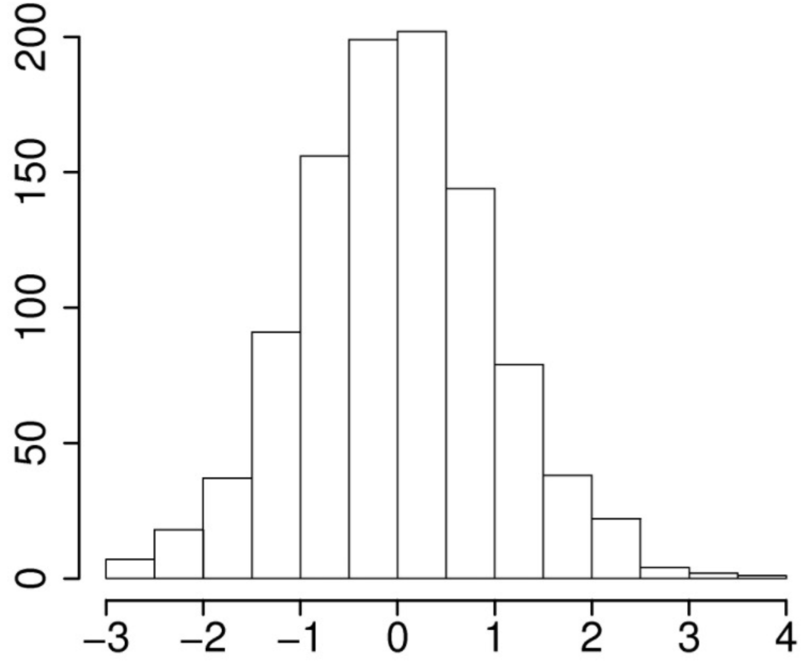
Consider the data set with the Box-and-Whisker plot seen at the right:

What is Q_1	32
What is Q_2	55
What is Q_3	77
What is the median	55
What is the minimum value	4
What is the IQR	45
Is there an outlier	yes



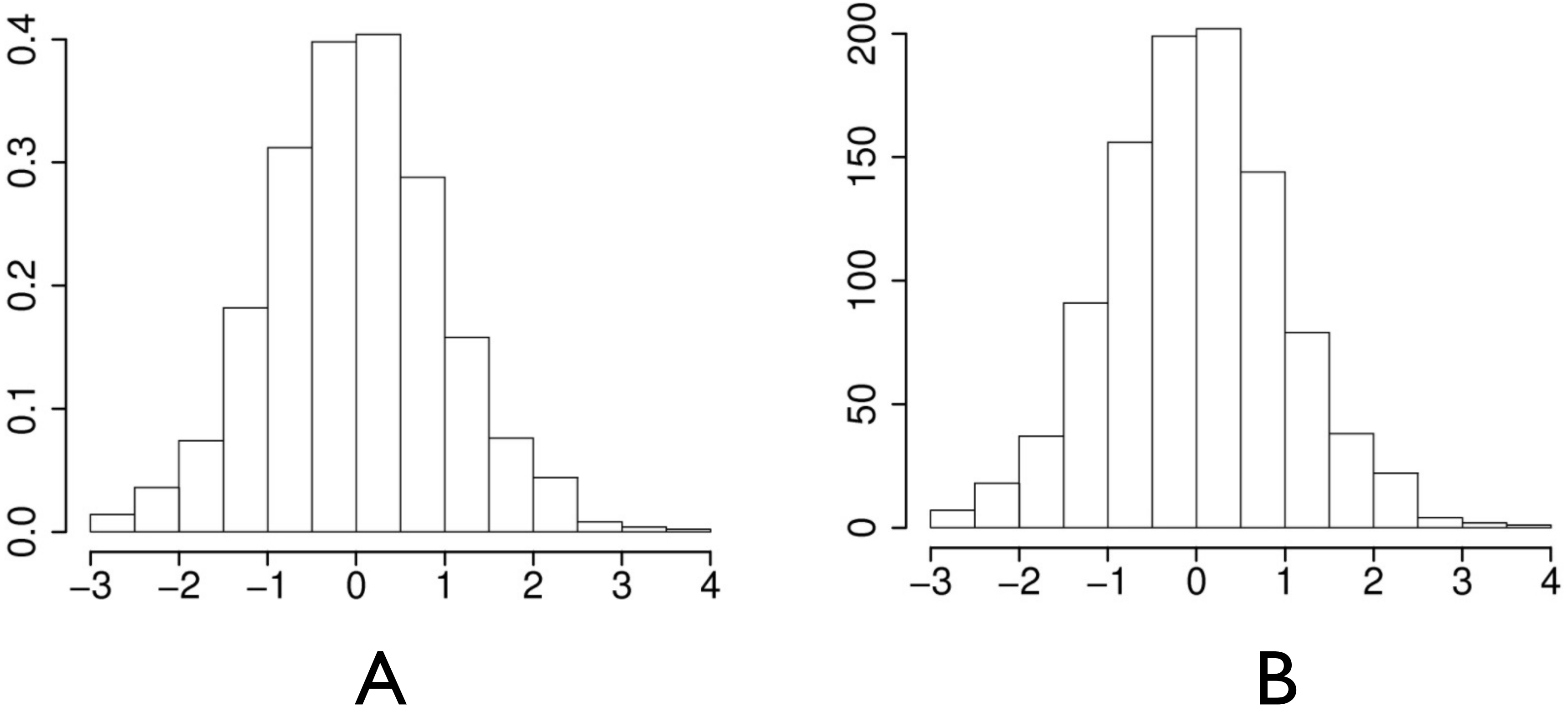


A



B

Which graph is a frequency histogram, and which one is a density histogram?



