

Quiz 10 — 4/25/2024

Instructions

This take-home quiz is due on **Thursday, April 25 at 23:59**.

You may use your own course materials, as well as any materials directly linked from the course website. **No collaboration allowed.**

Type your answers **directly in this Jupyter notebook**, and submit this notebook (just the `ipynb` file) using the submission form on the [course website](#).

Problem 1

The *Titanic* was a British luxury oceanliner that sank famously in the icy North Atlantic Ocean on its maiden voyage in April 1912. Of the approximately 2200 passengers on board, 1500 died. The high death rate was blamed largely on the inadequate supply of lifeboats, a result of the manufacturer's claim that the ship was "unsinkable". A partial dataset of the passenger list was compiled by Philip Hinde in his *Encyclopedia Titanica*. This dataset is named `Titanic` in the `Stat2Data` Library.

Run the cell below to load and preview this data.

```
In [1]: library(Stat2Data)
data(Titanic)
head(Titanic)
```

A data.frame: 6 × 6

	Name	PClass	Age	Sex	Survived	SexCode
	<fct>	<fct>	<dbl>	<fct>	<int>	<int>
1	Allen, Miss Elisabeth Walton	1st	29.00	female	1	1
2	Allison, Miss Helen Loraine	1st	2.00	female	0	1
3	Allison, Mr Hudson Joshua Creighton	1st	30.00	male	0	0
4	Allison, Mrs Hudson JC (Bessie Waldo Daniels)	1st	25.00	female	0	1
5	Allison, Master Hudson Trevor	1st	0.92	male	1	0
6	Anderson, Mr Harry	1st	47.00	male	1	0

a.

The dataset contains a variable named *Survived*, which is equal to 1 if the passenger survived, and 0 otherwise. It also contains a variable named *Age*, the passenger's age in years.

You are interested in the relationship between survival and the passenger's age. Using R, fit a suitable regression model, using *Survival* as the response variable. Provide **only** the summary output.

```
In [ ]:
```

Feedback. See Lesson 24 Part 1 for examples of how to fit logistic regression models in R.

b.

Is there a statistically significant relationship between survival and the passenger's age?

Perform an appropriate hypothesis test to answer this question:

- State the name of the test.
- Perform all four steps of the test, including an answer to the question above.

Use a significance level of 0.05.

Write your answer here. Double-click to edit.

Feedback. See Example 1 in Lesson 26 for a similar example.

c.

Use the code cell below to compute the odds ratio of survival corresponding to a **10-year increase** in the passenger's age.

Use the Markdown cell below to interpret the odds ratio.

Hint. In R, e^x is `exp(x)`.

In []:

Write your answer here. Double-click to edit.

Feedback. See Example 3e in Lesson 24 Part 1 for a similar example.

Some of you struggled when writing your interpretations of the odds ratio for this problem – it is a bit tricky in this case to get something that is accurate yet concise. Here's one way you could approach it:

The odds of survival *change by a factor of (insert number here)* with a 10-year increase in passenger age.

d.

Give a 95% confidence interval on the odds ratio of survival corresponding to a **1-year increase** in the passenger's age.

Use the code cell below to write R code that shows your work. Report your final answer in the Markdown cell below.

Hint. Remember from Lesson 26 that you can find the critical value with the `qnorm()` function.

In []:

Write your answer here. Double-click to edit.

Feedback. See Example 1 in Lesson 26 for a similar example.

Grading rubric

Problem	Weight
1a	1.0
1b	1.0
1c	1.0
1d	1.0
Max Score	40