Area Under the Normal Curve and Sampling Design and Experiments

Donna Dietz

American University

dietz@american.edu

STAT 202 - Spring 2020

Area Under the Normal Curve

The formulas for converting back and forth between raw data (from any Normal Distribution, $N(\mu, \sigma)$) and z-scores ($N(0, 1)$) are:

$$z = \frac{x - \mu}{\sigma}$$

$$x = \mu + z\sigma$$

Let's practice converting between these and finding μ and σ.

Fill in the missing values

$$z = \frac{x - \mu}{\sigma} \qquad x = \mu + z\sigma$$

raw data	mean of data	std dev	z-score
44	35	17	
	73	13	-1.7
12	10		0.2
33		13	1.2

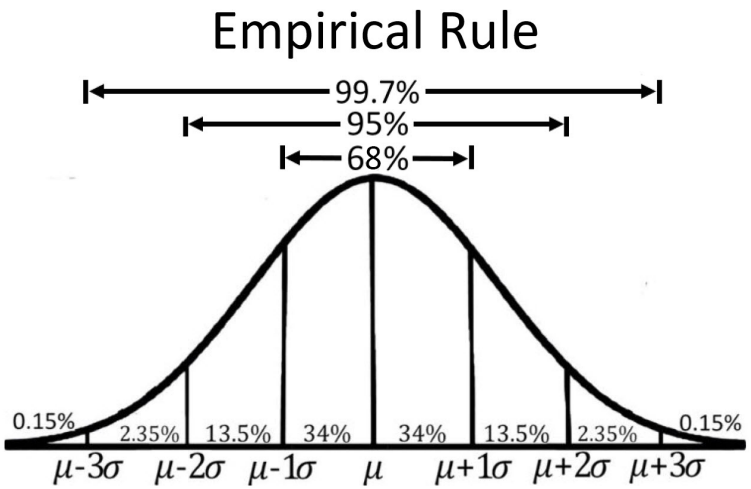
Fill in the missing values

$$z = \frac{x - \mu}{\sigma} \qquad x = \mu + z\sigma$$

raw data	mean of data	std dev	z-score
44	35	17	**0.5294**
50.9	73	13	-1.7
12	10	**10**	0.2
33	**17.4**	13	1.2

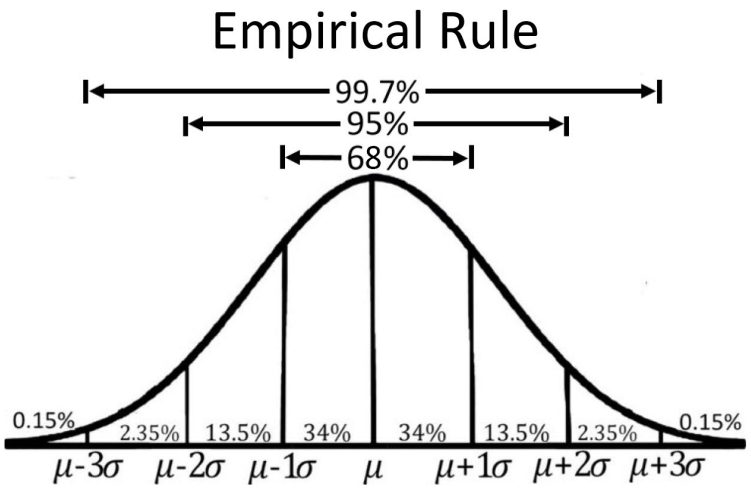
Using z-scores to estimate area

We often want to estimate what percent of our population falls above, below, or between two cutoff values. We can use Normal Tables and software to to this, but there are a few values that get used so frequently, we should commit them to memory. This graphic is from AndyMath.com.



The Empirical Rule

We use the Empirical Rule, otherwise known as the "68-95-99.7 Rule" to make quick calculations involving the most common cut-offs: z = -3, -2, -1, 0, 1, 2, and 3.

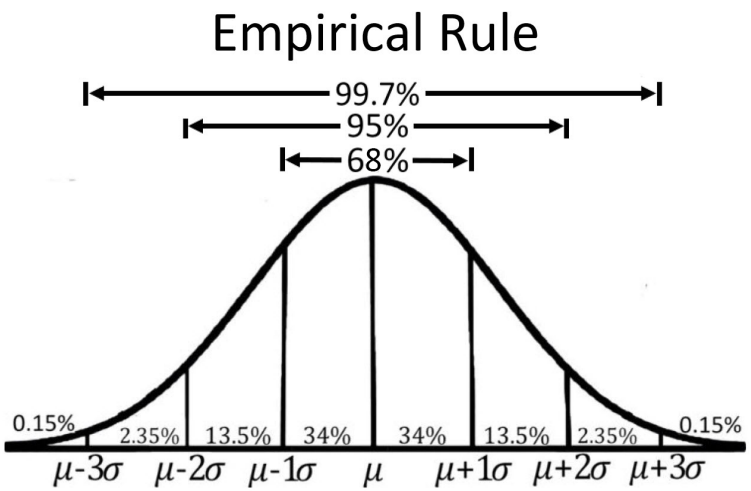


In this diagram, $z = -2$ is labeled with the more generic label of $\mu - 2\sigma$. Normalized z-scores are just another Normal Distribution with $\mu = 0$ and $\sigma = 1$ to make everything consistent.

Practice Question

What percent of a Normal Distribution falls between $z = 1$ and $z = 3$?

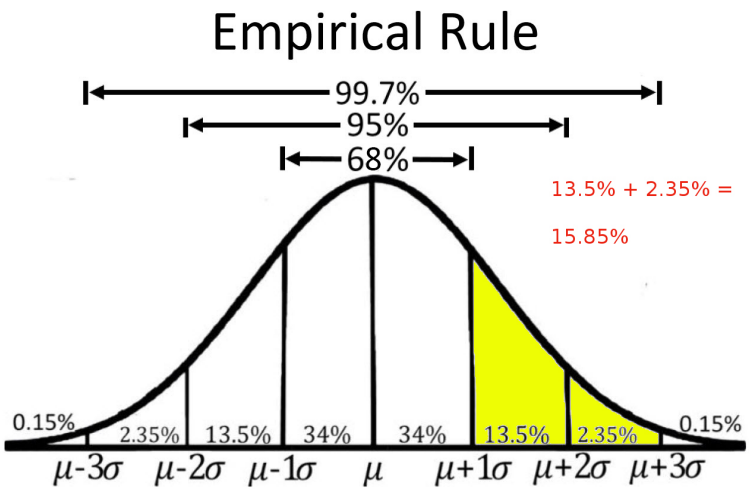
Use this diagram to help you figure it out!



Practice Question

What percent of a Normal Distribution falls between $z = 1$ and $z = 3$?

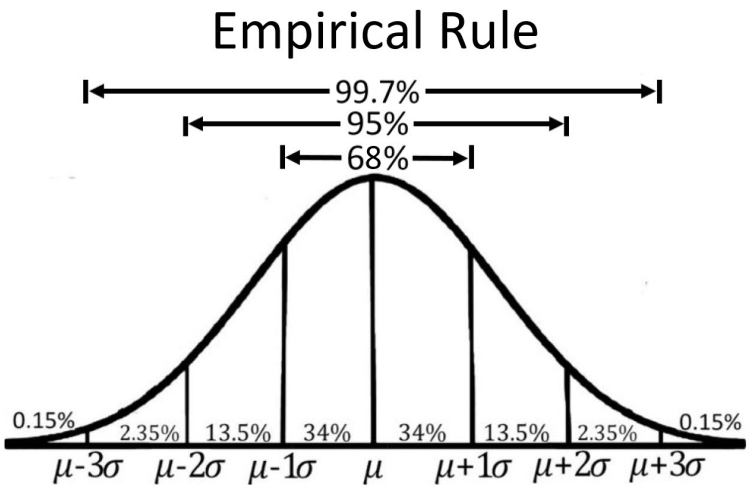
Use this diagram to help you figure it out!



$$13.5\% + 2.35\% = 15.85\%$$

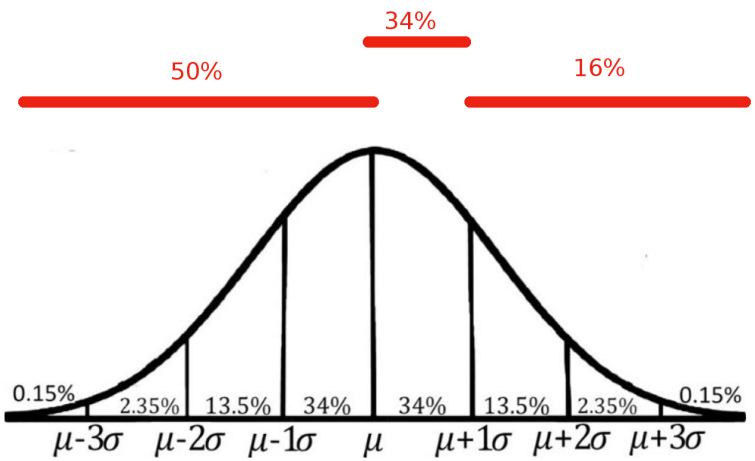
Practice Question 2

An entrance exam is given to a private school. The distribution is normal with $N(86, 10)$. What is the cutoff score for the top 16% of test-takers?



