**Statistical Analysis**

**Loading Data**

bacteria <- read.csv("~/univ/2014-2015/thesis/results february/design\_results\_febr\_counts.csv")

bacteria$LMG7866 <- ifelse(bacteria$LMG7866 > 30, 30, bacteria$LMG7866)

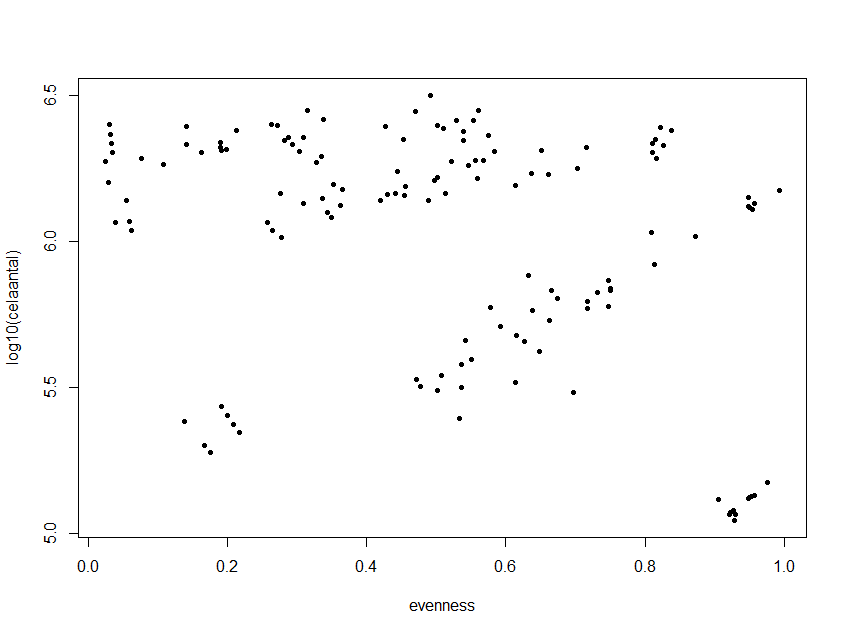
bacteria$LMG3203 <- ifelse(bacteria$LMG3203 > 30, 30, bacteria$LMG3203)

bacteria$LMG2954 <- ifelse(bacteria$LMG2954 > 30, 30, bacteria$LMG2954)

bacteria$LMG7878 <- ifelse(bacteria$LMG7878 > 30, 30, bacteria$LMG7878)

**Plot distribution**

plot(log10(celaantal)~evenness, bacteria, pch = 20)



**Make first model containing all parameters**

model1 <- glm(LMG7866~ evenness+log10(celaantal)+evenness:log10(celaantal), data = bacteria, family = poisson())

model2 <- glm(LMG2954~ evenness+log10(celaantal)+evenness:log10(celaantal), data = bacteria, family = poisson())

model3 <- glm(LMG3203~ evenness+log10(celaantal)+evenness:log10(celaantal), data = bacteria, family = poisson())

model4 <- glm(LMG7878~ evenness+log10(celaantal)+evenness:log10(celaantal), data = bacteria, family = poisson())

> summary(model1)

Call:

glm(formula = LMG7866 ~ evenness + log10(celaantal) + evenness:log10(celaantal),

family = poisson(), data = bacteria)

Deviance Residuals:

Min 1Q Median 3Q Max

-4.9506 -2.0801 -0.4649 1.2611 5.9059

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 4.6790 0.4702 9.950 < 2e-16 \*\*\*

evenness 0.1451 0.7418 0.196 0.845

log10(celaantal) -0.3751 0.0781 -4.804 1.56e-06 \*\*\*

evenness:log10(celaantal) -0.1670 0.1261 -1.324 0.185

> summary(model2)

Call:

glm(formula = LMG2954 ~ evenness + log10(celaantal) + evenness:log10(celaantal),

family = poisson(), data = bacteria)

Deviance Residuals:

