Homework 3 P3

Amy Ly

2/9/2023

Problem Context

In an experiment, flour beetles Tribolium castaneum were sprayed with one of thre 3 insecticides in a solution at different doses The number of insects killed after a 6-day period is recorded below:

Insecticide	2.00	2.64	3.48	4.59	6.06	8.00
DDT	3/50	5/49	19/49	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	49/50	50/50

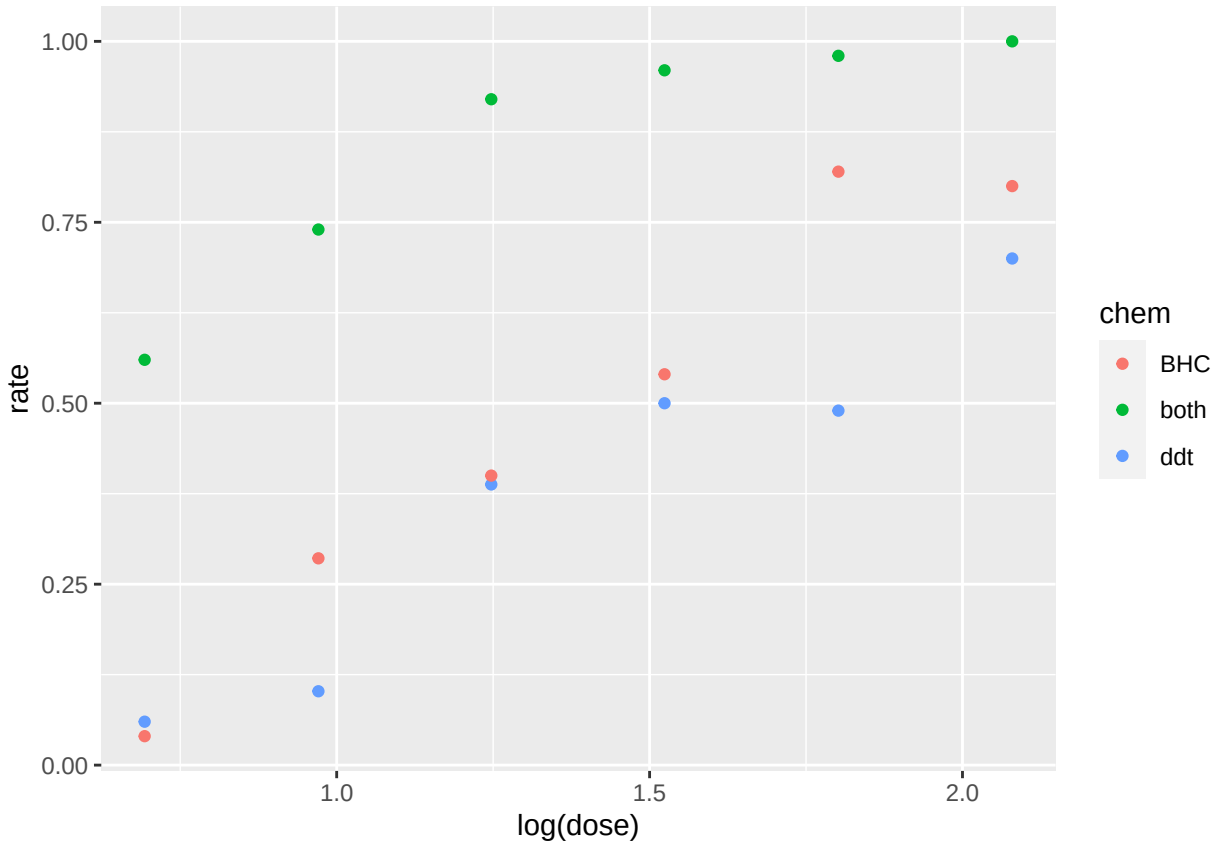
Note that the third insecticide is a combination of the first 2.

```
df <- data.frame(dose = c(2.00, 2.64, 3.48, 4.59, 6.06, 8.00),
                 ddt = c(3/50, 5/49, 19/49, 19/38, 24/49, 35/50),
                 BHC = c(2/50, 14/49, 20/50, 27/50, 41/50, 40/50),
                 both = c(28/50, 37/50, 46/50, 48/50, 49/50, 50/50)
                 )

df <- df %>%
  pivot_longer(!dose, names_to = "chem", values_to = "rate")

df <- df %>%
  mutate(
    dead = c(3, 2, 28, 5, 14, 37, 19, 20, 46, 19, 27, 48, 24, 41, 49, 35, 40, 50),
    tot = c(50, 50, 50, 49, 49, 50, 49, 50, 50, 38, 50, 50, 49, 50, 50, 50, 50, 50),
    alive = tot-dead
  )
```

