Models for Binary Outcomes

Logistic regression and variations

Contents

1 Models for Binary Outcomes

Configuring R

Functions from these packages will be used throughout this document:

```r
library(conflicted) # check for conflicting function definitions
# library(printr) # inserts help-file output into markdown output
library(rmarkdown) # Convert R Markdown documents into a variety of formats.
library(pander) # format tables for markdown
library(ggplot2) # graphics
library(ggfortify) # help with graphics
library(dplyr) # manipulate data
library(tibble) # `tibble`s extend `data.frame`s
library(magrittr) # `%>%` and other additional piping tools
library(haven) # import Stata files
library(knitr) # format R output for markdown
library(tidyr) # Tools to help to create tidy data
library(plotly) # interactive graphics
library(dobson) # datasets from Dobson and Barnett 2018
library(parameters) # format model output tables for markdown
library(haven) # import Stata files
```

```r
library(latex2exp) # use LaTeX in R code (for figures and tables)
library(fs) # filesystem path manipulations
library(survival) # survival analysis
library(survminer) # survival analysis graphics
library(KMsurv) # datasets from Klein and Moeschberger
library(parameters) # format model output tables for
library(webshot2) # convert interactive content to static for pdf
library(forcats) # functions for categorical variables ("factors")
library(stringr) # functions for dealing with strings
library(lubridate) # functions for dealing with dates and times
```

Here are some R settings I use in this document:

```r
rm(list = ls()) # delete any data that's already loaded into R

conflicts_prefer(dplyr::filter)
ggplot2::theme_set(
  ggplot2::theme_bw() +
        # ggplot2::labs(col = "") +
    ggplot2::theme(
      legend.position = "bottom",
      text = ggplot2::element_text(size = 12, family = "serif")))

knitr::opts_chunk$set(message = FALSE)
options('digits' = 6)

panderOptions("big.mark", ",")
pander::panderOptions("table.emphasize.rownames", FALSE)
pander::panderOptions("table.split.table", Inf)
conflicts_prefer(dplyr::filter) # use the `filter()` function from dplyr() by default
legend_text_size = 9
run_graphs = TRUE
```

```r
options(digits = 6)
```

Acknowledgements

This content is adapted from:

- Dobson and Barnett (2018), Chapter 7
- Vittinghoff et al. (2012), Chapter 5
- David Rocke[1]'s materials from the 2021 edition of Epi 204[2]
- Nahhas (2024) Chapter 6[3]

1.1 Introduction

—————————————————————

Exercise 1.1. What is logistic regression?

—————————————————————

Solution 1.1.

Definition 1.1. **Logistic regression** is a framework for modeling binary[4] outcomes, conditional on one or more *predictors* (a.k.a. *covariates*).

———————————————

[1]https://dmrocke.ucdavis.edu/
[2]https://dmrocke.ucdavis.edu/Class/EPI204-Spring-2021/EPI204-Spring-2021.html
[3]https://www.bookdown.org/rwnahhas/RMPH/blr.html
[4]probability.qmd#def-binary

Exercise 1.2 (Examples of binary outcomes)**.** What are some examples of binary outcomes in the health sciences?

Solution 1.2. Examples of binary outcomes include:

- exposure (exposed vs unexposed)
- disease (diseased vs healthy)
- recovery (recovered vs unrecovered)
- relapse (relapse vs remission)
- return to hospital (returned vs not)
- vital status (dead vs alive)

Logistic regression uses the Bernoulli[5] distribution to model the outcome variable, conditional on one or more covariates.

Exercise 1.3. Write down a mathematical definition of the Bernoulli distribution.

Solution 1.3. The **Bernoulli distribution** family for a random variable X is defined as:

$$\Pr(X = x) = 1_{x \in \{0,1\}} \pi^x (1 - \pi)^{1-x}$$
$$= \begin{cases} \pi, x = 1 \\ 1 - \pi, x = 0 \end{cases}$$

1.1.1 Logistic regression versus linear regression

Logistic regression differs from linear regression, which uses the Gaussian ("normal") distribution to model the outcome variable, conditional on the covariates.

Exercise 1.4. Recall: what kinds of outcomes is linear regression used for?

Solution 1.4. Linear regression is typically used for numerical outcomes that aren't event counts or waiting times for an event.

Examples of outcomes that are often analyzed using linear regression include:

- weight
- height
- income
- prices

1.2 Risk estimation and prediction

In Epi 203, you have already seen methods for modeling binary outcomes using one covariate that is also binary (such as exposure/non-exposure). In this section, we review one-covariate analyses, with a special focus on risk ratios and odds ratios, which are important concepts for interpreting logistic regression.

Example 1.1 (Oral Contraceptive Use and Heart Attack)**.**

---

[5][probability.qmd#def-bernoulli](probability.qmd#def-bernoulli)

- Research question: how does oral contraceptive (OC) use affect the risk of myocardial infarction (MI; a.k.a. heart attack)?

This was an issue when oral contraceptives were first developed, because the original formulations used higher concentrations of hormones. Modern OCs don't have this issue.

Table 1 contains simulated data for an imaginary follow-up (a.k.a. *prospective*) study in which two groups are identified, one using OCs and another not using OCs, and both groups are tracked for three years to determine how many in each groups have MIs.

```r
library(dplyr)
oc_mi <-
  tribble(
    ~OC, ~MI, ~Total,
    "OC use", 13, 5000,
    "No OC use", 7, 10000
  ) |>
  mutate(`No MI` = Total - MI) |>
  relocate(`No MI`, .after = MI)

totals <-
  oc_mi |>
  summarize(across(c(MI, `No MI`, Total), sum)) |>
  mutate(OC = "Total")

tbl_oc_mi <- bind_rows(oc_mi, totals)

tbl_oc_mi |> pander::pander()
```

Table 1: Simulated data from study of oral contraceptive use and heart attack risk

OC	MI	No MI	Total
OC use	13	4,987	5,000
No OC use	7	9,993	10,000
Total	20	14,980	15,000

Exercise 1.5. Estimate the probabilities of MI for OC users and non-OC users in Example 1.1.

Solution 1.5.

$$\hat{p}(MI|OC) = \frac{13}{5000} = 0.0026$$

$$\hat{p}(MI|\neg OC) = \frac{7}{10000} = 7 \times 10^{-4}$$

Exercise 1.6. What does the term "controls" mean in the context of study design?

Solution 1.6.

Definition 1.2 (Two meanings of "controls"). Depending on context, "controls" can mean either:

- individuals who don't experience an *exposure* of interest,
- or individuals who don't experience an *outcome* of interest.

Exercise 1.7. What types of studies do the two definitions of controls correspond to?

Solution 1.7.

Definition 1.3 (cases and controls in retrospective studies). In *retrospective case-control studies*, participants who experience the outcome of interest are called **cases**, while participants who don't experience that outcome are called **controls**.

Definition 1.4 (treatment groups and control groups in prospective studies). In *prospective studies*, the group of participants who experience the treatment or exposure of interest is called the **treatment group**, while the participants who receive the baseline or comparison treatment (for example, clinical trial participants who receive a placebo or a standard-of-care treatment rather than an experimental treatment) are called **controls**.

1.3 Comparing probabilities

1.3.1 Risk differences

The simplest comparison of two probabilities, π_1, and π_2, is the difference of their values:

Definition 1.5 (Risk difference). The **risk difference** of two probabilities, π_1, and π_2, is the difference of their values:

$$\delta(\pi_1, \pi_2) \stackrel{\text{def}}{=} \pi_1 - \pi_2$$

Example 1.2 (Difference in MI risk). In Example 1.1, the maximum likelihood estimate of the difference in MI risk between OC users and OC non-users is:

$$\begin{aligned}
\hat{\delta}(\pi(OC), \pi(\neg OC)) &= \delta(\hat{\pi}(OC), \hat{\pi}(\neg OC)) \\
&= \hat{\pi}(OC) - \hat{\pi}(\neg OC) \\
&= 0.0026 - 7 \times 10^{-4} \\
&= 0.0019
\end{aligned}$$

Exercise 1.8 (interpreting risk differences). How can we interpret the preceding relative risk estimate in prose?

Solution 1.8 (interpreting risk differences). "The difference in risk of MI between OC users and non-users was 0.0019."

or

"The difference in risk of MI between OC users and non-users was 0.19 percentage points[6]."

See the note about working with percentages in the Appendix[7].

1.3.2 Risk ratios

Exercise 1.9. If π_1 and π_2 are two probabilities, what do we call the following ratio?

$$\rho(\pi_1, \pi_2) = \frac{\pi_1}{\pi_2}$$

Solution 1.9.

---

[6]https://en.wikipedia.org/wiki/Percentage_point
[7]notation.qmd#percent-sign

Definition 1.6 (Relative risk ratios). The ratio of two probabilities π_1 and π_2,

$$\rho(\pi_1, \pi_2) = \frac{\pi_1}{\pi_2}$$

is called the:

- **risk ratio**,
- **relative risk ratio**,
- **probability ratio**,
- or **rate ratio**

of π_1 compared to π_2.

Exercise 1.10.

Above, we estimated that:

$$\hat{p}(MI|OC) = 0.0026$$

$$\hat{p}(MI|\neg OC) = 7 \times 10^{-4}$$

Now, estimate the *relative risk* of MI for OC versus non-OC.

Solution 1.10.

The *relative risk* of MI for OC versus non-OC is:

```r
rr <- (13 / 5000) / (7 / 10000)
```

$$\begin{aligned}
\hat{\rho}(OC, \neg OC) &= \frac{\hat{p}(MI|OC)}{\hat{p}(MI|\neg OC)} \\
&= \frac{0.0026}{7 \times 10^{-4}} \\
&= 3.714286
\end{aligned}$$

Exercise 1.11. How can we interpret the preceding relative risk estimate in prose?

Solution 1.11.

We might summarize this result by saying that:

- "The estimated probability of MI among OC users was 3.714286 <u>times as high as</u> the estimated probability among OC non-users."

or

- "The estimated probability of MI among OC users was 2.714286 <u>times higher than</u>, the estimated probability among OC non-users."

see also Section 8.1.4[8] which uses similar phrasing.

---

[8]https://link.springer.com/chapter/10.1007/978-1-4614-1353-0_8#Sec5_8

1.3.3 Relative risk differences

The second approach above, where we subtract 1 from the risk ratio, is actually reporting a slightly different metric:

Definition 1.7 (Relative risk difference).

Sometimes, we divide the risk difference by the comparison probability; the result is called the **relative risk difference**:

$$\xi(\pi_1, \pi_2) \overset{\text{def}}{=} \frac{\delta(\pi_1, \pi_2)}{\pi_2}$$

Theorem 1.1 (Relative risk difference equals risk ratio minus 1).

$$\xi(\pi_1, \pi_2) = \rho(\pi_1, \pi_2) - 1$$

Proof.

$$\begin{aligned}
\xi(\pi_1, \pi_2) &\overset{\text{def}}{=} \frac{\delta(\pi_1, \pi_2)}{\pi_2} \\
&= \frac{\pi_1 - \pi_2}{\pi_2} \\
&= \frac{\pi_1}{\pi_2} - 1 \\
&= \rho(\pi_1, \pi_2) - 1
\end{aligned}$$

$\square$

1.3.4 Changing reference groups in risk comparisons

Risk differences, risk ratios, and relative risk differences are defined by two probabilities, plus a choice of which probability is the **baseline** or **reference** probability (i.e., which probability is the subtrahend of the risk difference or the denominator of the risk ratio).

$$\delta(\pi_2, \pi_1) = -\delta(\pi_1, \pi_2)$$

$$\rho(\pi_2, \pi_1) = (\rho(\pi_1, \pi_2))^{-1}$$
$$\xi(\pi_2, \pi_1) = (\xi(\pi_1, \pi_2) + 1)^{-1} - 1$$

Exercise 1.12. Prove the relationships above.

Example 1.3 (Switching the reference group in a risk ratio). Above, we estimated that the risk ratio of OC versus non-OC is:

$$\rho(OC, \neg OC) = 3.714286$$

In comparison, the risk ratio for non-OC versus OC is:

$$\rho(\neg OC, OC) = \frac{\hat{\mathrm{p}}(MI|\neg OC)}{\hat{\mathrm{p}}(MI|OC)}$$
$$= \frac{7 \times 10^{-4}}{0.0026}$$
$$= 0.269231$$
$$= \frac{1}{\rho(OC, \neg OC)}$$

1.4 Odds and odds ratios

1.4.1 Odds and probabilities

In logistic regression, we will make use of a mathematically-convenient transformation of probability, called *odds*:

Definition 1.8 (Odds).

The **odds** of an event A, is the probability that the event occurs, divided by the probability that it doesn't occur. We can represent odds with the Greek letter ω ("omega"). [9] Thus, in mathematical notation:

$$\omega \stackrel{\text{def}}{=} \frac{\Pr(A)}{\Pr(\neg A)} \tag{1}$$

This course is about regression models, which are conditional probability models (**?@def-regression-model**). Accordingly, we define conditional odds in terms of conditional probabilities:

Definition 1.9 (Conditional odds).

The **conditional odds** of an event A given a condition B, is the (conditional) probability that event A occurs (given condition B), divided by the (conditional) probability that it doesn't occur (given condition B). We can represent conditional odds using $\omega(A|B)$, $\omega(B)$ or ω_B ("omega bee"). Thus, in mathematical notation:

$$\omega(B) \stackrel{\text{def}}{=} \frac{\Pr(A|B)}{\Pr(\neg A|B)} \tag{2}$$

Example 1.4 (Computing odds from probabilities). In Exercise 1.5, we estimated that the probability of MI, given OC use, is $\pi(OC) \stackrel{\text{def}}{=} \Pr(MI|OC) = 0.0026$. If this estimate is correct, then the odds of MI, given OC use, is:

$$\omega(OC) \stackrel{\text{def}}{=} \frac{\Pr(MI|OC)}{\Pr(\neg MI|OC)}$$
$$= \frac{\Pr(MI|OC)}{1 - \Pr(MI|OC)}$$
$$= \frac{\pi(OC)}{1 - \pi(OC)}$$
$$= \frac{0.0026}{1 - 0.0026}$$
$$\approx 0.002607$$

Exercise 1.13 (Computing odds from probabilities). Estimate the odds of MI, for non-OC users.

---

[9]The name "omega" is a contraction of "o mega", which means "long o" in Greek, in contrast with "omicron" (*o*, "short o"). See https://www.etymonline.com/search?q=omega and https://en.wikipedia.org/wiki/Omega for more details.

Solution.

$$\omega(\neg OC) = 7.004903 \times 10^{-4}$$

Exercise 1.14. Find a general formula for converting probabilities into odds.

Solution 1.12. Using Definition 1.8 and **?@cor-p-neg**:

$$\omega \overset{\text{def}}{=} \frac{\Pr(A)}{\Pr(\neg A)}$$
$$= \frac{\pi}{1 - \pi}$$

Theorem 1.2. *If π is the probability of an event A and ω is the corresponding odds of A, then:*

$$\omega = \frac{\pi}{1 - \pi} \tag{3}$$

Proof. By Solution 1.12. $\qquad\square$

The mathematical relationship between odds ω and probabilities π, which is represented in Equation 3, is a core component of logistic regression models, as we will see in the rest of this chapter. Let's give the expression on the righthand side of Equation 3 its own name and symbol, so that we can refer to it concisely:

Definition 1.10 (Odds function)**.** The **odds function** is defined as:

$$\text{odds}\{\pi\} \overset{\text{def}}{=} \frac{\pi}{1 - \pi} \tag{4}$$

We can use the odds function (Definition 1.10) to simplify Equation 3 (in Theorem 1.2) into a more concise expression, which is easier to remember and manipulate:

Corollary 1.1. *If π is the probability of an outcome A and ω is the corresponding odds of A, then:*

$$\omega = odds\{\pi\} \tag{5}$$

In other words, the odds function rescales probabilities into odds.

Proof. By Theorem 1.2 and Definition 1.10. $\qquad\square$

Exercise 1.15. Graph the odds function.

Solution 1.13.

Figure 1 graphs the odds function.

```
odds <- function(pi) pi / (1 - pi)
library(ggplot2)
ggplot() +
  geom_function(
    fun = odds,
    arrow = arrow(ends = "last"),
```

```
  mapping = aes(col = "odds function")
) +
xlim(0, .99) +
xlab("Probability") +
ylab("Odds") +
geom_abline(aes(
  intercept = 0,
  slope = 1,
  col = "y=x"
)) +
theme_bw() +
labs(colour = "") +
theme(legend.position = "bottom")
```


Figure 1: Odds versus probability

Theorem 1.3 (One-sample MLE for odds)**.** *Let $X_1, ... X_n$ be a set of n iid Bernoulli trials, and let $X = \sum_{i=1}^{n} X_i$ be their sum.*

Then the maximum likelihood estimate of the odds of $X = 1$, ω, is:

$$\hat{\omega} = \frac{x}{n-x}$$

Proof.

$$1 - \hat{\pi} = 1 - \frac{x}{n}$$
$$= \frac{n}{n} - \frac{x}{n}$$
$$= \frac{n-x}{n}$$

Thus, the estimated odds is:

$$\frac{\hat{\pi}}{1 - \hat{\pi}} = \frac{\left(\frac{x}{n}\right)}{\left(\frac{n-x}{n}\right)}$$
$$= \frac{x}{n - x} \tag{6}$$

That is, the odds estimate can be computed directly as "# events" divided by "# nonevents", without needing to compute $\hat{\pi}$ and $1 - \hat{\pi}$ first.

□

Example 1.5 (Calculating odds using the shortcut)**.** In Example 1.4, we calculated

$$\omega(OC) = 0.002607$$

Let's recalculate this result using our shortcut.

Solution 1.14.
$$\omega(OC) = \frac{13}{5000 - 13}$$
$$= 0.002607$$

Same answer as in Example 1.4!

Theorem 1.4 (Simplified expressions for odds function)**.**

Two equivalent expressions for the odds function are:

$$odds\{\pi\} = \frac{1}{\pi^{-1} - 1}$$
$$= \left(\pi^{-1} - 1\right)^{-1} \tag{7}$$

Exercise 1.16. Prove Theorem 1.4.

Solution 1.15. Starting from Definition 1.10:

$$\text{odds}\{\pi\} = \frac{\pi}{1 - \pi}$$
$$= \frac{\pi}{1 - \pi}\frac{\pi^{-1}}{\pi^{-1}}$$
$$= \frac{\pi\pi^{-1}}{(1 - \pi)\pi^{-1}}$$
$$= \frac{1}{(\pi^{-1} - \pi\pi^{-1})}$$
$$= \frac{1}{(\pi^{-1} - 1)}$$
$$= \left(\pi^{-1} - 1\right)^{-1}$$

Corollary 1.2 (Odds of a non-event)**.** *If π is the odds of event A and ω is the corresponding odds of A, then the odds of $\neg A$ are:*

$$\omega(\neg A) = \frac{1 - \pi}{\pi}$$
$$= \pi^{-1} - 1$$

Proof. Left to the reader. □

Odds of rare events

Exercise 1.17. What odds value corresponds to the probability $\pi = 0.2$, and what is the numerical difference between these two values?

Solution.
$$\omega = \frac{\pi}{1 - \pi} = \frac{.2}{.8} = .25$$

Exercise 1.18. Find the difference between an odds ω and its corresponding probability π, as a function of π.

Solution 1.16.
$$\omega - \pi = \frac{\pi}{1 - \pi} - \pi$$
$$= \frac{\pi}{1 - \pi} - \frac{\pi(1 - \pi)}{1 - \pi}$$
$$= \frac{\pi}{1 - \pi} - \frac{\pi - \pi^2}{1 - \pi}$$
$$= \frac{\pi - (\pi - \pi^2)}{1 - \pi}$$
$$= \frac{\pi - \pi + \pi^2}{1 - \pi}$$
$$= \frac{\pi^2}{1 - \pi}$$
$$= \frac{\pi}{1 - \pi}\pi$$
$$= \omega\pi$$

Theorem 1.5. *Let* $\omega = \frac{\pi}{1-\pi}$. *Then:*

$$\omega - \pi = \frac{\pi^2}{1 - \pi}$$

Proof. By Solution 1.16. □

For rare events (small π), odds and probabilities are nearly equal (see Figure 1), because $1 - \pi \approx 1$ and $\pi^2 \approx 0$.

For example, in Example 1.4, the probability and odds differ by 6.777622×10^{-6}.

1.4.2 The inverse odds function

Exercise 1.19. If π is the probability of an event A and ω is the corresponding odds of A, how can we compute π from ω?

For example, if $\omega = 3/2$, what is π?

Solution 1.17. Starting from Theorem 1.2, we can solve Equation 3 for π in terms of ω:

$$\omega = \frac{\pi}{1 - \pi}$$
$$(1 - \pi)\omega = \pi$$
$$\omega - \pi\omega = \pi$$
$$\omega = \pi + \pi\omega$$
$$\omega = (1 + \omega)\pi$$
$$\pi = \frac{\omega}{1 + \omega}$$

So if $\omega = 3/2$,

$$\pi = \frac{3/2}{1 + 3/2}$$
$$= \frac{3/2}{5/2}$$
$$= \frac{3}{5}$$

Theorem 1.6. *If π is the probability of an event and ω is the corresponding odds of that event, then:*

$$\pi = \frac{\omega}{1 + \omega} \tag{8}$$

Proof. By Theorem 1.2 and Solution 1.17. $\qquad\square$

Definition 1.11 (inverse odds function)**.**

$$\text{invodds}\{\omega\} \overset{\text{def}}{=} \frac{\omega}{1 + \omega} \tag{9}$$

can be called the **inverse-odds function**.

Corollary 1.3.
$$\pi = invodds\{\omega\}$$

Proof. By Definition 1.11 and Theorem 1.6. $\qquad\square$

Corollary 1.4.
$$invodds\{\omega\} = odds^{-1}\{\omega\}$$

Proof. Using Corollary 1.1 and Theorem 1.6:

$$\mathrm{invodds}\{\mathrm{odds}\{\pi\}\} = \mathrm{invodds}\{\omega\}$$
$$= \frac{\omega}{1 + \omega}$$
$$= \pi$$

Likewise (not shown):

$$\mathrm{odds}\{\mathrm{invodds}\{\omega\}\} = \omega$$

$\square$

The inverse-odds function converts odds into their corresponding probabilities (Figure 2). Its domain of inputs is $\omega \in [0, \infty)$ and its range of outputs is $\pi \in [0, 1]$.

I haven't seen anyone give the inverse-odds function a concise name; maybe prob() or prob() or risk()?

```r
odds_inv <- function(omega) (1 + omega^-1)^-1
library(ggplot2)
ggplot() +
  geom_function(fun = odds_inv, aes(col = "inverse-odds")) +
  xlab("Odds") +
  ylab("Probability") +
  xlim(0, 5) +
  ylim(0, 1) +
  geom_abline(aes(intercept = 0, slope = 1, col = "x=y"))
```


Figure 2: The inverse odds function, invodds$\{\omega\}$

Exercise 1.20. What probability corresponds to an odds of $\omega = 1$, and what is the numerical difference between these two values?

Solution.

$$\pi = \mathrm{invodds}\{1\}$$
$$= \frac{1}{1+1}$$
$$= \frac{1}{2}$$
$$= .5$$
$$\omega - \pi = 1 - .5$$
$$= .5$$

Lemma 1.1 (Simplified expression for inverse odds function)**.**

Equivalent expressions for the inverse odds function are:

$$\mathrm{invodds}\{\omega\} = \frac{1}{1+\omega^{-1}} \tag{10}$$
$$= (1+\omega^{-1})^{-1}$$

Exercise 1.21. Prove that Equation 10 is equivalent to Definition 1.11.

Solution 1.18. Analogous to Solution 1.15.

Lemma 1.2 (One minus inverse-odds)**.**

$$1 - \pi = \frac{1}{1+\omega}$$

Proof. By Theorem 1.6:

$$1 - \pi = 1 - \frac{\omega}{1+\omega}$$
$$= \frac{1+\omega}{1+\omega} - \frac{\omega}{1+\omega}$$
$$= \frac{(1+\omega)-\omega}{1+\omega}$$
$$= \frac{1+\omega-\omega}{1+\omega}$$
$$= \frac{1}{1+\omega}$$

$\square$

Corollary 1.5.

$$1 + \omega = \frac{1}{1-\pi}$$

1.4.3 Odds ratios

Now that we have defined odds, we can introduce another way of comparing event probabilities: odds ratios.

Definition 1.12 (Odds ratio)**.** The **odds ratio** for two conditional odds, ω_1 and ω_2, is the ratio of those odds:

$$\theta(\omega_1, \omega_2) \stackrel{\text{def}}{=} \frac{\omega_1}{\omega_2}$$

There's a 1:1 mapping between probability and odds, and according to that mapping, the odds are equal between two covariate patterns IF and ONLY IF the probabilities are also equal between those patterns. So, testing whether an odds ratio $= 1$ is equivalent to testing whether the corresponding risk ratio $= 1$, and also equivalent to testing whether the risk difference $= 0$. Therefore, in **hypothesis testing**, if the null hypothesis is no effect, then we can shift between RD, RR, and OR. But when we're talking about **point estimates** and **CIs**, we need to limit our conclusions to the effect measure(s) that we actually estimated, because the *sizes* of RDs, RRs, and ORs don't have a simple relationship to each other, except when pi_1=pi_2 (as shown by Figure 3).

