# Generalized Linear Regression Homework 3

Brian Cervantes Alvarez

November 12, 2024

### Toxicity of Insecticides

Flour beetles *Tribolium castaneum* were sprayed with one of three insecticides in solution at different doses. The number of insects killed after a six-day period is recorded below:

| Insecticide | 2.00 | 2.64 | 3.48 | 4.59 | 6.06 | 8.00 |
|---|---|---|---|---|---|---|
| **DDT** | 3/50 | 5/49 | 19/47 | 19/38 | 24/49 | 35/50 |
| $\gamma$-BHC | 2/50 | 14/49 | 20/50 | 27/50 | 41/50 | 40/50 |
| **DDT +** $\gamma$-BHC | 28/50 | 37/50 | 46/50 | 48/50 | 48/50 | 50/50 |

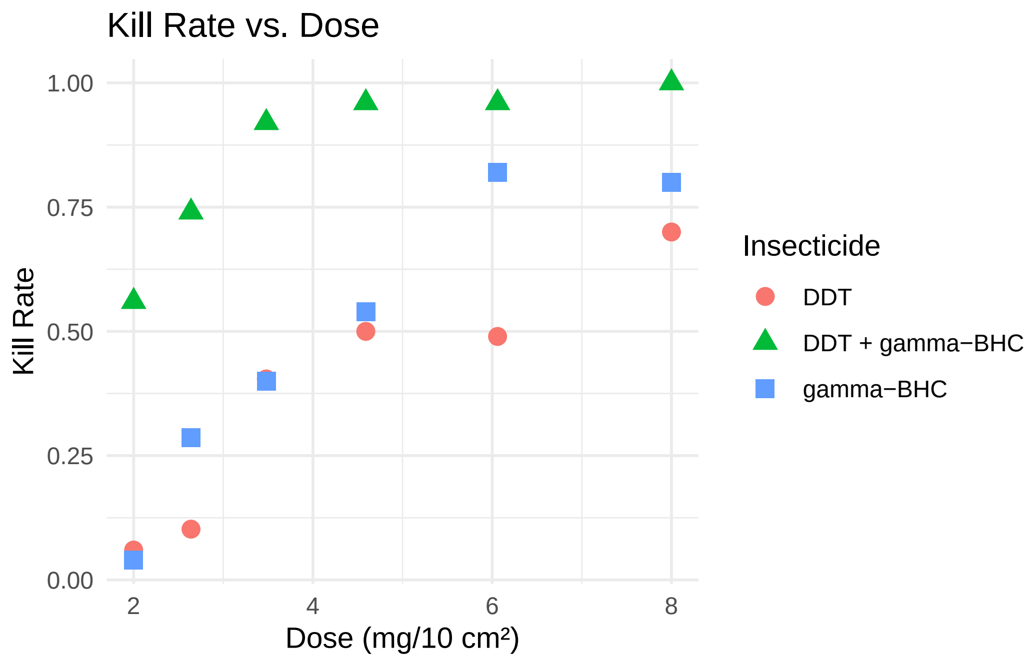
## Part A

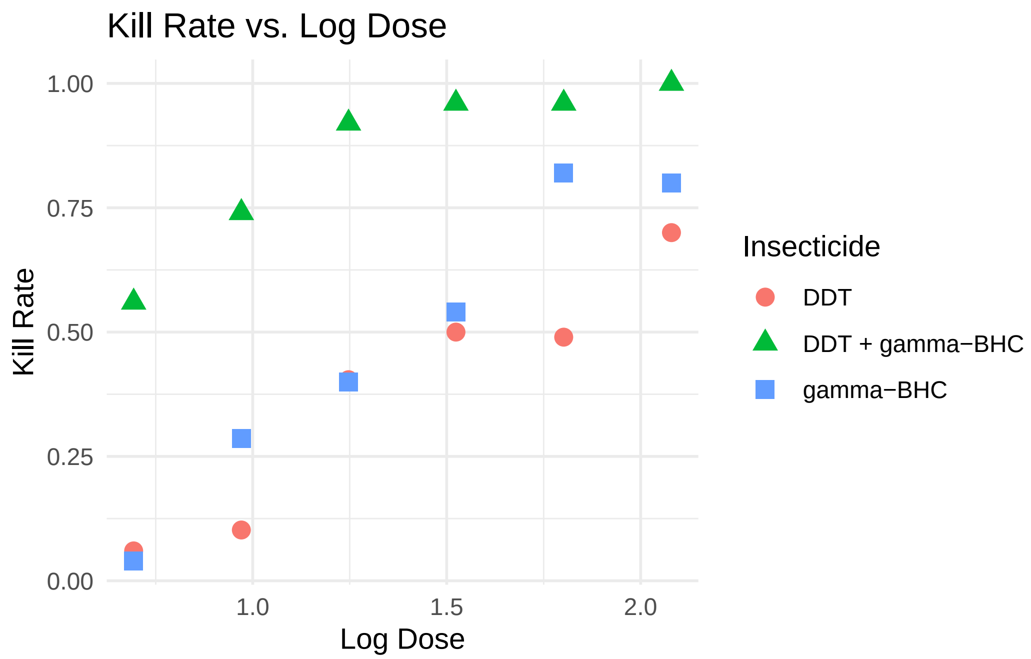
Investigate graphically the relationship between dose, either in original units or in log units, and the kill rate.

