Binary and Binomial Response Data

Binary Data

For each observation i, the response Y_i can take only two values, coded as 1 and 0,

yes/no, success/failure etc

Assume π_i is the success probability for observation i, Y_i has a Bernoulli distribution:

$$Y_i \sim \text{Bern}(\pi_i)$$

Binomial Data

Each observation i is a count of Y_i successes out of m_i trials. Assume π_i is the success probability for observation i, Y_i has a Binomial distribution:

$$Y_i \sim \text{Bin}(m_i, \pi_i)$$

$$\text{Bin}(1, \pi_i) \equiv \text{Bern}(\pi_i)$$

Binomial Regression Models

For data of the form: Y_i successes out of m_i, $i = 1, \ldots, n$ trials, assume

$$Y_i \sim \text{Bin}(m_i, \pi_i)$$

with mean and variance:

$$\mathbf{E}[Y_i] = \mu_i = m_i \pi_i \quad e \quad \text{Var}(Y_i) = m_i \pi_i (1 - \pi_i).$$

- Want to model π_i as a function of explanatory variables,
- Want to specify the correct distribution for maximum likelihood estimation

  Note: the variance is not constant.

Model π_i as a linear function of the explanatory variables

$$\pi_i = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

may give fitted values for π_i outside the range $[0, 1]$.

Solution is to transform the success probability. If $F(\theta)$ is a strictly increasing function of θ such that

$$\lim_{\theta \to -\infty} F(\theta) = 0 \quad \text{e} \quad \lim_{\theta \to \infty} F(\theta) = 1$$

the F defines a transformation from $(-\infty, \infty)$ to $[0, 1]$.

$F(.)$ can be any cumulative distribution function for a random variable on $(-\infty, \infty)$,

for example, N$(0, 1)$ distribution $F(.) = \Phi(.)$.

Define the **linear preditor** $\eta_i = \boldsymbol{\beta}' \mathbf{x}_i$, then, $\pi_i = F(\eta_i)$ and

$$\mathbf{E}[Y_i] = \mu_i = m_i \pi_i = m_i F(\eta_i).$$

The inverse of $F(.)$ is called **link function**

$$g(.) = F^{-1}(.)$$

$$g(\pi_i) = \eta_i$$

Examples

Logit link: the canonical link

$$g(\mu_i) = \log\left(\frac{\mu_i}{m_i - \mu_i}\right) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i$$

$$\pi_i = F(\eta_i) = \frac{e^{\eta_i}}{1+e^{\eta_i}}$$

$F(.)$ is the cumulative distribution function for logistic distribution.

Probit link

$$g(\mu_i) = \Phi^{-1}(\mu_i/m_i) = \Phi^{-1}(\pi_i)$$

$$\pi_i = F(\eta_i) = \Phi(\eta_i)$$

$F(.)$ is the cumulative distribution function for $N(0,1)$ distribution.

Complementary log-log (CLL)

$$g(\mu_i) = \log[-\log(1-\pi_i)].$$

$$\pi_i = F(\eta_i) = 1 - e^{-e^{\eta_i}}$$

$F(.)$ is the cumulative distribution function for extreme value (Gumbel) distribution.

Specification of Binomial distribution in R

Binomial distribution - depends on the number of trials m_i.

```
response <- cbind(y, m-y)
glm(response~linear predictor,
family=binomial(link=cloglog))
```

For binary data $m_i = 1$, for all i $(Y_i = 0, 1)$

```
m <- rep(1, nobs)
```

Link function: The most common are

- logit – logistic cdf (*default*), $\log(\pi_i/(1-\pi_i)) = \eta_i$
- probit – normal cdf, $\Phi^{-1}(\pi_i) = \eta_i$
- cloglog – complementary log-log, extreme value cdf, $\log(-\log(1-\pi_i)) = \eta_i$

Other links may be specified:

- log – $\log(\pi_i) = \eta_i$
- cauchit – Cauchy cdf

or, you can define your own, using family=quasi

Analysis of Quantal Bioassay Data

- Bioassay or biological assay – the measurement of the potency of a stimulus by means of the reactions it produces in living matter
- The stimulus may be physiological, chemical, biological, phisiological etc. Examples:
  - insects exposed to chemical stimuli (insecticides or biological stimuli (eg a virus)
  - other agricultural bioassays may involve large animals, fungi, plants or plant parts such as leaves.
- The response here is quantal or binary – just 2 possible responses. Examples:
  - an insect dies or survives
  - a seed germinates or fails to germinate
  - a cutting roots or fails to root

- The tolerance of an individual (eg an insect) is the dose or concentration or intensity of the stimulus above which the individual responds (eg dies) and below which it does not respond.

- Tolerance varies between individuals – it is a random variable.

- Let
  τ: tolerance of a randomly chosen individual;
  $f(\tau)$: the probability density function of τ;
  x: dose of the stimulus.

- Then, for an individual chosen at random from the population

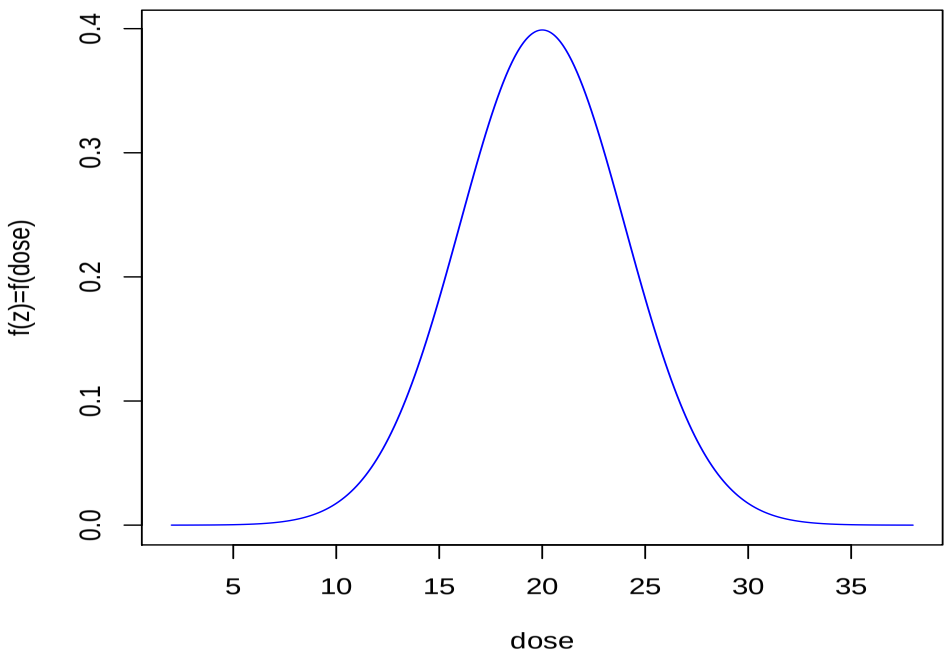
$$Pr(\text{response}|x) = Pr(\tau < x) = \int_0^x f(\tau)d\tau$$

- In practice it is often reasonable to assume that tolerance have a log-normal distribution, that is,

$$\log(\tau) \sim \mathsf{N}(\mu, \sigma^2)$$

$$\begin{aligned} \pi(x) &= Pr(\text{death}|x) = Pr(\tau < x) \\ &= Pr(\log(\tau) < \log(x)) = \Phi\left[\frac{\log(x) - \mu}{\sigma}\right] \end{aligned}$$

that is, $\Phi^{-1}(\pi(x)) = \alpha + \beta \log(x)$ where $\alpha = -\frac{\mu}{\sigma}$ and $\beta = \frac{1}{\sigma}$.



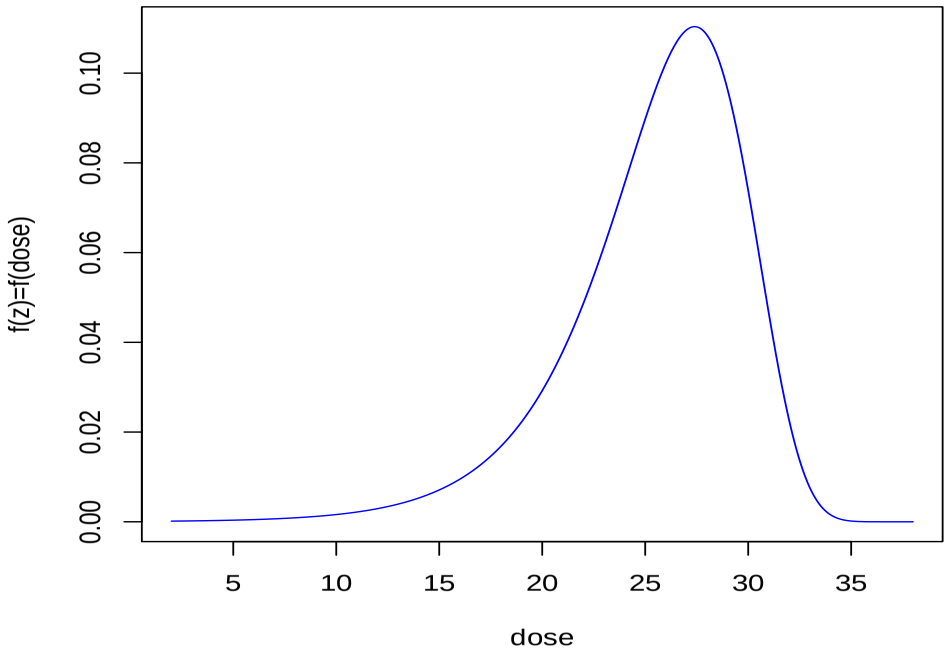
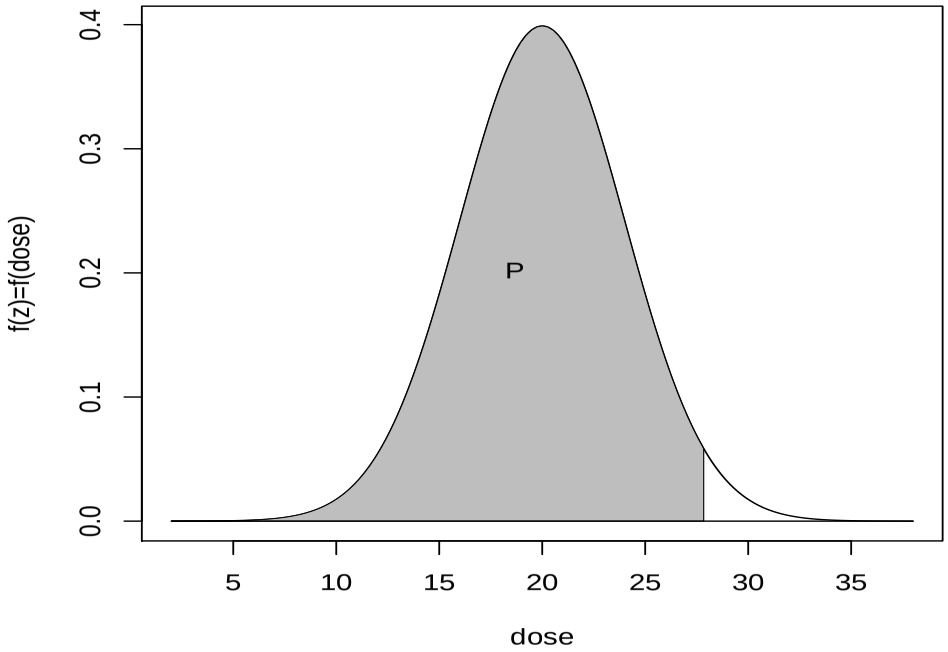


Figure 1: Distribuições de tolerância



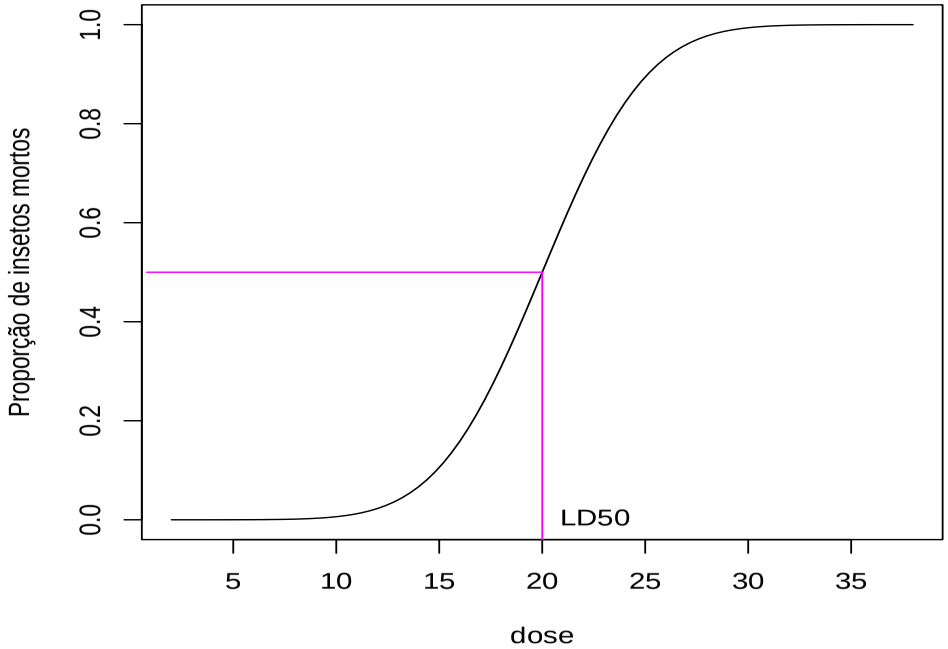


Figure 2: Distribuições acumuladas

- The function $\Phi^{-1}(.)$, the inverse of the cumulative normal function, is called the **probit** transformation.
- An alternative transformation, which in practice usually gives very similar results, is the **logit** transformation

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

- This is the inverse of the cumulative distribution of the logistic distribution.
- A different transformation is the **complementary log-log**

$$\text{clolog}(\pi) = \log(-\log(1-\pi))$$

- This is the inverse of the cumulative distribution of the extreme value (Gumbel) distribution which is asymetric.
- The exposed theory shows one type of biological justification for the link function.

