Regression modeling

BP

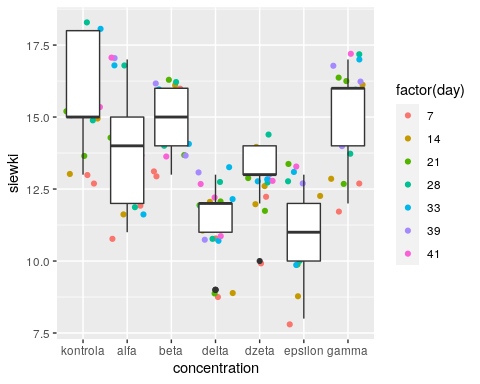
16 03 2020

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

library(forcats)  
library(ggplot2)  
library(sandwich)  
  
  
kielkowanie <- read.csv("./kielkowanie.csv")  
kielkowanie$concentration = fct\_relevel(kielkowanie$concentration, "kontrola", "alfa")  
  
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) 2.664375 0.078013 34.153 < 2e-16 \*\*\*  
## concentrationalfa -0.125037 0.086640 -1.443 0.148972   
## concentrationbeta -0.062860 0.085317 -0.737 0.461260   
## concentrationdelta -0.329879 0.088970 -3.708 0.000209 \*\*\*  
## concentrationdzeta -0.197390 0.086016 -2.295 0.021744 \*   
## concentrationepsilon -0.363780 0.089774 -4.052 5.07e-05 \*\*\*  
## concentrationgamma -0.051616 0.083088 -0.621 0.534455   
## day 0.003848 0.001977 1.946 0.051605 .   
## ---  
## 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) 2.664375490 0.0341266291 0.000000e+00 2.597487297  
## concentrationalfa -0.125036908 0.0381736591 1.054840e-03 -0.199857279  
## concentrationbeta -0.062859663 0.0316322519 4.690006e-02 -0.124858877  
## concentrationdelta -0.329878830 0.0333405091 4.409692e-23 -0.395226227  
## concentrationdzeta -0.197389647 0.0307565600 1.382519e-10 -0.257672504  
## concentrationepsilon -0.363780381 0.0376434066 4.293992e-22 -0.437561458  
## concentrationgamma -0.051615979 0.0330446071 1.182858e-01 -0.116383409  
## day 0.003847757 0.0007413728 2.102451e-07 0.002394666  
## UL  
## (Intercept) 2.7312636826  
## concentrationalfa -0.0502165360  
## concentrationbeta -0.0008604491  
## concentrationdelta -0.2645314318  
## concentrationdzeta -0.1371067890  
## concentrationepsilon -0.2899993045  
## concentrationgamma 0.0131514509  
## day 0.0053008475

## 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) 2.66259 0.15250 17.460 <2e-16 \*\*\*  
## concentrationalfa -0.20585 0.22766 -0.904 0.366   
## concentrationbeta -0.02353 0.21695 -0.108 0.914   
## concentrationdelta -0.32721 0.23561 -1.389 0.165   
## concentrationdzeta -0.17768 0.22591 -0.787 0.432   
## concentrationepsilon -0.36000 0.23788 -1.513 0.130   
## concentrationgamma -0.09764 0.22113 -0.442 0.659   
## ---  
## 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

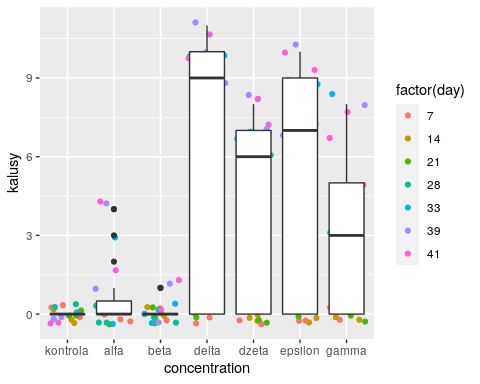
## Estimate Robust SE Pr(>|z|) LL  
## (Intercept) 2.66258783 0.07595317 3.197976e-269 2.5137196  
## concentrationalfa -0.20585205 0.07945503 9.575325e-03 -0.3615839  
## concentrationbeta -0.02353050 0.09576139 8.058989e-01 -0.2112228  
## concentrationdelta -0.32721291 0.09243252 4.000979e-04 -0.5083807  
## concentrationdzeta -0.17768118 0.10927819 1.039597e-01 -0.3918664  
## concentrationepsilon -0.36000273 0.12106929 2.943990e-03 -0.5972985  
## concentrationgamma -0.09763847 0.08416536 2.460167e-01 -0.2626026  
## UL  
## (Intercept) 2.81145604  
## concentrationalfa -0.05012020  
## concentrationbeta 0.16416184  
## concentrationdelta -0.14604517  
## concentrationdzeta 0.03650408  
## concentrationepsilon -0.12270693  
## concentrationgamma 0.06732564

