Confidence Intervals for a Proportion

Donna Dietz

American University

dietz@american.edu

STAT 202 - Spring 2020

Binary (Binomial) Experiments

Recall: Binary (Binomial) Experiments

$$\mu = np$$
$$\sigma = \sqrt{np(1-p)}$$

Where μ is the expected number of success cases in n trials, each having a success probability of p. Of course σ is the standard deviation, but it is for the μ above.

We still have to avoid $np < 10$ and $n(1-p) < 10$ as before.

What if we want the proportion?

If instead of looking at the total success count, we wish to consider the expected success proportion, we know how to do that! Just divide the entire distribution by n, right?

$$\mu = p$$

$$\sigma = \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}}$$

We still have to avoid $np < 10$ and $n(1-p) < 10$ as before.

95% Confidence Interval

Then, if we want the 95% Confidence Interval

$$\mu \pm 2\sigma$$

$$p \pm 2\sqrt{\frac{p(1-p)}{n}}$$

Which means that we can expect around 95% of the time:

$$p - 2\sigma \leq \hat{p} \leq p + 2\sigma$$

So, let's experiment!

Let's try this 10,000 times!

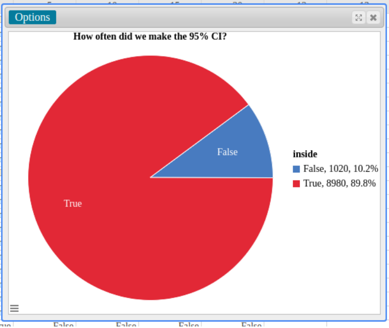
- Pick a p, and some $10 \leq n \leq 120$
- Run the experiment, find the *wins* and $\hat{p}$
- Calculate the 95% Confidence Interval
- See if you've captured the true proportion
- See if $np < 5$ and $n(1-p) < 5$
- And again for 10, 15, and 20.

Results!

experiments.txt

StatCrunch 3 Applets Edit Data Stat Graph

Row	p	n	wins	pHat	L	H	inside
1	0.9698632871	11	11	1	1	1	Fa
2	0.12519735	49	5	0.10204082	0.017284262	0.18679737	Tr
3	0.066261224	31	1	0.032258065	-0.029939612	0.094455741	Tr
4	0.30820085	77	23	0.2987013	0.19647073	0.40093187	Tr
5	0.59398566	94	50	0.53191489	0.43104173	0.63278805	Tr
6	0.095428095	99	10	0.1010101	0.041649501	0.1603707	Tr
7	0.036948586	74	3	0.040540541	-0.0043958612	0.085476942	Tr
8	0.17269724	34	4	0.11764706	0.0093470932	0.22594702	Tr
9	0.93967862	93	90	0.96774194	0.93183209	1.0036518	Tr
10	0.84481315	19	17	0.89473684	0.75674128	1.0327324	Tr
11	0.57976793	99	58	0.58585859	0.48882788	0.68288929	Tr
12	0.10991101	32	1	0.03125	-0.029035373	0.091535373	Fa
13	0.51827768	22	12	0.54545455	0.33738301	0.75352608	Tr
14	0.18616864	24	6	0.25	0.076758839	0.42324116	Tr
15	0.79377726	12	10	0.83333333	0.62247091	1.0441958	Tr
16	0.88876097	58	49	0.84482759	0.7516453	0.93800988	Tr
17	0.50968515	114	54	0.47368421	0.38202602	0.56534241	Tr
18	0.84201268	16	15	0.9375	0.81888989	1.0561101	Tr
19	0.65650885	108	67	0.62037037	0.52884326	0.71189748	Tr
20	0.2077235	108	23	0.21296296	0.13574936	0.29017657	Tr
21	0.55024758	92	49	0.5326087	0.43065415	0.63456324	Tr
22	0.40874699	62	27	0.43548387	0.31206418	0.55890356	Tr
23	0.92190756	69	64	0.92753623	0.86636356	0.9887089	True



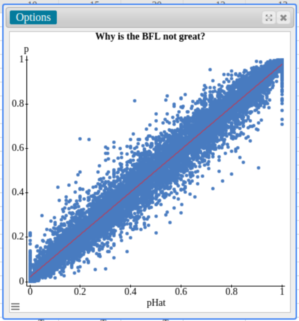
We failed quite a lot! Why?

