Lecture 12: Modelling Proportions — Dose-Response Experiments
MATH3823 Generalised Linear Models

Richard P Mann

MATH3823 Generalised Linear Models

Reading

Course notes: Chapter 5, Section 5.5

www.richardpmann.com/MATH3823

Dose-Response Studies

Setting:

- Subjects exposed to varying doses of a treatment/substance
- Binary response: success/failure, survival/death
- Goal: Model how response probability depends on dose

Examples:

- Drug efficacy studies
- Toxicology (beetle mortality!)
- Clinical trials
- Pesticide effectiveness

The Beetle Mortality Data (Revisited)

Dose (x_i)	Beetles (m_i)	Killed (y_i)	Proportion
1.6907	59	6	0.10
1.7242	60	13	0.22
1.7552	62	18	0.29
1.7842	56	28	0.50
1.8113	63	52	0.83
1.8369	59	53	0.90
1.8610	62	61	0.98
1.8839	60	60	1.00

Note: Dose is $\log_{10}$(concentration).

Logistic Regression Model

Model:

$$Y_i \sim \text{Binomial}(m_i, p_i), \quad i = 1, \ldots, 8$$

$$\text{logit}(p_i) = \alpha + \beta x_i$$

Probability of death:

$$p_i = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}$$

Interpretation:

- $\beta > 0$: Higher dose $\Rightarrow$ higher mortality
- e^β: Odds ratio per unit increase in dose

R Implementation

```r
# Load data
beetle <- data.frame(
  dose = c(1.6907, 1.7242, 1.7552, 1.7842,
           1.8113, 1.8369, 1.8610, 1.8839),
  died = c(6, 13, 18, 28, 52, 53, 61, 60),
  total = c(59, 60, 62, 56, 63, 59, 62, 60)
)

# Fit logistic regression
y <- cbind(beetle$died, beetle$total - beetle$died)
model <- glm(y ~ dose, data = beetle,
             family = binomial)

summary(model)
```

Model Output

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -60.717      5.181  -11.72   <2e-16 ***
dose          34.270      2.912   11.77   <2e-16 ***

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  11.116  on 6  degrees of freedom
AIC: 41.43
```

Fitted model:

$$\text{logit}(\hat{p}) = -60.72 + 34.27 \times \text{dose}$$

Deviance: 11.12 on 6 df — reasonable fit.

Sequential Hypothesis Tests

Test 1: Null model adequacy

- H_0: All beetles have same mortality rate
- Null deviance: 284.2 on 7 df
- $p < 0.001 \Rightarrow$ Reject (mortality varies)

Test 2: Dose effect

- H_0: $\beta = 0$ (no dose effect)
- $\Delta D = 284.2 - 11.1 = 273.1$ on 1 df
- $p < 0.001 \Rightarrow$ Strong dose effect

Test 3: Model adequacy

- H_0: Logistic dose model is correct
- Residual deviance: 11.1 on 6 df
- $p \approx 0.085$ — marginally adequate

Predictions

Probability at a given dose:

$$\hat{p}(x) = \frac{\exp(-60.72 + 34.27x)}{1 + \exp(-60.72 + 34.27x)}$$

```
# Predict at new doses
new_dose <- data.frame(dose = c(1.7, 1.75, 1.8))
predict(model, newdata = new_dose,
        type = "response")
```

Example predictions:

- Dose 1.70: $\hat{p} = 0.14$
- Dose 1.75: $\hat{p} = 0.36$
- Dose 1.80: $\hat{p} = 0.70$

Finding the Lethal Dose

LD50: Dose at which 50% mortality is expected.

For logistic model:

$$\text{logit}(0.5) = 0 = \alpha + \beta x_{50}$$

Solving:

$$\boxed{x_{50} = -\frac{\alpha}{\beta} = -\frac{-60.72}{34.27} = 1.772}$$

General formula for LDp (dose for mortality p):

$$x_p = \frac{\text{logit}(p) - \alpha}{\beta} = \frac{1}{\beta}\left(\log\frac{p}{1-p} - \alpha\right)$$

LD Values in R