# 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) -2.116e+01 1.189e+03 -0.018 0.986   
## concentrationalfa 1.768e+01 1.189e+03 0.015 0.988   
## concentrationbeta 1.574e+01 1.189e+03 0.013 0.989   
## concentrationdelta 1.955e+01 1.189e+03 0.016 0.987   
## concentrationdzeta 1.919e+01 1.189e+03 0.016 0.987   
## concentrationepsilon 1.937e+01 1.189e+03 0.016 0.987   
## concentrationgamma 1.891e+01 1.189e+03 0.016 0.987   
## 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

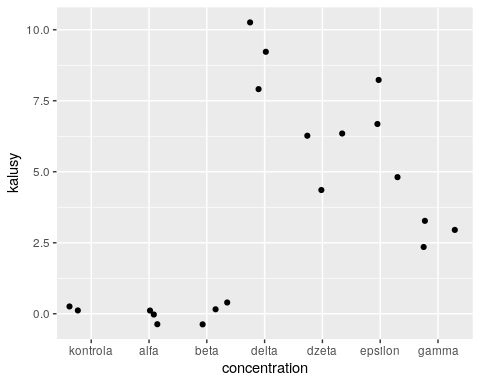
## Estimate Robust SE Pr(>|z|) LL  
## (Intercept) -21.1630214 0.462975572 0.000000e+00 -22.07045350  
## concentrationalfa 17.6848141 0.368574936 0.000000e+00 16.96240727  
## concentrationbeta 15.7389040 0.528859774 1.284711e-194 14.70233884  
## concentrationdelta 19.5480965 0.145062571 0.000000e+00 19.26377385  
## concentrationdzeta 19.1889554 0.562039540 1.778929e-255 18.08735795  
## concentrationepsilon 19.3692173 0.278923701 0.000000e+00 18.82252682  
## concentrationgamma 18.9053802 0.048978200 0.000000e+00 18.80938289  
## day 0.1041413 0.008324455 6.558437e-36 0.08782541  
## UL  
## (Intercept) -20.2555893  
## concentrationalfa 18.4072210  
## concentrationbeta 16.7754692  
## concentrationdelta 19.8324191  
## concentrationdzeta 20.2905529  
## concentrationepsilon 19.9159077  
## concentrationgamma 19.0013774  
## 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 1.812e+04 -0.001 0.999  
## concentrationalfa -4.542e-10 2.339e+04 0.000 1.000  
## concentrationbeta -4.542e-10 2.339e+04 0.000 1.000  
## concentrationdelta 2.350e+01 1.812e+04 0.001 0.999  
## concentrationdzeta 2.298e+01 1.812e+04 0.001 0.999  
## concentrationepsilon 2.320e+01 1.812e+04 0.001 0.999  
## concentrationgamma 2.228e+01 1.812e+04 0.001 0.999  
##   
## (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

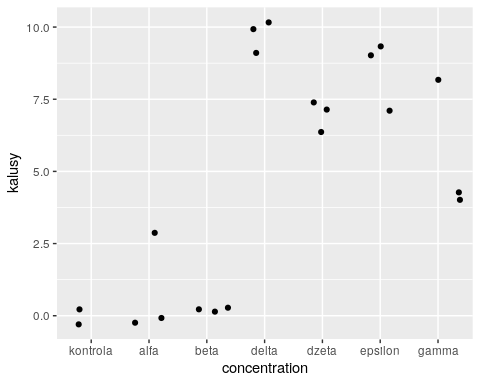
r.est\_m2d28 <- cbind(Estimate= coef(m2\_d28), "Robust SE" = std.err\_m2d28,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m2\_d28)/std.err\_m2d28), lower.tail=FALSE),  
LL = coef(m2\_d28) - 1.96 \* std.err\_m2d28,  
UL = coef(m2\_d28) + 1.96 \* std.err\_m2d28)  
  
r.est\_m2d28

## Estimate Robust SE Pr(>|z|) LL UL  
## (Intercept) -2.130259e+01 NaN NaN NaN NaN  
## concentrationalfa -4.541712e-10 9.42816069 1.000000000 -18.4791949 18.47919  
## concentrationbeta -4.541728e-10 NaN NaN NaN NaN  
## concentrationdelta 2.349981e+01 0.05237825 0.000000000 23.3971483 23.60247  
## concentrationdzeta 2.297656e+01 6.99506297 0.001020973 9.2662381 36.68688  
## concentrationepsilon 2.319971e+01 NaN NaN NaN NaN  
## concentrationgamma 2.228341e+01 11.27846127 0.048183154 0.1776303 44.38920