An *odds ratio* is a *ratio of odds*. An odds is a ratio of probabilities, so odds ratios are ratios of ratios:

Theorem 1.7.

$$\theta(\omega_1, \omega_2) = \frac{\omega_1}{\omega_2}$$

$$= \frac{\left(\frac{\pi_1}{1-\pi_1}\right)}{\left(\frac{\pi_2}{1-\pi_2}\right)}$$

Example 1.6 (Calculating odds ratios)**.** In Example 1.1, the odds ratio for OC users versus OC-non-users is:

$$\theta(\omega(OC), \omega(\neg OC)) = \frac{\omega(OC)}{\omega(\neg OC)}$$

$$= \frac{0.0026}{7 \times 10^{-4}}$$

$$= 3.714286$$

A shortcut for calculating odds ratio estimates

The general form of a two-by-two table is shown in Table 2.

Table 2: A generic 2x2 table

	Event	Non-Event	Total
Exposed	a	b	a+b
Non-exposed	c	d	c+d
Total	a+c	b+d	a+b+c+d

From this table, we have:

- $\hat{\pi}(Event|Exposed) = a/(a + b)$

- $\hat{\pi}(\neg Event|Exposed) = b/(a + b)$

- $\hat{\omega}(Event|Exposed) = \frac{\frac{a}{a+b}}{\frac{b}{a+b}} = \frac{a}{b}$

- $\hat{\omega}(Event|\neg Exposed) = \frac{c}{d}$ (see Exercise 1.22)

- $\theta(Exposed, \neg Exposed) = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$

Exercise 1.22. Given Table 2, show that $\hat{\omega}(Event|\neg Exposed) = \frac{c}{d}$.

Properties of odds ratios

Odds ratios have a special property: we can swap a covariate with the outcome, and the odds ratio remains the same.

Theorem 1.8 (Odds ratios are reversible)**.** *For any two events A, B:*

$$\theta(A|B) = \theta(B|A)$$

Proof.

$$\theta(A|B) \overset{\text{def}}{=} \frac{\omega(A|B)}{\omega(A|\neg B)}$$

$$= \frac{\left(\frac{\mathrm{p}(A|B)}{\mathrm{p}(\neg A|B)}\right)}{\left(\frac{\mathrm{p}(A|\neg B)}{\mathrm{p}(\neg A|\neg B)}\right)}$$

$$= \left(\frac{\mathrm{p}(A|B)}{\mathrm{p}(\neg A|B)}\right)\left(\frac{\mathrm{p}(A|\neg B)}{\mathrm{p}(\neg A|\neg B)}\right)^{-1}$$

$$= \left(\frac{\mathrm{p}(A|B)}{\mathrm{p}(\neg A|B)}\right)\left(\frac{\mathrm{p}(\neg A|\neg B)}{\mathrm{p}(A|\neg B)}\right)$$

$$= \left(\frac{\mathrm{p}(A|B)}{\mathrm{p}(\neg A|B)} \cdot \frac{\mathrm{p}(B)}{\mathrm{p}(B)}\right)\left(\frac{\mathrm{p}(\neg A|\neg B)}{\mathrm{p}(A|\neg B)} \cdot \frac{\mathrm{p}(\neg B)}{\mathrm{p}(\neg B)}\right)$$

$$= \left(\frac{\mathrm{p}(A, B)}{\mathrm{p}(\neg A, B)}\right)\left(\frac{\mathrm{p}(\neg A, \neg B)}{\mathrm{p}(A, \neg B)}\right)$$

$$= \left(\frac{\mathrm{p}(B, A)}{\mathrm{p}(B, \neg A)}\right)\left(\frac{\mathrm{p}(\neg B, \neg A)}{\mathrm{p}(\neg B, A)}\right)$$

$$= \left(\frac{\mathrm{p}(B, A)}{\mathrm{p}(\neg B, A)}\right)\left(\frac{\mathrm{p}(\neg B, \neg A)}{\mathrm{p}(B, \neg A)}\right)$$

$$= [\text{reverse the preceding steps}]$$

$$= \theta(B|A)$$

$\square$

Example 1.7. In Example 1.1, we have:

$$\theta(MI; OC) \overset{\text{def}}{=} \frac{\omega(MI|OC)}{\omega(MI|\neg OC)}$$

$$\overset{\text{def}}{=} \frac{\left(\frac{\Pr(MI|OC)}{\Pr(\neg MI|OC)}\right)}{\left(\frac{\Pr(MI|\neg OC)}{\Pr(\neg MI|\neg OC)}\right)}$$

$$= \frac{\left(\frac{\Pr(MI, OC)}{\Pr(\neg MI, OC)}\right)}{\left(\frac{\Pr(MI, \neg OC)}{\Pr(\neg MI, \neg OC)}\right)}$$

$$= \left(\frac{\Pr(MI, OC)}{\Pr(\neg MI, OC)}\right)\left(\frac{\Pr(\neg MI, \neg OC)}{\Pr(MI, \neg OC)}\right)$$

$$= \left(\frac{\Pr(MI, OC)}{\Pr(MI, \neg OC)}\right)\left(\frac{\Pr(\neg MI, \neg OC)}{\Pr(\neg MI, OC)}\right)$$

$$= \left(\frac{\Pr(OC, MI)}{\Pr(\neg OC, MI)}\right)\left(\frac{\Pr(\neg OC, \neg MI)}{\Pr(OC, \neg MI)}\right)$$

$$= \left(\frac{\Pr(OC|MI)}{\Pr(\neg OC|MI)}\right)\left(\frac{\Pr(\neg OC|\neg MI)}{\Pr(OC|\neg MI)}\right)$$

$$= \frac{\left(\frac{\Pr(OC|MI)}{\Pr(\neg OC|MI)}\right)}{\left(\frac{\Pr(OC|\neg MI)}{\Pr(\neg OC|\neg MI)}\right)}$$

$$\overset{\text{def}}{=} \frac{\omega(OC|MI)}{\omega(OC|\neg MI)}$$

$$\overset{\text{def}}{=} \theta(OC; MI)$$

Exercise 1.23. For Table 2, show that $\hat{\theta}(Exposed, Unexposed) = \hat{\theta}(Event, \neg Event)$.

Conditional odds ratios have the same reversibility property:

Theorem 1.9 (Conditional odds ratios are reversible)**.** *For any three events A, B, C:*

$$\theta(A|B, C) = \theta(B|A, C)$$

Proof. Apply the same steps as for Theorem 1.8, inserting C into the conditions (RHS of |) of every expression. $\square$

Odds Ratios vs Probability (Risk) Ratios

When the outcome is rare (i.e., its probability is small) for both groups being compared in an odds ratio, the odds of the outcome will be similar to the probability of the outcome, and thus the risk ratio will be similar to the odds ratio.

Case 1: rare events

For rare events, odds ratios and probability (a.k.a. risk, a.k.a. prevalence) ratios will be close:

$$\pi_1 = .01$$
$$\pi_2 = .02$$

```
pi1 <- .01
pi2 <- .02
pi2 / pi1
#> [1] 2
odds(pi2) / odds(pi1)
#> [1] 2.02041
```

Example 1.8. In Example 1.1, the outcome is rare for both OC and non-OC participants, so the odds for both groups are similar to the corresponding probabilities, and the odds ratio is similar the risk ratio.

Case 2: frequent events

$$\pi_1 = .4$$

$$\pi_2 = .5$$

For more frequently-occurring outcomes, this won't be the case:

```
pi1 <- .4
pi2 <- .5
pi2 / pi1
#> [1] 1.25
odds(pi2) / odds(pi1)
#> [1] 1.5
```

Figure 3 compares risk differences, risk ratios, and odds ratios as functions of the underlying probabilities being compared.

```r
if (run_graphs) {
  RD <- function(p1, p2) p2 - p1
  RR <- function(p1, p2) p2 / p1
  odds <- function(p) p / (1 - p)
  OR <- function(p1, p2) odds(p2) / odds(p1)
  OR_minus_RR <- function(p1, p2) OR(p2, p1) - RR(p2, p1)

  n_ticks <- 201
  probs <- seq(.001, .99, length.out = n_ticks)
  RD_mat <- outer(probs, probs, RD)
  RR_mat <- outer(probs, probs, RR)
  OR_mat <- outer(probs, probs, OR)

  opacity <- .3
  z_min <- -1
  z_max <- 5
  plotly::plot_ly(
    x = ~probs,
    y = ~probs
  ) |>
    plotly::add_surface(
      z = ~ t(RD_mat),
      contours = list(
        z = list(
          show = TRUE,
          start = -1,
          end = 1,
          size = .1
        )
      ),
      name = "Risk Difference",
      showscale = FALSE,
      opacity = opacity,
      colorscale = list(c(0, 1), c("green", "green"))
    ) |>
    plotly::add_surface(
      opacity = opacity,
      colorscale = list(c(0, 1), c("red", "red")),
      z = ~ t(RR_mat),
      contours = list(
        z = list(
          show = TRUE,
          start = z_min,
          end = z_max,
          size = .2
        )
      ),
      showscale = FALSE,
      name = "Risk Ratio"
    ) |>
    plotly::add_surface(
      opacity = opacity,
      colorscale = list(c(0, 1), c("blue", "blue")),
      z = ~ t(OR_mat),
      contours = list(
        z = list(
          show = TRUE,
          start = z_min,
          end = z_max,
          size = .2
        )
      ),
      showscale = FALSE,
```

21

Odds Ratios in Case-Control Studies

Table 1 simulates a follow-up study in which two populations were followed and the number of MI's was observed. The risks are $P(MI|OC)$ and $P(MI|\neg OC)$ and we can estimate these risks from the data.

But suppose we had a case-control study in which we had 100 women with MI and selected a comparison group of 100 women without MI (matched as groups on age, etc.). Then MI is not random, and we cannot compute P(MI|OC) and we cannot compute the risk ratio. However, the odds ratio however can be computed.

The disease odds ratio is the odds for the disease in the exposed group divided by the odds for the disease in the unexposed group, and we cannot validly compute and use these separate parts.

We can still validly compute and use the exposure odds ratio, which is the odds for exposure in the disease group divided by the odds for exposure in the non-diseased group (because exposure can be treated as random):

$$\hat{\theta}(OC|MI) = \frac{\hat{\omega}(OC|MI)}{\hat{\omega}(OC|\neg MI)}$$

And these two odds ratios, $\hat{\theta}(MI|OC)$ and $\hat{\theta}(OC|MI)$, are mathematically equivalent, as we saw in Section 1.4.3:

$$\hat{\theta}(MI|OC) = \hat{\theta}(OC|MI)$$

Exercise 1.24. Calculate the odds ratio of MI with respect to OC use, assuming that Table 1 comes from a case-control study. Confirm that the result is the same as in Example 1.6.

Solution.

```
tbl_oc_mi |> pander::pander()
```

Table 3: Simulated data from study of oral contraceptive use and heart attack risk

OC	MI	No MI	Total
OC use	13	4,987	5,000
No OC use	7	9,993	10,000
Total	20	14,980	15,000

- $\omega(OC|MI) = P(OC|MI)/(1-P(OC|MI) = \frac{13}{7} = 1.857143$

- $\omega(OC|\neg MI) = P(OC|\neg MI)/(1-P(OC|\neg MI) = \frac{4987}{9993} = 0.499049$

- $\theta(OC, MI) = \frac{\omega(OC|MI)}{\omega(OC|\neg MI)} = \frac{13/7}{4987/9993} = 3.721361$

This is the same estimate we calculated in Example 1.6.

Odds Ratios in Cross-Sectional Studies

- If a cross-sectional study is a uniform probability sample of a population (which it rarely is), then we can estimate prevalence (sometimes called "prevalence risk" or just "risk") using standard methods (Lee 1994), and we can thus also estimate prevalence differences, prevalence ratios, and prevalence odds ratios comparing sub-populations.

- If the cross-sectional study is a stratified probability sample, then we can estimate prevalence, prevalence differences, prevalence ratios, and prevalence odds ratios using specialized methods for complex surveys (Lumley 2010).

- If the study has sampling biases that we cannot adjust for with survey weights, such as in a convenience sample, then we need to treat it in the same way as a case-control study, and we cannot validly estimate prevalence, prevalence differences, or prevalence ratios; we can only validly estimate prevalence *odds* ratios.

1.5   The logit and expit functions

1.5.1   The logit function

Definition 1.13 (log-odds)**.**

If ω is the odds of an event A, then the **log-odds** of A, which we will represent by η ("eta"), is the natural logarithm of the odds of A:

$$\eta \overset{\text{def}}{=} \log\{\omega\} \tag{11}$$

Theorem 1.10. *If π is the probability of an event A, ω is the corresponding odds of A, and η is the corresponding log-odds of A, then:*

$$\eta = log\left\{\frac{\pi}{1-\pi}\right\} \tag{12}$$

Proof. Apply Definition 1.13 and then Theorem 1.2. $\qquad\square$

Definition 1.14 (logit function)**.**

The **logit function** of a probability π is the natural logarithm of the odds function of π:

$$\text{logit}(\pi) \overset{\text{def}}{=} \log\{\text{odds}\{\pi\}\}$$

The **logit function** is a composite function[10].

Exercise 1.25 (Compose the logit function)**.** Mathematically expand the definition of the logit function.

Solution 1.19 (Compose the logit function)*.*

Theorem 1.11 (Expanded expression for logit)*.*

$$logit(\pi) = log\left\{\frac{\pi}{1-\pi}\right\} \tag{13}$$

Proof. Apply Definition 1.14 and then Definition 1.8 (details left to the reader). $\qquad\square$

Corollary 1.6. *If π is the probability of an event A and η is the corresponding log-odds of A, then:*

$$\eta = logit\{\pi\}$$

---

[10]https://en.wikipedia.org/wiki/Function_composition

Proof. Apply Theorem 1.10 and Theorem 1.11. □

Figure 4 shows the shape of the logit() function.

```r
odds <- function(pi) pi / (1 - pi)

logit <- function(p) log(odds(p))

library(ggplot2)
logit_plot <-
  ggplot() +
  geom_function(
    fun = logit,
    arrow = arrow(ends = "both")
  ) +
  xlim(.001, .999) +
  ylab("logit(p)") +
  xlab("p") +
  theme_bw()
print(logit_plot)
```


Figure 4: The logit function
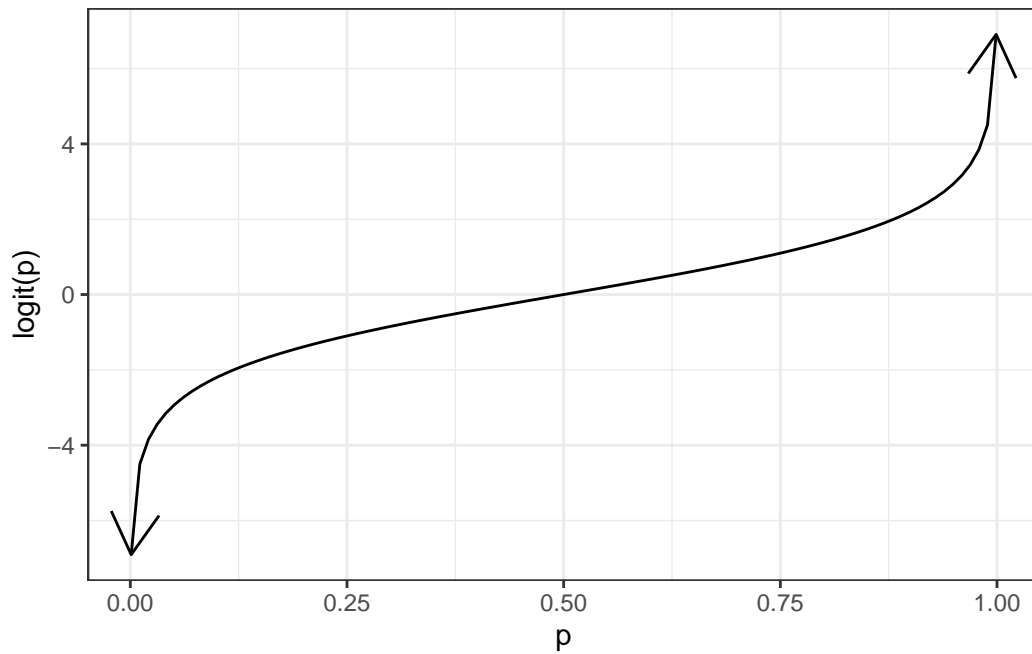
1.5.2   The expit function

Lemma 1.3.

If ω is the odds of an event A and η is the corresponding log-odds of A, then:

$$\omega = exp\{\eta\}$$

Proof. Start from Definition 1.13 and solve for ω. □

24

Theorem 1.12.

If π is the probability of an event A, ω is the corresponding odds of A, and η is the corresponding log-odds of A, then:

$$\pi = \frac{exp\{\eta\}}{1 + exp\{\eta\}}$$

Proof. Apply Theorem 1.6 and then Lemma 1.3. □

Definition 1.15 (expit, logistic, inverse-logit)**.** The **expit function** of a log-odds η, also known as the **inverse-logit function** or **logistic function**, is the inverse-odds of the exponential of η:

$$\text{expit}(\eta) \overset{\text{def}}{=} \text{invodds}\{\exp\{\eta\}\}$$

Theorem 1.13 (Expressions for expit function)**.**

$$
\begin{aligned}
expit(\eta) &= \frac{exp\{\eta\}}{1 + exp\{\eta\}} \\
&= \frac{1}{1 + exp\{-\eta\})} \\
&= (1 + exp\{-\eta\})^{-1}
\end{aligned}
$$

Proof. Apply definitions and Lemma 1.1. Details left to the reader. □

Theorem 1.14. *If π is the probability of an event A, ω is the corresponding odds of A, and η is the corresponding log-odds of A, then:*

$$\pi = expit\{\eta\}$$

Proof. Apply Theorem 1.12 and Theorem 1.13. □

Figure 5 graphs the expit function.

```r
expit <- function(eta) {
  exp(eta) / (1 + exp(eta))
}
library(ggplot2)
expit_plot <-
  ggplot() +
  geom_function(
    fun = expit,
    arrow = arrow(ends = "both")
  ) +
  xlim(-8, 8) +
  ylim(0, 1) +
  ylab(expression(expit(eta))) +
  xlab(expression(eta)) +
```

```
    theme_bw()
print(expit_plot)
```


Figure 5: The expit function

Theorem 1.15 (logit and expit are each others' inverses)**.**

$$logit\{expit\{\eta\}\} = \eta$$

$$expit\{logit\{\pi\}\} = \pi$$

Proof. Left to the reader. $\qquad\square$

1.5.3 Diagram of expit and logit

$$\left[\pi \stackrel{\text{def}}{=} \Pr(Y=1|\tilde{X}=\tilde{x})\right] \underset{\xleftarrow[\frac{\omega}{1+\omega}]{}}{\overset{\frac{\pi}{1-\pi}}{\xrightarrow{\hspace{1cm}}}} \overbrace{\left[\omega \stackrel{\text{def}}{=} \text{odds}(Y=1|\tilde{X}=\tilde{x})\right]}^{\text{logit}(\pi)} \underset{\xleftarrow[\exp\{\eta\}]{}}{\overset{\log\{\omega\}}{\xrightarrow{\hspace{1cm}}}} \left[\eta(\tilde{x}) \stackrel{\text{def}}{=} \text{log-odds}(Y=1|\tilde{X}=\tilde{x})\right]$$
$$\underbrace{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx}}_{\text{expit}(\eta)}$$

Figure 6: Diagram of logistic regression link and inverse link functions

1.6 Introduction to logistic regression

- In Example 1.1, we estimated the risk and the odds of MI for two groups, defined by oral contraceptive use.

26

- If the predictor is quantitative (dose) or there is more than one predictor, the task becomes more difficult.

- In this case, we will use logistic regression, which is a generalization of the linear regression models you have been using that can account for a binary response instead of a continuous one.

1.6.1 Independent binary outcomes - general

Exercise 1.26. Let $\tilde{y}$ represent a data set of mutually independent binary outcomes, each with a potentially different event probability π_i:

$$\tilde{y} = (y_1, ..., y_n)$$
$$y_i \sim_{\perp\!\!\!\perp} \text{Ber}(\pi_i)$$

Write the likelihood of $\tilde{y}$.

Solution 1.20.

$$\pi_i \stackrel{\text{def}}{=} \mathrm{P}(Y_i = 1)$$
$$\mathrm{P}(Y_i = 0) = 1 - \pi_i$$
$$\mathrm{P}(Y_i = y_i) = \mathrm{P}(Y_i = 1)^{y_i} \mathrm{P}(Y_i = 0)^{1-y_i}$$
$$= (\pi_i)^{y_i} (1 - \pi_i)^{1-y_i}$$
$$\mathcal{L}_i(\pi_i) \stackrel{\text{def}}{=} \mathrm{P}(Y_i = y_i)$$
$$\mathcal{L}(\tilde{\pi}) \stackrel{\text{def}}{=} \mathrm{P}(Y_1 = y_1, ..., Y_n = y_n)$$
$$= \prod_{i=1}^{n} \mathrm{P}(Y_i = y_i)$$
$$= \prod_{i=1}^{n} \mathcal{L}_i(\pi_i)$$
$$= \prod_{i=1}^{n} (\pi_i)^{y_i} (1 - \pi_i)^{1-y_i}$$

Exercise 1.27. Write the log-likelihood of $\tilde{y}$.

Solution 1.21.

$$\ell(\tilde{\pi}) \stackrel{\text{def}}{=} \log\{\mathcal{L}(\tilde{\pi})\}$$
$$= \log\left\{ \prod_{i=1}^{n} \mathcal{L}_i(\pi_i) \right\}$$
$$= \sum_{i=1}^{n} \log\{\mathcal{L}_i(\pi_i)\}$$
$$= \sum_{i=1}^{n} \ell_i(\pi_i)$$
$$\ell_i(\pi_i) \stackrel{\text{def}}{=} \log\{\mathcal{L}_i(\pi_i)\}$$
$$= y_i \log\{\pi_i\} + (1 - y_i)\log\{1 - \pi_i\}$$

1.6.2 Modeling π_i as a function of X_i

If there are only a few distinct X_i values, we can model π_i separately for each value of X_i.

Otherwise, we need regression.

