At a small college, the average student GPA is 2.8, and the standard deviation is .5. Students with GPAs below -1.8 (as measured by Z-score) are put on academic probation. What is the corresponding GPA for academic probation? (3pts)

$$\mu = 2.8 \quad \sigma = .5$$

$$Z = \frac{X - \mu}{\sigma}$$

So

$$X = -1.8(.5) + 2.8$$

$$X = 1.9$$

If the school wants to graduate approximately the top 2.5% of students with honors, where should the administration set the cutoff GPA (give both Z-score and GPA value)? (3pts)

For normal dist, 95% of values fall within 2 SDs.

So, upper cutoff can be set at

$$Z = 2$$

(don't need table!)

$$X = z(\sigma) + \mu$$

$$X = 2(.5) + 2.8$$

Imagine that, in subsequent years, the spread in the distribution of GPAs among the student body decreases. How would you expect the cutoff GPAs in the previous two questions to change? (3pts)

Both should move closer to the mean!

$$Z$$ remains the same.
A study of government run health care facilities examined the effectiveness of 5 features in various facilities to prevent falls by wheelchair users. Some of the wheelchair users had access to no features (35.6%); others had access to at least one (61.5%), while others had access to all 5 (2.9%). The results of the study are summarized below.

Total $N=306$, proportion of wheelchair users who had experienced a fall.

<table>
<thead>
<tr>
<th>Facility Features</th>
<th>Falls</th>
</tr>
</thead>
<tbody>
<tr>
<td>All 5</td>
<td>22.2%</td>
</tr>
<tr>
<td>At least 1, not all</td>
<td>13.9%</td>
</tr>
<tr>
<td>None</td>
<td>18.3%</td>
</tr>
</tbody>
</table>

What is the probability that any wheelchair user had a fall? (3pts)

$$0.356(0.183) + 0.615(0.134) + 0.099(0.222)$$

$$= 0.157$$

$P(\text{fall})$

What is the probability that a wheelchair user did not have a fall and had access to none of the features? (3pts)

$$P(\text{no fall } \cap \text{ no features}) = P(\text{no fall } | \text{ no features}) \cdot P(\text{no features})$$

$$= 0.817(0.356)$$

$$= 0.291$$

Of users in facilities with no features installed, what was the probability of a fall? (3pts)

$$P(\text{fall } | \text{ no features}) = \frac{0.183}{0.291}$$

Given these results, what would your suggestion be regarding the use of features to prevent falls among wheelchair users? Explain your rationale in one to two sentences. (3pts)

"Could make a variety of arguments, based on the data."
A large non-profit is considering investing in new software to test for spam in incoming email. This software been previously been shown to correctly identify 90% of all spam email. However, 10% of email that is valid is incorrectly identified as spam. Right now, the best estimate is that 15% of all incoming email is spam. If the software identifies an individual message as spam, what is the probability that it is actually spam? (8pts)

\[ P(\text{spam}) = 0.15 \quad P(\text{label} \mid \text{spam}) = 0.9 \]
\[ P(\text{not spam}) = 0.85 \quad P(\text{label} \mid \text{not spam}) = 0.1 \]

Find \( P(\text{spam} \mid \text{label}) \)

\[
P(\text{spam} \mid \text{label}) = \frac{P(\text{label} \mid \text{spam}) \cdot P(\text{spam})}{P(\text{label} \mid \text{spam}) \cdot P(\text{spam}) + P(\text{label} \mid \text{not spam}) \cdot P(\text{not spam})} = \frac{0.9 \cdot 0.15}{0.9 \cdot 0.15 + 0.1 \cdot 0.85} = 0.614
\]

Would you recommend that the organization purchase this software? Why or why not? (2pts)

Could reasonably argue either way.

Now imagine that a different estimate claims that only 3% of incoming email is spam (instead of 15%), would you expect the non-profit to be more or less likely to purchase the software? Explain why. (5pts)

Less likely!

As base rate drops, prob. of correctly detecting spam drops.
A recent study was reported in the *NYTimes* that claimed that cranberry juice does indeed help to prevent bladder infections, confirming previous scientific notions. If you were reading this article, what would be 3 pieces of information that you would like to evaluate this claim, and why? (5pts)

[USE ONLY AVAILABLE SPACE BELOW]

- Sample size
- Sampling method
- How much needed
- How measuring “help”
  ... etc.

Imagine you are considering purchasing a home. You are considering two different neighborhoods with comparable attributes in terms of amenities and distance from your job. Furthermore, your real estate agent tells you that these neighborhoods have almost identical average home values.

However, you realize that you might like more (statistical) information about home prices in these neighborhoods. What are two other pieces of information you might ask about, and why? (5pts)

[USE ONLY AVAILABLE SPACE BELOW]

- Info about spread
  (SD, for example)
- Changes over time
- Measures of median and/or mode
  (other descriptive/s)