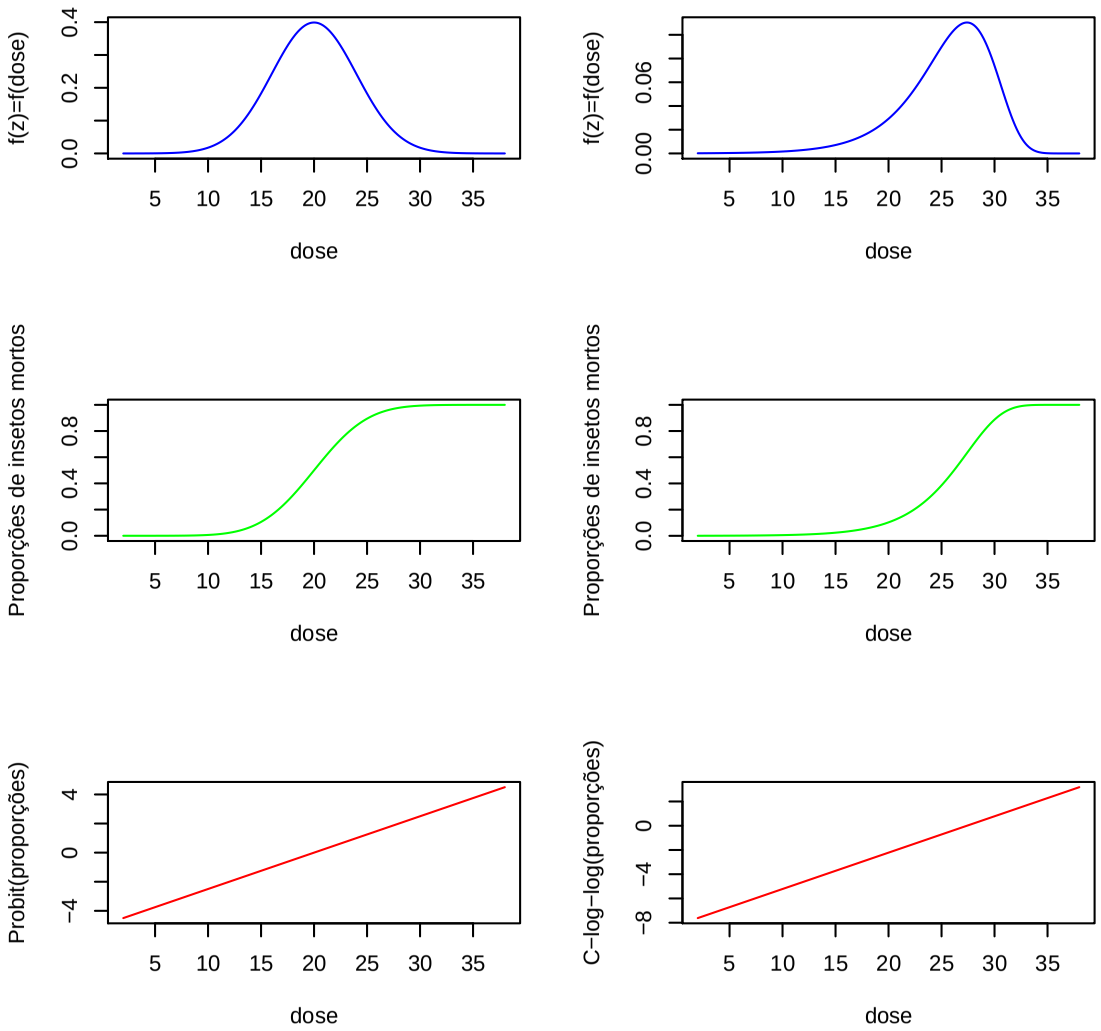
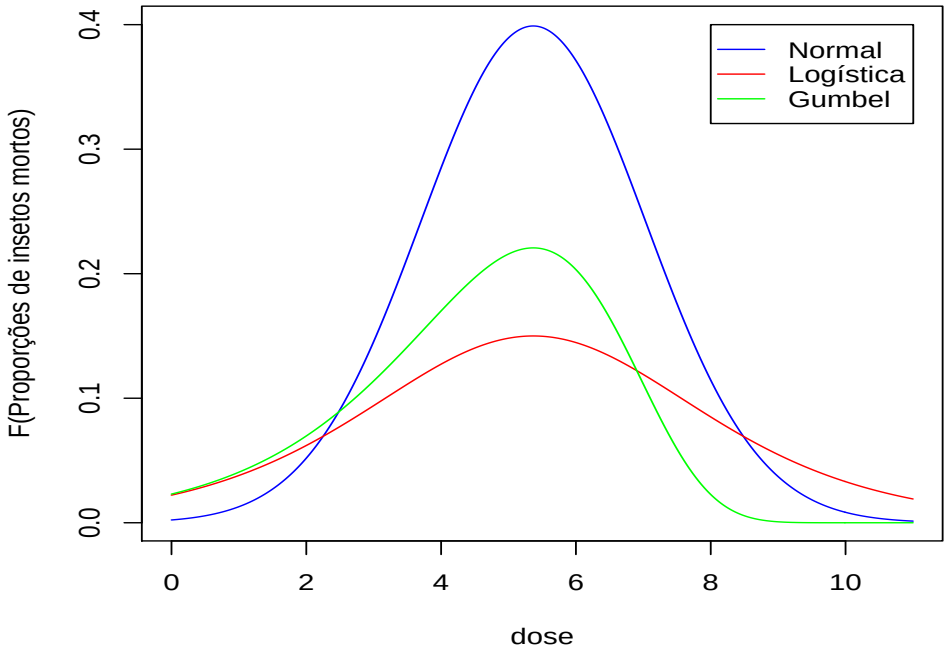


Figure 3: Distribuições de tolerância, sigmóides e retas



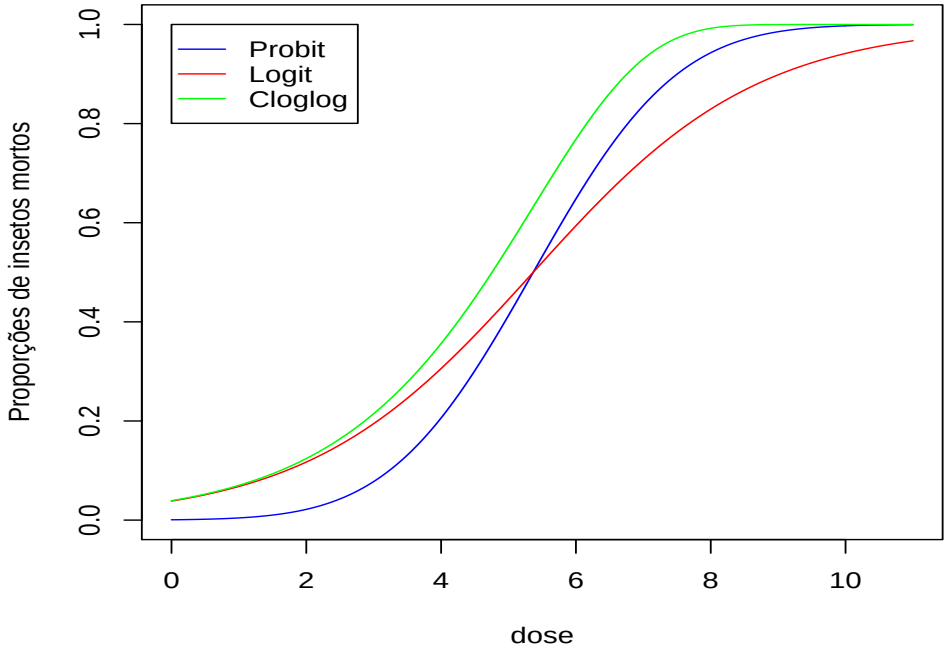


Figure 4: Distribuições de tolerância e sigmóides

- In a typical insect bioassay, there are k batches of insects.
- The ith batch consists of m_i insects which receive a dose x_i of the insecticide.
- Y_i of the m_i insects are killed.
- usually we assume independence between individual insects that receive the same dose, as well as between insects that receive different doses. It follows that

$$Y_i \sim \text{Binomial}(m_i, \pi_i), \text{ where } \pi_i = \pi(x_i)$$

- The likelihood function is:

$$L = \prod_{i=1}^{k} \binom{m}{y} \pi_i^{y_i} (1 - \pi_i)^{m_i - y_i}$$

where, in general $\pi_i = F^{-1}(\alpha + \beta z_i)$ and $F^{-1}(.)$ is the inverse of a cumulative function (eg the logit, probit or complementary log-log transformation) and z_i is called the dose-metameter. This may be the same as the dose ($z_i = x_i$) or it may be a transformation of the dose (eg $z_i = \log(x_i)$).

Rotenone toxicity

Toxicity of Rotenone to *Macrosiphoniela sanborni*, the chrysanthemum aphis.

Dose (d_i)	m_i	y_i
0.0	49	0
2.6	50	6
3.8	48	16
5.1	46	24
7.7	49	42
10.2	50	44

Considerations

- Response variable: Y_i – number of dead insects out of m_i insects (Martin, 1942).
- Distribution: Binomial.
- Systematic component: regression model, completely randomized experiment.
- Aim: Lethal doses.

Estimation of lethal or effective doses

- Response probability modelled as a function of a single explanatory variable x.
- Effective dose, $\hat{\theta}_p$ – the concentration of a chemical that is expected to produce a response in 100p% of the individuals exposed to it.
- LD = "lethal dose", the response is death. $LD_{50\%}$, median dose, expected to kill 50% of the individuals.