Part a

Perform exploratory data analysis and plot the relationship between dose and kill rate.

```
ggplot(data=df, aes(x=log(dose), y=rate, color=chem)) +
  geom_point()
```

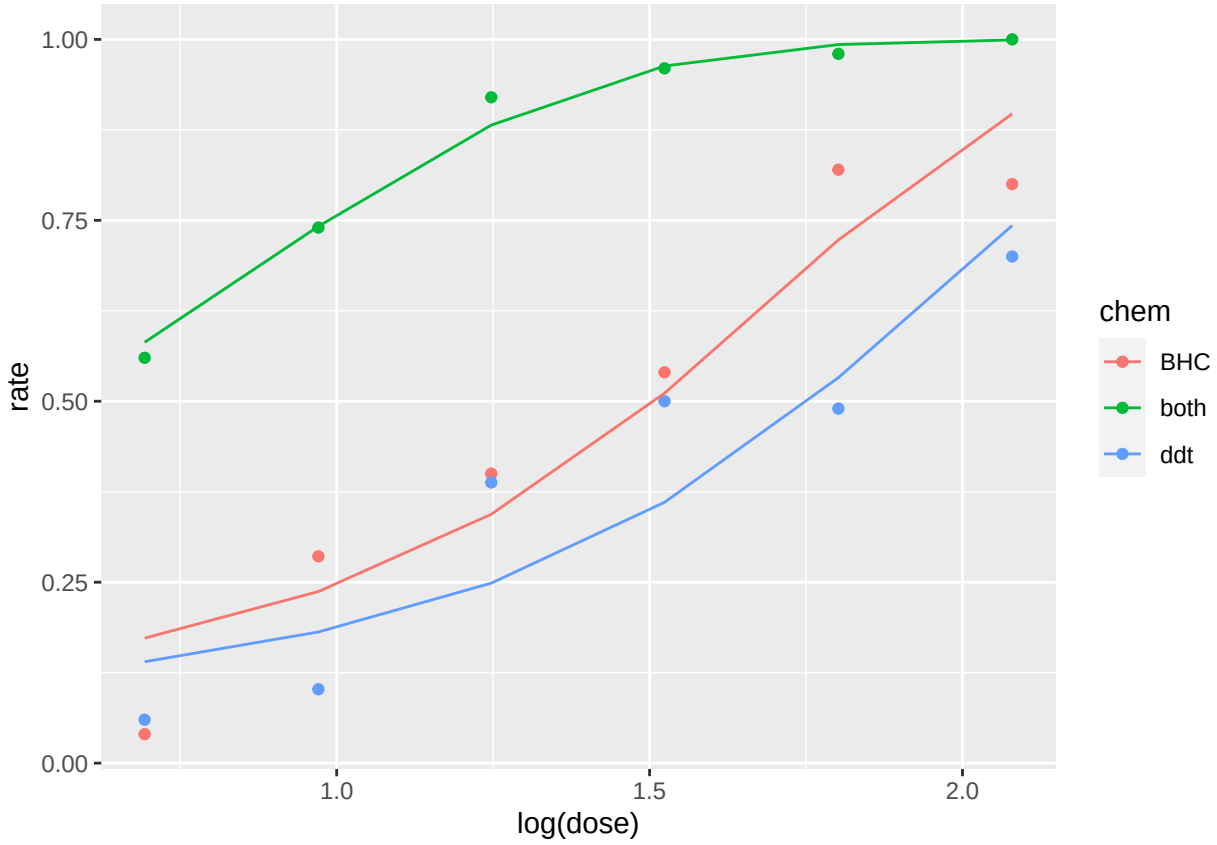


Part b

Plot the logistic fitted curve for each of the insecticides.

```
#data wrangling

#remember that the order of the y matrix matters. c(dead, alive)

logit.y_ddt <- df %>%
  filter(chem == "ddt") %>%
  glm(cbind(dead, alive) ~ dose, family=binomial, data=.)

df_ddt <- df %>%
  filter(chem == "ddt") %>%
  select(dose) %>%
  mutate(pred_rate = as.numeric(logit.y_ddt$fit),
         chem = "ddt")

logit.y_BHC <- df %>%
  filter(chem == "BHC") %>%
  glm(cbind(dead, alive) ~ dose, family=binomial, data=.)

df_BHC <- df %>%
  filter(chem == "BHC") %>%
  select(dose) %>%
```

```
  mutate(pred_rate = as.numeric(logit.y_BHC$fit),
         chem = "BHC")

logit.y_both <- df %>%
  filter(chem == "both") %>%
  glm(cbind(dead, alive) ~ dose, family=binomial, data=.)

df_both <- df %>%
  filter(chem == "both") %>%
  select(dose) %>%
  mutate(pred_rate = as.numeric(logit.y_both$fit),
         chem = "both")

#merge the dataframes into df

df_temp <- rbind(df_BHC, df_both, df_ddt)

df <- merge(df, df_temp, by = c("chem", "dose"))

#plot the logistic fitted curves

ggplot(data=df, aes(x=log(dose), y=rate, color=chem)) +
   geom_point() +
  geom_line(aes(x=log(dose), y=pred_rate, color=chem))
```



```
#lines(data$temperature, logit.y$fit)
```

Part c

Consider two models, one in which the relationship is described by three parallel straight lines in the log dose and one in which the three lines are straight but not parallel. Assess the evidence against the hypothesis of parallelism.