Practice Question 2

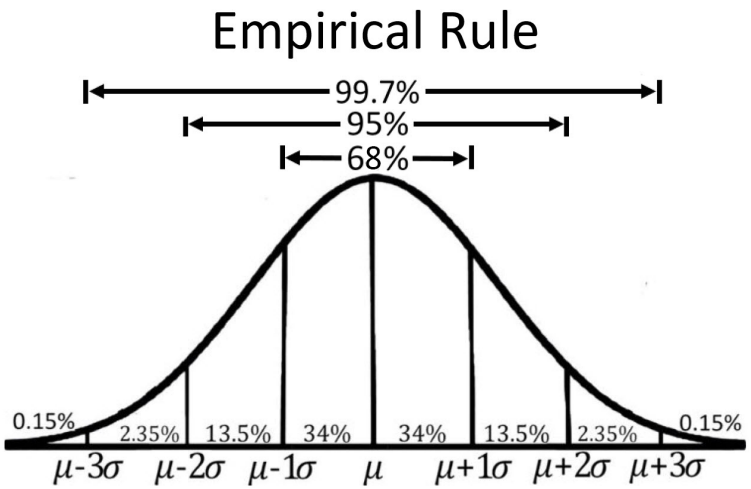
An entrance exam is given by a private school. The distribution of scores is normal with $N(86, 10)$. What is the cutoff score for the top 16% of test-takers?



The cutoff should be at $z = 1$ or $\mu + \sigma$, or $86 + 10 = 96$.

Practice Question 3

You have 700 students taking STAT 202 in various sections of the course. You estimate the amount of time students spend studying is about $N(2, 0.4)$ hours per week for the course. Roughly how many students do you estimate will spend more than 3.2 hours per week studying?



Practice Question 3

The students above $z = 3$ are the ones we are considering, because

$$z = \frac{x - \mu}{\sigma} = \frac{1.2}{.4} = 3$$

This is $0.15\% = 0.0015$, or $0.0015 \times 700 = 1.05$, or about one student.

Note: I do not believe this model is accurate!

Revisiting addition of distributions

We can revisit our old friend, the addition of distributions for independent events!

Let's take a hypothetical trip to Pizza Hut. We will buy a Coke with $N(130, 17)$ Calories, and a Pan Pizza with $N(625, 22)$ Calories.

- Find the expected mean, variance, and standard deviation
- Use the empirical rule to find the 95% confidence interval for expected Calories for the total

Pizza Hut Question

The expected mean: $130+625 = 755$ Calories

The variance of the sum: $17^2 + 22^2 = 773$

The square root of the variance is the standard deviation: 27.803

The 95% confidence interval is found with $\mu \pm 2\sigma$.

Low End: $755 - 2 \times 27.8 = 699.4$, High End: $755 + 2 \times 27.8 = 810.6$

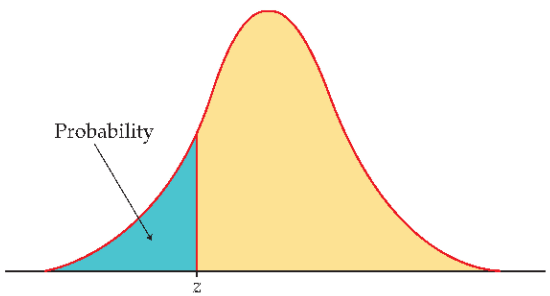
What if the Empirical Rule isn't enough?

If the Empirical Rule doesn't give you the information you need, you have to rely on either Normal Tables, or computers. Luckily, those are both easy fixes!

Use a Normal Table to determine the answers to these questions:

- What percent falls below $z = -1.3$?
- What percent falls above $z = 1.3$?
- What percent falls above $z = -1.3$?
- What percent falls below $z = 1.3$?

Using the "Tail" Chart



−1.9	.0287	.0281	.0274
−1.8	.0359	.0351	.0344
−1.7	.0446	.0436	.0427
−1.6	.0548	.0537	.0526
−1.5	.0668	.0655	.0643
−1.4	.0808	.0793	.0778
−1.3	.0968	.0951	.0934
−1.2	.1151	.1131	.1112
−1.1	.1357	.1335	.1314
−1.0	.1587	.1562	.1539
−0.9	.1841	.1814	.1788

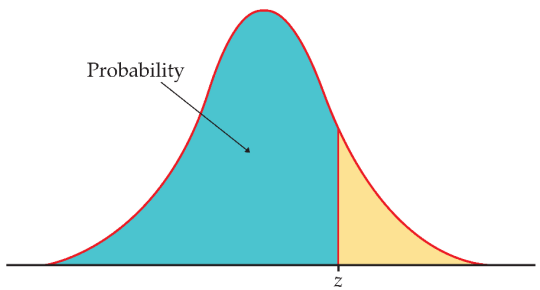
The "Tail" chart tells you how much area is in the tail which is cut off by the $z - score$ listed. If your $z - score$ only contains one digit after the decimal, the $z - score$ will be in the first column of the table.

Using the "Tail" Chart

- What percent falls below $z = -1.3$? ANSWER: 0.0968 or 9.68%.
- What percent falls above $z = 1.3$? ANSWER: 0.0968 or 9.68%.
- What percent falls above $z = -1.3$?
- What percent falls below $z = 1.3$?

For the other two questions, you can either use the "Hump" chart or just subtract from one. So, the other answers will be $1 - .0968 = .9032$

Using the "Hump" Chart



1.0	.8413	.8438	.8461
1.1	.8643	.8665	.8686
1.2	.8849	.8869	.8888
1.3	.9032	.9049	.9066
1.4	.9192	.9207	.9222
1.5	.9332	.9345	.9357
1.6	.9452	.9463	.9474
1.7	.9554	.9564	.9573
1.8	.9641	.9649	.9656
1.9	.9713	.9719	.9726
2.0	.9772	.9778	.9783
2.1	.9821	.9826	.9830

The "Hump" chart tells you how much area is in the larger part (hump plus a tail) which is cut off by the $z-score$ listed. If your $z-score$ only contains one digit after the decimal, the $z-score$ will be in the first column of the table.

Using the "Hump" Chart

- What percent falls below $z = -1.3$?
- What percent falls above $z = 1.3$?
- What percent falls above $z = -1.3$? ANSWER: 0.9032 or 90.32%.
  item What percent falls below $z = 1.3$? ANSWER: 0.9032 or 90.32%.

For the other two questions, you can either use the "Tail" chart or just subtract from one. So, the other answers will be $1 - .9032 = .0968$

Using the Normal Tool

There are many Statistics Calculators online. They don't all work as you might expect, so if you find a new one, please check it against the Empirical Rule to make sure you understand how it works!

We will use StatCrunch:

$$\textit{Stat} > \textit{Calculators} > \textit{Normal}$$

.

Use StatCrunch to verify the example was just did with $z = -1.3$ and $z = 1.3$. To use z-scores, set $N(\mu = 0, \sigma = 1)$ in your Normal Calculator.

Sampling Design and Experiments

This section is about concepts, best practices, and vocabulary. Exam questions on this section would be either short-answer or matching type questions, not calculations.

Observational Studies vs. Experiments

Observational Study

An observational study is done in a pre-existing situation, and the researcher does not interfere. They only observe. Examples include watching how many people enter a grocery store at various times of day, or after the announcement of an upcoming storm.

Experiment

An experiment involves a research manipulation of groups receiving treatments. Best practices indicate that the groups should usually be determined randomly.

Ways to be random

SRS

Because we can never really tell if there are important hidden traits we are unaware of, we consider SRS (simple random sample) to be the gold standard for getting a representative sample of a population.

Stratified random sample

It sometimes makes sense to take a *stratifed* random sample. For example, we may wish to interview an equal number of First-Year, Sophomore, Junior, and Senior students. However, even with stratification, the selection within each stratum should be random.

Control group

Control Group

If you are attempting to determine cause and effect, it's usually wise to have a control group which either receives no treatment, or receives the usual expected treatment. For example, you may need to treat certain medical conditions, so the control group may receive the standard recommended treatment while the other group/s receive supplemental or different treatments.

Placebos

If a control group is used but you don't want subjects should not know they are in the control group, they are often given **placebos** or sugar pills. This is to make it possible to determine the effect of the actual treatments as opposed to the effect of belief in a treatment.

Placebo effect

Placebo effect

The placebo effect is the surprising effect that people tend to report side effects or often even improvement in their illness when they are simply given inert treatments. When drug tests are performed, ideally you have two groups, one given a treatment, and one given the placebo, so you can compare what the two groups report. Effect reported by the placebo group are attributed to the so-called "placebo effect".

Double-Blind Studies

Double-Blind Study

If neither the subjects nor the interviewers know which subjects are in which treatment groups, this is called a double-blind study. After the study is done and results are collected, the subjects can be informed of which group they were in.

Survey or poll?

Polls

A poll usually contains just one quick question, often multiple choice, which can be given quickly as people pass by. Those happy/sad face buttons you see in many stores can be considered as a type of poll.

Survey or poll?

Surveys

A survey typically involves multiple questions, some may be free-response, and a respondent is often rewarded for their efforts (but not always). You might be shopping one day and someone asks you if you have an hour to spend in exchange for a store gift certificate, for example.

Ethical issues...

It's very upsetting but important to talk about times where ethics were not followed.

https://www.cdc.gov/tuskegee/timeline.htm

In 1932, the Public Health Service, working with the Tuskegee Institute, began a study to record the natural history of syphilis in hopes of justifying treatment programs for blacks. It was called the "Tuskegee Study of Untreated Syphilis in the Negro Male."

ok... but...