### 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) -1.930e+01 6.666e+03 -0.003 0.998  
## concentrationalfa 1.930e+01 6.666e+03 0.003 0.998  
## concentrationbeta 1.616e-07 8.605e+03 0.000 1.000  
## concentrationdelta 2.157e+01 6.666e+03 0.003 0.997  
## concentrationdzeta 2.120e+01 6.666e+03 0.003 0.997  
## concentrationepsilon 2.142e+01 6.666e+03 0.003 0.997  
## concentrationgamma 2.098e+01 6.666e+03 0.003 0.997  
##   
## (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

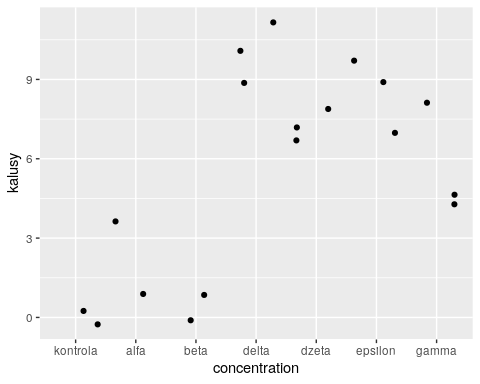
cov.m2\_d33 <- vcovHC(m2\_d33, type="HC0")  
std.err\_m2d33 <- sqrt(diag(cov.m2\_d33))  
r.est\_m2d33 <- cbind(Estimate= coef(m2\_d33), "Robust SE" = std.err\_m2d33,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m2\_d33)/std.err\_m2d33), lower.tail=FALSE),  
LL = coef(m2\_d33) - 1.96 \* std.err\_m2d33,  
UL = coef(m2\_d33)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

## Estimate Robust SE Pr(>|z|) LL UL  
## (Intercept) -1.930259e+01 1.908237 4.720675e-24 -23.042729 -15.562441  
## concentrationalfa 1.930259e+01 3.305859 5.254603e-09 12.823102 25.782069  
## concentrationbeta 1.615820e-07 2.192199 9.999999e-01 -4.296709 4.296709  
## concentrationdelta 2.157127e+01 1.819650 2.036368e-32 18.004754 25.137783  
## concentrationdzeta 2.119971e+01 1.819890 2.324630e-31 17.632720 24.766690  
## concentrationepsilon 2.142285e+01 2.573895 8.567415e-17 16.378015 26.467683  
## concentrationgamma 2.097656e+01 2.675609 4.508354e-15 15.732367 26.220756

### 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) -19.30 6665.62 -0.003 0.998  
## concentrationalfa 20.22 6665.62 0.003 0.998  
## concentrationbeta 18.61 6665.62 0.003 0.998  
## concentrationdelta 21.61 6665.62 0.003 0.997  
## concentrationdzeta 21.30 6665.62 0.003 0.997  
## concentrationepsilon 21.46 6665.62 0.003 0.997  
## concentrationgamma 21.04 6665.62 0.003 0.997  
##   
## (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))

## Warning in sqrt(diag(cov.m2\_d39)): wyprodukowano wartości NaN

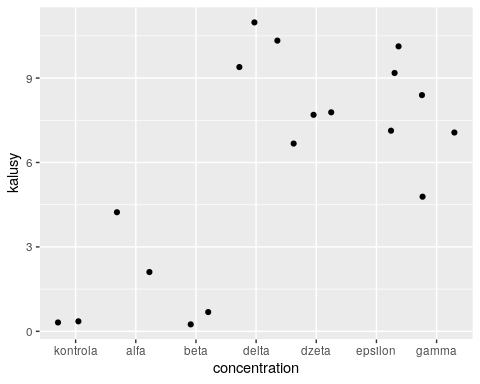
r.est\_m2d39 <- cbind(Estimate= coef(m2\_d39), "Robust SE" = std.err\_m2d39,  
"Pr(>|z|)" = 2 \* pnorm(abs(coef(m2\_d39)/std.err\_m2d39), lower.tail=FALSE),  
LL = coef(m2\_d39) - 1.96 \* std.err\_m2d39,  
UL = coef(m2\_d39) + 1.96 \* std.err\_m2d39)  
  