Results

experiments.txt

StatCrunch Applets Edit Data Stat Graph

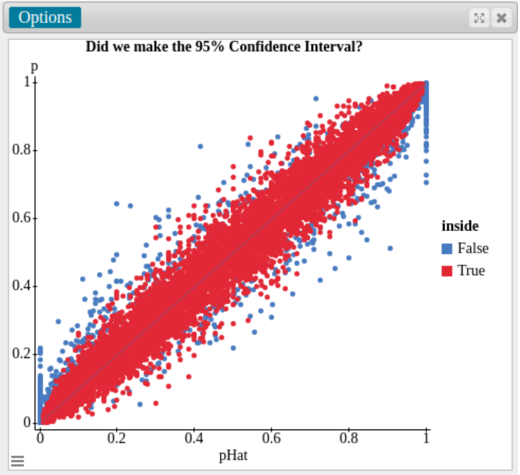
Row	p	n	wins	pHat	L	H	inside	np5
223	0.38672385	35	17	0.48571429	0.32013168	0.65129689	True	True
224	0.17464849	71	11	0.15492958	0.070762836	0.23909632	True	True
225	0.19377663	74	8	0.10810811	0.03735827	0.17885795	False	True
226	0.45290745	14	8	0.57142857	0.31219894	0.8306582	True	True
227	0.93263021	23	22	0.95652174	0.87317749	1.039866	True	False
228	0.14431341	29	5	0.17241379	0.034930572	0.30989701	True	False
229	0.05462214	97	4	0.041237113	0.001666759	0.080807468	True	False
230	0.63522019	113	74	0.65486726	0.56721026	0.74252425	True	True
231	0.66010805	84	58	0.69047619	0.59161227	0.78934011	True	True
232	0.74680244	86	67	0.77906977	0.69138531	0.86675422	True	True
233	0.17002632	61	10	0.16393443	0.07102786	0.25684099	True	True
234	0.0057332701	86	0	0	0	0	False	False
235	0.37889889	100	42	0.42	0.32326253	0.51673747	True	True
236	0.20667507	98	24	0.24489796	0.15975702	0.3300389	True	True
237	0.28493798	20	4	0.2	0.024692271	0.37530773	True	False
238	0.36067411	113	51	0.45132743	0.35957463	0.54308024	True	True
239	0.037523656	14	0	0	0	0	False	False
240	0.4228259	70	30	0.42857143	0.31264041	0.54450244	True	True
241	0.26148641	98	27	0.2755102	0.18705406	0.36396635	True	True
242	0.48604683	62	28	0.4516129	0.32773695	0.57548886	True	True
243	0.77570354	43	35	0.81395349	0.69763938	0.9302676	True	True
244	0.85472874	114	91	0.79824561	0.72457685	0.87191438	True	True
245	0.66038402	74	48	0.64864865	0.53987698	0.75742032	True	True



A plot of p versus $\hat{p}$ values.
Note: $\hat{p} = 0$ and $\hat{p} = 1$ can happen. $p = 0$ and $p = 1$ can't.
Look at lines 234 and 239. Here, the line shown is an LSR line.

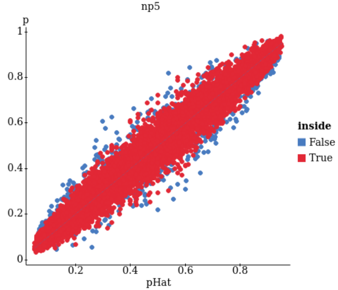
So, a low n or a low or high p (or both) can trigger this type of thing.
Look at the collections of dots where $p = 0$ and $p = 1$ in the previous slide.

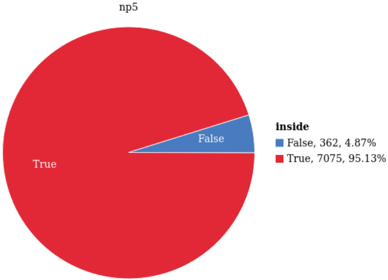
Results



This basically shows that $\hat{p} = 0$ and $\hat{p} = 1$ are not very meaningful.
(This is a $p = \hat{p}$ line, not an LSR line.)

