

# ddPCR\_data

Lars Martin Myhre

2022-07-28

```
library(readr)
library(plyr)
library(dplyr)
library(tidyverse)
library(lme4)
library(sjPlot)
library(glm)
library(multcomp)
library(emmeans)
library(knitr)

#Importing data from csv-file
rawdata <- read.csv2("C:/Users/larsm/OneDrive/Skrivebord/ddPCR_data/rådata.csv", header=FALSE
)
```

```
#Selecting relevant data rows
data <- rawdata %>% dplyr::select(V2, V3, V4, V5, V7, V14, V15, V16, V25)

#Removing redundant row(first row) and renaming column names to more comprehensible names
data <- data[-c(1),]
data <- data %>% dplyr::rename("sample" = V2,
                             "location" = V3,
                             "comment" = V4,
                             "contry" = V5,
                             "copies_per_µL" = V7,
                             "accepected droplets" = V14,
                             "positive droplets" = V15,
                             "negative droplets" = V16,
                             "copies_per_L" = V25)

#Changing desimal sign form comma to period, making it possible for R to read data as numeric
data
data$"copies_per_µL" <- gsub(",", ".", data$"copies_per_µL")
data$"copies_per_L" <- gsub(",", ".", data$"copies_per_L")

#Transforming copy data to numeric. Further, removing triplicates that failed ddPCR, and nega
tive + positive controls from ddPCR (all of which were marked with "lab" in location to make
it easier to remove in R).
field_data <- data %>% dplyr::mutate(`copies_per_µL` = as.numeric(`copies_per_µL`)) %>%
  dplyr::mutate(`copies_per_L` = as.numeric(`copies_per_L`)) %>%
  dplyr::mutate(sample = as.numeric(sample)) %>%
  drop_na(`copies_per_µL`) %>%
  dplyr::filter(!location == "lab")
```

```
#Separating data that are useful for plots that will be merged in to a data set for means later
```

```
additional_data <- field_data %>% dplyr::select(sample, location, contry, comment)
additional_data <- distinct(additional_data)
```

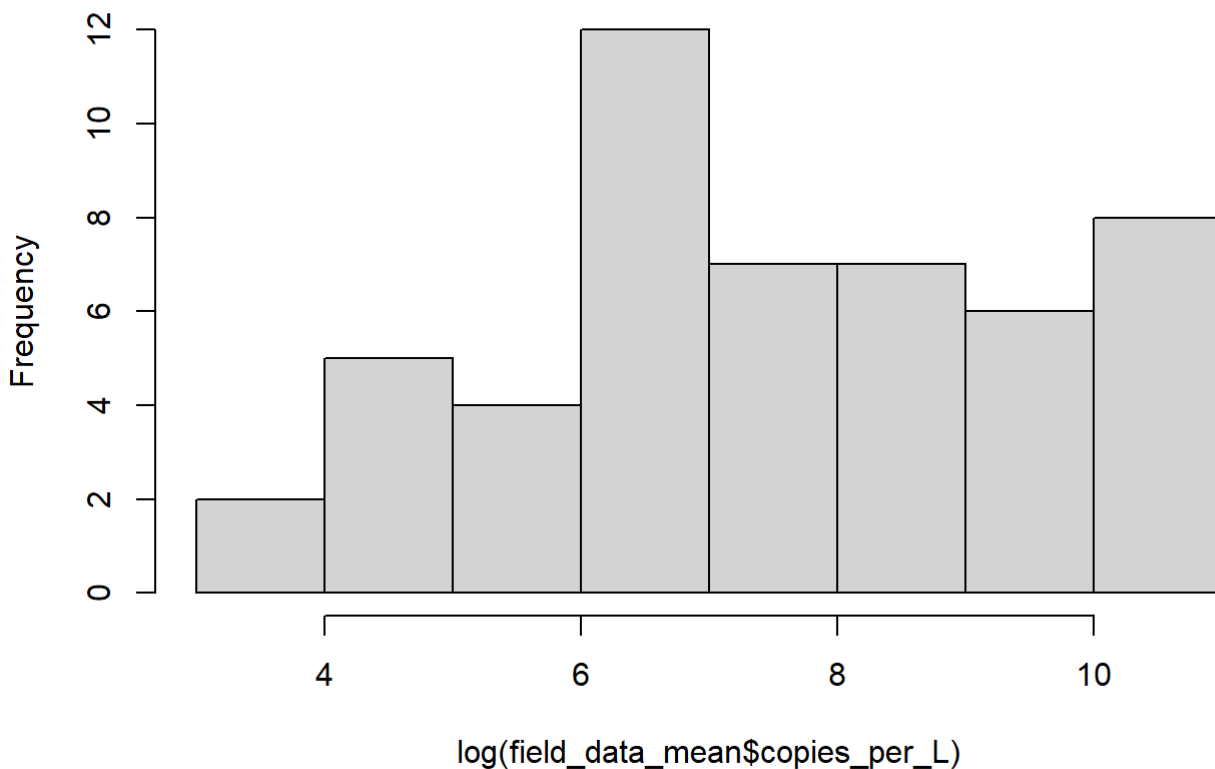
```
#Calculating means for all Lab-triplicates and merging with useful data
```

```
field_data_mean <- field_data %>%
  group_by(sample) %>%
  dplyr::summarise(across(`copies_per_L`, mean)) %>%
  merge(additional_data)
```

```
#Personal preference to investigate data
```

```
hist(log(field_data_mean$copies_per_L))
```