Table 4: Mortality rates of adult flour beetles after five hours' exposure to gaseous carbon disulphide (Bliss 1935)

```r
library(glmx)
library(dplyr)
data(BeetleMortality, package = "glmx")
beetles <- BeetleMortality |>
  mutate(
    pct = died / n,
    survived = n - died,
    dose_c = dose - mean(dose)
  )
beetles
#> # A tibble: 8 x 6
#>    dose  died     n   pct survived   dose_c
#>   <dbl> <int> <int> <dbl>    <int>    <dbl>
#> 1  1.69     6    59 0.102       53  -0.103
#> 2  1.72    13    60 0.217       47  -0.0692
#> 3  1.76    18    62 0.290       44  -0.0382
#> 4  1.78    28    56 0.5         28  -0.00923
#> 5  1.81    52    63 0.825       11   0.0179
#> 6  1.84    53    59 0.898        6   0.0435
#> 7  1.86    61    62 0.984        1   0.0676
#> 8  1.88    60    60 1            0   0.0905
```

$$\pi(x) \equiv \mathrm{E}(Y = 1 | X = x)$$
$$= f(x^\top \beta)$$

Typically, we use the expit inverse-link:

$$\pi(\tilde{x}) = \mathrm{expit}(\tilde{x}'\beta) \tag{14}$$

1.6.3 Meet the beetles

```
library(ggplot2)
plot1 <-
  beetles |>
  ggplot(aes(x = dose, y = pct)) +
  geom_point(aes(size = n)) +
  xlab("Dose (log mg/L)") +
  ylab("Mortality rate (%)") +
  scale_y_continuous(labels = scales::percent) +
  scale_size(range = c(1, 2)) +
  theme_bw(base_size = 18)

print(plot1)
```


Figure 7: Mortality rates of adult flour beetles after five hours' exposure to gaseous carbon disulphide (Bliss 1935)

1.6.4 Why don't we use linear regression?

```r
beetles_long <- beetles |>
  reframe(
    .by = everything(),
    outcome = c(
      rep(1, times = died),
      rep(0, times = survived)
    )
  ) |>
  as_tibble()

lm1 <- beetles_long |> lm(formula = outcome ~ dose)
f_linear <- function(x) predict(lm1, newdata = data.frame(dose = x))

range1 <- range(beetles$dose) + c(-.2, .2)

plot2 <-
  plot1 +
  geom_function(
    fun = f_linear,
    aes(col = "Straight line")
  ) +
  labs(colour = "Model", size = "")

plot2 |> print()
```


Figure 8: Mortality rates of adult flour beetles after five hours' exposure to gaseous carbon disulphide (Bliss 1935)

1.6.5 Zoom out

```
(plot2 + expand_limits(x = c(1.6, 2))) |> print()
```


Figure 9: Mortality rates of adult flour beetles after five hours' exposure to gaseous carbon disulphide (Bliss 1935)

1.6.6 log transformation of dose?

```
lm2 <- beetles_long |> lm(formula = outcome ~ log(dose))
f_linearlog <- function(x) predict(lm2, newdata = data.frame(dose = x))

plot3 <- plot2 +
  expand_limits(x = c(1.6, 2)) +
  geom_function(fun = f_linearlog, aes(col = "Log-transform dose"))
(plot3 + expand_limits(x = c(1.6, 2))) |> print()
```
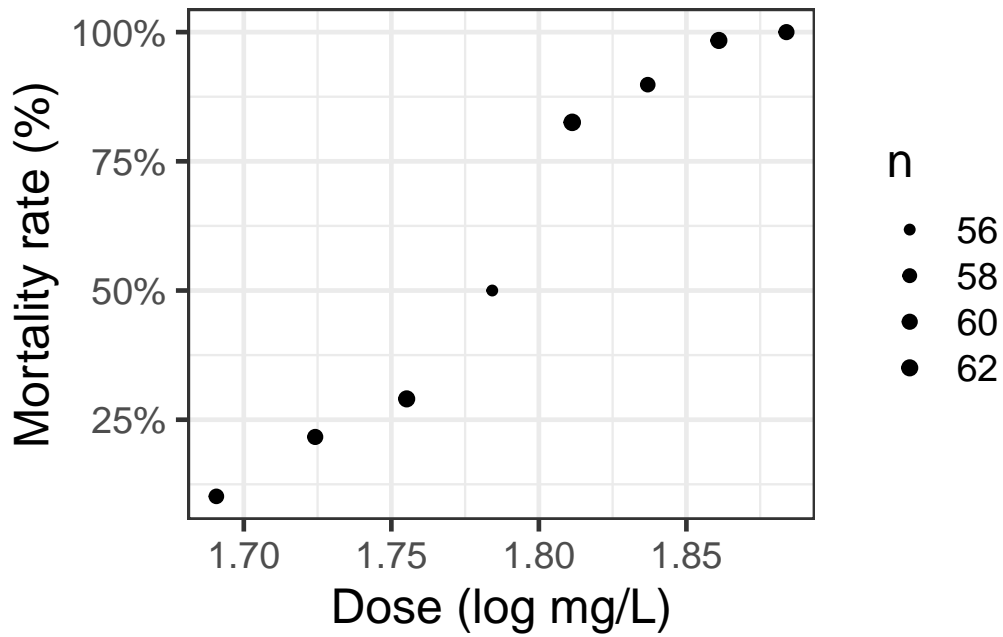

Figure 10: Mortality rates of adult flour beetles after five hours' exposure to gaseous carbon disulphide (Bliss 1935)

1.6.7 Logistic regression

```
beetles_glm_grouped <- beetles |>
  glm(formula = cbind(died, survived) ~ dose, family = "binomial")
f <- function(x) {
  beetles_glm_grouped |>
    predict(newdata = data.frame(dose = x), type = "response")
}

plot4 <- plot3 + geom_function(fun = f, aes(col = "Logistic regression"))
plot4 |> print()
```
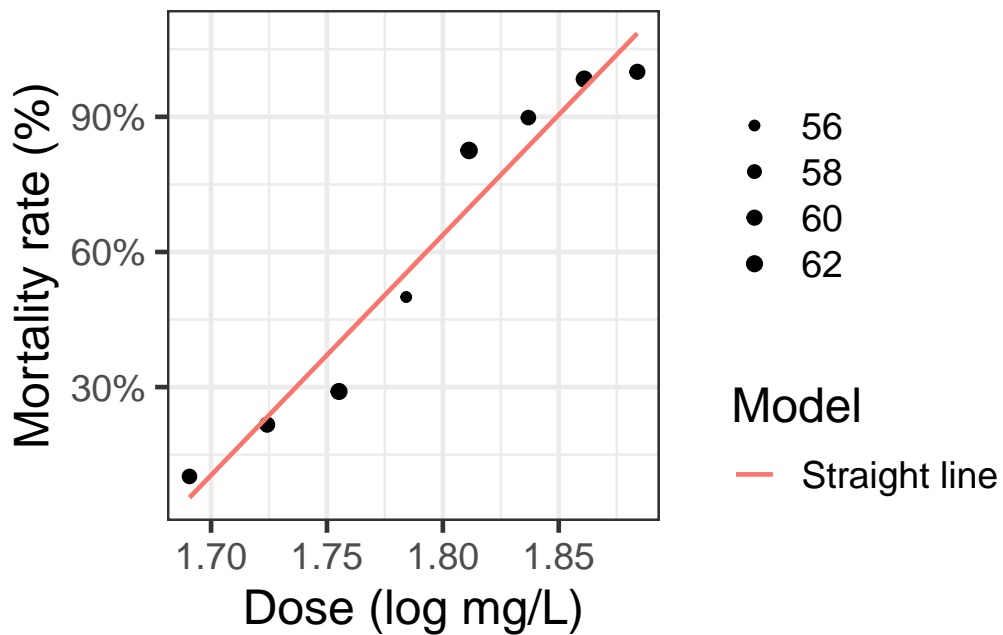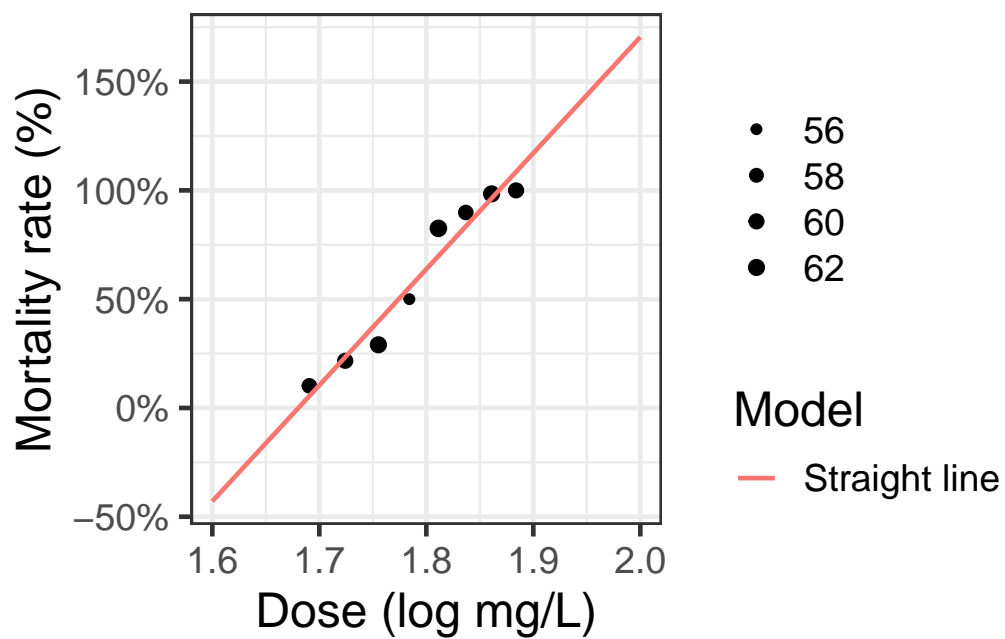

Figure 11: Mortality rates of adult flour beetles after five hours' exposure to gaseous carbon disulphide (Bliss 1935).

1.6.8 Three parts to regression models

- What distribution does the outcome have for a specific subpopulation defined by covariates? (outcome model)

- How does the combination of covariates relate to the mean? (link function)

- How do the covariates combine? (linear predictor, interactions)

1.6.9 Fitting and manipulating logistic regression models in R

```
library(glmx)
library(dplyr)
data(BeetleMortality)
beetles <- BeetleMortality |>
  mutate(
    pct = died / n,
    survived = n - died
  )
```

Table 6

```
fitted(beetles_glm_grouped)
#>        1        2        3        4        5        6        7        8
#> 0.058601 0.164028 0.362119 0.605315 0.795172 0.903236 0.955196 0.979049
predict(beetles_glm_grouped, type = "response")
#>        1        2        3        4        5        6        7        8
#> 0.058601 0.164028 0.362119 0.605315 0.795172 0.903236 0.955196 0.979049
```

```
beetles_glm_grouped <-
  beetles |>
  glm(
    formula = cbind(died, survived) ~ dose,
    family = "binomial"
  )

library(parameters)
beetles_glm_grouped |>
  parameters() |>
  print_md()
```

Table 5: logistic regression model for beetles data with grouped (binomial) data

Parameter	Log-Odds	SE	95% CI	z	p
(Intercept)	-60.72	5.18	(-71.44, -51.08)	-11.72	< .001
dose	34.27	2.91	(28.85, 40.30)	11.77	< .001

Fitted values

Fitted values are provided on the probability scale (Table 6)

Count scale

For grouped data, we can convert to the count scale by multiplying by the group size:

```
beetles$n * fitted(beetles_glm_grouped)
#>        1        2        3        4        5        6        7        8
#>  3.45746  9.84167 22.45138 33.89763 50.09582 53.29091 59.22216 58.74296
```

Logit scale

```
predict(beetles_glm_grouped, type = "link")
#>         1         2         3         4         5         6         7         8
#> -2.776615 -1.628559 -0.566179  0.427661  1.356386  2.233707  3.059622  3.844412
```

Converting between logit and probability scales works as expected:

```
predict(beetles_glm_grouped, type = "link") |> arm::invlogit()
#>        1        2        3        4        5        6        7        8
#> 0.058601 0.164028 0.362119 0.605315 0.795172 0.903236 0.955196 0.979049
predict(beetles_glm_grouped, type = "response")
#>        1        2        3        4        5        6        7        8
```

```
#> 0.058601 0.164028 0.362119 0.605315 0.795172 0.903236 0.955196 0.979049

predict(beetles_glm_grouped, type = "response") |> arm::logit()
#>         1         2         3         4         5         6         7         8
#> -2.776615 -1.628559 -0.566179  0.427661  1.356386  2.233707  3.059622  3.844412
predict(beetles_glm_grouped, type = "link")
#>         1         2         3         4         5         6         7         8
#> -2.776615 -1.628559 -0.566179  0.427661  1.356386  2.233707  3.059622  3.844412
```

_____

`type = "terms"` is confusing, because the variables get centered:

```
predict(beetles_glm_grouped, type = "terms")
#>        dose
#> 1 -3.520419
#> 2 -2.372363
#> 3 -1.309983
#> 4 -0.316144
#> 5  0.612582
#> 6  1.489902
#> 7  2.315817
#> 8  3.100608
#> attr(,"constant")
#> [1] 0.743804
coef(beetles_glm_grouped)["dose"] * beetles$dose
#> [1] 57.9408 59.0889 60.1513 61.1451 62.0738 62.9512 63.7771 64.5619
```

_____

We can construct the link-scale predictions from the terms:

```
terms_pred <- predict(beetles_glm_grouped, type = "terms")
terms_pred + attr(terms_pred, "constant")
#>        dose
#> 1 -2.776615
#> 2 -1.628559
#> 3 -0.566179
#> 4  0.427661
#> 5  1.356386
#> 6  2.233707
#> 7  3.059622
#> 8  3.844412
#> attr(,"constant")
#> [1] 0.743804
predict(beetles_glm_grouped, type = "link")
#>         1         2         3         4         5         6         7         8
#> -2.776615 -1.628559 -0.566179  0.427661  1.356386  2.233707  3.059622  3.844412
```

Individual observations

_____

```
beetles_glm_ungrouped <-
  beetles_long |>
  glm(
    formula = outcome ~ dose,
    family = "binomial"
  )
```

Table 7: `beetles` data in long format

```
beetles_long
#> # A tibble: 481 x 7
#>     dose  died     n   pct survived dose_c outcome
#>    <dbl> <int> <int> <dbl>    <int>  <dbl>   <dbl>
#>  1  1.69     6    59 0.102       53 -0.103       1
#>  2  1.69     6    59 0.102       53 -0.103       1
#>  3  1.69     6    59 0.102       53 -0.103       1
#>  4  1.69     6    59 0.102       53 -0.103       1
#>  5  1.69     6    59 0.102       53 -0.103       1
#>  6  1.69     6    59 0.102       53 -0.103       1
#>  7  1.69     6    59 0.102       53 -0.103       0
#>  8  1.69     6    59 0.102       53 -0.103       0
#>  9  1.69     6    59 0.102       53 -0.103       0
#> 10  1.69     6    59 0.102       53 -0.103       0
#> # i 471 more rows
```

```
beetles_glm_ungrouped |>
  parameters() |>
  print_md()
```

Table 8: logistic regression model for beetles data with individual Bernoulli data

Parameter	Log-Odds	SE	95% CI	z	p
(Intercept)	-60.72	5.18	(-71.44, -51.08)	-11.72	< .001
dose	34.27	2.91	(28.85, 40.30)	11.77	< .001

Exercise 1.28. Compare this model with the grouped-observations model (Table 5).

―――――――――――――――――

Solution 1.22.

They seem the same! But not quite:

```
logLik(beetles_glm_grouped)
#> 'log Lik.' -18.7151 (df=2)
logLik(beetles_glm_ungrouped)
#> 'log Lik.' -186.235 (df=2)
```

The difference is due to the binomial coefficient $\binom{n}{x}$ which isn't included in the individual-observations (Bernoulli) version of the model.

1.7 Derivatives of logistic regression functions

In order to interpret logistic regression models and find their MLEs, we will need to compute various derivatives. This section compiles some useful results.

1.7.1 Derivatives of odds function

Theorem 1.16 (Derivative of odds function)**.**

$$odds'\{\pi\} = \frac{\partial \omega}{\partial \pi} = \frac{1}{(1-\pi)^2}$$

―――――――――――――――――

36

Proof. We can use Theorem 1.2 and the quotient rule (**?@thm-quotient-rule**):

$$
\begin{aligned}
\frac{\partial \omega}{\partial \pi} &= \frac{\partial}{\partial \pi}\left(\frac{\pi}{1-\pi}\right) \\
&= \frac{\frac{\partial}{\partial \pi}\pi}{1-\pi} - \left(\frac{\pi}{(1-\pi)^2} \cdot \frac{\partial}{\partial \pi}(1-\pi)\right) \\
&= \frac{1}{1-\pi} - \frac{\pi}{(1-\pi)^2} \cdot (-1) \\
&= \frac{1}{1-\pi} + \frac{\pi}{(1-\pi)^2} \\
&= \frac{1-\pi}{(1-\pi)^2} + \frac{\pi}{(1-\pi)^2} \\
&= \frac{1-\pi+\pi}{(1-\pi)^2} \\
&= \frac{1}{(1-\pi)^2}
\end{aligned}
$$

$\square$

Corollary 1.7.

$$
\frac{\partial \omega}{\partial \pi} = (1+\omega)^2
$$

Proof. By Theorem 1.16 and Corollary 1.5. $\square$

1.7.2 Derivatives of inverse-odds function

Theorem 1.17 (Derivative of inverse odds function)**.**

$$
invodds'\{\omega\} = \frac{\partial \pi}{\partial \omega} = (1-\pi)^2 = \frac{1}{(1+\omega)^2} \tag{15}
$$

Proof. By Theorem 1.16 and Corollary 1.7.

Or for a direct approach, use the quotient rule (**?@thm-quotient-rule**) again:

$$
\begin{aligned}
\frac{\partial \pi}{\partial \omega} &= \frac{\partial}{\partial \omega}\frac{\omega}{1+\omega} \\
&= \frac{\frac{\partial}{\partial \omega}\omega}{1+\omega} - \frac{\omega}{(1+\omega)^2} \cdot \frac{\partial}{\partial \omega}(1+\omega) \\
&= \frac{1}{1+\omega} - \frac{\omega}{(1+\omega)^2} \cdot 1 \\
&= \frac{1}{1+\omega} - \frac{\omega}{(1+\omega)^2} \\
&= \frac{1+\omega}{(1+\omega)^2} - \frac{\omega}{(1+\omega)^2} \\
&= \frac{1+\omega-\omega}{(1+\omega)^2} \\
&= \frac{1}{(1+\omega)^2}
\end{aligned}
$$

$\square$

1.7.3 Derivatives of logit function

Lemma 1.4 (Derivative of log-odds by odds)**.**

$$\frac{\partial \eta}{\partial \omega} = \omega^{-1}$$

Proof. Using Definition 1.13:

$$\frac{\partial \eta}{\partial \omega} = \frac{\partial}{\partial \omega} \log \omega$$
$$= \omega^{-1}$$

$\square$

Theorem 1.18 (Derivative of log-odds by odds)**.**

$$\frac{\partial \eta}{\partial \omega} = \frac{1 - \pi}{\pi}$$

Proof. Using Theorem 1.2 and Lemma 1.4:

$$\frac{\partial \eta}{\partial \omega} = \omega^{-1}$$
$$= \frac{1 - \pi}{\pi}$$

$\square$

Theorem 1.19 (Derivative of log-odds by probability)**.**

$$\frac{\partial \eta}{\partial \pi} = \frac{1}{(\pi)(1 - \pi)}$$

Proof. Using Theorem 1.18, Theorem 1.16, and the chain rule (**?@thm-chain-rule**):

$$\frac{\partial \eta}{\partial \pi} = \frac{\partial \eta}{\partial \omega} \frac{\partial \omega}{\partial \pi}$$
$$= \frac{1 - \pi}{\pi} \frac{1}{(1 - \pi)^2}$$
$$= \frac{1}{(\pi)(1 - \pi)}$$

$\square$

Corollary 1.8 (Derivative of logit function)**.**

$$logit'(\pi) = \frac{1}{(\pi)(1 - \pi)}$$

Proof. By Theorem 1.19 and Corollary 1.6.

$\square$

1.7.4 Derivatives of expit function

Lemma 1.5.

$$\frac{\partial \omega}{\partial \eta} = \omega$$

Proof. Using Lemma 1.3 and **?@thm-deriv-exp**:

$$\begin{aligned}
\frac{\partial \omega}{\partial \eta} &= \frac{\partial}{\partial \eta}\exp\{\eta\} \\
&= \exp\{\eta\} \\
&= \omega
\end{aligned}$$

$\square$

Theorem 1.20.

$$\frac{\partial \omega}{\partial \eta} = \frac{\pi}{1 - \pi} \tag{16}$$

Proof. Use Lemma 1.5 and Theorem 1.2. $\square$

Theorem 1.21.

$$\frac{\partial \pi}{\partial \eta} = \pi(1 - \pi)$$

Proof. By the chain rule (**?@thm-chain-rule**), Theorem 1.20, and Theorem 1.17:

$$\begin{aligned}
\frac{\partial \pi}{\partial \eta} &= \frac{\partial \omega}{\partial \eta}\frac{\partial \pi}{\partial \omega} \\
&= \frac{\pi}{1 - \pi}(1 - \pi)^2 \\
&= \pi(1 - \pi)
\end{aligned}$$

Alternatively, by Theorem 1.19:

$$\begin{aligned}
\frac{\partial \pi}{\partial \eta} &= \left(\frac{\partial \eta}{\partial \pi}\right)^{-1} \\
&= \left(\frac{1}{(\pi)(1 - \pi)}\right)^{-1} \\
&= \pi(1 - \pi)
\end{aligned}$$

$\square$

Corollary 1.9. *If $\pi = \Pr(Y = 1|\tilde{X} = \tilde{x})$, then:*

$$\frac{\partial \pi}{\partial \eta} = Var(Y|X = x)$$

1.8 Understanding logistic regression models

Lemma 1.6. *By ?@thm-deriv-lincom:*

$$\frac{\partial \eta}{\partial \tilde{x}} = \frac{\partial}{\partial \tilde{x}} \tilde{x} \cdot \tilde{\beta}$$
$$= \tilde{\beta}$$

Exercise 1.29. Consider a logistic regression model with a single predictor, X:

$$Y_i | X_i \sim_{\perp\!\!\!\perp} \text{Ber}(\pi(X_i))$$
$$\pi(x) = \text{expit}\{\eta(x)\} = \pi(\omega(\eta(x))) \tag{17}$$
$$\eta(x) = \alpha + \beta x$$

Find the derivative of $\pi(x) = \text{E}[Y|X = x]$ with respect to x:

$$\frac{\partial \pi}{\partial x} = ?$$

Solution 1.23. By Theorem 1.21, Lemma 1.6, and the chain rule (**?@thm-chain-rule**):

$$\frac{\partial \pi}{\partial x} = \frac{\partial \pi}{\partial \eta} \frac{\partial \eta}{\partial x}$$
$$= \pi(1 - \pi)\beta$$
$$= \text{Var}(Y|X = x) \cdot \beta$$

The slope is steepest at $\pi = 0.5$, i.e., at $\eta = 0$, which for a unipredictor model occurs at $x = -\alpha/\beta$. The slope goes to 0 as x goes to $-\infty$ or $+\infty$ (compare with Figure 5).

> **ℹ Note**
>
> **In order to interpret** β_j: differentiate or difference $\eta(\tilde{x})$ with respect to x_j (depending on whether x_j is continuous or discrete, respectively):
>
> $$\frac{\partial}{\partial x_j} \eta(\tilde{x})$$
>
> **In order to find the MLE** $\hat{\tilde{\beta}}$: differentiate the log-likelihood function $\ell(\tilde{\beta})$ with respect to $\tilde{\beta}$:
>
> $$\frac{\partial}{\partial \tilde{\beta}} \ell(\tilde{\beta})$$

Exercise 1.30 (General formula for odds ratios in logistic regression)**.** Consider the generic logistic regression model:

- $Y_i | \tilde{X}_i \sim_{\perp\!\!\!\perp} \text{Ber}(\pi(\tilde{X}_i))$
- $\text{logit}\{\pi(\tilde{x})\} = \eta(\tilde{x})$
- $\eta(\tilde{x}) = \tilde{x}'\tilde{\beta}$

Let $\tilde{x}$ and $\tilde{x}^*$ be two covariate patterns, representing two individuals or two subpopulations.

Find a concise formula to compute the odds ratio comparing covariate patterns $\tilde{x}$ and $\tilde{x}^*$:

$$\theta(\tilde{x}, \tilde{x}^*) \stackrel{\text{def}}{=} \frac{\omega(\tilde{x})}{\omega(\tilde{x}^*)} \tag{18}$$

Solution 1.24 (General formula for odds ratios in logistic regression).

$$\begin{aligned}
\theta(\tilde{x}, \tilde{x}^*) &\stackrel{\text{def}}{=} \frac{\omega(\tilde{x})}{\omega(\tilde{x}^*)} \\
&= \frac{\exp\{\eta(\tilde{x})\}}{\exp\{\eta(\tilde{x}^*)\}} \\
&= \exp\{\eta(\tilde{x}) - \eta(\tilde{x}^*)\}
\end{aligned}$$

Solution 1.24 is more concrete than Equation 18, but it doesn't yet completely explain how to compute $\theta(\tilde{x}, \tilde{x}^*)$, so let's mark it as a lemma:

Lemma 1.7 (General formula for odds ratios in logistic regression)**.**

$$\theta(\tilde{x}, \tilde{x}^*) = exp\{\eta(\tilde{x}) - \eta(\tilde{x}^*)\} \tag{19}$$

Proof. By Solution 1.24. $\square$

Definition 1.16 (Difference in log-odds)**.**

Let $\tilde{x}$ and $\tilde{x}^*$ be two covariate patterns, representing two individuals or two subpopulations.

Then we can define the difference in log-odds between $\tilde{x}$ and $\tilde{x}^*$, denoted $\Delta\eta(\tilde{x}, \tilde{x}^*)$ or $\Delta\eta$ for short, as:

$$\Delta\eta \stackrel{\text{def}}{=} \eta(\tilde{x}) - \eta(\tilde{x}^*)$$

Corollary 1.10 (Shorthand general formula for odds ratios in logistic regression)**.**

$$\theta(\tilde{x}, \tilde{x}^*) = exp\{\Delta\eta\} \tag{20}$$

Proof. By Lemma 1.7 and Definition 1.16. $\square$

Exercise 1.31 (Difference in log-odds)**.** Find a concise expression for the difference in log-odds:

$$\Delta\eta \stackrel{\text{def}}{=} \eta(\tilde{x}) - \eta(\tilde{x}^*)$$

Solution 1.25 (Difference in log-odds).

$$\begin{aligned}
\Delta\eta &\stackrel{\text{def}}{=} \eta(\tilde{x}) - \eta(\tilde{x}^*) \\
&= (\tilde{x} \cdot \tilde{\beta}) - (\tilde{x}^* \cdot \tilde{\beta}) \\
&= (\tilde{x}^\top \tilde{\beta}) - ((\tilde{x}^*)^\top \tilde{\beta}) \\
&= (\tilde{x}^\top - (\tilde{x}^*)^\top)\tilde{\beta} \\
&= (\tilde{x} - \tilde{x}^*)^\top \tilde{\beta} \\
&= (\tilde{x} - \tilde{x}^*) \cdot \tilde{\beta}
\end{aligned}$$

Lemma 1.8 (Difference in log-odds)**.**

$$\Delta\eta = (\tilde{x} - \tilde{x}^*) \cdot \tilde{\beta}$$

Proof. By Solution 1.25. □

Definition 1.17 (Difference in covariate patterns)**.**

Let $\tilde{x}$ and $\tilde{x}^*$ be two covariate patterns, representing two individuals or two subpopulations. The difference in covariate patterns, denoted $\Delta\tilde{x}$, is defined as:

$$\Delta\tilde{x} \stackrel{\text{def}}{=} \tilde{x} - \tilde{x}^*$$

Corollary 1.11 (Difference in log-odds)**.**

$$\Delta\eta = (\Delta\tilde{x}) \cdot \tilde{\beta}$$

Proof. By Lemma 1.8 and Definition 1.17. □

Exercise 1.32. Find an expression for the odds ratio $\theta(\tilde{x}, \tilde{x}^*)$ in terms of $\Delta\tilde{x}$ and $\tilde{\beta}$.