B is a frequency histogram, and A is a density histogram.

Consider a population of N elements.

Population $\{4, 3, 2, 3, 3, 1, 4, 1, 4, 5\}$

A population has parameters: μ, σ (3, 1.6)

You then sample n elements of that population.

Sample $\{1, 1, 3, 4, 4, 5\}$

A sample has statistics: $\bar{x}$, s (3, 2.8)

The sample mean, is $\bar{x} = 3$

The sample variance is $s = 2.8$.

BTW, `numpy.var(set_x, ddof = 1)`

You are collecting data from people who live in a campus dormitory.
The dorm is 60% female, and 40% male.

You are attempting to determine the opinion of the resident's concerning signage on the bathrooms. So, you are going to give a questionnaire to a sample of students.

You randomly select 16 males and 24 females from the dorm registry to give them the questionnaire.

What type of sample was inferred?

You are collecting data from people who live in a campus dormitory.
The dorm is 60% female, and 40% male.

You are attempting to determine the opinion of the resident's concerning signage on the bathrooms. So, you are going to give a questionnaire to a sample of students.

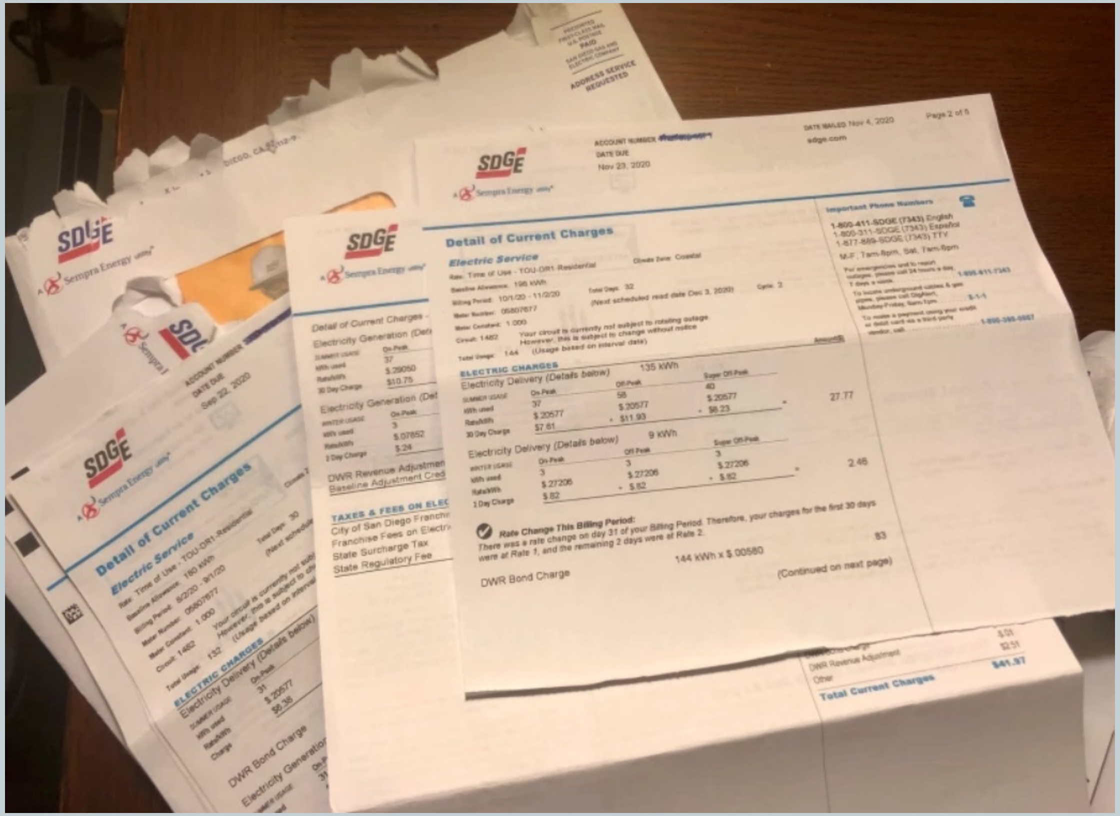
You randomly select 16 males and 24 females from the dorm registry to give them the questionnaire.

What type of sample was inferred?

