

GSIN-1ug-flow-analysis 10/2

CTB

2024-10-21

```
data <- read.csv("/Users/csneffin/CSU Dropbox/Casey-Tyler Berezin/Cytek Data/GSIN-1ug-Data/20241002-GSIN-1ug-flow-analysis-10/2")
rename(., "Sample"="X") %>%
separate(., "Sample", sep = "-(?!mEGFP|Unsigned|Signed)", into=c("Date", "Exp", "Mixed", "Rack", "Growth"), remove=F) %>%
select(Group:GFP..Cells.Min.GFP.A) %>%
separate(., "Sample", sep = "\\s", extra = "merge", into=c("Well", "Sample")) %>%
mutate(Group = as.factor(Group),
       Sample = factor(as.factor(Sample),
                      levels = c("Unstained (Cells)", "Unstained",
                                   "Green GFP (Cells)", "CMV-mEGFP-Unsigned",
                                   "CMV-mEGFP-Signed", "pV18", "pV19", "pV20")),
       Rep = as.integer(Rep),
       GFP..Cells.Mean.GFP.A = as.integer(GFP..Cells.Mean.GFP.A),
       GFP..Cells.Min.GFP.A = as.integer(GFP..Cells.Min.GFP.A),
       GFP..Cells.Max.GFP.A = as.integer(GFP..Cells.Max.GFP.A)
       ) %>%
