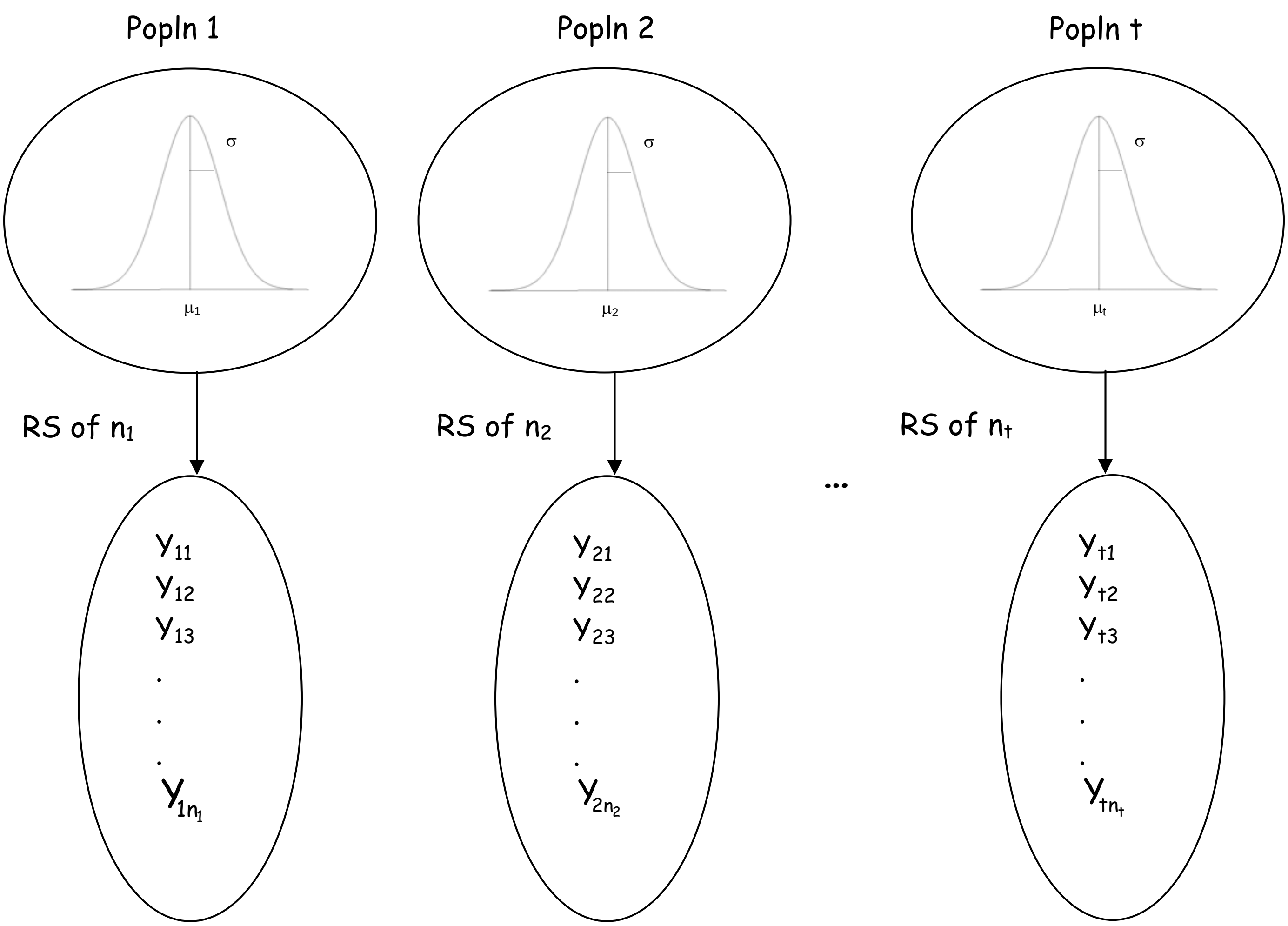
Week 6.2--IES 612-STA 4-573-STA 4-576.doc

THE GENERAL ONE-WAY ANOVA

ANOVA Data AND ASSUMPTIONS

Assume we take independent RS's of measurements from each of the "t" populations.



	Sample Values				
Population (aka Sample)	1	2	…	Sample variance	Sample Average
1	y_{11}	y_{12}	…	s_1^2	$\bar{y}_1$
2	y_{21}	y_{22}	…	s_2^2	$\bar{y}_2$
…	…	…		…	…
t	y_{t1}	y_{t2}	…	s_t^2	$\bar{y}_t$

MORE NOTATION

Notationally, Y_{ij} represents the j^{th} sample value from the i^{th} population.

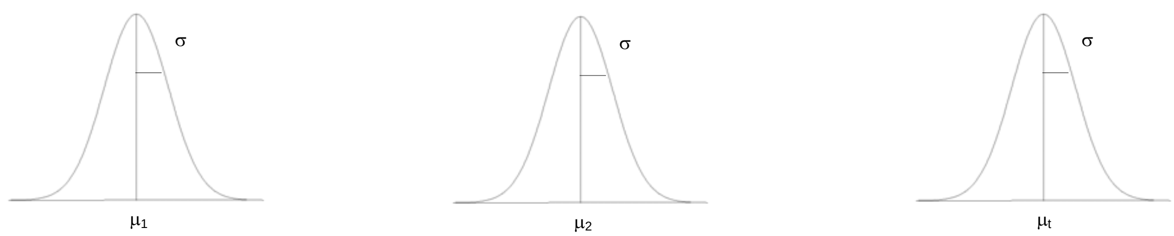
i = 1,2, …, t (populations)
j = 1, 2, …, n_i (n_i = number of observations from the i^{th} population)

$n_T = n_1 + n_2 + \ldots + n_t$

Assume Y_{ij} ~ independent $N(\mu_i, \sigma^2)$, i = 1,2, …, t; j = 1, 2, …, n_i

QUESTIONS OF INTEREST

Popln 1 Popln 2 Popln t



-

-

So our main or overall or initial hypothesis is

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_t$

H_A: $\mu_i \neq \mu_j$ [at least two population means differ]

SOME QUICK TERMINOLOGY

RESPONSE: Continuous measurement of interest

FACTOR: Categorical variable that "defines" the populations

LEVELS: The different values that the FACTOR can take on

Also known as **TREATMENTS**.

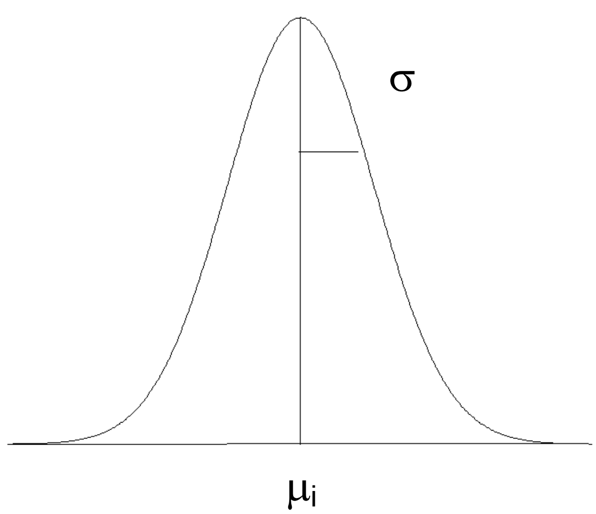
DESIGNED EXPERIMENTS:

versus OBSERVATIONAL STUDIES:

COMPLETELY RANDOMIZED DESIGNS (CRD):

ONE-WAY ANOVA MODEL (aka One-Factor ANOVA)

Model: $Y_{ij} = \mu_i + \varepsilon_{ij}$, random errors, ε_{ij}, are independent $N(0, \sigma^2)$



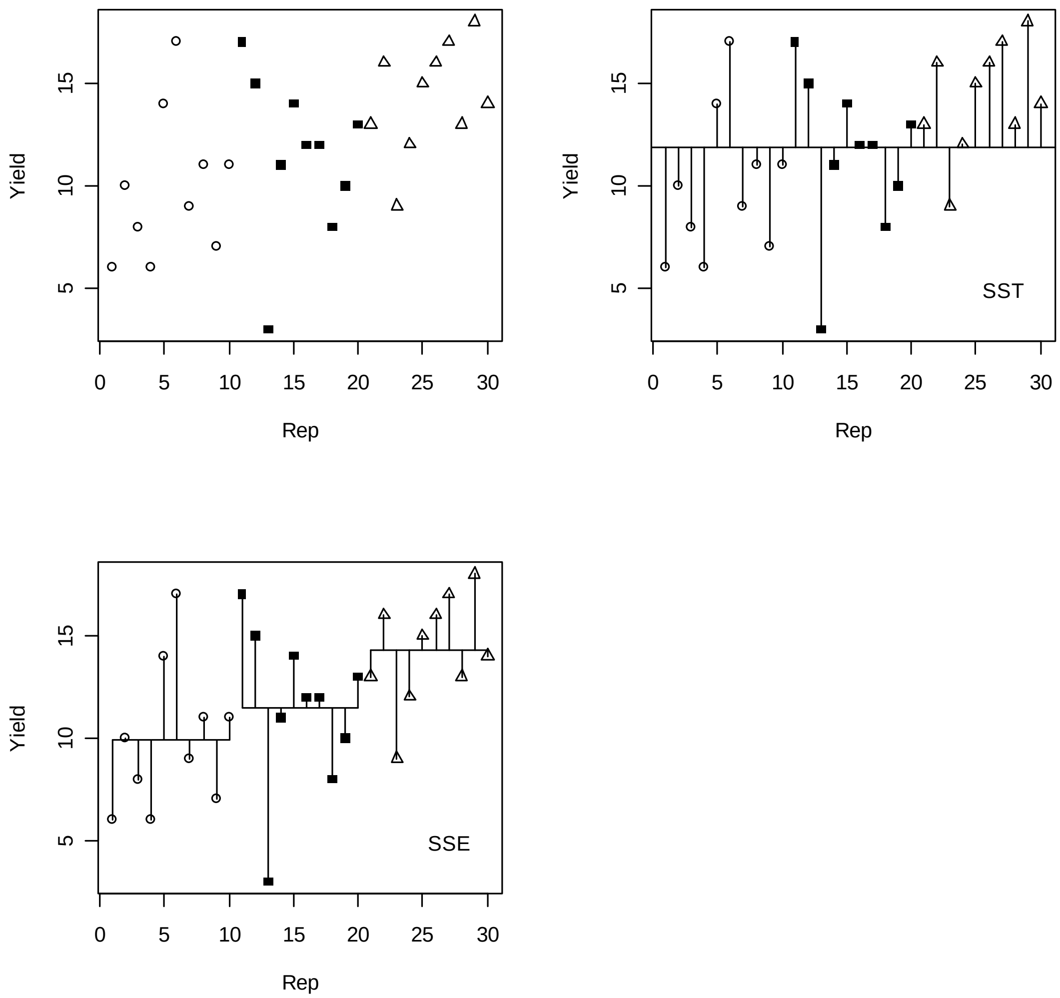
$$i^{th}\text{-sample-average}: \bar{y}_{i\bullet} = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

$$\text{Overall-average}: \bar{y}_{\bullet\bullet} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i} y_{ij}}{n_T}$$