Results



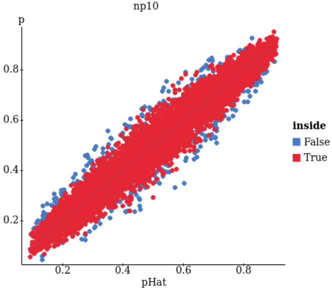


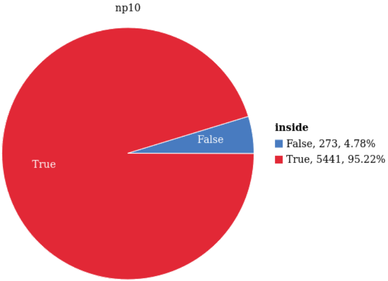
What if we only eliminate cases where $np \leq 5$ or $n(1-p) \leq 5$?

Results

That pretty much gets rid of most of the problem, but let's just look at the progression anyhow, to see how much more we cut down on the issues if we increase that threshold.

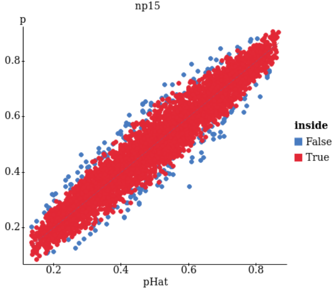
Results

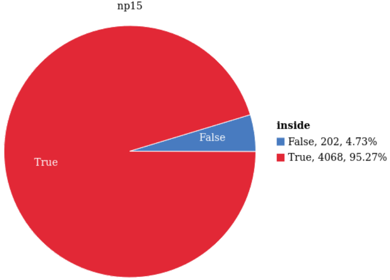




What if we eliminate cases where $np \leq 10$ or $n(1-p) \leq 10$?

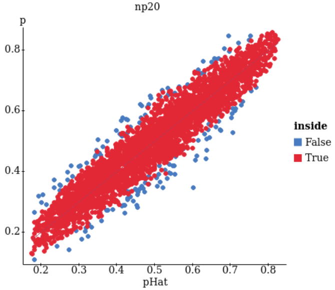
Results

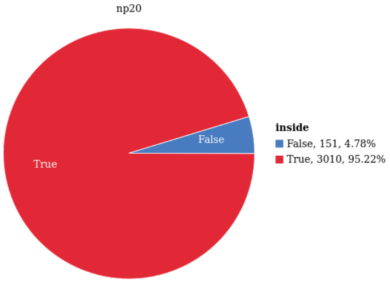




What if we eliminate cases where $np \leq 15$ or $n(1-p) \leq 15$?

Results



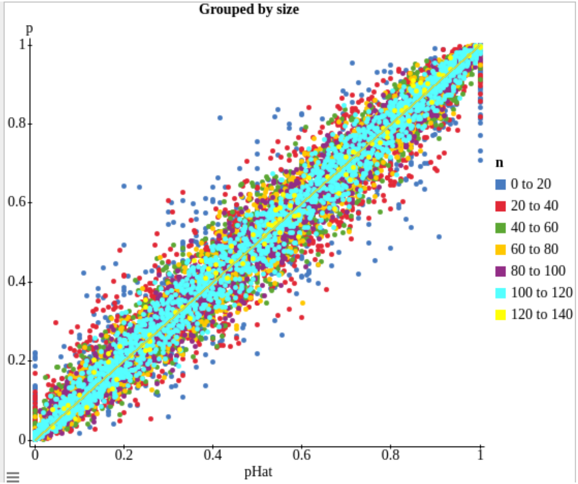


What if we eliminate cases where $np \leq 20$ or $n(1 - p) \leq 20$?

Lesson

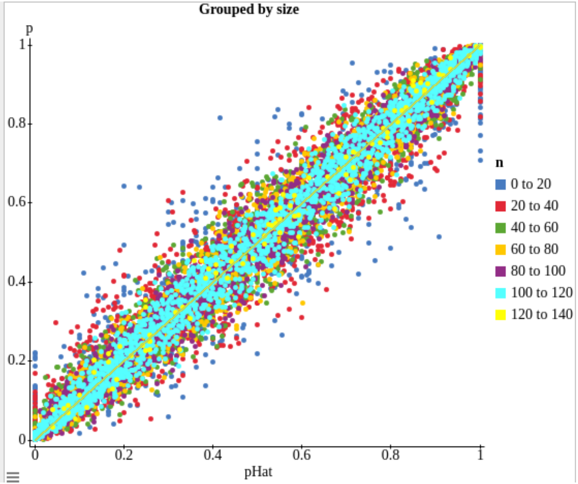
Of course we are progressively getting rid of bad cases, but we're also making it much harder to find low or high p values. What if we tried trimming this a different way?