mutate(Sample = str_replace_all(Sample, "-", "_")) #for multcompLetters
```

```
## Warning: Expected 7 pieces. Missing pieces filled with 'NA' in 2 rows [23, 24].
```

```
## Warning: There were 3 warnings in 'mutate()'.
## The first warning was:
## i In argument: 'GFP..Cells.Mean.GFP.A = as.integer(GFP..Cells.Mean.GFP.A)'.
## Caused by warning:
## ! NAs introduced by coercion
## i Run 'dplyr::last_dplyr_warnings()' to see the 2 remaining warnings.
```

```
head(data)
```

##	Group	Well	Sample	Rep	All.Events.Count	All.Events...Parent
## 1	GSIN	A3	CMV_mEGFP_Signed	1	10000	100
## 2	GSIN	A4	pV18	1	10000	100
## 3	GSIN	A5	pV19	1	10000	100
## 4	GSIN	A6	pV20	1	10000	100
## 5	GSIN	B3	CMV_mEGFP_Signed	2	10000	100
## 6	GSIN	B4	pV18	2	10000	100
##	All.Events...Grand.Parent		All.Events...Total	All.Events.Count.uL		
## 1			100	100	558.97	
## 2			100	100	465.12	
## 3			100	100	606.43	
## 4			100	100	436.30	

## 5	100	100	646.83
## 6	100	100	713.78
##	All.Events.Mean.GFP.A	All.Events.Max.GFP.A	All.Events.Min.GFP.A Cells.Count
## 1	519580	5863031	-219 7397
## 