PARTITIONING THE TOTAL SUM OF SQUARES

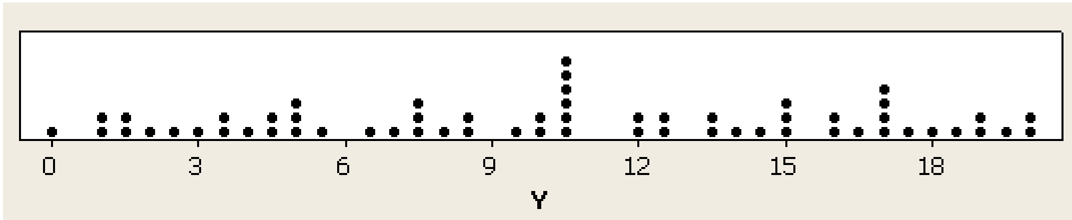
Data on Crop Yield and Soil Type from the R Book.

sand	6	10	8	6	14	17	9	11	7	11
clay	17	15	3	11	14	12	12	8	10	13
loam	13	16	9	12	15	16	17	13	18	14




```
par(mfrow=c(2,2))
plot(Rep, Yield, pch=(1+14*(Rep>10)+9*(Rep>20)))
plot(Rep, Yield, pch=(1+14*(Rep>10)+9*(Rep>20)))
for (i in 1:30) lines(c(i,i), c(Yield[i], mean(Yield)))
abline(h=mean(Yield))
text(27,5, "SST")
plot(Rep, Yield, pch=(1+14*(Rep>10)+9*(Rep>20)))
for (i in 1:10) lines(c(i,i), c(Yield[i], mean(Yield[1:10])))
segments(1,mean(Yield[1:10]) , 10, mean(Yield[1:10]))
for (i in 11:20) lines(c(i,i), c(Yield[i], mean(Yield[11:20])))
segments(11,mean(Yield[11:20]) , 20, mean(Yield[11:20]))
for (i in 21:30) lines(c(i,i), c(Yield[i], mean(Yield[21:30])))
segments(21,mean(Yield[21:30]) , 30, mean(Yield[21:30]))
text(27,5, "SSE")
```

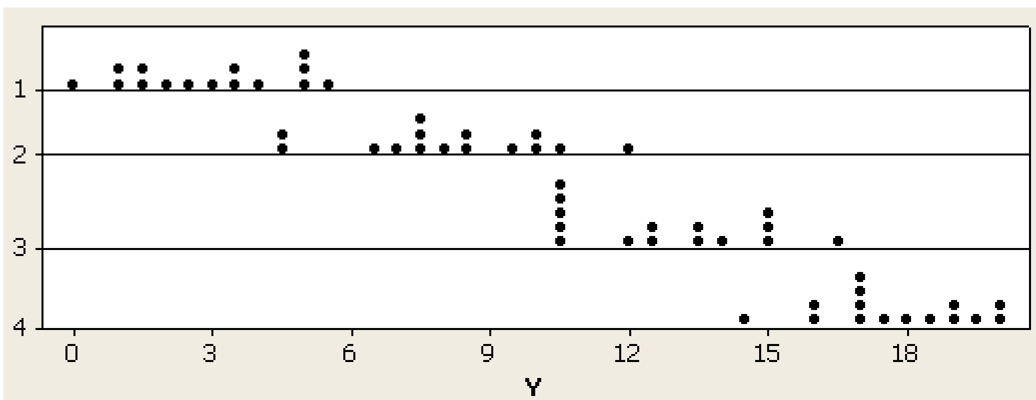
PARTITIONING THE TOTAL SUM OF SQUARES



Where's the average of these observations?

Define the TOTAL VARIATION or SUM OF SQUARES TOTAL as: $\sum(y-\bar{y})^2$

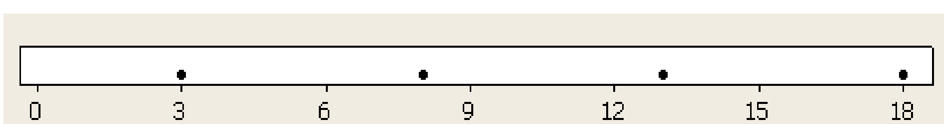
Now suppose the Y's above actually came from four different groups/poplns, so that:



Can we find the average for each group?

Then find the total variation of the Y's in that group around this average? $\sum\left(y_{ij}-\bar{y}_i\right)^2$

Then combine them to obtain the TOTAL VARIATION **WITHIN** GROUPS: $\sum\sum\left(y_{ij}-\bar{y}_i\right)^2$

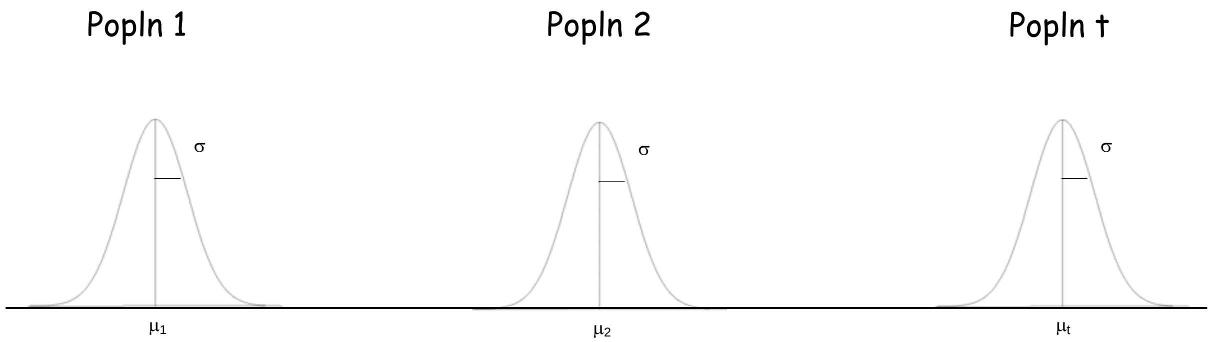


Can we find the average of these four points and the variation of these four points around this average, but we'll weight each deviation by the number of obs in that group.

This is the TOTAL VARIATION **BETWEEN** GROUPS: $\sum n_i\left(\bar{y}_i-\bar{y}\right)^2$

PARTITIONING THE TOTAL SUM OF SQUARES

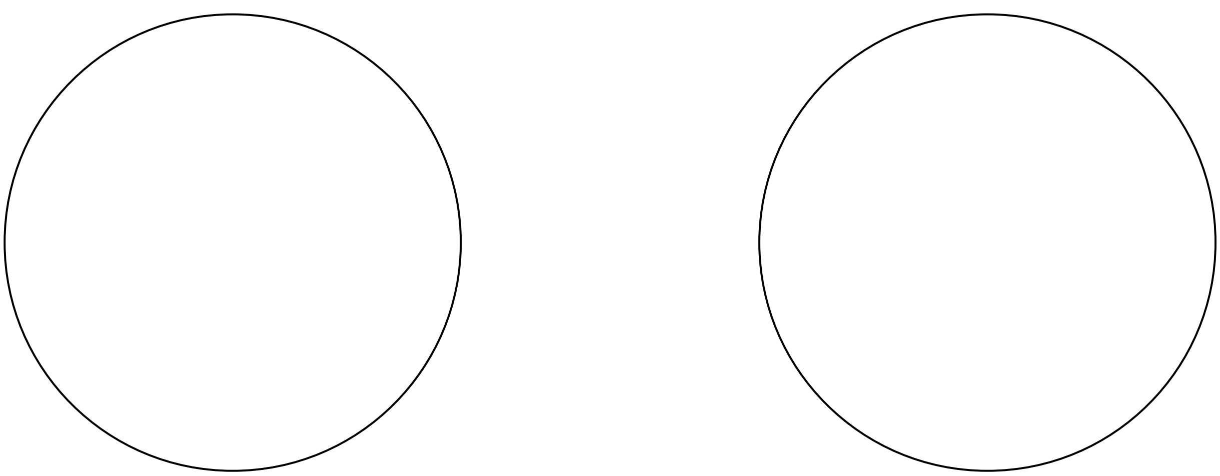
Defn: $SSTot = SST = TSS = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{..}\right)^2$.



Now define: Sum of Squares **<u>between(among)</u>** groups = $SSB = \sum_{i=1}^{t} n_i\left(\bar{y}_{i} - \bar{y}_{..}\right)^2$ and

Sum of Squares **<u>within</u>** groups = $SSW = \sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{i}\right)^2$.

These sums of squares add so that <u>SST = SSB + SSW</u>.



TEST STATISTIC?

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_t$

H_a: $\mu_i \neq \mu_j$ [at least two population means differ]

$$F_{obs} = \frac{S_B^2}{S_W^2}$$

where the **between(among)** group $S_B^2 = \frac{SSB}{t-1} = \frac{\sum_{i=1}^{t} n_i\left(\bar{y}_i - \bar{y}_{..}\right)^2}{t-1}$ variability is

and the **within** group variability is $S_W^2 = \frac{SSW}{n_T - t} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2}{n_1 + n_2 + \ldots + n_t - t}$.

Reject H_0 if: $F_{obs} > F_{\alpha, t-1, n_T - t}$ OR p-value = $\Pr(F_{t-1, n_T - t} > F_{obs}) < \alpha$.