Min 1Q Median 3Q Max

-5.8875 -1.2946 -0.0201 1.1664 3.3172

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 3.93425 0.36735 10.710 < 2e-16 \*\*\*

evenness -0.69644 0.54291 -1.283 0.19956

log10(celaantal) -0.18190 0.06073 -2.995 0.00274 \*\*

evenness:log10(celaantal) 0.09944 0.09131 1.089 0.27617

> summary(model3)

Call:

glm(formula = LMG3203 ~ evenness + log10(celaantal) + evenness:log10(celaantal),

family = poisson(), data = bacteria)

Deviance Residuals:

Min 1Q Median 3Q Max

-3.9653 -1.9889 -0.0366 1.3460 4.3942

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 3.53298 0.56219 6.284 3.29e-10 \*\*\*

evenness -0.48302 0.82445 -0.586 0.55796

log10(celaantal) -0.26375 0.09316 -2.831 0.00464 \*\*

evenness:log10(celaantal) 0.06682 0.13917 0.480 0.63114

> summary(model4)

Call:

glm(formula = LMG7878 ~ evenness + log10(celaantal) + evenness:log10(celaantal),

family = poisson(), data = bacteria)

Deviance Residuals:

Min 1Q Median 3Q Max

-4.5625 -1.8298 -0.4654 0.8858 6.6271

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 2.55928 0.49101 5.212 1.87e-07 \*\*\*

evenness 2.34979 0.73269 3.207 0.00134 \*\*

log10(celaantal) -0.03172 0.08104 -0.391 0.69548

evenness:log10(celaantal) -0.48977 0.12379 -3.957 7.60e-05 \*\*\*

**Evaluation first models.**

Model1: interaction factor not significant, reevaluate after removing interaction parameter

Model2: interaction factor not significant, reevaluate after removing interaction parameter

Model3: interaction factor not significant, reevaluate after removing interaction parameter

Model4: interaction factor significant. Since variable of log(cellcount) always between 5 and 6, when evenness increases, the number of colonies will decrease

**Update model after removing interaction parameters for model 1, 2 and 3**

model1 <- glm(LMG7866~ evenness+log10(celaantal), data = bacteria, family = poisson())

model2 <- glm(LMG2954~ evenness+log10(celaantal), data = bacteria, family = poisson())

model3 <- glm(LMG3203~ evenness+log10(celaantal), data = bacteria, family = poisson())

> summary(model1)

Call:

glm(formula = LMG7866 ~ evenness + log10(celaantal), family = poisson(),

data = bacteria)

Deviance Residuals:

Min 1Q Median 3Q Max

-5.0313 -2.1372 -0.4718 1.2432 5.9759

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 5.19089 0.26320 19.72 <2e-16 \*\*\*

evenness -0.83349 0.06515 -12.79 <2e-16 \*\*\*

log10(celaantal) -0.46123 0.04266 -10.81 <2e-16 \*\*\*

> summary(model2)

Call:

glm(formula = LMG2954 ~ evenness + log10(celaantal), family = poisson(),

data = bacteria)

Deviance Residuals:

Min 1Q Median 3Q Max

-5.8677 -1.3303 -0.0291 1.1514 3.3764

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 3.58839 0.18717 19.171 < 2e-16 \*\*\*

evenness -0.10719 0.04505 -2.379 0.0173 \*

log10(celaantal) -0.12420 0.03004 -4.135 3.55e-05 \*\*\*

> summary(model3)

Call:

glm(formula = LMG3203 ~ evenness + log10(celaantal), family = poisson(),

data = bacteria)

Deviance Residuals:

Min 1Q Median 3Q Max

-3.9972 -2.0002 -0.0158 1.3364 4.3869

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 3.29985 0.28588 11.543 < 2e-16 \*\*\*

evenness -0.08857 0.06981 -1.269 0.205

log10(celaantal) -0.22474 0.04592 -4.895 9.85e-07 \*\*\*

**Evaluation updated models.**

Model1: Eveness-parameter and log10(cellcount) parameter is negative and significant. This means that when the evenness or the cellcount increases, the number of colonies formed by pathogen decreases.