Let BHC be the reference category, then assume we have a logit model that looks like:

$log\frac{Y_i}{1-Y_i} = p_i = \alpha + +\beta_1 I_{DDT} + \beta_2 I_{both} + \beta_3 ldose_i + \beta_4 I_{DDT} ldose_i + \beta_5 I_{both} ldose_i + \varepsilon$

If the lines are parallel, then $\beta_4 = 0$ and $\beta_5 = 0$ where all the lines would have different α intercept values.

```
df <- df %>%
  mutate(ldose = log(dose))

logit.y <- df %>%
  glm(cbind(dead, alive) ~ chem*ldose, family=binomial, data=.)

logit.y_constrain <- df %>%
  glm(cbind(dead, alive) ~ chem + ldose, family=binomial, data=.)

# Do LRT test for beta_4 and beta_5

anova(logit.y_constrain,logit.y, test = "LRT")
```

```
## Analysis of Deviance Table
##
## Model 1: cbind(dead, alive) ~ chem + ldose
## Model 2: cbind(dead, alive) ~ chem * ldose
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        14     21.877
## 2        12     17.289  2   4.5881   0.1009
```

The p-value from the likelihood ratio test indicates that there is little evidence for parallelism (p-value = 0.1009).

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 formula chem + ldose and chem + ldose -1. Explain the relationship between the two covariance matrices.

In these two models, we are only looking at how α changes.

In the model $y \sim chem + ldose$, BHC was considered as the reference category.

When the intercept is removed in $y \sim chem + ldose - 1$, BHC shows up as a coefficient and the estimate is very similar to that of the intercept.