ANS: A stratified Random Sample

Your bills are due at the beginning of the month.
During the first week of the month, you typically receive your bills in the mail.

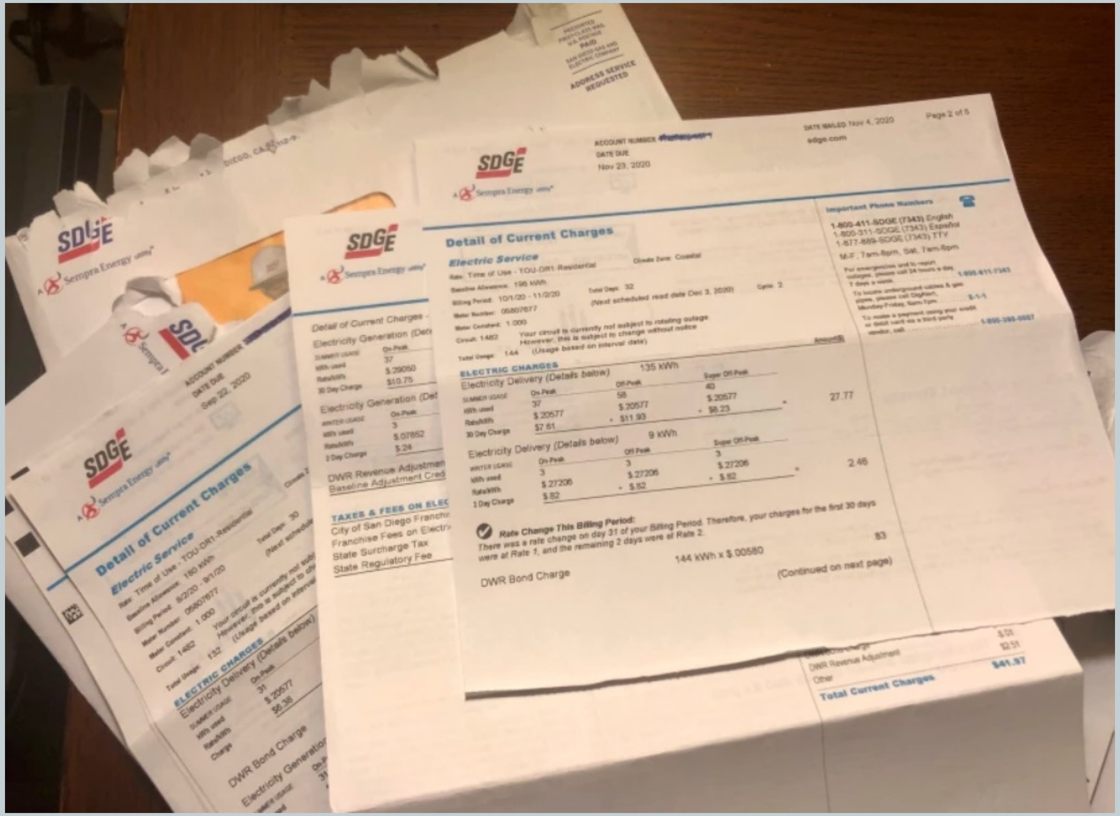
You decide to pay the bills in the order they are received.
How many ways can you pay the utilities bill, the phone bill, and the rent?



Your bills are due at the beginning of the month.
During the first week of the month, you typically receive your bills in the mail.

You decide to pay the bills in the order they are received.
How many ways can you pay the utilities bill, the phone bill, and the rent?

ANS: $3! = 6$



Given the set of numbers:

$\mathrm{U} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$

Consider the subsets: $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

$B = \{5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$

Which set has the larger cardinality: $(A \cup B)^c$ or $A^c \cap B^c$

Given the set of numbers:

$U = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20\}$

Consider the subsets: $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

$B = \{5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$

Which set has the larger cardinality: $(A \cup B)^c$ or $A^c \cap B^c$

ANS: DeMorgan's Law: $(A \cup B)^c = A^c \cap B^c$

Consider events E and F.

Suppose the probability that at least one of these events occurs is $3/4$.

What is the probability that neither occur?

Q

Consider events E and F.

Suppose the probability that at least one of these events occurs is ¾.

What is the probability that neither occur?

ANS: $P(neither) = 1 - P(at\ least\ one) = 1 - \frac{3}{4} = \mathbf{\frac{1}{4}}$

Consider two events called A and B.

$P(A) = \frac{2}{3}$ and $P(B) = \frac{1}{6}$ and $P(A \cap B) = \frac{1}{9}$, then $P(A \cup B) = ?$

Consider two events called A and B.

$P(A) = \frac{2}{3}$ and $P(B) = \frac{1}{6}$ and $P(A \cap B) = \frac{1}{9}$, then $P(A \cup B) = ?$

ANS: $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{2}{3} + \frac{1}{6} - \frac{1}{9} = \frac{13}{18}$

Suppose you design an experiment with a coin being tossed 4 times.