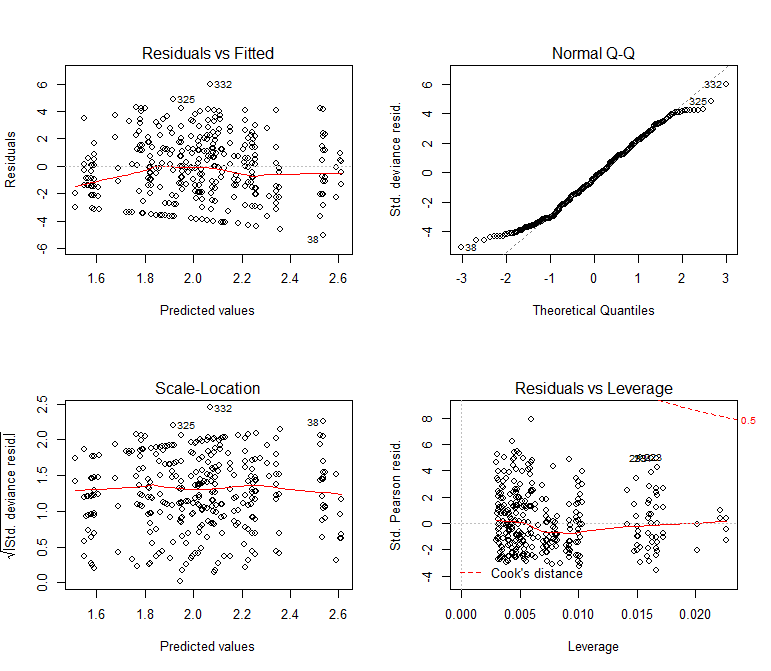
Model2: Eveness-parameter and log10(cellcount) parameter is negative and significant. This means that when the evenness or the cellcount increases, the number of colonies formed by pathogen decreases.

Model3: log10(cellcount) parameter is negative and significant. This means that when the cellcount increases, the number of colonies formed by pathogen decreases. However, the evenness parameter is nog significant.

**Check models on normality conditions of residuals and equality of variances**

par(mfrow = c(2,2))

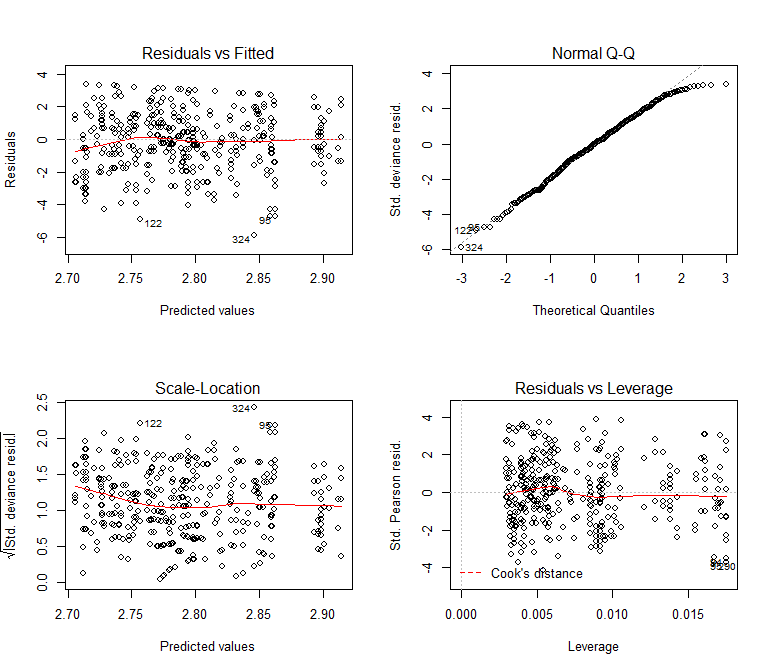
plot(model1)



In the left top plot the residuals are scattered without a real pattern so the assumptions that the relationship is linear is OK. Residuals form a band around the 0 line, suggesting that the variances are equal. However there are 3 outliers from the random pattern of residuals, but as the data is biological, noise is to be expected.

