Basic Probability

Donna Dietz

American University

dietz@american.edu

STAT 202 - Spring 2020

Review

Things you (probably) already know:

- Mutually exclusive events can be combined into new categories
- Probabilites for mutually exclusive events can be added
- The joint probabilites (overall probabilities) for independent events can be multiplied.
- The probability of something happening or not happening adds to 100%
- If an outcome is broken down into non-overlapping outcomes that cover all possible outcomes, the sum of their probabilities is 100%

A formula

This idea that something either does or does not happen is a recurring theme.

$$\bar{P} + P = 1$$

$$\bar{P} = 1 - P$$

$$P = 1 - \bar{P}$$

We place a bar over the letter P to mean **not P** or **P does not happen**.

Sample Exercise:

For a certain type of candy, 13% are green, 18% are red, and the rest are yellow. You draw one at random. What is the probability you draw a yellow candy?

Sample Exercise:

For a certain type of candy, 13% are green, 18% are red, and the rest are yellow. You draw one at random. What is the probability you draw a yellow candy?

$$1 - 0.13 - 0.18 = 0.69$$

The probability is 69%.

Example:

For a certain type of candy, 13% are green, 18% are red, and the rest are yellow. You draw one at random. What is the probability you do not draw a yellow candy?

Previously, we showed that yellow happened 69% of the time, so non-yellow will happen 31% of the time.

Consecutive draws

If you pretend to draw from a pile (or box or whatever) repeatedly, you need to be aware of whether this is a **selection with replacement** or a **selection without replacement**.

Selection with replacement

You have a box full of balls, identical except for color. You will draw a ball randomly, check the color, then return it to the box, mixing the contents after replacing the ball. If you have 17 red balls and 23 white balls, what is the probability of drawing 2 red balls in a row?

Verbal Cues:

If you are explicitly told you will replace the item, the exercise is a selection with replacement question. Also, if you are told you have an infinite supply of something, this is another cue that you are doing a selection with replacement question.

Selection with replacement

You have a box full of balls, identical except for color. You will draw a ball randomly, check the color, then return it to the box, mixing the contents after replacing the ball. If you have 17 red balls and 23 white balls, what is the probability of drawing 2 red balls in a row?

Answer

$$\frac{17}{40} \times \frac{17}{40} = 0.180625 \approx 18.06\%$$

Selection without replacement

You have a box full of balls, identical except for color. You will draw a ball randomly, check the color, but **do not return the ball to the box**. If you have 17 red balls and 23 white balls **initially**, what is the probability of drawing 2 red balls in a row?

Verbal Cues:

If you are not explicitly told you will replace the item, the exercise is a selection without replacement question. You may or may not be explicitly told not to return the items. However, if you are confused during an exam, it's ok to ask. (You may or may not get an answer.)

Selection without replacement

You have a box full of balls, identical except for color. You will draw a ball randomly, check the color, but **do not return the ball to the box**. If you have 17 red balls and 23 white balls **initially**, what is the probability of drawing 2 red balls in a row?

Answer

$$\frac{17}{40} \times \frac{16}{39} = 0.174358974 \approx 17.44\%$$

The probability dropped slightly, because your probability of drawing a red ball would drop slightly if you've already removed one!

With or without replacement?

When you are replacing the selected items, the denominators will remain the same for all your fractions.

When you are not replacing the items, your denominators will decrease by one with each draw. Watch out for the numerators as well, as those also change.

Selection without replacement

You have a box full of balls, identical except for color. You will draw a ball randomly, check the color, but **do not return the ball to the box**. You have 17 red balls and 23 white balls **initially**. If you draw two balls, what is the probability of getting one of each color?

Answer

$$\frac{17}{40} \cdot \frac{23}{39} + \frac{23}{40} \cdot \frac{17}{39} \approx 50.13\%$$

In this case, the two terms actually have the same value, but that's not always the case. In the first term, we presumed that we drew a red followed by a white ball. In the second term, we presumed that we drew a white followed by a red ball. Those two events are distinct, so they are disjoint and their probabilites can be added.