2	543127	5840190	-270 6544
## 3	344771	5862167	-455 6601
## 4	128178	5907923	-434 6853
## 5	570001	5842935	-323 6263
## 6	292603	5839461	-435 6258
##	Cells...Parent	Cells...Grand.Parent	Cells...Total Cells.Count.uL
## 1	73.97	73.97	73.97 413.47
## 2	65.44	65.44	65.44 304.37
## 3	66.01	66.01	66.01 400.30
## 4	68.53	68.53	68.53 299.00
## 5	62.63	62.63	62.63 405.11
## 6	62.58	62.58	62.58 446.68
##	Cells.Mean.GFP.A	Cells.Max.GFP.A	Cells.Min.GFP.A GFP..Cells.Count
## 1	402790	4426931	-219 3357
## 2	378930	4607429	-233 2981
## 3	224530	4591333	-455 2255
## 4	76591	4403215	-434 1125
## 5	383838	4435806	-323 2783
## 6	189501	4459408	-435 1895
##	GFP..Cells...Parent	GFP..Cells...Grand.Parent	GFP..Cells...Total
## 1	45.38	33.57	33.57
## 2	45.55	29.81	29.81
## 3	34.16	22.55	22.55
## 4	16.42	11.25	11.25
## 5	44.44	27.83	27.83
## 6	30.28	18.95	18.95
##	GFP..Cells.Count.uL	GFP..Cells.Mean.GFP.A	GFP..Cells.Max.GFP.A
## 1	187.65	867575	4103897
## 2	138.65	800108	4155224
## 3	136.75	636217	4146679
## 4	49.08	435549	4150323
## 5	