ANOVA TABLE

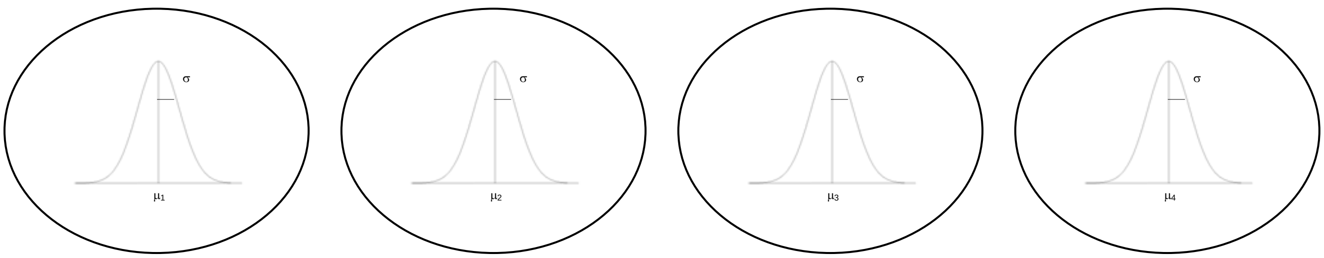
Source	SS	df	MS	F_{obs}	p-value
Between	SSB	t-1	SSB/(t-1)	MSB/MSW	$\Pr(F_{t-1, n_T - t} > F_{obs})$
Within	SSW	n_T-t	SSW/(n_T-t)		
Total	TSS	n_T-1			

Example Bacteria growth in meat under different packaging conditions (revisited)

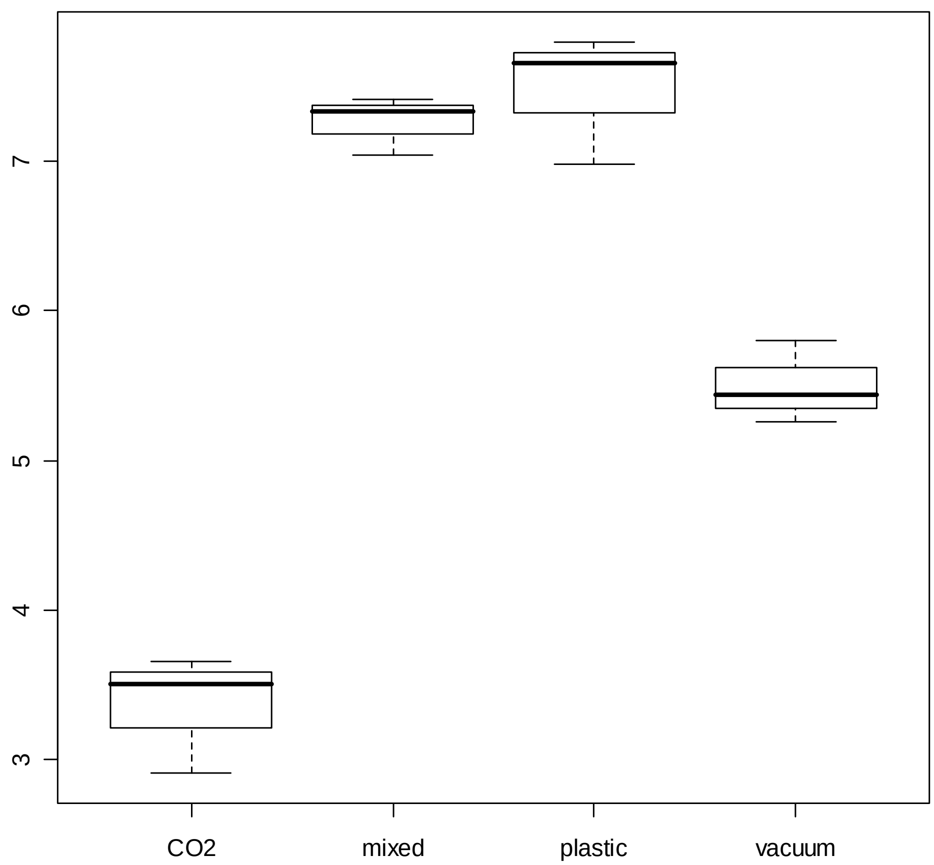
Using R

```
> #
> #  Bacteria Count and meat packaging Example
> #
>
> MeatData = as.data.frame(scan(what=list(Condition="", logCount=0), sep=",",
dec="."))
1: plastic, 7.66
2: plastic, 6.98
3: plastic, 7.80
4: vacuum, 5.26
5: vacuum, 5.44
6: vacuum, 5.80
7: mixed, 7.41
8: mixed, 7.33
9: mixed, 7.04
10: CO2, 3.51
11: CO2, 2.91
12: CO2, 3.66
13:
Read 12 records
> #
> # Alternative Data Entry from a .csv file
> #
> # MeatData = read.table(file("C:\\ ... \\Bacteria Data.csv"), header=TRUE)
> #
>
> attach(MeatData)
> MeatData
   Condition logCount
1    plastic     7.66
2    plastic     6.98
3    plastic     7.80
4     vacuum     5.26
5     vacuum     5.44
6     vacuum     5.80
7      mixed     7.41
8      mixed     7.33
9      mixed     7.04
10       CO2     3.51
11       CO2     2.91
12       CO2     3.66
> by(logCount, Condition, mean)
INDICES: CO2
[1] 3.36
------------------------------------------------------------
INDICES: mixed
[1] 7.26
------------------------------------------------------------
INDICES: plastic
[1] 7.48
------------------------------------------------------------
INDICES: vacuum
[1] 5.5
```

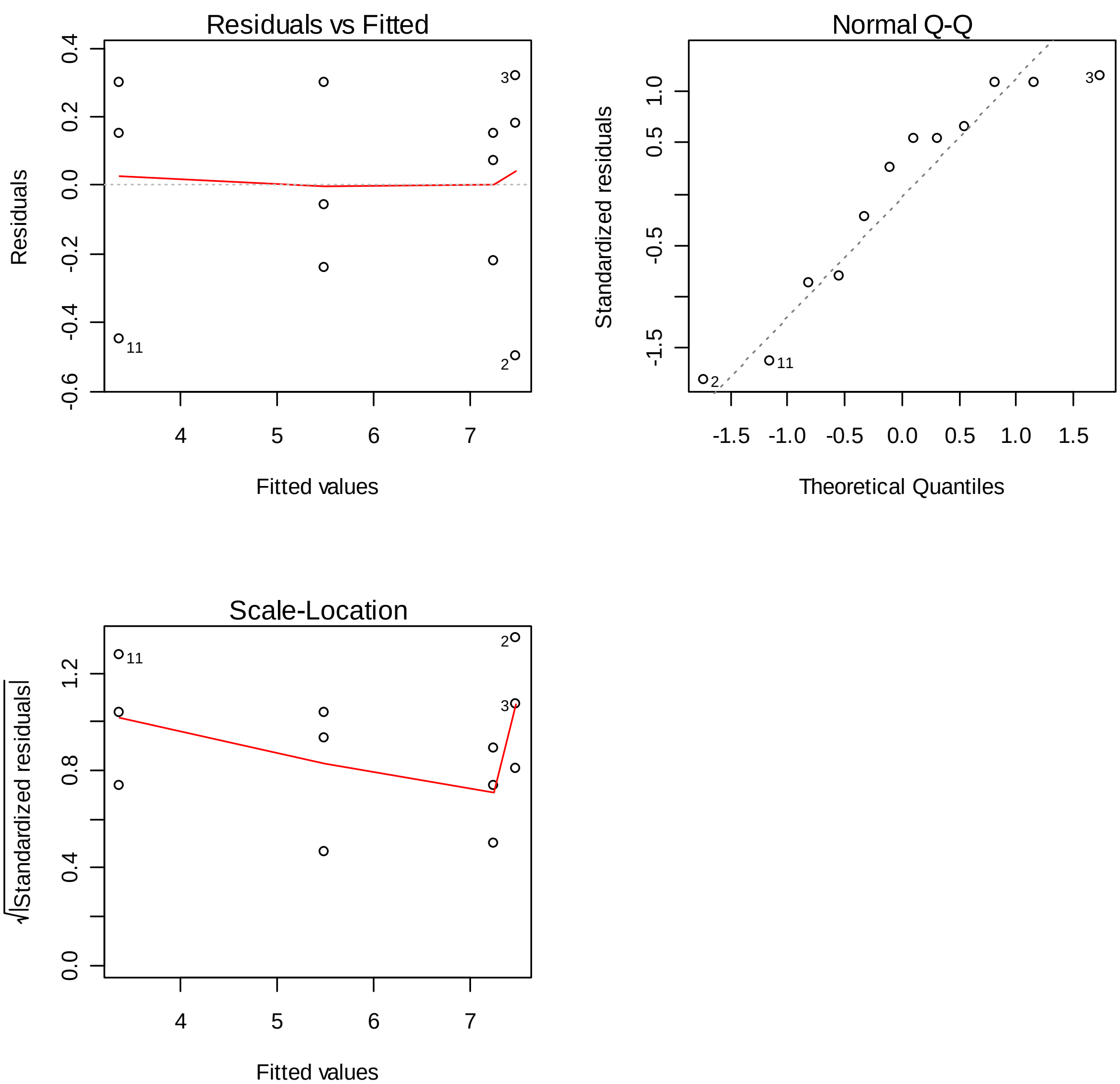
```
> by(logCount, Condition, sd)
INDICES: CO2
[1] 0.3968627
------------------------------------------------------------
INDICES: mixed
[1] 0.1946792
------------------------------------------------------------
INDICES: plastic
[1] 0.4386342
------------------------------------------------------------
INDICES: vacuum
[1] 0.2749545
> numSummary(logCount, groups=Condition, statistics=c("mean", "sd"))
Loading required package: abind
        mean        sd n
CO2     3.36 0.3968627 3
mixed   7.26 0.1946792 3
plastic 7.48 0.4386342 3
vacuum  5.50 0.2749545 3
```