Logit:
$\text{logit}(p) = \log \dfrac{p}{1-p} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\theta}_p \Rightarrow \hat{\theta}_p = \dfrac{1}{\hat{\beta}_1}(\log \dfrac{p}{1-p} - \hat{\beta}_0)$

Probit:
$\text{probit}(p) = \Phi^{-1}(p) = \hat{\beta}_0 + \hat{\beta}_1 \hat{\theta}_p \Rightarrow \hat{\theta}_p = \dfrac{1}{\hat{\beta}_1}[\Phi^{-1}(p) - \hat{\beta}_0]$

Complementary log-log:
$\log[-\log(1-p)] = \hat{\beta}_0 + \hat{\beta}_1 \hat{\theta}_p \Rightarrow$
$\hat{\theta}_p = \dfrac{1}{\hat{\beta}_1}\{\log[-\log(1-p)] - \hat{\beta}_0\}$

Particular case

$\hat{\theta}_{50} = -\dfrac{\hat{\beta}_0}{\hat{\beta}_1}$ logit and probit,

$\hat{\theta}_{50} = \dfrac{\log(\log 2) - \hat{\beta}_0}{\hat{\beta}_1}$ clolog

Note: $\hat{\theta}_p = \log \hat{d}_p \Rightarrow \hat{d}_p = e^{\hat{\theta}_p}$

Confidence interval for lethal or effective doses

Delta method

- $\hat{\boldsymbol{\theta}} \sim \mathsf{N}(\boldsymbol{\theta}, \Sigma)$ where Σ is the variance-covariance matrix of the parameter estimates (inverse of Fisher's information matrix).
- $\hat{\theta}_p = \dfrac{F^{-1}(p) - \hat{\beta}_0}{\hat{\beta}_1} = g(\hat{\beta}_0, \hat{\beta}_1) = g(\hat{\boldsymbol{\theta}}) =$
  $g(\beta_0, \beta_1) + (\hat{\beta}_0 - \beta_0)\dfrac{\partial g}{\partial \beta_0} + (\hat{\beta}_1 - \beta_1)\dfrac{\partial g}{\partial \beta_1}$
- $g(\hat{\boldsymbol{\theta}}) \sim \mathsf{N}(g(\boldsymbol{\theta}), \boldsymbol{\gamma}^T \Sigma \boldsymbol{\gamma})$, where
  $\boldsymbol{\gamma}^T = \left(\frac{\partial g}{\partial \beta_0}, \frac{\partial g}{\partial \beta_1}\right) = \left(-\frac{1}{\beta_1}, -\frac{F^{-1}(p) - \hat{\beta}_0}{\beta_1^2}\right)$
- $$\Sigma = \widehat{\mathsf{Var}}(\hat{\theta}_p) = \frac{1}{\hat{\beta}_1^2}[\widehat{\mathsf{Var}}(\hat{\beta}_0) + \hat{\theta}_p^2 \widehat{\mathsf{Var}}(\hat{\beta}_1) + 2\hat{\theta}_p \widehat{\mathsf{Cov}}(\hat{\beta}_0, \hat{\beta}_1)]$$
- $CI(\hat{\theta}_p) : \hat{\theta}_p \mp z_{\alpha/2}\sqrt{\widehat{\mathsf{Var}}(\hat{\theta}_p)}$
- Always Symmetric!!

- negative lower limit is possible - replace by zero
- alternative is to use a confidence interval for $\log(\hat{\theta}_p)$

$$CI(\log(\hat{\theta}_p)) : \log(\hat{\theta}_p) \mp z_{\alpha/2} s.e.(\log(\hat{\theta}_p))$$

  where $s.e.(\log(\hat{\theta}_p)) \approx (\hat{\theta}_p^{-1}) s.e.(\hat{\theta}_p)$, and the $CI(\hat{\theta}_p)$ follows by exponentiating the limits
- not generally recommended - problem with the assumed normal distribution
- explanotory variable used is log(dose)

  $s.e.\big(\log(\hat{\theta}_p)\big) \approx \frac{1}{\beta_1^2}[\widehat{\mathsf{Var}}(\hat{\beta}_0) + \hat{\theta}^2 \widehat{\mathsf{Var}}(\hat{\beta}_1) + 2\hat{\theta}\widehat{\mathsf{Cov}}(\hat{\beta}_0, \hat{\beta}_1)]^{1/2}$

  and $s.e.(\hat{\theta}_p) \approx (\hat{\theta}_p) s.e.(\log(\hat{\theta}_p))$, and the $CI(\hat{\theta}_p)$

Fieller's theorem

- General theorem about ratios of (possibly correlated) normally distributed random variables.
- $\hat{\theta}_p = \dfrac{F^{-1}(p) - \hat{\beta}_0}{\hat{\beta}_1} = g(\hat{\beta}_0, \hat{\beta}_1)$

  where $\hat{\beta}_0 \sim \mathsf{N}(\beta_0, \mathsf{Var}(\hat{\beta}_0))$,
  $\hat{\beta}_1 \sim \mathsf{N}(\beta_1, \mathsf{Var}(\hat{\beta}_1))$ and $\mathsf{Cov}(\hat{\beta}_0, \hat{\beta}_1)$.
- $\hat{\beta}_1\theta_p + \hat{\beta}_0 - F^{-1}(p) \sim$
  $\mathsf{N}(0, \mathsf{Var}(\hat{\beta}_0) + 2\theta_p\mathsf{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \theta_p^2\mathsf{Var}(\hat{\beta}_1))$
- $\dfrac{\hat{\beta}_1\theta_p + \hat{\beta}_0 - F^{-1}(p)}{\sqrt{\mathsf{Var}(\hat{\beta}_0) + 2\theta_p\mathsf{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \theta_p^2\mathsf{Var}(\hat{\beta}_1))}} \sim \mathsf{N}(0, 1)$
- $CI(\hat{\theta}_p)$: solution of

$$\frac{[\hat{\beta}_1\theta_p + \hat{\beta}_0 - F^{-1}(p)]^2}{\mathsf{Var}(\hat{\beta}_0) + 2\theta_p\mathsf{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \theta_p^2\mathsf{Var}(\hat{\beta}_1))} < z^2_{\alpha/2}$$

- may not give an interval
- similar results as delta method under conditions
- an alternative – likelihood ratio interval

Rotenone Data Analysis

Model	d.f.	Deviance	X^2
Constant	5	163.7	135.7
Dose	4	10.2	9.7

Analysis of deviance

Source	d.f.	Deviance	p-value
Linear Regression	1	153.5	< 0.0001
Error	4	10.2	0.0527
Total	5	163.7	

$$log\frac{\hat{p}}{1-\hat{p}} = -3,226 + 0,6051dose$$

$$\hat{p} = \frac{e^{-3.226+0.6051dose}}{1+e^{-3.226+0.6051dose}}$$

$$\hat{LD}_{50} = \frac{3.226}{0.6051} = 5.33$$

$$IC[LD_{50}]_{95\%} : (4.8; 5.9)$$

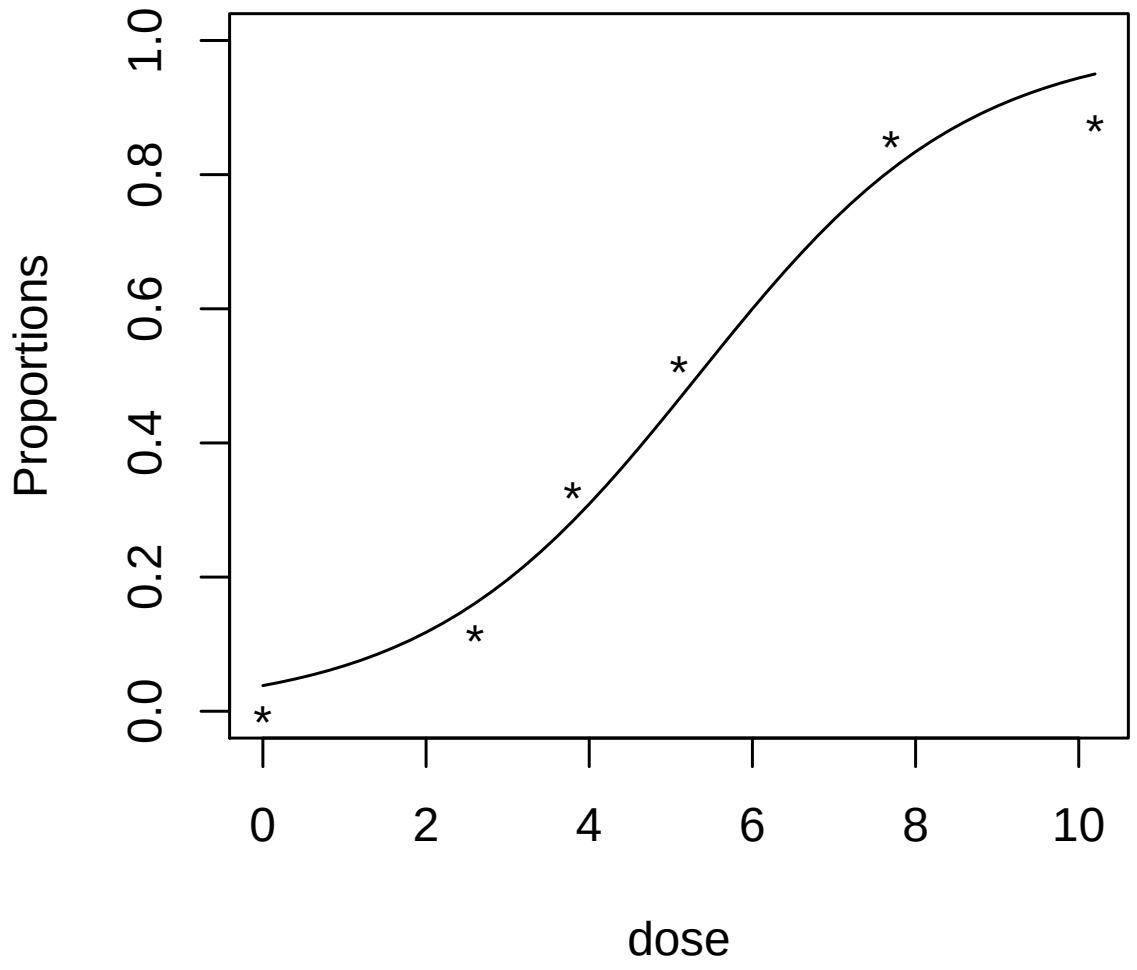


Figure 5: Rotenone - Observed values and fitted curve

Rotenone analysis in R

```
dose <- c(0,2.6,3.8,5.1,7.7,10.2)
y <- c(0,6,16,24,42,44)
m <- c(49,50,48,46,49,50)
Rotenone.dat <- data.frame(dose, y, m)
attach(Rotenone.dat)
plot(dose,y/m, xlab="Dose", ylab="Observed proportions", pch="*")
resp<-cbind(y,m-y)

Rotenon1<-glm(resp~1, family=binomial)
Rotenon2<-glm(resp~dose, family=binomial)
summary(Rotenon2)
anova(Rotenon1, Rotenon2, test="Chisq")

plot(c(0,10.2), c(0,1), type="n", xlab="dose", ylab="Proportions")
points(dose,y/m,pch="*")
x<-seq(0,10.2,0.2)
lp<-predict(Rotenon2,data.frame(dose=x))
pe<-exp(lp)/(1+exp(lp))
lines(x,pe,lty=1)
title(sub="Observed values and fitted curve")

library(MASS)
# variance-covariance matrix of estimated parameters
vcov(Rotenon2)
# Doses LD50 and LD90
dose.p(Rotenon2); dose.p(Rotenon2, p=0.9)
# Doses LD25, LD50, LD75
dose.p(Rotenon2, p = 1:3/4)
```

Exemplo: Toxicity of cypermethrin to moths

Batches of 20 pyrethroid-resistant moths (*Heliothis virescens*, a cotton crop pest) of each sex were exposed to a range of doses of cypermethrin two days after emergence from pupation. The number of moths which were either knocked down or dead was recorded after 72h.

Doses (d_i)	Males	Females
1.0	1	0
2.0	4	2
4.0	9	6
8.0	13	10
16.0	18	12
32.0	20	16

Considerations

- Response variable: Y_i – number of dead insects out of 20 insects.
- Distribution: Binomial.
- Systematic component: regression model, completely randomized experiment.
- Aim: Lethal doses.

