Regression modeling

BP

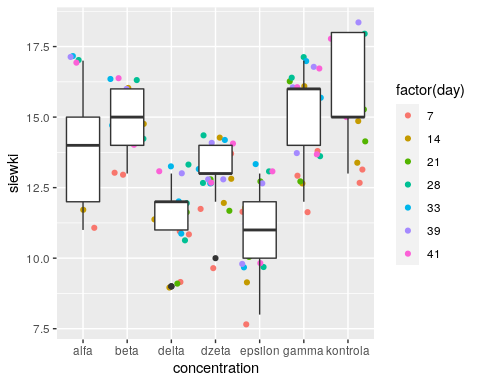
16 03 2020

# 1. Czy dodatek hormonu ma wpływ na liczbę powstających siewek?

library(ggplot2)  
library(sandwich)  
  
kielkowanie <- read.csv("./kielkowanie.csv")  
  
ggplot(data = kielkowanie) + geom\_jitter(aes(x=concentration, y=siewki, color = factor(day))) + geom\_boxplot(aes(x=concentration, y=siewki))

## Warning: Removed 8 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 8 rows containing missing values (geom\_point).



## 1.1 Przez cały okres trwania eksperymentu

### Poisson regression

m1 = glm(siewki ~ concentration + day, family="poisson", data=kielkowanie)  
summary(m1)

##   
## Call:  
## glm(formula = siewki ~ concentration + day, family = "poisson",   
## data = kielkowanie)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.73195 -0.28574 -0.02549 0.28656 0.74438   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.539339 0.079077 32.112 < 2e-16 \*\*\*  
## concentrationbeta 0.062177 0.085555 0.727 0.46738   
## concentrationdelta -0.204842 0.089157 -2.298 0.02159 \*   
## concentrationdzeta -0.072353 0.086208 -0.839 0.40131   
## concentrationepsilon -0.238743 0.089959 -2.654 0.00796 \*\*   
## concentrationgamma 0.073421 0.083287 0.882 0.37803   
## concentrationkontrola 0.125037 0.086640 1.443 0.14897   
## day 0.003848 0.001977 1.946 0.05161 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 51.757 on 138 degrees of freedom  
## Residual deviance: 17.940 on 131 degrees of freedom  
## (8 observations deleted due to missingness)  
## AIC: 651.1  
##   
## Number of Fisher Scoring iterations: 4

cov.m1 <- vcovHC(m1, type="HC0")  
std.err <- sqrt(diag(cov.m1))  
r.est <- cbind(Estimate= coef(m1), "Robust SE" = std.err,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m1)/std.err), lower.tail=FALSE),  
LL = coef(m1) - 1.96 \* std.err,  
UL = coef(m1) + 1.96 \* std.err)  
  
r.est

## Estimate Robust SE Pr(>|z|) LL  
## (Intercept) 2.539338582 0.0317842076 0.000000e+00 2.4770415349  
## concentrationbeta 0.062177245 0.0321017552 5.276002e-02 -0.0007421952  
## concentrationdelta -0.204841922 0.0339351742 1.577159e-09 -0.2713548633  
## concentrationdzeta -0.072352739 0.0313775761 2.111767e-02 -0.1338527880  
## concentrationepsilon -0.238743474 0.0381695995 3.980081e-10 -0.3135558887  
## concentrationgamma 0.073420929 0.0336561852 2.914654e-02 0.0074548056  
## concentrationkontrola 0.125036908 0.0381736591 1.054840e-03 0.0502165360  
## day 0.003847757 0.0007413728 2.102451e-07 0.0023946662  
## UL  
## (Intercept) 2.601635629  
## concentrationbeta 0.125096685  
## concentrationdelta -0.138328980  
## concentrationdzeta -0.010852690  
## concentrationepsilon -0.163931058  
## concentrationgamma 0.139387052  
## concentrationkontrola 0.199857279  
## day 0.005300847

## W poszczególne dni:

### Dzień 7

m1\_d7 = glm(siewki ~ concentration, family="poisson", data= subset(kielkowanie, day == 7))  
summary(m1\_d7)