The top right plot shows the Q-Q plot. If the values are more or less on the diagonal line, the residuals can be assumed to be normal distributed.

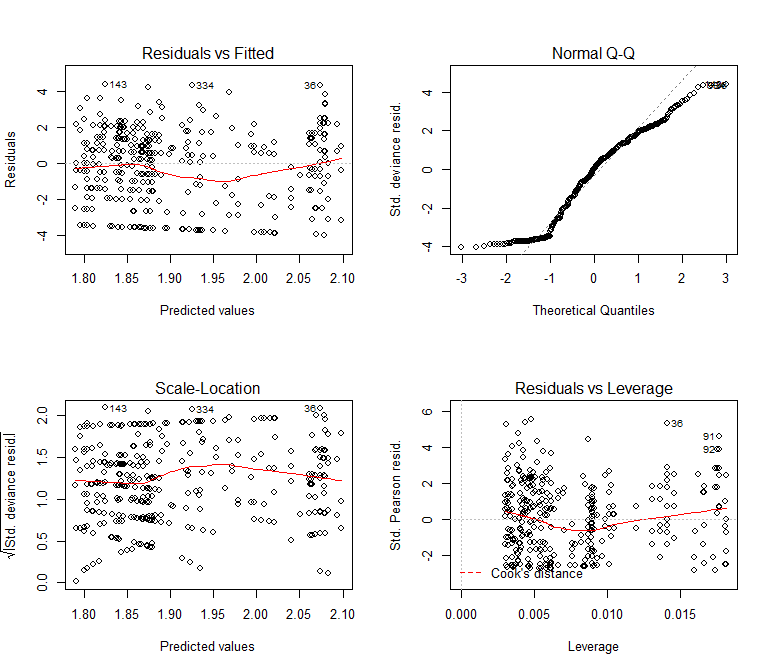
plot(model2)



In the left top plot the residuals are scattered without a real pattern so the assumptions that the relationship is linear is OK. Residuals form a band around the 0 line, suggesting that the variances are equal. However there are 3 outliers from the random pattern of residuals, but as the data is biological, noise is to be expected.

The top right plot shows the Q-Q plot. If the values are more or less on the diagonal line, the residuals can be assumed to be normal distributed.

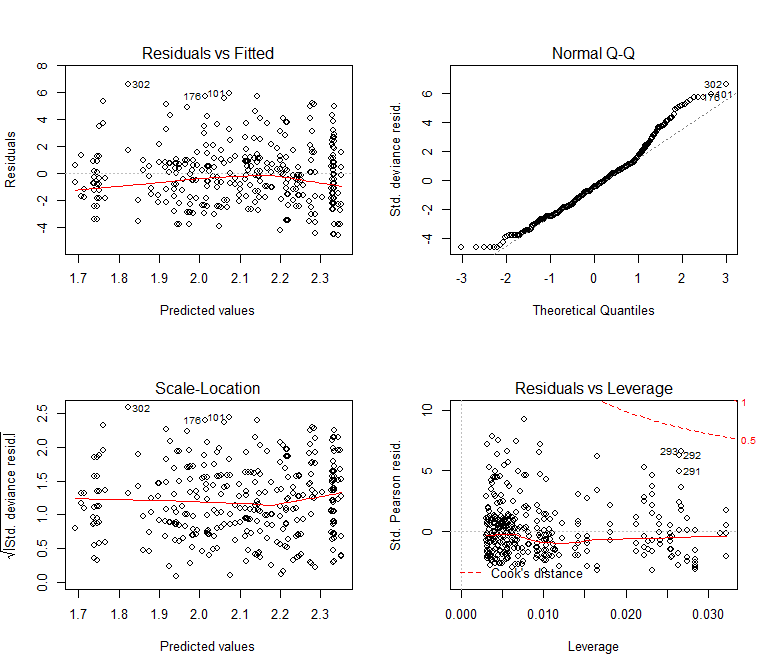
plot(model3)



In the left top plot the residuals are scattered without a real pattern so the assumptions that the relationship is linear is OK. Residuals form a band around the 0 line, suggesting that the variances are equal. However there are 3 outliers from the random pattern of residuals, but as the data is biological, noise is to be expected.