180.01	840776	4174397
## 6	135.26	615727	4007979
##	GFP..Cells.Min.GFP.A		
## 1	31541		
## 2	31260		
## 3	31266		
## 4	31719		
## 5	31344		
## 6	31246		

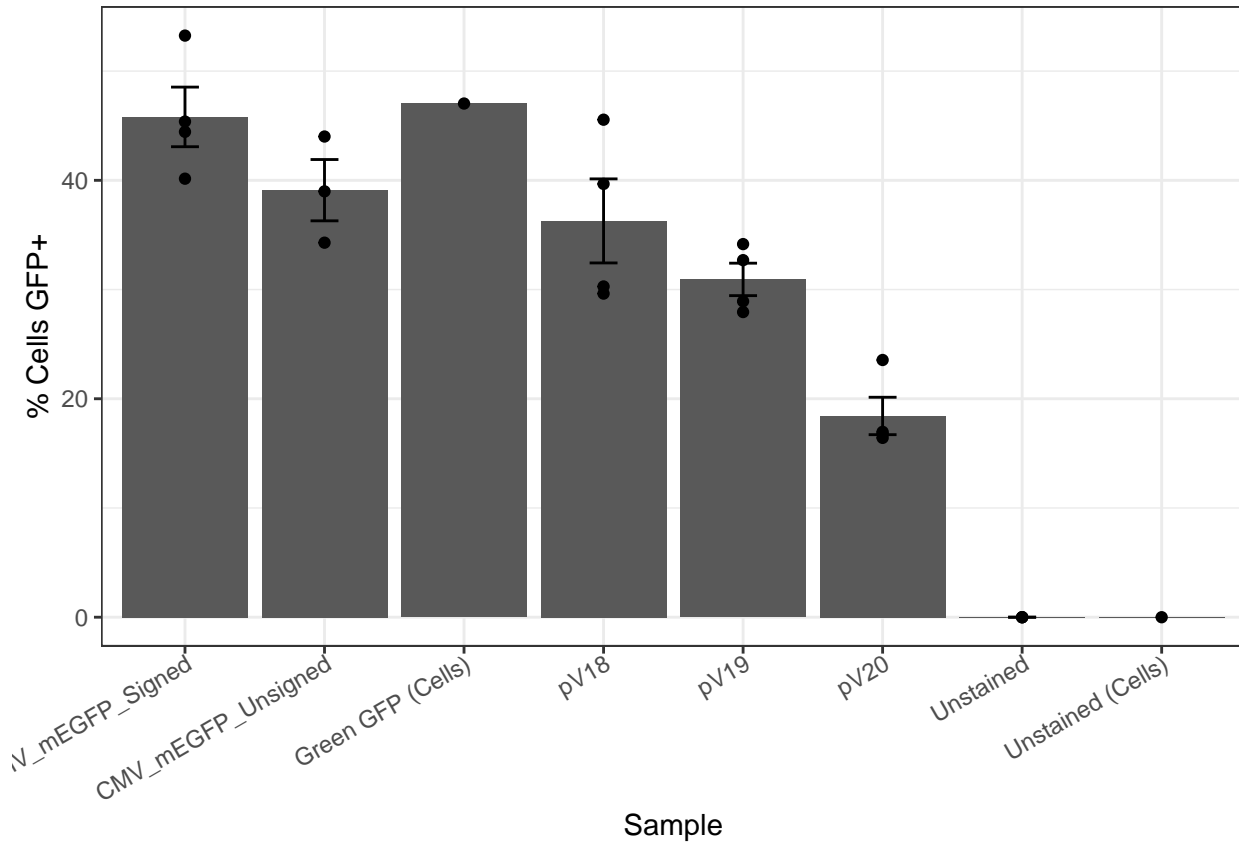
```
ggplot(data = data, mapping = aes(x = Sample, y = GFP..Cells...Parent)) +
  geom_bar(position = "dodge", stat = "summary", fun.y = "mean") +
  stat_summary(geom="errorbar", width=0.2) +
  geom_point() +
  scale_y_continuous(name="% Cells GFP+") +
  theme_bw() %+replace%
  theme(axis.text.x = element_text(angle = 30, vjust = 1, hjust=1))
```

```
## Warning in geom_bar(position = "dodge", stat = "summary", fun.y = "mean"):
```

```
## Ignoring unknown parameters: 'fun.y'
```

```
## No summary function supplied, defaulting to 'mean_se()'
```

```
## No summary function supplied, defaulting to 'mean_se()'
```



```
ggplot(data = data, mapping = aes(x = Sample, y = GFP..Cells.Mean.GFP.A)) +  
  geom_bar(position = "dodge", stat = "summary", fun.y = "mean") +  
  stat_summary(geom="errorbar", width=0.2) +  
  geom_point() +  