##   
## Call:  
## glm(formula = siewki ~ concentration, family = "poisson", data = subset(kielkowanie,   
## day == 7))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.6555 -0.2810 0.0000 0.2052 0.6841   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.45674 0.16903 14.534 <2e-16 \*\*\*  
## concentrationbeta 0.18232 0.22887 0.797 0.426   
## concentrationdelta -0.12136 0.24664 -0.492 0.623   
## concentrationdzeta 0.02817 0.23738 0.119 0.906   
## concentrationepsilon -0.15415 0.24881 -0.620 0.536   
## concentrationgamma 0.10821 0.23284 0.465 0.642   
## concentrationkontrola 0.20585 0.22766 0.904 0.366   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 7.3218 on 20 degrees of freedom  
## Residual deviance: 3.0946 on 14 degrees of freedom  
## AIC: 108.19  
##   
## Number of Fisher Scoring iterations: 4

cov.m1\_d7 <- vcovHC(m1\_d7, type="HC0")  
std.err\_m1d7 <- sqrt(diag(cov.m1\_d7))  
r.est\_m1d7 <- cbind(Estimate= coef(m1\_d7), "Robust SE" = std.err\_m1d7,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m1\_d7)/std.err\_m1d7), lower.tail=FALSE),  
LL = coef(m1\_d7) - 1.96 \* std.err\_m1d7,  
UL = coef(m1\_d7) + 1.96 \* std.err\_m1d7)  
  
r.est\_m1d7

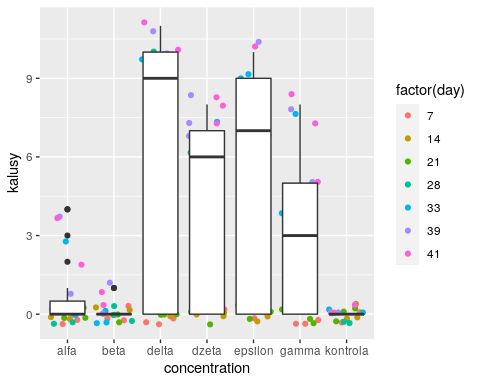
## Estimate Robust SE Pr(>|z|) LL  
## (Intercept) 2.45673577 0.02332847 0.000000000 2.41101196  
## concentrationbeta 0.18232156 0.06281384 0.003701145 0.05920643  
## concentrationdelta -0.12136086 0.05761167 0.035158132 -0.23427974  
## concentrationdzeta 0.02817088 0.08195765 0.731053273 -0.13246613  
## concentrationepsilon -0.15415068 0.09712418 0.112478593 -0.34451408  
## concentrationgamma 0.10821358 0.04311777 0.012082649 0.02370276  
## concentrationkontrola 0.20585205 0.07945503 0.009575325 0.05012020  
## UL  
## (Intercept) 2.502459581  
## concentrationbeta 0.305436679  
## concentrationdelta -0.008441976  
## concentrationdzeta 0.188807881  
## concentrationepsilon 0.036212715  
## concentrationgamma 0.192724414  
## concentrationkontrola 0.361583913

# 2. Czy dodatek hormonu ma wpływ na liczbę powstających kalusów?

ggplot(data = kielkowanie) + geom\_jitter(aes(x=concentration, y=kalusy, color = factor(day))) + geom\_boxplot(aes(x=concentration, y=kalusy))

## Warning: Removed 8 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 8 rows containing missing values (geom\_point).



## 2.1 Przez cały okres trwania eksperymentu

### Poisson regression

m2 = glm(kalusy ~ concentration + day, family="poisson", data=kielkowanie)  
summary(m2)