Solution 1.26. Combine Corollary 1.10 and Corollary 1.11:

$$\theta(\tilde{x}, \tilde{x}^*) = \exp\{\Delta\eta\}$$
$$= \exp\left\{\Delta\tilde{x} \cdot \tilde{\beta}\right\}$$

Theorem 1.22. *The odds ratio comparing covariate patterns $\tilde{x}$ and $\tilde{x}^*$ is:*

$$\theta(\tilde{x}, \tilde{x}^*) = exp\left\{(\Delta\tilde{x}) \cdot \tilde{\beta}\right\} \tag{21}$$

Proof. By Solution 1.26. □

Corollary 1.12.

$$log\{\theta(\tilde{x}, \tilde{x}^*)\} = \Delta\eta$$

1.9 Estimating logistic regression models

1.9.1 Model

Assume:

- $Y_i | \tilde{X}_i \sim_{\perp\!\!\!\perp} \text{Ber}(\pi(X_i))$
- $\pi(\tilde{x}) = \text{expit}\{\eta(\tilde{x})\}$
- $\eta(\tilde{x}) = \tilde{x} \cdot \tilde{\beta}$

1.9.2 Likelihood function

Exercise 1.33. Compute and graph the likelihood for the `beetles` data model:

Table 9: Mortality rates of adult flour beetles after five hours' exposure to gaseous carbon disulphide (Bliss 1935)

```
library(glmx)
library(dplyr)
data(BeetleMortality)
beetles <- BeetleMortality |>
  mutate(
    pct = died / n,
    survived = n - died,
    dose_c = dose - mean(dose)
  )
beetles_long <-
  beetles |>
  reframe(
    .by = everything(),
    outcome = c(
      rep(1, times = died),
      rep(0, times = survived)
    )
  )
beetles
#> # A tibble: 8 x 6
#>    dose  died     n   pct survived   dose_c
#>   <dbl> <int> <int> <dbl>    <int>    <dbl>
#> 1  1.69     6    59 0.102       53 -0.103
#> 2  1.72    13    60 0.217       47 -0.0692
#> 3  1.76    18    62 0.290       44 -0.0382
#> 4  1.78    28    56 0.5         28 -0.00923
#> 5  1.81    52    63 0.825       11  0.0179
#> 6  1.84    53    59 0.898        6  0.0435
#> 7  1.86    61    62 0.984        1  0.0676
#> 8  1.88    60    60 1            0  0.0905
```

```
beetles_glm <-
  beetles |>
  glm(
    formula = cbind(died, survived) ~ dose,
    family = "binomial"
  )
equatiomatic::extract_eq(beetles_glm)
```

$$\log\left[\frac{P(\text{died} = 60)}{1 - P(\text{died} = 60)}\right] = \alpha + \beta_1(\text{dose}) \tag{22}$$

```
beetles_glm |>
  parameters::parameters() |>
  parameters::print_md()
```

Table 10: Fitted logistic regression model for `beetles` data

Parameter	Log-Odds	SE	95% CI	z	p
(Intercept)	-60.72	5.18	(-71.44, -51.08)	-11.72	< .001
dose	34.27	2.91	(28.85, 40.30)	11.77	< .001

Solution 1.27.

```
odds_inv <- function(omega) (1 + omega^-1)^-1
lik_beetles0 <- function(beta_0, beta_1) {
  beetles |>
    mutate(
      eta = beta_0 + beta_1 * dose,
      omega = exp(eta),
      pi = odds_inv(omega),
      Lik = pi^died * (1 - pi)^survived,
      # llik = died*eta + n*log(1 - pi)
    ) |>
    pull(Lik) |>
    prod()
}

lik_beetles <- Vectorize(lik_beetles0)
```

1.9.3 Log-likelihood function

Exercise 1.34. Find the log-likelihood function for the general logistic regression model.

Solution 1.28.

$$\begin{aligned}
\ell(\tilde{\beta}, \tilde{y}) &= \log\left\{\mathcal{L}(\tilde{\beta}, \tilde{y})\right\} \\
&= \sum_{i=1}^{n} \ell_i(\pi(\tilde{x}_i))
\end{aligned} \tag{23}$$

Using Theorem and Corollary :

$$
\begin{aligned}
\ell_i(\pi_i) &= y_i \log\{\pi_i\} + (1 - y_i)\log\{1 - \pi_i\} \\
&= y_i \log\{\pi_i\} + (1 \cdot \log\{1 - \pi_i\} - y_i \cdot \log\{1 - \pi_i\}) \\
&= y_i \log\{\pi_i\} + (\log\{1 - \pi_i\} - y_i \log\{1 - \pi_i\}) \\
&= y_i \log\{\pi_i\} + \log\{1 - \pi_i\} - y_i \log\{1 - \pi_i\} \\
&= y_i \log\{\pi_i\} - y_i \log\{1 - \pi_i\} + \log\{1 - \pi_i\} \\
&= (y_i \log\{\pi_i\} - y_i \log\{1 - \pi_i\}) + \log\{1 - \pi_i\} \\
&= y_i (\log\{\pi_i\} - \log\{1 - \pi_i\}) + \log\{1 - \pi_i\} \\
&= y_i \left(\log\left\{ \frac{\pi_i}{1 - \pi_i} \right\} \right) + \log\{1 - \pi_i\} \\
&= y_i \operatorname{logit}(\pi_i) + \log\{1 - \pi_i\} \\
&= y_i \eta_i + \log\{1 - \pi_i\} \\
&= y_i \eta_i + \log\left\{ (1 + \omega_i)^{-1} \right\} \\
&= y_i \eta_i - \log\{1 + \omega_i\}
\end{aligned}
$$

Lemma 1.9.

$$
\ell_i(\pi_i) = y_i \eta_i - log\{1 + \omega_i\}
$$

Exercise 1.35. Compute and graph the log-likelihood for the `beetles` data.

Solution 1.29.

```r
odds_inv <- function(omega) (1 + omega^-1)^-1
llik_beetles0 <- function(beta_0, beta_1) {
  beetles |>
    mutate(
      eta = beta_0 + beta_1 * dose,
      omega = exp(eta),
      pi = odds_inv(omega), # need for next line:
      llik = died*eta + n*log(1 - pi)
    ) |>
    pull(llik) |>
    sum()
}


llik_beetles <- Vectorize(llik_beetles0)

# to check that we implemented it correctly:
# ests = coef(beetles_glm_ungrouped)
# logLik(beetles_glm_ungrouped)
# llik_beetles(ests[1], ests[2])
```

Let's center dose:

```r
beetles_glm_grouped_centered <- beetles |>
  glm(
    formula = cbind(died, survived) ~ dose_c,
    family = "binomial"
  )

beetles_glm_ungrouped_centered <- beetles_long |>
```

```
  mutate(died = outcome) |>
  glm(
    formula = died ~ dose_c,
    family = "binomial"
  )

equatiomatic::extract_eq(beetles_glm_ungrouped_centered)
```

$$\log\left[\frac{P(\text{died}=1)}{1-P(\text{died}=1)}\right] = \alpha + \beta_1(\text{dose\_c}) \tag{24}$$

```
beetles_glm_grouped_centered |>
  parameters::parameters() |>
  parameters::print_md()
```

Table 11: Fitted logistic regression model for `beetles` data, with `dose` centered

Parameter	Log-Odds	SE	95% CI	z	p
(Intercept)	0.74	0.14	(0.48, 1.02)	5.40	< .001
dose c	34.27	2.91	(28.85, 40.30)	11.77	< .001

```
odds_inv <- function(omega) (1 + omega^-1)^-1
lik_beetles0 <- function(beta_0, beta_1) {
  beetles |>
    mutate(
      eta = beta_0 + beta_1 * dose_c,
      omega = exp(eta),
      pi = odds_inv(omega),
      Lik = (pi^died) * (1 - pi)^(survived)
    ) |>
    pull(Lik) |>
    prod()
}

lik_beetles <- Vectorize(lik_beetles0)
```

```
odds_inv <- function(omega) (1 + omega^-1)^-1
llik_beetles0 <- function(beta_0, beta_1) {
  beetles |>
    mutate(
      eta = beta_0 + beta_1 * dose_c,
      omega = exp(eta),
      pi = odds_inv(omega),
      llik = died * eta + n*log(1 - pi)
    ) |>
    pull(llik) |>
    sum()
}

llik_beetles <- Vectorize(llik_beetles0)
```

1.9.4 Score function

As usual, by independence, we have:

46

Lemma 1.10.

$$\tilde{\ell}'(\tilde{\beta}) \overset{def}{=} \frac{\partial}{\partial \tilde{\beta}} \ell(\tilde{\beta})$$

$$= \frac{\partial}{\partial \tilde{\beta}} \sum_{i=1}^{n} \ell_i(\tilde{\beta})$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \tilde{\beta}} \ell_i(\tilde{\beta})$$

$$= \sum_{i=1}^{n} \tilde{\ell}'_i(\tilde{\beta})$$

Starting from Lemma 1.9, we can apply the vector chain rule (**?@thm-chain-vec**):

Lemma 1.11.

$$\tilde{\ell}'_i(\tilde{\beta}) = \frac{\partial}{\partial \tilde{\beta}} \ell_i(\tilde{\beta})$$

$$= \frac{\partial}{\partial \tilde{\beta}} (y_i \eta_i - log\{1 + \omega_i\})$$

$$= \frac{\partial}{\partial \tilde{\beta}} y_i \eta_i - \frac{\partial}{\partial \tilde{\beta}} log\{1 + \omega_i\}$$

$$= \frac{\partial \eta_i}{\partial \tilde{\beta}} y_i - \frac{\partial \omega_i}{\partial \tilde{\beta}} \frac{1}{1 + \omega_i}$$

Lemma 1.12. *By **?@thm-deriv-lincom**:*

$$\frac{\partial \eta}{\partial \tilde{\beta}} = \frac{\partial}{\partial \tilde{\beta}} (\tilde{x} \cdot \tilde{\beta}) \tag{25}$$

$$= \tilde{x}$$

Lemma 1.12 is very similar to Lemma 1.6, but not quite the same; Lemma 1.6 differentiates by $\tilde{x}$, whereas Lemma 1.12 differentiates by $\tilde{\beta}$.

Theorem 1.23.

*To derive $\frac{\partial \omega}{\partial \tilde{\beta}}$, we can apply the vector chain rule (**?@thm-chain-vec**) again along with Lemma 1.5 and Lemma 1.12:*

$$\frac{\partial \omega}{\partial \tilde{\beta}} = \frac{\partial \eta}{\partial \tilde{\beta}} \frac{\partial \omega}{\partial \eta}$$

$$= \tilde{x} \omega$$

Corollary 1.13.

$$\frac{\partial \omega}{\partial \tilde{\beta}} = \tilde{x} \frac{\pi}{1 - \pi}$$

Now we can combine Lemma 1.11, Lemma 1.12, and Theorem 1.23:

$$\ell_i'(\tilde\beta) = \frac{\partial \eta_i}{\partial \tilde\beta} y_i - \frac{\partial \omega_i}{\partial \tilde\beta}\frac{1}{1+\omega_i}$$

$$= \tilde x_i y_i - \tilde x \omega_i \frac{1}{1+\omega_i}$$

$$= \tilde x_i y_i - \tilde x \frac{\omega_i}{1+\omega_i}$$

$$= \tilde x_i y_i - \tilde x_i \pi_i$$

$$= \tilde x_i (y_i - \pi_i)$$

$$= \tilde x_i (y_i - \mu_i)$$

$$= \tilde x_i (y_i - \mathrm{E}[Y_i | \tilde X_i = \tilde x_i])$$

$$= \tilde x_i \; \varepsilon(y_i | \tilde X_i = \tilde x_i)$$

$$= \tilde x_i \varepsilon_i$$

Theorem 1.24.

$$\ell_i'(\tilde\beta) = \tilde x_i \varepsilon_i \tag{26}$$

This last expression is essentially the same as we found in linear regression[11].

Finally, combining Lemma 1.10 and Theorem 1.24, we have:

Theorem 1.25.

$$\tilde\ell'(\tilde\beta) = \sum_{i=1}^n \ell_i'(\tilde\beta)$$

$$= \sum_{i=1}^n \tilde x_i \varepsilon_i \tag{27}$$

$$= \mathbf{X}^\top \tilde\varepsilon$$

The score function is vector-valued; its components are:

$$\frac{\partial \ell}{\partial \tilde\beta} = \begin{pmatrix} \frac{\partial \ell}{\partial \beta_0} \\ \frac{\partial \ell}{\partial \beta_1} \\ \vdots \\ \frac{\partial \ell}{\partial \beta_p} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n 1\varepsilon_i \\ \sum_{i=1}^n x_{i,1}\varepsilon_i \\ \vdots \\ \sum_{i=1}^n x_{i,p}\varepsilon_i \end{pmatrix} = \begin{pmatrix} \tilde 1 \cdot \tilde\varepsilon \\ \tilde x_1 \cdot \tilde\varepsilon \\ \vdots \\ \tilde x_p \cdot \tilde\varepsilon \end{pmatrix}$$

Thus, the score equation $\tilde\ell' = 0$ means that for the MLE $\hat{\tilde\beta}$:

1. the sum of the errors (aka deviations) equals 0:

$$\sum_{i=1}^n \varepsilon_i = 0$$

2. the sums of the errors times each covariate also equal 0:

$$\tilde x_j \cdot \tilde\varepsilon = \sum_{i=1}^n x_{ij}\varepsilon_i = 0, \forall j \in \{1:p\}$$

Example 1.9. In our model for the `beetles` data, we only have an intercept plus one covariate, gas concentration (c):

$$\tilde x = (1, c)$$

---

[11]Linear-models-overview.qmd#eq-scorefun-linreg

If c_i is the gas concentration for the beetle in observation i, and $\tilde{c} = (c_1, c_2, ...c_n)$, then the score equation $\tilde{\ell}' = 0$ means that for the MLE $\hat{\tilde{\beta}}$:

1. the sum of the errors (aka deviations) equals 0:

$$\sum_{i=1}^{n} \varepsilon_i = 0$$

2. the weighted sum of the error times the gas concentrations equals 0:

$$\sum_{i=1}^{n} c_i \varepsilon_i = 0$$

Exercise 1.36. Implement and graph the score function for the beetles data

---

Solution 1.30.

```r
odds_inv <- function(omega) (1 + omega^-1)^-1

score_fn_beetles_beta0_0 <- function(beta_0, beta_1) {
  beetles |>
    mutate(
      eta = beta_0 + beta_1 * dose_c,
      omega = exp(eta),
      pi = odds_inv(omega),
      mu = pi * n,
      epsilon = died - mu,
      score = epsilon
    ) |>
    pull(score) |>
    sum()
}
score_fn_beetles_beta_0 <- Vectorize(score_fn_beetles_beta0_0)

score_fn_beetles_beta1_0 <- function(beta_0, beta_1) {
  beetles |>
    mutate(
      eta = beta_0 + beta_1 * dose_c,
      omega = exp(eta),
      pi = odds_inv(omega),
      mu = pi * n,
      epsilon = died - mu,
      score = dose_c * epsilon
    ) |>
    pull(score) |>
    sum()
}
score_fn_beetles_beta_1 <- Vectorize(score_fn_beetles_beta1_0)
```

1.9.5 Hessian function

$$\ell''(\tilde{\beta}) = \sum_{i=1}^{n} \ell_i''(\tilde{\beta}) \tag{28}$$

$$\begin{aligned}
\ell_i''(\tilde{\beta}) &= \frac{\partial}{\partial \tilde{\beta}^\top} \tilde{\ell}_i' \\
&= \frac{\partial}{\partial \tilde{\beta}^\top} \tilde{x}_i \varepsilon_i \\
&= \tilde{x}_i \frac{\partial}{\partial \tilde{\beta}^\top} \varepsilon_i \\
&= \tilde{x}_i \varepsilon_i'
\end{aligned} \tag{29}$$

Theorem 1.26. *Using Lemma 1.12 and Theorem 1.21:*

$$\begin{aligned}
\frac{\partial \pi}{\partial \tilde{\beta}} &= \frac{\partial \eta}{\partial \tilde{\beta}} \frac{\partial \pi}{\partial \eta} \\
&= \tilde{x} \pi (1 - \pi)
\end{aligned}$$

Using Theorem 1.26:

$$\begin{aligned}
\varepsilon_i' &= \frac{\partial \varepsilon_i}{\partial \tilde{\beta}^\top} \\
&= \frac{\partial}{\partial \tilde{\beta}^\top} \varepsilon_i \\
&= \frac{\partial}{\partial \tilde{\beta}^\top} (y_i - \mu_i) \\
&= \frac{\partial}{\partial \tilde{\beta}^\top} y_i - \frac{\partial}{\partial \tilde{\beta}^\top} \mu_i \\
&= 0 - \frac{\partial}{\partial \tilde{\beta}^\top} \mu_i \\
&= -\frac{\partial \mu_i}{\partial \tilde{\beta}^\top} \\
&= -\frac{\partial \pi_i}{\partial \tilde{\beta}^\top} \\
&= -\pi_i (1 - \pi_i) \tilde{x}_i^\top \\
&= -\mathrm{Var}(Y_i | X_i = x_i) \tilde{x}_i^\top
\end{aligned}$$

Returning to Equation 29:

$$\begin{aligned}
\ell_i''(\tilde{\beta}) &= \tilde{x}_i \varepsilon_i' \\
&= -\tilde{x}_i \mathrm{Var}(Y_i | X_i = x_i) \tilde{x}_i^\top
\end{aligned} \tag{30}$$

Returning to Equation 28:

$$\begin{aligned}
\ell''(\tilde{\beta}) &= \sum_{i=1}^n \ell_i''(\tilde{\beta}) \\
&= -\sum_{i=1}^n \tilde{x}_i \mathrm{Var}(Y_i | X_i = x_i) \tilde{x}_i' \\
&= -\mathbf{X}^\top \mathbf{D} \mathbf{X}
\end{aligned} \tag{31}$$

where $\mathbf{D} \overset{\text{def}}{=} \mathrm{diag}(\mathrm{Var}(Y_i | X_i = x_i))$ is the diagonal matrix whose i^{th} diagonal element is $\mathrm{Var}(Y_i | X_i = x_i)$.

Compare with **?@eq-lm-hess** from linear regression:

$$\begin{aligned}
\ell''(\tilde{\beta}) &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\tilde{x}_i\tilde{x}_i' \\
&= -\mathbf{X}^\top\mathbf{D}^{-1}\mathbf{X}
\end{aligned}$$

(32)

Exercise 1.37. Determine the elements of the Hessian matrix for logistic regression.

Solution 1.31. The components of the Hessian are:

$$\begin{aligned}
\ell''(\beta) &= \frac{\partial^2}{\partial\beta^\top\partial\beta}\ell \\
&= \frac{\partial}{\partial\beta^\top}\ell' \\
&= \begin{bmatrix} \frac{\partial\ell'}{\partial\beta_0} & \frac{\partial\ell'}{\partial\beta_1} & \cdots & \frac{\partial\ell'}{\partial\beta_p} \end{bmatrix} \\
&= \begin{bmatrix}
\frac{\partial^2\ell}{\partial\beta_0^2} & \frac{\partial^2\ell}{\partial\beta_0\partial\beta_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_0\partial\beta_p} \\
\frac{\partial^2\ell}{\partial\beta_1\partial\beta_0} & \frac{\partial^2\ell}{\partial\beta_1^2} & \cdots & \frac{\partial^2\ell}{\partial\beta_0\partial\beta_p} \\
\vdots & \ddots & \ddots & \vdots \\
\frac{\partial^2\ell}{\partial\beta_p\partial\beta_0} & \frac{\partial^2\ell}{\partial\beta_p\partial\beta_1} & \cdots & \frac{\partial^2\ell}{\partial\beta_p^2}
\end{bmatrix}
\end{aligned}$$

Exercise 1.38. Determine the Hessian for the `beetles` model.

Solution 1.32. In the `beetles` model, the Hessian is:

$$\begin{aligned}
\ell''(\beta) &= \begin{bmatrix} \frac{\partial\ell'}{\partial\beta_0} & \frac{\partial\ell'}{\partial\beta_1} \end{bmatrix} \\
&= \begin{bmatrix}
\frac{\partial^2\ell}{\partial\beta_0^2} & \frac{\partial^2\ell}{\partial\beta_0\partial\beta_1} \\
\frac{\partial^2\ell}{\partial\beta_1\partial\beta_0} & \frac{\partial^2\ell}{\partial\beta_1^2}
\end{bmatrix} \\
&= \begin{bmatrix}
-\sum_{i=1}^{n}\pi_i(1-\pi_i) & -\sum_{i=1}^{n}c_i\pi_i(1-\pi_i) \\
-\sum_{i=1}^{n}c_i\pi_i(1-\pi_i) & -\sum_{i=1}^{n}c_i^2\pi_i(1-\pi_i)
\end{bmatrix}
\end{aligned}$$

Setting $\ell'(\tilde{\beta};\tilde{y}) = 0$ gives us:

$$\sum_{i=1}^{n}\{\tilde{x}_i(y_i - \text{expit}\{\tilde{x}_i'\beta\})\} = 0$$

(33)

In general, the estimating equation $\ell'(\tilde{\beta};\tilde{y}) = 0$ cannot be solved analytically.

Instead, we can use the Newton-Raphson method[12]:

$$\hat{\theta}^* \leftarrow \hat{\theta}^* - \left(\ell''\left(\tilde{y};\hat{\theta}^*\right)\right)^{-1}\ell'\left(\tilde{y};\hat{\theta}^*\right)$$

We make an iterative series of guesses, and each guess helps us make the next guess better (i.e., higher log-likelihood). You can see some information about this process like so:

---

[12]intro-MLEs.qmd#sec-newton-raphson

Table 12: Fitted model for `beetles` data

```
beetles_glm_ungrouped |> summary()
#>
#> Call:
#> glm(formula = outcome ~ dose, family = "binomial", data = beetles_long,
#>     control = glm.control(trace = TRUE))
#>
#> Coefficients:
#>             Estimate Std. Error z value Pr(>|z|)
#> (Intercept)   -60.72       5.18   -11.7   <2e-16 ***
#> dose           34.27       2.91    11.8   <2e-16 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> (Dispersion parameter for binomial family taken to be 1)
#>
#>     Null deviance: 645.44  on 480  degrees of freedom
#> Residual deviance: 372.47  on 479  degrees of freedom
#> AIC: 376.5
#>
#> Number of Fisher Scoring iterations: 5
```

Table 13: Parameter estimate covariance matrix for `beetles` data

```
beetles_glm_ungrouped |> vcov()
#>             (Intercept)      dose
#> (Intercept)    26.8393 -15.08189
#> dose          -15.0819   8.48041
```

```
beetles_glm_ungrouped <-
  beetles_long |>
  glm(
    control = glm.control(trace = TRUE),
    formula = outcome ~ dose,
    family = "binomial"
  )
#> Deviance = 383.249 Iterations - 1
#> Deviance = 372.921 Iterations - 2
#> Deviance = 372.472 Iterations - 3
#> Deviance = 372.471 Iterations - 4
#> Deviance = 372.471 Iterations - 5
```

After each iteration of the fitting procedure, the deviance $(2(\ell_{\text{full}} - \ell(\hat{\beta}))$) is printed. You can see that the algorithm took 5 iterations to converge to a solution where the likelihood wasn't changing much anymore.

Table 12 and Table 13 show the fitted model and the covariance matrix of the estimates, respectively.

1.10 Inference for logistic regression models

1.10.1 Inference for individual predictor coefficients

Wald tests and confidence intervals

(to be added)

1.10.2 Inference for odds ratios

Exercise 1.39. Given a maximum likelihood estimate $\hat\beta$ and a corresponding estimated covariance matrix $\hat\Sigma \stackrel{\text{def}}{=} \widehat{\text{Cov}}(\hat\beta)$, calculate a 95% confidence interval for the odds ratio comparing covariate patterns $\tilde x$ and $\tilde x^*$, $\theta(\tilde x, \tilde x^*)$.

Solution 1.33.

By **?@thm-dist-mle**, a 95% confidence interval for $\theta(\tilde x, \tilde x^*)$ can be constructed as:

$$\hat\theta \pm 1.96 * \widehat{\text{SE}}\left(\hat\theta\right) \tag{34}$$

However, $\widehat{\text{SE}}\left(\hat\theta\right)$ seems difficult to compute; doing so would require using the delta method[13].

Instead, using the invariance property of MLEs, we can first calculate a confidence interval for the logarithm of the odds ratio,

$$\log\{\theta(\tilde x, \tilde x^*)\} \in (L, R) \tag{35}$$

and then exponentiate the endpoints of that log-odds-scale confidence interval:

$$\theta(\tilde x, \tilde x^*) \in (e^L, e^R) \tag{36}$$

Exercise 1.40. Find a 95% confidence interval for the natural logarithm of the odds ratio, $\log\{\theta(\tilde x, \tilde x^*)\}$

Solution 1.34. From Corollary 1.12, we know:

$$\log\{\theta(\tilde x, \tilde x^*)\} = \Delta\eta$$

By **?@thm-dist-mle**, a 95% confidence interval for $\Delta\eta$ can be constructed as:

$$\widehat{\Delta\eta} \pm 1.96 * \widehat{\text{SE}}\left(\widehat{\Delta\eta}\right)$$

Exercise 1.41.