Selection without replacement

You have a box full of balls, identical except for color. You will draw a ball randomly, check the color, but **do not return the ball to the box**. You have 17 red balls and 23 white balls **initially**. If you draw three balls, what is the probability of getting exactly one red?

Answer

$$\frac{17}{40}\cdot\frac{23}{39}\cdot\frac{22}{38}+\frac{23}{40}\cdot\frac{17}{39}\cdot\frac{22}{38}+\frac{23}{40}\cdot\frac{22}{39}\cdot\frac{17}{38}\approx 43.53\%$$

In the first term, we presumed that we drew a red followed by white balls. In the second term, we presumed that we drew a red ball second. Finally, we presume we drew the red ball last. Those three events are distinct, so they are disjoint and their probabilites can be added.

Selection with replacement

You have a box full of balls, identical except for color. You will draw a ball randomly, check the color, and return the ball to the box. You have 17 red balls and 23 white balls. If you draw three balls, what is the probability of getting exactly one red?

Answer

$$\frac{17}{40}\cdot\frac{23}{40}\cdot\frac{23}{40}+\frac{23}{40}\cdot\frac{17}{40}\cdot\frac{23}{40}+\frac{23}{40}\cdot\frac{23}{40}\cdot\frac{17}{40}\approx 42.15\%$$

In the first term, we presumed that we drew a red followed by white balls. In the second term, we presumed that we drew a red ball second. Finally, we presume we drew the red ball last. Those three events are distinct, so they are disjoint and their probabilites can be added.

Simplification

Notice in the above exercise, that the three terms are equal. The fractions are simply in different orders. So, it makes sense to simplify the process and calculate:

Answer

$$3 \cdot \frac{17}{40} \cdot \frac{23}{40} \cdot \frac{23}{40} \approx 42.15\%$$

To make these simplifications, we have to understand that one Red and two White balls can be ordered three ways:

$$RWW, WRW, WWR$$

In how many ways can the letters: A, B, C be placed in a row?

- ABC
- ACB
- BAC
- BCA
- CAB
- CBA

Note: $3 \cdot 2 \cdot 1 = 6$.

Factorials

In how many ways can four children form a queue (line)?

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24.$$

$$n! = (n)(n-1)(n-2)\cdots(4)(3)(2)(1)$$

Club Officer Questions

If you have 10 people and wish to elect a President, Vice-President, and Secretary, in how many ways is this possible?

$$10 \cdot 9 \cdot 8 = 720$$

Club Officer Questions

These questions are *permutations of n things taken k at a time*.

$$P(n,k) = (n)(n-1)(n-2)\cdots(n-(k-1))$$

Don't let the $n-(k-1)$ fool you. Just count out k factors.

Committee Questions

If you have 10 people and wish to form a committee of 3 persons, in how many ways is this possible?

You first solve the Club Officer Problem, then you divide by the overcount factor. In this case, you will overcount by 3! or 6, because there are 6 ways to arrange 3 people, as we have shown!

$$\frac{720}{6} = 120$$

Committee Questions

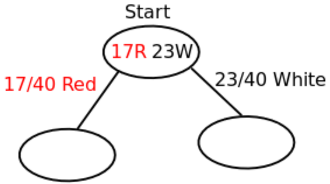
Committee questions are *combinations of n things taken k at a time*.

$$C(n,k) = \frac{P(n,k)}{k!}$$

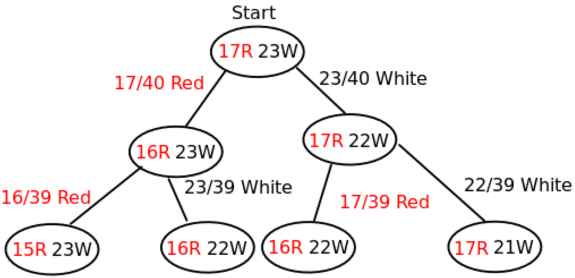
Tree Diagrams

Tree diagrams are often used to sort out slighty more complex probability questions.