r.est\_m2d39

## Estimate Robust SE Pr(>|z|) LL UL  
## (Intercept) -19.30259 NaN NaN NaN NaN  
## concentrationalfa 20.21888 NaN NaN NaN NaN  
## concentrationbeta 18.60944 NaN NaN NaN NaN  
## concentrationdelta 21.60517 NaN NaN NaN NaN  
## concentrationdzeta 21.29502 NaN NaN NaN NaN  
## concentrationepsilon 21.46207 NaN NaN NaN NaN  
## concentrationgamma 21.03719 NaN NaN NaN NaN

### 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) -19.30 6665.62 -0.003 0.998  
## concentrationalfa 20.40 6665.62 0.003 0.998  
## concentrationbeta 18.61 6665.62 0.003 0.998  
## concentrationdelta 21.61 6665.62 0.003 0.997  
## concentrationdzeta 21.34 6665.62 0.003 0.997  
## concentrationepsilon 21.46 6665.62 0.003 0.997  
## concentrationgamma 21.20 6665.62 0.003 0.997  
##   
## (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))

## Warning in sqrt(diag(cov.m2\_d41)): wyprodukowano wartości NaN

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

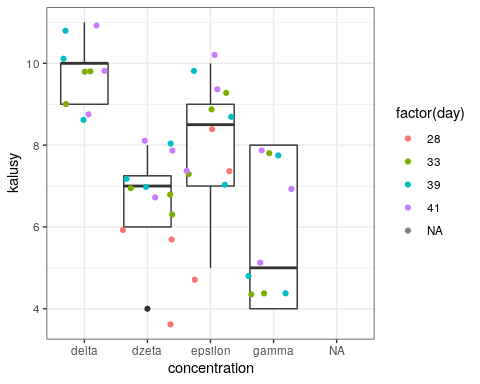
## Estimate Robust SE Pr(>|z|) LL UL  
## (Intercept) -19.30259 1.409326 1.067995e-42 -22.06486 -16.54031  
## concentrationalfa 20.40120 NaN NaN NaN NaN  
## concentrationbeta 18.60944 NaN NaN NaN NaN  
## concentrationdelta 21.60517 NaN NaN NaN NaN  
## concentrationdzeta 21.33947 NaN NaN NaN NaN  
## concentrationepsilon 21.46207 NaN NaN NaN NaN  
## concentrationgamma 21.19971 NaN NaN NaN NaN

## 3.Zmieniony zestaw danych

sub\_kielkowanie = subset(kielkowanie,  
 kielkowanie$concentration == "delta" |   
 kielkowanie$concentration == "epsilon" |  
 kielkowanie$concentration == "dzeta" |  
 kielkowanie$concentration == "gamma")  
   
sub\_kielkowanie\_day = subset(sub\_kielkowanie, kielkowanie$day >= 21)  
  
ggplot(data = sub\_kielkowanie\_day) +  
 geom\_boxplot(aes(x=concentration, y=kalusy)) +   
 geom\_jitter(aes(x=concentration, y=kalusy, color = factor(day))) +  
 theme\_bw()

## Warning: Removed 63 rows containing non-finite values (stat\_boxplot).

## Warning: Removed 63 rows containing missing values (geom\_point).



modelX = glm(kalusy ~ concentration, family = "poisson", data = sub\_kielkowanie\_day)  
summary(modelX)

##   
## Call:  
## glm(formula = kalusy ~ concentration, family = "poisson", data = sub\_kielkowanie\_day)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.16748 -0.35568 0.03527 0.33940 0.82448   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.2914 0.1060 21.617 < 2e-16 \*\*\*  
## concentrationdzeta -0.3819 0.1536 -2.487 0.01289 \*   
## concentrationepsilon -0.2016 0.1468 -1.374 0.16959   
## concentrationgamma -0.5183 0.1735 -2.987 0.00281 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 21.524 on 41 degrees of freedom  
## Residual deviance: 10.556 on 38 degrees of freedom  
## (63 observations deleted due to missingness)  
## AIC: 180.5  
##   
## Number of Fisher Scoring iterations: 4