Exam #2

1. I always adore using student-created questions on assessments; here’s one from one of your classmates!

It’s about the SAT (the “Scholastic Aptitude Test”), which is a standardized test folks sometimes have to take to get into college. The part your classmate is talking about is a multiple choice one; it consists of 45 questions like the one at right, where you have a choice of 4 answers (A through D). According to the College Board (the folks that make these tests), the answer choice letters are randomly arranged.

This was your classmate’s question!

“I was always told that if you don’t know an answer to a multiple-choice question on the SAT, you should always guess ‘A’ (or any consistent letter), rather than just randomly guessing. Is that true?”

a. (5 points) Shoot from the hip first: what do you think? Do you think it’s better to stick with one answer choice? Randomly guess? Mix it up? Does it even matter? Write your thoughts here in a sentence or two. If you’re not sure right now, just say so!

OK! So, I figure the best thing to do is to look at some data! I found a College Board site that actually gives a number of practice tests along with their answer keys. At right is the first answer key I found.

b. (1 point) What percent of the answers are A?

c. (1 point) What percent of the answers are B?

d. (1 point) What percent of the answers are C?

e. (1 point) What percent of the answers are D?

OK! Here we have it! We should always choose “C”, since there were more C’s than any other letter!

f. (5 points) Please respond to that statement and gently explain to me why it’s not necessarily correct. 😊 If you need help with inspiration, you can either read the opening paragraphs again, or check out some more answer keys. 😊

Hey! I think my last sentence there actually gave me an idea for how we should proceed! Read on!

---

a There are also “fill in the blank” type questions, but we won’t talk about those here (as they don’t apply).

b https://blog.prepscholar.com/free-sat-practice-tests-massive-collection
If you remember back to the first exam, you might remember that we used a spreadsheet to simulate the game of Woot, because the chance of getting Woots was pretty tough to visualize. While this problem isn’t necessarily as tricky, there is a somewhat confounding mental twist happening that’s been bothering folks for years. So, let’s look at another Monte Carlo spreadsheet model that examines our two strategies! Make sure you close any open spreadsheets you may have open, and then open up this spreadsheet! Watch this video to get started!

Now, do the following in the spreadsheet!

- Click on over to the “Deeper Dive” tab. There, you’ll see the cumulative results (“right” and “wrong”) for both strategies.
- In cells F8, F9, F10, F11 and J8 of that sheet, add some formulas to calculate the percentage of the time that the simulator was correct for each of the scenarios!

(g. (5 points) Does any percentage jump out at you as being different? If so, which one(s)? If not, say so!

Keep going in the spreadsheet!

- Make a bar graph of the Strategy 1 right %’s. Be sure that your axes are well-labeled (especially that horizontal one – make sure it has A through D on it)! Add a nice title, too! If you want, you can also turn off the vertical axes and add the % right to the top of the bars; just make sure that your vertical axis starts at 0% before you get rid of it! Ask me to help if you get stuck!
- Make a pie chart of the right and wrong percentages for Strategy #2 (the easiest way is to just select all the data from H7 down to I8, and form the graph from that). Make sure that the pie chart sections are labeled, you’ve got a nice title, and you add the percentages.

(if you want, you can keep running trials after you’ve created these graphs. It’s kinda fun to watch them change. 😊)

(h. (5 points) Make sure I can see everything you’ve done in that sheet so far, take a screenshot, and add it as your answer to this question!

OK – here are my results, after running about 10000 trials (2000 for each – yours will be different):

---

When I was very, very young, my Pop Pop would take me to the local bowling alley, where he would buy a “pick 3” lottery ticket. He told me that he always played the same number – 701 – because that was my house address number. When I asked him if that was better than changing, he said, “Well, I win every few years, so it’s gotta be better!” Thank you, Pop Pop, for giving me one of my first mathematical pauses. ❤

Probably. 😊
Now, if you show this to a random member of a statistically challenged population, they might say this:

“Clearly, it’s better to always pick answer choice D, because that’s the highest percentage out of all 5 ‘right’ percentages!”

- An adorable misuser of statistics

We’re now going to gently explain why that last statement makes no sense. 😊 In fact, we’ll show this adorable misuser of stats why the highest percentage (25.74% for Strategy 1 D) and the lowest (24.01% for Strategy 1 A) are, essentially, indistinguishable from one another.

There’s a very important value that need to be added and subtracted to and from each of those two percentages – a Margin of Error!

i. (5 points) Remind me, again, why we need to do that!