$|\Omega| = ?$

In the experiment mentioned above, what is the set of outcomes corresponding to A = "We throw tails 3 times"

Therefore, $P(A) = \ ?$

Suppose you design an experiment with a coin being tossed 4 times.

$|\Omega| = 2^4 = 16$

In the experiment mentioned above, what is the set of outcomes corresponding to A = "We throw tails 3 times"

$$\{(T, T, T, H),\ (T, T, H, T),\ (T, H, T, T),\ (H, T, T, T)\}$$

Therefore, $P(A) = \frac{4}{16} = 0.25$

Suppose you own a small business with 10 employees:

Al, Betty, Carl, Don, Eli, Fran, Gob, Hank, Ivy, Joe

Some of the employees are trained in various tasks.
Employees who can operate the fry machine: Al, Betty, Carl, Don, Eli, Fran
Employees who can manage the cash register: Betty, Don, Eli, Hank

F: "Can operate the fry machine"
C: "Can manage the cash register"

If you randomly pick an employee from your roster of all employees:

What is $P(C)$

What is $P(C|F)$

What if $P(C \cap F)$

What is $\frac{P(C \cap F)}{P(F)}$

Suppose you own a small business with 10 employees:

Al, Betty, Carl, Don, Eli, Fran, Gob, Hank, Ivy, Joe

Some of the employees are trained in various tasks.

Employees who can operate the fry machine: Al, Betty, Carl, Don, Eli, Fran

Employees who can manage the cash register: Betty, Don, Eli, Hank

F: "Can operate the fry machine"

C: "Can manage the cash register"

If you randomly pick an employee from your roster of all employees:

$$P(C) = \frac{4}{10} = 0.4$$

$$P(C|F) = \frac{3}{6} = 0.5$$

$$P(C \cap F) = \frac{3}{10} = 0.3$$

$\frac{P(C \cap F)}{P(F)} = 0.5$ This is equivalent to $P(C|F)$

Experiment: Roll three dice.

A: The sum is 6

B: At least one of the die is a 2.

$P(A) = ?$

$P(B) = ?$

$P(A|B) = ?$

$P(A \cap B) = ?$

Are the events independent?

Experiment: Roll three dice.

A: The sum is 6

B: At least one of the die is a 2.

Possible rolls to sum 6:

$(1,1,4)$, $(1,2,3)$, $(1,3,2)$, $(1,4,1)$

$(2,1,3)$, $(2,2,2)$, $(2,3,1)$

$(3,1,2)$, $(3,2,1)$

$(4,1,1)$

$P(A) = ?$ $\quad \frac{10}{6^3}$

$P(B) = ?$ $\quad 1 - P(no\ 2's) = 1 - \left(\frac{5}{6}\right)^3$

$P(A|B) = ?$ $\quad \frac{7}{10}$

$P(A \cap B) = ?$ $\quad \frac{7}{6^3}$

Are the events independent? No, because $P(A|B) \neq P(A)$,

or because $P(A \cap B) \neq P(A)P(B)$.

Suppose Z represent the number of times a 1 appeared in two independent throws of a 4-sided die.

[1] Describe the probability distribution of Z with F_Z.

Recall, F_Z is the cumulative distribution function as opposed to f which is the probability mass function.

[2] What type of distribution is this?

[3] If S is the sum of the dice, then what are the outcomes in $\{S = 5\}$?

[4] $P(\{S = 5\}) = \ ?$

Suppose Z represent the number of times a 1 appeared in two independent throws of a 4-sided die.

[1] Describe the probability distribution of Z with F_Z.

a	0	1	2
$p(a)$	$^9/_{16}$	$^6/_{16}$	$^1/_{16}$

$$F_Z(a) = \begin{cases} 0, & x < 0 \\ ^9/_{16}, & 0 \le x < 1 \\ ^{15}/_{16}, & 1 \le x < 2 \\ 1, & 2 \le x \end{cases}$$

[2] What type of distribution is this?

Z is the sum of two independent $Ber(^1/_4)$ distributed random variables, so Z has a $Bin(2, ^1/_4)$ distribution.

[3] If S is the sum of the dice, then what are the outcomes in $\{S = 5\}$?

$\{(1,4),\ (2,3),\ (3,2),\ (4,1)\}$

[4] $P(\{S = 5\}) = \ ^4/_{16} = 0.25$

Suppose that the probability a defendant is acquitted, given they are innocent, is 80%.

The probability a defendant is convicted given they are guilty is 82%.

Furthermore, 85% of defendants are in fact guilty.

What is the probability that a defendant is innocent, given they were convicted?

Suppose that the probability a defendant is acquitted, given they are innocent, is 80%.

The probability a defendant is convicted given they are guilty is 82%.

Furthermore, 85% of defendants are in fact guilty.