How can we estimate the standard error of $\widehat{\Delta\eta}$?

$$\widehat{\text{SE}}\left(\widehat{\Delta\eta}\right) = ?$$

Solution 1.35.

$$\text{SE}\left(\widehat{\Delta\eta}\right) = \sqrt{\text{Var}\left(\widehat{\Delta\eta}\right)} \tag{37}$$

By Lemma 1.8 and **?@thm-var-lincom**:

$$\begin{aligned}
\text{Var}\left(\widehat{\Delta\eta}\right) &= \text{Var}\left((\Delta\tilde x) \cdot \hat\beta\right) \\
&= (\Delta\tilde x)^\top \text{Cov}\left(\hat\beta\right)(\Delta\tilde x) \\
&= (\Delta\tilde x)^\top \Sigma(\Delta\tilde x)
\end{aligned} \tag{38}$$

---

[13]https://en.wikipedia.org/wiki/Delta_method

where $\Sigma \stackrel{\text{def}}{=} \text{Cov}(\hat{\beta})$.

Expanding Equation 38 out of matrix-vector notation, we have:

$$
\begin{aligned}
(\Delta \tilde{x})^{\top} \Sigma (\Delta \tilde{x}) &= \sum_{i=1}^{p} \sum_{j=1}^{p} (\Delta \tilde{x})_i \Sigma_{ij} (\Delta \tilde{x})_j \\
&= \sum_{i=1}^{p} \sum_{j=1}^{p} (\Delta x_i) \Sigma_{ij} (\Delta x_j) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{p} (x_i - x_i^*) \text{Cov}(\hat{\beta}_i, \hat{\beta}_j)(x_j - x_j^*)
\end{aligned}
$$

Combining Equation 38 and MLE invariance:

Theorem 1.27 (Estimated variance and standard error of difference in log-odds).

$$
\widehat{Var}(\Delta \hat{\eta}) = \Delta \tilde{x}^{\top} \hat{\Sigma} (\Delta \tilde{x}) \tag{39}
$$

$$
\widehat{SE}(\Delta \hat{\eta}) = \sqrt{\Delta \tilde{x}^{\top} \hat{\Sigma} (\Delta \tilde{x})} \tag{40}
$$

Note: on the RHS, we have plugged in $\hat{\Sigma}$, our estimate of Σ.

Compare this result with confidence intervals[14].

1.11 Multiple logistic regression

1.11.1 Coronary heart disease (WCGS) study data

Let's use the data from the Western Collaborative Group Study (WCGS) (Rosenman et al. (1975)) to explore multiple logistic regression:

From Vittinghoff et al. (2012):

"The **Western Collaborative Group Study (WCGS)** was a large epidemiological study designed to investigate the association between the"type A" behavior pattern and coronary heart disease (CHD)".

Exercise 1.42. What is "type A" behavior?

Solution 1.36. *From Wikipedia, "Type A and Type B personality theory":*

"The hypothesis describes Type A individuals as outgoing, ambitious, rigidly organized, highly status-conscious, impatient, anxious, proactive, and concerned with time management….

The hypothesis describes Type B individuals as a contrast to those of Type A. Type B personalities, by definition, are noted to live at lower stress levels. They typically work steadily and may enjoy achievement, although they have a greater tendency to disregard physical or mental stress when they do not achieve."

---

[14]Linear-models-overview.html#sec-se-fitted

Study design

from ?faraway::wcgs:

3154 healthy young men aged 39-59 from the San Francisco area were assessed for their personality type. All were free from coronary heart disease at the start of the research. Eight and a half years later change in CHD status was recorded.

Details (from faraway::wcgs)

The WCGS began in 1960 with 3,524 male volunteers who were employed by 11 California companies. Subjects were 39 to 59 years old and free of heart disease as determined by electrocardiogram. After the initial screening, the study population dropped to 3,154 and the number of companies to 10 because of various exclusions. The cohort comprised both blue- and white-collar employees.

1.11.2   Baseline data collection

socio-demographic characteristics:

- age
- education
- marital status
- income
- occupation
- physical and physiological including:
- height
- weight
- blood pressure
- electrocardiogram
- corneal arcus

----

biochemical measurements:

- cholesterol and lipoprotein fractions;
- medical and family history and use of medications;

----

behavioral data:

- Type A interview,
- smoking,
- exercise
- alcohol use.

Later surveys added data on:

- anthropometry
- triglycerides
- Jenkins Activity Survey
- caffeine use

Average follow-up continued for 8.5 years with repeat examinations.

1.11.3   Load the data

Here, I load the data:

```
### load the data directly from a UCSF website:
library(haven)
url <- paste0(
```

Table 14: `wcgs` data

```
wcgs
#> # A tibble: 3,154 x 22
#>      age arcus behpat   bmi chd69  chol   dbp dibpat height     id lnsbp lnwght
#>    <dbl> <lgl> <fct>  <dbl> <fct> <dbl> <dbl> <fct>   <dbl> <dbl> <dbl>  <dbl>
#>  1    50 TRUE  A1      31.3 No      249    90 Type A     67  2343  4.88   5.30
#>  2    51 FALSE A1      25.3 No      194    74 Type A     73  3656  4.79   5.26
#>  3    59 TRUE  A1      28.7 No      258    94 Type A     70  3526  5.06   5.30
#>  4    51 TRUE  A1      22.1 No      173    80 Type A     69 22057  4.84   5.01
#>  5    44 FALSE A1      22.3 No      214    80 Type A     71 12927  4.84   5.08
#>  6    47 FALSE A1      27.1 No      206    76 Type A     64 16029  4.75   5.06
#>  7    40 FALSE A1      23.2 No      190    78 Type A     70  3894  4.80   5.09
#>  8    41 FALSE A1      23.0 No      212    84 Type A     70 11389  4.87   5.08
#>  9    50 TRUE  A1      27.2 No      130    70 Type A     71 12681  4.72   5.27
#> 10    43 FALSE A1      28.4 No      233    80 Type A     68 10005  4.79   5.23
#> # i 3,144 more rows
#> # i 10 more variables: ncigs <dbl>, sbp <dbl>, smoke <fct>, t1 <dbl>,
#> #   time169 <dbl>, typchd69 <fct>, uni <dbl>, weight <dbl>, wghtcat <fct>,
#> #   agec <fct>
```

```
  # I'm breaking up the url into two chunks for readability
  "https://regression.ucsf.edu/sites/g/files/",
  "tkssra6706/f/wysiwyg/home/data/wcgs.dta"
)
wcgs <- haven::read_dta(url)
```

1.11.4 Data cleaning

Now let's do some data cleaning

```
library(arsenal) # provides `set_labels()`
library(forcats) # provides `as_factor()`
library(haven)
library(plotly)
wcgs <- wcgs |>
  mutate(
    age = age |>
      arsenal::set_labels("Age (years)"),
    arcus = arcus |>
      as.logical() |>
      arsenal::set_labels("Arcus Senilis"),
    time169 = time169 |>
      as.numeric() |>
      arsenal::set_labels("Observation (follow up) time (days)"),
    dibpat = dibpat |>
      as_factor() |>
      relevel(ref = "Type B") |>
      arsenal::set_labels("Behavioral Pattern"),
    typchd69 = typchd69 |>
      labelled(
        label = "Type of CHD Event",
        labels =
          c(
            "None" = 0,
            "infdeath" = 1,
            "silent" = 2,
```

```
      "angina" = 3
    )
  ),

# turn stata-style labelled variables in to R-style factors:
across(
  where(is.labelled),
  haven::as_factor
)
)
```

1.11.5   What's in the data

Table 15 summarizes the data.

1.11.6   Data by age and personality type

For now, we will look at the interaction between age and personality type (`dibpat`). To make it easier to visualize the data, we summarize the event rates for each combination of age:

```
library(dplyr)
odds <- function(pi) pi / (1 - pi)
chd_grouped_data <-
  wcgs |>
  summarize(
    .by = c(age, dibpat),
    n = sum(chd69 %in% c("Yes", "No")),
    x = sum(chd69 == "Yes")
  ) |>
  mutate(
    `n - x` = n - x,
    `p(chd)` = (x / n) |>
      labelled(label = "CHD Event by 1969"),
    `odds(chd)` = `p(chd)` / (1 - `p(chd)`),
    `logit(chd)` = log(`odds(chd)`)
  )

chd_grouped_data
#> # A tibble: 42 x 8
#>      age dibpat      n      x `n - x` `p(chd)`  `odds(chd)` `logit(chd)`
#>    <dbl> <fct>   <int> <int>   <int> <dbl+lbl>        <dbl>        <dbl>
#> 1     50 Type A     76     8      68 0.105            0.118        -2.14
#> 2     51 Type A     67    11      56 0.164            0.196        -1.63
#> 3     59 Type A     30     7      23 0.233            0.304        -1.19
#> 4     44 Type A    113     9     104 0.0796           0.0865       -2.45
#> 5     47 Type A     72     7      65 0.0972           0.108        -2.23
#> 6     40 Type A    133     9     124 0.0677           0.0726       -2.62
#> 7     41 Type A    108     7     101 0.0648           0.0693       -2.67
#> 8     43 Type A     97     7      90 0.0722           0.0778       -2.55
#> 9     54 Type A     53     7      46 0.132            0.152        -1.88
#> 10    48 Type A     80    12      68 0.15             0.176        -1.73
#> # i 32 more rows
```

1.11.7   Graphical exploration

```
library(ggplot2)
library(scales)
chd_plot_probs <-
```

Table 15: Baseline characteristics by CHD status at end of follow-up

```r
library(gtsummary)
wcgs |>
  dplyr::select(
    -dplyr::all_of(c("id", "uni", "t1"))
  ) |>
  gtsummary::tbl_summary(
    by = "chd69",
    missing_text = "Missing"
  ) |>
  gtsummary::add_p() |>
  gtsummary::add_overall() |>
  gtsummary::bold_labels() |>
  gtsummary::separate_p_footnotes()
```

Characteristic	Overall N = 3,154[1]	No N = 2,897[1]	Yes N = 2
Age (years)	45.0 (42.0, 50.0)	45.0 (41.0, 50.0)	49.0 (44.0, 5
Arcus Senilis	941 (30%)	839 (29%)	102 (40%
Missing	2	0	2
Behavioral Pattern			
A1	264 (8.4%)	234 (8.1%)	30 (12%
A2	1,325 (42%)	1,177 (41%)	148 (58%
B3	1,216 (39%)	1,155 (40%)	61 (24%
B4	349 (11%)	331 (11%)	18 (7.0%
Body Mass Index (kg/m2)	24.39 (22.96, 25.84)	24.39 (22.89, 25.84)	24.82 (23.63,
Total Cholesterol	223 (197, 253)	221 (195, 250)	245 (222, 2
Missing	12	12	0
Diastolic Blood Pressure	80 (76, 86)	80 (76, 86)	84 (80, 90
Behavioral Pattern			
Type B	1,565 (50%)	1,486 (51%)	79 (31%
Type A	1,589 (50%)	1,411 (49%)	178 (69%
Height (inches)	70.00 (68.00, 72.00)	70.00 (68.00, 72.00)	70.00 (68.00,
Ln of Systolic Blood Pressure	4.84 (4.79, 4.91)	4.84 (4.77, 4.91)	4.87 (4.82, 4
Ln of Weight	5.14 (5.04, 5.20)	5.13 (5.04, 5.20)	5.16 (5.09, 5
Cigarettes per day	0 (0, 20)	0 (0, 20)	20 (0, 30
Systolic Blood Pressure	126 (120, 136)	126 (118, 136)	130 (124, 1
Current smoking	1,502 (48%)	1,343 (46%)	159 (62%
Observation (follow up) time (days)	2,942 (2,842, 3,037)	2,952 (2,864, 3,048)	1,666 (934, 2
Type of CHD Event			
None	0 (0%)	0 (0%)	0 (0%)
infdeath	2,897 (92%)	2,897 (100%)	0 (0%)
silent	135 (4.3%)	0 (0%)	135 (53%
angina	71 (2.3%)	0 (0%)	71 (28%
4	51 (1.6%)	0 (0%)	51 (20%
Weight (lbs)	170 (155, 182)	169 (155, 182)	175 (162, 1
Weight Category			
< 140	232 (7.4%)	217 (7.5%)	15 (5.8%
140-170	1,538 (49%)	1,440 (50%)	98 (38%
170-200	1,171 (37%)	1,049 (36%)	122 (47%
> 200	213 (6.8%)	191 (6.6%)	22 (8.6%
RECODE of age (Age)			
35-40	543 (17%)	512 (18%)	31 (12%
41-45	1,091 (35%)	1,036 (36%)	55 (21%
46-50	750 (24%)	680 (23%)	70 (27%
51-55	528 (17%)	463 (16%)	65 (25%

```
chd_grouped_data |>
ggplot() +
aes(
    x = age,
    y = `p(chd)`,
    col = dibpat
  ) +
geom_point(aes(size = n), alpha = .7) +
scale_size(range = c(1, 4)) +
geom_line() +
theme_bw() +
ylab("P(CHD Event by 1969)") +
scale_y_continuous(
  labels = scales::label_percent(),
  sec.axis = sec_axis(
    ~ odds(.),
    name = "odds(CHD Event by 1969)"
  )
) +
theme(legend.position = "bottom")

print(chd_plot_probs)
```
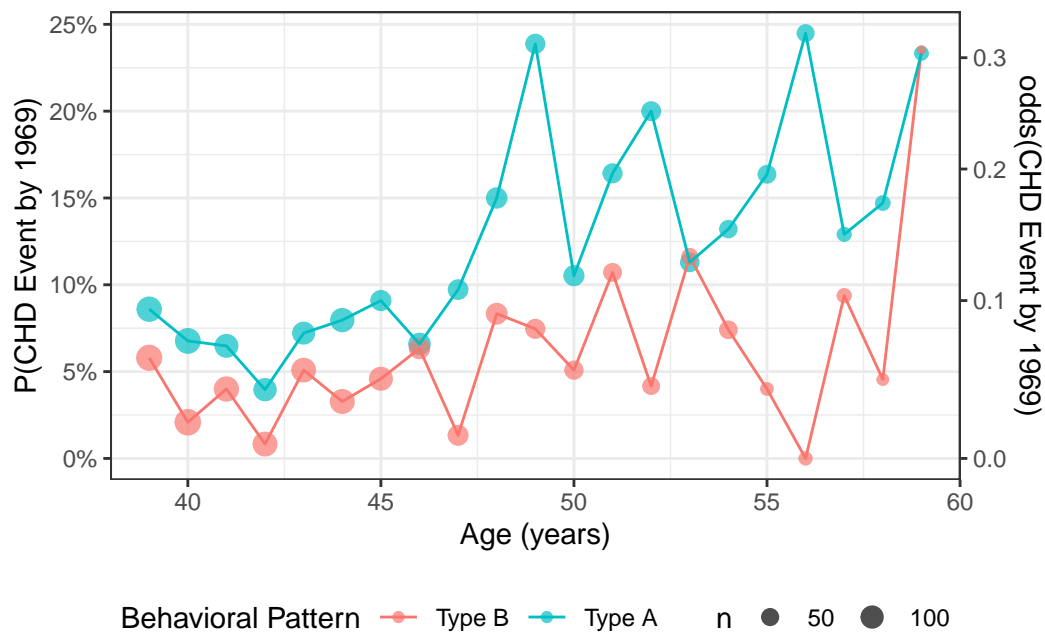
Figure 12: CHD rates by age group, probability scale

Odds scale

```
odds_inv <- function(omega) omega / (1 + omega)
trans_odds <- trans_new(
  name = "odds",
  transform = odds,
  inverse = odds_inv
)
```

59

```
chd_plot_odds <- chd_plot_probs +
  scale_y_continuous(
    trans = trans_odds, # this line changes the vertical spacing
    name = chd_plot_probs$labels$y,
    sec.axis = sec_axis(
      ~ odds(.),
      name = "odds(CHD Event by 1969)"
    )
  )

print(chd_plot_odds)
```
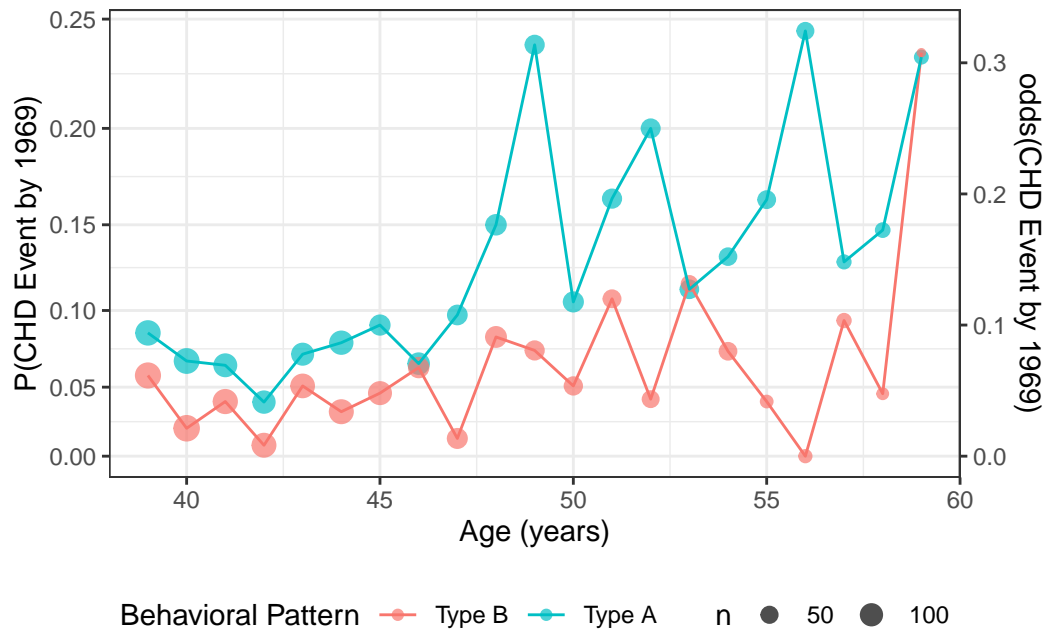

Figure 13: CHD rates by age group, odds spacing

Log-odds (logit) scale

```
logit <- function(pi) log(odds(pi))
expit <- function(eta) odds_inv(exp(eta))
trans_logit <- trans_new(
  name = "logit",
  transform = logit,
  inverse = expit
)

chd_plot_logit <-
  chd_plot_probs +
  scale_y_continuous(
    trans = trans_logit, # this line changes the vertical spacing
    name = chd_plot_probs$labels$y,
    breaks = c(seq(.01, .1, by = .01), .15, .2),
    minor_breaks = NULL,
    sec.axis = sec_axis(
      ~ logit(.),
```

```
      name = "log(odds(CHD Event by 1969))"
    )
  )

print(chd_plot_logit)
```
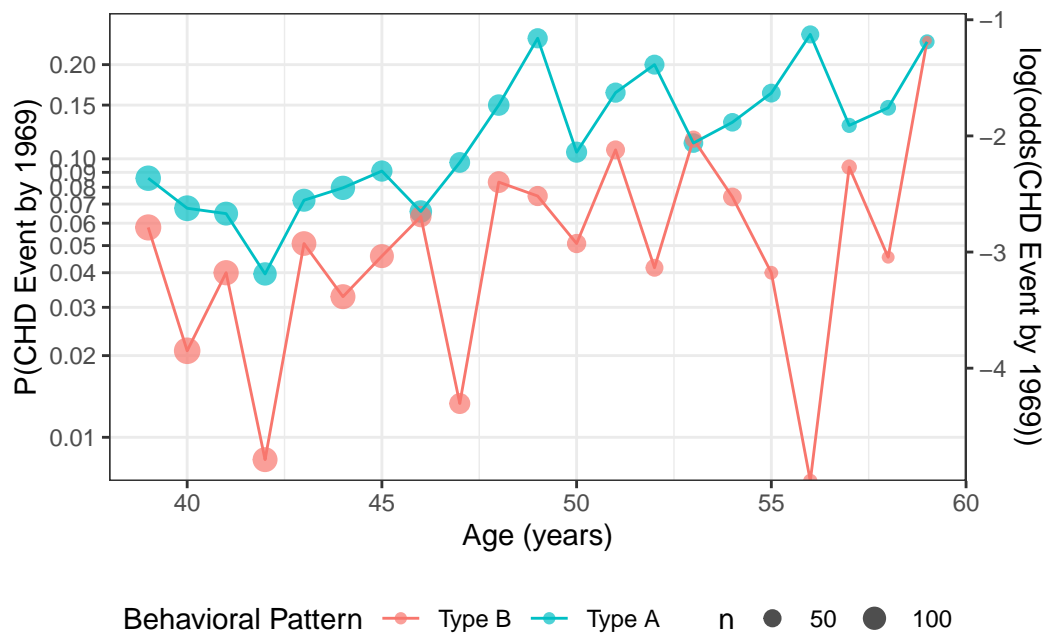

Figure 14: CHD data (logit-scale)

1.11.8 Logistic regression models for CHD data

For the `wgcs` dataset, let's consider a **logistic regression model** for the outcome of Coronary Heart Disease (Y; `chd` in computer output):

- $Y = 1$ if an individual developed CHD by the end of the study;
- $Y = 0$ if they have not developed CHD by the end of the study.

Let's include an intercept, two covariates, plus their interaction:

- A: age at study enrollment (`age`, recorded in years)
- P: personality type (`dibpat`):
    - $P = 1$ represents "Type A personality",
    - $P = 0$ represents "Type B personality".
- PA: the interaction of personality type and age (`dibpat:age`)
- $\tilde{X} = (1, A, P, PA)$

```
chd_glm_contrasts <-
  wcgs |>
  glm(
    "data" = _,
    "formula" = chd69 == "Yes" ~ dibpat * age,
    "family" = binomial(link = "logit")
  )

library(equatiomatic)
equatiomatic::extract_eq(chd_glm_contrasts)
```

61

$$\log\left[\frac{P(\text{chd69}=\text{Yes})}{1-P(\text{chd69}=\text{Yes})}\right] = \alpha + \beta_1(\text{dibpat}_{\text{Type A}}) + \beta_2(\text{age}) + \beta_3(\text{dibpat}_{\text{Type A}} \times \text{age}) \quad (41)$$

Or in more formal notation:

$$
\begin{aligned}
Y_i|\tilde{X}_i &\sim_{\perp\!\!\!\perp} \text{Ber}(\pi(\tilde{X}_i)) \\
\pi(\tilde{x}) &= \text{expit}(\eta(\tilde{x})) \\
\eta(\tilde{x}) &= \beta_0 + \beta_P p + \beta_A a + \beta_{PA} pa
\end{aligned}
\quad (42)
$$

1.11.9 Models superimposed on data

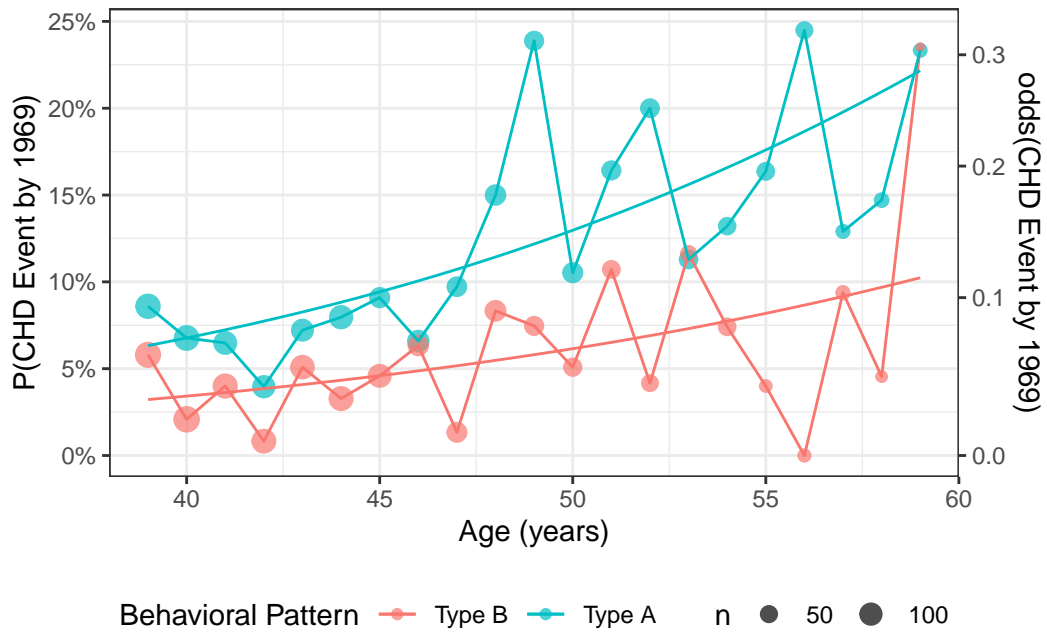We can graph our fitted models on each scale (probability, odds, log-odds).