  scale_y_continuous(name="GFP Intensity") +  
  theme_bw() %+replace%  
  theme(axis.text.x = element_text(angle = 30, vjust = 1, hjust=1))
```

```
## Warning in geom_bar(position = "dodge", stat = "summary", fun.y = "mean"):
```

```
## Ignoring unknown parameters: 'fun.y'
```

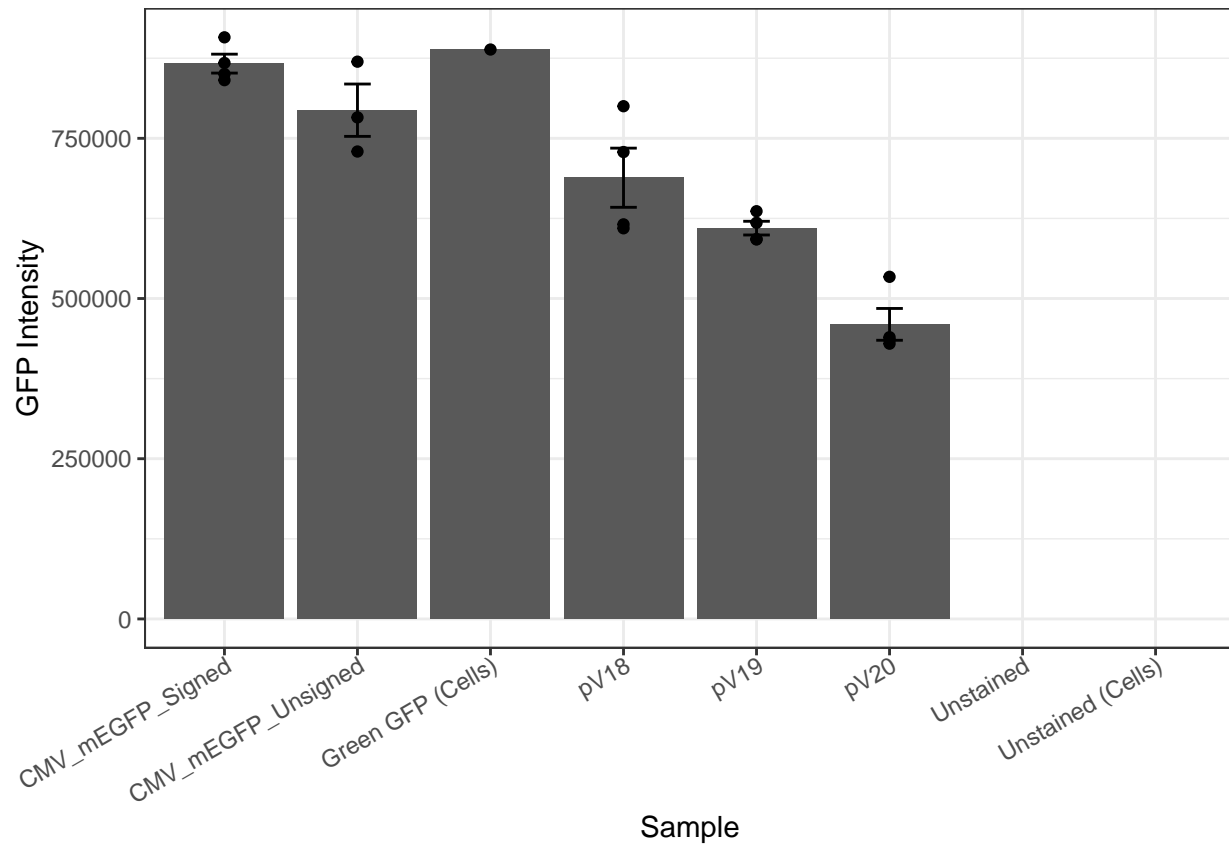
```
## Warning: Removed 4 rows containing non-finite values ('stat_summary()').
```

```
## No summary function supplied, defaulting to 'mean_se()'
```

```
## Warning: Removed 4 rows containing non-finite values ('stat_summary()').
```

```
## No summary function supplied, defaulting to 'mean_se()'
```

```
## Warning: Removed 4 rows containing missing values ('geom_point()').
```



```
data %>% group_by(Sample) %>% summarise(
  mean_int = mean(GFP..Cells.Mean.GFP.A),
  sd_mean_int = sd(GFP..Cells.Mean.GFP.A),
  min_int = min(GFP..Cells.Min.GFP.A),
  max_int = max(GFP..Cells.Max.GFP.A),
  mean_perc = mean(GFP..Cells...Parent),
  sd_perc = sd(GFP..Cells...Parent),
  min_perc = min(GFP..Cells...Parent),