### Histogram of log(field\_data\_mean\$copies\_per\_L)



```
#Creating a box-plot for all stations to visualize the differences between stations. The stations were reordered in increasing signal because this was the easiest way to sort them by country.
```

```
plot_data <- field_data_mean %>% filter(!comment == "field control", !comment == "field control1") %>%
```

```
  mutate(sampleID = paste0(location, "_", comment))
```

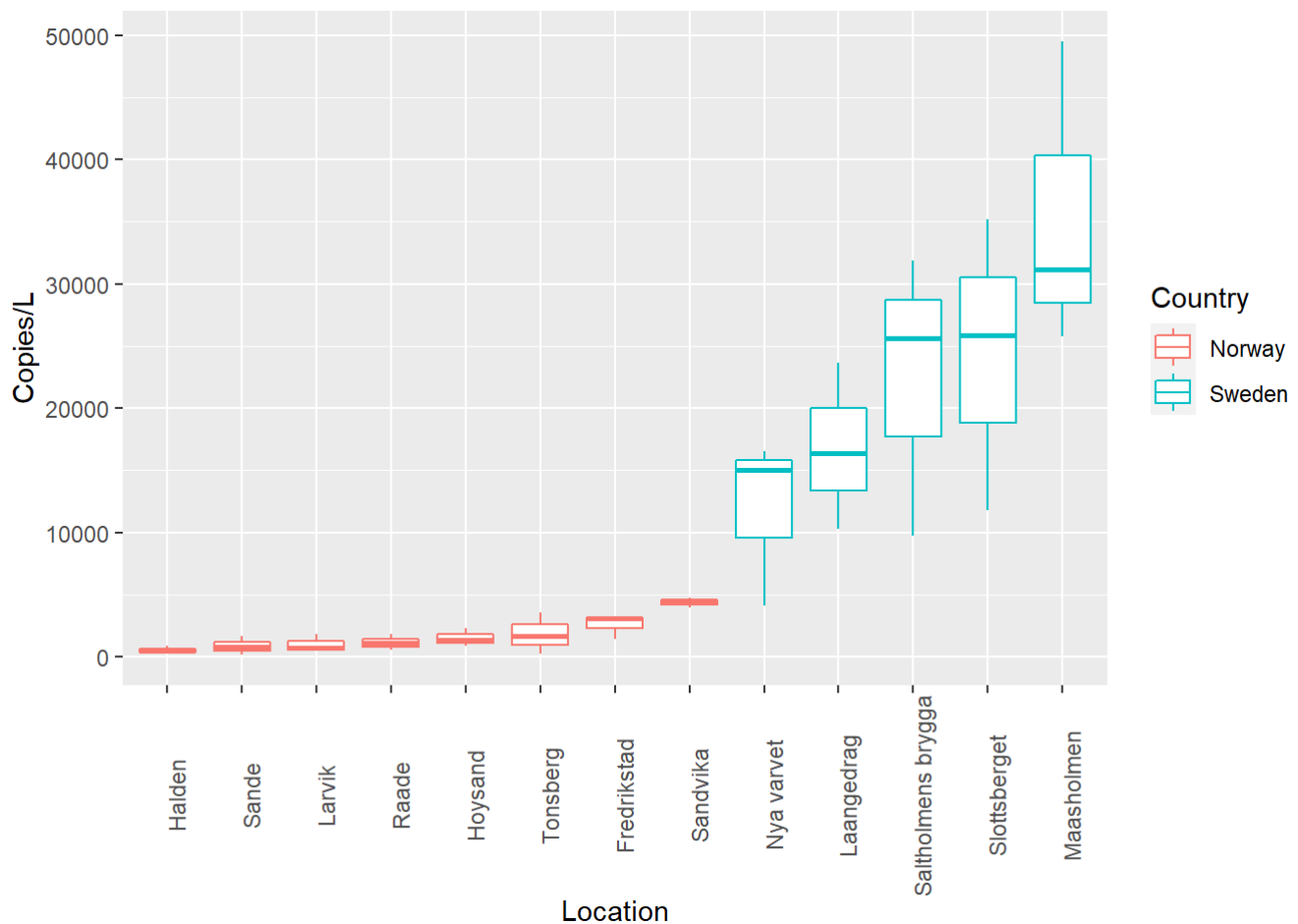
```
p1 <- ggplot(plot_data) +
```

```
  geom_boxplot(aes(x= reorder(location, copies_per_L), y= copies_per_L, color= contry)) +
```

```
  labs(y= "Copies/L", x= "Location", color= "Country") +
```

```
  theme(axis.text.x = element_text(angle = 90))
```

```
p1
```



```

#Separating controls from samples to