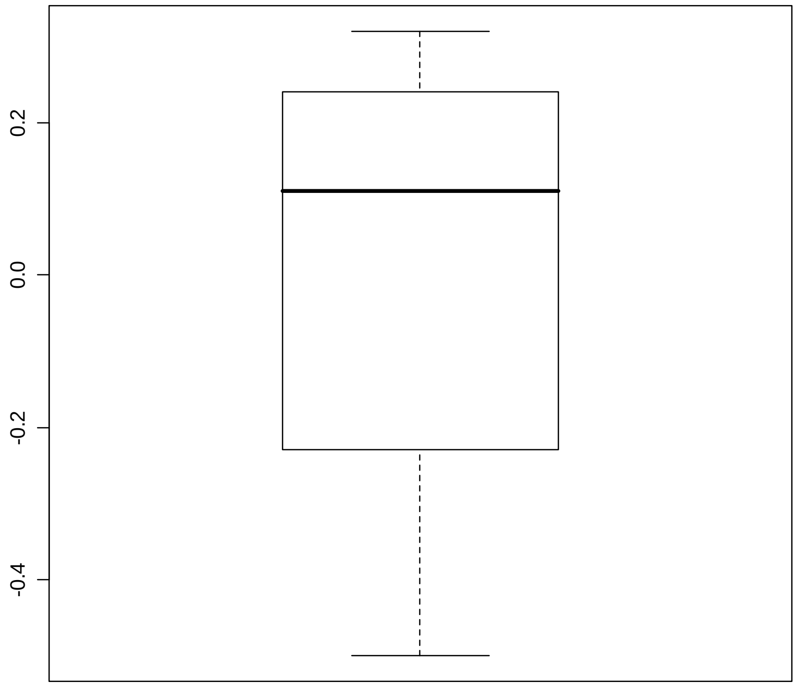

```
> boxplot(logCount ~ Condition)
```



```
> MeatData$iplas = as.numeric(Condition=="plastic")
> MeatData$imixed = as.numeric(Condition=="mixed")
> MeatData$ivac = as.numeric(Condition=="vacuum")
> MeatData
   Condition logCount iplas imixed ivac
1    plastic     7.66     1      0    0
2    plastic     6.98     1      0    0
3    plastic     7.80     1      0    0
4     vacuum     5.26     0      0    1
5     vacuum     5.44     0      0    1
6     vacuum     5.80     0      0    1
7      mixed     7.41     0      1    0
8      mixed     7.33     0      1    0
9      mixed     7.04     0      1    0
10       CO2     3.51     0      0    0
11       CO2     2.91     0      0    0
12       CO2     3.66     0      0    0
> MeatReg = lm(logCount ~ iplas + imixed + ivac, data=MeatData)
> par(mfrow=c(2,2))
> plot(MeatReg)
hat values (leverages) are all = 0.3333333
 and there are no factor predictors; no plot no. 5
> par(mfrow=c(1,1))
```

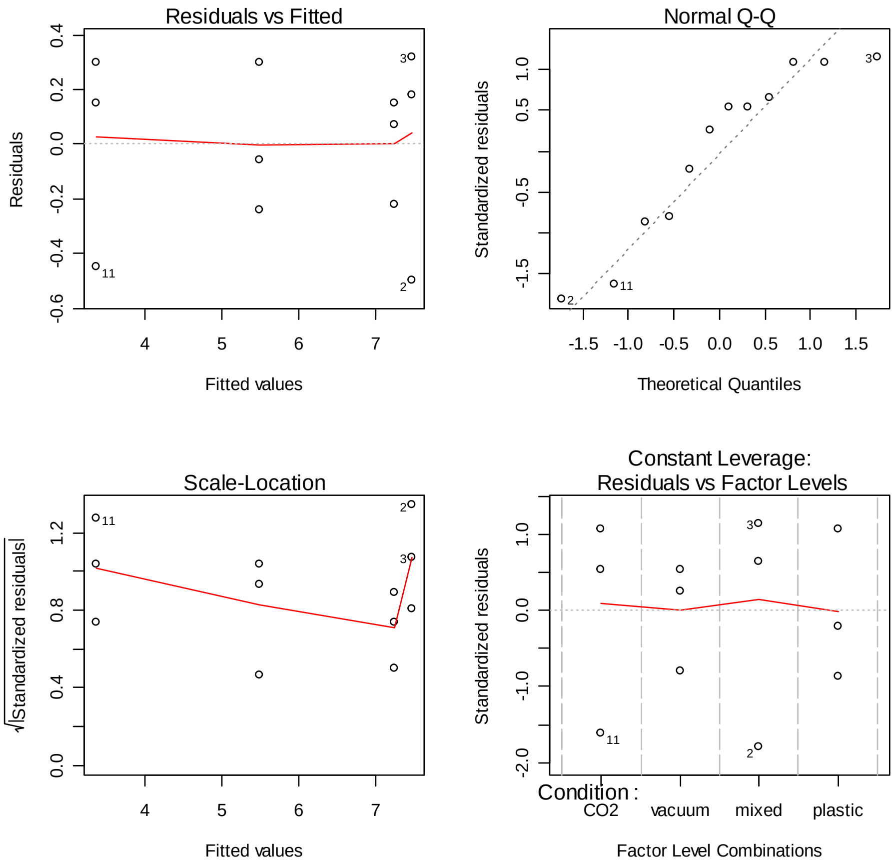


```
> shapiro.test(MeatReg$residuals)

        Shapiro-Wilk normality test

data:  MeatReg$residuals
W = 0.8942, p-value = 0.1336

> boxplot(MeatReg$residuals)
```

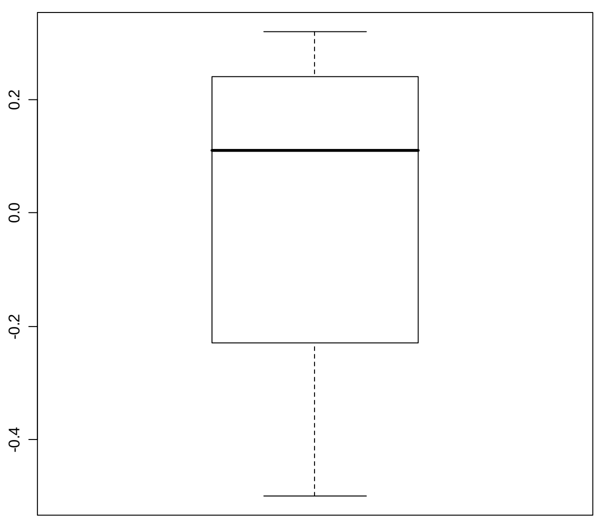



```
> summary(MeatReg)

Call:
lm(formula = logCount ~ iplas + imixed + ivac, data = MeatData)

Residuals:
   Min     1Q Median     3Q    Max
-0.500 -0.225  0.110  0.210  0.320

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.3600     0.1965   17.10 1.39e-07 ***
iplas         4.1200     0.2779   14.82 4.22e-07 ***
imixed        3.9000     0.2779   14.03 6.45e-07 ***
ivac          2.1400     0.2779    7.70 5.74e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3404 on 8 degrees of freedom
Multiple R-squared: 0.9726,     Adjusted R-squared: 0.9623
F-statistic: 94.58 on 3 and 8 DF,  p-value: 1.376e-06
```

```
> MeatANOVA = lm(logCount ~ Condition, data=MeatData)
> par(mfrow=c(2,2))
> plot(MeatANOVA)
> par(mfrow=c(1,1))
```

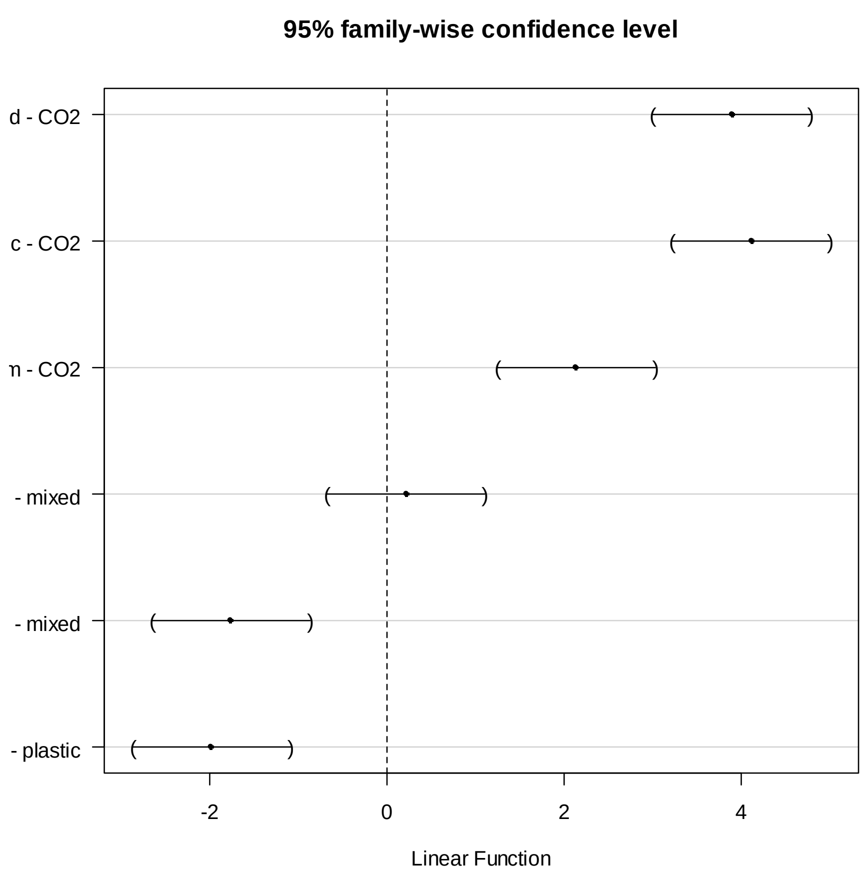



```
> shapiro.test(MeatANOVA$residuals)

        Shapiro-Wilk normality test

data:  MeatANOVA$residuals
W = 0.8942, p-value = 0.1336

> boxplot(MeatANOVA$residuals)
```