The top right plot shows the Q-Q plot. If the values are more or less on the diagonal line, the residuals can be assumed to be normal distributed.

plot(model4)



In the left top plot the residuals are scattered without a real pattern so the assumptions that the relationship is linear is OK. Residuals form a band around the 0 line, suggesting that the variances are equal. However there are 3 outliers from the random pattern of residuals, but as the data is biological, noise is to be expected.

The top right plot shows the Q-Q plot. If the values are more or less on the diagonal line, the residuals can be assumed to be normal distributed.

**Test whether the pathogen counts the same between replicas**

aov1 <- aov(LMG7866 ~ replica, data=bacteria)

aov2 <- aov(LMG3203 ~ replica, data=bacteria)

aov3 <- aov(LMG2954 ~ replica, data=bacteria)

aov4 <- aov(LMG7878 ~ replica, data=bacteria)

> summary(aov1)

Df Sum Sq Mean Sq F value Pr(>F)

replica 1 146 145.99 3.432 0.0647 .

Residuals 386 16421 42.54

> summary(aov2)

Df Sum Sq Mean Sq F value Pr(>F)

replica 1 8 8.082 0.306 0.58

Residuals 386 10181 26.375

> summary(aov3)

Df Sum Sq Mean Sq F value Pr(>F)

replica 1 18 18.19 0.373 0.542

Residuals 386 18803 48.71

> summary(aov4)

Df Sum Sq Mean Sq F value Pr(>F)

replica 1 156 155.97 3.457 0.0637 .

Residuals 386 17414 45.12

The Anova tests between the 2 replicas for each pathogen suggest there is no significant difference between the replicas.

**Explanation increase colony growth pathogen**

By looking at the correlation table, it can be seen whether an increase in pathogen counts is caused by one bacteria.

> signif(cor(bacteria)[16:19,], digits = 2)

id b1 b2 b3 b4 b5 b6 b7 b8 b9

LMG7866 -0.080 -0.200 -0.084 -0.0730 -0.0260 0.130 -0.130 -0.0720 0.0081 -0.013

LMG3203 0.055 -0.075 0.014 0.0008 -0.1500 0.046 0.022 0.0015 0.0250 0.051

LMG7878 0.100 -0.029 -0.039 -0.0960 0.0043 -0.062 -0.039 0.1300 -0.0320 0.052

LMG2954 0.081 -0.070 -0.098 0.0052 0.0190 -0.029 -0.046 -0.0015 -0.0890 0.041

b10 evenness celaantal gini\_thas gini\_wittebolle LMG7866 LMG3203 LMG7878

LMG7866 0.0054 -0.2200 -0.150 0.23000 0.23000 1.000 0.24 0.14

LMG3203 -0.0960 0.0036 -0.082 0.00085 0.00085 0.240 1.00 0.23

LMG7878 -0.0910 -0.1500 -0.083 0.15000 0.15000 0.140 0.23 1.00

LMG2954 -0.0410 -0.0370 -0.120 0.04400 0.04400 0.031 0.13 0.18

LMG2954 replica

LMG7866 0.031 0.094

LMG3203 0.130 -0.028

LMG7878 0.180 0.094

LMG2954 1.000 0.031

> signif(cor(bacteria)[2:11,16:19], digits = 2)