Results



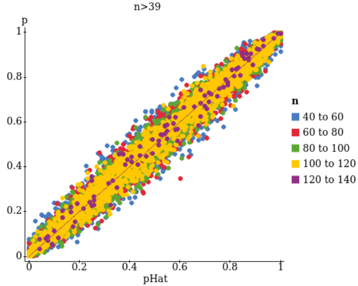
It looks like we could just eliminate small samples and trim this just as well! (The blue and red dots are farthest away from the $p = \hat{p}$ line.)

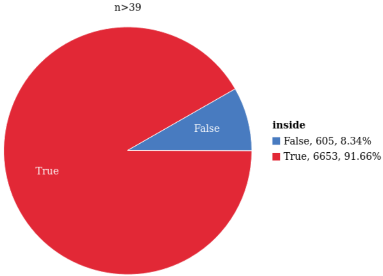
Results



It looks like we could just eliminate small samples and trim this just as well! (The blue and red dots are farthest away from the $p = \hat{p}$ line.)

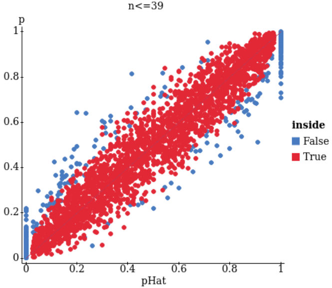
Results

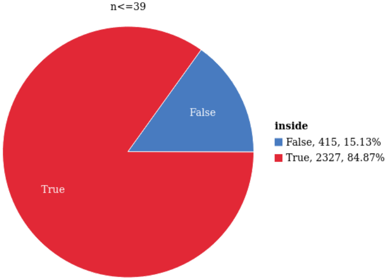




Regardless of how we trim this, we get rid of perfectly good intervals as well!

Results





Look at what we tossed out with this size restriction!

Conclusions

Since the shape of the cloud of points appears to be roughly symmetric about the $p = \hat{p}$ line, we should be able to take this formula:

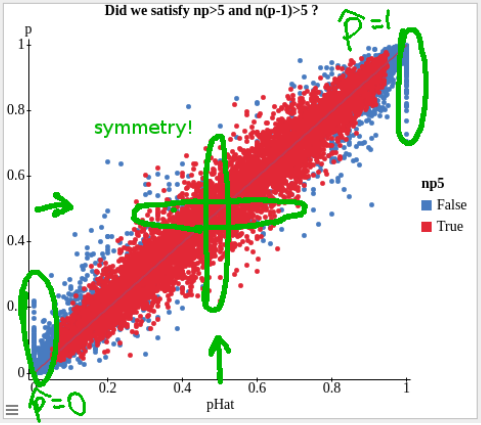
$$p - 2\sqrt{np(1-p)} \leq \hat{p} \leq p + 2\sqrt{np(1-p)}$$

and turn it into:

$$\hat{p} - 2\sqrt{n\hat{p}(1-\hat{p})} \leq p \leq \hat{p} + 2\sqrt{n\hat{p}(1-\hat{p})}$$

(again at a 95% confidence level).

Results



Note symmetry everywhere except $\hat{p} = 0$ and $\hat{p} = 1$.

Margin of Error

Margin of Error

The margin of error in the above formula is $2\sqrt{np(1-p)}$. The margin of error is half the size of your confidence interval.

Confidence Level

Confidence Level

The confidence level gives you a sense of roughly how likely you are to have captured the parameter you were after. But, you have to be very careful how to state this, or you risk getting it wrong.

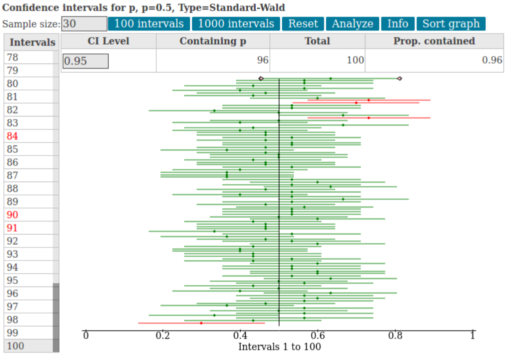
There are many reasons to be careful about how you talk about **the meaning of a confidence level**, including the fact that you can never really eliminate design flaws.