Due to the different parameterization, the two covariance matrices will be different because the α values are different for each chemical.

```r
logit.y_a <- df %>%
  glm(cbind(dead, alive) ~ chem +ldose, family=binomial, data=.)

logit.y_b <- df %>%
  glm(cbind(dead, alive) ~ chem +ldose -1, family=binomial, data=.)

logit.y_a
```

```
##
## Call:  glm(formula = cbind(dead, alive) ~ chem + ldose, family = binomial,
##     data = .)
##
## Coefficients:
## (Intercept)     chemboth      chemddt        ldose
##     -3.8993       2.4649      -0.6297       2.7361
##
## Degrees of Freedom: 17 Total (i.e. Null);  14 Residual
## Null Deviance:       414.9
## Residual Deviance: 21.88     AIC: 92.57
```

```r
logit.y_b
```

```
##
## Call:  glm(formula = cbind(dead, alive) ~ chem + ldose - 1, family = binomial,
##     data = .)
##
## Coefficients:
##  chemBHC  chemboth   chemddt     ldose
##   -3.899    -1.434    -4.529     2.736
##
## Degrees of Freedom: 18 Total (i.e. Null);  14 Residual
## Null Deviance:       434.1
## Residual Deviance: 21.88     AIC: 92.57
```

```r
vcov(logit.y_a)
```

```
##             (Intercept)    chemboth      chemddt        ldose
## (Intercept)  0.11179636 -0.04388970 -0.010869212 -0.066169254
## chemboth    -0.04388970  0.05778147  0.016537436  0.017932528
## chemddt     -0.01086921  0.01653744  0.040151818 -0.005523201
## ldose       -0.06616925  0.01793253 -0.005523201  0.047002580
```

```r
vcov(logit.y_b)
```

```
##              chemBHC    chemboth     chemddt       ldose
## chemBHC   0.11179636  0.06790666  0.10092715 -0.06616925
## chemboth  0.06790666  0.08179843  0.07357488 -0.04823673
## chemddt   0.10092715  0.07357488  0.13020975 -0.07169245
## ldose    -0.06616925 -0.04823673 -0.07169245  0.04700258
```

Part e

On the assumption that 3 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. I will be defining the following variables:

x_1 = amount of γ-BHC

x_2 = amount of both

x_3 = amount of DDT

If we are estimating the potency of the combination relative to BHC for example, then:

$$x_2 = x_1 p_1(x_2) = P_2(x_1) log(\frac{P_2(x_2)}{1 - P_2(x_2)}) = log(\frac{P_1(x_1)}{1 - P_1(x_1)})\alpha_1 + \beta log(x_2) = \alpha_1 + \beta log(x_1)\frac{x_2}{x_1} = e^{(\alpha_2 - \alpha_1)/\beta} \tag{1}$$

Then, when we calculate ∇g:

$$\nabla g = (\frac{dg(\alpha,\beta)}{d\alpha_1}, \frac{dg(\alpha,\beta)}{d\alpha_2}, \frac{dg(\alpha,\beta)}{d\alpha_3}, \frac{dg(\alpha,\beta)}{d\beta}) = (\frac{-e^{-(\alpha_1-\alpha_2)/\beta}}{\beta}, \frac{e^{-(\alpha_1-\alpha_2)/\beta}}{\beta}, 0, \frac{(\alpha_1 - \alpha_2)e^{-(\alpha_1-\alpha_2)/\beta}}{\beta^2}) \tag{2}$$

```
mod <- df %>%
  glm(cbind(dead, alive) ~ chem + ldose-1, family=binomial, data=.)

a_1 <- mod$coefficients[1]   #alpha for BHC
a_2 <- mod$coefficients[2]   #alpha for both
a_3 <- mod$coefficients[3]   #alpha for DDT
b <- mod$coefficients[4]     #beta
V <- vcov(mod)
z <- qnorm(0.95) #for 90# confidence interval

#x_both / x_BHC

theta1 <- exp((a_1-a_2)/b)
g.prime1 <- c(-exp(-(a_1-a_2)/b), exp(-(a_1-a_2)/b), 0,  (a_1-a_2)*exp(-(a_1-a_2)/b)/b^2)
theta.sd1 <- sqrt(g.prime1%*%V%*%g.prime1)

theta1
```

```
##   chemBHC
## 0.4062076
```

```
CI1 <- c(theta1-z*theta.sd1, theta1 +z*theta.sd1)
CI1
```

```
## [1] -0.5088628  1.3212780
```

```
#x_both / x_DDT
theta2 <- exp((a_3-a_2)/b)
g.prime2 <- c(0, exp(-(a_3-a_2)/b), -exp(-(a_3-a_2)/b),  (a_3-a_2)*exp(-(a_3-a_2)/b)/b^2)
```

```
theta.sd2 <- sqrt(g.prime2%*%V%*%g.prime2)

theta2
```

```
##   chemddt
## 0.3226938
```

```
CI2 <- c(theta2-z*theta.sd2, theta2 + z*theta.sd2)
CI2
```

```
## [1] -0.8562846  1.5016723
```

The potency of combination relative to BHC is:

0.4062076 with a CI of (-0.5088628, 1.321278).

The potency of combination relative to BHC is:

0.3226938 with a CI of (-0.8562846, 1.5016723).