LMG7866 LMG3203 LMG7878 LMG2954

b1 -0.2000 -0.0750 -0.0290 -0.0700

b2 -0.0840 0.0140 -0.0390 -0.0980

b3 -0.0730 0.0008 -0.0960 0.0052

b4 -0.0260 -0.1500 0.0043 0.0190

b5 0.1300 0.0460 -0.0620 -0.0290

b6 -0.1300 0.0220 -0.0390 -0.0460

b7 -0.0720 0.0015 0.1300 -0.0015

b8 0.0081 0.0250 -0.0320 -0.0890

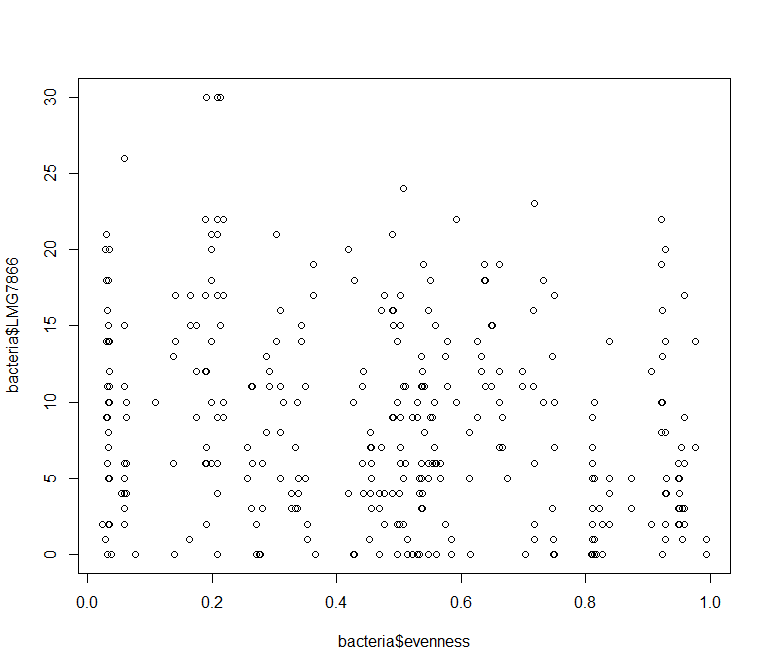
b9 -0.0130 0.0510 0.0520 0.0410

b10 0.0054 -0.0960 -0.0910 -0.0410

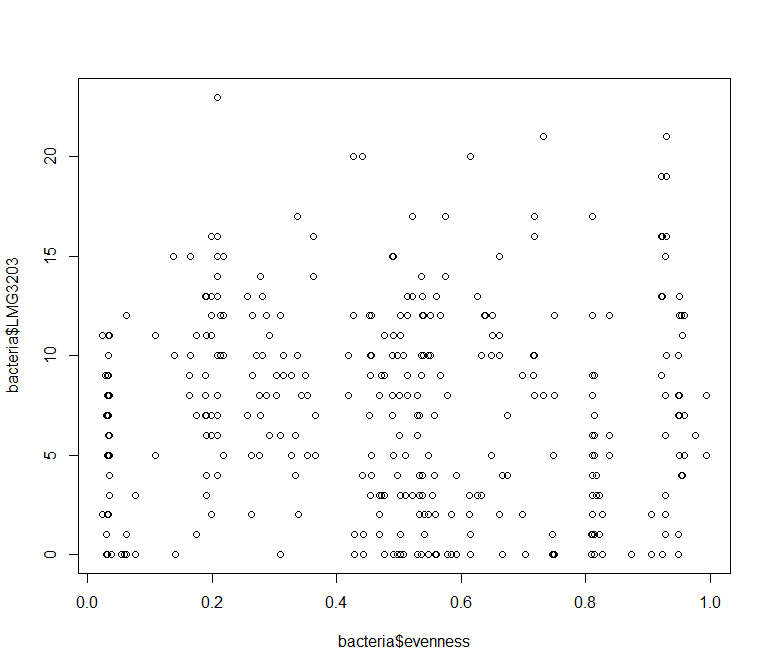
There does not seem to be one specific bacteria responsible for increase in pathogen growth.