OK! Here you go – I calculated the MOE’s for you!

<table>
<thead>
<tr>
<th>The statistic of...</th>
<th>...has a MOE of</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.74%</td>
<td>1.88%</td>
</tr>
<tr>
<td>24.01%</td>
<td>1.83%</td>
</tr>
</tbody>
</table>

j. (5 points) Calculate the confidence intervals for both of those statistics (you may have to go back and review the intro to stats, day 2, video 4 if you forgot!)

k. (5 points) Using those two intervals, explain in a sentence or two why the statistical percentages of 25.74% and 24.01% are indistinguishable (this video might help!).

And, in the same way, all of the statistics that we got up there (and, very likely, yours too!¹) can be shown to be statistically indistinguishable from each other – meaning that there’s no clear advantage, at all, over using one method or the other (in fact, you didn’t even really need the simulator to show this – you can show, pretty readily, that the chance of getting a guess right, with either strategy, is 25%). And, since 25% is contained in both of the confidence intervals you got in g, you can show, another way, that neither method is better than the other (nor statistically than 25%).

2. OK! For this question, I decided to chat about COCC tuition; quite a few of you got all riled up about the cost of COCC tuition (as my students often do when we study it). In particular, you all seemed to get really up in arms about how much COCC’s tuition had risen over the past few years.... there are the past 5 years’ worth of increases at right.

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>In-District Tuition (per credit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-2017</td>
<td>$93</td>
</tr>
<tr>
<td>2017-2018</td>
<td>$95</td>
</tr>
<tr>
<td>2018-2019</td>
<td>$99</td>
</tr>
<tr>
<td>2019-2020</td>
<td>$106</td>
</tr>
<tr>
<td>2020-2021</td>
<td>$109</td>
</tr>
</tbody>
</table>

I can hear you all right now: “How come we have to pay so much for tuition? Our tuition went up WAY faster than inflation!!! What gives? I have a buddy at (insert Oregon community college here) who actually pays less than we do at COCC – why?” ² In this question, I want to break down those concerns, and also give you one more skill that you can use to measure financial quantities more effectively!

---

¹ If you hop over to the “MOE” tab, I had Excel keep track of the margins on your data, too.

² To list a few of the questions that a) you all asked and b) I’ve heard over the years. 😊
I went online and gathered up as much data as I could\(^8\) about all the 17 Oregon community college tuitions; it’s in *this spreadsheet for this question* (note: to keep it simpler, I only used in-district tuition). Use that data to answer the following!

a. *(1 point)* Where did COCC rank (1 through 17, “1” being “most expensive”) in 2014? 

b. *(1 point)* Where did COCC rank (same ranking system) in 2020?

Ah, so we’re kinda “in the middle”. Bummer! I was hoping we’d be more affordable.

Oh! But wait! We forgot to adjust for cost of living in those various places!

We spoke about cost of living before – it’s a way to see how two (or more) places compare when it comes to buying stuff in them. I found *this rad tool* to do just that. Here’s how it works, using Pendleton (hometown of Blue Mountain Community College) as an example:

### 2020 Cost of Living Calculator: Bend, Oregon vs Pendleton, Oregon

- Overall, Pendleton, Oregon is 30.0% cheaper than Bend, Oregon

So that means that things in Pendleton only cost 70% of what they do in Bend.

c. *(5 points)* Using that same tool, how does Portland compare to Bend? Is it more expensive, or cheaper? And by how much?

So, clearly, some towns in Oregon are “cheaper to live in” than Bend, and others are “pricier to live in”. And that got me thinking: I think, to be fair when comparing tuitions, we need to adjust all of the tuitions in all of the towns so that we’re “paying with only one town’s dollars”. So, I decided to use Bend as the base dollar amount. *Watch this video so you can see what I mean!*

So, for Pendleton, I multiplied all the tuitions in their unadjusted list by about 1.43 to get what their tuition *should* have been each year. Like this:

<table>
<thead>
<tr>
<th>Year</th>
<th>Blue Mountain (unadjusted)</th>
<th>Blue Mountain (adjusted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>$90.00</td>
<td>$128.57</td>
</tr>
<tr>
<td>2015</td>
<td>$94.00</td>
<td>$134.29</td>
</tr>
<tr>
<td>2016</td>
<td>$96.00</td>
<td>$137.14</td>
</tr>
<tr>
<td>2017</td>
<td>$103.00</td>
<td>$147.14</td>
</tr>
<tr>
<td>2018</td>
<td>$108.00</td>
<td>$154.29</td>
</tr>
<tr>
<td>2019</td>
<td>$108.00</td>
<td>$154.29</td>
</tr>
<tr>
<td>2020</td>
<td>$110.00</td>
<td>$157.14</td>
</tr>
</tbody>
</table>

d. *(5 points)* *(w)* You do Portland! What multiplier would I have to use to get “Portland dollars” into “Bend dollars?” Since Portland is more expensive than Bend (as you saw in C), your number will be less than 1! Show me (or explain to me) how you figured it!