The study initially involved 600 black men ... The study was conducted without the benefit of patients' informed consent. Researchers told the men they were being treated for "bad blood," a local term used to describe several ailment... In truth, they did not receive the proper treatment needed to cure their illness... Although originally projected to last 6 months, the study actually went on for 40 years.

then...

In July 1972, an Associated Press story about the Tuskegee Study caused a public outcry that led the Assistant Secretary for Health and Scientific Affairs to appoint an Ad Hoc Advisory Panel to review the study.

Informed Consent, anyone?

The panel found that the men had agreed freely to be examined and treated. However, there was no evidence that researchers had informed them of the study or its real purpose. In fact, the men had been misled and had not been given all the facts required to provide informed consent.

Not good!

The men were never given adequate treatment for their disease. Even when penicillin became the drug of choice for syphilis in 1947, researchers did not offer it to the subjects.

Reparations

In the summer of 1973, a class-action lawsuit was filed on behalf of the study participants and their families. In 1974, a $10 million out-of-court settlement was reached.

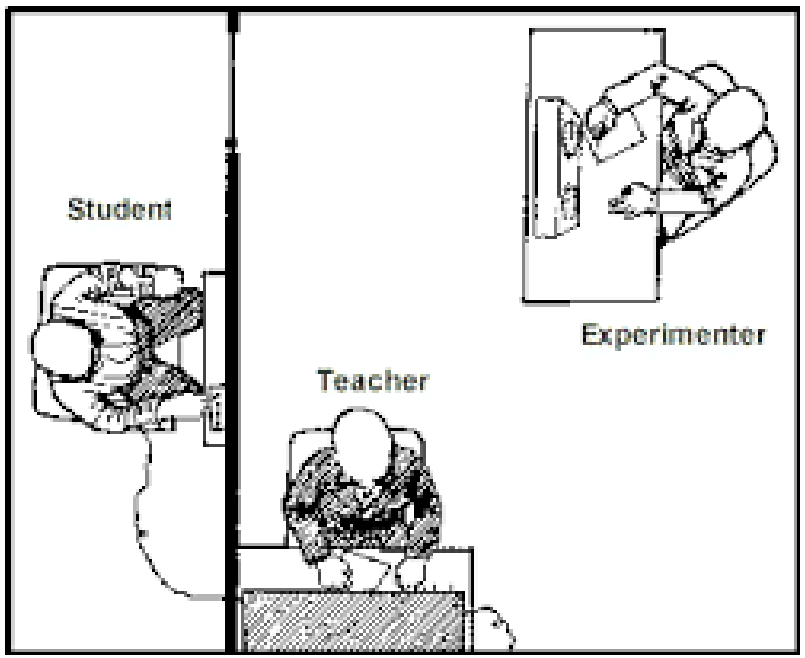
As of 2009...

The last widow receiving THBP benefits died in January 2009. There are 12 offspring currently receiving medical and health benefits.

Milgram shock experiment

https://www.simplypsychology.org/milgram.html
Milgram (1963) examined justifications for acts of genocide offered by those accused at the World War II, Nuremberg War Criminal trials. Their defense often was based on "obedience" - that they were just following orders from their superiours.

Milgram shock experiment

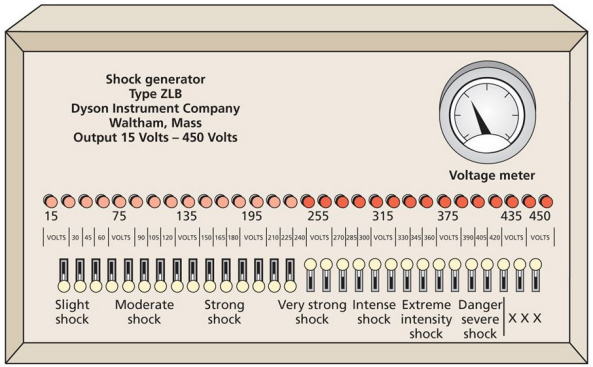


But again, the true purpose of the research was hidden from those participating...

Milgram shock experiment

So the “Teachers” were encouraged to cause physical pain and harm to the “Students”, even though in reality, they were part of the reserach team, and no physical harm was being done to them.

Milgram shock experiment



There were four prods and if one was not obeyed, then the experimenter (Mr. Williams) read out the next prod, and so on.

> **Prod 1**: Please continue.
>
> **Prod 2:** The experiment requires you to continue.
>
> **Prod 3**: It is absolutely essential that you continue.
>
> **Prod 4**: You have no other choice but to continue.

Results:

65% (two-thirds) of participants (i.e., teachers) continued to the highest level of 450 volts. All the participants continued to 300 volts.

Milgram shock experiment

Milgram argued that he had to trick his participants, due to the nature of the study. He also followed up on his subjects to reassure them. Does this make it all ok? What do you think?

Further reading...

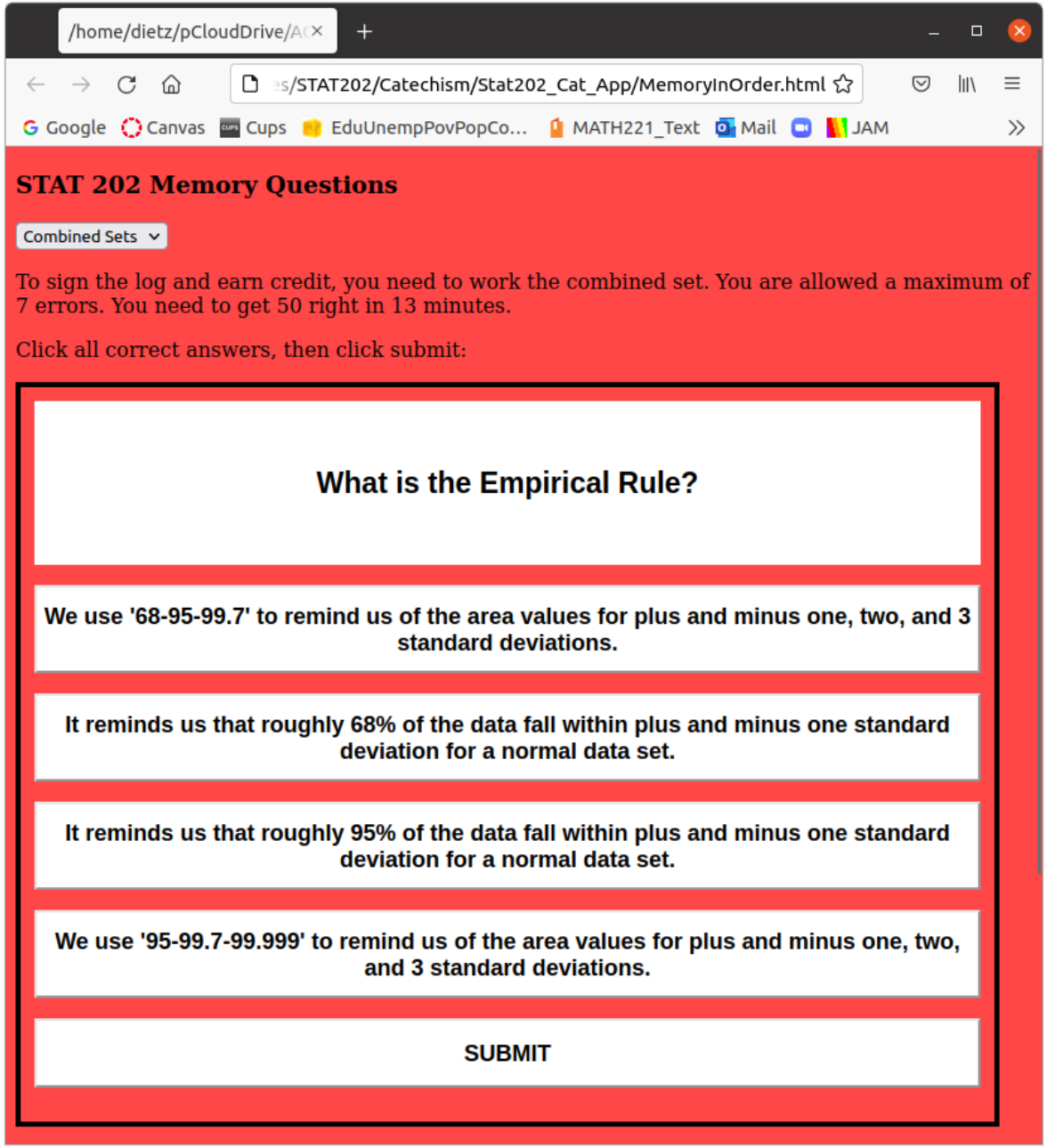
https://behavioralscientist.org/
how-would-people-behave-in-milgrams-experiment-today/