**Visualising pathogen vs evenness**

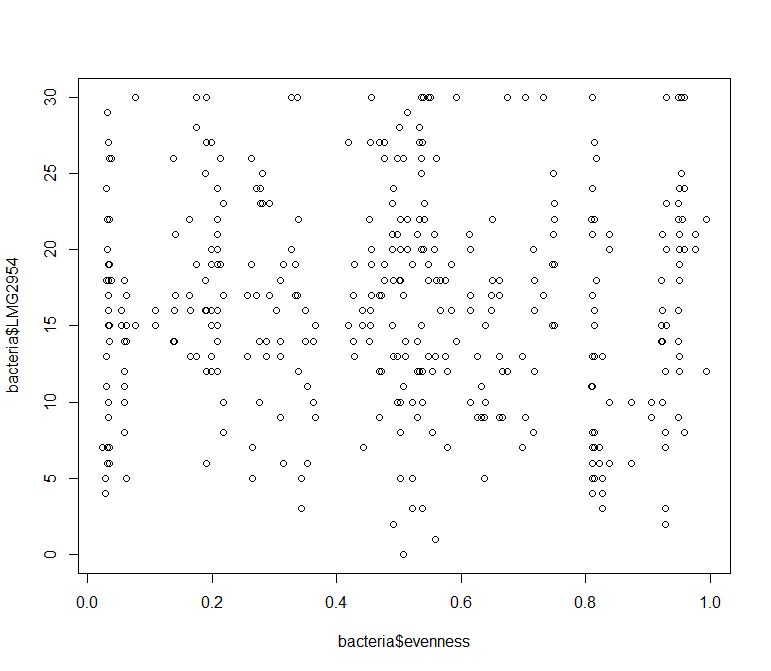
**LMG7866**



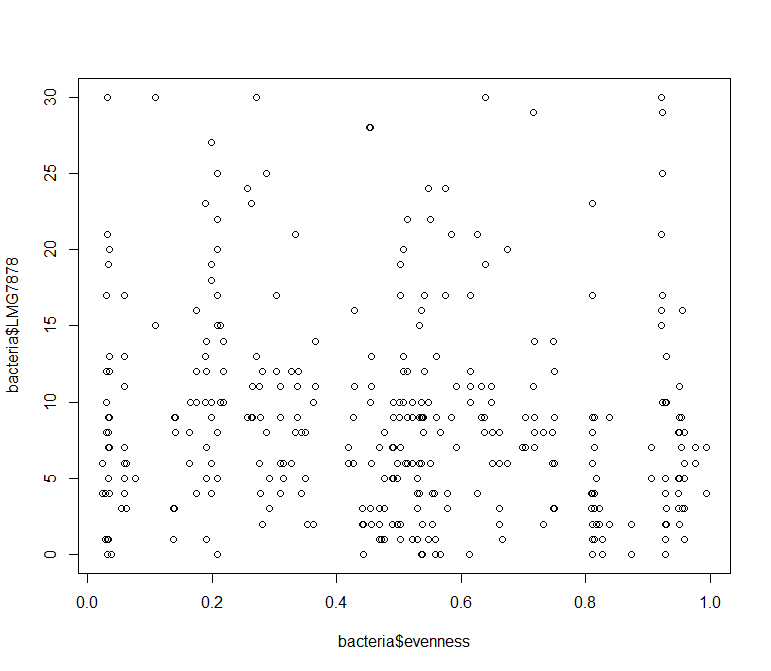
**LMG3203**



**LMG2954**



**LMG7878**



There is no immediate pattern visible in the data.

**Conclusion**

For 2 pathogen the statistical analysis suggests that the number of pathogen colonies grown decreases when the evenness (p-value = <2e-16 & p-value = 0.0173) and the total cellcount of the initial starting community increases (p-value = <2e-16 & p-value = 3.55e-05). For a third pathogen, the evenness variable does not a play a role (p-value = 0.205). While for a fourth pathogen there is an interaction factor present (p = 7.60e-05), indicating that as the evenness increases the colonies formed by pathogen decreases, but this is not linearly distinguishable from the parameter of the cellcount.

After checking for conditions of normality of residuals and equality of variances, these seem satisfied, as the noise is caused by the biological results.

The anova-test between replicas suggests there's no significant difference between replicas.