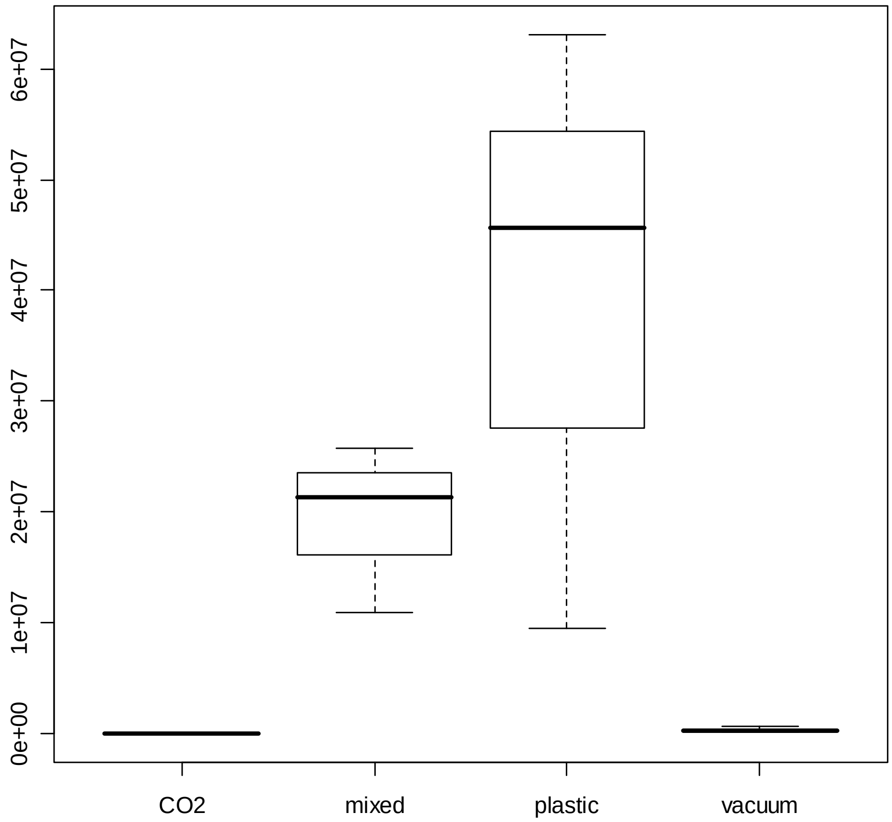
```
> summary(MeatANOVA)

Call:
lm(formula = logCount ~ Condition, data = MeatData)

Residuals:
   Min     1Q Median     3Q    Max 
-0.500 -0.225  0.110  0.210  0.320 

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)            3.3600     0.1965   17.10 1.39e-07 ***
Condition[T.mixed]     3.9000     0.2779   14.03 6.45e-07 ***
Condition[T.plastic]   4.1200     0.2779   14.82 4.22e-07 ***
Condition[T.vacuum]    2.1400     0.2779    7.70 5.74e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 0.3404 on 8 degrees of freedom
Multiple R-squared: 0.9726,     Adjusted R-squared: 0.9623 
F-statistic: 94.58 on 3 and 8 DF,  p-value: 1.376e-06 

> anova(MeatANOVA)
Analysis of Variance Table

Response: logCount
          Df Sum Sq Mean Sq F value    Pr(>F)    
Condition  3 32.873  10.958  94.584 1.376e-06 ***
Residuals  8  0.927   0.116                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 
```

```
> library(multcomp)
Loading required package: mvtnorm
> confint(MeatANOVA)
                          2.5 %   97.5 %
(Intercept)          2.906844 3.813156
Condition[T.mixed]   3.259141 4.540859
Condition[T.plastic] 3.479141 4.760859
Condition[T.vacuum]  1.499141 2.780859
> plot(print(confint(glht(MeatANOVA, linfct=mcp(Condition="Tukey")))))

         Simultaneous Confidence Intervals for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = logCount ~ Condition, data = MeatData)

Estimated Quantile = 3.2024

Linear Hypotheses:
                      Estimate lwr     upr
mixed - CO2 == 0       3.9000   3.0100  4.7900
plastic - CO2 == 0     4.1200   3.2300  5.0100
vacuum - CO2 == 0      2.1400   1.2500  3.0300
plastic - mixed == 0   0.2200  -0.6700  1.1100
vacuum - mixed == 0   -1.7600  -2.6500 -0.8700
vacuum - plastic == 0 -1.9800  -2.8700 -1.0900

95% family-wise confidence level
```

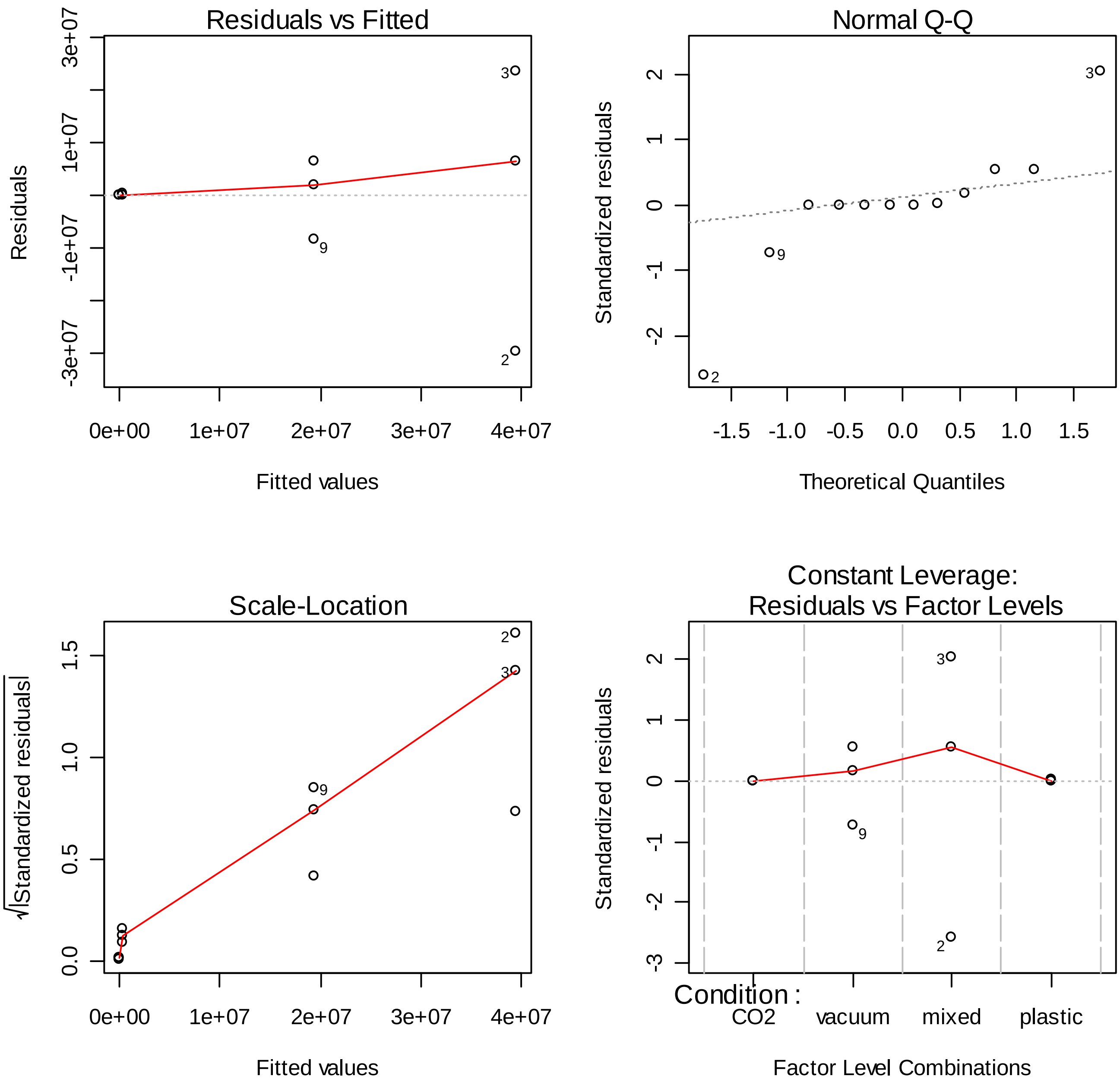


WHAT IF WE USED THE BACTERIAL COUNTS RATHER THAN THE logCOUNTS?

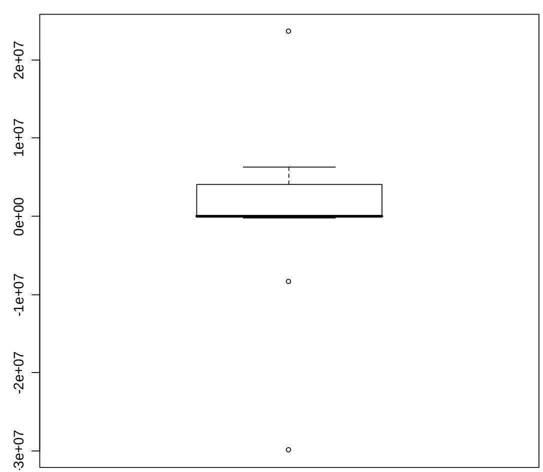
```
> MeatData$Count = 10^logCount
> detach(MeatData)
> attach(MeatData)
> numSummary(Count, groups=Condition, statistics=c("mean", "sd"))
                mean           sd n
CO2         2873.216     1905.102 3
mixed   19349453.561  7576411.476 3
plastic 39451493.090 27315819.586 3
vacuum    362783.434   236899.326 3
> boxplot(Count ~ Condition)
```



```
> MeatANOVACount = lm(Count ~ Condition, data=MeatData)
> par(mfrow=c(2,2))
> plot(MeatANOVACount)
> par(mfrow=c(1,1))
```

