Questions 3 and 4: 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 a combination of age group, alcohol consumption group, and tobacco consumption group, followed by the number of cancer cases and number of controls for study subjects in that combined grouping.

Question 3a

1.0/1.0 point (graded)

How many groups are in the study?

88 ✔ **Answer:** 88

88

Explanation
You can find the number of groups using `nrow(esoph)`.

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Question 3b

1.0/1.0 point (graded)
How many cases are there?
Save this value as `all_cases` for later problems.

200 ✔ **Answer:** 200

200

Explanation
You can find the number of cases using this code:

```
all_cases <- sum(esoph$ncases)
all_cases
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Question 3c

1.0/1.0 point (graded)
How many controls are there?
Save this value as `all_controls` for later problems.

975 ✔ **Answer:** 975

975

Explanation
You can find the number of controls using this code:

```
all_controls <- sum(esoph$ncontrols)
all_controls
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

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

Question 4a

0.0/1.0 point (graded)
What is the probability that a subject in the highest alcohol consumption group is a cancer case?

0.918 ✖ **Answer:** 0.402

0.918

Explanation
You can find the probability using this code:

```
esoph %>%
  filter(alcgp == "120+") %>%
  summarize(ncases = sum(ncases), ncontrols = sum(ncontrols)) %>%
  mutate(p_case = ncases / (ncases + ncontrols)) %>%
  pull(p_case)
```

Submit You have used 10 of 10 attempts

Answers are displayed within the problem

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?

0.065 ✔ **Answer:** 0.0653

0.065

Explanation
You can find the probability using this code:

```
esoph %>%
  filter(alcgp == "0-39g/day") %>%
  summarize(ncases = sum(ncases), ncontrols = sum(ncontrols)) %>%
  mutate(p_case = ncases / (ncases + ncontrols)) %>%
  pull(p_case)
```

Submit You have used 5 of 10 attempts

Answers are displayed within the problem

Question 4c

0.0/1.0 point (graded)

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

0.366 ✖ **Answer:** 0.61

0.366

Explanation

You can find the probability using this code:

```
tob_cases <- esoph %>%
  filter(tobgp != "0-9g/day") %>%
  pull(ncases) %>%
  sum()

tob_cases/all_cases
```

Submit You have used 10 of 10 attempts

Answers are displayed within the problem

Question 4d

1/1 point (graded)

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

0.461 ✔ **Answer:** 0.462

0.461

Explanation

You can find the probability using this code:

```
tob_controls <- esoph %>%
  filter(tobgp != "0-9g/day") %>%
  pull(ncontrols) %>%
  sum()

tob_controls/all_controls
```

Submit

You have used 1 of 10 attempts

Answers are displayed within the problem

© All Rights Reserved