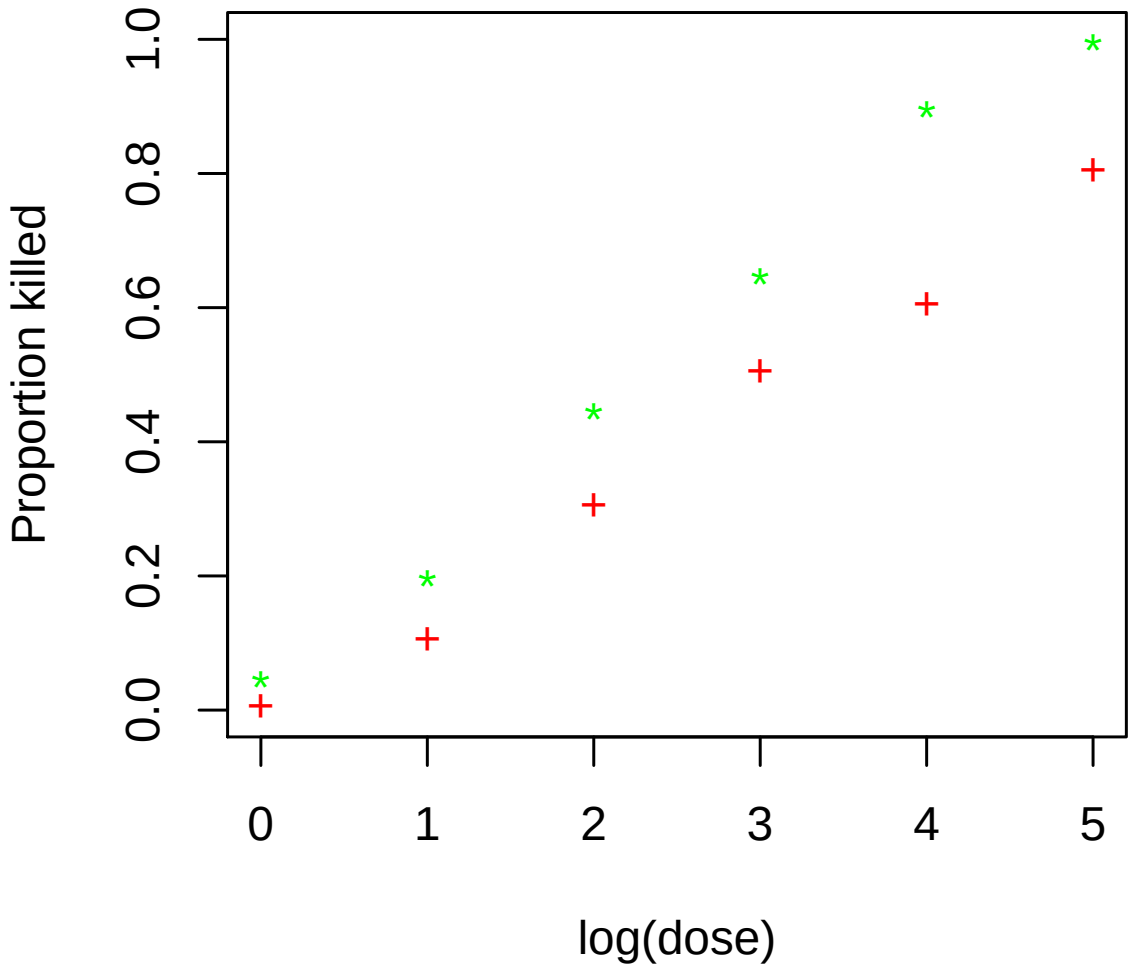


Figure 6: Cypermethrin toxicity - Scatterplot

Model: Estimates and Tests

Cypermethrin toxicity data - deviance residuals

Terms fitted in model	d.f.	Deviance	X^2
Constant	11	124.9	101.4
Sex	10	118.8	97.4
Dose	6	15.2	12.9
Sex + Dose	5	5.0	3.7

Cypermethrin toxicity data - Analysis of deviance

Source	d.f.	Deviance	p-value
Sex	1	6.1	0.0144
Sex|Dose	1	10.1	0.0002
Dose	5	109.7	< 0.0001
Dose|Sex	5	113.8	< 0.0001
Residual	5	5.0	0.5841
Total	11	124.9	

Models

logit$(p) = \alpha_j + \beta_j \ \log(\text{dose})$ – different logistic regression lines

logit$(p) = \alpha_j + \beta \ \log(\text{dose})$ – common slope

logit$(p) = \alpha + \beta_j \ \log(\text{dose})$ – common intercept

logit$(p) = \alpha + \beta \ \log(\text{dose})$ – same logistic regression line.

Residual Deviances

Terms fitted in model	d.f.	Deviance	X^2
Constant	11	124.9	101.4
Sex + Sex log(dose)	8	4.99	3.51
Sex + log(dose)	9	6.75	5.31
Const. + Sex log(dose)	9	5.04	3.50
Const. + log(dose)	10	16.98	14.76

Analysis of deviance

Source	d.f.	Deviance	p-value
Sex	1	6.1	0.0144
Linear Regression	1	112.0	< 0.0001
Residual	9	6.8	0.7473
Total	11	124.9	

Regression equations

males

$$\log \frac{\hat{p}}{1-\hat{p}} = -2.372 + 1.535 \log(\text{dose})$$

females

$$\log \frac{\hat{p}}{1-\hat{p}} = -3.473 + 1.535 \log(\text{dose})$$

Lethal Doses

males

$$\log(\hat{LD}_{50}) = \frac{2.372}{1.535} = 1.55 \Rightarrow LD_{50} = 4.69$$

females

$$\log(\hat{LD}_{50}) = \frac{3.473}{1.535} = 2.26 \Rightarrow LD_{50} = 9.61$$

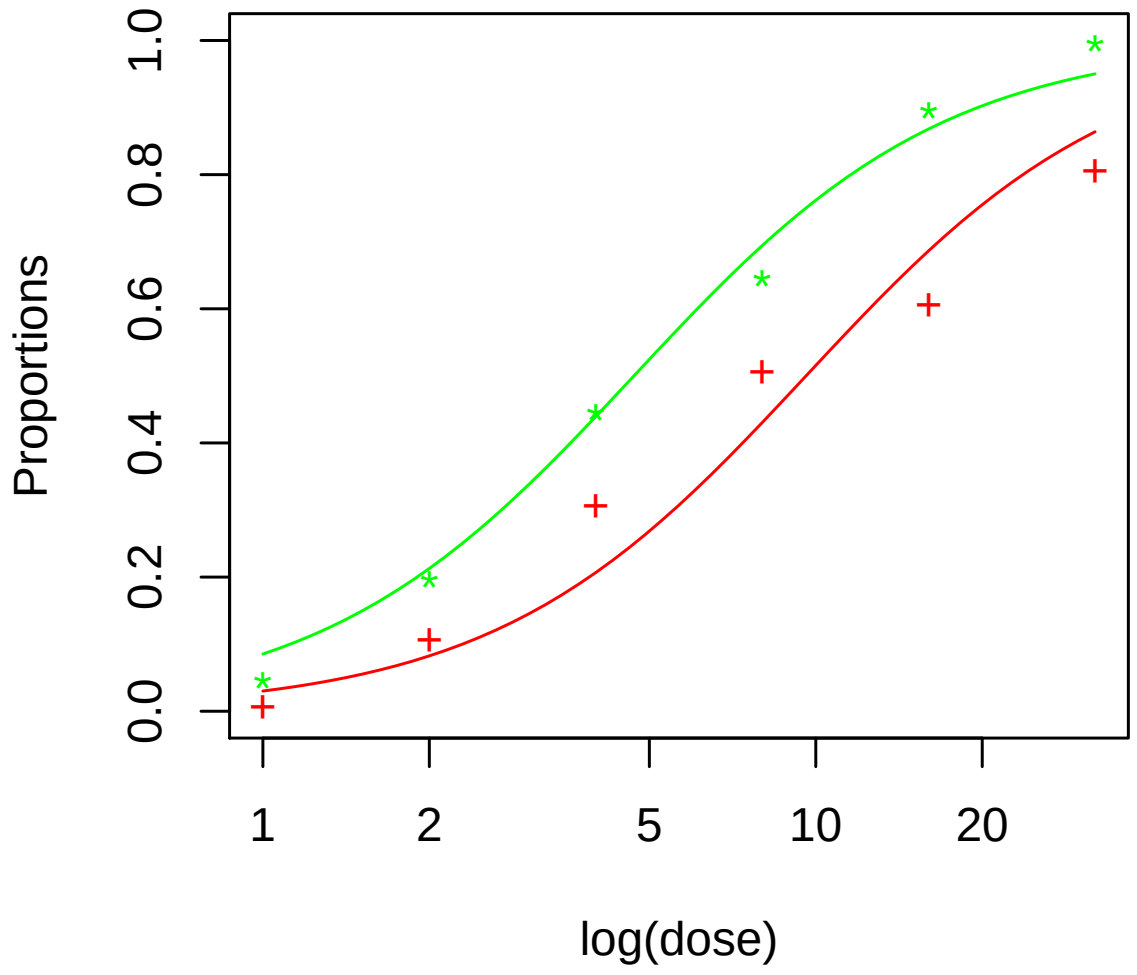


Figure 7: Cypermetrin - Observed proportions and fitted curves

Cypermethrin analysis in R

```
y <- c(1, 4, 9, 13, 18, 20, 0, 2, 6, 10, 12, 16)
sex<-factor(rep(c("M","F"), c(6,6)))
ldose<-rep(0:5,2)
dose<-2**ldose
dose<-factor(dose)
Cyper.dat <- data.frame(sex, dose, ldose, y)
attach(Cyper.dat)

plot(ldose,y/20, pch=c(rep("*",6),rep("+",6)), col=c(rep("green",6),
     rep("red",6)), xlab="log(dose)", ylab="Proportion killed")
resp<-cbind(y,20-y)

mod1<-glm(resp~1, family=binomial)
mod2<-glm(resp~dose, family=binomial)
mod3<-glm(resp~sex, family=binomial)
mod4<-glm(resp~dose+sex, family=binomial)
anova(mod1, mod2, mod4, test="Chisq")
anova(mod1, mod3, mod4, test="Chisq")

mod5<-glm(resp~ldose, family=binomial)
mod6<-glm(resp~sex+ldose-1, family=binomial)
mod7<-glm(resp~ldose/sex, family=binomial)
mod8<-glm(resp~ldose*sex, family=binomial)
anova(mod1, mod5, mod6, mod8, test="Chisq")
anova(mod1, mod5, mod7, mod8, test="Chisq")
summary(mod6)

plot(c(1,32), c(0,1), type="n", xlab="log(dose)", ylab="Proportions", log="x")
points(2**ldose,y/20, pch=c(rep("*",6),rep("+",6)), col=c(rep("green",6),
       rep("red",6)))
ld<-seq(0,5,0.1)
lines(2**ld, predict(mod6,data.frame(ldose=ld,
      sex=factor(rep("M",length(ld)),levels=levels(sex))), type="response"),
      col="green")

lines(2**ld, predict(mod6,data.frame(ldose=ld,
      sex=factor(rep("F",length(ld)),levels=levels(sex))), type="response"),
      col="red")
```

Exemplo: Mortality of confused flour beetles

Duplicate batches of flour beetles (*Tribolium confusum*) were exposed to increasing doses of carbon disulphid, CS_2. After 4 hours the number of killed insects were counted.

Concentration	Replicate 1		Replicate 2	
of CS_2	y	m	y	m
49.06	2	29	4	30
52.99	7	30	6	30
56.91	9	28	9	34
60.84	14	27	14	29
64.76	23	30	29	33
68.69	29	31	24	28
72.61	29	30	32	32
76.54	29	29	31	31

Considerations

- Response variable: Y_i – number of dead insects out of m_i insects.
- Distribution: Binomial.
- Systematic component: regression model, completely randomized experiment.
- Aim: Lethal doses.