What is the probability that a defendant is innocent, given they were convicted?

$$P(I|C) = \frac{P(C|I) \cdot P(I)}{P(C|I) \cdot P(I) + P(C|G) \cdot P(G)} = \frac{(.2)(.15)}{(.2)(.15) + (.82)(.85)} = 0.3729$$

Total Probability

Suppose there are two bags in a box, which contain the following marbles:

Bag 1: 7 red marbles and 3 green marbles

Bag 2: 2 red marbles and 8 green marbles

If we randomly select one of the bags and then randomly select one marble from that bag, what is the probability that it's a green marble?

Total Probability

Suppose there are two bags in a box, which contain the following marbles:

Bag 1: 7 red marbles and 3 green marbles

Bag 2: 2 red marbles and 8 green marbles

If we randomly select one of the bags and then randomly select one marble from that bag, what is the probability that it's a green marble?

ANS:

$$P(G) = P(G|B1) \cdot P(B1) + P(G|B2) \cdot P(B2)$$

$$P(G) = \frac{3}{10} \cdot \frac{1}{2} + \frac{8}{10} \cdot \frac{1}{2} = \frac{11}{20} = 0.55$$

The probability density function f of a continuous random variable X is given by:

$$f(x) = \begin{cases} cx + 3, & -3 \leq x \leq -2 \\ 3 - cx, & 2 \leq x \leq 3 \\ 0, & elsewhere \end{cases}$$

What is c ?

The probability density function f of a continuous random variable X is given by:

$$f(x) = \begin{cases} cx + 3, & -3 \leq x \leq -2 \\ 3 - cx, & 2 \leq x \leq 3 \\ 0, & elsewhere \end{cases}$$

What is c ?

ANS:

$$\int_{-3}^{-2} cx + 3 \ dx + \int_{2}^{3} 3 - cx \ dx = 1$$

work, work, work,…

$$c = 1$$

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\, dx$$

for $-\infty < a < \infty$

The above indicates that X has a ___________ distribution

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\, dx$$

for $-\infty < a < \infty$

The above indicates that X has a ___________ distribution

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\, dx$$

for $-\infty < a < \infty$

The above indicates that X has a $N(\mu, \sigma^2)$ distribution

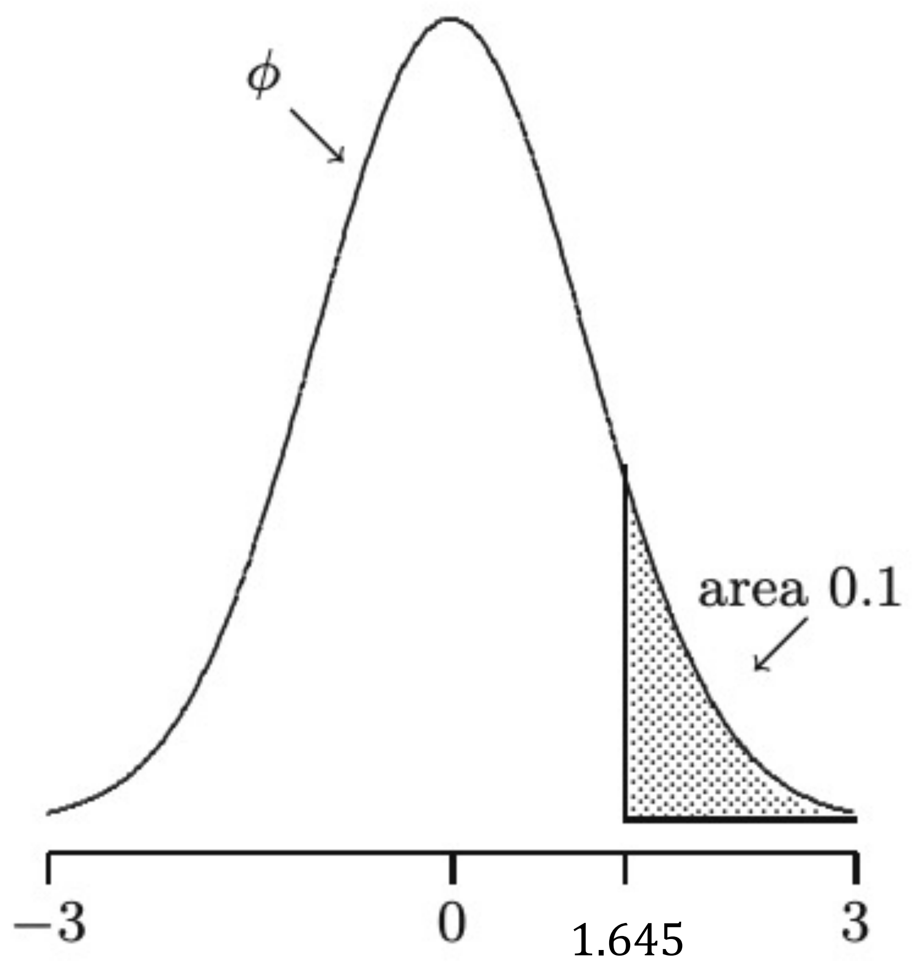
---

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

for $-\infty < a < \infty$

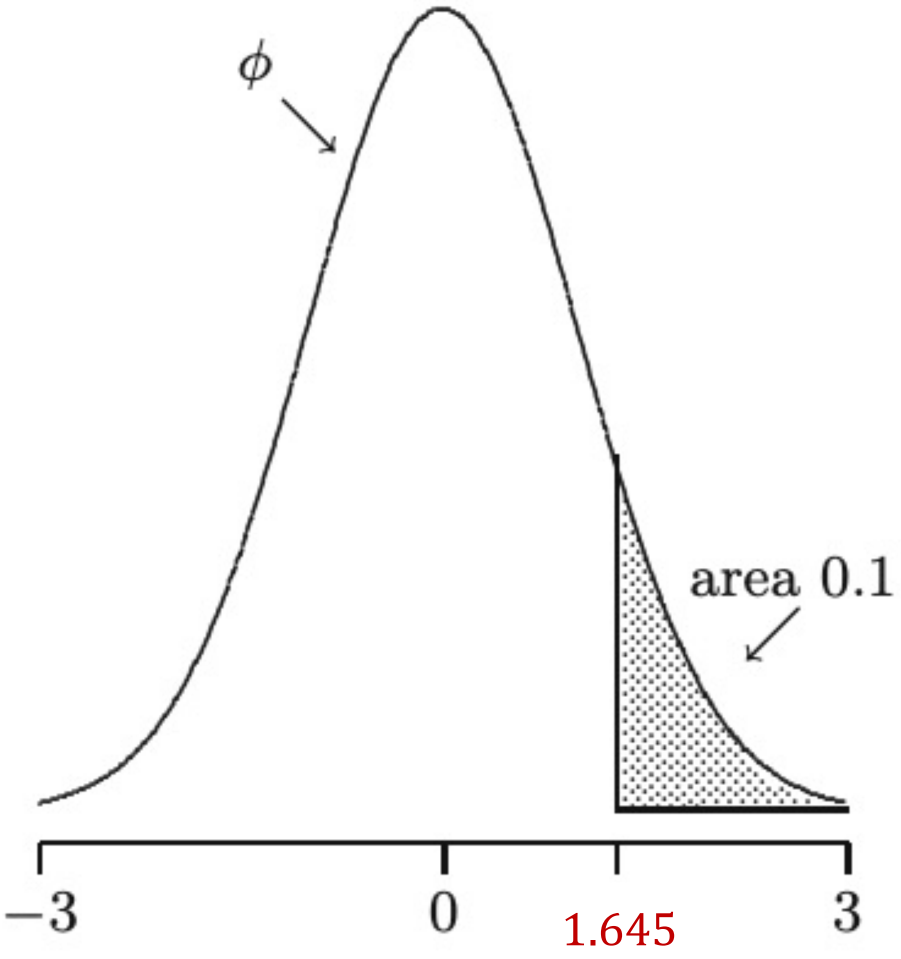
The above indicates that X has a $N(0, 1)$ distribution

What is the 90^{th} percentile of the the standard normal distribution?