control <- field_data_mean %>% filter(!comment == "1") %>% filter(!comment == "2") %>% filter(
(!comment == "3")
control$"comment" <- gsub("field control","cont.", control$"comment")
control$"comment" <- gsub("cont.1","cont.", control$"comment")
control <- control %>% mutate(sampleID = paste0(location, "_", comment))

#Separating sample from controls
no_control <- field_data_mean %>% filter(!comment == "field control") %>% filter(!comment ==
"field control1")
no_control$"comment" <- gsub("1","samp", no_control$"comment")
no_control$"comment" <- gsub("2","samp", no_control$"comment")
no_control$"comment" <- gsub("3","samp", no_control$"comment")
no_control <- no_control %>% mutate(sampleID = paste0(location, "_", comment))

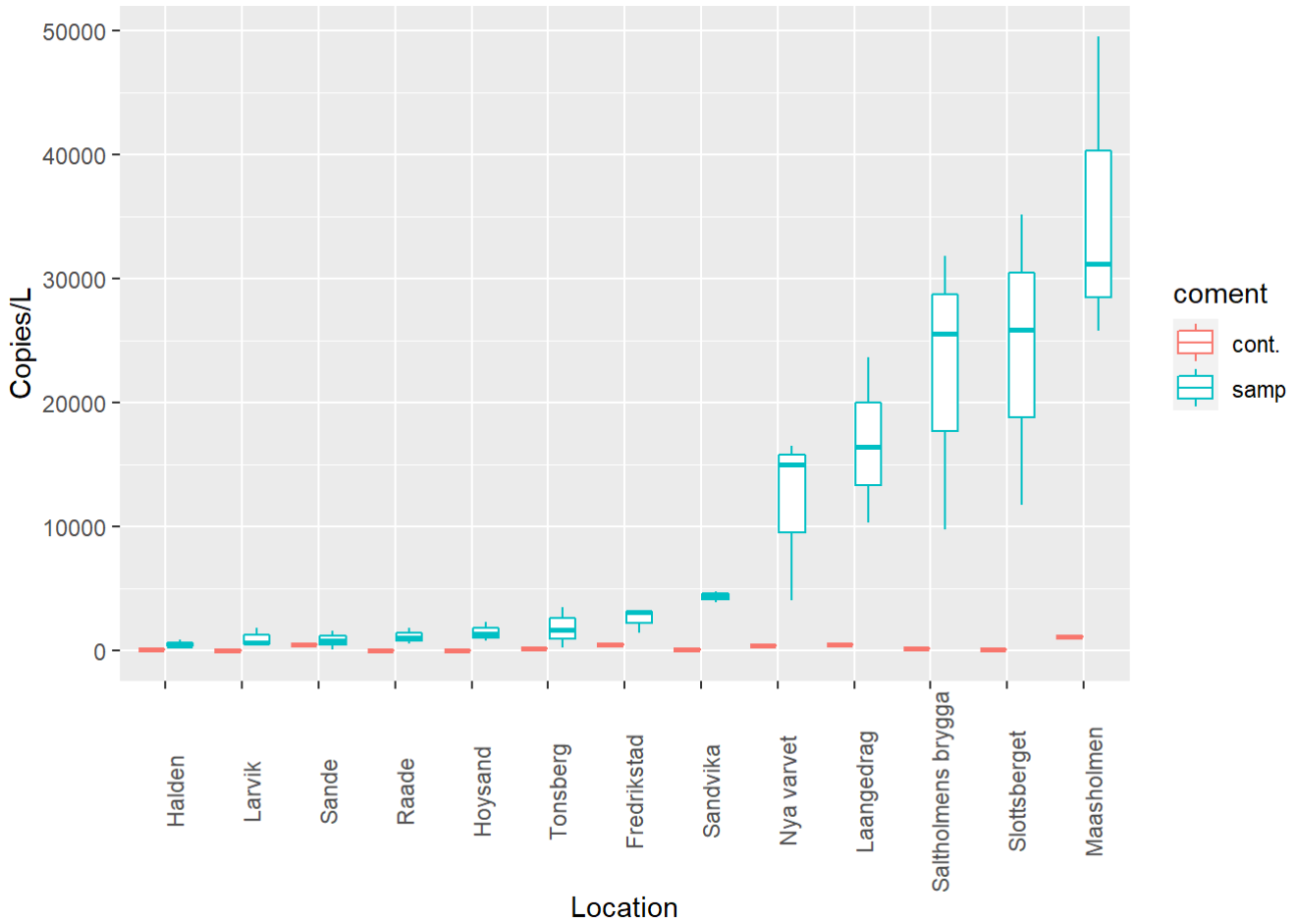
#Merging data
all <- bind_rows(control, no_control)