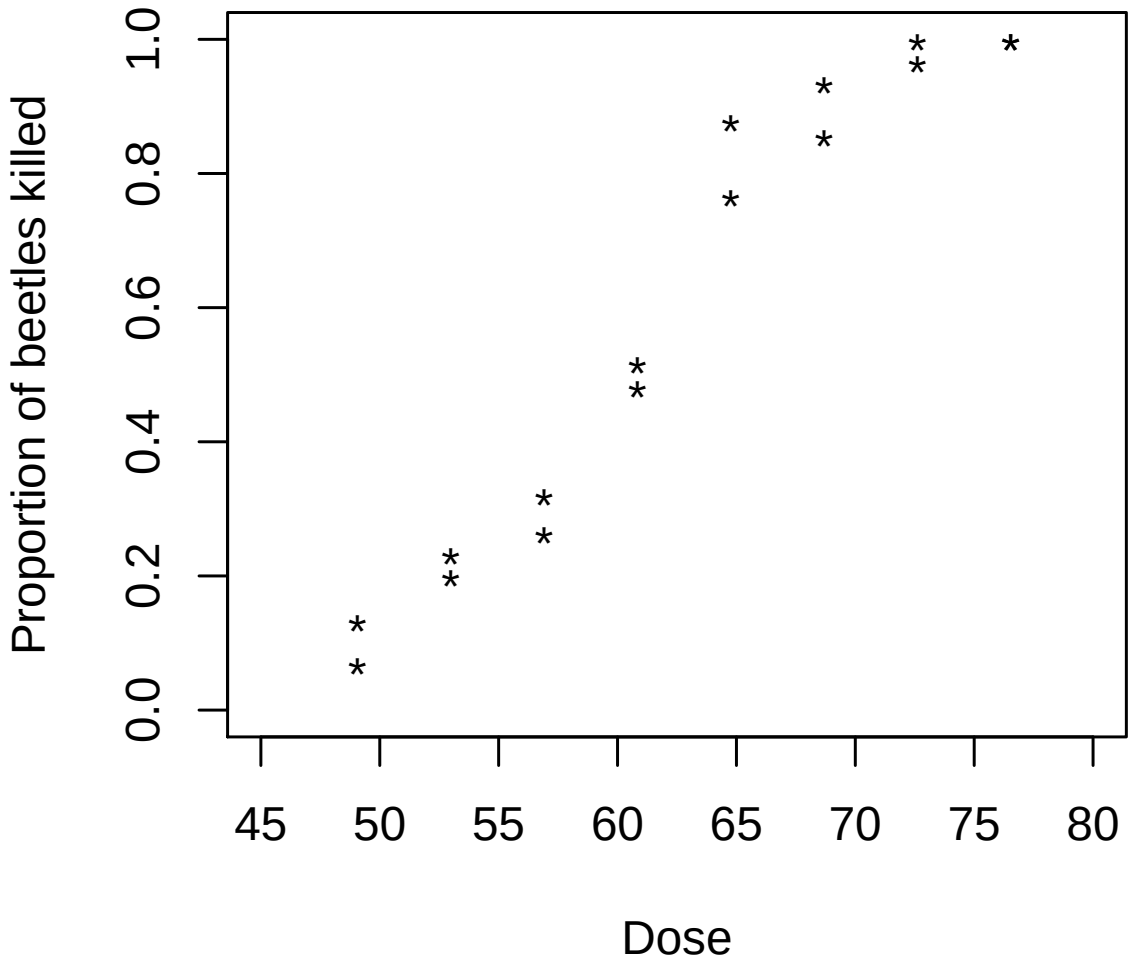


Figure 8: CS2 - Scatterplot

Models: Estimates e Tests - CS2

Logistic link: Residual Deviances

Terms fitted in model	d.f.	Deviance	p-value	X^2
$\eta = \beta_0$	15	289.14	0.00	241.02
$\eta = \beta_0 + d < 1 >$	14	12.50	0.56	10.84
$\eta = \beta_0 + d < 2 >$	13	7.93	0.85	7.36
$\eta = \beta_0 + Dose$	8	4.94	0.76	4.52

Analysis of deviance

Source	d.f.	Deviance	p-value
Linear Regression	1	276.64	0.00
Quadratic Regression	1	4.58	0.03
Lack of fit	5	2.99	0.70
(Dose)	(7)	(284.20)	
Residual	8	4.94	
Total	15	289.14	

$$\log \frac{\hat{p}}{1-\hat{p}} = 7.968 - 0.517\text{dose} + 0.00637\text{dose}^2$$

$$\hat{p} = \frac{e^{7.968-0.517\text{dose}+0.00637\text{dose}^2}}{1+e^{7.968-0.517\text{dose}+0.00637\text{dose}^2}}$$

$$\hat{LD}_{90} = 67.80$$

Models: Estimates and Tests

Complementary log-log link: Residual Deviances

Terms fitted in model	d.f.	Deviance	p-value	X^2
$\eta = \beta_0$	15	289.14	0.00	241.02
$\eta = \beta_0 + d < 1 >$	14	8.67	0.85	8.62
$\eta = \beta_0 + d < 2 >$	13	8.30	0.82	7.87
$\eta = \beta_0 + Dose$	8	4.94	0.76	4.52

Analysis of Deviance

Source	d.f.	Deviance	p-value
Linear Regression	1	280.47	0.00
Quadratic Regression	1	0.37	0.54
Lack of fit	6	3.36	0.64
(Dose)	(7)	4.94	
Residual	8	4.94	
Total	15	289.14	

$\log[-\log(1-\hat{p})] = -9.755 + 0.1554$dose

$\hat{p} = e^{e^{-9.755+0.1554}\text{dose}}$

$\hat{LD}_{50} = 60.40(0.56)$ and $\hat{LD}_{90} = 68.13(0.67)$

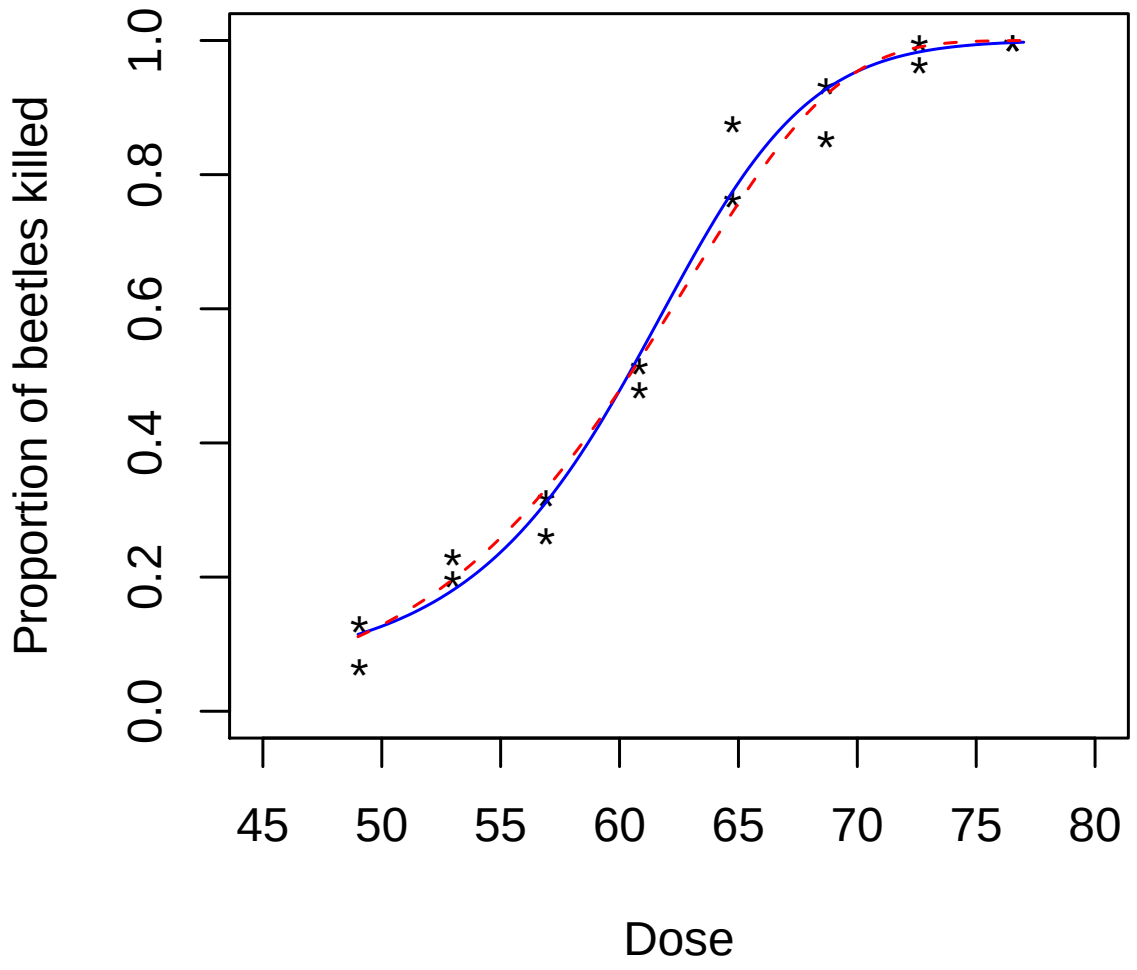


Figure 9: CS2 - Observed proportions and fitted curves

CS2 analysis in R

```
# *** CS2 Example ***
# Collett (1991), page 109

y <- c(2, 4, 7, 6, 9, 9, 14, 14, 23, 29, 29, 24, 29, 32, 29, 31)
m <- c(29, 30, 30, 30, 28, 34, 27, 29, 30, 33, 31, 28, 30, 32, 29, 31)
dose <- rep(c(49.06, 52.99, 56.91, 60.84, 64.76, 68.69, 72.61, 76.54),
times=c(2,2,2,2,2,2,2,2))
resp<-cbind(y,m-y)

## Logistic link function
modl <- glm(resp ~ poly(I(dose))+poly(I(dose),2)+ factor(dose), family=binomial)
anova(modl, test="Chisq")

mod1l<-glm(resp ~ 1, family=binomial)
1-pchisq(deviance(mod1l), df.residual(mod1l))
print(sum(residuals(mod1l, 'pearson')^2))
mod2l<-glm(resp ~ poly(I(dose)), family=binomial)
1-pchisq(deviance(mod2l), df.residual(mod2l))
print(sum(residuals(mod2l, 'pearson')^2))
mod3l<-glm(resp ~ poly(I(dose),2), family=binomial)
1-pchisq(deviance(mod3l), df.residual(mod3l))
print(sum(residuals(mod3l, 'pearson')^2))
mod4l<-glm(resp ~ factor(dose), family=binomial)
1-pchisq(deviance(mod4l), df.residual(mod4l))
print(sum(residuals(mod4l, 'pearson')^2))

summary(mod5l<-glm(resp ~ I(dose)+I(dose^2), family=binomial))
library(MASS)
# variance-covariance matrix of estimated parameters
vcov(mod5l)

## Link function test
LP2 <-  (predict(mod2l))^2
mod6l <- update(mod2l , .~. +LP2, family=binomial)
anova(mod6l, test="Chisq")

## Complementary log-log link function
modc <- glm(resp ~ poly(I(dose))+poly(I(dose),2)+ factor(dose),
family=binomial(link="cloglog"))
anova(modc, test="Chisq")

mod1c <- glm(resp ~ 1, family=binomial(link="cloglog"))
1-pchisq(deviance(mod1c), df.residual(mod1c))
print(sum(residuals(mod1c, 'pearson')^2))
mod2c <- glm(resp ~ poly(I(dose)), family=binomial(link="cloglog"))
1-pchisq(deviance(mod2c), df.residual(mod2c))
print(sum(residuals(mod2c, 'pearson')^2))
```

```
mod3c <- glm(resp ~ poly(I(dose),2), family=binomial(link="cloglog"))
1-pchisq(deviance(mod3c), df.residual(mod3c))
print(sum(residuals(mod3c, 'pearson')^2))
mod4c <- glm(resp ~ factor(dose), family=binomial(link="cloglog"))
1-pchisq(deviance(mod4c), df.residual(mod4c))
print(sum(residuals(mod4c, 'pearson')^2))

summary(mod5c <- glm(resp ~ I(dose), family=binomial(link="cloglog")))
library(MASS)
# variance-covariance matrix of estimated parameters
vcov(mod5c)

# Doses LD50 and LD90
dose.p(mod5c); dose.p(mod5c, p=0.9)

# Doses LD25, LD50, LD75
dose.p(mod5c, p = 1:3/4)

## Link function test
LP2 <-  (predict(mod2c))^2
mod6c <- update(mod2c , .~. +LP2, family=binomial(link="cloglog"))
anova(mod6c, test="Chisq")

# Plotting
plot(c(45,80), c(0,1), type="n", xlab="Dose",
ylab="Proportion of beetles killed")
points(dose, y/m, pch="*")
d<-seq(49,77,0.1)
lines(d, predict(mod3l,data.frame(dose=d), type="response"), col="blue", lty=1)
lines(d, predict(mod2c,data.frame(dose=d), type="response"), col="red", lty=2)
legend(45,1.0,c("Observed","Logit", "Cloglog"), lty=c(-1,1,2), pch=c("*"," "," "),
       col=c("black","blue", "red"))
```

Residuals for GLMs

Use in similar way to residuals from a normal model. For a glm with variance function $V(\mu)$

- **Pearson Residuals**
  Residuals based on the generalized Pearson χ^2 statistic

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

  also a standardized version

$$r_i^{PS} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)(1 - h_{ii})}}$$

- **Deviance Residuals**
  Writing the deviance as $D = \sum_i d_i^2$, the deviance residual is defined as

$$r_i^D = \text{sgn}(y_i - \hat{\mu}_i) d_i$$

  and a standardized version as

$$r_i^{DS} = \frac{\text{sgn}(y_i - \hat{\mu}_i) d_i}{\sqrt{1 - h_{ii}}}$$

- **Deletion residuals or Jackknifed residuals**

$$r_i^J = \text{sign}(y_i - \hat{\mu}_i)\sqrt{(1-h_{ii})(r_i^D)^2 + h_{ii}(r_i^P)^2}$$

where r_i^D and r_i^P are obtained in the 1st iteration.