Part f

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

For a fixed $log\tau$ where $\tau = \frac{x_2}{x_1}$ which is the potency of combination relative to BHC, I am trying to solve $\alpha_1 - \alpha_2 + \beta log\tau = 0$

The variance of $\hat{\alpha_1} - \hat{\alpha_2} + \hat{\beta} log\tau = 0$ can be found with:

$Var(\alpha_1) + Var(\alpha_2) + log^2\tau Var(\beta) - 2Cov(\alpha_1, \alpha_2) - 2log\tau Cov(\alpha_1, \beta) - 2log\tau Cov(\alpha_2, \beta)$

The (1-a)100% CI for $log\tau$ is then given by:

$\frac{\alpha_1^2 - 2\alpha_1\alpha_2 + 2\alpha_1\beta log\tau - 2\alpha_2\beta log\tau + \alpha_2^2 + \beta^2 log^2\tau}{Var(\alpha_1) + Var(\alpha_2) + log^2\tau Var(\beta) - 2Cov(\alpha_1,\alpha_2) + 2log\tau Cov(\alpha_1,\beta) - 2log\tau Cov(\alpha_2,\beta)}$

```
mod <- df %>%
  glm(cbind(dead, alive) ~ chem + ldose-1, family=binomial, data=.)

# CI for LD50

a_1 <- mod$coefficients[1]   #alpha for BHC
a_2 <- mod$coefficients[2]   #alpha for both
a_3 <- mod$coefficients[3]   #alpha for DDT
b <- mod$coefficients[4]     #beta
V <- vcov(mod)

#log (x_BHC/x_both) CI

z <- qnorm(0.95)
tau_1 <- as.numeric(polyroot(c(a_1^2 - 2*a_1*a_2 + a_2^2 -V[1,1]*z^2 - V[2,2]*z^2 + 2*V[1,2]*z^2, 2*b*a_

#transformed interval to get CI for x_BHC/x_both
theta1
```

```
##   chemBHC
## 0.4062076
```

```
tau_1
```

```
## [1] 0.7566686 1.0630628
```

```
# log(x_DDT/x_both) CI

tau_2 <- as.numeric(polyroot(c(a_3^2 - 2*a_3*a_2 + a_2^2 +(-V[3,3] - V[2,2] + 2*V[3,2])*z^2, 2*b*a_3 -2

#transformed interval to get CI for x_DDT/x_both
theta2
```

```
##   chemddt
## 0.3226938
```

```
tau_2
```

```
## [1] 0.9789905 1.3049549
```

The potency of combination relative to BHC is:

0.4062076 with a CI of (0.7566686, 1.0630628).

The potency of combination relative to BHC is:

0.3226938 with a CI of (0.9789905, 1.3049549).

Part g

Redo part e and part f using the c-log-log link.

```
mod <- df %>%
  glm(cbind(dead, alive) ~ chem + ldose-1, family=binomial(link="cloglog"), data=.)

a_1 <- mod$coefficients[1]   #alpha for BHC
a_2 <- mod$coefficients[2]   #alpha for both
a_3 <- mod$coefficients[3]   #alpha for DDT
b <- mod$coefficients[4]     #beta
V <- vcov(mod)
z <- qnorm(0.95) #for 90% confidence interval

###Delta method CI

#x_both / x_BHC

theta3 <- exp((a_1-a_2)/b)
g.prime1 <- c(-exp(-(a_1-a_2)/b), exp(-(a_1-a_2)/b), 0,  (a_1-a_2)*exp(-(a_1-a_2)/b)/b^2)
theta.sd1 <- sqrt(g.prime1%*%V%*%g.prime1)

theta3
```

```
##   chemBHC
## 0.4177552
```

```
CI3 <- c(theta1-z*theta.sd1, theta1 +z*theta.sd1)
CI3
```

```
## [1] -0.1012795  0.9136947
```

```
#x_both / x_DDT
theta4 <- exp((a_3-a_2)/b)
g.prime2 <- c(0, exp(-(a_3-a_2)/b), -exp(-(a_3-a_2)/b),  (a_3-a_2)*exp(-(a_3-a_2)/b)/b^2)
theta.sd2 <- sqrt(g.prime2%*%V%*%g.prime2)

theta4
```

```
##   chemddt
## 0.3248904
```

```
CI4 <- c(theta2-z*theta.sd2, theta2 + z*theta.sd2)
CI4
```

```
## [1] -0.3788437  1.0242313
```

```
#Fieller's CI

#log (x_BHC/x_both) CI

z <- qnorm(0.95)
tau_3 <- polyroot(c(a_1^2 - 2*a_1*a_2 + a_2^2 +(-V[1,1] - V[2,2] + 2*V[1,2])*z^2, 2*b*a_1 -2*b*a_2+ (-2

#transformed interval to get CI for x_BHC/x_both
theta3
```

```
##   chemBHC
## 0.4177552
```

```
tau_3
```

```
## [1] 0.7446513-0i 1.0156092+0i
```

```
# log(x_DDT/x_both) CI

tau_4 <- polyroot(c(a_3^2 - 2*a_3*a_2 + a_2^2 +(-V[3,3] - V[2,2] + 2*V[3,2])*z^2, 2*b*a_3 -2*b*a_2+ (-2

#transformed interval to get CI for x_DDT/x_both
theta4
```

```
##   chemddt
## 0.3248904
```

```
tau_4
```

```
## [1] 0.980210-0i 1.289477+0i
```