Let's revisit the same balls and boxes problem using a tree diagram.

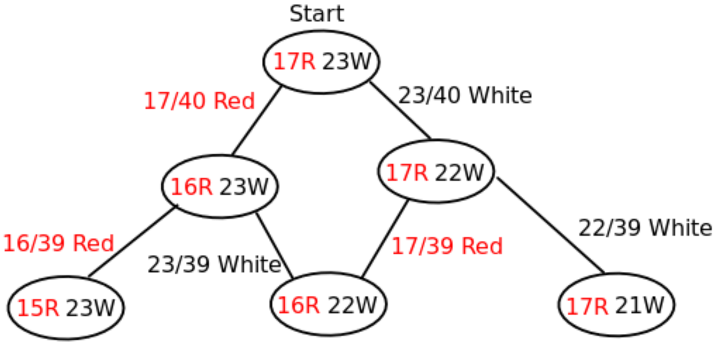


We start the top with the original 40 balls, colored as given. We have branches which indicate the probabilities for each transition.

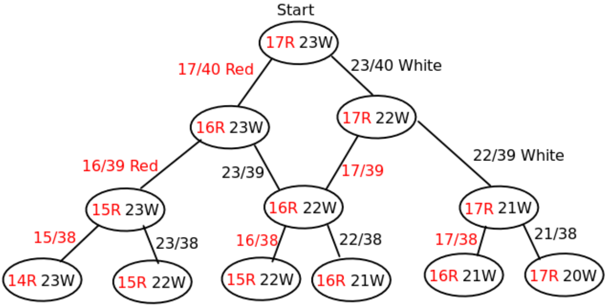


Each oval represents some combination of remaining red and white balls. The branches are labeled with probabilities. In this problem, we can arrive at the same combination from different paths, so we choose to combine the ovals.

Balls and Boxes

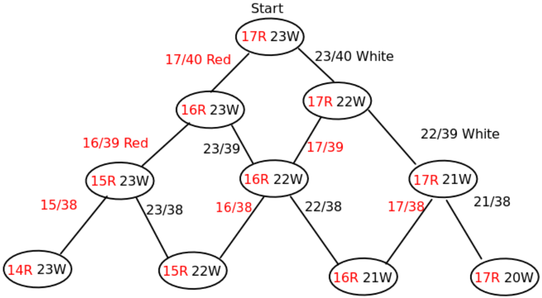


Balls and Boxes

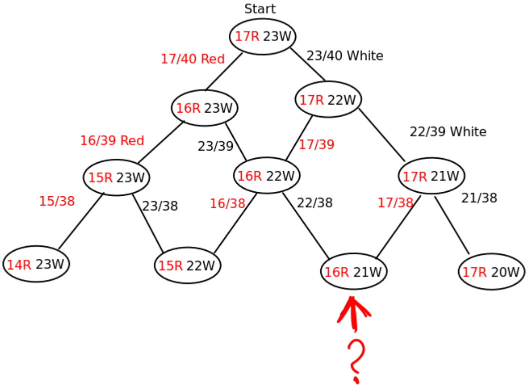


We have now filled out the tree sufficiently to answer our previous question.

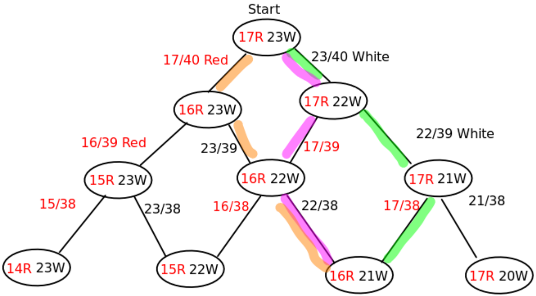
Balls and Boxes



But we choose to combine the combinations.