```

```

p2 <- ggplot(all) +
  geom_boxplot(aes(x= reorder(location, copies_per_L), y= copies_per_L, color= comment)) +
  labs(y= "Copies/L", x= "Location", color= "coment") +
  theme(axis.text.x = element_text(angle = 90))
p2

```



```
mod5 <- lm(copies_per_L ~ contry, data = plot_data)
mod4 <- lm(copies_per_L ~ location, data = plot_data)
mod2 <- lm(copies_per_L ~ contry*location, data = plot_data)
mod6 <- lm(copies_per_L ~ contry+location, data = plot_data)
```

```
AIC(mod5, mod4, mod2, mod6)
```

```
##      df      AIC
## mod5  3 809.4425
## mod4 14 805.5179
## mod2 14 805.5179
## mod6 14 805.5179
```

*# Although "mod5" showed a higher value, this does not test significance between Locations. "Mod5" will only be able to show significance between Norway and Sweden. The other models show the same amounts, and it is logical to use "mod2" since this tests for significance between all stations.*

```
kable(anova(mod2), digits = 10)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
contry	1	3855035259	3855035259	96.418723	0.0000000003
location	11	981293875	89208534	2.231205	0.0453044097
Residuals	26	1039537895	39982227	NA	NA

```
# The ANOVA show significance between both countries and Locations. "emmeans()" will list the  
significance between all the Locations.
```