With the Delta Method:

The potency of combination relative to BHC is:

0.4177552 with a CI of (-0.1012795, 0.9136947).

The potency of combination relative to BHC is:

0.3248904 with a CI of (-0.3788437, 1.0242313).

With the Fieller's Method:

The potency of combination relative to BHC is:

0.4177552 with a CI of (0.744651341116445-0i, 1.01560918110637+0i).

The potency of combination relative to BHC is:

0.3248904 with a CI of (0.980209971516138-0i, 1.28947674202969+0i).

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.

We are going to solve the following equation to the the ldose required to have a 99% kill rate.

$$p(x) = 0.99 = \frac{e^{\alpha_2+\beta_2 ldose}}{1+e^{\alpha_2+\beta_2 ldose}} 0.99(1+e^{\alpha_2+\beta_2 ldose}) = e^{\alpha_2+\beta_2 ldose} 0.99 = 0.01 e^{\alpha_2+\beta_2 ldose} 99 = e^{\alpha_2+\beta_2 ldose} log(99) = \alpha_2+\beta_2 ldose \beta \tag{3}$$

Note that $exp(\frac{log(99)-\alpha_2}{\beta_2}) = exp(\frac{log(99)-(\beta_1+\beta_2)}{\beta_4+\beta_5})$ based on our logistic model parameters.

```
logit.y <- df %>%
  glm(cbind(dead, alive) ~ chem*ldose, family=binomial, data=.)

a_2 <- logit.y$coef[1] +logit.y$coef[2]
b_2 <- logit.y$coef[4]+logit.y$coef[5]
V <- vcov(logit.y)
z <- qnorm(0.95)


dose <- exp((log(99) -a_2)/b_2)
dose # Combination dose in terms of mg/10cm^2
```

```
## (Intercept)
##    6.629417
```

```
theta5 <- (log(99)-a_2)/b_2

g.prime3 <- c(-(1/b_2), -(1/b_2), 0,-((log(99) -a_2)/b_2^2), -((log(99) -a_2)/b_2^2), 0)

theta.sd3 <- sqrt(g.prime3%*%V%*%g.prime3)

CI <- c(theta3-z*theta.sd3, theta3 +z*theta.sd3) #for ldose

theta5
```

```
## (Intercept)
##    1.891517
```

```
CI5 <- exp(CI) #cI for dose
CI5
```

```
## [1] 1.158215 1.990986
```

The desired drug dose is :

1.8915169 with a CI of (1.1582154, 1.9909861).

Part i

Give a brief summary of your conclusions regarding the effectiveness of these 3 insecticides.

Based on the Fieller's confidence intervals, 1 is a possible value and it suggests that it may be possible that there is no difference in effectiveness. It is possible that the combination dose is just as potent relative to γ-BHC or DDT by itself since it produces the same effect at a similar concentration levels.

Based on the visualizations of the plot, the general trend of kill rate for the combination is higher than γ-BHC or DDT.

There is suggestive evidence to suggest that the combination of chemicals is more effective at killing pests.