  max_perc = max(GFP..Cells...Parent)
)
```

```
## # A tibble: 8 x 9
##   Sample      mean_int sd_mean_int min_int max_int mean_perc sd_perc min_perc
##   <chr>          <dbl>      <dbl>   <int>   <int>    <dbl>   <dbl>   <dbl>
## 1 CMV_mEGFP_Sig~ 866565      29510.  31229 4191918    45.8    5.46    40.2
## 2 CMV_mEGFP_Uns~ 793865      70756.  31264 4180302    39.1    4.86    34.3
## 3 Green GFP (Ce~ 888548         NA    31251 4161831    47.0    NA     47.0
## 4 Unstained      NA         NA         NA     NA         0      0         0
## 5 Unstained (Ce~ NA         NA         NA     NA         0      NA         0
## 6 pV18          688510.    92344.   31227 4178046    36.3    7.70    29.6
## 7 pV19          609832.    21363.   31266 4193757    30.9    2.97    27.9
## 8 pV20          459664.    49638.   31232 4150323    18.4    3.42    16.4
## # i 1 more variable: max_perc <dbl>
```

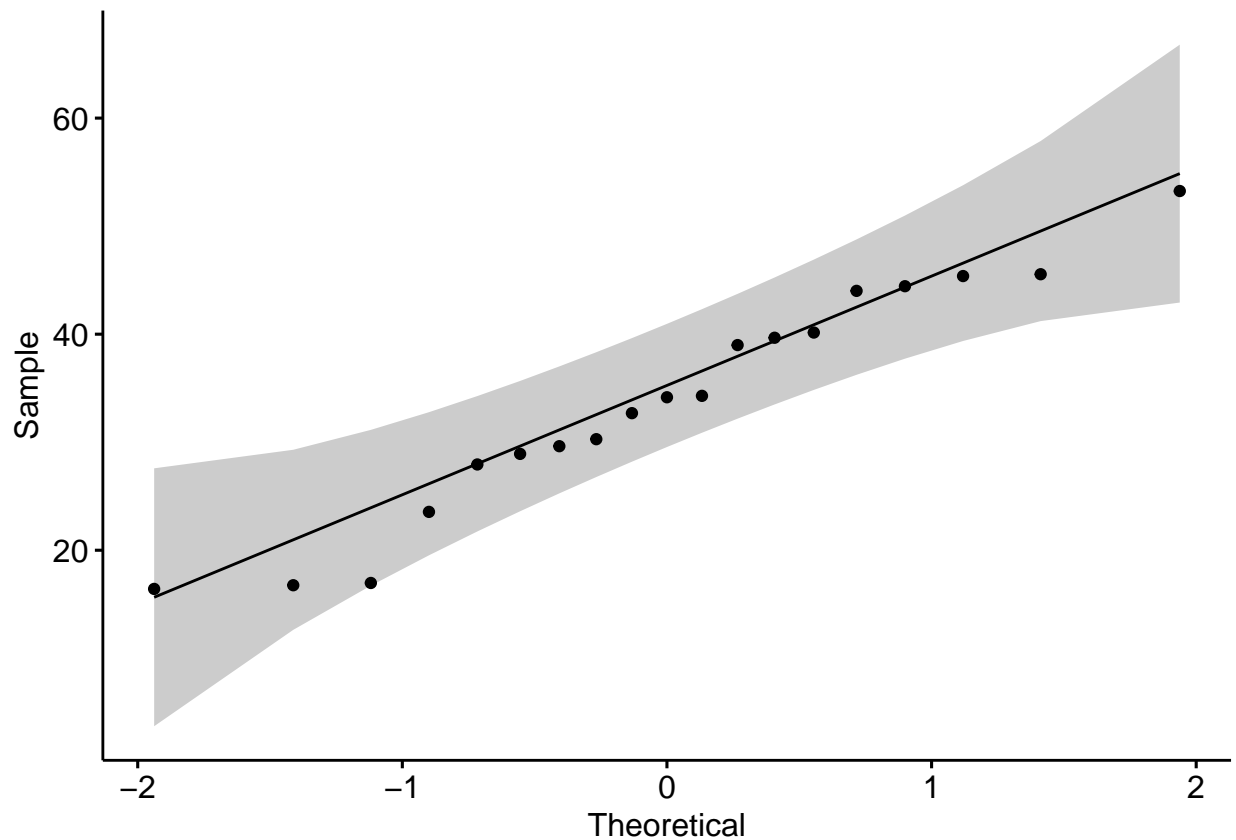
```
sample_data_only <- data %>% filter(Group != "Reference Group")