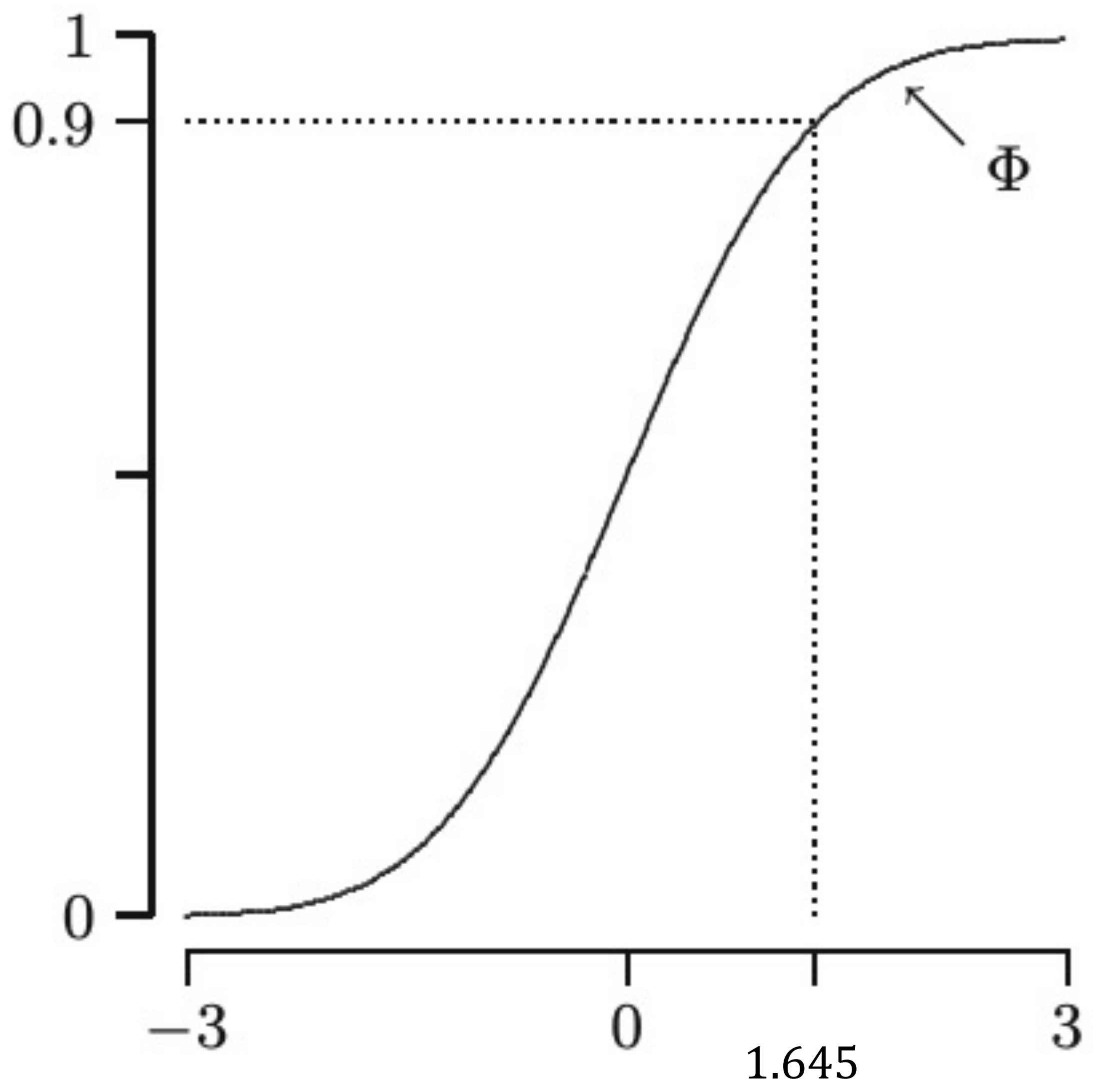


What is the 90th percentile of the the standard normal distribution?