---

probability scale

```
curve_type_A <- function(x) { # nolint: object_name_linter
  chd_glm_contrasts |> predict(
    type = "response",
    newdata = tibble(age = x, dibpat = "Type A")
  )
}

curve_type_B <- function(x) { # nolint: object_name_linter
  chd_glm_contrasts |> predict(
    type = "response",
    newdata = tibble(age = x, dibpat = "Type B")
  )
}

chd_plot_probs_2 <-
  chd_plot_probs +
  geom_function(
    fun = curve_type_A,
    aes(col = "Type A")
  ) +
  geom_function(
    fun = curve_type_B,
    aes(col = "Type B")
  )
print(chd_plot_probs_2)
```
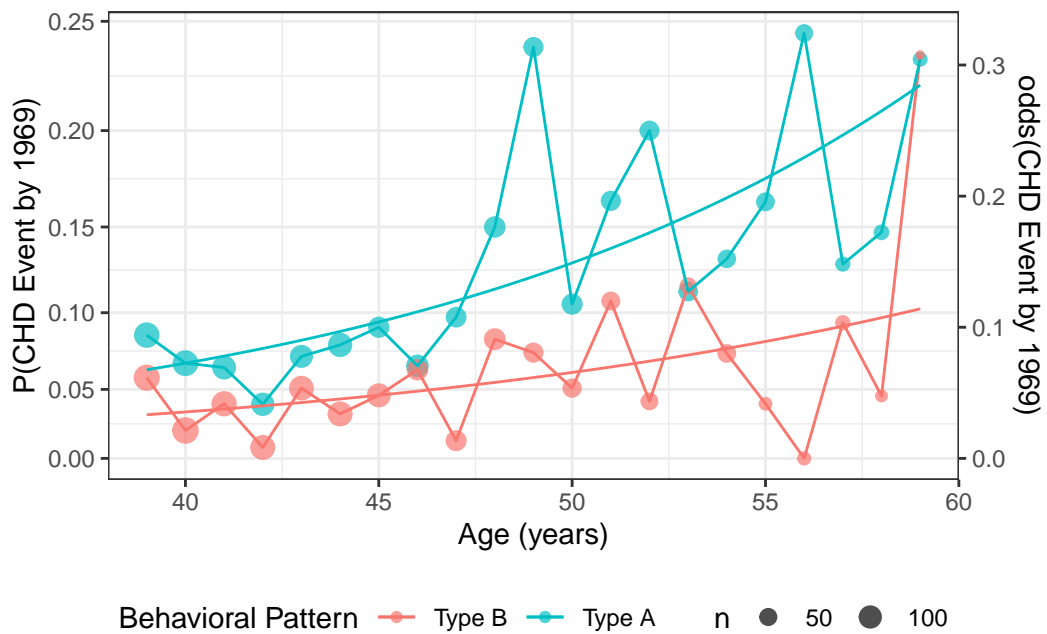
```
chd_plot_odds_2 <-
  chd_plot_odds +
  geom_function(
    fun = curve_type_A,
    aes(col = "Type A")
  ) +
  geom_function(
    fun = curve_type_B,
    aes(col = "Type B")
  )
print(chd_plot_odds_2)
```


odds scale

log-odds (logit) scale

```
chd_plot_logit_2 <-
  chd_plot_logit +
  geom_function(
    fun = curve_type_A,
    aes(col = "Type A")
  ) +
  geom_function(
    fun = curve_type_B,
    aes(col = "Type B")
  )

print(chd_plot_logit_2)
```
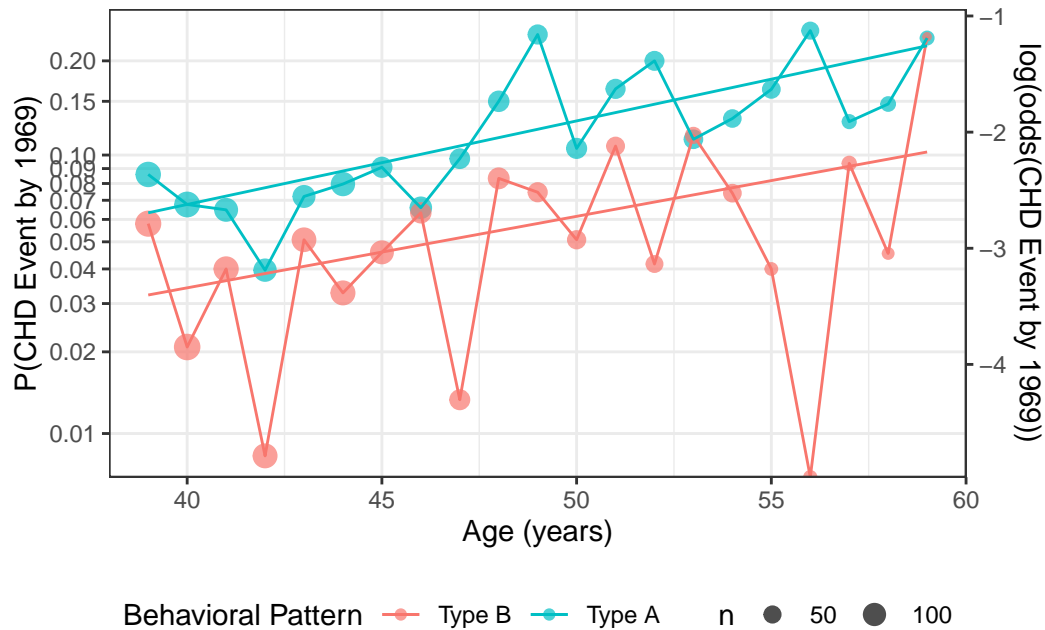

Figure 15

1.11.10 Interpreting the model parameters

Exercise 1.43. For Equation 42, derive interpretations of β_0, β_P, β_A, and β_{PA} on the odds **and** log-odds scales, State the interpretations concisely in math **and** in words.

Solution 1.37.

```
# include: false
age_offset = 0L
```

$$\eta(P = 0, A = 0) = \beta_0 + \beta_P 0 + \beta_A 0$$
$$= \beta_0 + 0 + 0$$
$$= \beta_0$$

Therefore:

$$\beta_0 = \eta(P = 0, A = 0) \tag{43}$$

β_0 is the natural logarithm of the odds ("log-odds") of experiencing CHD for a 0 year-old person with a type B personality; that is,

—————————————————————

e^{β_0} is the odds of experiencing CHD for a 0 year-old with a type B personality,

$$
\begin{aligned}
\exp\{\beta_0\} &= \frac{\Pr(Y = 1 | P = 0, A = 0)}{1 - \Pr(Y = 1 | P = 0, A = 0)} \\
&= \frac{\Pr(Y = 1 | P = 0, A = 0)}{\Pr(Y = 0 | P = 0, A = 0)}
\end{aligned}
$$

—————————————————————

$$
\begin{aligned}
\frac{\partial}{\partial a}\eta(P = 0, A = a) &= \frac{\partial}{\partial a}(\beta_0 + \beta_P 0 + \beta_A a + \beta_{PA}(0 \cdot a)) \\
&= \frac{\partial}{\partial a}\beta_0 + \frac{\partial}{\partial a}\beta_P 0 + \frac{\partial}{\partial a}\beta_A a + \frac{\partial}{\partial a}\beta_{PA}(0 \cdot a)) \\
&= 0 + 0 + \beta_A + 0 \\
&= \beta_A
\end{aligned}
$$

Therefore:

$$\beta_A = \frac{\partial}{\partial a}\eta(P = 0, A = a) \tag{44}$$

β_A is the slope of the log-odds of CHD with respect to age, for individuals with personality type B.

Alternatively:

$$\beta_A = \eta(P = 0, A = a + 1) - \eta(P = 0, A = a)$$

That is, β_A is the difference in log-odds of experiencing CHD experiencing CHD per one-year difference in age between two individuals with type B personalities.

—————————————————————

$$
\begin{aligned}
\exp\{\beta_A\} &= \exp\{\eta(P = 0, A = a + 1) - \eta(P = 0, A = a)\} \\
&= \frac{\exp\{\eta(P = 0, A = a + 1)\}}{\exp\{\eta(P = 0, A = a)\}} \\
&= \frac{\omega(P = 0, A = a + 1)}{\omega(P = 0, A = a)} \\
&= \frac{\text{odds}(Y = 1 | P = 0, A = a + 1)}{\text{odds}(Y = 1 | P = 0, A = a)} \\
&= \theta(\Delta a = 1 | P = 0)
\end{aligned}
$$

- The odds ratio of experiencing CHD (aka "the odds ratio") differs by a factor of e^{β_A} per one-year difference in age between individuals with type B personality.

—————————————————————

β_P is the difference in log-odds of experiencing CHD for a 0 year-old person with type A personality compared to a 0 year-old person with type B personality; that is,

$$\beta_P = \eta(P = 1, A = 0) - \eta(P = 0, A = 0) \tag{45}$$

- e^{β_P} is the ratio of the odds (aka "the odds ratio") of experiencing CHD, for a 0-year old individual with type A personality vs a 0-year old individual with type B personality; that is,

$$\exp\{\beta_P\} = \frac{\text{odds}(Y = 1 | P = 1, A = 0)}{\text{odds}(Y = 1 | P = 0, A = 0)}$$

$$\frac{\partial}{\partial a}\eta(P = 1, A = a) = \beta_A + \beta_{PA}$$

$$\frac{\partial}{\partial a}\eta(P = 0, A = a) = \beta_A$$

Therefore:

$$\frac{\partial}{\partial a}\eta(P = 1, A = a) - \frac{\partial}{\partial a}\eta(P = 0, A = a) = \beta_A + \beta_{PA} - \beta_A$$
$$= \beta_{PA}$$

That is,

$$\beta_{PA} = \frac{\partial}{\partial a}\eta(P = 1, A = a) - \frac{\partial}{\partial a}\eta(P = 0, A = a)$$
$$= \frac{\partial}{\partial a}\eta(P = 1, A = a) - \frac{\partial}{\partial a}\eta(P = 0, A = a)$$

β_{PA} is the difference in the slopes of log-odds over age between participants with Type A personalities and participants with Type B personalities.

Accordingly, the odds ratio of experiencing CHD per one-year difference in age differs by a factor of $e^{\beta_{PA}}$ for participants with type A personality compared to participants with type B personality; that is,

$$\theta(\Delta a = 1 | P = 1) = \exp\{\beta_{PA}\} \times \theta(\Delta a = 1 | P = 0)$$

or equivalently:

$$\exp\{\beta_{PA}\} = \frac{\theta(\Delta a = 1 | P = 1)}{\theta(\Delta a = 1 | P = 0)}$$

See Section 5.1.1[15] of Vittinghoff et al. (2012) for another perspective, also using the `wcgs` data as an example.

1.11.11 Interpreting the model parameter estimates

Table 16 shows the fitted model.

```
library(parameters)
chd_glm_contrasts |>
  parameters() |>
  print_md()
```

---

[15]https://link.springer.com/chapter/10.1007/978-1-4614-1353-0_5#Sec2_5

Table 16: CHD model (corner-point parametrization)

Parameter	Log-Odds	SE	95% CI	z	p
(Intercept)	-5.80	0.98	(-7.73, -3.90)	-5.95	< .001
dibpat (Type A)	0.30	1.18	(-2.02, 2.63)	0.26	0.797
age	0.06	0.02	(0.02, 0.10)	3.01	0.003
dibpat (Type A) × age	0.01	0.02	(-0.04, 0.06)	0.42	0.674

We can get the corresponding odds ratio estimates ($e^{\hat{\beta}}$s) by passing `exponentiate = TRUE` to `parameters()`:

```
chd_glm_contrasts |>
  parameters(exponentiate = TRUE) |>
  print_md()
```

Table 17: Odds ratio estimates for CHD model

Parameter	Odds Ratio	SE	95% CI	z	p
(Intercept)	3.02e-03	2.94e-03	(4.40e-04, 0.02)	-5.95	< .001
dibpat (Type A)	1.36	1.61	(0.13, 13.88)	0.26	0.797
age	1.06	0.02	(1.02, 1.11)	3.01	0.003
dibpat (Type A) × age	1.01	0.02	(0.96, 1.06)	0.42	0.674

1.11.12 Stratified parametrization

We could instead use a stratified parametrization:

```
chd_glm_strat <- glm(
  "formula" = chd69 == "Yes" ~ dibpat + dibpat:age - 1,
  "data" = wcgs,
  "family" = binomial(link = "logit")
)
equatiomatic::extract_eq(chd_glm_strat)
```

$$\log\left[\frac{P(\text{chd69} = \text{Yes})}{1 - P(\text{chd69} = \text{Yes})}\right] = \beta_1(\text{dibpat}_{\text{Type B}}) + \beta_2(\text{dibpat}_{\text{Type A}}) + \beta_3(\text{dibpat}_{\text{Type B}} \times \text{dibpat}_{\text{age}}) + \beta_4(\text{dibpat}_{\text{Type A}} \times \text{dibp}$$

(46)

```
chd_glm_strat |>
  parameters() |>
  print_md()
```

Table 18: CHD model, stratified parametrization

Parameter	Log-Odds	SE	95% CI	z	p
dibpat (Type B)	-5.80	0.98	(-7.73, -3.90)	-5.95	< .001
dibpat (Type A)	-5.50	0.67	(-6.83, -4.19)	-8.18	< .001
dibpat (Type B) × age	0.06	0.02	(0.02, 0.10)	3.01	0.003
dibpat (Type A) × age	0.07	0.01	(0.05, 0.10)	5.24	< .001

Again, we can get the corresponding odds ratios (e^{β}s) by passing `exponentiate = TRUE` to `parameters()`:

```
chd_glm_strat |>
  parameters(exponentiate = TRUE) |>
  print_md()
```

Table 19: Odds ratio estimates for CHD model

Parameter	Odds Ratio	SE	95% CI	z	p
dibpat (Type B)	3.02e-03	2.94e-03	(4.40e-04, 0.02)	-5.95	< .001
dibpat (Type A)	4.09e-03	2.75e-03	(1.08e-03, 0.02)	-8.18	< .001
dibpat (Type B) $\times$ age	1.06	0.02	(1.02, 1.11)	3.01	0.003
dibpat (Type A) $\times$ age	1.07	0.01	(1.05, 1.10)	5.24	< .001

Compare with Table 16.

Exercise 1.44. If I give you model 1, how would you get the coefficients of model 2?

1.12 Model comparisons for logistic models

1.12.1 Deviance test

We can compare the maximized log-likelihood of our model, $\ell(\hat{\beta}; \mathbf{x})$, versus the log-likelihood of the full model (aka saturated model aka maximal model), ℓ_{full}, which has one parameter per covariate pattern. With enough data, $2(\ell_{\text{full}} - \ell(\hat{\beta}; \mathbf{x})) \dot\sim \chi^2(N-p)$, where N is the number of distinct covariate patterns and p is the number of β parameters in our model. A significant p-value for this **deviance** statistic indicates that there's some detectable pattern in the data that our model isn't flexible enough to catch.

> 🔥 Caution
>
> The deviance statistic needs to have a large amount of data **for each covariate pattern** for the χ^2 approximation to hold. A guideline from Dobson is that if there are q distinct covariate patterns $x_1..., x_q$, with $n_1, ..., n_q$ observations per pattern, then the expected frequencies $n_k \cdot \pi(x_k)$ should be at least 1 for every pattern $k \in 1 : q$.

If you have covariates measured on a continuous scale, you may not be able to use the deviance tests to assess goodness of fit.

1.12.2 Hosmer-Lemeshow test

If our covariate patterns produce groups that are too small, a reasonable solution is to make bigger groups by merging some of the covariate-pattern groups together.

Hosmer and Lemeshow (1980) proposed that we group the patterns by their predicted probabilities according to the model of interest. For example, you could group all of the observations with predicted probabilities of 10% or less together, then group the observations with 11%-20% probability together, and so on; $g = 10$ categories in all.

Then we can construct a statistic

$$X^2 = \sum_{c=1}^{g} \frac{(o_c - e_c)^2}{e_c}$$

where o_c is the number of events *observed* in group c, and e_c is the number of events expected in group c (based on the sum of the fitted values $\hat{\pi}_i$ for observations in group c).

If each group has enough observations in it, you can compare X^2 to a χ^2 distribution; by simulation, the degrees of freedom has been found to be approximately $g - 2$.

For our CHD model, this procedure would be:

```r
wcgs <-
  wcgs |>
  mutate(
    pred_probs_glm1 = chd_glm_contrasts |> fitted(),
    pred_prob_cats1 = pred_probs_glm1 |>
      cut(
        breaks = seq(0, 1, by = .1),
        include.lowest = TRUE
      )
  )

HL_table <- # nolint: object_name_linter
  wcgs |>
  summarize(
    .by = pred_prob_cats1,
    n = n(),
    o = sum(chd69 == "Yes"),
    e = sum(pred_probs_glm1)
  )

library(pander)
HL_table |> pander()
```

pred_prob_cats1	n	o	e
(0.1,0.2]	785	116	108
(0.2,0.3]	64	12	13.77
[0,0.1]	2,305	129	135.2

```r
X2 <- HL_table |> # nolint: object_name_linter
  summarize(
    `X^2` = sum((o - e)^2 / e)
  ) |>
  pull(`X^2`)
print(X2)
#> [1] 1.11029

pval1 <- pchisq(X2, lower = FALSE, df = nrow(HL_table) - 2)
```
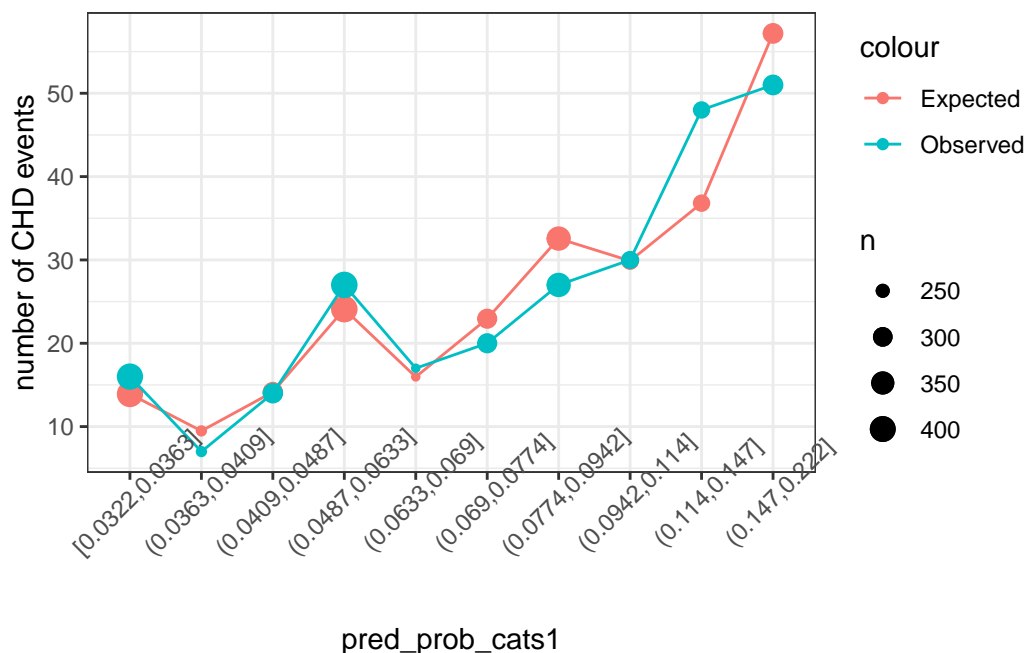
Our statistic is $X^2 = 1.110287$; $p(\chi^2(1) > 1.110287) = 0.29202$, which is our p-value for detecting a lack of goodness of fit.

Unfortunately that grouping plan left us with just three categories with any observations, so instead of grouping by 10% increments of predicted probability, typically analysts use deciles of the predicted probabilities:

```r
wcgs <-
  wcgs |>
  mutate(
    pred_probs_glm1 = chd_glm_contrasts |> fitted(),
    pred_prob_cats1 = pred_probs_glm1 |>
      cut(
        breaks = quantile(pred_probs_glm1, seq(0, 1, by = .1)),
        include.lowest = TRUE
      )
  )

HL_table <- # nolint: object_name_linter
```

```
  wcgs |>
  summarize(
    .by = pred_prob_cats1,
    n = n(),
    o = sum(chd69 == "Yes"),
    e = sum(pred_probs_glm1)
  )

HL_table |> pander()
```

pred_prob_cats1	n	o	e
(0.114,0.147]	275	48	36.81
(0.147,0.222]	314	51	57.19
(0.0774,0.0942]	371	27	32.56
(0.0942,0.114]	282	30	29.89
(0.0633,0.069]	237	17	15.97
(0.069,0.0774]	306	20	22.95
(0.0487,0.0633]	413	27	24.1
(0.0409,0.0487]	310	14	14.15
[0.0322,0.0363]	407	16	13.91
(0.0363,0.0409]	239	7	9.48

```
X2 <- HL_table |> # nolint: object_name_linter
  summarize(
    `X^2` = sum((o - e)^2 / e)
  ) |>
  pull(`X^2`)

print(X2)
#> [1] 6.78114

pval1 <- pchisq(X2, lower = FALSE, df = nrow(HL_table) - 2)
```

Now we have more evenly split categories. The p-value is 0.56042, still not significant.

Graphically, we have compared:

```
HL_plot <- # nolint: object_name_linter
  HL_table |>
  ggplot(aes(x = pred_prob_cats1)) +
  geom_line(
    aes(y = e, x = pred_prob_cats1, group = "Expected", col = "Expected")
  ) +
  geom_point(aes(y = e, size = n, col = "Expected")) +
  geom_point(aes(y = o, size = n, col = "Observed")) +
  geom_line(aes(y = o, col = "Observed", group = "Observed")) +
  scale_size(range = c(1, 4)) +
  theme_bw() +
  ylab("number of CHD events") +
  theme(axis.text.x = element_text(angle = 45))
```

```
print(HL_plot)
```


number of CHD events

colour
- Expected
- Observed

n
- 250
- 300
- 350
- 400

pred_prob_cats1

1.12.3 Comparing models

- AIC $= -2 * \ell(\hat{\theta}) + 2 * p$ [lower is better]
- BIC $= -2 * \ell(\hat{\theta}) + p * \log(n)$ [lower is better]
- likelihood ratio [higher is better]

1.13 Residual-based diagnostics

1.13.1 Logistic regression residuals only work for grouped data

```
library(haven)
url <- paste0(
  # I'm breaking up the url into two chunks for readability
  "https://regression.ucsf.edu/sites/g/files/",
  "tkssra6706/f/wysiwyg/home/data/wcgs.dta"
)
library(here) # provides the `here()` function
library(fs) # provides the `path()` function
here::here() |>
  fs::path("Data/wcgs.rda") |>
  load()
chd_glm_contrasts <-
  wcgs |>
  glm(
    "data" = _,
    "formula" = chd69 == "Yes" ~ dibpat * age,
    "family" = binomial(link = "logit")
  )
library(ggfortify)
chd_glm_contrasts |> autoplot()
```
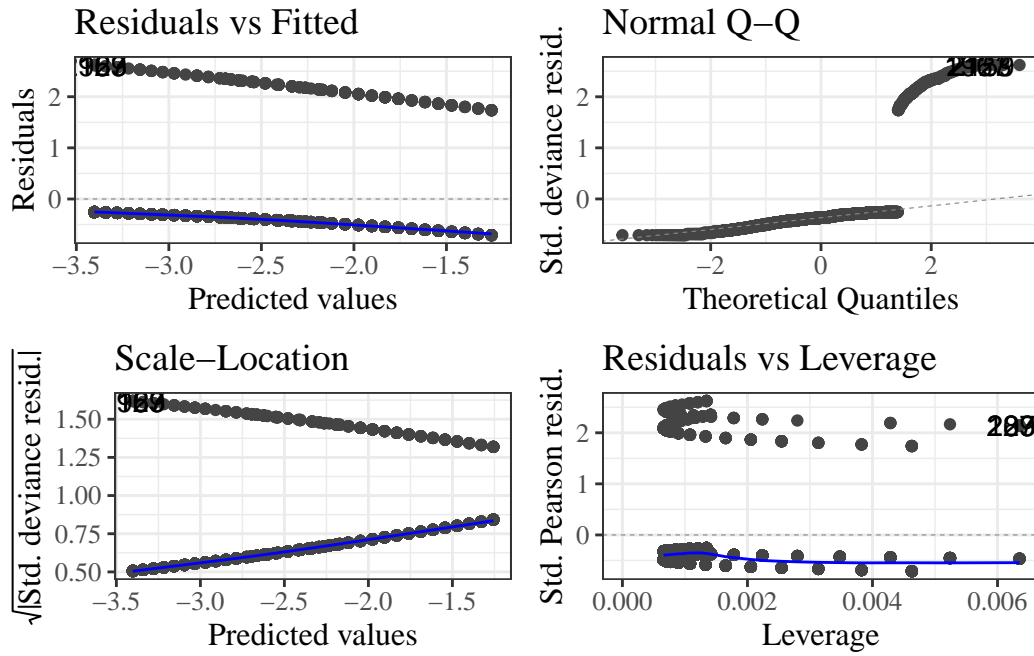
Figure 16: Residual diagnostics for WCGS model with individual-level observations

---

Residuals only work if there is more than one observation for most covariate patterns.

