Congratulations! You passed!

Grade received 100% **To pass** 80% or higher

Go to next item

1. What is the sensitivity and specificity of a pneumonia model that always outputs positive? In other words, the models says that every patient has the disease.

1 / 1 point

- ○ sensitivity = 0.5, specificity = 0.5
- ◉ sensitivity = 1.0, specificity = 0.0
- ○ sensitivity = 0.0, specificity = 1.0
- ○ sensitivity = 1.0, specificity = 1.0

Correct

- Sensitivity tells us how good the model is at correctly identifying those patients who actually have the disease and label them as having the disease.
- Specificity tells us how good the model is at correctly identifying the healthy patients as not having the disease.

A sensitivity of 1 would mean that the model identifies all the diseased patients as having the disease, and does not identify any healthy patients as healthy. This is what the model is doing in this example.

2. In some studies, you may have to compute the Positive predictive value (PPV) from the sensitivity, specificity and prevalence.

1 / 1 point

Given a sensitivity = 0.9, specificity = 0.8, and prevalence = 0.2, what is the PPV (positive predictive value)?

HINT: please check the reading item "Calculating PPV in terms of sensitivity, specificity and prevalence"

- ◉ 0.53
- ○ 0.02
- ○ 0.9
- ○ 0.18

Correct

$PPV = \frac{sensitivity \times prevalence}{sensitivity \times prevalence + (1 - specificity) \times (1 - prevalence)}$

The numerator is (sensitivity * prevalence) = 0.9*0.2 = 0.18.

The denominator is

0.18 + 0.2 * 0.8 = 0.34.

Therefore the PPV is 0.18/0.34 ~ 0.52

3. If sensitivity = 0.9, specificity = 0.8, and prevalence = 0.2, then what is the accuracy?

1 / 1 point

Hint: You can watch the video "Sensitivity, Specificity and Prevalence" to find the equation.

- ◉ 0.82
- ○ 0.75
- ○ 0.52
- ○ 0.44

Correct

The equation for accuracy is:

$$Accuracy = (Sensitivity \times Prevalence) + (Specificity \times (1 - Prevalence))$$

So accuracy = (0.9*0.2) + (0.8*0.8) = 0.82

4. What is the sensitivity and specificity of a model which randomly assigns a score between 0 and 1 to each example (with equal probability) if we use a threshold of 0.7?

1 / 1 point

- ○ Not enough information to answer the question.
- ○ Sensitivity = 0.7, Specificity = 0.3
- ○ Sensitivity = 0.5, Specificity = 0.5
- ◉ Sensitivity = 0.3, Specificity = 0.7

Correct

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$Sensitivity = P(\hat{pos}|pos) = P(score > 0.7|pos)$

Our score is independent of the input data (it randomly assigns 0 or 1 predictions) so

$P(score > 0.7|pos) = P(score > 0.7) = 0.3$

Similarly,

$specificity = P(\hat{neg}|neg) = P(score < 0.7|neg) = P(score < 0.7) = 0.7$

5. What is the PPV and sensitivity associated with the following confusion matrix? **1 / 1 point**

Recall that

$PPV = \frac{\text{TruePositives}}{\text{positive predictions}}$

$Sensitivity = \text{How many actual positives are predicted positive?}$

	Test Positive	Test Negative
Disease Positive	30	20
Disease Negative	70	10

◉ PPV = 0.3, Sensitivity = 0.6

○ Not enough information is given

○ PPV = 0.6, Sensitivity = 0.33

○ PPV = 0.4, Sensitivity = 0.2

Correct

$PPV = P(pos|\hat{pos})$

$PPV = \frac{TP}{TP+FP}$

$PPV = \frac{30}{30+70} = 0.3$

Sensitivity = P(predict positive | actual positive)

$Sensitivity = \frac{TP}{TP+FN}$

$Sensitivity = \frac{30}{30+20} = 0.6$

6. You have a model such that the lowest score for a positive example is higher than the maximum score for a negative example. What is its ROC AUC? **1 / 1 point**

HINT 1: watch the video "Varying the threshold".

HINT 2: draw a number line and choose values for the score that is the lowest prediction for any positive example, and choose another number that is the score for the highest prediction for any negative example. Draw a few circles for "positive" examples and a few "x" for the negative examples. What do you notice about the model's ability to identify positive and negative examples?

○ 0.52

○ Not enough information is given

○ 0.82

◉ 1.0

Correct

The model perfectly discriminates between positive and negative examples.

Pretend that the score predictions for all positive examples is 0.5 or higher, and the score predictions for all the negative examples are less than 0.5. Then all the positive examples have prediction scores of 0.5 or higher. All the negative examples have prediction scores less than 0.5. They are perfectly separated.

For any thresholds > 0.5, the specificity will be 1.0 (it correctly identifies all the negative examples), and the sensitivity will range from 0 to 1, so the points will run along the line y=1 (in the plot of the ROC curve, it will be the top horizontal edge of the chart.

At the threshold 0.5, the sensitivity (ability to correctly identify positive examples) will be 1.0 and the specificity will also be 1.0, so the point will be at the top right corner of the ROC curve.

At any threshold < 0.5, the sensitivity (ability to identify positive examples) will be 1.0 and the specificity will range from 1 to 0, so the point will be along the line x = 1 (the right side edge of the ROC Curve chart.

So the ROC curve is a box with width 1 and height 1, so the area under it is 1.0.

7. For every specificity, as we vary the threshold, the sensitivity of model 1 is at least as high as model 2. Which of the following must be true? **1 / 1 point**

○ None of the above

○ The accuracy of model 2 is higher than model 1

○ The ROC of model 2 is higher than model 1

◉ The ROC of model 1 is at least as high as model 2

Correct

Note that because specificity determines the x-axis location, and since the sensitivity of model 1 is at least as high as the sensitivity of model 2, the ROC curve for model 1 never goes underneath the curve for model 2. Therefore if we compute the area under the two curves, the area for model 1 must be at least as high as model 2.

8. You want to measure the proportion of people with high blood pressure in a population. You sample 1000 people and find that 55% have high blood pressure with a 90% confidence interval of (50%, 60%). What is the correct interpretation of this result? **1 / 1 point**

HINT: Please watch the video "Confidence interval" to help you answer this question.

◉ If you repeated this sampling, the true proportion would be in the confidence interval about 90% of the time

○ There is a 5% chance that the true mean is less than 50%

○ With 90% probability, the proportion of people with high blood pressure is between 50% and 60%

○ If we repeated this sampling, the middle of the confidence interval would be 55%, 90% of the time

Correct

Confidence intervals are created so that 90% of the time you repeat the experiment, the interval will contain the true parameter value.

9. One experiment calculates a confidence interval using 1000 samples, and the another computes it using 10000 samples. Which interval do you expect to be tighter (assume they use the normal approximation)?

1 / 1 point

- ○ Cannot say with confidence
- ◉ 10,000 samples
- ○ 1,000 samples
- ○ Not enough information

Correct

When we're using a normal approximation, the width of our confidence interval depends on the variance of the normal distribution. Recall that the variance of each sample is identical, but the variance of the average is divided by n. Therefore since dividing by a larger number makes a quantity smaller, the variance of the average of 10000 samples should be less than that for 1000 samples, so the second confidence interval should be tighter.