- **Anscombe Residuals**

$$r_i^A = \frac{A(y_i) - A(\hat{\mu}_i)}{\sqrt{V[A(y_i) - A(\hat{\mu}_i)]}} = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{V(\hat{\mu}_i)}}$$

where $A(.) = \dfrac{d\mu}{V^{1/3}(\mu)}$

- **Measure of influence – Cook Distance**

$$C_i = \frac{h_{ii}}{p(1-h_{ii})}(r_i^{PS})^2$$

Added Variable plot

(Wang, 1985)

- Aim: to see if individual observations are responsible for large score statistics, or if evidence spread throughout data;

- from a glm fit of Y with $\eta = X\boldsymbol{\beta}$
  $\mathbf{z} = \eta + (\mathbf{y} - \mu)\text{diag}\{\frac{d\eta}{d\mu}\} \Rightarrow s = \mathbf{z} - \eta =$
  $(\mathbf{y} - \mu)\text{diag}\{\frac{d\eta}{d\mu}\}$
  $W = V^{-1}\text{diag}\{(\frac{d\mu}{d\eta})^2\}$
  $W^{-1/2}s = r^P$ a vector of Pearson residuals with elements $r_i^P = W_i^{-1/2}\dfrac{(\mathbf{y}_i - \mu_i)}{V(\mu_i)}\dfrac{d\eta_i}{d\mu_i} = \dfrac{(\mathbf{y}_i - \mu_i)}{V_i^{1/2}(\mu_i)}$;

- from a weighted normal regression fit of U with $\eta = X\boldsymbol{\beta}$ and weights W
  $(I - H)W^{1/2}U$ – a vector of ordinary residuals

- plot $W^{-1/2}s$ against $(I - H)W^{-1/2}U$
  to test the extra covariate U in the linear predictor

$$\eta = X\boldsymbol{\beta} + \gamma U$$

- Strong linearity – evidence spread throughout data

- Isolated points – evidence concentrated in particular cases.

Partial residual plot

(Wang, 1987)

- Aim: to assess need for transformation of single covariate U
- from a glm fit of Y with $\eta = X\boldsymbol{\beta} + \gamma U$ get:
  $W = V^{-1}\text{diag}\{(\frac{d\mu}{d\eta})^2\}$
  s with elements $\dfrac{(\mathbf{y}_i - \mu_i)}{V(\mu_i)}\dfrac{d\eta_i}{d\mu_i}$ and
  $\hat{\gamma}$
- plot $W^{-1}s + \hat{\gamma}U$ against U
- curvature suggests need for transformation

Verificação da função de ligação

Seja a função de ligação $g_0(\mu) = g(\mu, \lambda_0) = X\boldsymbol{\beta}$, aninhada em uma família paramétrica $g(\mu, \lambda)$, indexada pelo parâmetro escalar λ, por exemplo,

$$g(\mu, \lambda) = \begin{cases} \dfrac{\mu^\lambda - 1}{\lambda} & \lambda \neq 0 \\ log(\mu) & \lambda = 0 \end{cases}$$

que inclui as ligações identidade, logarítmica etc, ou então, a família de Aranda-Ordaz,

$$\mu = \begin{cases} 1 - (1 + \lambda e^\eta)^{-\frac{1}{\lambda}} & \lambda e^\eta > -1 \\ 1 & c.c. \end{cases}$$

que inclui as ligações logística, complemento log-log etc.

A expansão de Taylor para $g(\mu, \lambda)$ em torno de λ_0, produz

$$g(\mu, \lambda) \simeq g(\mu, \lambda_0) + (\lambda - \lambda_0) u(\lambda_0) = X\boldsymbol{\beta} + \gamma u(\lambda_0)$$

em que $u(\lambda_0) = \left.\dfrac{\partial g(\mu, \lambda)}{\partial \lambda}\right|_{\lambda = \lambda_0}$.

De uma forma geral usa-se $u = \hat{\eta}^2$.

Justificativa

Suponha que a função de ligação usada foi $\eta = g(\mu)$ e que a ligação verdadeira seja $g^*(\mu)$. Então,

$$g(\mu) = g[{g^*}^{-1}(\eta)] = h(\eta)$$

A hipótese nula é $H_0 : h(\eta) = \eta$ e a alternativa é $H_a : h(\eta) =$ não linear.

Fazendo-se a expansão de $g(\mu)$ em série de Taylor, tem-se:

$$g(\mu) \simeq h(0) + \eta h'(0) + \eta^2 \frac{h''(0)}{2}$$

então, a variável adicionada é $\hat{\eta}^2$, desde que o modelo tenha termos para o qual a média geral seja marginal.

Testes formais

- Teste da razão de verossimilhança
- Teste escore
- Teste de Wald

Gráficos

- Gráfico da variável adicionada
- Gráfico de resíduos parciais

CS2 toxicity

Mortality of adult beetles after five hours' exposure to gaseous carbon disulphid (Bliss, 1935).

Log dosage	Number exposed	Number killed
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

Considerations

- Response variable: Y_i – number of killed beetles out of m_i exposed beetles (Pregibon, 1980).
- Distribution: Binomial.
- Systematic component: regression model, completely randomized experiment.
- Aim: Lethal doses.

- Binomial logistic model with $\eta = \beta_0 + \beta_1 log(dose)$ gives deviance 11.23 on 6 d.f. $\Rightarrow$ lack of fit
- Deviance for link test 8.037 on 1 d.f. $\Rightarrow$ need for different link
- Added variable plot for $U = \hat{\eta}^2 \Rightarrow$ evidence that $\gamma \neq 0$, spread throughout data

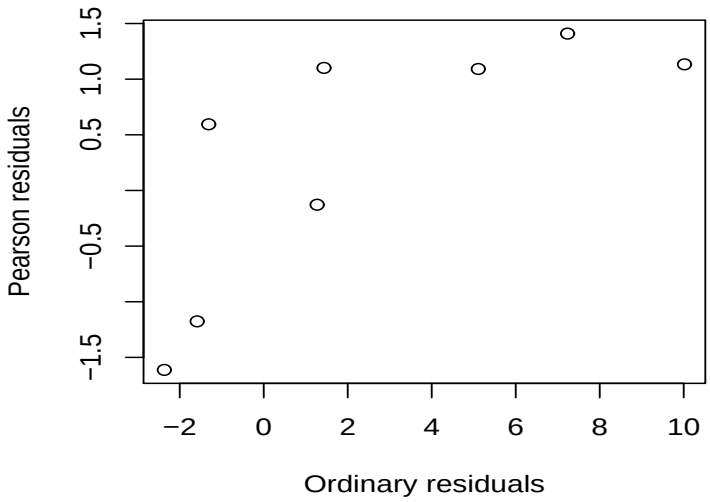


Figure 10: CS2 - Added variable plot

- C-loglog link with a linear lp gives good fit
- Logistic link with a quadratic lp gives good fit

