**Esophageal cancer and alcohol/tobacco use, part 1**

Case-control studies help determine whether certain exposures are associated with outcomes such as developing cancer. The built-in dataset **esoph** contains data from a case-control study in France comparing people with esophageal cancer (cases, counted in **ncases**) to people without esophageal cancer (controls, counted in **ncontrols**) that are carefully matched on a variety of demographic and medical characteristics. The study compares alcohol intake in grams per day (**alcgp**) and tobacco intake in grams per day (**tobgp**) across cases and controls grouped by age range (**agegp**).

The dataset is available in base R and can be called with the variable name **esoph**:

head(esoph)

You will be using this dataset to answer the following four multi-part questions (Questions 3-6).

You may wish to use the **tidyverse** package:

library(tidyverse)

The following three parts have you explore some basic characteristics of the dataset.

Each row contains one group of the experiment. Each group has a different combination of age, alcohol consumption, and tobacco consumption. The number of cancer cases and number of controls (individuals without cancer) are reported for each group.

**Question 3a**

1.0/1.0 point (graded)

How many groups are in the study? correct

88 Loading

You have used 1 of 10 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 3b**

1.0/1.0 point (graded)

How many cases are there?

Save this value as all\_cases for later problems.

correct

200 Loading

You have used 1 of 10 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 3c**

1.0/1.0 point (graded)

How many controls are there?

Save this value as all\_controls for later problems. Remember from the instructions that controls are individuals without cancer.

correct

975 Loading

You have used 1 of 10 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

The following four parts ask you to explore some probabilities within this dataset related to alcohol and tobacco consumption.

**Question 4a**

1.0/1.0 point (graded)

What is the probability that a subject in the highest alcohol consumption group is a cancer case?

Remember that the total number of individuals in the study includes people with cancer (cases) and people without cancer (controls), so you must add both values together to get the denominator for your probability.

correct

0.402 Loading

You have used 1 of 10 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 4b**

1.0/1.0 point (graded)

What is the probability that a subject in the lowest alcohol consumption group is a cancer case? correct

0.0653 Loading

You have used 1 of 10 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 4c**

1.0/1.0 point (graded)

Given that a person is a case, what is the probability that they smoke 10g or more a day? correct

0.61 Loading

You have used 1 of 10 attempts Some problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

**Question 4d**

1.0/1.0 point (graded)

Given that a person is a control, what is the probability that they smoke 10g or more a day? correct

0.462 Loading

You have used 1 of 10 attempts