* 1. **R Code from mixed effect fits**

**R Code & results from mixed effect fits using lme4 library.**

dataset<-read.csv("Tambjamines\_New\_numbers\_classified\_cleaned.csv")
ds\_alkyl <- subset(dataset, R.type =="alkyl")

ds\_OMe\_alkyl <- subset(dataset, R.type =="alkyl" & Ring.substituent == "OMe")
#subset

#plot the points for subset 1 and then label them by their Rgroup (NH.substituent) + colour by top substituent

#plot with labels and title, labels each point with Nh.substituent and colour by top substituent
plot(ds\_OMe\_alkyl$ALOGPs,ds\_OMe\_alkyl$Log.1.EC50., main="Plot of Log(1/EC50) against ALOGPs,
alkyl chain only, OMe ring substituent", ylim=c(0.4,3.0), xlab="ALOGPs", ylab="Log(1/EC50)", text(ds\_OMe\_alkyl$ALOGPs,ds\_OMe\_alkyl$Log.1.EC50., ds\_OMe\_alkyl$NH.substituent, cex=0.5, pos=3), col=(ifelse(ds\_OMe\_alkyl$Enamine.Substituent=='NH', "red", "blue")))

plot(ds\_OBn\_alkyl$ALOGPs,ds\_OBn\_alkyl$Log.1.EC50., main="Plot of Log(1/EC50) against ALOGPs,
alkyl chain only, OBn ring substituent", ylim=c(0.4,3.0), xlab="ALOGPs", ylab="Log(1/EC50)", text(ds\_OBn\_alkyl$ALOGPs,ds\_OBn\_alkyl$Log.1.EC50., ds\_OBn\_alkyl$NH.substituent, cex=0.5, pos=3), col=(ifelse(ds\_OBn\_alkyl$Enamine.Substituent=='NH', "red", "blue")))

#plot with the 2 different Enamine.Substituents in different colours

plot(ds\_alkyl$ALOGPs,ds\_alkyl$Log.1.EC50.,col=ds\_alkyl$Enamine.Substituent, pch=c(16, 17)[as.numeric(ds\_alkyl$Ring.substituent)], main="Plot of Log(1/EC50) vs ALOGPs - alkyl R chain", sub="Points coloured by Enamine Substituent type", ylim=c(0.4,3.0),xlab="ALOGPs",ylab="Log(1/EC50)")
# triangle – OMe
# circle – OBn
# black – NH
# green - NH-Ph

> # Try a parabolic fit, with the maximum and curator the same for both sets but different intercept
> # This uses fixed and random effects by group so use lme4 library
>
> library(lme4)
>
> # formula has the random or grouping effects of Enamine.Substituent
> fit\_l\_alkyl <- lmer(data = ds\_alkyl, Log.1.EC50. ~ (1|Enamine.Substituent) + ALOGPs + I(ALOGPs^2))
>
> summary(fit\_l\_alkyl)
Linear mixed model fit by REML ['lmerMod']
Formula: Log.1.EC50. ~ (1 | Enamine.Substituent) + ALOGPs + I(ALOGPs^2)
Data: ds\_alkyl

REML criterion at convergence: 16.6

Scaled residuals:

 Min 1Q Median 3Q Max

-2.3650 -0.3438 0.1081 0.5481 1.6244

Random effects:

 Groups Name Variance Std.Dev.

 Enamine.Substituent (Intercept) 0.03059 0.1749

 Residual 0.06238 0.2498

Number of obs: 28, groups: Enamine.Substituent, 2

Fixed effects:

 Estimate Std. Error t value

(Intercept) -0.73933 0.32341 -2.286

ALOGPs 1.37279 0.14548 9.436

I(ALOGPs^2) -0.15605 0.01661 -9.393

Correlation of Fixed Effects:

 (Intr) ALOGPs

ALOGPs -0.881

I(ALOGPs^2) 0.814 -0.979

>
> coef(fit\_l\_alkyl)
$Enamine.Substituent

 (Intercept) ALOGPs I(ALOGPs^2)

NH -0.6254821 1.372787 -0.1560464

NH-Ph -0.8531819 1.372787 -0.1560464

attr(,"class")