```
emmeans(mod2, pairwise ~ location) #Post Hoc test
```

```
## NOTE: A nesting structure was detected in the fitted model:
```

```
## location %in% contry
```

```

## $emmeans
## location      contry emmean  SE df lower.CL upper.CL
## Fredrikstad   Norway  2574 3651 26   -4930  10078
## Halden         Norway   571 3651 26   -6933   8075
## Hoysand        Norway  1499 3651 26   -6005   9003
## Larvik         Norway   988 3651 26   -6516   8492
## Raade          Norway  1152 3651 26   -6352   8656
## Sande          Norway   855 3651 26   -6649   8359
## Sandvika       Norway  4368 3651 26   -3136  11873
## Tonsberg       Norway  1814 3651 26   -5690   9318
## Laangedrag     Sweden 16782 3651 26    9278  24286
## Maasholmen     Sweden 35487 3651 26   27983  42992
## Nya varvet     Sweden 11895 3651 26    4391  19399
## Saltholmens brygga Sweden 22399 3651 26   14894  29903
## Slottsberget   Sweden 24256 3651 26   16752  31760
##
## Confidence level used: 0.95
##
## $contrasts
## contrast                estimate  SE df t.ratio
## Fredrikstad Norway - Halden Norway          2003 5163 26  0.388
## Fredrikstad Norway - Hoysand Norway         1075 5163 26  0.208
## Fredrikstad Norway - Larvik Norway          1586 5163 26  0.307
## Fredrikstad Norway - Raade Norway           1422 5163 26  0.276
## Fredrikstad Norway - Sande Norway           1719 5163 26  0.333
## Fredrikstad Norway - Sandvika Norway       -1794 5163 26 -0.348
## Fredrikstad Norway - Tonsberg Norway         760 5163 26  0.147
## Fredrikstad Norway - Laangedrag Sweden     -14208 5163 26 -2.752
## Fredrikstad Norway - Maasholmen Sweden     -32913 5163 26 -6.375
## Fredrikstad Norway - Nya varvet Sweden      -9320 5163 26 -1.805
## Fredrikstad Norway - Saltholmens brygga Sweden -19824 5163 26 -3.840
## Fredrikstad Norway - Slottsberget Sweden   -21682 5163 26 -4.200
## Halden Norway - Hoysand Norway             -928 5163 26 -0.180
## Halden Norway - Larvik Norway              -417 5163 26 -0.081
## Halden Norway - Raade Norway               -581 5163 26 -0.112
## Halden Norway - Sande Norway               -284 5163 26 -0.055
## Halden Norway - Sandvika Norway            -3797 5163 26 -0.736
## Halden Norway - Tonsberg Norway            -1243 5163 26 -0.241
## Halden Norway - Laangedrag Sweden          -16211 5163 26 -3.140
## Halden Norway - Maasholmen Sweden          -34916 5163 26 -6.763
## Halden Norway - Nya varvet Sweden          -11323 5163 26 -2.193
## Halden Norway - Saltholmens brygga Sweden  -21827 5163 26 -4.228
## Halden Norway - Slottsberget Sweden       -23685 5163 26 -4.588
## Hoysand Norway - Larvik Norway              511 5163 26  0.099
## Hoysand Norway - Raade Norway              347 5163 26  0.067
## Hoysand Norway - Sande Norway              644 5163 26  0.125
## Hoysand Norway - Sandvika Norway           -2869 5163 26 -0.556
## Hoysand Norway - Tonsberg Norway           -315 5163 26 -0.061
## Hoysand Norway - Laangedrag Sweden         -15283 5163 26 -2.960
## Hoysand Norway - Maasholmen Sweden         -33988 5163 26 -6.583
## Hoysand Norway - Nya varvet Sweden        -10396 5163 26 -2.014
## Hoysand Norway - Saltholmens brygga Sweden -20900 5163 26 -4.048
## Hoysand Norway - Slottsberget Sweden      -22757 5163 26 -4.408
## Larvik Norway - Raade Norway               -164 5163 26 -0.032
## Larvik Norway - Sande Norway               133 5163 26  0.026

```

