Week 3 Quiz

TOTAL POINTS 8

1. Let $f(x)$ be the probability that a person with feature x dies within 5 years.

   1 / 1 point

   Let $S_x(t)$ be the survival function of a person with feature x. Assume t is measured in years.

   Which of the following is true?

   - ○ $f(x) = S_x(5)$
   - ○ f(x) = S_x(0)
   - ◉ f(x) = 1-S_x(5)

   Correct
   Recall that S(t) is the probability that you live at least t years or more. Therefore, $S_x 5$ is the probability that you live past 5 years.

   $f(x)$ is the complement of that (probability of dying within 5 years). So it is 1 - S_x(5).

2. The survival function is always:

   1 / 1 point

   - ◉ Decreasing
   - ○ Increasing
   - ○ Linear

   Correct
   The survival function is always decreasing. As time moves forward, it is less likely that you live for longer.

3. Which of the following is a difference between survival data and classification datasets?

1 / 1 point

- ○ Survival data can be used to build prognostic models
- ◉ In survival data the labels are amounts of time and in classification data the labels are binary
- ○ Classification dataset contain information on other features

✓ **Correct**

Both survival data and classification data can be used to build prognostic models (we did this last week!).

Both types of data can contain feature information.

Survival data includes time, and is therefore not binary, unlike classification datasets.

4. Which of the following is an example of censoring?

0.667 / 1 point

- [x] Patient does not have the event by the end of the study period.

✓ **Correct**

If a patient does not have the event by the time the study ends, that is an example of right censoring.

Are the other options examples of right censoring?

- [x] The patient withdraws from a study before having an event, and before the study ends.

✓ **Correct**

If a patient withdraws from a study before the study ends, their data is right censored.

Are the other options examples of right censoring?

- [ ] Death due to other, unrelated causes (such as an automobile accident)

You didn't select all the correct answers

5. Estimate $P(T > 2|T >= 2)$ from the following dataset:

1 / 1 point

i	T_i
1	3
2	5
3	4+
4	2

Hint: $P(T > 2|T >= 2) = (1 - P(T = 2|T >= 2))$.

- 1/2
- 0
- 3/4
- 1/4

Correct

Recall that $P(T > 2|T >= 2) = (1 - P(T = 2|T >= 2))$.

There are 4 individuals who we know live at least 2 years, and 1 of them dies at 2 years. The

6. Compute the probability of surviving up to 4 years $S(4)$ given the following dataset using the Kaplan Meier estimate:

1 / 1 point

i	T_i
1	3
2	5
3	4+
4	2

The Kaplan Meier Estimator is

$S(t) = \prod_{i=0}^{N} \left(1 - \frac{d_i}{n_i}\right)$

- 1/2
- 1/4
- 0
- 3/4

Correct

We write out the formula:

7. Compute S(5) given the following dataset using the Kaplan Meier estimate (note, it's the same dataset as in the previous question).

1 / 1 point

i	T_i
1	3
2	5
3	4+
4	2

The Kaplan Meier Estimator is

$S(t) = \prod_{i=0}^{N} \left(1 - \frac{d_i}{n_i}\right)$

Hint: since we're using the same dataset as in the previous question, you may notice that

$S(5) = S(4) \times (\frac{d_5}{n_5})$

- (selected) 0
- 3/4
- 1/2
- 1/4

Correct
$S(5) = (1 - P(T = 2|T \geq 2)) \times (1 - P(T = 3|T \geq 3)) \times (1 - P(T = 4|T \geq 4)$

8. True or False: If t is larger than the longest survival time recorded in the dataset, then $S(t) = 0$ according to the Kaplan-Meier estimate.

1 / 1 point

The Kaplan Meier Estimator is

$S(t) = \prod_{i=0}^{N} \left(1 - \frac{d_i}{n_i}\right)$

- (selected) False
- True

Correct
This is true only if the last observation is not censored. If the last observation is censored, and if all the other the terms in the Kaplan-Meier estimate are greater than 0, then $S(t)$ will be greater than 0 as well.