### Solution

For each observation, the kill rate can be calculated for each insecticide.

**Plotting Kill Rate vs. Dose and Log Dose:**



## From the plots

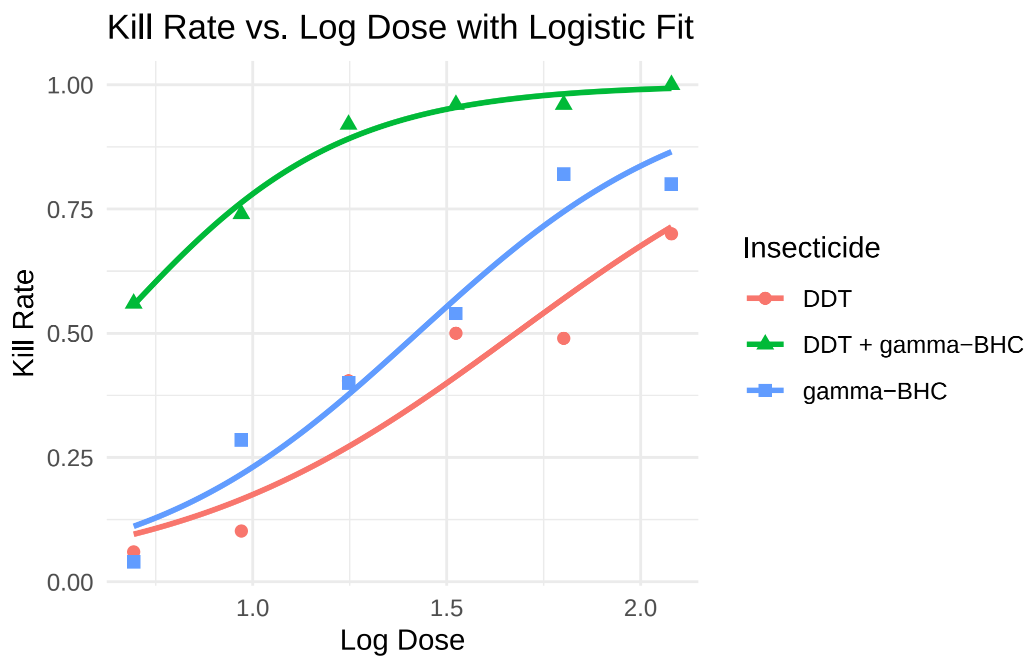
The relationship between dose and kill rate is non-linear and shows a sigmoidal shape. Next, when plotting kill rate against log dose, the relationship appears more linear, especially in the middle range of doses. Lastly, The combination of DDT and gamma-BHC achieves higher kill rates at lower doses compared to the individual insecticides.

## Part B

### Solution

**Fitting Logistic Regression Models:**

We fit logistic regression models for each insecticide separately, using the log dose as the predictor.




### What we see from the plots

The logistic regression curves fit the data points well for each insecticide. Additionally, the combination insecticide shows a steeper curve, indicating higher potency. To end, the individual insecticides (DDT and gamma-BHC) have similar but less steep curves compared to the combination.

## Part C

**Consider two models:**

1. One in which the relationship is described by three parallel straight lines in the log dose.
2. One in which the three lines are straight but not parallel.