Here we will create the grouped-data version of our CHD model from the WCGS study:

```r
library(dplyr)
wcgs_grouped <-
  wcgs |>
  summarize(
    .by = c(dibpat, age),
    n = n(),
    chd = sum(chd69 == "Yes"),
    no_chd = sum(chd69 == "No")
  ) |>
  mutate(p_chd = chd/n)

chd_glm_contrasts_grouped <- glm(
  "formula" = cbind(chd, no_chd) ~ dibpat*age,
  "data" = wcgs_grouped,
  "family" = binomial(link = "logit")
)
chd_glm_contrasts_grouped |> equatiomatic::extract_eq()
```
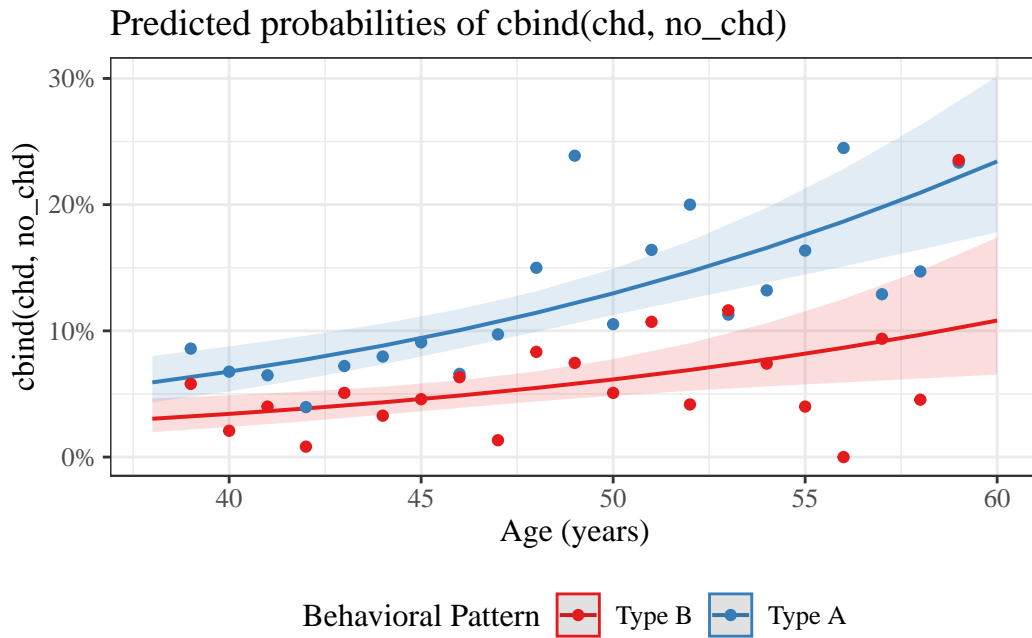
$$\log\left[\frac{P(\text{chd})}{1 - P(\text{chd})}\right] = \alpha + \beta_1(\text{dibpat}_{\text{Type A}}) + \beta_2(\text{age}) + \beta_3(\text{dibpat}_{\text{Type A}} \times \text{age}) \qquad (47)$$

```r
library(parameters)
chd_glm_contrasts_grouped |>
  parameters() |>
  print_md()
```
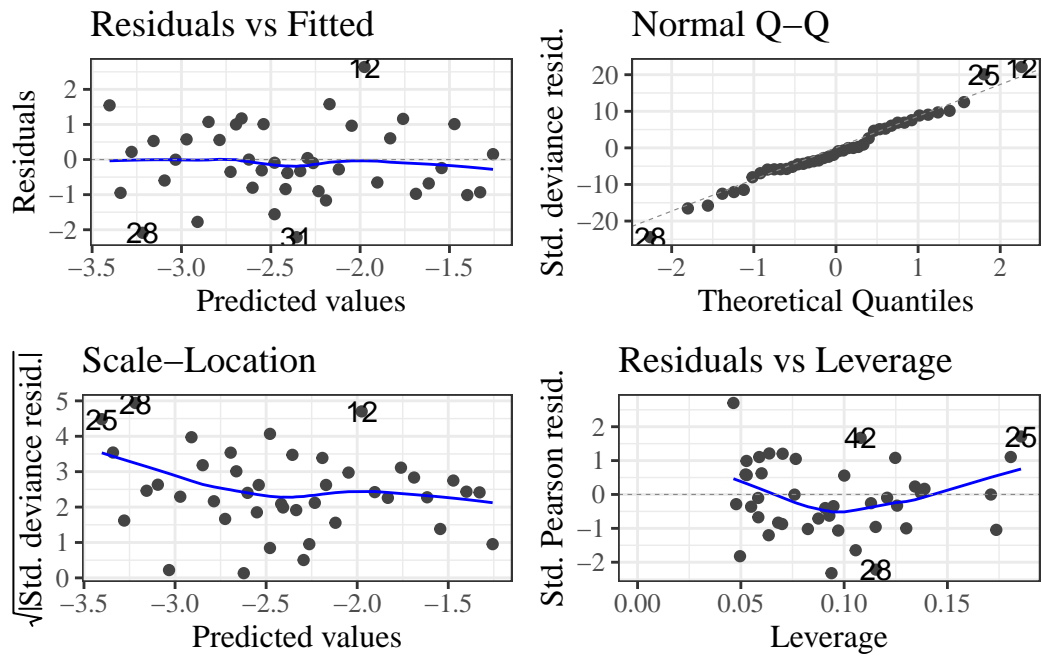
Table 22: CHD model with grouped `wcgs` data

Parameter	Log-Odds	SE	95% CI	z	p
(Intercept)	-5.80	0.98	(-7.73, -3.90)	-5.95	< .001
dibpat (Type A)	0.30	1.18	(-2.02, 2.63)	0.26	0.797
age	0.06	0.02	(0.02, 0.10)	3.01	0.003
dibpat (Type A) × age	0.01	0.02	(-0.04, 0.06)	0.42	0.674

```
chd_glm_contrasts_grouped |>
  sjPlot::plot_model(type = "pred", terms = c("age", "dibpat")) +
  geom_point(data = wcgs_grouped |> mutate(group_col = dibpat),
             aes(x = age, y = p_chd))
```


Figure 17: CHD model with grouped `wcgs` data

```
library(ggfortify)
chd_glm_contrasts_grouped |> autoplot()
```

1.13.2 (Response) residuals

$$e_k \overset{\text{def}}{=} \bar{y}_k - \hat{\pi}(x_k)$$

(k indexes the covariate patterns)

We can graph these residuals e_k against the fitted values $\hat{\pi}(x_k)$:

```r
odds <- function(pi) pi/(1-pi)
logit <- function(pi) log(odds(pi))
wcgs_grouped <-
  wcgs_grouped |>
  mutate(
    fitted = chd_glm_contrasts_grouped |> fitted(),
    fitted_logit = fitted |> logit(),
    response_resids = chd_glm_contrasts_grouped |> resid(type = "response")
  )

wcgs_response_resid_plot <-
  wcgs_grouped |>
  ggplot(
    mapping = aes(
      x = fitted,
      y = response_resids
    )
  ) +
  geom_point(
    aes(col = dibpat)
  ) +
  geom_hline(yintercept = 0) +
  geom_smooth(
    se = TRUE,
    method.args = list(
      span = 2 / 3,
      degree = 1,
      family = "symmetric",
      iterations = 3
```

```
  ),
  method = stats::loess
)
```

① Don't worry about these options for now; I chose them to match `autoplot()` as closely as I can. `plot.glm` and `autoplot` use `stats::lowess` instead of `stats::loess`; `stats::lowess` is older, hard to use with `geom_smooth`, and hard to match exactly with `stats::loess`; see https://support.bioconductor.org/p/2323/.]

```
wcgs_response_resid_plot |> print()
```


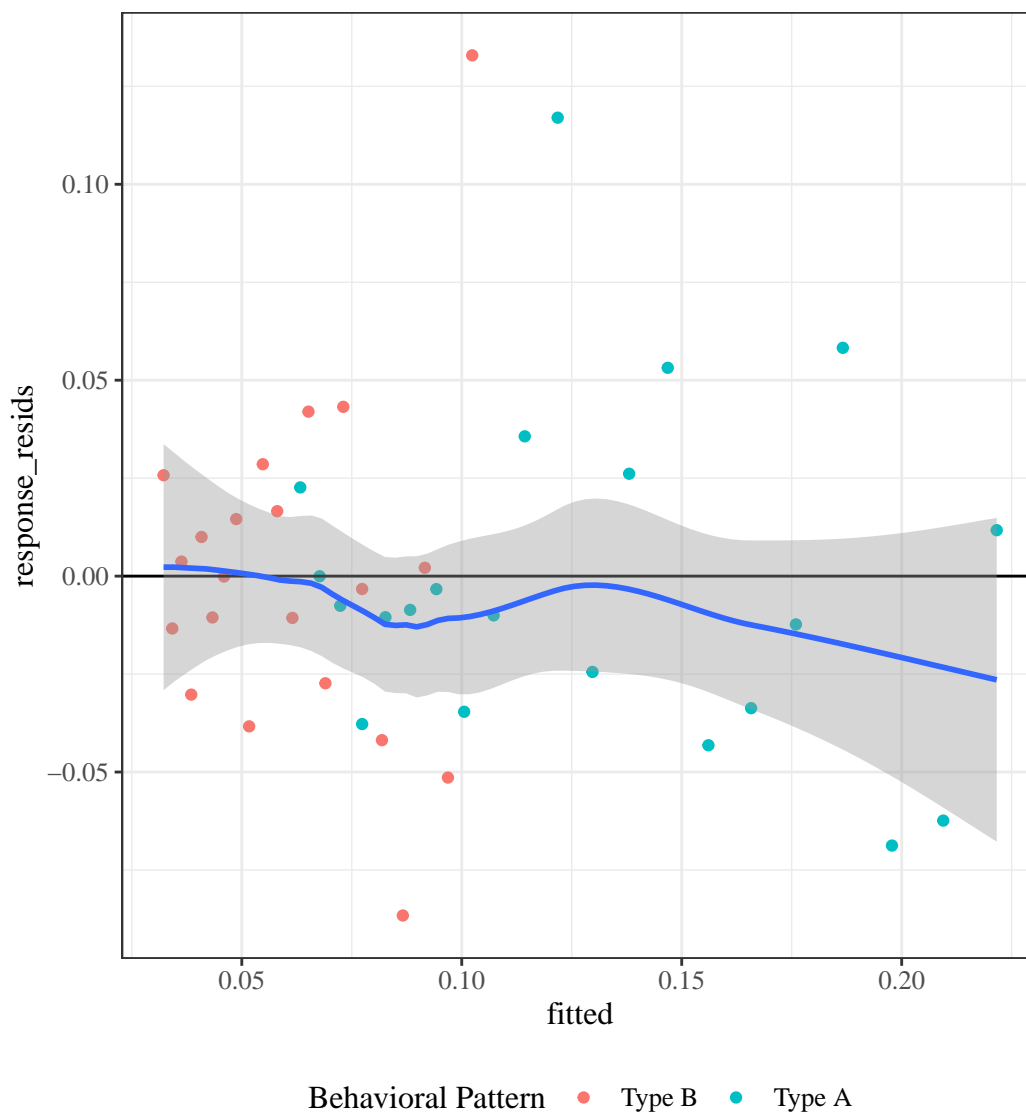
Figure 18: residuals plot for `wcgs` model

We can see a slight fan-shape here: observations on the right have larger variance (as expected since $var(\bar{y}) = \pi(1-\pi)/n$ is maximized when $\pi = 0.5$).

1.13.3 Pearson residuals

The fan-shape in the response residuals plot isn't necessarily a concern here, since we haven't made an assumption of constant residual variance, as we did for linear regression.

However, we might want to divide by the standard error in order to make the graph easier to interpret. Here's one way to do that:

The Pearson (chi-squared) residual for covariate pattern k is:

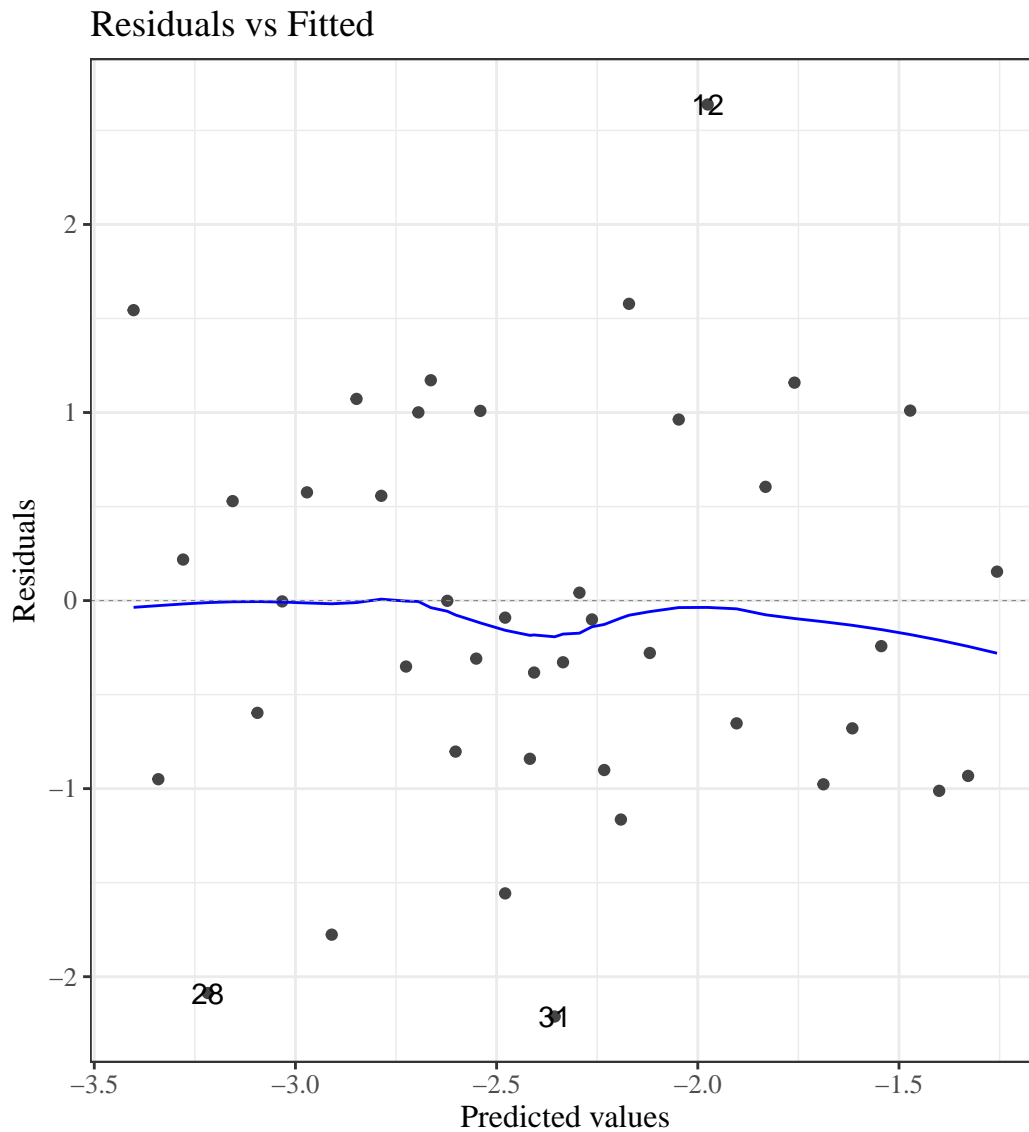$$X_k = \frac{\bar{y}_k - \hat{\pi}_k}{\sqrt{\hat{\pi}_k(1 - \hat{\pi}_k)/n_k}}$$

where

$$
\begin{aligned}
\hat{\pi}_k &\overset{\text{def}}{=} \hat{\pi}(x_k) \\
&\overset{\text{def}}{=} \hat{P}(Y = 1 | X = x_k) \\
&\overset{\text{def}}{=} \text{expit}(x_i'\hat{\beta}) \\
&\overset{\text{def}}{=} \text{expit}(\hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij})
\end{aligned}
$$

Let's take a look at the Pearson residuals for our CHD model from the WCGS data (graphed against the fitted values on the logit scale):

```
library(ggfortify)
```

```
autoplot(chd_glm_contrasts_grouped, which = 1, ncol = 1) |> print()
```
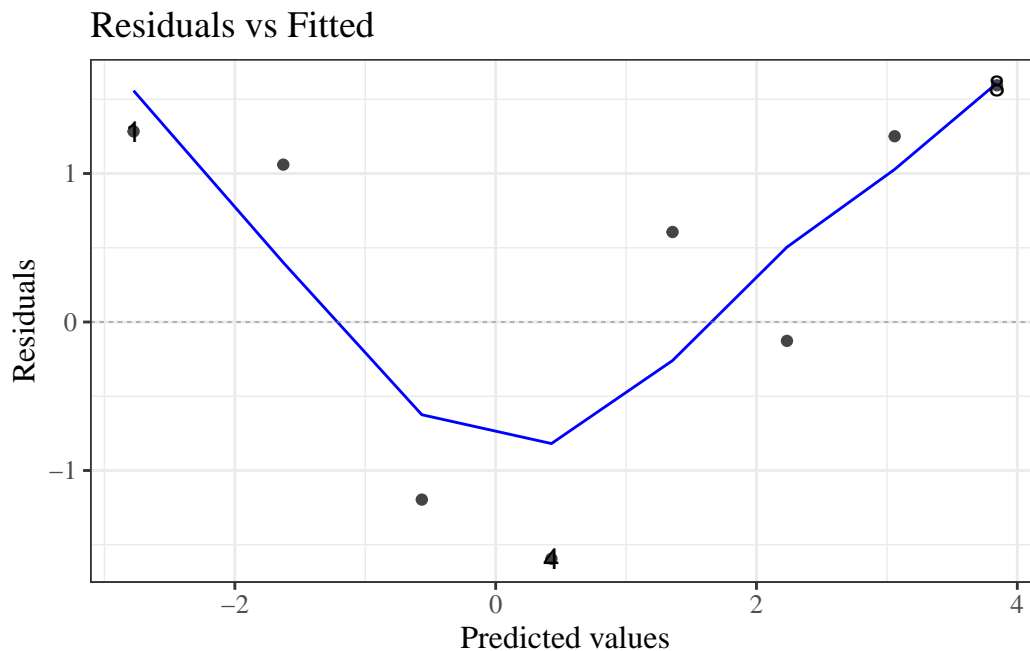

Residuals vs Fitted

The fan-shape is gone, and these residuals don't show any obvious signs of model fit issues.

Pearson residuals plot for `beetles` data

If we create the same plot for the `beetles` model, we see some strong evidence of a lack of fit:

```
library(glmx)
library(dplyr)
data(BeetleMortality)
beetles <- BeetleMortality |>
  mutate(
    pct = died / n,
    survived = n - died,
    dose_c = dose - mean(dose)
  )
beetles_glm_grouped <- beetles |>
  glm(
    formula = cbind(died, survived) ~ dose,
    family = "binomial"
  )
autoplot(beetles_glm_grouped, which = 1, ncol = 1) |> print()
```

Residuals vs Fitted

Pearson residuals with individual (ungrouped) data
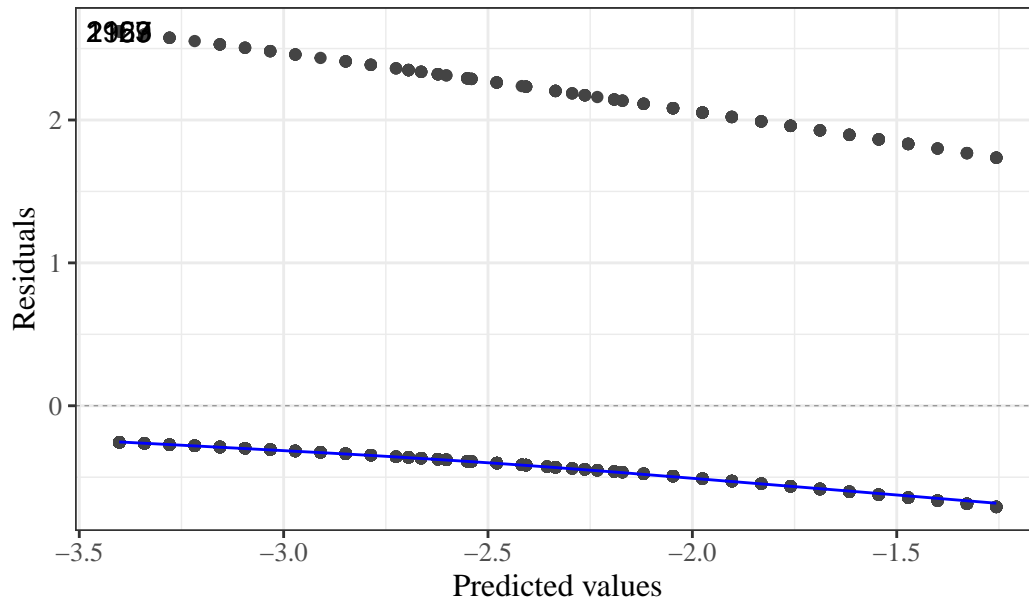
What happens if we try to compute residuals without grouping the data by covariate pattern?

```
library(ggfortify)
```

```
chd_glm_strat <- glm(
  "formula" = chd69 == "Yes" ~ dibpat + dibpat:age - 1,
  "data" = wcgs,
  "family" = binomial(link = "logit")
)

autoplot(chd_glm_strat, which = 1, ncol = 1) |> print()
```

Residuals vs Fitted

Meaningless.

Residuals plot by hand (*optional section*)
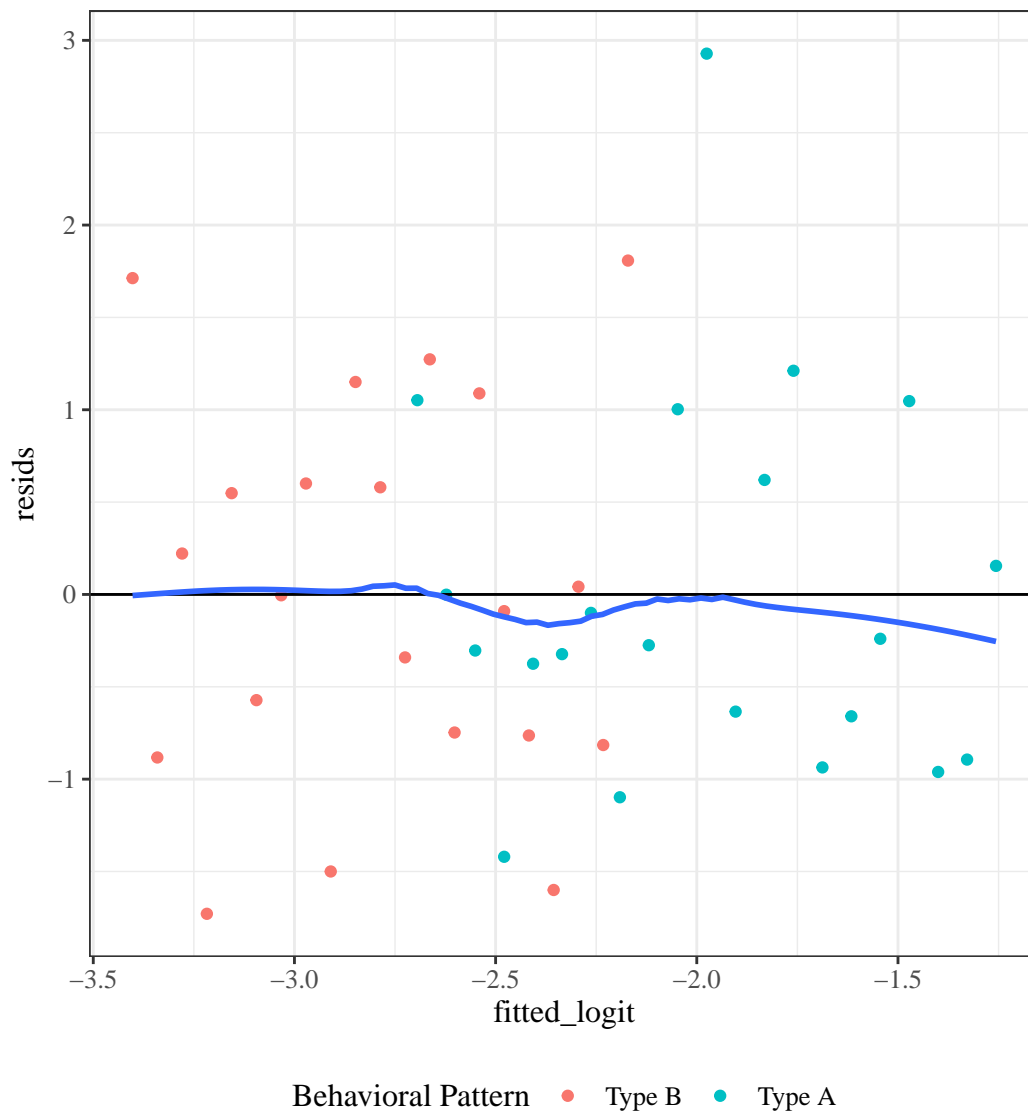
If you want to check your understanding of what these residual plots are, try building them yourself:

```
wcgs_grouped <-
  wcgs_grouped |>
  mutate(
    fitted = chd_glm_contrasts_grouped |> fitted(),
    fitted_logit = fitted |> logit(),
    resids = chd_glm_contrasts_grouped |> resid(type = "pearson")
  )

wcgs_resid_plot1 <-
  wcgs_grouped |>
  ggplot(
    mapping = aes(
      x = fitted_logit,
      y = resids
    )
  ) +
  geom_point(
    aes(col = dibpat)
  ) +
  geom_hline(yintercept = 0) +
  geom_smooth(
    se = FALSE,
    method.args = list(
      span = 2 / 3,
      degree = 1,
      family = "symmetric",
      iterations = 3,
      surface = "direct"
    ),
    method = stats::loess
  )
```

```
# plot.glm and autoplot use stats::lowess, which is hard to use with
# geom_smooth and hard to match exactly;
# see https://support.bioconductor.org/p/2323/
```

```
wcgs_resid_plot1 |> print()
```


1.13.4 Pearson chi-squared goodness of fit test

The Pearson chi-squared goodness of fit statistic is:

$$X^2 = \sum_{k=1}^{m} X_k^2$$

Under the null hypothesis that the model in question is correct (i.e., sufficiently complex), $X^2 \dot\sim \chi^2(N-p)$.

```
x_pearson <- chd_glm_contrasts_grouped |>
  resid(type = "pearson")

chisq_stat <- sum(x_pearson^2)
```

```
pval <- pchisq(
  chisq_stat,
  lower = FALSE,
  df = length(x_pearson) - length(coef(chd_glm_contrasts_grouped))
)
```

For our CHD model, the p-value for this test is 0.265236; no significant evidence of a lack of fit at the 0.05 level.