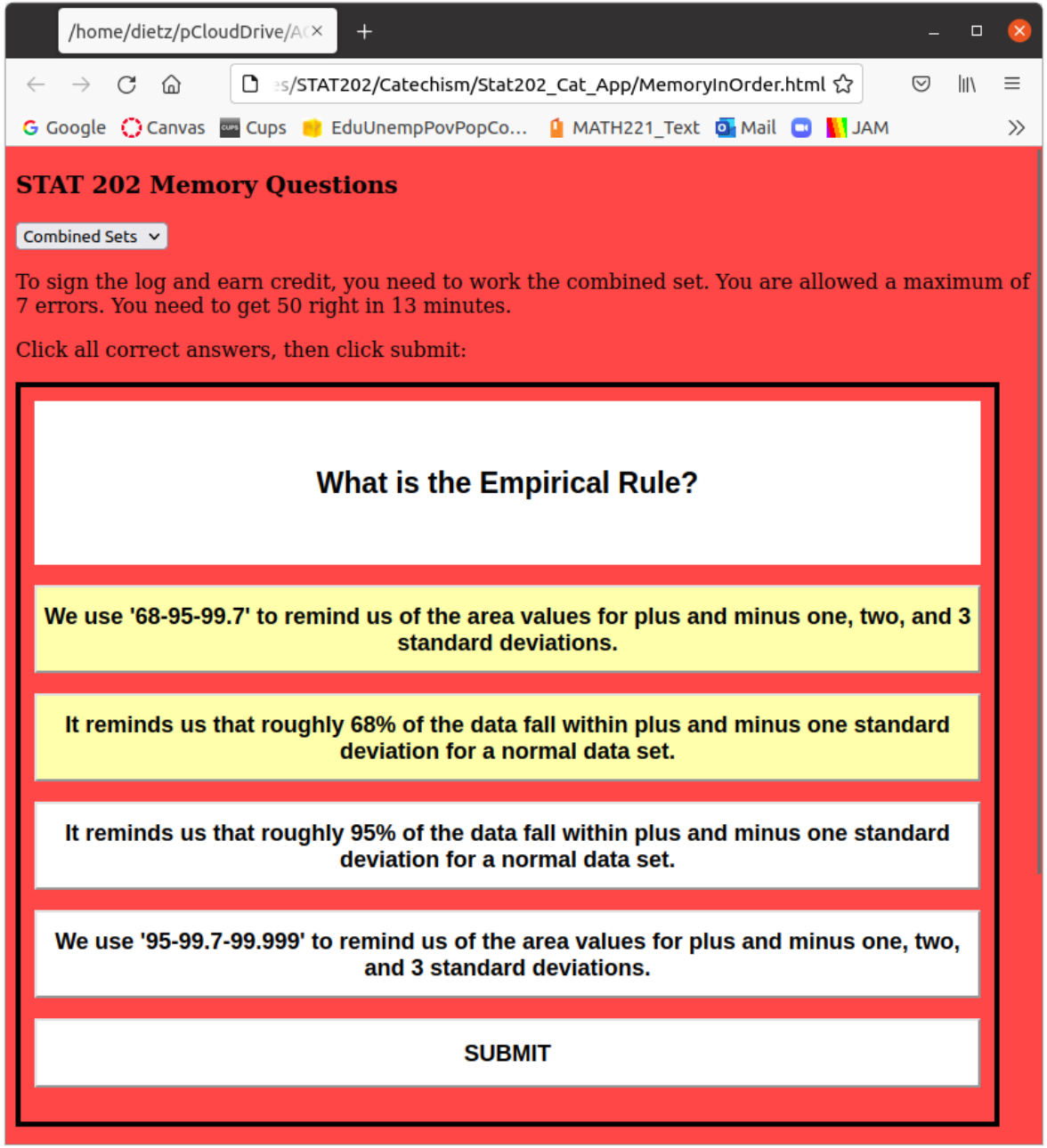
Milgram's experiment would be illegal today, but the story will continue.
For more reading, the Internet is at your service.

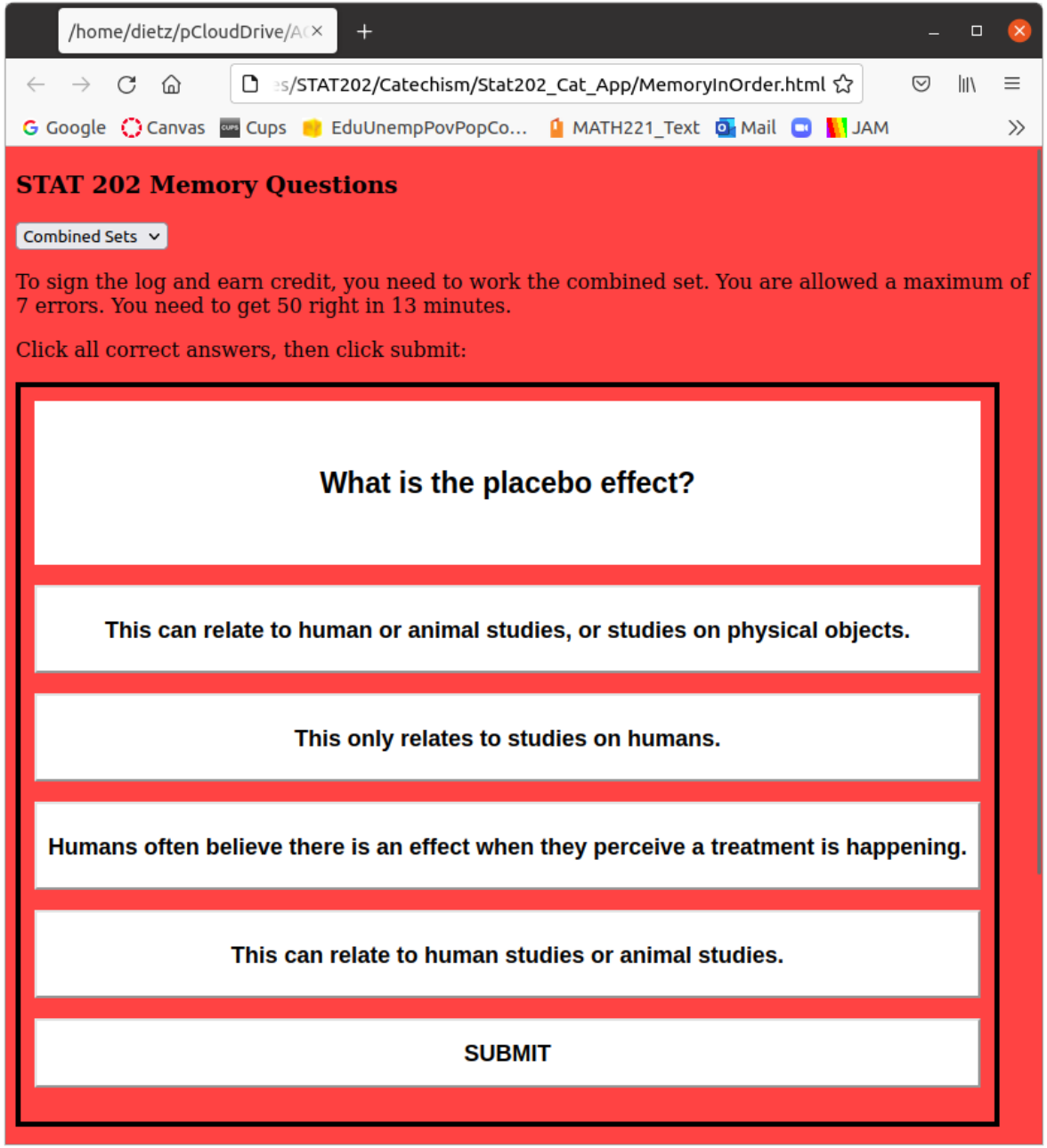
MEMORY QUESTIONS

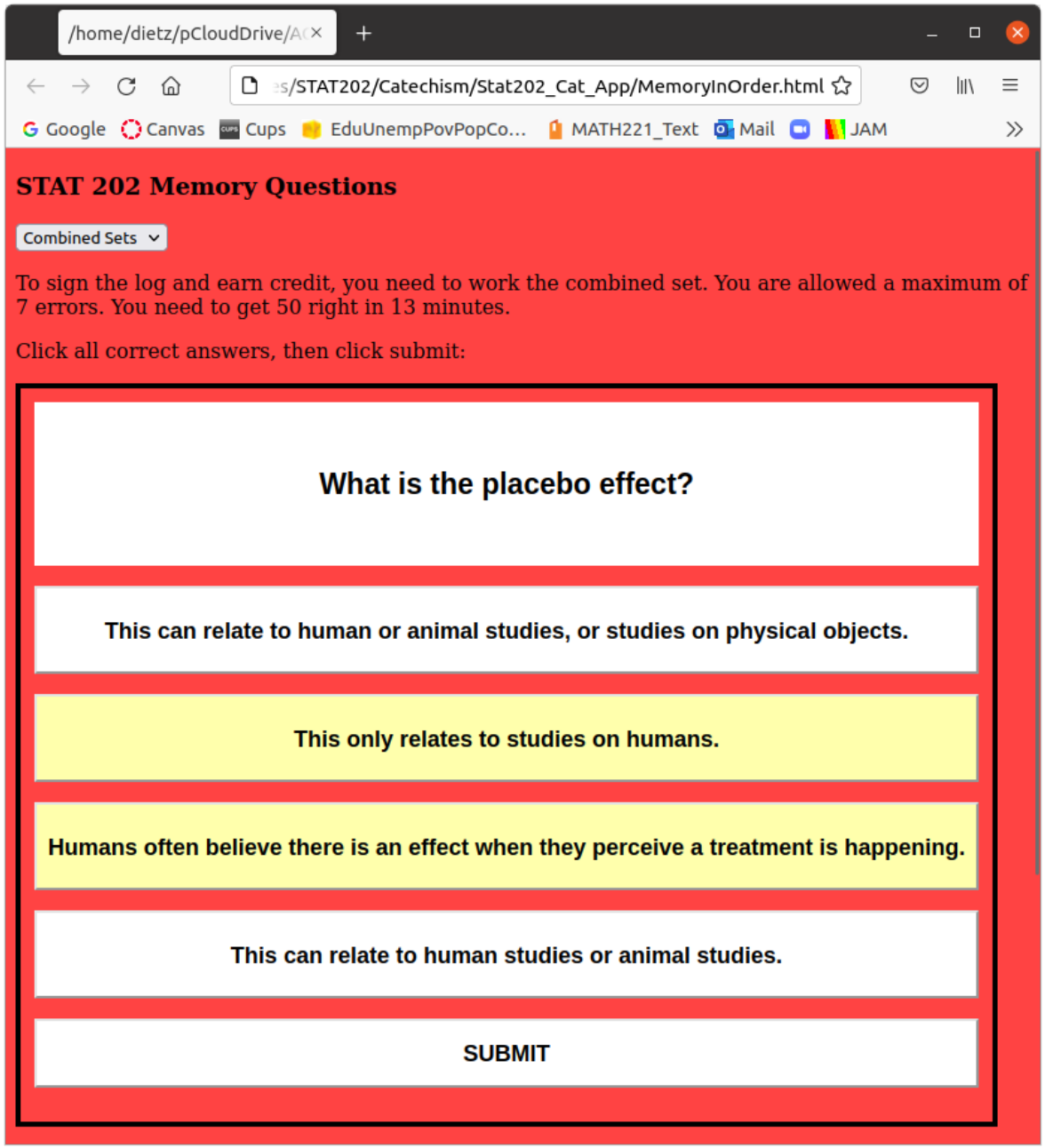
Six today!