[1] "coef.mer"
>
> # calculate their 3 fitted curves and plot
> # define a fine x grid to calculate the curves
> x <- (0:80)/10
>
> # coef(fit\_l\_alkyl) gives the coefficients for each group (fixed and random combined).
> # (coef(fit\_l\_alkyl)$Enamine.Substituent)gives the table of pluses in effect a matrix from which we can
> # extract the numbers, first index is the group and can be a numeric index or use the name
> # (i.e. "NH") and second index is the coefficient (intercept, ALOGP etc)
>
> # coef(fit\_l\_alkyl) is a matrix of coefficients?
>
> # y1 takes the first line - NH
> y1 <- (coef(fit\_l\_alkyl)$Enamine.Substituent)[1,1]+(coef(fit\_l\_alkyl)$Enamine.Substituent)[1,2]\*x+(coef(fit\_l\_alkyl)$Enamine.Substituent)[1,3]\*x\*x
>
>
> #add line to the plot
> lines(x,y1,col="black")
>
> # y2 takes the second line - NH-Ph
> y2 <- (coef(fit\_l\_alkyl)$Enamine.Substituent)[2,1]+(coef(fit\_l\_alkyl)$Enamine.Substituent)[2,2]\*x+(coef(fit\_l\_alkyl)$Enamine.Substituent)[2,3]\*x\*x
>
> lines(x,y2,col="green")
>
>
> #
> #Just the OMe subset
> #
> #plot with the 2 different Enamine.Substituents in different colours
> plot(ds\_OMe\_alkyl$ALOGPs,ds\_OMe\_alkyl$Log.1.EC50.,col=ds\_OMe\_alkyl$Enamine.Substituent, pch=c(16, 17)[as.numeric(ds\_OMe\_alkyl$Ring.substituent)], main="Plot of Log(1/EC50) vs ALOGPs -
+ alkyl R chain, OMe ring substituent", sub="Points coloured by Enamine Substituent type", ylim=c(0.4,3.0),xlab="ALOGPs",ylab="Log(1/EC50)")
> # triangle - OMe
> # black - NH
> # green - NH-Ph
>
>
> # Try a parabolic fit, with the maximum and curator the same for both sets but different intercept
> # This uses fixed and random effects by group so use lme4 library
>
> library(lme4)
>
> # formula has the random(?) or grouping effects of Top.Substituent
> fit\_l\_OMe\_alkyl <- lmer(data = ds\_OMe\_alkyl, Log.1.EC50. ~ (1|Top.Substituent) + ALOGPs + I(ALOGPs^2))
>
> summary(fit\_l\_OMe\_alkyl)
Linear mixed model fit by REML ['lmerMod']
Formula: Log.1.EC50. ~ (1 | Enamine.Substituent) + ALOGPs + I(ALOGPs^2)
Data: ds\_OMe\_alkyl

REML criterion at convergence: -5.4

Scaled residuals:

 Min 1Q Median 3Q Max

-1.3440 -0.5360 -0.1963 0.2020 2.1088

Random effects:

 Groups Name Variance Std.Dev.

 Enamine.Substituent (Intercept) 0.06455 0.2541

 Residual 0.01738 0.1318

Number of obs: 20, groups: Enamine.Substituent, 2

Fixed effects:

 Estimate Std. Error t value

(Intercept) -1.13221 0.24732 -4.578

ALOGPs 1.63726 0.08750 18.712

I(ALOGPs^2) -0.19202 0.01069 -17.960

Correlation of Fixed Effects:

 (Intr) ALOGPs

ALOGPs -0.649

I(ALOGPs^2) 0.586 -0.971

> coef(fit\_l\_OMe\_alkyl)

$Enamine.Substituent

 (Intercept) ALOGPs I(ALOGPs^2)

NH -0.9556585 1.637261 -0.1920198

NH-Ph -1.3087691 1.637261 -0.1920198

attr(,"class")

[1] "coef.mer"

>
> # calculate their 3 fitted curves and plot
> # define a fine x grid to calculate the curves
> x <- (0:80)/10
>
> # coef(fit\_l\_OMe\_alkyl) gives the coefficients for each group (fixed and random combined).
> # (coef(fit\_l\_OMe\_alkyl)$Enamine.Substituent)gives the table of pluses in effect a matrix from which we can
> # extract the numbers, first index is the group and can be a numeric index or use the name
> # (i.e. "NH") and second index is the coefficient (intercept, ALOGP etc)
>
> # coef(fit\_l\_OMe\_alkyl) is a matrix of coefficients?
>
> # y1 takes the first line - NH
> y1 <- (coef(fit\_l\_OMe\_alkyl)$Enamine.Substituent)[1,1]+(coef(fit\_l\_OMe\_alkyl)$Enamine.Substituent)[1,2]\*x+(coef(fit\_l\_OMe\_alkyl)$Enamine.Substituent)[1,3]\*x\*x
>
>
> #add line to the plot
> lines(x,y1,col="black")
>
> # y2 takes the second line - NH-Ph
> y2 <- (coef(fit\_l\_OMe\_alkyl)$Enamine.Substituent)[2,1]+(coef(fit\_l\_OMe\_alkyl)$Enamine.Substituent)[2,2]\*x+(coef(fit\_l\_OMe\_alkyl)$Enamine.Substituent)[2,3]\*x\*x
>
> lines(x,y2,col="green")