\(^8\) I actually asked HECC (the group that keeps this data) for older data (pre-2014). Still haven’t heard back from them. I *did* go ahead and check every CC’s website to get the 2020 data, though.
So then I did that for all the towns that each of the Oregon community colleges are located in! Click on over to the “Adjusted Tuition” tab, and voila! You’ll see all of the Oregon community colleges’ tuitions, adjusted for cost of living! You’ll notice that a couple schools’ tuitions went down; they’re the ones that are in cities that are more expensive, overall, than Bend.

e. (2 points) Using these more “apples/apples” data, where did COCC rank (1 to 17, with, again, 1 being most expensive) in 2014?

f. (2 points) How about in 2020?

Yay! I feel much better! And you should, too – because the board of directors at COCC does a very good job of keeping your tuition as low as they possibly can. 😊

One last question for this topic:

g. (5 points) I discovered something somewhat shady as I was researching Chemeketa for this chart. Find the “tuition” page on their website, and tell me why saying their in-district tuition per credit is “$95” is…well, kinda dishonest.

3. “There is broad consensus that widespread SARS-CoV-2 testing is essential to safely reopening the United States. A large concern has been testing availability, but test accuracy may prove a larger long-term problem.”

Friends, I’ve tried, throughout this term, to mention COVID as little as possible. I know we’re all dealing with it, and we’ve all been affected by it. But you know, friends…I think we have to chat about it now. I decided to include this for two reasons:

• It sure seems here to stay.
• We need to talk about diagnosing those who might have it to try to keep it from growing out of control.

a. (5 points) Here’s a scenario: suppose you got to get tested for C-19, and the test comes back positive. How worried should you be? What is the % chance, do you think, that you have it? Use a scale of 0% to 100%, “0%” meaning “there’s no way I have it” to “100%” meaning “I absolutely have it”. (you’re just estimating here, based on your intuition; we’ll be working out the numbers soon!) Also write a sentence or two about why you answered the way you did!

Now, in a perfect world, if you tested positive for something, that would mean you had it, without a doubt. Conversely, testing negative for something would mean you wouldn’t have it. But we don’t live in a perfect world: we live in a world where tests are sometimes wrong – and tests like these can be wrong in two different ways. But, they can also be right in two different ways.

Confused? Totally understandable. 😊 We’ll walk through the possibilities here in a chart.

---

h Woloshin, Et. Al, NEJM.
<table>
<thead>
<tr>
<th>Do you have COVID?</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>This is a “correct” result! You have C19, and the test told you that you have C19. This is called a <strong>true positive</strong>.</td>
<td>This is called a <strong>false positive</strong>. You don’t actually have COVID, but the test says that you do. A “wrong” result!</td>
</tr>
<tr>
<td>You take a COVID test and it comes back...</td>
<td>You have COVID, but the test says you don’t. This is called a <strong>false negative</strong>. Another “wrong” result!</td>
<td>You don’t have COVID, and the test says you don’t. This is good! And it’s called a <strong>true negative</strong>.</td>
</tr>
<tr>
<td>Negative</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

So, again: in an ideal world, you’d want those **green boxes** to happen all the time (depending on whether or not you have COVID) and you’d want the **red boxes** to happen...well, never! But that’s the problem: those false positive and negatives do happen!

“OK, Sean – what’s the big deal?” You might ask. “So what if you get a false positive test result? The test comes back positive, it was wrong, and maybe you were a little inconvenienced because you had to quarantine. Big deal!”

b. **(5 points)** What about a false **negative** COVID test (or, I would argue, a false negative test for any infectious disease). Why can **that** be a huge deal?

So, yeah...it **is** a big deal to “get the test right” – but “getting it right” means, actually, 4 different things (as we just saw):

- Maximizing the chances of true positives
- Maximizing the chances of true negatives
- Minimizing the chances of false positives
- Minimizing the chances of false negatives

So, that’s what we’ll study in this question: the mathematics of how some of the currently available COVID tests work!