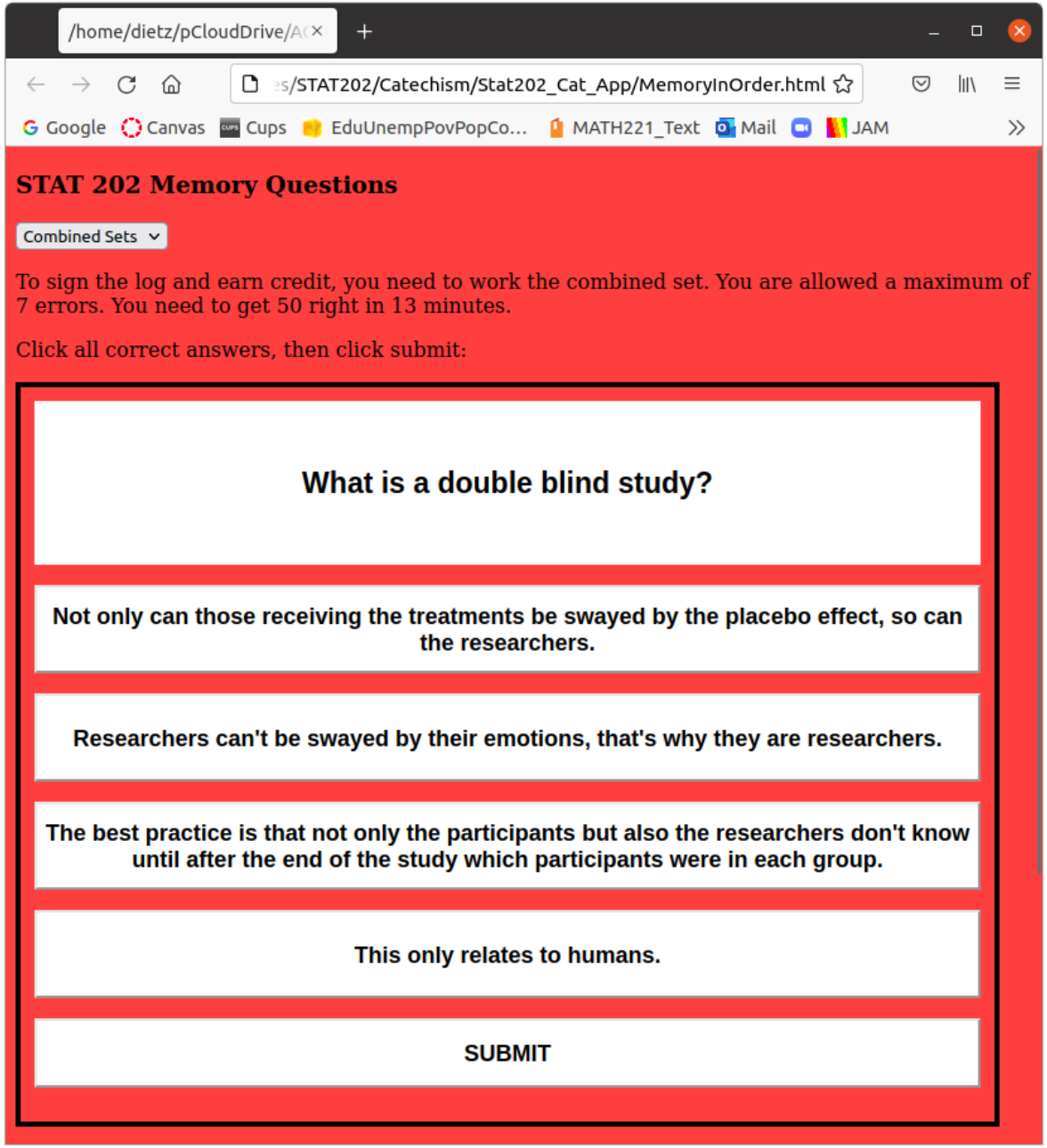
This finishes all the memory questions for Exam 1 study, and it puts us more than halfway through the entire set of memory questions.

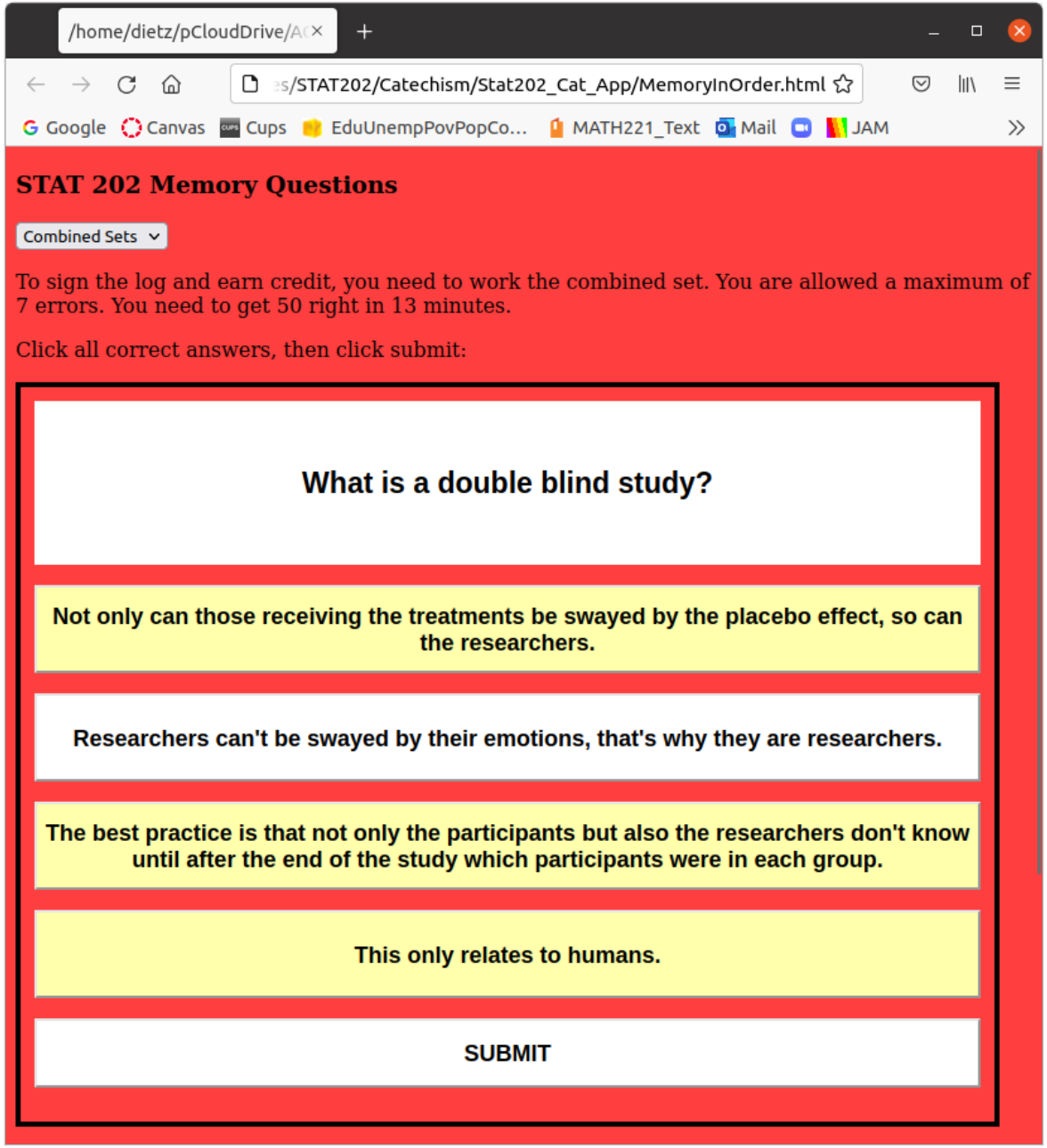

/home/dietz/pCloudDrive/A
/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
What is the Empirical Rule?
We use '68-95-99.7' to remind us of the area values for plus and minus one, two, and 3 standard deviations.
It reminds us that roughly 68% of the data fall within plus and minus one standard deviation for a normal data set.
It reminds us that roughly 95% of the data fall within plus and minus one standard deviation for a normal data set.
We use '95-99.7-99.999' to remind us of the area values for plus and minus one, two, and 3 standard deviations.
SUBMIT