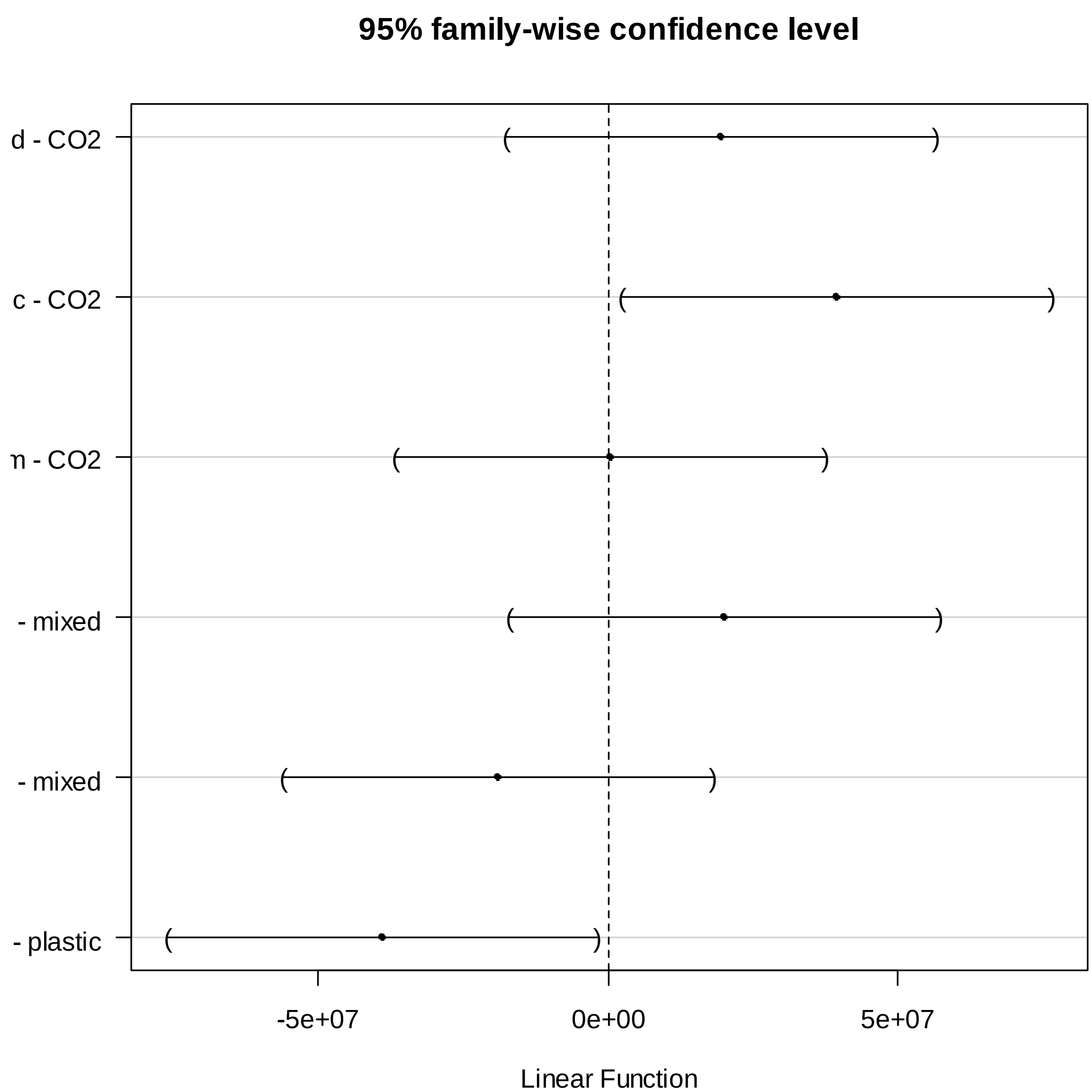



```
> shapiro.test(MeatANOVACount$residuals)

        Shapiro-Wilk normality test

data:  MeatANOVACount$residuals
W = 0.8102, p-value = 0.01227

> boxplot(MeatANOVACount$residuals)
```



```
> summary(MeatANOVACount)

Call:
lm(formula = Count ~ Condition, data = MeatData)

Residuals:
      Min        1Q    Median        3Q       Max
-29901567   -110724      1030   3086957  23644241

Coefficients:
                     Estimate Std. Error  t value Pr(>|t|)
(Intercept)              2873    8183378 0.000351  0.99973
Condition[T.mixed]   19346580   11573044    1.672  0.13313
Condition[T.plastic] 39448620   11573044    3.409  0.00924 **
Condition[T.vacuum]    359910   11573044    0.031  0.97595
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14170000 on 8 degrees of freedom
Multiple R-squared: 0.6634,     Adjusted R-squared: 0.5371
F-statistic: 5.255 on 3 and 8 DF,  p-value: 0.02699

> anova(MeatANOVACount)
Analysis of Variance Table

Response: Count
          Df     Sum Sq    Mean Sq F value  Pr(>F)
Condition  3 3.1673e+15 1.0558e+15  5.2552 0.02699 *
Residuals  8 1.6072e+15 2.0090e+14
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> library(multcomp)
> plot(print(confint(glht(MeatANOVACount, linfct=mcp(Condition="Tukey")))))

         Simultaneous Confidence Intervals for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = Count ~ Condition, data = MeatData)

Estimated Quantile = 3.2009

Linear Hypotheses:
                      Estimate       lwr            upr
mixed - CO2 == 0       19346580.3451 -17697030.1975  56390190.8876
plastic - CO2 == 0     39448619.8736   2405009.3311  76492230.4161
vacuum - CO2 == 0        359910.2172 -36683700.3253  37403520.7598
plastic - mixed == 0   20102039.5285 -16941571.0140  57145650.0710
vacuum - mixed == 0   -18986670.1278 -56030280.6703  18056940.4147
vacuum - plastic == 0 -39088709.6563 -76132320.1989  -2045099.1138

95% family-wise confidence level
```



```
title “One-way ANOVA/ CRD example + contrasts + multiple comparisons”;
title2 “Bacteria in meat data”;
data meat;
  input condition $ logcount @@;
  ivac = (condition=”vacuum”);
  imix = (condition=”mixed”);
  iCO2 = (condition=”CO2”);
  cards;
plastic 7.66 plastic 6.98 plastic 7.80
 vacuum 5.26  vacuum 5.44  vacuum 5.80
  mixed 7.41   mixed 7.33   mixed 7.04
    CO2 3.51     CO2 2.91     CO2 3.66
;
run;

proc print data=meat;
run;

proc sort out=smeat; by condition;
proc univariate plot; by condition;
  title3 summary statistics and boxplot;
  var logcount;
run;

proc reg data=meat;
  title3  Regression with indicators;
  model logcount = ivac imix iCO2;
run;

proc glm data=meat order=data;
title3  One-way anova + contrast + model adequacy;
  class condition;
  model logcount=condition;
  output out=new p=yhat r=resid;
  contrast 'plastic vs. rest' condition 3 -1 -1 -1;
  estimate 'plastic vs. rest' condition 3 -1 -1 -1;
  contrast 'CO2 vs. plastic'  condition -1  0  0 1;
  estimate 'CO2 vs. plastic'  condition -1  0  0 1;
  contrast 'CO2 vs. vacuum'   condition  0 -1  0 1;
  estimate 'CO2 vs. vacuum'   condition  0 -1  0 1;
  contrast 'CO2 vs. mixed'    condition  0  0 -1 1;
  estimate 'CO2 vs. mixed'    condition  0  0 -1 1;
  lsmeans condition / stderr pdiff;
  means condition / lsd clm;
  means condition / bon scheffe tukey;
  means condition / bon tukey cldiff;
run;

proc plot data=new;
  plot logcount*condition yhat*condition='p' /overlay;
  plot resid*condition resid*yhat / vref=0;
  run;

proc univariate plot;
  var resid;
run;

* construct the normal scores - Z[(i-.375)/(n+.25)];
```

```
* note not multiplied by sqrt(mse);
proc rank data=new normal=blom out=rnew;
  var resid;
  ranks nscore;
run;

* generate plot analogous to univariate's normal prob. plot;

proc plot;
  plot resid*nscore;
run;

data moremeat; set meat;
  count = exp(logcount);
  title3 raw count data analyzed;
run;

proc glm data=moremeat;
  class condition;
  model count=condition;
  output out=mnew p=yhat r=resid;
  lsmeans condition / stderr pdiff;
*  means condition / clm bon scheffe lsd tukey snk;
run;

proc plot data=mnew;
  plot count*condition yhat*condition='p' /overlay;
  plot resid*condition resid*yhat / vref=0;
run;

proc univariate data=mnew plot;
  var resid;
run;

proc rank data=mnew normal=blom out=rnew;
  var resid;
  ranks nscore;
run;

proc plot;
  plot resid*nscore;
run;
```

```
proc print data=meat;
run;
```

Obs	condition	logcount	ivac	imix	iCO2
1	plastic	7.66	0	0	0
2	plastic	6.98	0	0	0
3	plastic	7.80	0	0	0
4	vacuum	5.26	1	0	0
5	vacuum	5.44	1	0	0
6	vacuum	5.80	1	0	0
7	mixed	7.41	0	1	0
8	mixed	7.33	0	1	0
9	mixed	7.04	0	1	0
10	CO2	3.51	0	0	1
11	CO2	2.91	0	0	1
12	CO2	3.66	0	0	1

```
proc sort out=smeat; by condition;
run;
proc univariate plot; by condition;
  title3 summary statistics and boxplot;
  var logcount;
run;
```

```
The UNIVARIATE Procedure
Variable:  logcount

Schematic Plots
       8 +
         |                                 *-----*
         |                     +-----+     |  +  |
         |                     *--+--*     |     |
       7 +                     +-----+     +-----+
         |
         |
         |
       6 +
         |                                             +-----+
         |                                             *--+--*
         |                                             +-----+
       5 +
         |
         |
         |
       4 +
         |         +-----+
         |         *-----*
         |         |  +  |
       3 +         +-----+
         |
         |
         |
       2 +
          ------------+-----------+-----------+-----------+-----------
 condition               CO2       mixed     plastic      vacuum
```

```
proc reg data=meat;
  title3  Regression with indicators;
  model logcount = ivac imix iCO2;
run;
```

```
The REG Procedure
Model: MODEL1
Dependent Variable: logcount

                            Analysis of Variance
                                   Sum of           Mean
Source                   DF       Squares         Square    F Value    Pr > F
Model                     3      32.87280       10.95760      94.58    <.0001
Error                     8       0.92680        0.11585
Corrected Total          11      33.79960


Root MSE              0.34037    R-Square     0.9726
Dependent Mean        5.90000    Adj R-Sq     0.9623
Coeff Var             5.76894

                        Parameter Estimates

                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|

Intercept     1        7.48000        0.19651      38.06      <.0001
ivac          1       -1.98000        0.27791      -7.12      <.0001
imix          1       -0.22000        0.27791      -0.79      0.4514
iCO2          1       -4.12000        0.27791     -14.83      <.0001
```

```
proc glm data=meat order=data;
title3  One-way anova + contrast + model adequacy;
  class condition;
  model logcount=condition;
  output out=new p=yhat r=resid;
run;
```

```
The GLM Procedure
             Class Level Information
Class          Levels    Values
condition           4    plastic vacuum mixed CO2
Number of observations    1

The GLM Procedure
Dependent Variable: logcount
                                        Sum of
Source                      DF         Squares     Mean Square    F Value    Pr > F
Model                        3     32.87280000     10.95760000      94.58    <.0001
Error                        8      0.92680000      0.11585000
Corrected Total             11     33.79960000

R-Square      Coeff Var      Root MSE    logcount Mean
0.972580      5.768940       0.340367         5.900000

Source                      DF       Type I SS     Mean Square    F Value    Pr > F
condition                      3     32.87280000     10.95760000      94.58    <.0001

Source                      DF     Type III SS     Mean Square    F Value    Pr > F
condition                      3     32.87280000     10.95760000      94.58    <.0001
```

```
contrast 'plastic vs. rest' condition 3 -1 -1 -1;
estimate 'plastic vs. rest' condition 3 -1 -1 -1;
contrast 'CO2 vs. plastic'  condition -1  0  0 1;
estimate 'CO2 vs. plastic'  condition -1  0  0 1;
contrast 'CO2 vs. vacuum'   condition  0 -1  0 1;
estimate 'CO2 vs. vacuum'   condition  0 -1  0 1;
contrast 'CO2 vs. mixed'    condition  0  0 -1 1;
estimate 'CO2 vs. mixed'    condition  0  0 -1 1;
```

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
plastic vs. rest	1	9.98560000	9.98560000	86.19	<.0001
CO2 vs. plastic	1	25.46160000	25.46160000	219.78	<.0001
CO2 vs. vacuum	1	6.86940000	6.86940000	59.30	<.0001
CO2 vs. mixed	1	22.81500000	22.81500000	196.94	<.0001