Now, part of what makes this question interesting (and not-at-all trivial) is that no one really knows, for sure, what the infection rate for COVID is. In the hour or so of my searching of various CDC sites, I found that it appears as though the **infection rate might average around 2%**. So, we’ll use that figure!

So now, let’s break the US down into, say, small groups of 100,000 people. Assuming that, indeed, the infection rate is 2%, that means that 2000 of them will get COVID, and the other 98000 won’t.¹ Let’s redraw that table from above with this new information!

---

¹I can’t emphasize this enough: this is using the 2% as an **average** infection rate across the entire country. As we’ve all seen, COVID tends to pop up in “hot spots”, and other places have vastly lower numbers. We’ll deal with that last!
Do you have COVID?

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive</strong></td>
<td>2000</td>
</tr>
<tr>
<td><strong>Negative</strong></td>
<td>0</td>
</tr>
</tbody>
</table>

OK! That’s coming along nicely¹.

But remember – if all 100,000 of those people take a COVID test, some will come back positive, and the rest will come back negative. And, in an *ideal* testing world, all the folks with COVID would get positive test results, and all the folks without it would get negative test results:

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive</strong></td>
<td>2000</td>
</tr>
<tr>
<td><strong>Negative</strong></td>
<td>0</td>
</tr>
</tbody>
</table>

But, the tests *aren’t* perfect, and we *don’t* live in an ideal world...so we need to deal with false positives and false negatives! I took a trip back out to the interwebs and found that the false positive rate for the most effective COVID test (the painful sounding “deep nose swab” one) was between *0.8%* and *4%*.² Let’s assume, for purposes of illustration, it’s *1%*.

So what *that* means is this: 1% of people who *don’t* have COVID will still test positive. And that means the chart would look more like this:

---

¹ I mean, other than the fact that we’re talking about a deadly pandemic. The *math* looks good, though.
² [https://www.thelancet.com/journals/lanres/article/PIIS2213-2600%2820%2930453-7/fulltext](https://www.thelancet.com/journals/lanres/article/PIIS2213-2600%2820%2930453-7/fulltext)
Do you have COVID?

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>980</td>
</tr>
<tr>
<td>Negative</td>
<td>97020</td>
</tr>
</tbody>
</table>

(we’ll deal with this column in a sec!)

You take a COVID test and it comes back...

(in case you’re not seeing how I got those numbers: 98000 people don’t have COVID. Assuming they all get tested, 1% of them – that is, 980 – will falsely test positive. The rest will test negative)

So, that’s kinda wild – even with just a 1% false positive rate, you still have almost 1000 people testing positive! But that kinda makes sense – most people don’t have COVID. #thankfully

OK – now it’s time to deal with false negatives. Remember, these are the ones we should really be nervous about! According to more sites I was able to track down\(^1\), the false negative rate is likely between 2% and 37%. Let’s assume it’s on the low end of that range at 2%; the, we can extend our last chart to look like this!

Do you have COVID?

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>1960</td>
</tr>
<tr>
<td>Negative</td>
<td>40</td>
</tr>
</tbody>
</table>

You take a COVID test and it comes back...

OK! Now we can look again at your answer from #2: up there, you estimate your chance of having COVID, assuming you tested positive. We’re going to figure out that chance right now, based on the above chart!

---

\(^1\) https://www.health.harvard.edu/blog/which-test-is-best-for-covid-19-2020081020734#:~:text=The%20reported%20rate%20of%20false,5%25%20or%20lower.

\(^m\) https://www.npr.org/sections/health-shots/2020/04/21/838794281/study-raises-questions-about-false-negatives-from-quick-covid-19-test
Chance you have COVID, assuming you tested positive
\[
= \frac{\text{those who have COVID and tested positive}}{\text{all those who tested positive}} = \frac{1960}{1960 + 980} \approx 67\%
\]

Hmmm…that seems low, doesn’t it? If I tested positive for something, I’d like to be sure (at least, mostly sure) that I had it! So why is this number so low?