## Larvik Norway - Sandvika Norway	-3380	5163	26	-0.655
## Larvik Norway - Tonsberg Norway	-826	5163	26	-0.160
## Larvik Norway - Laangedrag Sweden	-15794	5163	26	-3.059
## Larvik Norway - Maasholmen Sweden	-34499	5163	26	-6.682
## Larvik Norway - Nya varvet Sweden	-10907	5163	26	-2.113
## Larvik Norway - Saltholmens brygga Sweden	-21410	5163	26	-4.147
## Larvik Norway - Slottsberget Sweden	-23268	5163	26	-4.507
## Raade Norway - Sande Norway	297	5163	26	0.058
## Raade Norway - Sandvika Norway	-3217	5163	26	-0.623
## Raade Norway - Tonsberg Norway	-663	5163	26	-0.128
## Raade Norway - Laangedrag Sweden	-15630	5163	26	-3.027
## Raade Norway - Maasholmen Sweden	-34336	5163	26	-6.651
## Raade Norway - Nya varvet Sweden	-10743	5163	26	-2.081
## Raade Norway - Saltholmens brygga Sweden	-21247	5163	26	-4.115
## Raade Norway - Slottsberget Sweden	-23104	5163	26	-4.475
## Sande Norway - Sandvika Norway	-3514	5163	26	-0.681
## Sande Norway - Tonsberg Norway	-960	5163	26	-0.186
## Sande Norway - Laangedrag Sweden	-15927	5163	26	-3.085
## Sande Norway - Maasholmen Sweden	-34633	5163	26	-6.708
## Sande Norway - Nya varvet Sweden	-11040	5163	26	-2.138
## Sande Norway - Saltholmens brygga Sweden	-21544	5163	26	-4.173
## Sande Norway - Slottsberget Sweden	-23401	5163	26	-4.533
## Sandvika Norway - Tonsberg Norway	2554	5163	26	0.495
## Sandvika Norway - Laangedrag Sweden	-12414	5163	26	-2.404
## Sandvika Norway - Maasholmen Sweden	-31119	5163	26	-6.028
## Sandvika Norway - Nya varvet Sweden	-7526	5163	26	-1.458
## Sandvika Norway - Saltholmens brygga Sweden	-18030	5163	26	-3.492
## Sandvika Norway - Slottsberget Sweden	-19887	5163	26	-3.852
## Tonsberg Norway - Laangedrag Sweden	-14968	5163	26	-2.899
## Tonsberg Norway - Maasholmen Sweden	-33673	5163	26	-6.522
## Tonsberg Norway - Nya varvet Sweden	-10080	5163	26	-1.952
## Tonsberg Norway - Saltholmens brygga Sweden	-20584	5163	26	-3.987
## Tonsberg Norway - Slottsberget Sweden	-22441	5163	26	-4.347
## Laangedrag Sweden - Maasholmen Sweden	-18705	5163	26	-3.623
## Laangedrag Sweden - Nya varvet Sweden	4887	5163	26	0.947
## Laangedrag Sweden - Saltholmens brygga Sweden	-5616	5163	26	-1.088
## Laangedrag Sweden - Slottsberget Sweden	-7474	5163	26	-1.448
## Maasholmen Sweden - Nya varvet Sweden	23593	5163	26	4.570
## Maasholmen Sweden - Saltholmens brygga Sweden	13089	5163	26	2.535
## Maasholmen Sweden - Slottsberget Sweden	11232	5163	26	2.175
## Nya varvet Sweden - Saltholmens brygga Sweden	-10504	5163	26	-2.035
## Nya varvet Sweden - Slottsberget Sweden	-12361	5163	26	-2.394
## Saltholmens brygga Sweden - Slottsberget Sweden	-1857	5163	26	-0.360
## p.value				
## 1.0000				
## 1.0000				
## 1.0000				
## 1.0000				
## 1.0000				
## 1.0000				
## 1.0000				
## 0.2818				
## 0.0001				
## 0.8302				
## 0.0315				
## 0.0135				