**Assess the evidence against the hypothesis of parallelism.**

### Solution

**Model Definitions:**

- **Parallel Lines Model (Common Slope):**

$$\text{logit}(p) = \beta_0 + \beta_{\text{chem}} + \beta_1 \times \text{LogDose}$$

- **Non-Parallel Lines Model (Separate Slopes):**

$$\text{logit}(p) = \beta_0 + \beta_{\text{chem}} + \beta'_{\text{chem}} \times \text{LogDose}$$

**Fitting the Models:**

```
# Parallel lines model
parallel_model <- glm(cbind(Killed, Total - Killed) ~
                        Insecticide + LogDose,
                      data = data, family = binomial)

# Non-parallel lines model (interaction between Insecticide and LogDose)
nonparallel_model <- glm(cbind(Killed, Total - Killed) ~
                           Insecticide * LogDose,
                         data = data, family = binomial)
```

**Likelihood Ratio Test:**

We compare the two models using a likelihood ratio test.

```r
# Perform likelihood ratio test
anova(parallel_model, nonparallel_model, test = "Chisq")
```

```
Analysis of Deviance Table

Model 1: cbind(Killed, Total - Killed) ~ Insecticide + LogDose
Model 2: cbind(Killed, Total - Killed) ~ Insecticide * LogDose
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        14     22.685
2        12     19.292  2   3.3925   0.1834
```

**Results:** The p-value from the likelihood ratio test is greater than 0.05. This suggests that there is no significant difference between the models. Hence, we fail to reject the null hypothesis that the slopes are equal across insecticides. The assumption of parallelism is appropriate for this data.

## Part D

Let `chem` be a 3-level factor, and let `ldose` be the log dose. Explain the relationship between the regression coefficients in the model formulae:

1. `chem + ldose`
2. `chem + ldose - 1`

Explain the relationship between the two covariance matrices.

### Solution

**Model 1: `chem + ldose`**: Includes an intercept term; The reference level of `chem` is absorbed into the intercept, and the coefficients for `chem` represent differences from the reference level.

**Model 2: `chem + ldose - 1`**: Does not include the intercept term, and each level of `chem` has its own intercept. Lastly, the coefficients for `chem` are the actual intercepts for each insecticide.

**Relationship Between Coefficients:**

- In Model 1, the intercept represents the baseline for the reference insecticide at ldose = 0.
- In Model 2, the intercepts are directly estimated for each insecticide.

**Covariance Matrices:**

The covariance matrices differ due to the different parameterizations; by omitting the intercept, it changes the dependencies among the estimated coefficients. In Model 2, the covariance between the intercepts and slopes may be different compared to Model 1.

## Part E

Assuming that three parallel straight lines suffice, estimate the potency of the combination relative to each of the components. Use the Delta method to obtain a 90% confidence interval for each of these relative potencies.

## Solution

```
Call:
glm(formula = cbind(Killed, Total - Killed) ~ Insecticide + LogDose,
    family = binomial, data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0299  -0.7354  -0.3101   1.0198   2.3051

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -4.4541     0.3575 -12.460  < 2e-16 ***
InsecticideDDT + gamma-BHC   3.0314     0.2521  12.022  < 2e-16 ***
Insecticidegamma-BHC         0.6144     0.1999   3.074  0.00211 **
LogDose                      2.6938     0.2146  12.551  < 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: 407.129  on 17  degrees of freedom
Residual deviance:  22.685  on 14  degrees of freedom
AIC: 93.943

Number of Fisher Scoring iterations: 4
```

**Calculating Relative Potency:**

The relative potency $R$ of the combination compared to an insecticide is:

$$R = \exp\left(-\frac{\beta_{\text{comb}} - \beta_{\text{comp}}}{\beta_1}\right)$$

Where $\beta_{\text{comb}}$ is the coefficient for the combination, $\beta_{\text{comp}}$ is the coefficient for the component insecticide, and $\beta_1$ is the common slope.

**Using the Delta Method to Find Variance:**

The variance of $\log(R)$ is:

$$\mathrm{Var}[\log(R)] = \frac{\mathrm{Var}[\beta_{\mathrm{comb}} - \beta_{\mathrm{comp}}] + (\log(R))^2 \times \mathrm{Var}[\beta_1]}{\beta_1^2}$$

**Calculations (with R)**

```
Relative Potency (Combination vs DDT): 0.325
90% Confidence Interval: [0.307, 0.344]

Relative Potency (Combination vs gamma-BHC): 0.408
90% Confidence Interval: [0.386, 0.43]

Relative Potency (gamma-BHC vs DDT): 0.796
90% Confidence Interval: [0.761, 0.833]
```

## Part F

Use Fieller's method to obtain a 90% confidence interval for each of the above relative potencies.