Dependent Variable: logcount

Parameter	Estimate	Standard Error	t Value	Pr > |t|
plastic vs. rest	6.32000000	0.68073490	9.28	<.0001
CO2 vs. plastic	-4.12000000	0.27790886	-14.83	<.0001
CO2 vs. vacuum	-2.14000000	0.27790886	-7.70	<.0001
CO2 vs. mixed	-3.90000000	0.27790886	-14.03	<.0001

```
lsmeans condition / stderr pdiff;
```

The GLM Procedure
Least Squares Means

condition	logcount LSMEAN	Standard Error	Pr > |t|	LSMEAN Number
plastic	7.48000000	0.19651124	<.0001	1
vacuum	5.50000000	0.19651124	<.0001	2
mixed	7.26000000	0.19651124	<.0001	3
CO2	3.36000000	0.19651124	<.0001	4

Least Squares Means for effect condition
Pr > |t| for H0: LSMean(i)=LSMean(j)
Dependent Variable: logcount

i/j	1	2	3	4
1		<.0001	0.4514	<.0001
2	<.0001		0.0002	<.0001
3	0.4514	0.0002		<.0001
4	<.0001	<.0001	<.0001	

NOTE: To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used

```
means condition / lsd clm;
```

t Confidence Intervals for logcount

Alpha	0.05
Error Degrees of Freedom	8
Error Mean Square	0.11585
Critical Value of t	2.30600
Half Width of Confidence Interval	0.453156

condition	N	Mean	95% Confidence Limits	
plastic	3	7.4800	7.0268	7.9332
mixed	3	7.2600	6.8068	7.7132
vacuum	3	5.5000	5.0468	5.9532
CO2	3	3.3600	2.9068	3.8132

```
means condition / bon scheffe tukey;
```

Tukey's Studentized Range (HSD) Test for logcount

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	8
Error Mean Square	0.11585
Critical Value of Studentized Range	4.52880
Minimum Significant Difference	0.89

Means with the same letter are not significantly different.

	Mean	N	condition
A	7.4800	3	plastic
A			
A	7.2600	3	mixed
B	5.5000	3	vacuum
C	3.3600	3	CO2

Bonferroni (Dunn) t Tests for logcount

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	8
Error Mean Square	0.11585
Critical Value of t	3.47888
Minimum Significant Difference	0.9668

Means with the same letter are not significantly different.

	Mean	N	condition
A	7.4800	3	plastic
A			
A	7.2600	3	mixed

```
B        5.5000      3    vacuum

C        3.3600      3    CO2
```

Scheffe's Test for logcount

```
NOTE: This test controls the Type I experimentwise error rate.

Alpha                               0.05
Error Degrees of Freedom               8
Error Mean Square                0.11585
Critical Value of F              4.06618
Minimum Significant Difference    0.9706

Means with the same letter are not significantly different.
           Mean      N    condition
A        7.4800      3    plastic
A
A        7.2600      3    mixed

B        5.5000      3    vacuum

C        3.3600      3    CO
```

```
means condition / bon tukey cldiff;
```

Tukey's Studentized Range (HSD) Test for logcount

```
NOTE: This test controls the Type I experimentwise error rate.
Alpha                                    0.05
Error Degrees of Freedom                    8
Error Mean Square                     0.11585
Critical Value of Studentized Range   4.52880
Minimum Significant Difference           0.89

Comparisons significant at the 0.05 level are indicated by ***.

                     Difference
     condition          Between     Simultaneous 95%
   Comparison             Means    Confidence Limits
plastic - mixed          0.2200     -0.6700   1.1100
plastic - vacuum         1.9800      1.0900   2.8700   ***
plastic - CO2            4.1200      3.2300   5.0100   ***
mixed   - plastic       -0.2200     -1.1100   0.6700
mixed   - vacuum         1.7600      0.8700   2.6500   ***
mixed   - CO2            3.9000      3.0100   4.7900   ***
vacuum  - plastic       -1.9800     -2.8700  -1.0900   ***
vacuum  - mixed         -1.7600     -2.6500  -0.8700   ***
vacuum  - CO2            2.1400      1.2500   3.0300   ***
CO2     - plastic       -4.1200     -5.0100  -3.2300   ***
CO2     - mixed         -3.9000     -4.7900  -3.0100   ***
CO2     - vacuum        -2.1400     -3.0300  -1.2500   **
```

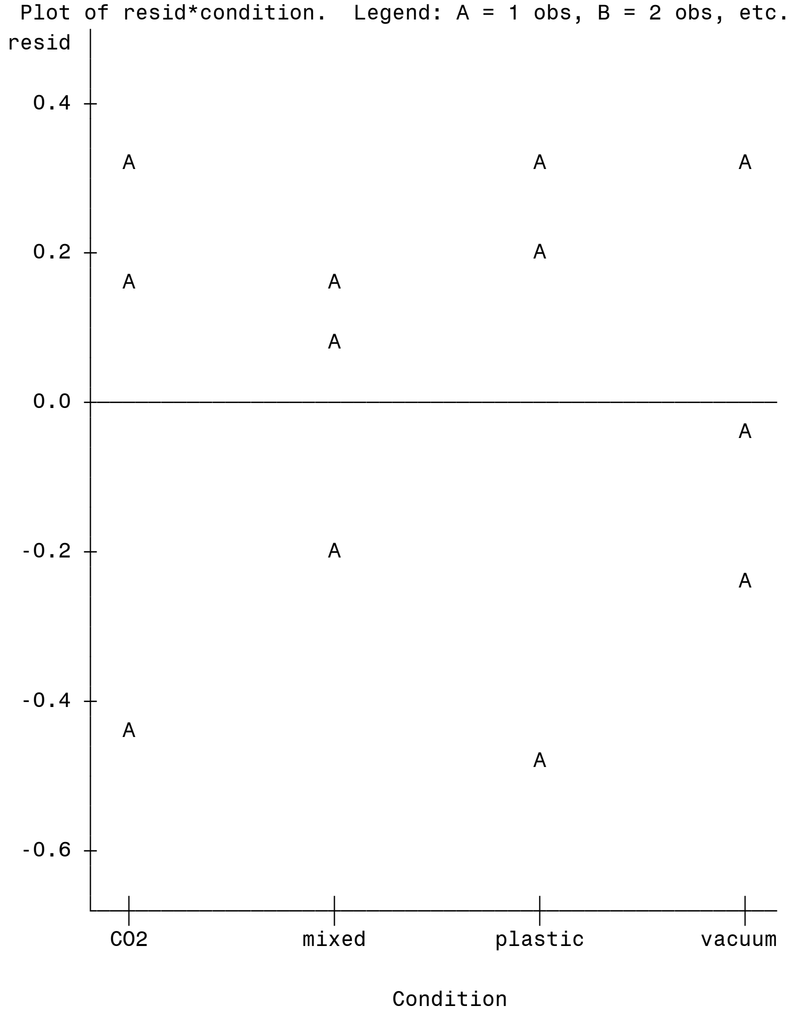
Bonferroni (Dunn) t Tests for logcount

```
NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate
than Tukey's for all pairwise comparisons.
Alpha                               0.05
Error Degrees of Freedom               8
Error Mean Square                0.11585
Critical Value of t              3.47888
Minimum Significant Difference    0.9668
```

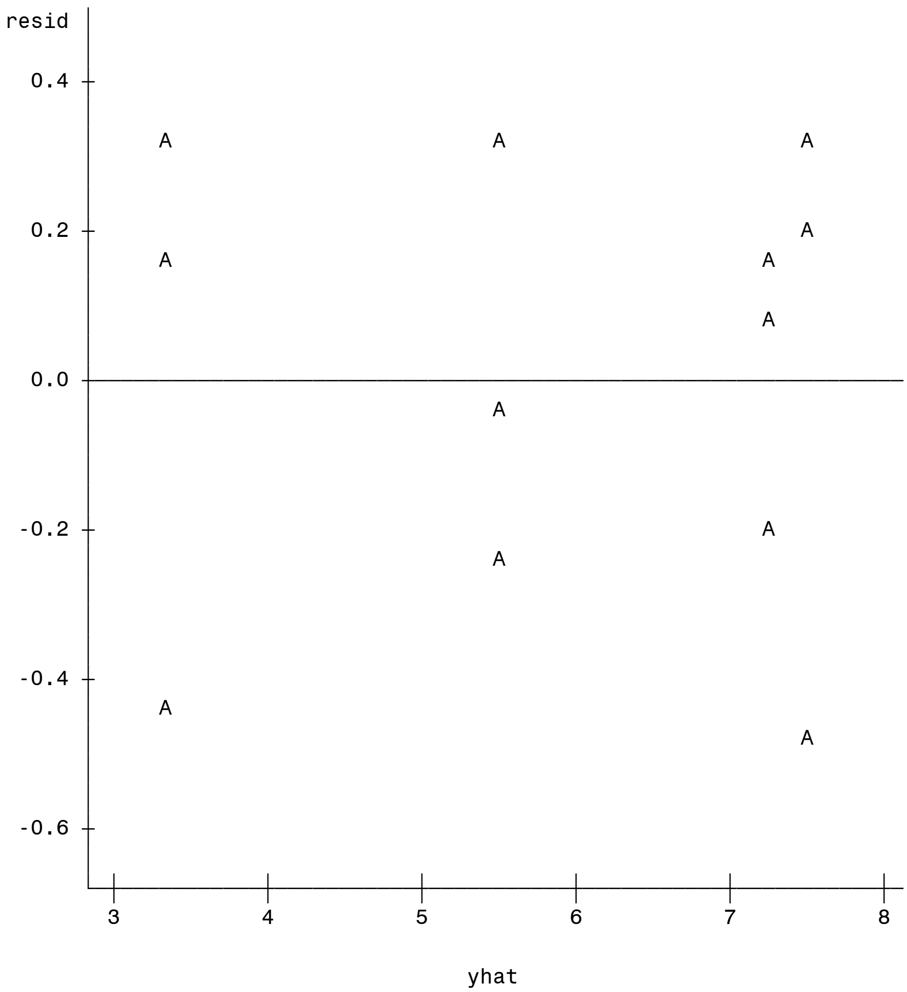
```
Comparisons significant at the 0.05 level are indicated by ***.
                     Difference
     condition            Between     Simultaneous 95%
   Comparison              Means    Confidence Limits
plastic - mixed           0.2200     -0.7468    1.1868
plastic - vacuum          1.9800      1.0132    2.9468  ***
plastic - CO2             4.1200      3.1532    5.0868  ***
mixed   - plastic        -0.2200     -1.1868    0.7468
mixed   - vacuum          1.7600      0.7932    2.7268  ***
mixed   - CO2             3.9000      2.9332    4.8668  ***
vacuum  - plastic        -1.9800     -2.9468   -1.0132  ***
vacuum  - mixed          -1.7600     -2.7268   -0.7932  ***
vacuum  - CO2             2.1400      1.1732    3.1068  ***
CO2     - plastic        -4.1200     -5.0868   -3.1532  ***
CO2     - mixed          -3.9000     -4.8668   -2.9332  ***
CO2     - vacuum         -2.1400     -3.1068   -1.1732  ***
```

```
options ls=70;
proc plot data=new;
  plot logcount*condition yhat*condition='p' /overlay;
  plot resid*condition resid*yhat / vref=0;
run;
```

Plot of resid*condition. Legend: A = 1 obs, B = 2 obs, etc.