Standardized Pearson residuals

Especially for small data sets, we might want to adjust our residuals for leverage (since outliers in X add extra variance to the residuals):

$$r_{P_k} = \frac{X_k}{\sqrt{1 - h_k}}$$

where h_k is the leverage of X_k. The functions `autoplot()` and `plot.lm()` use these for some of their graphs.

1.13.5 Deviance residuals

For large sample sizes, the Pearson and deviance residuals will be approximately the same. For small sample sizes, the deviance residuals from covariate patterns with small sample sizes can be unreliable (high variance).

$$d_k = \text{sign}(y_k - n_k \hat{\pi}_k) \left\{ \sqrt{2[\ell_{\text{full}}(x_k) - \ell(\hat{\beta}; x_k)]} \right\}$$

Standardized deviance residuals

$$r_{D_k} = \frac{d_k}{\sqrt{1 - h_k}}$$

1.13.6 Diagnostic plots

Let's take a look at the full set of `autoplot()` diagnostics now for our `CHD` model:

```
chd_glm_contrasts_grouped |>
  autoplot(which = 1:6) |>
  print()
```
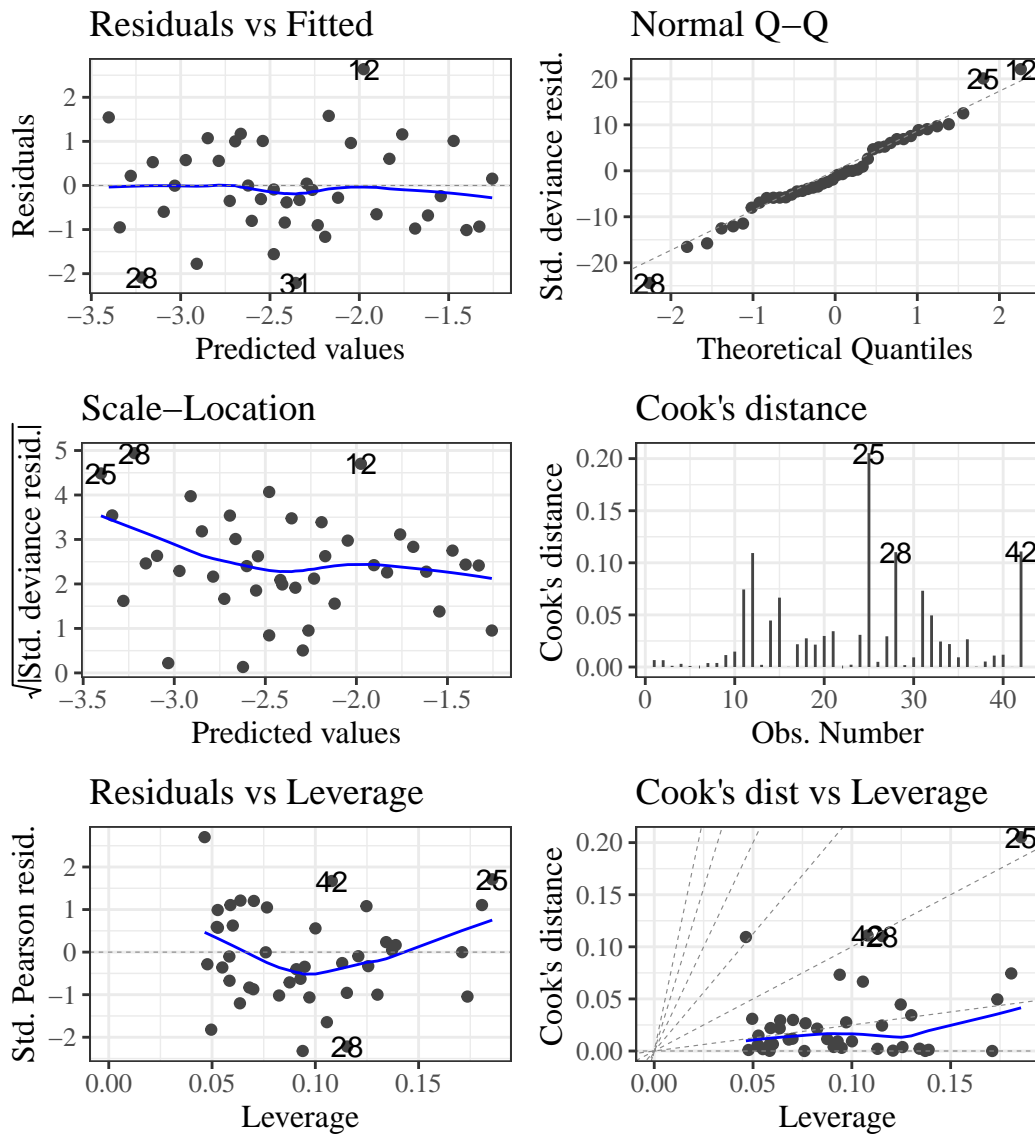
Figure 19: Diagnostics for CHD model

Things look pretty good here. The QQ plot is still usable; with large samples; the residuals should be approximately Gaussian.

Beetles

Let's look at the beetles model diagnostic plots for comparison:

```
beetles_glm_grouped |>
  autoplot(which = 1:6) |>
  print()
```
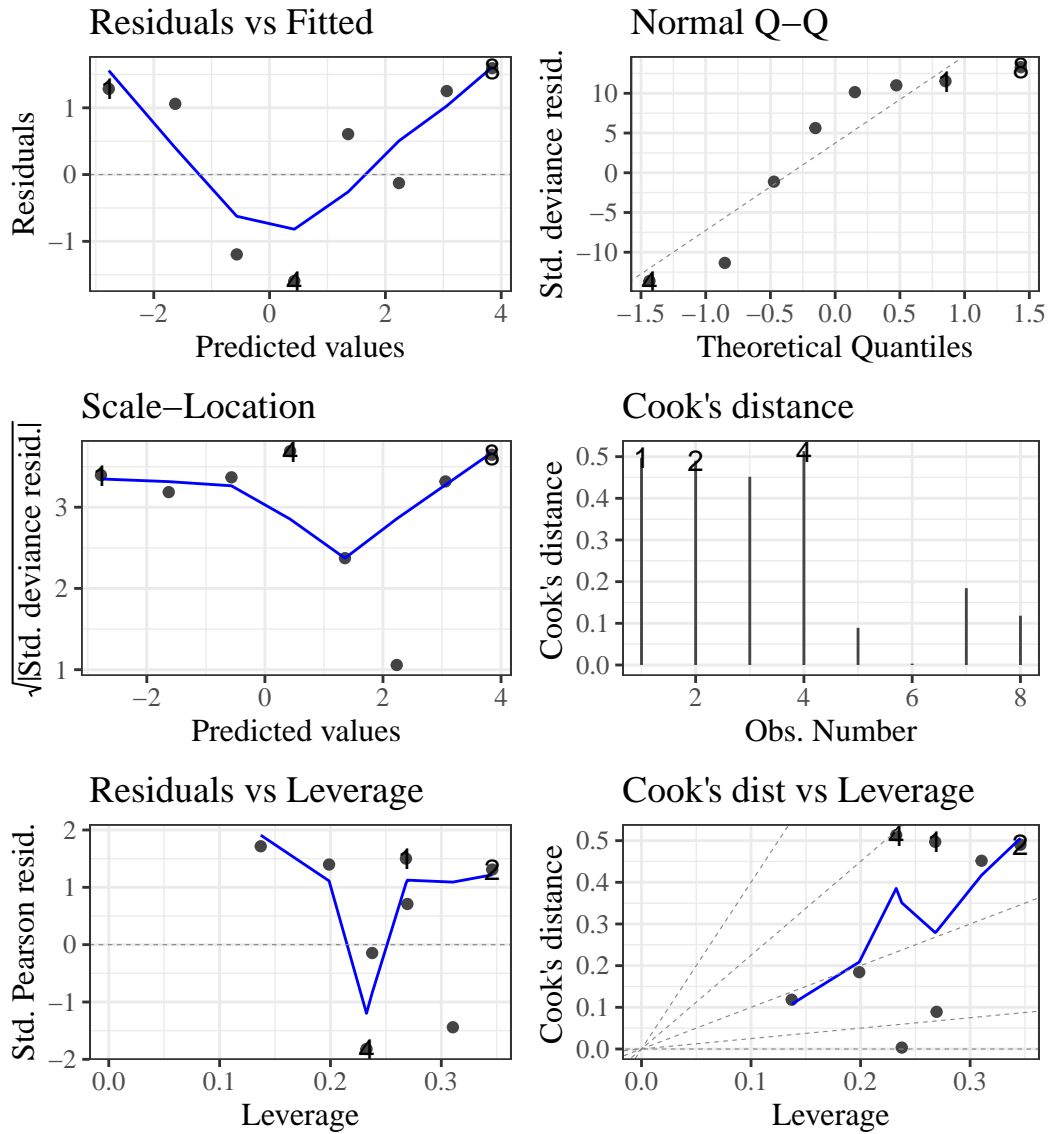
Figure 20: Diagnostics for logistic model of `BeetleMortality` data

Hard to tell much from so little data, but there might be some issues here.

1.14 Alternatives to reporting odds ratios

1.14.1 Objections to odds ratios

Some scholars have raised objections to the use of odds ratios as an effect measurement (Sackett, Deeks, and Altman 1996; Norton et al. 2024).

As we saw in Figure 3, the odds ratio is not very closely correlated with the risk difference, and the risk difference is typically the metric that ultimately matters for policy decisions.

Another objection is that odds ratios (and risk ratios, and risk differences) depend on the set of covariates in a logistic regression model, even when those covariates are independent of the exposure of interest and do not interact with that exposure. For example, consider the following model:

$$P(Y = y | X = x, C = c) = \pi(x, c)^y (1 - \pi(x, c))^{1-y}$$

$$\pi(x, c) = \text{expit}\{\eta_0 + \beta_X x + \beta_C c\}$$

Then:

$$
\begin{aligned}
\text{E}[Y|X = x] &= \text{E}[\text{E}[Y|X, C]|X = x] \\
&= \text{E}[\pi(X, C)|X = x] \\
&= \text{E}[\text{expit}\{\eta_0 + \beta_X X + \beta_C C\}|X = x] \\
&= \int_c \text{expit}\{\eta_0 + \beta_X x + \beta_C c\} \, \text{p}(C = c|X = x) \, \partial c \\
&= \int_c \pi(x, c) \, \text{p}(C = c|X = x) \, \partial c
\end{aligned}
$$

Since the expit{} function is nonlinear, we can't change the order of the expectation and expit{} operators:

$$\text{E}[\text{expit}\{\eta_0 + \beta_X X + \beta_C C\}|X] \neq \text{expit}\{\text{E}[\eta_0 + \beta_X X + \beta_C C]|X\}$$

––––––––––––––––––––––––––––––––––––––

In contrast, consider a model with an identity link function:

$$P(Y = y|X = x, C = c) = \pi(x, c)^y (1 - \pi(x, c))^{1-y}$$

$$\pi(x, c) = \eta_0 + \beta_X x + \beta_C c$$

Then:

$$
\begin{aligned}
\text{E}[Y|X = x] &= \text{E}[\text{E}[Y|X, C]|X = x] \\
&= \text{E}[\eta_0 + \beta_X X + \beta_C C|X = x] \\
&= \text{E}[\eta_0|X = x] + \text{E}[\beta_X X|X = x] + \text{E}[\beta_C C|X = x] \\
&= \eta_0 + \beta_X x + \beta_C \text{E}[C|X = x] \\
&= (\eta_0 + \beta_C \text{E}[C|X = x]) + \beta_X x
\end{aligned}
$$

If $C \perp\!\!\!\perp X$, then $\text{E}[C|X = x] = \text{E}[C]$, and we can simplify further:

$$
\begin{aligned}
\text{E}[Y|X = x] &= (\eta_0 + \beta_C \text{E}[C|X = x]) + \beta_X x \\
&= (\eta_0 + \beta_C \text{E}[C]) + \beta_X x \\
&= \eta_0^* + \beta_X x
\end{aligned}
$$

Then:

$$\frac{\partial}{\partial x} \text{E}[Y|X = x] = \beta_X = \frac{\partial}{\partial x} \text{E}[Y|X = x, C = c]$$

In other words, for a model with an identity link function, if covariates X and C are independent, then the slope with respect to X doesn't depend on whether C is included in the model (and an analogous result holds if X is discrete or categorical).

––––––––––––––––––––––––––––––––––––––

Exercise 1.45. What are the expressions for $\text{E}[Y|X = x]$ and $\frac{\partial}{\partial x}\text{E}[Y|X = x]$ for the model above, if $\text{E}[C|X = x] = \gamma_0 + \gamma_x x$?

––––––––––––––––––––––––––––––––––––––

Solution 1.38. Left to the reader.

———————————————

Exercise 1.46. What are the expressions for $E[Y|X = x]$ and $\frac{\partial}{\partial x}E[Y|X = x]$, if instead of the model above,

$$\pi(x, c) = \eta_0 + \beta_X x + \beta_C c + \beta_{XC} xc$$

and $E[C|X = x] = \gamma_0 + \gamma_x x?$

———————————————

Solution 1.39. Left to the reader.

Hint: does adding the interaction term change the functional form of $E[Y|X = x]$?

1.14.2 Deriving risk ratios and risk differences from logistic regression models

If you want to report risk differences or risk ratios instead of odds ratios, you can obtain estimates from logistic regression models, as long as you didn't stratify sampling by the outcome; in other words, not in case-control studies (see Section 1.4.3).

To compute risk ratios from logistic regression models:

- Apply the expit function to the linear predictor for each covariate pattern to compute the (estimated) risks,
- Then take the differences or ratios of the risks, as needed.

To quantify uncertainty for risk difference or risk ratio estimates derived from logistic regression models (e.g., to calculate SEs, CIs, and p-values), you will need to use the bootstrap, the multivariate delta method, or some other special technique.

1.14.3 Other link functions for Bernoulli outcomes

Alternatively, if you want to estimate risk ratios more directly from the model, you can sometimes change the link function from logit{} to log{}; then you can obtain risk ratios by exponentiating coefficients [16], just like we did for odds ratios with the logit link:

```
data(anthers, package = "dobson")
anthers_sum <- aggregate(
  anthers[c("n", "y")],
  by = anthers[c("storage")], FUN = sum
)

anthers_glm_log <- glm(
  formula = cbind(y, n - y) ~ storage,
  data = anthers_sum,
  family = binomial(link = "log")
)

anthers_glm_log |>
  parameters() |>
  print_md()
```

Parameter	Log-Risk	SE	95% CI	z	p
(Intercept)	-0.80	0.12	(-1.04, -0.58)	-6.81	< .001
storage	0.17	0.07	(0.02, 0.31)	2.31	0.021

———————————————

[16]or linear combinations of coefficients, depending on what covariate patterns you are contrasting

Now $\exp\{\beta\}$ gives us risk ratios instead of odds ratios:

```
anthers_glm_log |>
  parameters(exponentiate = TRUE) |>
  print_md()
```

Parameter	Risk Ratio	SE	95% CI	z	p
(Intercept)	0.45	0.05	(0.35, 0.56)	-6.81	< .001
storage	1.18	0.09	(1.03, 1.36)	2.31	0.021

Let's compare this model with a logistic model:

```
anthers_glm_logit <- glm(
  formula = cbind(y, n - y) ~ storage,
  data = anthers_sum,
  family = binomial(link = "logit")
)

anthers_glm_logit |>
  parameters(exponentiate = TRUE) |>
  print_md()
```

Parameter	Odds Ratio	SE	95% CI	z	p
(Intercept)	0.76	0.20	(0.45, 1.27)	-1.05	0.296
storage	1.49	0.26	(1.06, 2.10)	2.29	0.022

[to add: fitted plots on each outcome scale]

When I try to use `link ="log"` in practice, I often get errors about not finding good starting values for the estimation procedure. This is likely because the model is producing fitted probabilities greater than 1.

When this happens, you can try to fit Poisson regression models instead (we will see those soon!). But then the outcome distribution isn't quite right, and you won't get warnings about fitted probabilities greater than 1. In my opinion, the Poisson model for binary outcomes is confusing and not very appealing.

1.14.4 WCGS: link functions

```
wcgs_glm_logit_link <- chd_grouped_data |>
  mutate(type = relevel(dibpat, ref = "Type B")) |>
  glm(
    "formula" = cbind(x, `n - x`) ~ dibpat * age,
    "data" = _,
    "family" = binomial(link = "logit")
  )

wcgs_glm_identity_link <-
  chd_grouped_data |>
  mutate(type = relevel(dibpat, ref = "Type B")) |>
  glm(
    "formula" = cbind(x, `n - x`) ~ dibpat * age,
    "data" = _,
    "family" = binomial(link = "identity")
```

```
  )
wcgs_glm_identity_link |>
  coef() |>
  pander()
```

(Intercept)	dibpatType A	age	dibpatType A:age
-0.08257	-0.1374	0.002906	0.004194

```
library(ggfortify)
wcgs_glm_logit_link |> autoplot(which = c(1), ncol = 1) + facet_wrap(~dibpat)
wcgs_glm_identity_link |> autoplot(which = c(1), ncol = 1) + facet_wrap(~dibpat)
```


(a) Logistic link   (b) Identity link
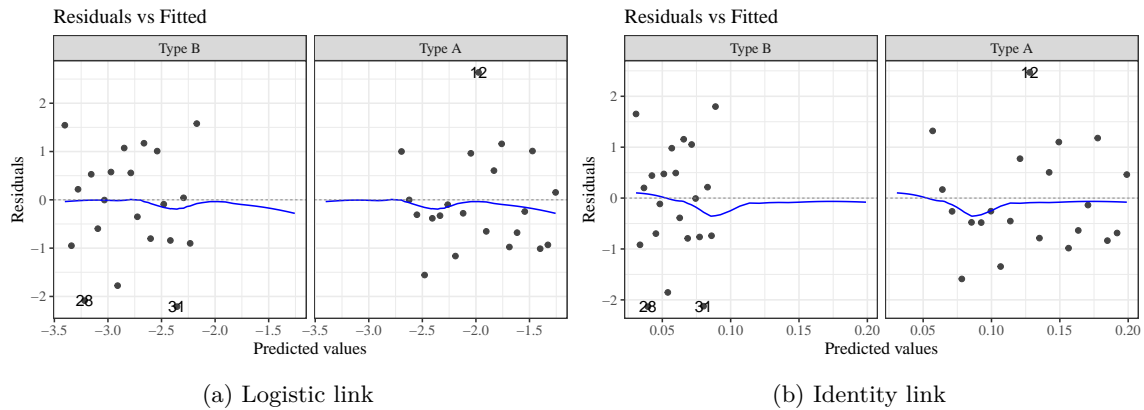
Figure 21: Residuals vs Fitted plot for `wcgs` models

```
beetles_lm <-
  beetles_long |>
  lm(formula = died ~ dose)

beetles_glm_grouped <- beetles |>
  glm(formula = cbind(died, survived) ~ dose, family = "binomial")


beetles <-
  beetles |> mutate(
    resid_logit = beetles_glm_grouped |> resid(type = "response")
  )
beetles_glm_grouped |> autoplot(which = c(1), ncol = 1)
beetles_lm |> autoplot(which = c(1), ncol = 1)
```
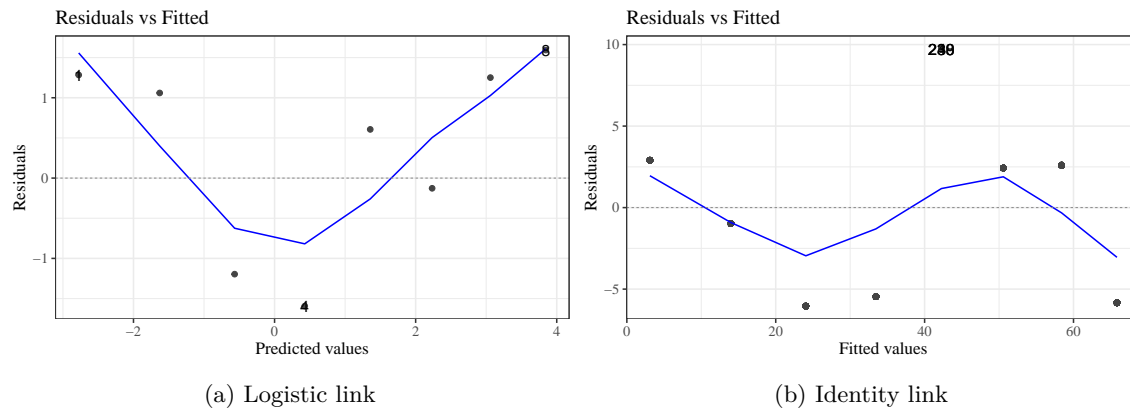
(a) Logistic link

(b) Identity link

Figure 22: Residuals vs Fitted plot for `BeetleMortality` models

1.14.5 Quasibinomial

See Hua Zhou[17]'s lecture notes[18]

1.15 Further reading

- Hosmer, Lemeshow, and Sturdivant (2013) is a classic textbook on logistic regression

Bliss, C. I. 1935. "The Calculation of the Dosage-Mortality Curve." *Annals of Applied Biology* 22 (1): 134–67. https://doi.org/10.1111/j.1744-7348.1935.tb07713.x.

Dobson, Annette J, and Adrian G Barnett. 2018. *An Introduction to Generalized Linear Models.* 4th ed. CRC press. https://doi.org/10.1201/9781315182780.

Hosmer, David W, Stanley Lemeshow, and Rodney X Sturdivant. 2013. *Applied Logistic Regression.* John Wiley & Sons. https://onlinelibrary.wiley.com/doi/book/10.1002/9781118548387.

Lee, James. 1994. "Odds Ratio or Relative Risk for Cross-Sectional Data?" *International Journal of Epidemiology* 23 (1): 201–3. https://doi.org/10.1093/ije/23.1.201.

Lumley, Thomas. 2010. *Complex Surveys : A Guide to Analysis Using R.* Wiley Series in Survey Methodology. Hoboken, N.J: John Wiley. https://doi.org/10.1002/9780470580066.

Nahhas, Ramzi W. 2024. *Introduction to Regression Methods for Public Health Using R.* CRC Press. https://www.bookdown.org/rwnahhas/RMPH/.

Norton, Edward C., Bryan E. Dowd, Melissa M. Garrido, and Matthew L. Maciejewski. 2024. "Requiem for Odds Ratios." *Health Services Research* 59 (4): e14337. https://doi.org/https://doi.org/10.1111/1475-6773.14337.

Rosenman, Ray H, Richard J Brand, C David Jenkins, Meyer Friedman, Reuben Straus, and Moses Wurm. 1975. "Coronary Heart Disease in the Western Collaborative Group Study: Final Follow-up Experience of 8 1/2 Years." *JAMA* 233 (8): 872–77. https://doi.org/10.1001/jama.1975.03260080034016.

Sackett, David L, Jonathan J Deeks, and Doughs G Altman. 1996. "Down with Odds Ratios!" *BMJ Evidence-Based Medicine* 1 (6): 164.

Vittinghoff, Eric, David V Glidden, Stephen C Shiboski, and Charles E McCulloch. 2012. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models.* 2nd ed. Springer. https://doi.org/10.1007/978-1-4614-1353-0.

---

[17]https://hua-zhou.github.io/

[18]https://ucla-biostat-200c-2020spring.github.io/slides/04-binomial/binomial.html#:~:text=0.05%20%27.%27%200.1%20%27%20%27%201-,Quasi%2Dbinomial,-Another%20way%20to