##   
## Call:  
## glm(formula = kalusy ~ concentration + day, family = "poisson",   
## data = kielkowanie)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.88249 -0.90811 -0.40353 -0.00006 2.71610   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.478e+00 3.705e-01 -9.388 < 2e-16 \*\*\*  
## concentrationbeta -1.946e+00 7.559e-01 -2.574 0.010 \*   
## concentrationdelta 1.863e+00 2.831e-01 6.583 4.62e-11 \*\*\*  
## concentrationdzeta 1.504e+00 2.896e-01 5.194 2.05e-07 \*\*\*  
## concentrationepsilon 1.684e+00 2.860e-01 5.889 3.89e-09 \*\*\*  
## concentrationgamma 1.221e+00 2.965e-01 4.117 3.84e-05 \*\*\*  
## concentrationkontrola -1.768e+01 1.189e+03 -0.015 0.988   
## day 1.041e-01 7.364e-03 14.143 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 759.07 on 138 degrees of freedom  
## Residual deviance: 145.58 on 131 degrees of freedom  
## (8 observations deleted due to missingness)  
## AIC: 362.41  
##   
## Number of Fisher Scoring iterations: 16

cov.m2 <- vcovHC(m2, type="HC0")  
std.err\_m2 <- sqrt(diag(cov.m2))  
r.est\_m2 <- cbind(Estimate= coef(m2), "Robust SE" = std.err\_m2,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m2)/std.err\_m2), lower.tail=FALSE),  
LL = coef(m2) - 1.96 \* std.err\_m2,  
UL = coef(m2) + 1.96 \* std.err\_m2)  
  
r.est\_m2

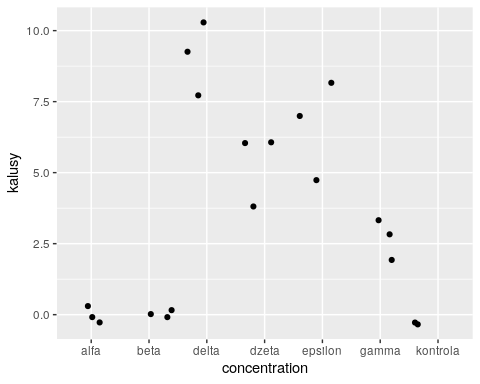
## Estimate Robust SE Pr(>|z|) LL  
## (Intercept) -3.4782072 0.400461216 3.770689e-18 -4.26311122  
## concentrationbeta -1.9459101 0.631802006 2.070446e-03 -3.18424208  
## concentrationdelta 1.8632823 0.312717398 2.547842e-09 1.25035624  
## concentrationdzeta 1.5041413 0.304516498 7.834681e-07 0.90728897  
## concentrationepsilon 1.6844031 0.310161027 5.612163e-08 1.07648751  
## concentrationgamma 1.2205660 0.307214881 7.097180e-05 0.61842484  
## concentrationkontrola -17.6848141 0.451330466 0.000000e+00 -18.56942186  
## day 0.1041413 0.008324455 6.558437e-36 0.08782541  
## UL  
## (Intercept) -2.6933033  
## concentrationbeta -0.7075782  
## concentrationdelta 2.4762084  
## concentrationdzeta 2.1009936  
## concentrationepsilon 2.2923187  
## concentrationgamma 1.8227072  
## concentrationkontrola -16.8002064  
## day 0.1204573

## W poszczególne dni:

### Dzień 28

ggplot(data= subset(kielkowanie, day == 28)) + geom\_jitter(aes(x=concentration, y=kalusy))

## Warning: Removed 1 rows containing missing values (geom\_point).



m2\_d28 = glm(kalusy ~ concentration, family="poisson", data= subset(kielkowanie, day == 28))  
summary(m2\_d28)

##   
## Call:  
## glm(formula = kalusy ~ concentration, family = "poisson", data = subset(kielkowanie,   
## day == 28))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.67566 -0.00003 -0.00003 0.20008 0.50048   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -2.130e+01 1.479e+04 -0.001 0.999  
## concentrationbeta -4.026e-10 2.092e+04 0.000 1.000  
## concentrationdelta 2.350e+01 1.479e+04 0.002 0.999  
## concentrationdzeta 2.298e+01 1.479e+04 0.002 0.999  
## concentrationepsilon 2.320e+01 1.479e+04 0.002 0.999  
## concentrationgamma 2.228e+01 1.479e+04 0.002 0.999  
## concentrationkontrola 6.961e-10 2.339e+04 0.000 1.000  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 85.6230 on 19 degrees of freedom  
## Residual deviance: 1.7341 on 13 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 58.287  
##   
## Number of Fisher Scoring iterations: 19

cov.m2\_d28 <- vcovHC(m2\_d28, type="HC0")  
std.err\_m2d28 <- sqrt(diag(cov.m2\_d28))