green_data_only <- data %>% filter(Sample != "Unstained")

green_samples_only <- sample_data_only %>% filter(Sample != "Unstained")
```

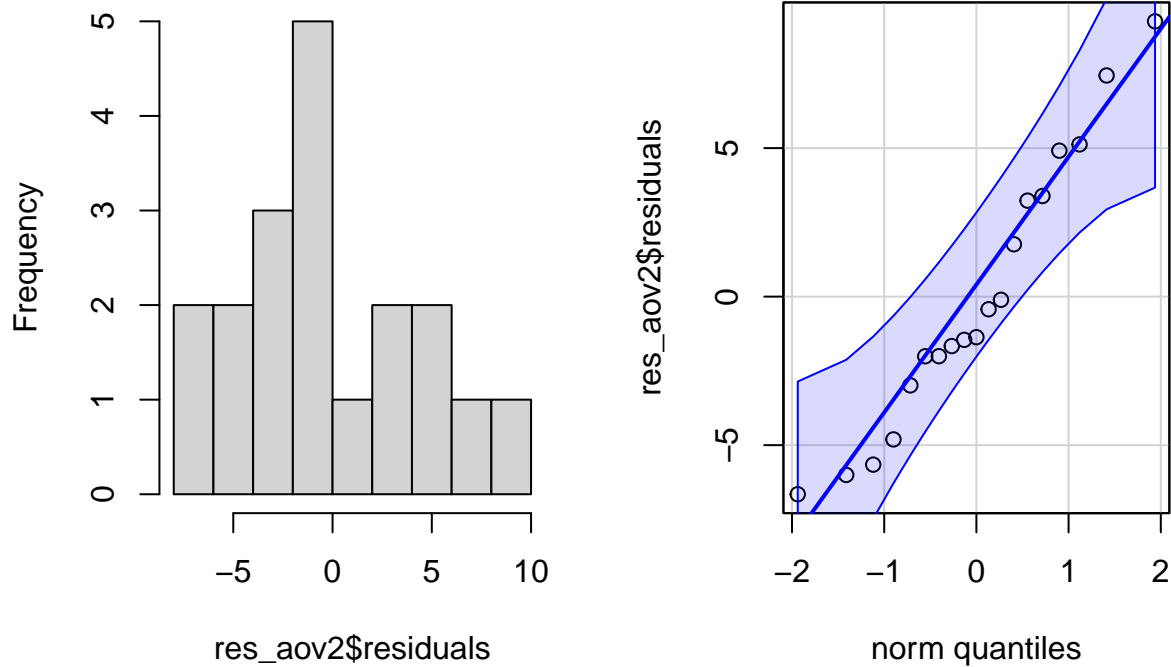
Using only green samples for stats - Dunnett's test compares to control (unstained) Samples with values 0 (unstained) make ANOVA calculations impossible

```
#Checking ANOVA assumptions for % GFP cells
ggqqplot(green_samples_only$GFP..Cells...Parent)
```



```
res_aov2 <- aov(GFP..Cells...Parent ~ Sample,
  data = green_samples_only
)
par(mfrow = c(1, 2)) # combine plots
# histogram
hist(res_aov2$residuals)
# QQ-plot
qqPlot(res_aov2$residuals,
  id = FALSE # id = FALSE to remove point identification
)
```

Histogram of res_aov2\$residual

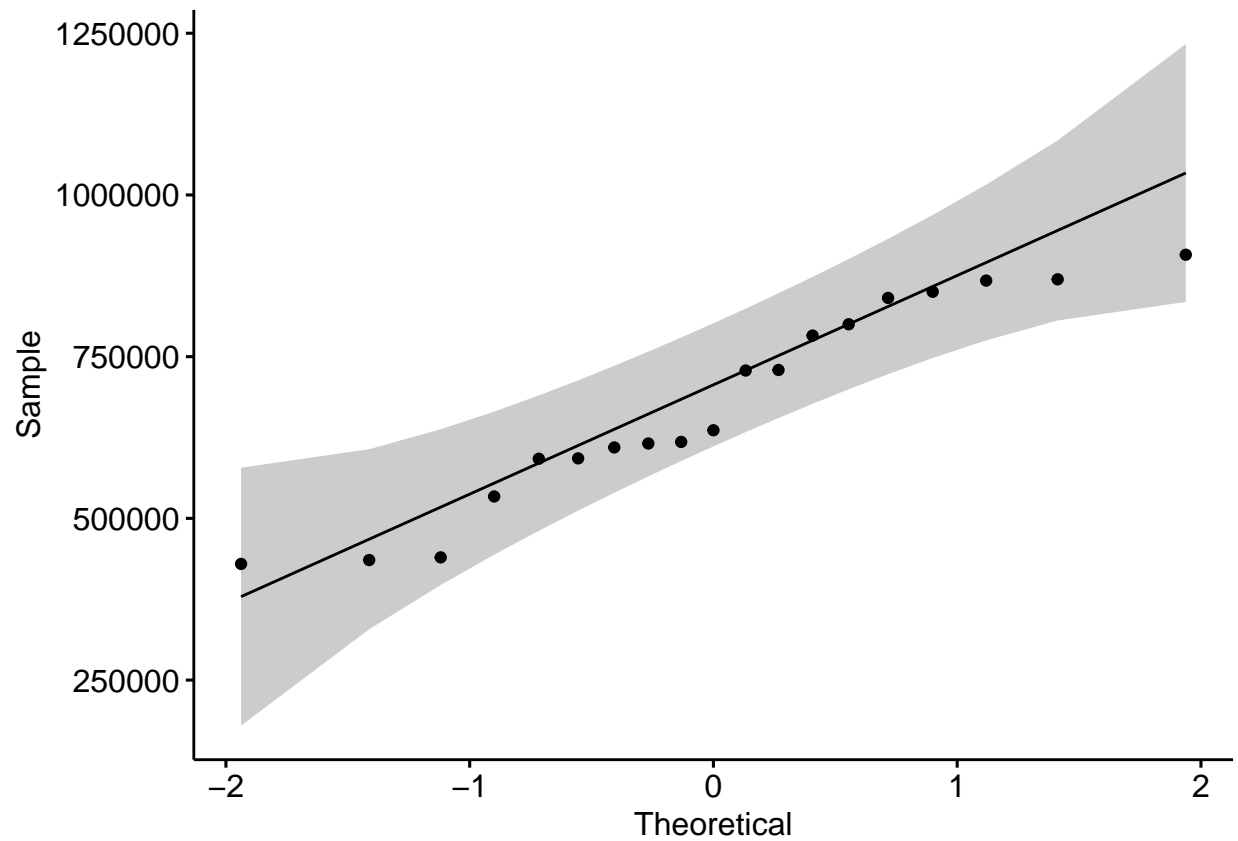


```
shapiro.test(green_samples_only$GFP..Cells...Parent)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  green_samples_only$GFP..Cells...Parent  
## W = 0.95623, p-value = 0.5006
```

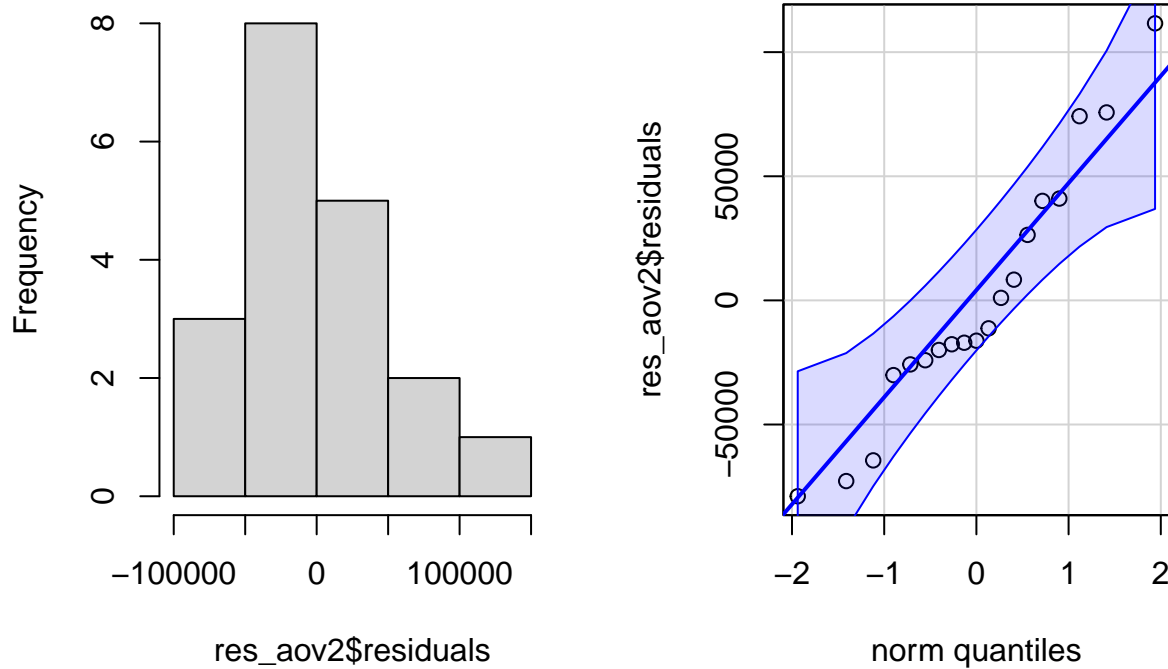
```
#Normally distributed!