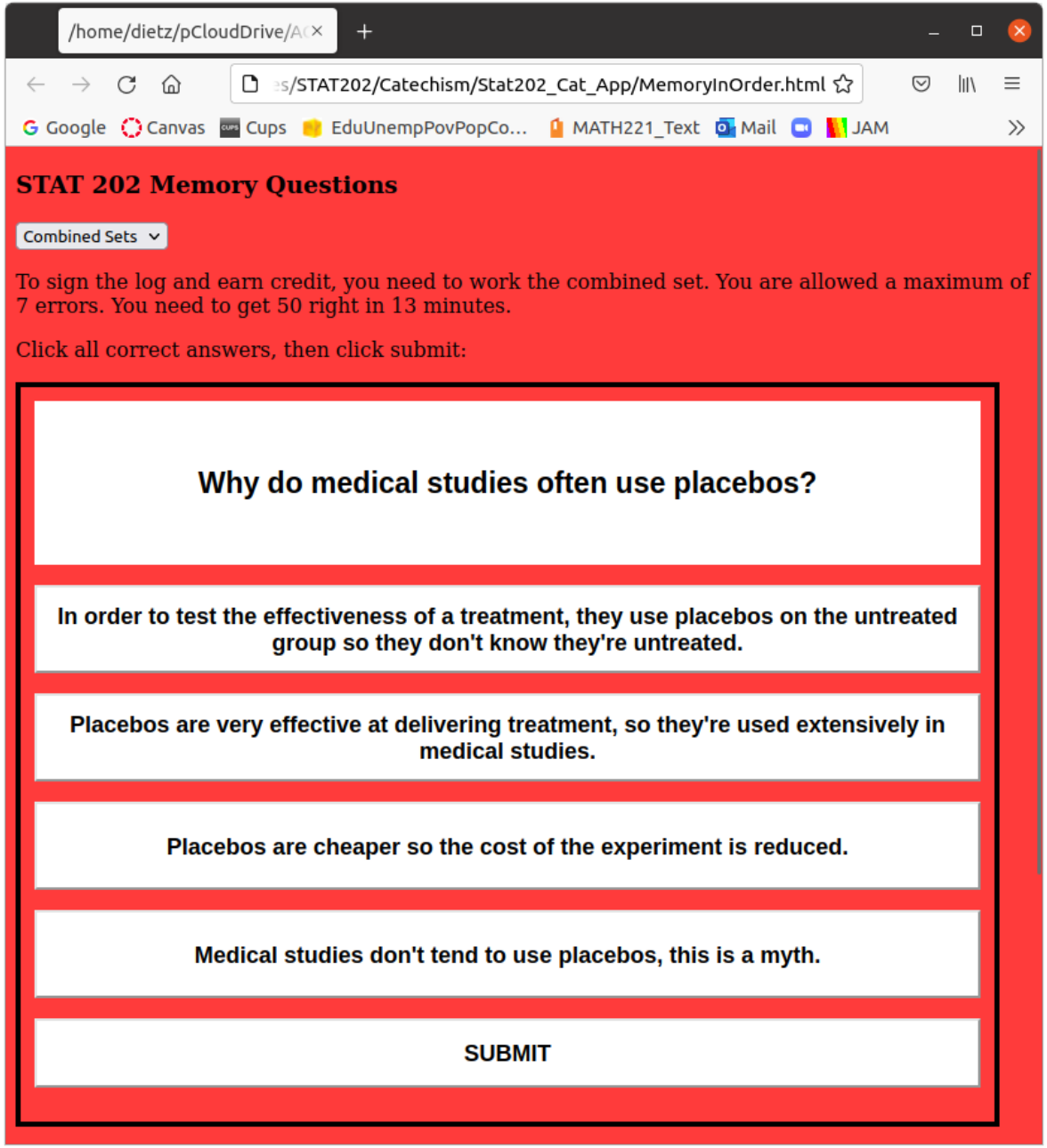

/home/dietz/pCloudDrive/A
/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
What is the Empirical Rule?
We use '68-95-99.7' to remind us of the area values for plus and minus one, two, and 3 standard deviations.
It reminds us that roughly 68% of the data fall within plus and minus one standard deviation for a normal data set.
It reminds us that roughly 95% of the data fall within plus and minus one standard deviation for a normal data set.
We use '95-99.7-99.999' to remind us of the area values for plus and minus one, two, and 3 standard deviations.
SUBMIT

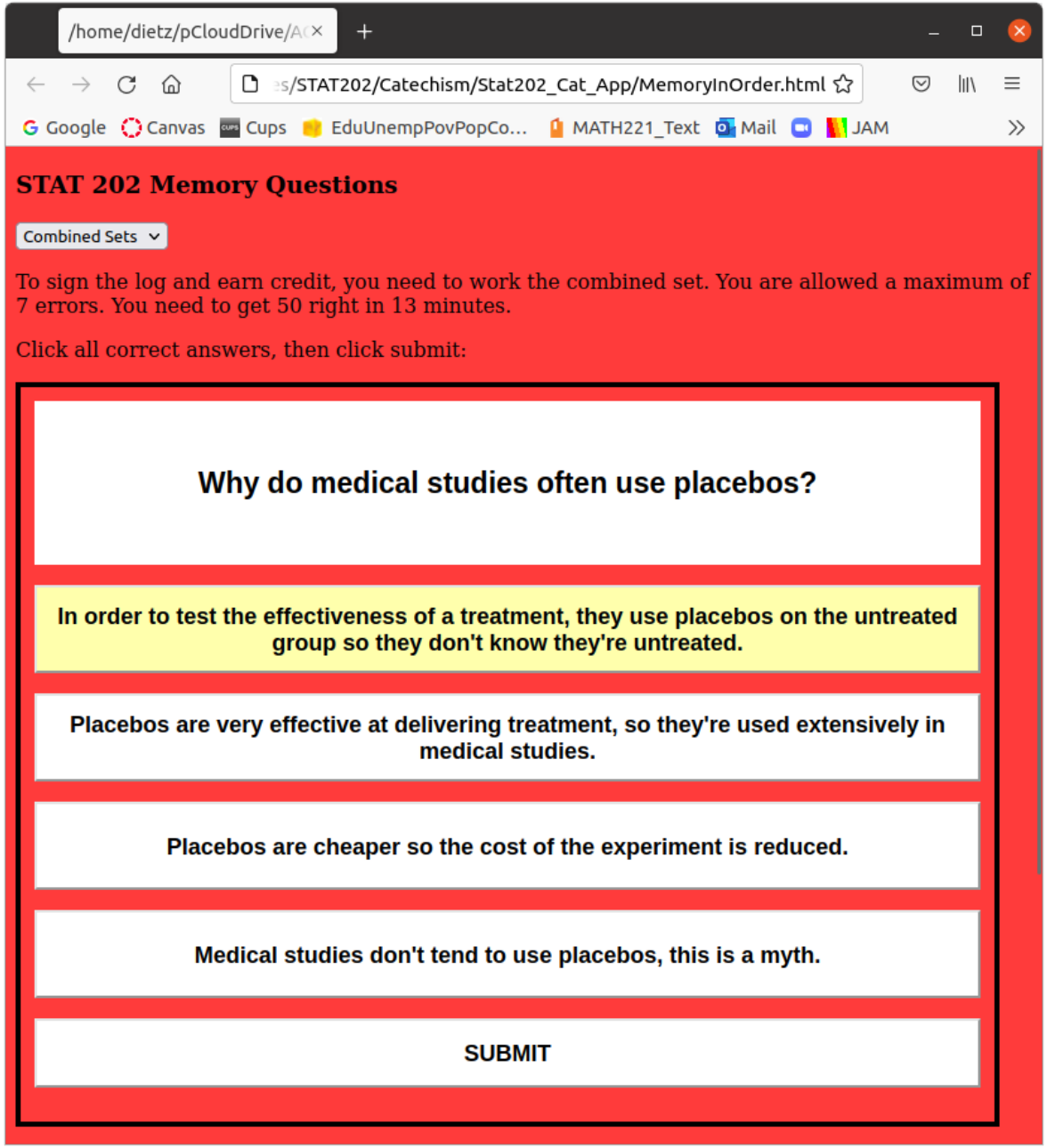

/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
What is the placebo effect?
This can relate to human or animal studies, or studies on physical objects.
This only relates to studies on humans.
Humans often believe there is an effect when they perceive a treatment is happening.
This can relate to human studies or animal studies.
SUBMIT