## Warning in sqrt(diag(cov.m2\_d28)): wyprodukowano wartości NaN

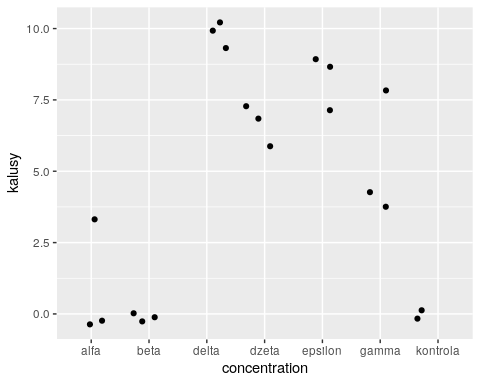
r.est\_m2d28 <- cbind(Estimate= coef(m2\_d28), "Robust SE" = std.err\_m2d28,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m2\_d28)/std.err\_m2d28), lower.tail=FALSE),  
LL = coef(m2\_d28) - 1.96 \* std.err\_m2d28,  
UL = coef(m2\_d28) + 1.96 \* std.err\_m2d28)  
  
r.est\_m2d28

## Estimate Robust SE Pr(>|z|) LL UL  
## (Intercept) -2.130259e+01 NaN NaN NaN NaN  
## concentrationbeta -4.025951e-10 NaN NaN NaN NaN  
## concentrationdelta 2.349981e+01 NaN NaN NaN NaN  
## concentrationdzeta 2.297656e+01 NaN NaN NaN NaN  
## concentrationepsilon 2.319971e+01 NaN NaN NaN NaN  
## concentrationgamma 2.228341e+01 NaN NaN NaN NaN  
## concentrationkontrola 6.961069e-10 NaN NaN NaN NaN

### Dzień 33

ggplot(data= subset(kielkowanie, day == 33)) + geom\_jitter(aes(x=concentration, y=kalusy))

## Warning: Removed 1 rows containing missing values (geom\_point).



m2\_d33 = glm(kalusy ~ concentration, family="poisson", data= subset(kielkowanie, day == 33))  
summary(m2\_d33)

##   
## Call:  
## glm(formula = kalusy ~ concentration, family = "poisson", data = subset(kielkowanie,   
## day == 33))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.41421 -0.31579 -0.00009 0.12805 1.60987   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.160e-15 5.774e-01 0.000 1.000000   
## concentrationbeta -1.930e+01 5.442e+03 -0.004 0.997170   
## concentrationdelta 2.269e+00 6.065e-01 3.741 0.000183 \*\*\*  
## concentrationdzeta 1.897e+00 6.191e-01 3.064 0.002183 \*\*   
## concentrationepsilon 2.120e+00 6.110e-01 3.470 0.000520 \*\*\*  
## concentrationgamma 1.674e+00 6.292e-01 2.661 0.007798 \*\*   
## concentrationkontrola -1.930e+01 6.666e+03 -0.003 0.997689   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 90.1690 on 19 degrees of freedom  
## Residual deviance: 8.9775 on 13 degrees of freedom  
## (1 observation deleted due to missingness)  
## AIC: 71.994  
##   
## Number of Fisher Scoring iterations: 17