```
## 1.0000
## 1.0000
## 1.0000
## 1.0000
## 0.9999
## 1.0000
## 0.1401
## <.0001
## 0.6038
## 0.0127
## 0.0053
## 1.0000
## 1.0000
## 1.0000
## 1.0000
## 0.1965
## <.0001
## 0.7159
## 0.0194
## 0.0082
## 1.0000
## 1.0000
## 1.0000
## 1.0000
## 0.1636
## <.0001
## 0.6550
## 0.0153
## 0.0064
## 1.0000
## 1.0000
## 1.0000
## 0.1736
## <.0001
## 0.6748
## 0.0165
## 0.0069
## 1.0000
## 1.0000
## 0.1558
## <.0001
## 0.6387
## 0.0144
## 0.0060
## 1.0000
## 0.4708
## 0.0001
## 0.9533
## 0.0681
## 0.0306
## 0.2192
## <.0001
## 0.7518
## 0.0224
## 0.0095
```



```
## 0.0512
## 0.9987
## 0.9953
## 0.9555
## 0.0055
## 0.3936
## 0.6151
## 0.7033
## 0.4770
## 1.0000
##
## P value adjustment: tukey method for comparing a family of 13 estimates
```

```
tab_model(mod2, file = "mod2.doc")
```

```
## Warning: Model matrix is rank deficient. Parameters locationSlottsberget,
## contrySweden:locationHalden, contrySweden:locationHoysand,
## contrySweden:locationLarvik, contrySweden:locationLaangedrag,
## contrySweden:locationMaasholmen, contrySweden:locationNya varvet,
## contrySweden:locationRaade, contrySweden:locationSalholmens brygga,
## contrySweden:locationSande, contrySweden:locationSandvika,
## contrySweden:locationSlottsberget, contrySweden:locationTonsberg were
## not estimable.
```

<i>Predictors</i>	<b>copies_per_L</b>		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	2574.17	-4929.90 – 10078.23	0.487
contry [Sweden]	21681.64	11069.29 – 32293.99	<b>&lt;0.001</b>
location [Halden]	-2002.96	-12615.31 – 8609.39	0.701
location [Hoysand]	-1075.17	-11687.52 – 9537.18	0.837
location [Larvik]	-1586.04	-12198.39 – 9026.31	0.761
location [Laangedrag]	-7473.68	-18086.03 – 3138.67	0.160
location [Maasholmen]	11231.67	619.32 – 21844.02	<b>0.039</b>
location [Nya varvet]	-12361.16	-22973.51 – -1748.81	<b>0.024</b>
location [Raade]	-1422.41	-12034.76 – 9189.94	0.785
location [Salholmens brygga]	-1857.30	-12469.65 – 8755.05	0.722
location [Sande]	-1719.37	-12331.72 – 8892.98	0.742
location [Sandvika]	1794.31	-8818.04 – 12406.66	0.731
location [Tonsberg]	-759.75	-11372.10 – 9852.60	0.884
Observations	39		
R <sup>2</sup> / R <sup>2</sup> adjusted	0.823 / 0.741		