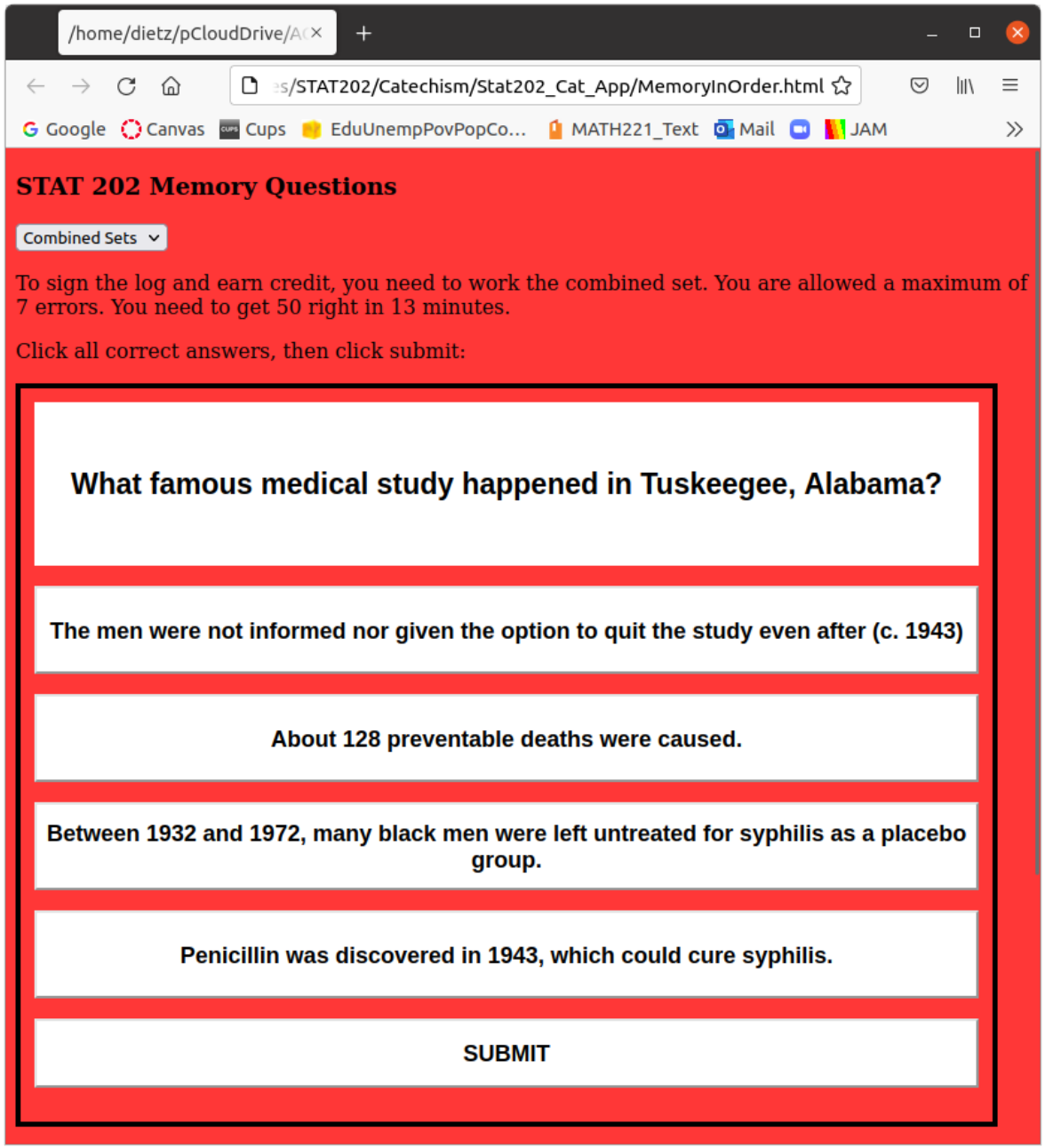

/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
What is the placebo effect?
This can relate to human or animal studies, or studies on physical objects.
This only relates to studies on humans.
Humans often believe there is an effect when they perceive a treatment is happening.
This can relate to human studies or animal studies.
SUBMIT

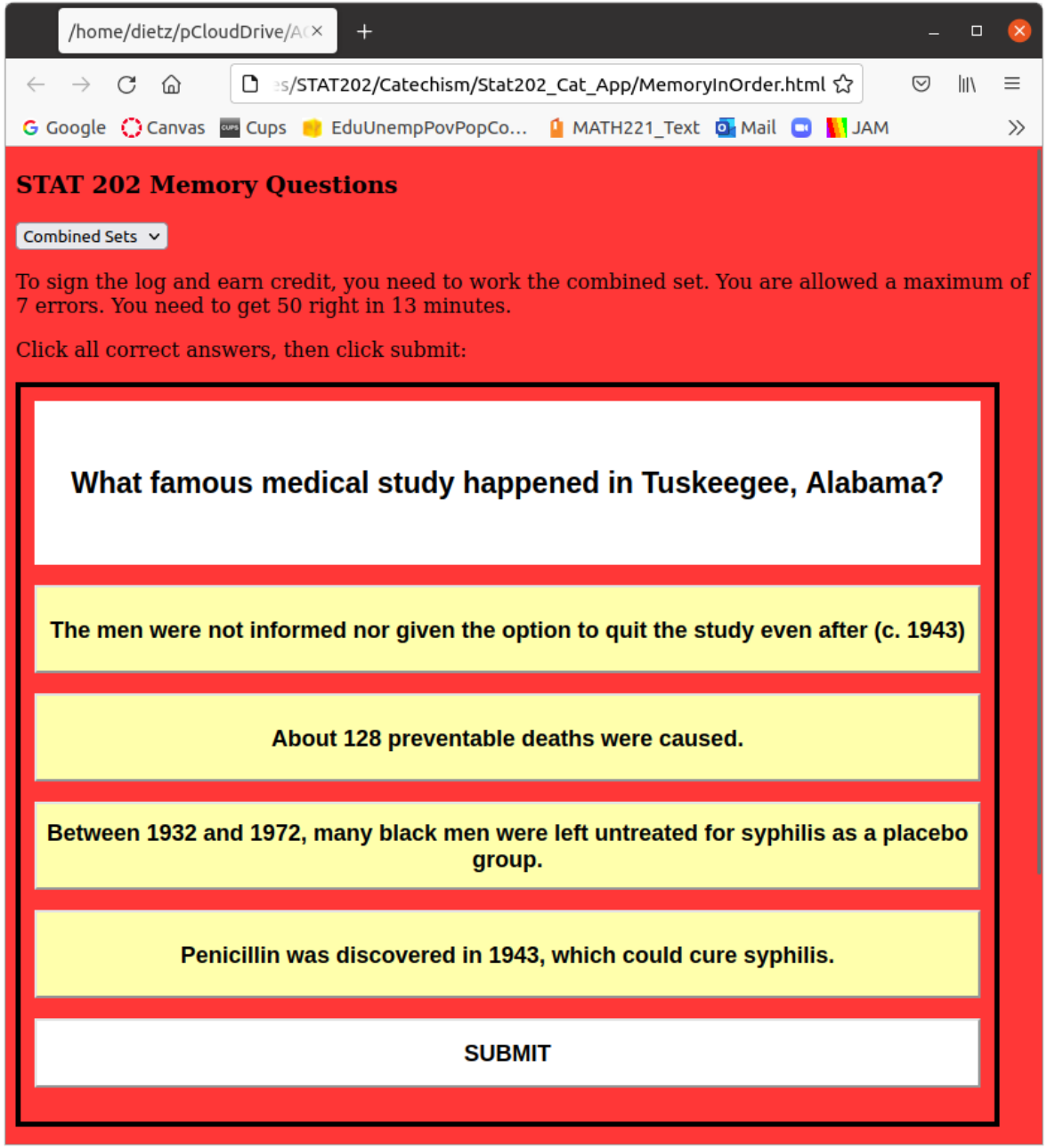

/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
What is a double blind study?
Not only can those receiving the treatments be swayed by the placebo effect, so can the researchers.
Researchers can't be swayed by their emotions, that's why they are researchers.
The best practice is that not only the participants but also the researchers don't know until after the end of the study which participants were in each group.
This only relates to humans.
SUBMIT