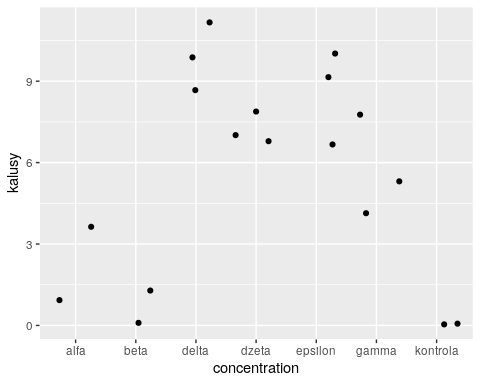
cov.m2\_d33 <- vcovHC(m2\_d33, type="HC0")  
std.err\_m2d33 <- sqrt(diag(cov.m2\_d33))  
r.est\_m2d33 <- cbind(Estimate= coef(m2\_d33), "Robust SE" = std.err\_m2d33,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m2\_d33)/std.err\_m2d33), lower.tail=FALSE),  
LL = coef(m2\_d33) - 1.96 \* std.err\_m2d33,  
UL = coef(m2\_d33) + 1.96 \* std.err\_m2d33)  
  
r.est\_m2d33

## Estimate Robust SE Pr(>|z|) LL  
## (Intercept) -3.160052e-15 0.8164966 1.000000e+00 -1.60033330  
## concentrationbeta -1.930259e+01 1.0000000 5.108831e-83 -21.26258509  
## concentrationdelta 2.268684e+00 0.8169819 5.487870e-03 0.66739908  
## concentrationdzeta 1.897120e+00 0.8175166 2.030903e-02 0.29478752  
## concentrationepsilon 2.120264e+00 0.8191052 9.639163e-03 0.51481734  
## concentrationgamma 1.673976e+00 0.8416254 4.670337e-02 0.02439063  
## concentrationkontrola -1.930259e+01 1.0801234 1.993995e-71 -21.41962705  
## UL  
## (Intercept) 1.600333  
## concentrationbeta -17.342585  
## concentrationdelta 3.869968  
## concentrationdzeta 3.499452  
## concentrationepsilon 3.725710  
## concentrationgamma 3.323562  
## concentrationkontrola -17.185543

### Dzień 39

ggplot(data= subset(kielkowanie, day == 39)) + geom\_jitter(aes(x=concentration, y=kalusy))

## Warning: Removed 3 rows containing missing values (geom\_point).



m2\_d39 = glm(kalusy ~ concentration, family="poisson", data= subset(kielkowanie, day == 39))  
summary(m2\_d39)

##   
## Call:  
## glm(formula = kalusy ~ concentration, family = "poisson", data = subset(kielkowanie,   
## day == 39))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.08047 -0.31276 -0.00009 0.29402 0.92238   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.9163 0.4472 2.049 0.04047 \*   
## concentrationbeta -1.6094 1.0954 -1.469 0.14178   
## concentrationdelta 1.3863 0.4830 2.870 0.00411 \*\*  
## concentrationdzeta 1.0761 0.4954 2.172 0.02985 \*   
## concentrationepsilon 1.2432 0.4883 2.546 0.01090 \*   
## concentrationgamma 0.8183 0.5087 1.608 0.10773   
## concentrationkontrola -20.2189 6665.6247 -0.003 0.99758   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 62.1034 on 17 degrees of freedom  
## Residual deviance: 5.6344 on 11 degrees of freedom  
## (3 observations deleted due to missingness)  
## AIC: 73.618  
##   
## Number of Fisher Scoring iterations: 17

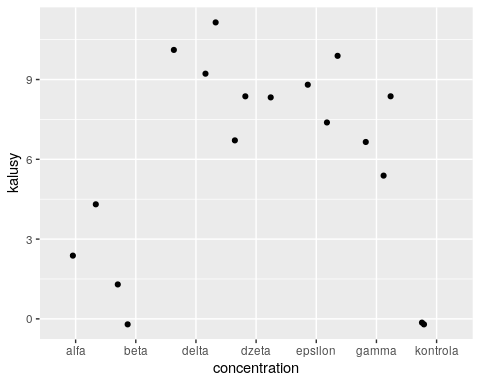
cov.m2\_d39 <- vcovHC(m2\_d39, type="HC0")  
std.err\_m2d39 <- sqrt(diag(cov.m2\_d39))  
r.est\_m2d39 <- cbind(Estimate= coef(m2\_d39), "Robust SE" = std.err\_m2d39,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m2\_d39)/std.err\_m2d39), lower.tail=FALSE),  
LL = coef(m2\_d39) - 1.96 \* std.err\_m2d39,  
UL = coef(m2\_d39) + 1.96 \* std.err\_m2d39)  
  
