Quiz 7

1. (Short answer) Suppose Charlie and Taylor decide to compete in a two-person foot race. The odds of Charlie beating Taylor in the foot race are 9 to 1 (odds = 9). Compute the **probability** that *Taylor* wins the foot race.
2. (Short answer) I fit a simple linear regression model with coronary heart disease (binary 1 = yes / 0 = no) as my outcome, and a binary indicator for arcus senilis as my predictor. The results of my regression are below.

Text

Description automatically generated

From the output provided, calculate the estimated probability of having coronary heart disease among individuals with arcus senilis.

1. (Multiple response) Suppose I am interested in the association between cholesterol and coronary heart disease. I know that having increased weight can lead to increased cholesterol (a causal relationship) and I know that higher weights can cause someone to be more at risk for coronary heart disease (again, a causal relationship). I determine that weight is a potential confounding variable, and so I decide to include it in my model. I fit a multiple logistic regression model in R and obtain the following results:

Text

Description automatically generated

Which of the following is an appropriate interpretation of the coefficient that corresponds to my scientific question? Select all that apply.

* We estimate that the odds of coronary heart disease are 1.0010256 times as large for those with one mg/dL higher cholesterol, for individuals of the same weight.
* We estimate that the odds of coronary heart disease are 1.0010256 times as large for those with one mg/dL higher cholesterol, for individuals who weigh 0 pounds.
* We estimate that the log odds of coronary heart disease are 0.0010251 times as large for those with one mg/dL higher cholesterol, for individuals of the same weight.
* We estimate that the difference in log odds of coronary heart disease is 0.0010251, comparing two groups of individuals of the same weight who differ by one mg/dL in cholesterol, with the group with higher cholesterol having higher log odds.
* We estimate that the difference in log odds of coronary heart disease is 0.0010251, comparing two groups of individuals who weigh 0 pounds and who differ by one mg/dL in cholesterol, with the group with higher cholesterol having higher log odds.

(Everyone given points for this question because code uses lm() instead of glm())

1. (Multiple choice) Suppose I am interested in whether or not frequent rock climbing is causally related to the number of pull-ups that a person can do. I plan to fit a linear regression model with number of pull-ups a person can do as the outcome, and number of times per week someone goes rock climbing as the predictor of interest. I also collect information on a binary variable, denoting whether or not a person does yoga. It is possible that someone who rock climbs would do yoga for injury prevention or to improve their flexibility, and therefore improve their climbing. Additionally, yoga is known to increase core strength, which may in turn allow someone to do more pull-ups.

What role does the binary indicator for whether someone does yoga play in the relationships between our predictor of interest and our outcome?

1. Confounder
2. Precision variable
3. Neither a confounder nor a precision variable
4. (Short answer) Suppose I am interested in predicting the total amount of time an animal sleeps (in hours) based on the following variables:

* vore (categorical: carnivore, omnivore, herbivore, insectivore)
* sleep\_cycle (length of sleep cycle, in hours)
* brainwt (brain weight in kilograms)

I fit a multiple linear regression model in R to address this question and get the following model:

What is the estimated average total sleep time in hours for an animal with a sleep cycle of 1 hour, that is a carnivore, and has a brain that weighs 0.01 kg?

1. (Multiple choice) Your friend is reading a scientific article that states, "The odds of completing your doctorate degree are 2.3 times higher for people who have dogs than for people who have cats (95% CI: (0.9, 5.2))." Which of the following is correct?
2. At a significance level of 0.05, we cannot reject the null hypothesis that the odds of completing your doctorate degree are the same for cat and dog owners.
3. At a significance level of 0.05, we can reject the null hypothesis that the odds of completing your doctorate degree are the same for cat and dog owners.
4. (Short answer) Suppose I am interested in the association between playing video games as a child and playing video games as an adult. I take a random sample of adults and ask them two questions: (1) Did you play video games as a child? (Binary, Yes/No), and (2) Do you play video games as an adult? (Binary, Yes/No)

I summarize my data in the following table:

|  |  |  |
| --- | --- | --- |
|  | Video Games as Child: Yes | Video Games as Child: No |
| Video Games as Adult: Yes | 10 | 0 |
| Video Games as Adult: No | 7 | 15 |

For each of the following two questions, either calculate the probability/odds and show your work, or explain why you cannot calculate the probability/odds:

What is the probability of not playing video games as an adult, for individuals who did not play video games as children?

What are the odds of not playing video games as an adult, for individuals who did not play video games as children?

1. (Short answer) Suppose I am interested in predicting the probability that Star Wars characters are human based on their age in years. I fit both a linear regression model and a logistic regression model with binary human status as my outcome, and age as a single quantitative predictor. I plot my data in a scatterplot, and include both fitted regression lines on my plot below, with the blue line being from the linear regression and the red line being form the logistic regression.

Chart, line chart

Description automatically generated

The character Yoda is recorded as 896 years old (the oldest character in my dataset). Should I use the linear regression model or logistic regression model for predicting the probability that characters are human based on their age? Justify your answer.

1. (True/False) Adjusted R2 will always increase by adding additional covariates to your model.
2. (Multiple choice) Suppose I am interested in the association between breast cancer (yes/no) and family history of breast cancer (yes/no). Specifically, I am interested in whether the probability of having breast cancer is different for individuals with a family history of breast cancer vs. those without. Which type of regression analysis will answer my statistical question using a single regression coefficient?
3. Linear regression
4. Logistic regression