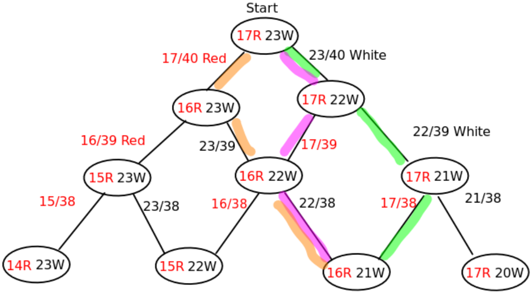


This oval marked with ? is our goal.
What is the probability of choosing 3 red balls?



These are our paths.

Balls and Boxes



$$\frac{17}{40} \times \frac{23}{39} \times \frac{22}{38} + \frac{23}{40} \times \frac{17}{39} \times \frac{22}{38} + \frac{23}{40} \times \frac{22}{39} \times \frac{17}{38}$$

This is the same solution we found before.

A recurring theme

In mathematics, computer science, and many other areas of science, it's common to have these algebraic expressions which are composed of *terms* where everything is multiplied together (or **and**ed together) which are then connected by addition (or **or**ed together).

$$(TraitA\ and\ TraitB)\ \ or(TraitC\ and\ TraitD)$$

Example: Let's say an *outfit* can be made up of jeans and a t-shirt or a dress shirt and slacks. I can find out how many outfits I can make if I know how many of each thing I have. We multiply the quantities for jeans and t-shirts and add that to the other product made by multiplying the number of shirts to the number of slacks.

Here it's the same

In this situation, we are breaking down all our success states into non-overlapping scenarios which we can then combine categories for. So those *terms* themselves are disjoint and can be added. However, within each scenario, there are multiple constraints which need to all be satisifed. If their probabilities are independent, we can find their joint probabilities with multiplication.

Sensitivity and Specificity

Sensitivity

The sensitivity of a medical test for a disease is the probability it will detect a disease when the disease is present.

Specificity

The specificity of a medical test for a disease is the probability it will give a correct result when the disease is **not** present.

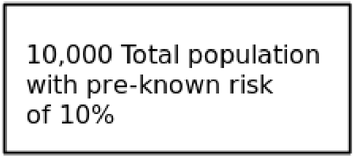
If you have a population with a pre-known risk, you can then predict (at a population level) the probability of a false positive or a false negative! It's risky to presume the group probabilities apply to individuals, but this information is important.

Example

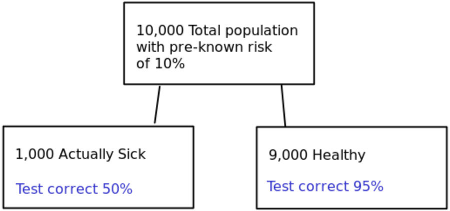
Let's presume there is a medical test for a disease which has a sensitivity of 50% and a specificity of 95%. If a population has a pre-known risk of 10%, what percent of those persons who test as sick are truly sick?

Sensitivity and Specificity

To make this easier to visualize, it is convenient to presume a large integer as the original population, but mathematically, this could be any number at all, even 1.

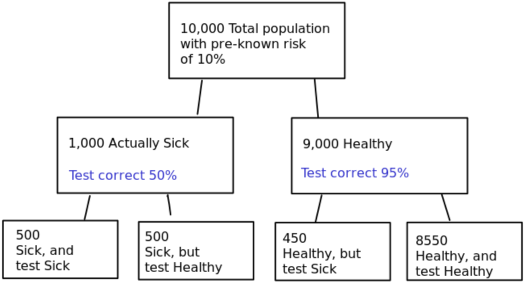


Sensitivity vs Specificity



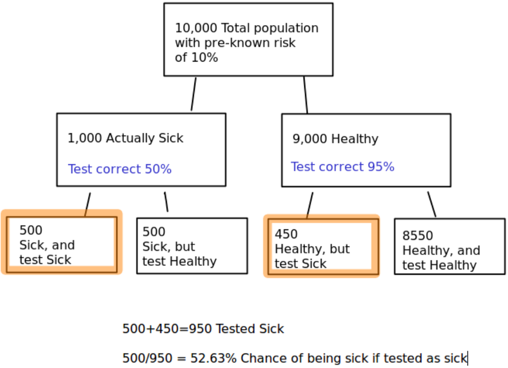
Of those who are sick, 50% will get the correct test results.
This is what sensitivity tells us.
Health persons will get the correct result 95% of the time.
This is what specificity tells us.