r.est\_m2d39

## Estimate Robust SE Pr(>|z|) LL  
## (Intercept) 0.9162907 0.4242641 3.079451e-02 0.08473316  
## concentrationbeta -1.6094379 0.8246211 5.097025e-02 -3.22569532  
## concentrationdelta 1.3862944 0.4268749 1.164064e-03 0.54961946  
## concentrationdzeta 1.0761394 0.4258843 1.150954e-02 0.24140627  
## concentrationepsilon 1.2431935 0.4323232 4.032459e-03 0.39584004  
## concentrationgamma 0.8183103 0.4582450 7.413998e-02 -0.07984985  
## concentrationkontrola -20.2188758 0.8246211 9.267604e-133 -21.83513323  
## UL  
## (Intercept) 1.747848307  
## concentrationbeta 0.006819493  
## concentrationdelta 2.222969261  
## concentrationdzeta 1.910872600  
## concentrationepsilon 2.090546998  
## concentrationgamma 1.716470493  
## concentrationkontrola -18.602618420

### Dzień 41

ggplot(data= subset(kielkowanie, day == 41)) + geom\_jitter(aes(x=concentration, y=kalusy))

## Warning: Removed 3 rows containing missing values (geom\_point).



m2\_d41 = glm(kalusy ~ concentration, family="poisson", data= subset(kielkowanie, day == 41))  
summary(m2\_d41)

##   
## Call:  
## glm(formula = kalusy ~ concentration, family = "poisson", data = subset(kielkowanie,   
## day == 41))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.00000 -0.30240 0.05626 0.26539 0.62153   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.0986 0.4082 2.691 0.00712 \*\*  
## concentrationbeta -1.7918 1.0801 -1.659 0.09715 .   
## concentrationdelta 1.2040 0.4472 2.692 0.00710 \*\*  
## concentrationdzeta 0.9383 0.4584 2.047 0.04068 \*   
## concentrationepsilon 1.0609 0.4529 2.342 0.01916 \*   
## concentrationgamma 0.7985 0.4655 1.715 0.08626 .   
## concentrationkontrola -20.4012 6665.6247 -0.003 0.99756   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 59.5657 on 17 degrees of freedom  
## Residual deviance: 3.6292 on 11 degrees of freedom  
## (3 observations deleted due to missingness)  
## AIC: 72.899  
##   
## Number of Fisher Scoring iterations: 17

cov.m2\_d41 <- vcovHC(m2\_d41, type="HC0")  
std.err\_m2d41 <- sqrt(diag(cov.m2\_d41))  
r.est\_m2d41 <- cbind(Estimate= coef(m2\_d41), "Robust SE" = std.err\_m2d41,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m2\_d41)/std.err\_m2d41), lower.tail=FALSE),  
LL = coef(m2\_d41) - 1.96 \* std.err\_m2d41,  
UL = coef(m2\_d41) + 1.96 \* std.err\_m2d41)  
  
r.est\_m2d41

## Estimate Robust SE Pr(>|z|) LL  
## (Intercept) 1.0986123 0.2357023 3.146504e-06 0.6366359  
## concentrationbeta -1.7917595 0.7453560 1.622132e-02 -3.2526572  
## concentrationdelta 1.2039728 0.2403701 5.476202e-07 0.7328474  
## concentrationdzeta 0.9382696 0.2383606 8.273213e-05 0.4710828  
## concentrationepsilon 1.0608720 0.2499178 2.187070e-05 0.5710331  
## concentrationgamma 0.7985077 0.2592725 2.071389e-03 0.2903336  
## concentrationkontrola -20.4011974 0.7453560 6.061532e-165 -21.8620951  
## UL  
## (Intercept) 1.5605887  
## concentrationbeta -0.3308617  
## concentrationdelta 1.6750982  
## concentrationdzeta 1.4054565  
## concentrationepsilon 1.5507109  
## concentrationgamma 1.3066818  
## concentrationkontrola -18.9402996