ANS: 1.645

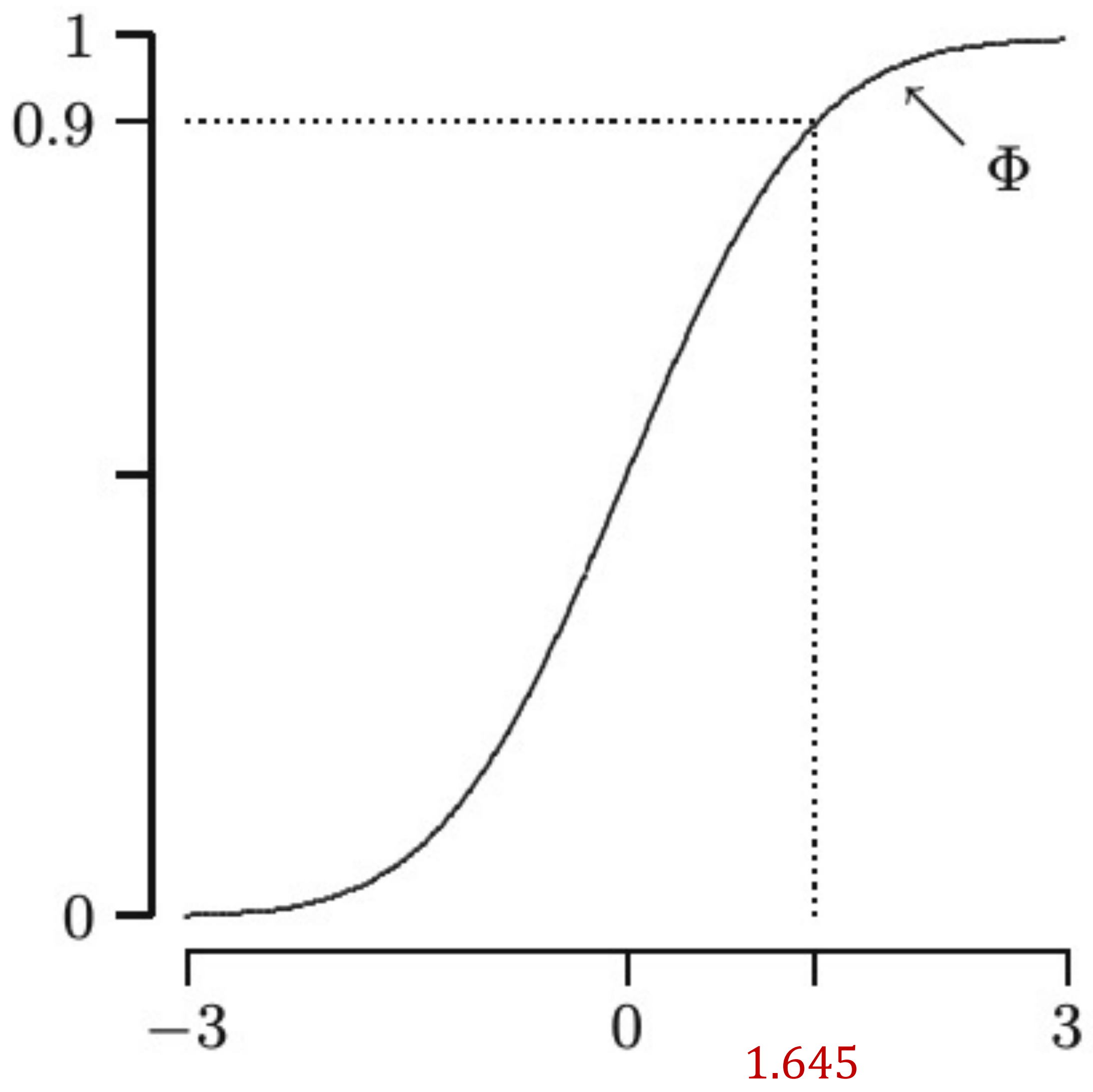


What is the 90th percentile of the standard normal distribution?



What is the 90th percentile of the standard normal distribution?

ANS: 1.645



Name the distribution

Suppose you are looking for easter eggs in various hiding places.

If your probability of success is 0.2, what is the probability you find an egg on your third try?

Name the distribution

Suppose you are looking for easter eggs in various hiding places.
If your probability of success is 0.2, what is the probability you find an egg on your third try?

ANS:

Geometric

$$f(x) = (1 - p)^{x-1} \cdot p$$

$$P(x = 3) = (1 - 0.2)^{3-1} \cdot (0.2) = 0.128$$

The binomial distribution versus the negative binomial distribution

The binomial

fixed number of trials

A random variable Y = the number of successes

The negative binomial

The number of trials, n, is not fixed

A random variable Y = the number of trials needed to make r successes.

Both

Each trial is independent

Only two outcomes are possible (Success, Failure)

Probability of success, p, for each trial is constant

Name the Distribution

Take a standard deck of cards, shuffle them, and choose a card.
Replace the card and repeat twenty times.
Y is the number of aces you draw.

Name the Distribution

Take a standard deck of cards, shuffle them, and choose a card.
Replace the card and repeat twenty times.
Y is the number of aces you draw.

ANS:

The Binomial

Name the Distribution

Take a standard deck of cards, shuffle them, and choose a card. Replace the card and repeat until you have drawn two aces.
Y is the number of draws needed to draw two aces.

Name the Distribution

Take a standard deck of cards, shuffle them, and choose a card. Replace the card and repeat until you have drawn two aces.
Y is the number of draws needed to draw two aces.

ANS:
As the number of trials isn't fixed (i.e. you stop when you draw the second ace), this makes it a negative binomial distribution.

Name the Distribution

You are surveying people exiting from a polling booth and asking them if they voted independent. The probability (p) that a person voted independent is 20%. What is the probability that 15 people must be asked before you can find 5 people who voted independent?

Name the Distribution

You are surveying people exiting from a polling booth and asking them if they voted independent. The probability (p) that a person voted independent is 20%. What is the probability that 15 people must be asked before you can find 5 people who voted independent?

ANS:

Negative Binomial

$$nb(10, 5, 0.2) = \binom{14}{4} (0.2)^5 (0.8)^{10}$$

The Exponential Distribution

The random variable for the exponential distribution is continuous and often measures a passage of time, although it can be used in other applications.

Typical questions may be,

"what is the probability that some event will occur within the next x hours or days," or "what is the probability that some event will occur between x_1 hours and x_2 hours," or "what is the probability that the event will take more than x_1 hours to perform?"

In short, the random variable X equals:

(a) the time between events or

(b) the passage of time to complete an action, e.g., wait on a customer.

$$f(x) = \frac{1}{\mu} e^{-\frac{1}{\mu}x}$$

The Poisson Distribution

Many experimental situations occur in which we observe the counts of events within a set unit of time, area, volume, length etc.

For example:

The number of cases of a disease in different towns;
The number of mutations in given regions of a chromosome;
The number of dolphin pod sightings along a flight path through a region;
The number of particles emitted by a radioactive source in a given time;
The number of births per hour during a given day.

In such situations we are often interested in whether the events occur randomly in time or space, or not.

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Name the distribution

Suppose, according to the latest police reports, 80% of all petty crimes are unresolved, and in your town, at least three of such petty crimes are committed. The three crimes are all independent of each other. From the given data, what is the probability that one of the three crimes will be resolved?

Name the distribution

Suppose, according to the latest police reports, 80% of all petty crimes are unresolved, and in your town, at least three of such petty crimes are committed. The three crimes are all independent of each other. From the given data, what is the probability that one of the three crimes will be resolved?

ANS:

Binomial Distribution

Number of fixed trials (n):	3 (Number of petty crimes)
Number of mutually exclusive outcomes:	2 (solved and unsolved)
The probability of success (p):	0.2 (20% of cases are solved)
Independent trials:	Yes

$$\binom{3}{1}(0.2)^1(0.8)^2 = 0.384$$

The 90-minute written exam is Wednesday!