Sensitivity vs Specificity



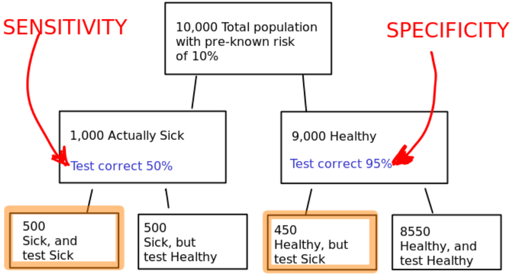
Notice that we cannot combine outcomes for this situation!
All the outcomes are different!

Sensitivity vs Specificity

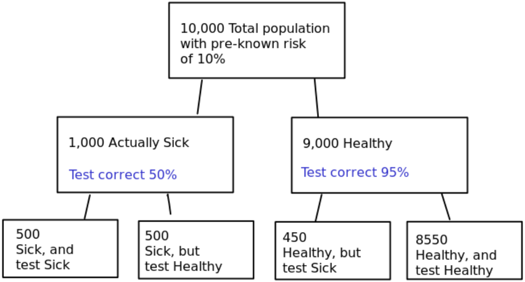


A total of 950 tested sick,
but only 500 were actually sick.

Sensitivity vs Specificity



False Negatives, False Positives



The 500 who were sick but tested healthy were false negatives.
The 450 who were healthy but tested sick were false positives.

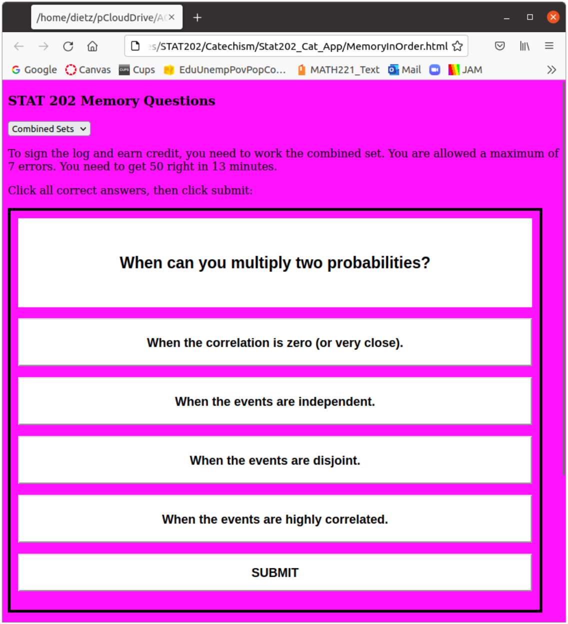
Power of a Statistical Test

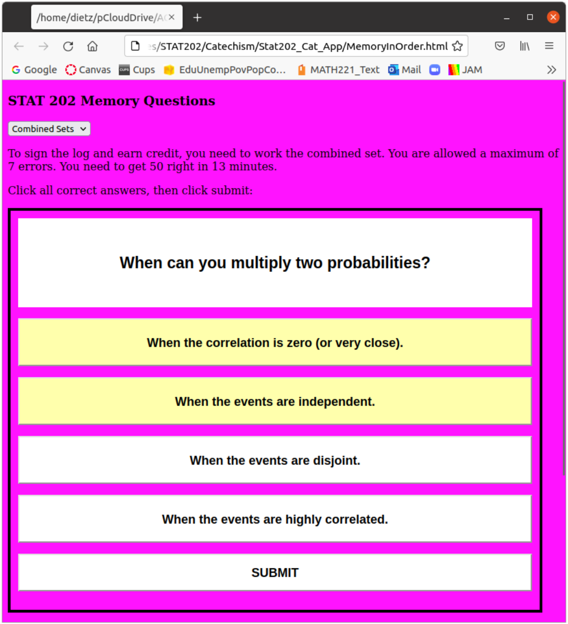
The power of a statistical test is the same as sensitivity. It's the probability of getting a correct result when what you are claiming is true is in fact true.