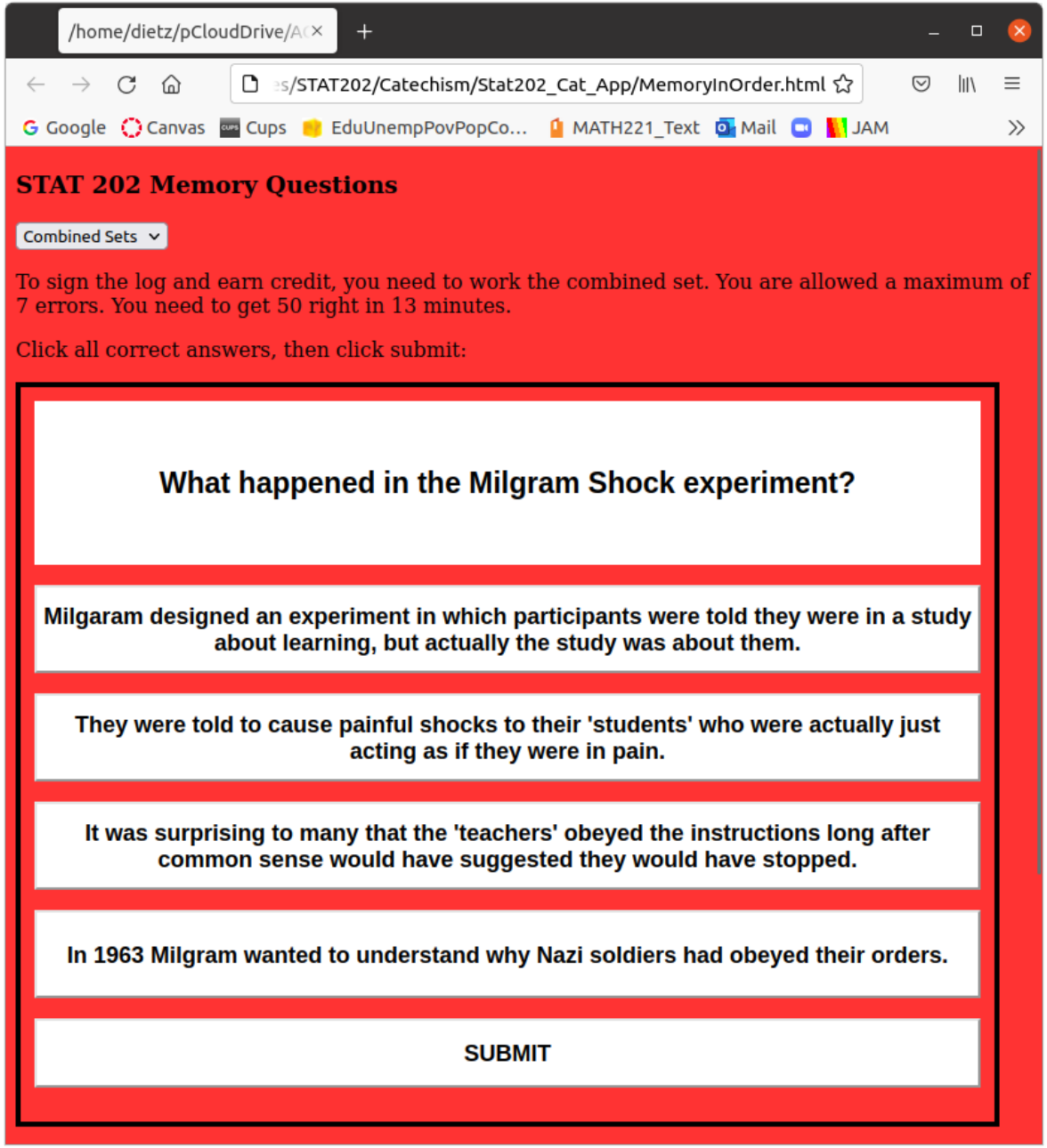

/home/dietz/pCloudDrive/A
s/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
What is a double blind study?
Not only can those receiving the treatments be swayed by the placebo effect, so can the researchers.
Researchers can't be swayed by their emotions, that's why they are researchers.
The best practice is that not only the participants but also the researchers don't know until after the end of the study which participants were in each group.
This only relates to humans.
SUBMIT

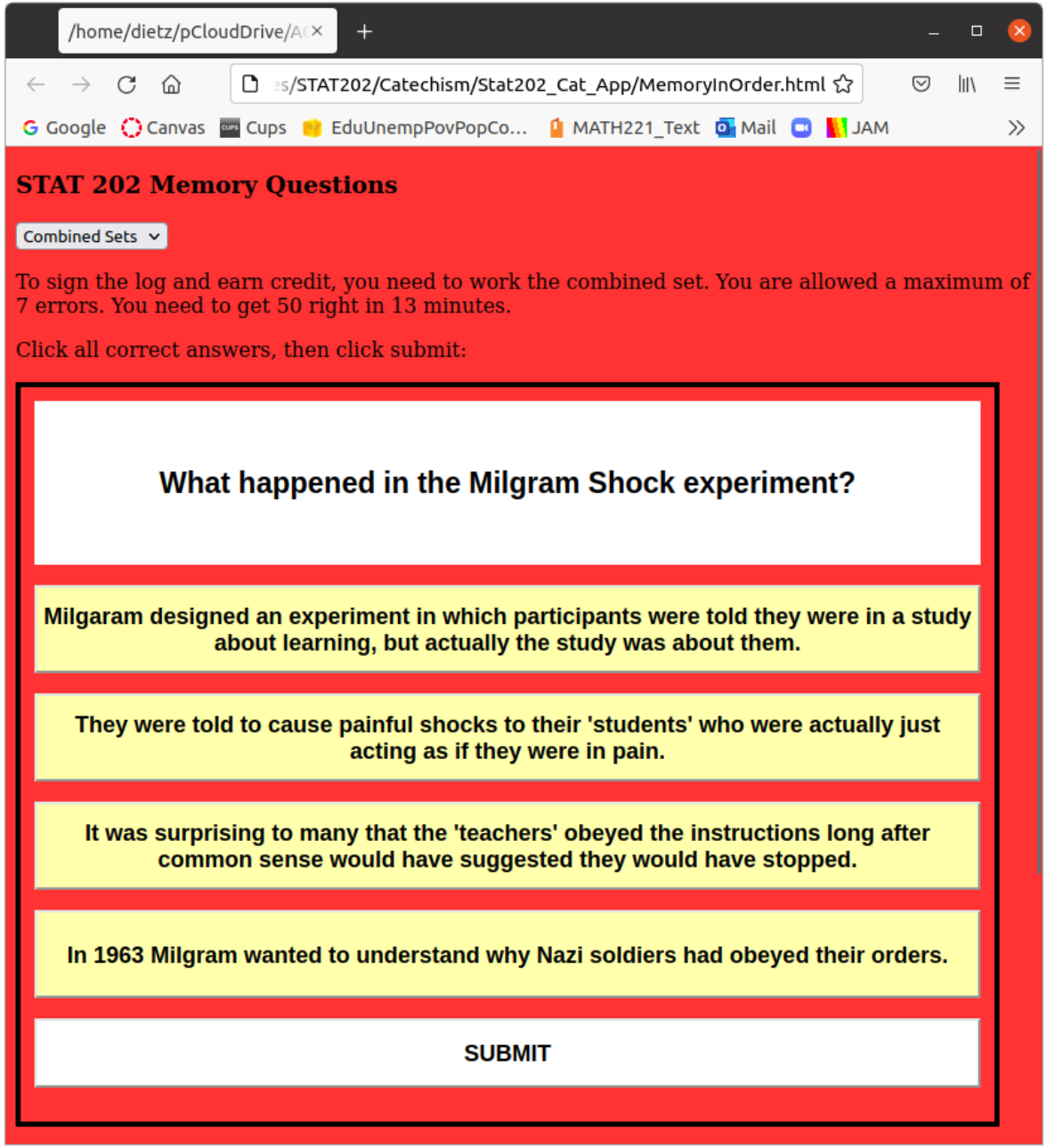

/home/dietz/pCloudDrive/A
/STAT202/Catechism/Stat202_Cat_App/MemoryInOrder.html
Google
Canvas
Cups
EduUnempPovPopCo...
MATH221_Text
Mail
JAM
STAT 202 Memory Questions
Combined Sets
To sign the log and earn credit, you need to work the combined set. You are allowed a maximum of 7 errors. You need to get 50 right in 13 minutes.
Click all correct answers, then click submit:
Why do medical studies often use placebos?
In order to test the effectiveness of a treatment, they use placebos on the untreated group so they don't know they're untreated.
Placebos are very effective at delivering treatment, so they're used extensively in medical studies.
Placebos are cheaper so the cost of the experiment is reduced.
Medical studies don't tend to use placebos, this is a myth.
SUBMIT


/home/dietz/pCloudDrive/A
/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:
Why do medical studies often use placebos?
In order to test the effectiveness of a treatment, they use placebos on the untreated group so they don't know they're untreated.
Placebos are very effective at delivering treatment, so they're used extensively in medical studies.
Placebos are cheaper so the cost of the experiment is reduced.
Medical studies don't tend to use placebos, this is a myth.
SUBMIT


/home/dietz/pCloudDrive/A
/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:
What famous medical study happened in Tuskeegee, Alabama?
The men were not informed nor given the option to quit the study even after (c. 1943)
About 128 preventable deaths were caused.
Between 1932 and 1972, many black men were left untreated for syphilis as a placebo group.
Penicillin was discovered in 1943, which could cure syphilis.
SUBMIT


/home/dietz/pCloudDrive/A
/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:
What famous medical study happened in Tuskeegee, Alabama?
The men were not informed nor given the option to quit the study even after (c. 1943)
About 128 preventable deaths were caused.
Between 1932 and 1972, many black men were left untreated for syphilis as a placebo group.
Penicillin was discovered in 1943, which could cure syphilis.
SUBMIT


/home/dietz/pCloudDrive/A
/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:
What happened in the Milgram Shock experiment?
Milgaram designed an experiment in which participants were told they were in a study about learning, but actually the study was about them.
They were told to cause painful shocks to their 'students' who were actually just acting as if they were in pain.
It was surprising to many that the 'teachers' obeyed the instructions long after common sense would have suggested they would have stopped.
In 1963 Milgram wanted to understand why Nazi soldiers had obeyed their orders.
SUBMIT


/home/dietz/pCloudDrive/A
/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:
What happened in the Milgram Shock experiment?
Milgaram designed an experiment in which participants were told they were in a study about learning, but actually the study was about them.
They were told to cause painful shocks to their 'students' who were actually just acting as if they were in pain.
It was surprising to many that the 'teachers' obeyed the instructions long after common sense would have suggested they would have stopped.
In 1963 Milgram wanted to understand why Nazi soldiers had obeyed their orders.
SUBMIT