```

```
#Checking ANOVA assumptions for Intensity  
ggqqplot(green_samples_only$GFP..Cells.Mean.GFP.A)
```



```
res_aov2 <- aov(GFP..Cells.Mean.GFP.A ~ Sample,
  data = green_samples_only
)
par(mfrow = c(1, 2)) # combine plots
# histogram
hist(res_aov2$residuals)
# QQ-plot
qqPlot(res_aov2$residuals,
  id = FALSE # id = FALSE to remove point identification
)
```

Histogram of res_aov2\$residual



```
shapiro.test(green_samples_only$GFP..Cells.Mean.GFP.A)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  green_samples_only$GFP..Cells.Mean.GFP.A
## W = 0.92787, p-value = 0.1582
```

```
#Normally distributed!
```

```
#Checking if % GFP cells is stat diff across groups
```

```
#Comparing all sample groups only to unsigned control group
```

```
DunnettTest(green_samples_only$GFP..Cells...Parent, green_samples_only$Sample, control="CMV_mEGFP_Unsigned")
```

```
##
##  Dunnett's test for comparing several treatments with a control :
##    95% family-wise confidence level
##
## $CMV_mEGFP_Unsigned
##               diff      lwr.ci    upr.ci    pval
## CMV_mEGFP_Signed-CMV_mEGFP_Unsigned  6.708333 -4.076803 17.493469 0.29010
## pV18-CMV_mEGFP_Unsigned             -2.814167 -13.599303  7.970969 0.87045
## pV19-CMV_mEGFP_Unsigned             -8.169167 -18.954303  2.615969 0.16098
```



```
## pV20-CMV_mEGFP_Unsigned          -20.671667 -31.456803 -9.886531 0.00056 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Checking if mean intensity is stat diff across groups
```

```
#Comparing all sample groups only to unsigned control group
```

```
DunnettTest(green_samples_only$GFP..Cells.Mean.GFP.A, green_samples_only$Sample, control="CMV_mEGFP_Uns
```

```
##
## Dunnett's test for comparing several treatments with a control :
## 95% family-wise confidence level
##
## $CMV_mEGFP_Unsigned
##              diff      lwr.ci      upr.ci    pval
## CMV_mEGFP_Signed-CMV_mEGFP_Unsigned  72700.0 -47834.54 193234.54 0.3125
## pV18-CMV_mEGFP_Unsigned             -105355.5 -225890.04  15179.04 0.0931 .
## pV19-CMV_mEGFP_Unsigned             -184033.5 -304568.04 -63498.96 0.0031 **
## pV20-CMV_mEGFP_Unsigned             -334201.2 -454735.79 -213666.71 2e-06 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Performing ANOVA for all pairwise comparisons
```

```
int_aov <- aov(GFP..Cells.Mean.GFP.A ~ Sample, data = green_samples_only)
summary(int_aov)
```

```
##              Df      Sum Sq  Mean Sq F value    Pr(>F)
## Sample         4 3.922e+11  9.805e+10   29.23 1.16e-06 ***
## Residuals     14 4.697e+10  3.355e+09
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tukey <- TukeyHSD(int_aov)
```

```
# Extract p-values from Tukey's HSD
```

```
tukey_pvalues <- tukey$Sample[, "p adj"]
```

```
# Generate groupings (a, b, c, etc.) based on Tukey HSD
```

```
group_letters <- multcompLetters(tukey_pvalues)
```

```
print(group_letters$Letters)
```

```
## CMV_mEGFP_Unsigned          pV18          pV19          pV20
##              "ab"          "ac"          "c"          "d"
## CMV_mEGFP_Signed
##              "b"
```