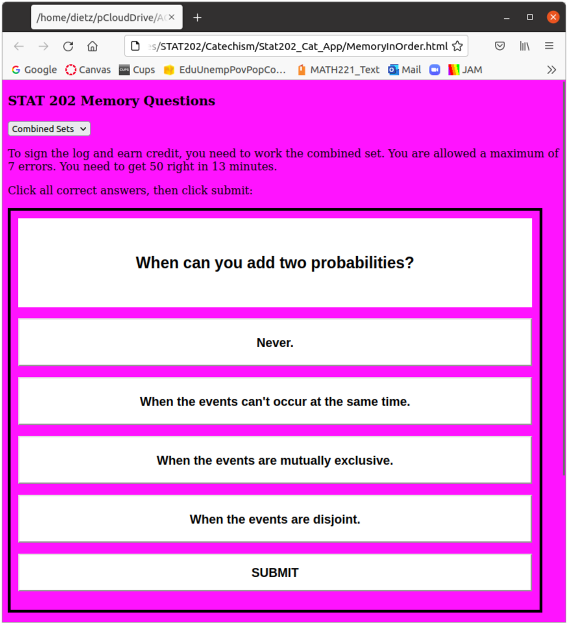
Sadly, we have the same issue as with sensitivity. If we don't know how likely the thing is to be true in the first place, we can't even assess the probability of a false accusation!

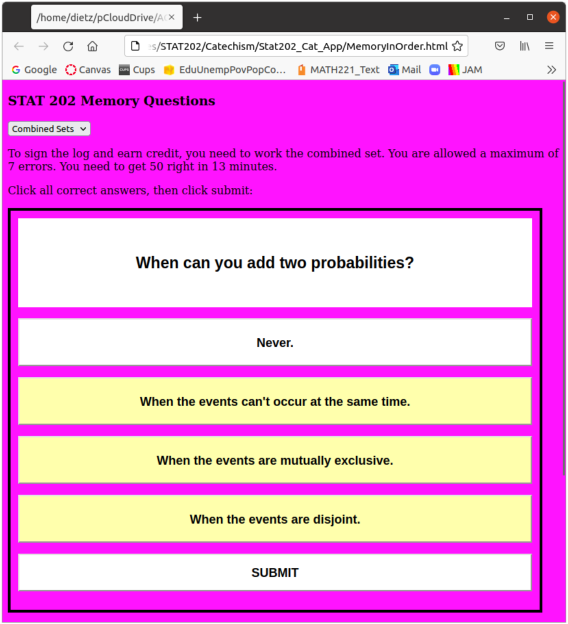
MEMORY QUESTIONS

5 today!


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
When can you multiply two probabilities?
When the correlation is zero (or very close).
When the events are independent.
When the events are disjoint.
When the events are highly correlated.
SUBMIT


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
When can you multiply two probabilities?
When the correlation is zero (or very close).
When the events are independent.
When the events are disjoint.
When the events are highly correlated.
SUBMIT


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
When can you add two probabilities?
Never.
When the events can't occur at the same time.
When the events are mutually exclusive.
When the events are disjoint.
SUBMIT


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
When can you add two probabilities?
Never.
When the events can't occur at the same time.
When the events are mutually exclusive.
When the events are disjoint.
SUBMIT

/home/dietz/pCloudDrive/A

s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html

Google Canvas Cups EduUnempPovPopCo... MATH221_Text Mail JAM

STAT 202 Memory Questions

Combined Sets

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

In combinatorics, would 'select 2 committee members from a group' be with or without replacement?

