

# AE 09: Probabilities, Odds, Odds ratios

2025-03-18

```
library(tidyverse)
library(knitr)
library(tidymodels)
```

## Introduction

This data comes from the 2023 [Pew Research Center's American Trends Panel](#). The survey aims to capture public opinion about a variety of topics including politics, religion, and technology, among others. We will use data from 11201 respondents in Wave 132 of the survey conducted July 31 - August 6, 2023.

A more complete analysis on this topic can be found in the Pew Research Center article [Growing public concern about the role of artificial intelligence in daily life](#) by Alec Tyson and Emma Kikuchi.

**The goal of this analysis is to understand the relationship between how much an respondent has heard about artificial intelligence (AI) and how concerned they are about increased use of AI in daily life.**

You will use the following variables:

- **ai\_heard** : Response to the question “How much have you heard or read about AI?”
  - A lot
  - A little
  - Nothing at all
  - Refused
- **ai\_concern**: Whether a respondent said they are “more concerned than excited” about in the increased use of AI in daily life (1: yes, 0: no)

```
pew_data <- pew_data |>
  mutate(ai_concern = if_else(CNCEXC_W132 == 2, 1, 0),
         ai_heard = case_when(AI_HEARD_W132 == 1 ~ "A lot",
                              AI_HEARD_W132 == 2 ~ "A little",
                              AI_HEARD_W132 == 3 ~ "Nothing at all",
                              TRUE ~ "Refused"
         ))

# Make factors and relevel
pew_data <- pew_data |>
  mutate(ai_concern = factor(ai_concern),
         ai_heard = factor(ai_heard, levels = c("A lot", "A little",
                                                "Nothing at all", "Refused"))
  )
```

#### Exercise 1

- What is the probability a randomly selected respondent has heard a lot about AI?
- What are the odds a randomly selected respondent has heard a lot about AI?

```
# add code here
```

#### Exercise 2

- What is the probability a randomly selected respondent who is concerned about increased use of AI in daily life has heard a lot about AI?
- What are the odds a randomly selected respondent who is concerned about increased use of AI in daily life has heard a lot about AI?

```
# add code here
```

#### Exercise 3

Make a plot to visualize the relationship between how much a respondent has heard about AI and being concerned with increased use of AI in daily life. Use the plot to describe the relationship between the two variables.

```
# add code here
```

#### Exercise 4

- How do the odds of being concerned about increased use of AI in daily life for a randomly selected respondent who has heard nothing about AI compare to the odds for a randomly selected respondent who has heard a lot about AI?
- How do the odds of being concerned about increased use of AI in daily life for a randomly selected respondent who has heard a little about AI compare to the odds for a randomly selected respondent who has heard a lot about AI?

```
# add code here
```

#### Exercise 5

We can use a logistic regression model to understand the relationship between how much someone has heard about AI and whether they are concerned about increased use of AI in daily life. (We will discuss this in detail next class, but will get a preview for now.)

Let  $p$  be the probability a randomly selected respondent is concerned about increased use of AI in daily life. The statistical model is

$$\begin{aligned}\log\left(\frac{p_i}{1-p_i}\right) &= \beta_0 + \beta_1 \mathbf{1}(ai\_heard_i = \text{A little}) \\ &\quad + \beta_2 \mathbf{1}(ai\_heard_i = \text{Nothing}) \\ &\quad + \beta_3 \mathbf{1}(ai\_heard_i = \text{Refused})\end{aligned}$$

The code and output to fit this model is shown below:

```
ai_concern_fit <- glm(ai_concern ~ ai_heard, data = pew_data,
                      family = "binomial")
tidy(ai_concern_fit) |>
  kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.065	0.031	-2.082	0.037
ai_heardA little	0.383	0.040	9.504	0.000
ai_heardNothing at all	-0.276	0.079	-3.505	0.000
ai_heardRefused	-0.697	0.459	-1.520	0.129

- Interpret the intercept in the context of the data in terms of the log-odds of being concerned about increased use of AI in daily life.
- Interpret the coefficient of `ai_heardA little` in the context of the data in terms of the log-odds of being concerned about increased use of AI in daily life.
- Interpret the coefficient of `ai_heardNothing at all` in the context of the data in terms of the odds of being concerned about the increased use of AI in daily life. How does this compare to your response to Exercise 4?

[add response here]

#### ! Submission

To submit the AE:

Render the document to produce the PDF with all of your work from today's class.

Push all your work to your AE repo on GitHub. You're done!