Think of it this way: of every 3 people that test positively, only 2 of them likely have COVID (the third is a false positive):

\[
\text{Chance you don’t have COVID, assuming you tested positive} = \frac{\text{those who don’t have COVID and tested positive}}{\text{all those who tested positive}} = \frac{980}{1960 + 980} \approx 33\%
\]

But, again…false positives, while annoying, aren’t necessarily the ones I’m worried about. No…I’m more concerned about the false negatives (especially after reading what y’all put up here in A). 

c. \text{(5 points) (w)} Using the numbers in the above chart (and a similar approach as I did right below that chart), find the % chance that, if someone tests negative for COVID, they don’t have it.

d. \text{(1 point)} That last value was the chance of a true negative from a COVID test. What’s the chance of a false negative?

And that seems pretty low, which is what we’d want! But, there are a couple of things to consider:

e. \text{(1 point)} Look at the chart again: of every 100000 people tested, how many will be falsely negative? Don’t overthink this one!

And that’s a small number, but still…those people can then go out and wander around, potentially infecting those around them. But, even more important that that is the fact that we used the absolute “best case scenario” false positive and false negative rates. Remember from above that false positive rates can be as high as 4% and false negative rates can be as high as 37%.

Use these new “worst case scenario” values (still use the infection rate of 2% for now) and find:

f. \text{(5 points) (w)} The chance that someone who tests positive actually has COVID (if you like, you can use this handy dandy little tool I built for you; it’ll get you the table like above; all you have to do is show me the fractions you create and give me the final percentage).

g. \text{(5 points) (w)} The chance that someone who tests negative actually doesn’t have COVID.

Whoa! Now, a positive test means you’re actually about 3 times more likely to not have COVID (while a negative test still means you’re pretty likely to not have it, either). Woof – with those false positive and false negative values, it’s really not even worth taking a test, honestly.

\(^{\circ}\) Of course, I’m typing this about a month and a half before you actually turn this in. But I have faith in your future answers. 😊
Unless...

As I said earlier, no one **really** knows the incidence rate of COVID. For the past few questions, I had you assume that it was 2%, but that was just a CDC average; in reality, it varies on where you are, and what you’re doing. For example, here in Central Oregon, my family and I are able to quarantine very effectively from other people and, therefore, the chance of us getting the disease is pretty low when compared to, say, a front-line healthcare worker.

In fact, the DOL has released a database\(^\circ\) detailing the numbers of various occupations and the risk of COVID transmission associated (not necessarily caused) by being in those professions; [here’s a cool visual.](https://www.onetonline.org/find(descriptor)/result/4.C.2.c.1.b?result-based=1) If you take a look at that scatterplot, you’ll see three pieces of information:

- On the **vertical** axis, the median salary of that group.
- On the **horizontal**, the “risk” axis…the higher the risk, the more likely to catch COVID.
- The size of the bubble. That’s how many folks in in those professions across the US.

It looks like Dental Hygienists are at the greatest risk for contracting COVID (according to this metric; they’re furthest to the right).

h. **(1 point)** Ballpark what their “risk score” is.
   
   I’m a “postsecondary teacher.”

i. **(1 point)** Ballpark what *my* risk score is.

j. **(2 points)** How many more times likely is a dental hygienist to contract COVID than me? You can round to the nearest whole number if you like.\(^p\)

   Now, what I want you to do is run those numbers through the spreadsheet and see the effect that extra risk has on the dental hygienist’s test results. We don’t actually **have** infection rates, so we’ll have to supply them. Let’s suppose that dental hygienists have an infection rate of just about 3% (since they’re higher than the average in the population).

k. **(1 point)** What’s a postsecondary teacher’s infection rate?

   Assume a dental hygienist takes a COVID test (for these last few questions, let’s assume false positive and false negative rates of 2%).

l. **(2 points)** If the test comes back positive, what’s the chance they actually have COVID?

m. **(2 points)** If their test comes back negative, what’s the chance they don’t have COVID?

   Now, assume I take a COVID test.

n. **(2 points)** Repeat l for me.

o. **(2 points)** Repeat m for me.

---

\(^\circ\) [https://www.onetonline.org/find(descriptor)/result/4.C.2.c.1.b?result-based=1](https://www.onetonline.org/find(descriptor)/result/4.C.2.c.1.b?result-based=1)

\(^p\) That visual you looked at should have made the following disclaimer: “We’re not exactly sure how these folks contracted COVID...just that these are the professions they held *while* they contracted it.”
So, just like almost everything else in world we live in, there are multiple factors influencing the outcomes. Your relative risk can affect the results of the test, as can how the test is processed (we didn’t get into that here). Heck, even which COVID test you take\(^q\) can change the outcome.

So, like so much you saw in this class: it depends. But now you have the math to analyze the variables as they come in!