A philosophical question

Many statisticians believe that you should never assign a probability to a confidence interval because once you do, you stop thinking in terms of: *Maybe we captured it, maybe we didn't!*

I like to parahprase it this way: You see two people sitting on a bench, talking closely to each other. Are they dating? Maybe! Maybe not! But you don't dare say, "There is a 95% chance they're dating!" because if they hear you, they may be very embarassed, and you may be as well! They either are or are not dating and that's that. This is how many statisticians like to think about confidence intervals.

Something we can all agree about that's accurate.



With this same process, if we were to resample 100 times or 1000 times, on average, we would expect to capture the parameter 95% of the time if we are using a 95% confidence interval. We avoid attributing a probability of success to a single interval after it's been calculated.

Worksheet

The worksheet will guide you through some experiments, especially to showcase what happens with small sample sizes. You might be surprised!

Next, you'll run experiments with a partner (or small group) to see how well your confidence intervals capture the proportions.

We will find out how to determine a good sample size if you know your confidence level and margin of error.

ME, C-level, n: Pick any two!

You can pick any two of these three, and the remaining one is determined:

- Confidence-level (C-level)
- Margin of Error (M.E.)
- sample size (n)

Examples

$$ME = z\sqrt{\frac{p(1-p)}{n}}$$

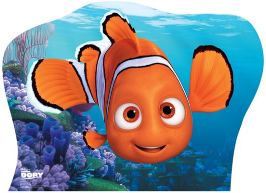
If you want to increase your confidence level, you will either need to increase your sample size or increase your margin of error, or both.

If you want to decrease your margin of error, you will either need to increase your sample size or decrease your confidence level, or both.

If you want to decrease your sample size, you will either need to decrease your confidence level or increase your margin of error, or both.

Note: $z = 1.96$ for 95% confidence-level and
$z = 2.576$ for a 99% confidence-level.

Finding n



We can use a little algebra to show that:

$$n \geq \frac{p(1-p)z^2}{(ME)^2}$$

This will be done in the worksheet! Always round up to the nearest integer, because you can't have part of a person!

Sample size surprise!

After you do a few excercises where you calculate the minimum required sample size, you'll find that to reduce the margin of error by a factor of 2, you have to increase the sample size by a factor of 4. To reduce the margin of error by a factor of 3, you'd have to increase the sample size by a factor of 9. You can actually see why this works by looking at the formula!

$$n \geq \frac{p(1-p)z^2}{(ME)^2}$$

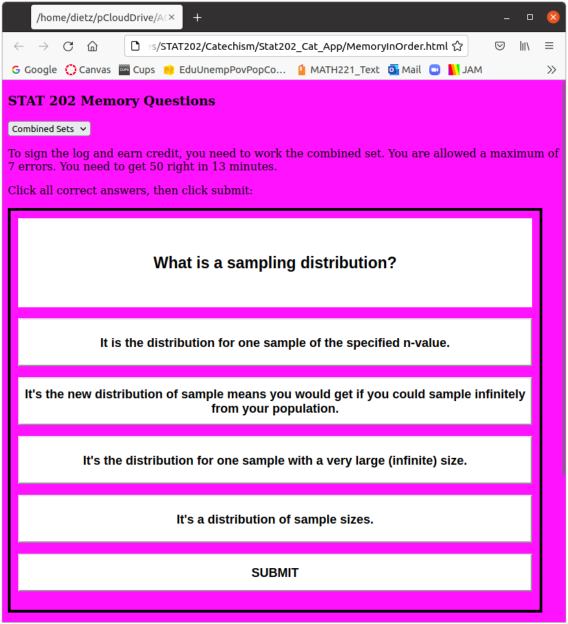
As such, it gets progressively harder to narrow the margin of error. Generally, most major public opinion polls are considered to be about as accurate as they are going to get, once they hit around 500 to 1000 responses.