```

mean_data_norway <- field_data_mean %>% filter(contry == "Norway")
mean_data_norway$comment <- gsub("field control1", "field control", mean_data_norway$comment)
mean_data_norway$comment <- gsub("1", "samp", mean_data_norway$comment)
mean_data_norway$comment <- gsub("2", "samp", mean_data_norway$comment)
mean_data_norway$comment <- gsub("3", "samp", mean_data_norway$comment)

control_norway <- mean_data_norway %>% filter(comment == "field control") %>%
  mutate(sampleID = "field control")
sample_norway <- mean_data_norway %>% filter(!comment == "field control") %>%
  mutate(sampleID = paste0(location, "_", comment))
mean_data_norway <- bind_rows(control_norway, sample_norway)

mod_nor1 <- lm(copies_per_L ~ sampleID, data = mean_data_norway)
anova(mod_nor1)

```

```

## Analysis of Variance Table
##
## Response: copies_per_L
##           Df   Sum Sq Mean Sq F value    Pr(>F)
## sampleID   8 46566387 5820798  10.765 3.581e-06 ***
## Residuals 23 12436188  540704
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```
kable(anova(mod_nor1), digits = 6)
```

	<b>Df</b>	<b>Sum Sq</b>	<b>Mean Sq</b>	<b>F value</b>	<b>Pr(&gt;F)</b>
sampleID	8	46566387	5820798.4	10.76522	4e-06
Residuals	23	12436188	540703.8	NA	NA

```
summary(mod_nor1)
```

```
##
## Call:
## lm(formula = copies_per_L ~ sampleID, data = mean_data_norway)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1584.9  -214.2   -79.2   333.6  1708.5
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)         179.5      260.0   0.691  0.49677
## sampleIDFredrikstad_samp 2394.6      497.8   4.810 7.47e-05 ***
## sampleIDHalden_samp     391.7      497.8   0.787  0.43943
## sampleIDHoysand_samp   1319.5      497.8   2.651  0.01429 *
## sampleIDLarvik_samp     808.6      497.8   1.624  0.11794
## sampleIDRaade_samp     972.2      497.8   1.953  0.06309 .
## sampleIDSande_samp     675.3      497.8   1.356  0.18812
## sampleIDSandvika_samp  4189.0      497.8   8.415 1.78e-08 ***
## sampleIDTonsberg_samp  1634.9      497.8   3.284  0.00325 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 735.3 on 23 degrees of freedom
## Multiple R-squared:  0.7892, Adjusted R-squared:  0.7159
## F-statistic: 10.77 on 8 and 23 DF,  p-value: 3.581e-06
```

```
tab_model(mod_nor1, file = "norway_control.doc")
```

<i>Predictors</i>	<b>copies_per_L</b>		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>
(Intercept)	179.52	-358.28 – 717.33	0.497
sampleID [Fredrikstad_samp]	2394.64	1364.83 – 3424.46	<b>&lt;0.001</b>
sampleID [Halden_samp]	391.69	-638.13 – 1421.50	0.439
sampleID [Hoysand_samp]	1319.47	289.66 – 2349.29	<b>0.014</b>
sampleID [Larvik_samp]	808.60	-221.21 – 1838.42	0.118
sampleID [Raade_samp]	972.23	-57.58 – 2002.05	0.063
sampleID [Sande_samp]	675.27	-354.54 – 1705.09	0.188
sampleID [Sandvika_samp]	4188.96	3159.14 – 5218.77	<b>&lt;0.001</b>
sampleID [Tonsberg_samp]	1634.89	605.08 – 2664.71	<b>0.003</b>
Observations	32		
R <sup>2</sup> / R <sup>2</sup> adjusted	0.789 / 0.716		