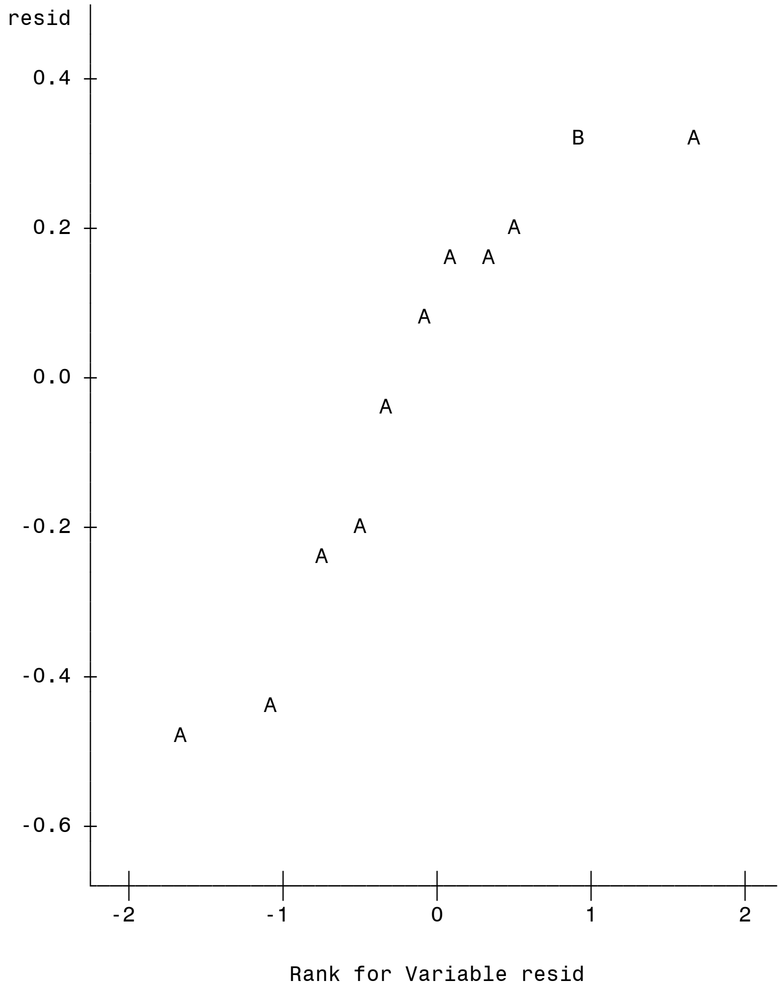


Plot of resid*yhat. Legend: A = 1 obs, B = 2 obs, etc.



```
proc univariate plot;
  var resid;

* construct the normal scores - Z[(i-.375)/(n+.25)];
* note not multiplied by sqrt(mse);
proc rank data=new normal=blom out=rnew;
  var resid;
  ranks nscore;

* generate plot analogous to univariate's normal prob. plot;

proc plot;
  plot resid*nscore;
```

Plot of resid*nscore. Legend: A = 1 obs, B = 2 obs, etc.




```
data moremeat; set meat;
  count = exp(logcount);
  title3 raw count data analyzed;

proc glm data=moremeat;
  class condition;
  model count=condition;
  output out=mnew p=yhat r=resid;
  lsmeans condition / stderr pdiff;
*  means condition / clm bon scheffe lsd tukey snk;
run;
```

The GLM Procedure
Dependent Variable: count

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	7282652.348	2427550.783	16.56	0.0009
Error	8	1172820.616	146602.577		
Corrected Total	11	8455472.964			

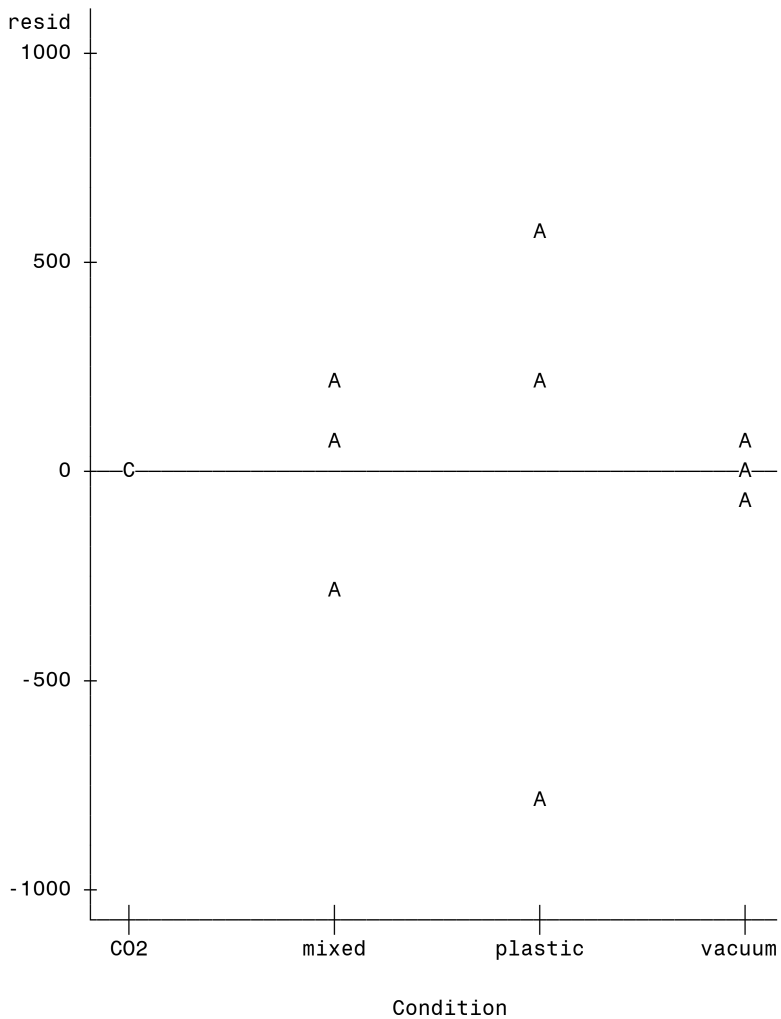
R-Square	Coeff Var	Root MSE	count Mean
0.861294	42.54159	382.8872	900.0303

Source	DF	Type I SS	Mean Square	F Value	Pr > F
condition	3	7282652.348	2427550.783	16.56	0.0009

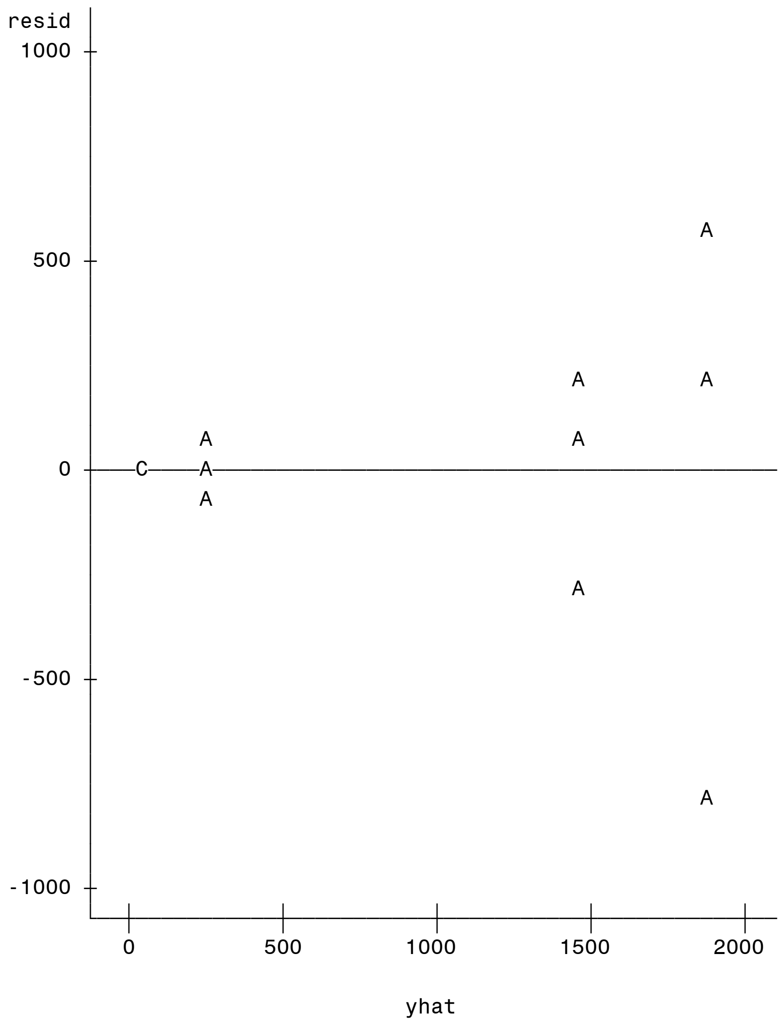
Source	DF	Type III SS	Mean Square	F Value	Pr > F
condition	3	7282652.348	2427550.783	16.56	0.0009

```
proc plot data=mnew;
  plot count*condition yhat*condition='p' /overlay;
  plot resid*condition resid*yhat / vref=0;
run;
```

Plot of resid*condition. Legend: A = 1 obs, B = 2 obs, etc.



Plot of resid*yhat. Legend: A = 1 obs, B = 2 obs, etc.



```
proc rank data=mnew normal=blom out=rnew;
  var resid;
  ranks nscore;
run;
proc plot;
  plot resid*nscore;
run;
```

Plot of resid*nscore. Legend: A = 1 obs, B = 2 obs, etc.