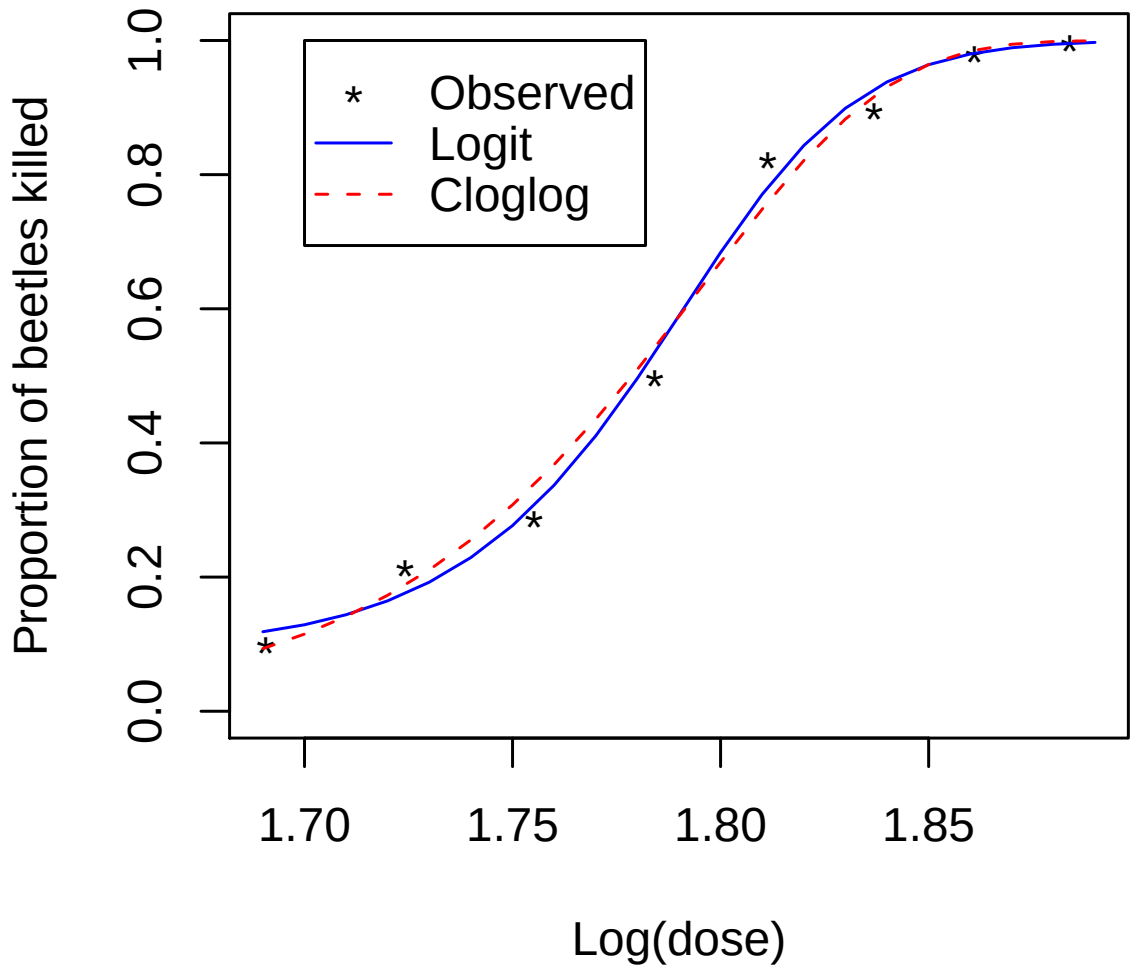


Figure 11: Beetle mortality - Observed proportions and fitted curves

Beetle mortality analysis in R

Pregibon (1980) - Applied Statistics, 29: 20

```
y <- c(6, 13, 18, 28, 52, 53, 61, 60)
m <- c(59, 60, 62, 56, 63, 59, 62, 60)
ldose <- c(1.6907, 1.7242, 1.7552, 1.7842, 1.8113, 1.8369, 1.8610, 1.8839)
beetle.dat <- data.frame(ldose, m, y)
attach(beetle.dat)
resp<-cbind(y,m-y)

## Logistic link function
modl<-glm(resp ~ poly(I(ldose))+poly(I(ldose),2), family=binomial(link="logit"))
anova(modl, test="Chisq")
mod1l<-glm(resp ~ 1, family=binomial(link="logit"))
1-pchisq(deviance(mod1l), df.residual(mod1l))
print(sum(residuals(mod1l, 'pearson')^2))
mod2l<-glm(resp ~ I(ldose), family=binomial(link="logit"))
1-pchisq(deviance(mod2l), df.residual(mod2l))
print(sum(residuals(mod2l, 'pearson')^2))

## Link function test
LP2l <-  (predict(mod2l,type = c("link")))^2
mod3l<-update(mod2l , .~. +LP2l, family=binomial(link="logit"))
anova(mod3l, test="Chisq")

mod4l<-glm(resp ~ poly(I(ldose),2), family=binomial(link="logit"))
1-pchisq(deviance(mod4l), df.residual(mod4l))
print(sum(residuals(mod4l, 'pearson')^2))

## Link function test
LP2q <-  (predict(mod4l,type = c("link")))^2
mod5l<-update(mod4l , .~. +LP2q, family=binomial(link="logit"))
anova(mod5l, test="Chisq")

## Added variable plot
rp <- residuals(mod2l, 'pearson')
W <- mod2l$fitted*(1- mod2l$fitted)*m
U <- LP2l
mod1n <- glm(U~I(ldose), family=gaussian, weights=W)
rU <- (U - mod1n$fitted)
plot(rU,rp, xlab="Ordinary residuals", ylab="Pearson residuals")

# Partial residual plot
resparc <- residuals(mod3l, 'pearson') + mod3l$coef[3]*U
plot(U,resparc, xlab="U", ylab="Residual + component")

summary(mod6l<-glm(resp ~ I(ldose)+I(ldose^2), family=binomial(link="logit")))
anova(mod6l, test="Chisq")
```

```
library(MASS)
# variance-covariance matrix of estimated parameters
vcov(mod6l)

## Complementary log-log link function
modc<-glm(resp ~ poly(I(ldose))+poly(I(ldose),2), family=binomial(link="cloglog"))
anova(modc, test="Chisq")

mod1c<-glm(resp ~ 1, family=binomial(link="cloglog"))
1-pchisq(deviance(mod1c), df.residual(mod1c))
print(sum(residuals(mod1c, 'pearson')^2))
mod2c<-glm(resp ~ I(ldose), family=binomial(link="cloglog"))
1-pchisq(deviance(mod2c), df.residual(mod2c))
print(sum(residuals(mod2c, 'pearson')^2))

## Link function test
LP2l <-  (predict(mod2c,type = c("link")))^2
mod3c<-update(mod2c , .~. +LP2l, family=binomial(link="cloglog"))
anova(mod3c, test="Chisq")

## Added variable plot
rp <- residuals(mod2c, 'pearson')
W <- mod2c$fitted*(1- mod2c$fitted)*m
U <- LP2l
mod1n <- glm(U~I(ldose), family=gaussian, weights=W)
rU <- (U - mod1n$fitted)
plot(rU,rp, xlab="Ordinary residuals", ylab="Pearson residuals")

# Partial residual plot
resparc <- residuals(mod3c, 'pearson') + mod3c$coef[3]*U
plot(U,resparc, xlab="U", ylab="Residual + component")

summary(mod5c<-glm(resp ~ I(ldose), family=binomial(link="cloglog")))
library(MASS)
# variance-covariance matrix of estimated parameters
vcov(mod5c)
# Doses LD50 and LD90
dose.p(mod5c); dose.p(mod5c, p=0.9)
# Doses LD25, LD50, LD75
dose.p(mod5c, p = 1:3/4)

# Plotting
plot(c(1.69,1.89), c(0,1), type="n", xlab="Log(dose)",
     ylab="Proportion of beetles killed")
points(ldose, y/m, pch="*")
ld<-seq(1.69,1.89,0.01)
lines(ld, predict(mod4l,data.frame(ldose=ld),type="response"), col="blue", lty=1)
lines(ld, predict(mod2c,data.frame(ldose=ld),type="response"), col="red", lty=2)
legend(1.7,1.0,c("Observed","Logit", "Cloglog"), lty=c(-1,1,2),
pch=c("*"," "," "), col=c("black","blue", "red"))
```

Half-normal Plots with simulation envelopes

- Fit a model and calculate, $d_{(i)}$, the ordered absolute values of some diagnostic.
- Simulate 19 samples for the response variable using the fitted model and the same values for the explanatory variables.
- Refit the model to each sample and calculate the ordered absolute values of the diagnostic of interest, $d^*_{j(i)}$, $j = 1, \ldots, 19$, $i = 1, \ldots, n$.
- For each i, calculate the mean, minimum and maximum of the $d^*_{j(i)}$.
- Plot these values and the observed $d_{(i)}$ against the half-normal order statistics.

Germination of Orobanche seed

O. aegyptiaca 75		*O. aegyptiaca 73*	
Bean	Cucumber	Bean	Cucumber
10/39	5/6	8/16	3/12
23/62	53/74	10/30	22/41
23/81	55/72	8/28	15/30
26/51	32/51	23/45	32/51
17/39	46/79	0/4	3/7
	10/13		

Considerations

- Response variable: Y_i – number of germinated seeds out of m_i seeds (Crowder, 1978).
- Distribution: Binomial.
- Systematic component: factorial 2×2 (2 species, 2 extracts), completely randomized experiment.
- Aim: to see how germination is affected by species and extracts.
- Problem: overdispersion.

Simple Binomial Model

- No of seeds germinating y_i as y-variable
- Binomial logit model
- `Species * Extract` interaction model

Analysis of deviance

Source	d.f.	Deviance	p-value
E	1	55,97	
E | S	1	56,49	
S	1	2,54	
S | E	1	3,06	
S.E	1	6,41	
Residual	17	33,28	0.0096
Total	20	98,72	

- Interaction significant
- Some evidence of overdispersion?

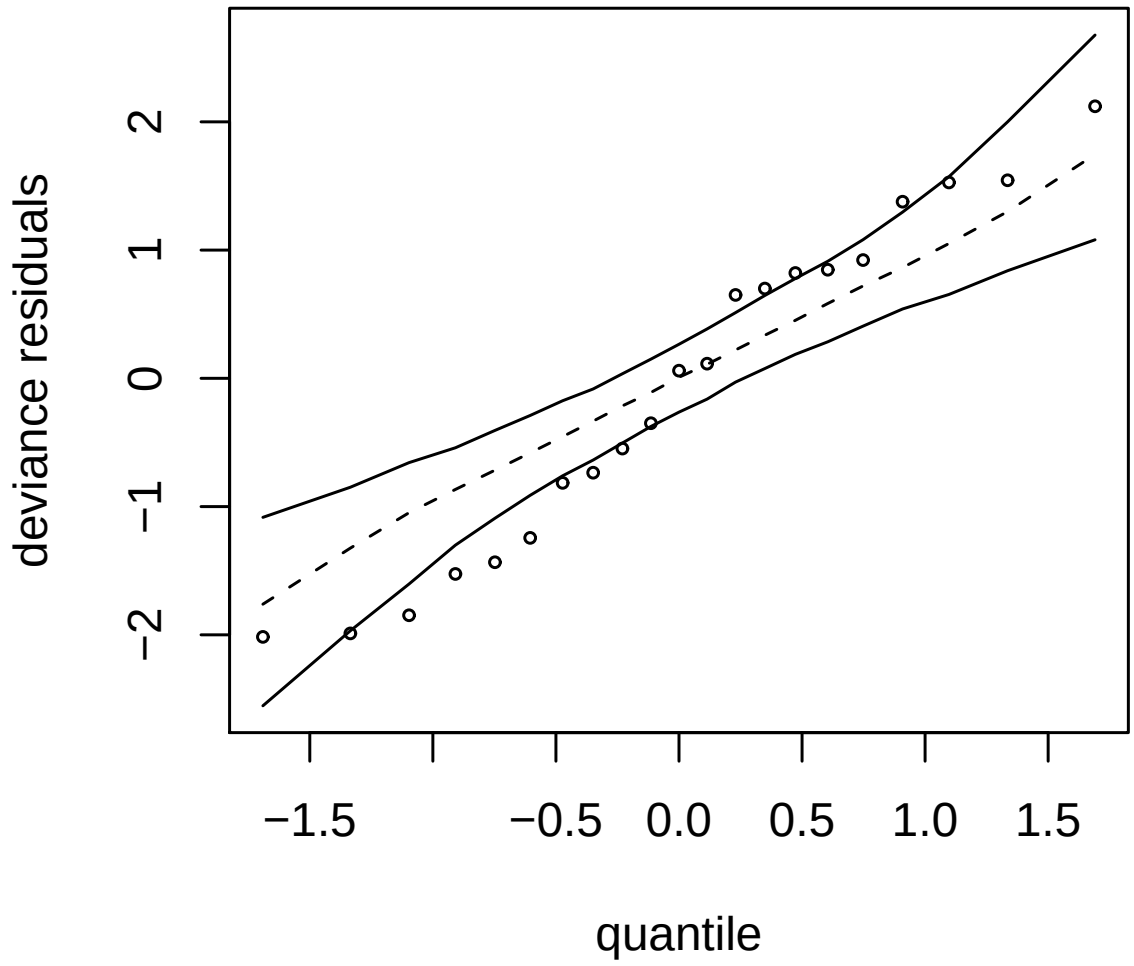


Figure 12: Orobanche – Normal plot with simulated envelope

Orobanche analysis in R

```
Species <- factor(c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
                    2, 2, 2, 2, 2, 2, 2, 2, 2, 2))
Extract <- factor(c(1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2,
                    1, 1, 1, 1, 1, 2, 2, 2, 2, 2))
y <- c(10, 23, 23, 26, 17,  5, 53, 55, 32, 46, 10,  8,
       10,  8, 23,  0,  3, 22, 15, 32,  3)
m <- c(39, 62, 81, 51, 39,  6, 74, 72, 51, 79, 13, 16,
       30, 28, 45,  4, 12, 41, 30, 51,  7)

orobanch <- data.frame(Species, Extract, m, y)
attach(orobanch)
resp<-cbind(y,m-y)
p<-y/m
boxplot(p~Species*Extract,xlab="Species*Extract", ylab="Observed proportions")

# Binomial fit
orobanch1.fit<-glm(resp~Species*Extract, family=binomial)
summary(orobanch1.fit, cor=T)
anova(orobanch1.fit, test="Chisq")

orobanch1.fit<-glm(resp~Extract*Species, family=binomial)
summary(orobanch1.fit, cor=T)
anova(orobanch1.fit, test="Chisq")

# Normal plot with simulated envelope
GLMBin<-glm(resp~Extract*Species,family=binomial())
res<-sort(summary(GLMBin)$deviance.resid)
env<-quantile(res,c(.05,0.5,.95))
for (i in 1:999){
y.<-rbinom(rep(1,length(GLMBin$fit)),m, GLMBin$fit)
resp.<-cbind(y.,m-y.)
res<-cbind(res,sort(summary(glm(resp.~Extract*Species,
family=binomial()))$deviance.resid))}
for (i in 1:nrow(res)){
env<-cbind(env,quantile(res[i,],c(.05,0.5,.95)))}
matplot(qnorm((1:nrow(res))/(nrow(res)+1)),cbind(t(env[,-1]),res[,1]),
col=c(1,1,1,1),pch=c(NULL,NULL,1),cex=.5,type=c("l","l","l","p"),lty=c(1,2,1,1),
xlab="quantile", ylab="deviance residuals")
```

Relative potency - Toxicity of insecticide to flour beetles

Table 1: Toxicity to the flour beetle of sprays of DDT, γ-BHC and the two together.

Insecticide	Deposit of Insecticide					
	2.00	2.64	3.48	4.59	6.06	8.00
DDT	3/50	5/49	19/47	19/50	24/49	35/50
γ-BHC	2/50	14/49	20/50	27/50	41/50	40/50
DDT + γ-BHC	28/50	37/50	46/50	48/50	48/50	50/50

Considerations

- Response variable: Y_i – number of dead insects out of m_i insects.
- Distribution: Binomial.
- Systematic component: ANOVA with regression model, completely randomized experiment.
- Aim: Lethal doses and comparison of insecticides.