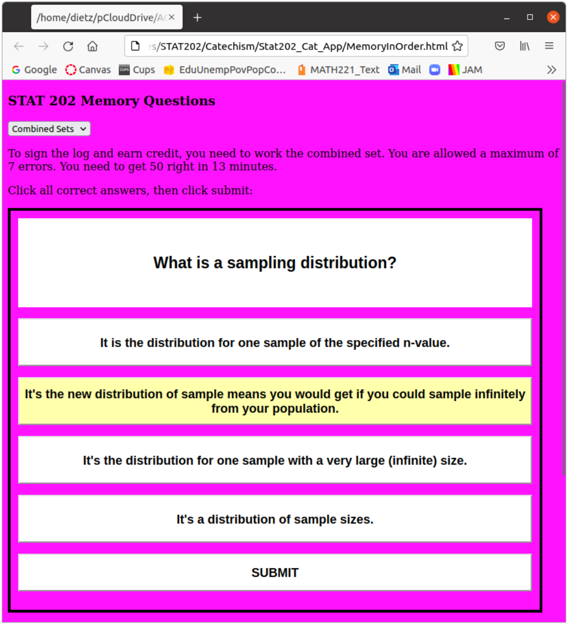
A nice article

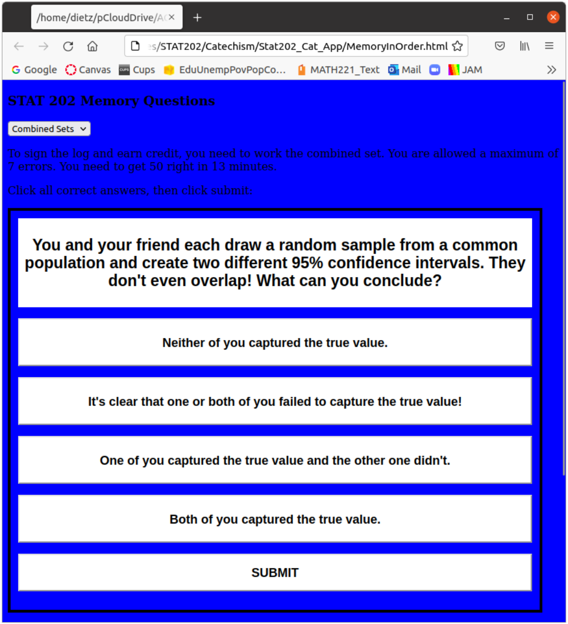
https://www.scientificamerican.com/article/howcan-a-poll-of-only-100/
Is a good read!

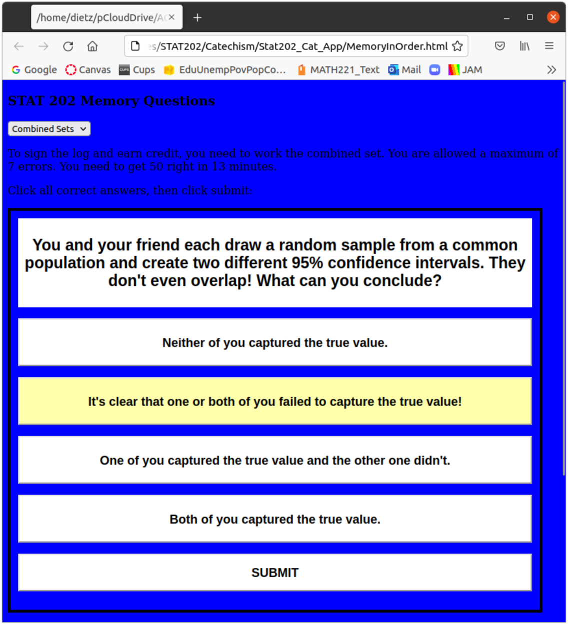
Andrew Gelman explains that a sample size of 250 will give a margin of error of 6%, for example, but a sample size of 100 will only give a 10% margin of error, so it's not very useful to most people. He explains that although you could make a grab for higher accuracy, there is not much point, because public opinion varies over time, so it would be very expensive and not generate useful information.

MEMORY QUESTIONS


/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 sampling distribution?
It is the distribution for one sample of the specified n-value.
It's the new distribution of sample means you would get if you could sample infinitely from your population.
It's the distribution for one sample with a very large (infinite) size.
It's a distribution of sample sizes.
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 sampling distribution?
It is the distribution for one sample of the specified n-value.
It's the new distribution of sample means you would get if you could sample infinitely from your population.
It's the distribution for one sample with a very large (infinite) size.
It's a distribution of sample sizes.
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:
You and your friend each draw a random sample from a common population and create two different 95% confidence intervals. They don't even overlap! What can you conclude?
Neither of you captured the true value.
It's clear that one or both of you failed to capture the true value!
One of you captured the true value and the other one didn't.
Both of you captured the true value.
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:
You and your friend each draw a random sample from a common population and create two different 95% confidence intervals. They don't even overlap! What can you conclude?
Neither of you captured the true value.
It's clear that one or both of you failed to capture the true value!
One of you captured the true value and the other one didn't.
Both of you captured the true value.
SUBMIT