### Solution

Using R:

```
Fieller's 90% CI for Relative Potency (Combination vs DDT):
Point Estimate: 0.325
90% Confidence Interval: [2.643, 3.673]

Fieller's 90% CI for Relative Potency (Combination vs gamma-BHC):
Point Estimate: 0.408
90% Confidence Interval: [2.121, 2.889]

Fieller's 90% CI for Relative Potency (gamma-BHC vs DDT):
Point Estimate: 0.796
90% Confidence Interval: [1.113, 1.423]
```

## Part G

Provide your answer to the previous two parts under the `c-log-log` link.

## Solution

**Using the Complementary Log-Log Link Function:**

We refit the parallel lines model using the complementary log-log link.

```r
# Fit the model with cloglog link
cloglog_model <- glm(cbind(Killed, Total - Killed) ~ Insecticide + LogDose,
                     data = data, family = binomial(link = "cloglog"))
summary(cloglog_model)
```

```
Call:
glm(formula = cbind(Killed, Total - Killed) ~ Insecticide + LogDose,
    family = binomial(link = "cloglog"), data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9184  -0.9624   0.1086   0.3664   1.9988

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -3.2356     0.2424 -13.350  < 2e-16 ***
InsecticideDDT + gamma-BHC   1.8618     0.1512  12.314  < 2e-16 ***
Insecticidegamma-BHC         0.4242     0.1342   3.161  0.00157 **
LogDose                      1.6655     0.1345  12.378  < 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: 407.129  on 17  degrees of freedom
Residual deviance:  29.848  on 14  degrees of freedom
AIC: 101.11

Number of Fisher Scoring iterations: 5
```

## Part H

Under the logistic model, estimate the combination dose required to give a 99% kill rate, and obtain a 90% confidence interval for this dose.

### Solution

**Calculating the Dose for 99% Kill Rate:**

The dose $D$ required for a kill rate $p$ satisfies:

$$\text{logit}(p) = \beta_{\text{comb}} + \beta_1 \times \log(D)$$

Solving for $\log(D)$:

$$\log(D) = \frac{\text{logit}(p) - \beta_{\text{comb}}}{\beta_1}$$

**Calculations:**

```
# Target kill rate
p <- 0.99
logit_p <- qlogis(p)
# Estimate log(D)
logD <- (logit_p - beta_comb) / beta_logdose
Dose_99 <- exp(logD)
# Variance of log(D)
var_logD <- (cov_matrix["InsecticideDDT + gamma-BHC",
                        "InsecticideDDT + gamma-BHC"] +
             (logD)^2 * cov_matrix["LogDose", "LogDose"] -
             2 * logD * cov_matrix["InsecticideDDT + gamma-BHC",
                                   "LogDose"]) / beta_logdose^2
# Standard error
se_logD <- sqrt(var_logD)
# 90% Confidence Interval
z_value <- qnorm(0.95)
lower_logD <- logD - z_value * se_logD
upper_logD <- logD + z_value * se_logD
lower_Dose <- exp(lower_logD)
upper_Dose <- exp(upper_logD)
```

```
Dose Estimate: 1.787
```

```
90% Confidence Interval: [1.552915, 2.056]
```

**Results:**

- **Estimated Dose for 99% Kill Rate:** 1.787 mg/10 $cm^2$
- **90% Confidence Interval:** (1.553), 2.056) mg/10 $cm^2$

## Part I

Provide a brief summary of your conclusions regarding the effectiveness of these three insecticides.

### Solution

The combination of DDT and gamma-BHC is significantly more effective than using either insecticide alone. This analysis shows that this mixture is much more potent, requiring lower doses to achieve high kill rates. The increased potency is confirmed by the confidence intervals from both the Delta method and Fieller's method. Additionally, the assumption of parallel lines in our model is appropriate, which simplifies the analysis without sacrificing accuracy. Practically speaking, using the combined insecticide could lead to more efficient pest control and potentially reduce environmental impact due to the lower doses needed.