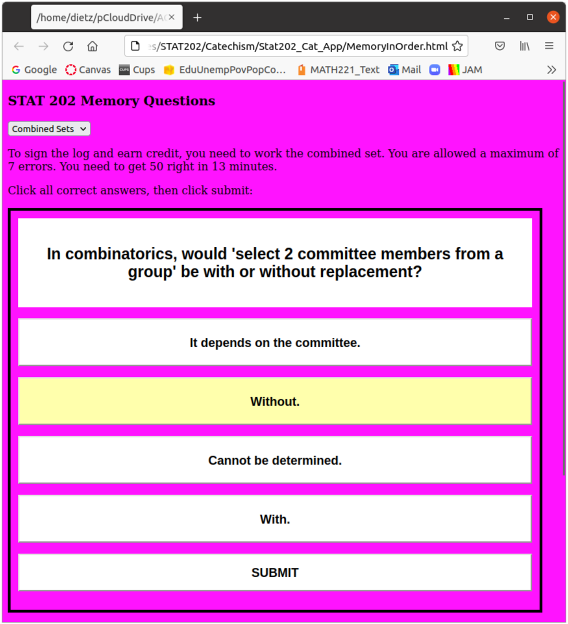
It depends on the committee.

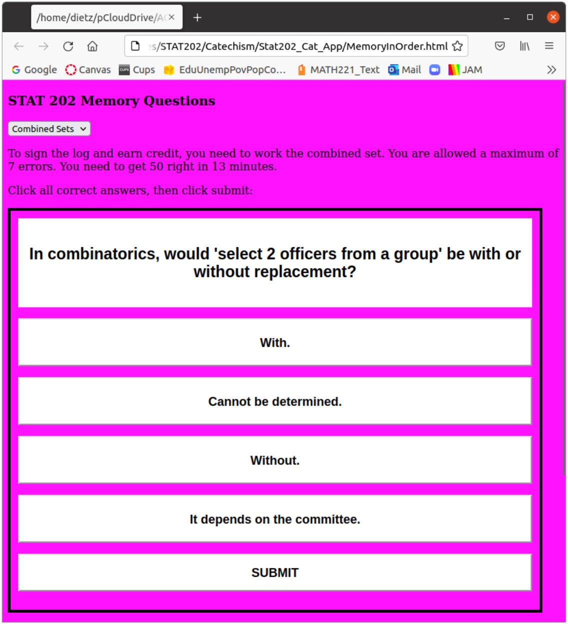
Without.

Cannot be determined.

With.

SUBMIT


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
In combinatorics, would 'select 2 committee members from a group' be with or without replacement?
It depends on the committee.
Without.
Cannot be determined.
With.
SUBMIT


/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
In combinatorics, would 'select 2 officers from a group' be with or without replacement?
With.
Cannot be determined.
Without.
It depends on the committee.
SUBMIT

STAT 202 Memory Questions

Combined Sets

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

In combinatorics, would 'select 2 officers from a group' be with or without replacement?

With.

Cannot be determined.

Without.

It depends on the committee.

SUBMIT

/home/dietz/pCloudDrive/A

s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html

Google Canvas Cups EduUnempPovPopCo... MATH221_Text Mail JAM

STAT 202 Memory Questions

Combined Sets

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

From a combinatorics perspective, how do the 'officer' selection and 'committee' selection differ?

In the 'officer' case, the order in which they're selected matters.

In the committee case, selection order matters.

In the 'officer' case, selection order does not matter.

In the committee case, the order in which they are selected doesn't matter.

SUBMIT

STAT 202 Memory Questions

Combined Sets

To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.

Click all correct answers, then click submit:

From a combinatorics perspective, how do the 'officer' selection and 'committee' selection differ?

In the 'officer' case, the order in which they're selected matters.

In the committee case, selection order matters.

In the 'officer' case, selection order does not matter.

In the committee case, the order in which they are selected doesn't matter.

SUBMIT