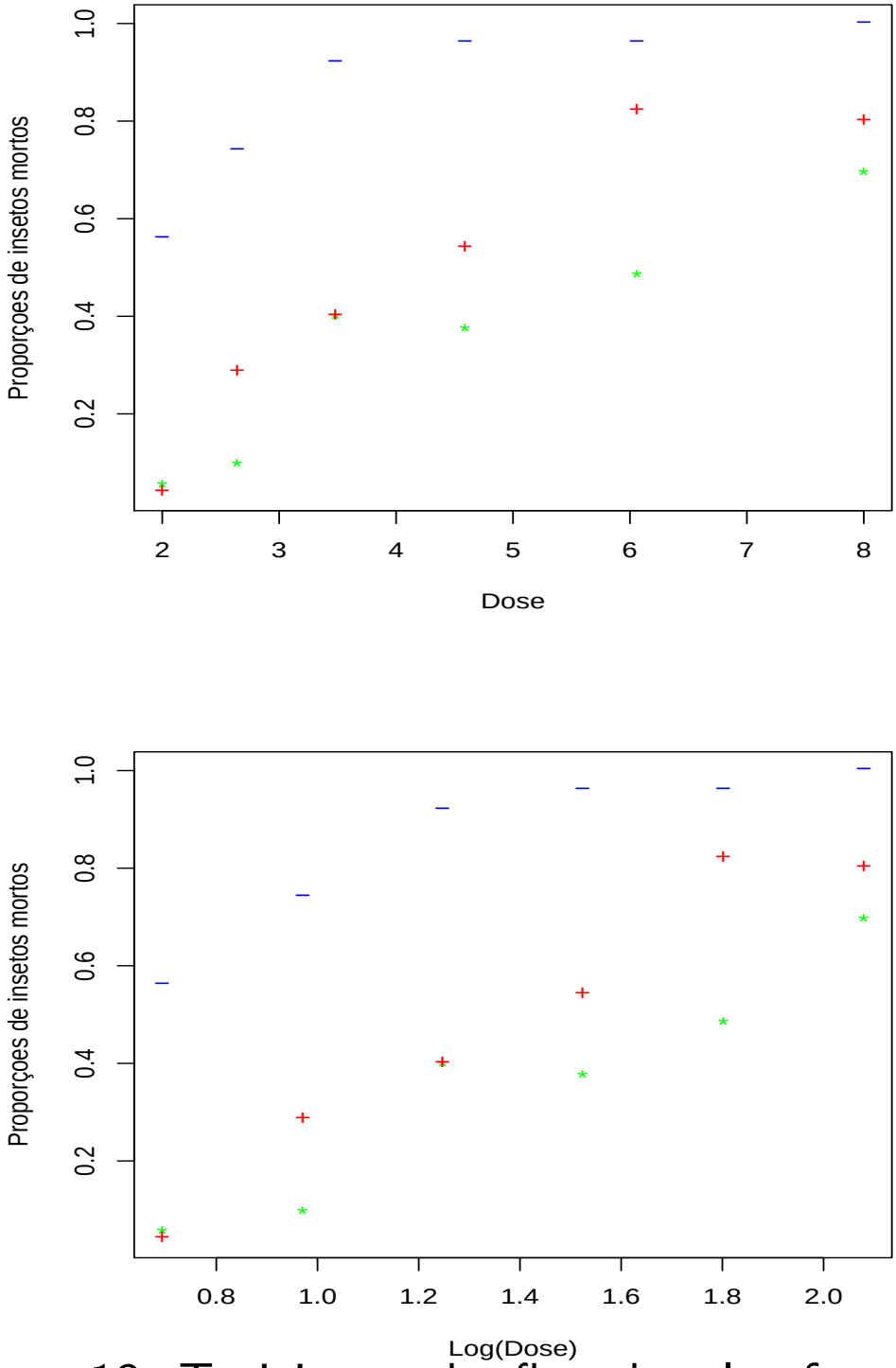


Figure 13: Toxicity to the flour beetle of sprays of DDT, γ-BHC and the two together - Scatterplot

Model: Estimates and Tests

Toxicity to the flour beetle of sprays of DDT, γ-BHC and the two together - deviance residuals

Terms fitted in model	d.f.	Deviance	X^2
Constant	17	413.6	347.1
Insecticide	15	234.7	215.0
Dose	12	242.6	218.9
Insecticide + Dose	10	12.8	11.8

Toxicity to the flour beetle of sprays of DDT, γ-BHC and the two together - Analysis of deviance

Source	d.f.	Deviance	p-value
Insecticide	2	178.9	< 0.0001
Insecticide|Dose	2	229.8	< 0.0001
Dose	5	171.0	< 0.0001
Dose|Insecticide	5	221.8	< 0.0001
Residual	10	12.8	0.2316
Total	17	413.6	

Models

logit$(p) = \alpha_j + \beta_j \ \log(\text{dose})$ – different logistic regression lines

logit$(p) = \alpha_j + \beta \ \log(\text{dose})$ – common slope

logit$(p) = \alpha + \beta_j \ \log(\text{dose})$ – common intercept

logit$(p) = \alpha + \beta \ \log(\text{dose})$ – same logistic regression line.

Residual Deviances

Terms fitted in model	d.f.	Deviance	X^2
Constant	17	413.6	347.1
Insecticide + Insecticide log(dose)	12	17.9	17.6
Insecticide + log(dose)	14	21.2	20.3
Const. + Insecticide log(dose)	14	24.7	28.0
Const. + log(dose)	16	246.8	219.8

Analysis of deviance

Source	d.f.	Deviance	p-value
Insecticide	2	178.9	< 0.0001
Linear Regression	1	213.4	< 0.0001
Residual	10	12.8	0.2316
Total	17	413.6	

Regression equations

DDT $\log \dfrac{\hat{p}}{1-\hat{p}} = -3.8425 + 2.6958 \log(\text{dose})$ γ-BHC

$$\log \frac{\hat{p}}{1-\hat{p}} = -4.5553 + 2.6958 \log(\text{dose})$$

DDT + γ-BHC

$$\log \frac{\hat{p}}{1-\hat{p}} = -1.4248 + 2.6958 \log(\text{dose})$$

Lethal Doses

DDT

$$\log(\hat{LD}_{50}) = \frac{3.8425}{2.6958} = 1.42 \Rightarrow LD_{50} = 4.16$$

γ-BHC

$$\log(\hat{LD}_{50}) = \frac{4.5553}{2.6958} = 1.69 \Rightarrow LD_{50} = 5.42$$

DDT + γ-BHC

$$\log(\hat{LD}_{50}) = \frac{1.4248}{2.6958} = 0.53 \Rightarrow LD_{50} = 1.70$$

Potency

Potência relativa da mistura em relação ao DDT

$$\frac{4.16}{1.696} = 2.45$$

Potência relativa da mistura em relação ao γ-BHC

$$\frac{5.417}{1.696} = 3.19$$

SINERGISMO

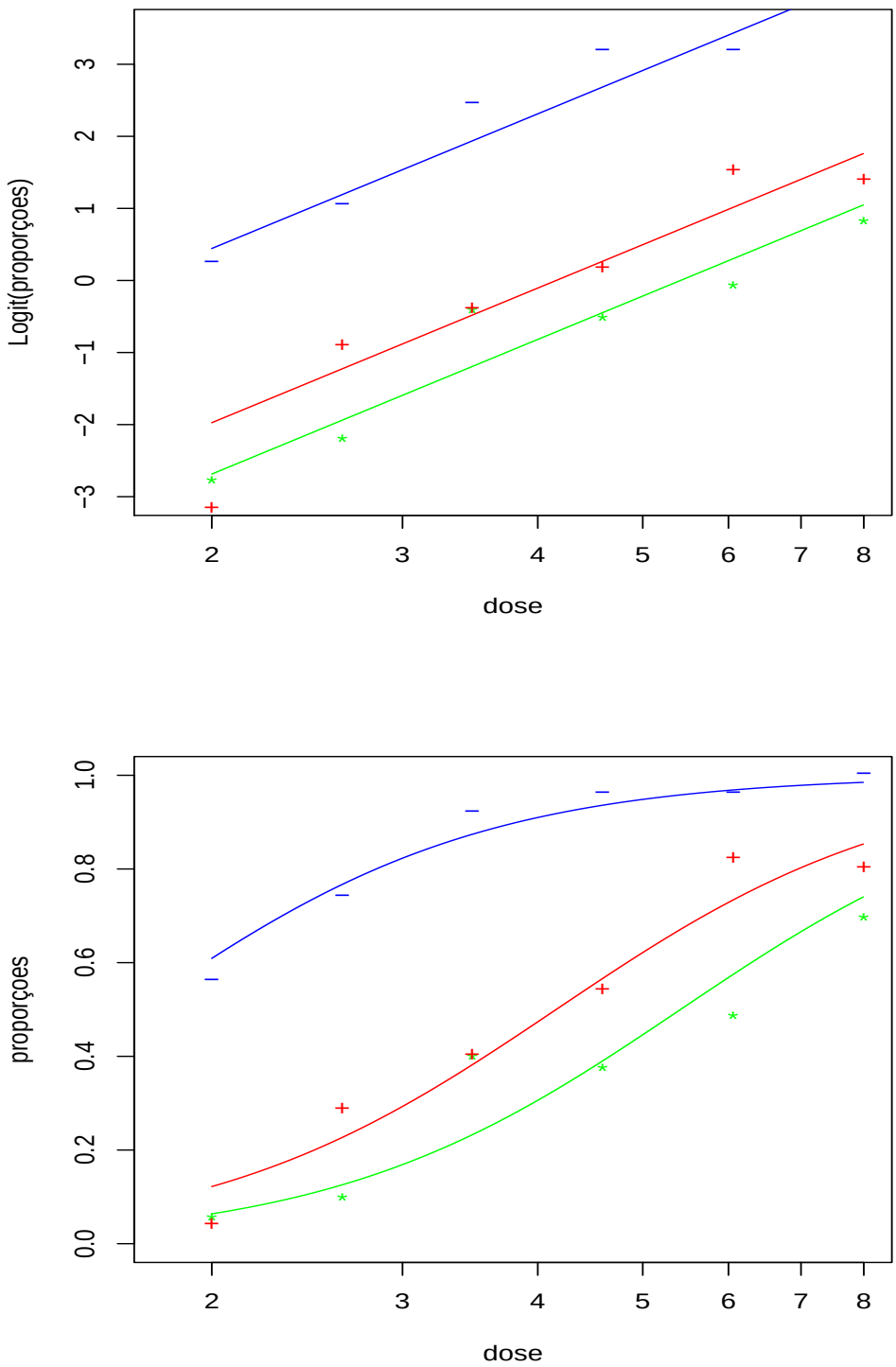


Figure 14: Toxicity to the flour beetle of sprays of DDT, γ-BHC and the two together - Scatterplot

These data show the effect of mixtures of two insecticides A and B on the tobacco budworm, *Heliothis virescens*. One microlitre of the mixture was painted on the body of each individual insect and the number of dead insects were recorded 96 hours later.

Table 2: Number of dead insects out of m insects for different mixtures of two insecticides

Mixture	Amount of A (ppm)	Amount of B (ppm)	Number of dead insects	Number of insects tested
B	0	30.00	26	30
B	0	15.00	19	30
B	0	7.50	7	30
B	0	3.75	5	30
A25:B75	6.50	19.50	23	30
A25:B75	3.25	9.75	11	30
A25:B75	1.625	4.875	3	30
A25:B75	0.812	2.438	0	30
A50:B50	13.00	13.00	15	30
A50:B50	6.50	6.50	5	30
A50:B50	3.25	3.25	4	29
A50:B50	1.625	1.625	0	29
A75:B25	19.50	6.50	20	30
A75:B25	9.75	3.25	13	30
A75:B25	4.875	1.625	6	29
A75:B25	2.438	0.812	0	30
A	30.00	0	23	30
A	15.00	0	21	30
A	7.50	0	13	30
A	3.75	0	5	30

Fonte: Gilitan *et al.* (1988)

Mixtures of insecticides

- In this last example: 2 insecticides A$\times$B, separetely, and and a mixture (1:3,1:1, 3:1). The total amount of insecticide was constant for each of of the 5 mixtures at one of the 4 levels.
- Why do experiments with mixtures? - to see whether mixing produces
  - SYNERGISM - improved performance
  - ANTAGONISM - reduced performance
- Steps for the analysis:
  - Preliminary analysis of A$\times$B when applied separately, using sequence of probit analysis with Chemical as a 2-level factor and log(dose), to test parallelism
  - Fit the same model sequence, with with Chemical as a 5-level factor, denoting the 5 mixtures, and log(dose), to test parallelism
  - Fit a Generalized non-linear model, a probit model with

additive behaviour - simple similar action - equivalent to dose $a + b\rho$ of A, where rho denotes potency of B relative to A

$$p_{ij} = \Phi_{ij}(\alpha + \beta log(a_i + \rho b_j))$$

If true response equivalent to more than $a + b\rho$ of A then SINERGISM

If true response equivalent to less than $a + b\rho$ of A then ANTAGONISM

multiplicative behaviour

$$p_{ij} = \Phi_{ij}(\alpha + \beta log(a_i + \rho b_j + \gamma\sqrt{a_i b_j}))$$

where $a_i + b_j$=fixed amount c, $\sqrt{a_i b_j}$ is maximized when $a_i = b_j = \frac{1}{2}c$. Also, when $a_i = 0$ or $b_j = o$. If

$\gamma > 0 \rightarrow$ SINERGISM

$\gamma < 0 \rightarrow$ ANTAGONISM

Model: Estimates and Tests

Insecticide mixtures - deviance residuals

Terms fitted in model	d.f.	Deviance	X^2
Constant	7	72.74	67.32
Chemical	6	72.32	67.03
log(Dose)	6	4.97	4.86
Chemical + log(Dose)	5	4.52	4.53
Chemical*log(Dose)	4	3.58	3.57

Cypermethrin toxicity data - Analysis of deviance

Source	d.f.	Deviance	p-value
Chemical	1	0.42	0.52
Chemical|log(Dose)	1	0.45	0.50
log(Dose)	1	67.77	< 0.0001
log(Dose)|Chemical	1	67.80	< 0.0001
log(Dose):Chemical	1	0.94	0.33
Residual	4	3.58	
Total	11	124.9	