```r
# Extract coefficients
alpha <- coef(model)[1]
beta <- coef(model)[2]

# LD50
LD50 <- -alpha / beta
LD50  # 1.772

# LD90 (90% mortality)
LD90 <- (log(0.9/0.1) - alpha) / beta
LD90  # 1.836

# LD10 (10% mortality)
LD10 <- (log(0.1/0.9) - alpha) / beta
LD10  # 1.708
```

Alternative Link Functions

Try different links for beetle data:

```r
# Logit (canonical)
m1 <- glm(y ~ dose, family = binomial(link = "logit"))

# Probit
m2 <- glm(y ~ dose, family = binomial(link = "probit"))

# Complementary log-log
m3 <- glm(y ~ dose, family = binomial(link = "cloglog"))
```

Compare deviances:

Link	Deviance	AIC
Logit	11.12	41.43
Probit	10.12	40.43
Cloglog	3.45	33.76

Residual Analysis

```r
# Deviance residuals
residuals(model, type = "deviance")

# Plot residuals vs fitted
plot(fitted(model), residuals(model, type = "deviance"),
     xlab = "Fitted values", ylab = "Deviance residuals")
abline(h = 0, lty = 2)
```
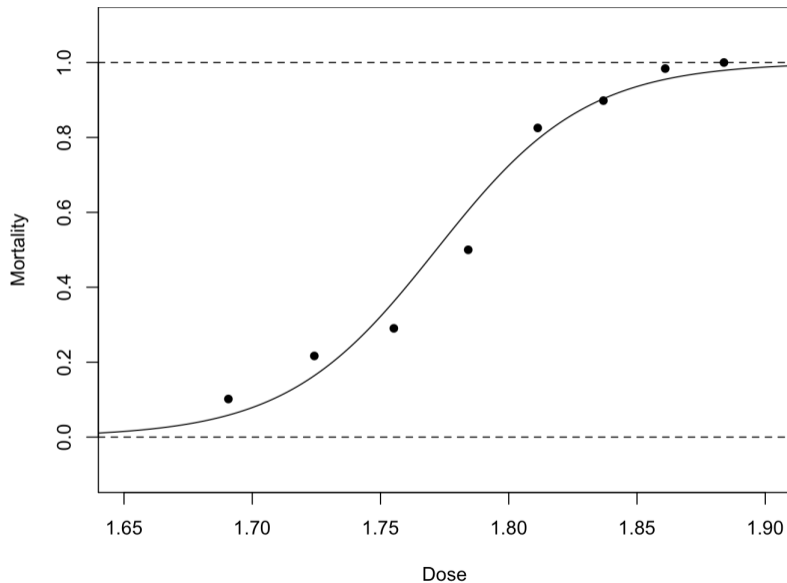
What to look for:

- Random scatter around 0

- No systematic patterns

- U-shaped pattern $\Rightarrow$ try different link

Visualizing the Fitted Model

```r
# Plot observed proportions
plot(beetle$dose, beetle$died/beetle$total,
     xlab = "Dose", ylab = "Mortality",
     pch = 16, ylim = c(0, 1))

# Add fitted curve
dose_seq <- seq(1.68, 1.90, length = 100)
pred <- predict(model,
                newdata = data.frame(dose = dose_seq),
                type = "response")
lines(dose_seq, pred, col = "blue", lwd = 2)
```

The Fitted Dose-Response Curve

Summary

Key points:
- Dose-response: model p(response) vs. dose
- Logistic regression: $\text{logit}(p) = \alpha + \beta x$
- β measures effect of dose on log-odds
- LD50 $= -\alpha/\beta$ (dose for 50% response)
- General: $x_p = (\text{logit}(p) - \alpha)/\beta$
- Compare link functions via deviance/AIC
- Check residuals for model adequacy

Next lecture: Loglinear models for count data.