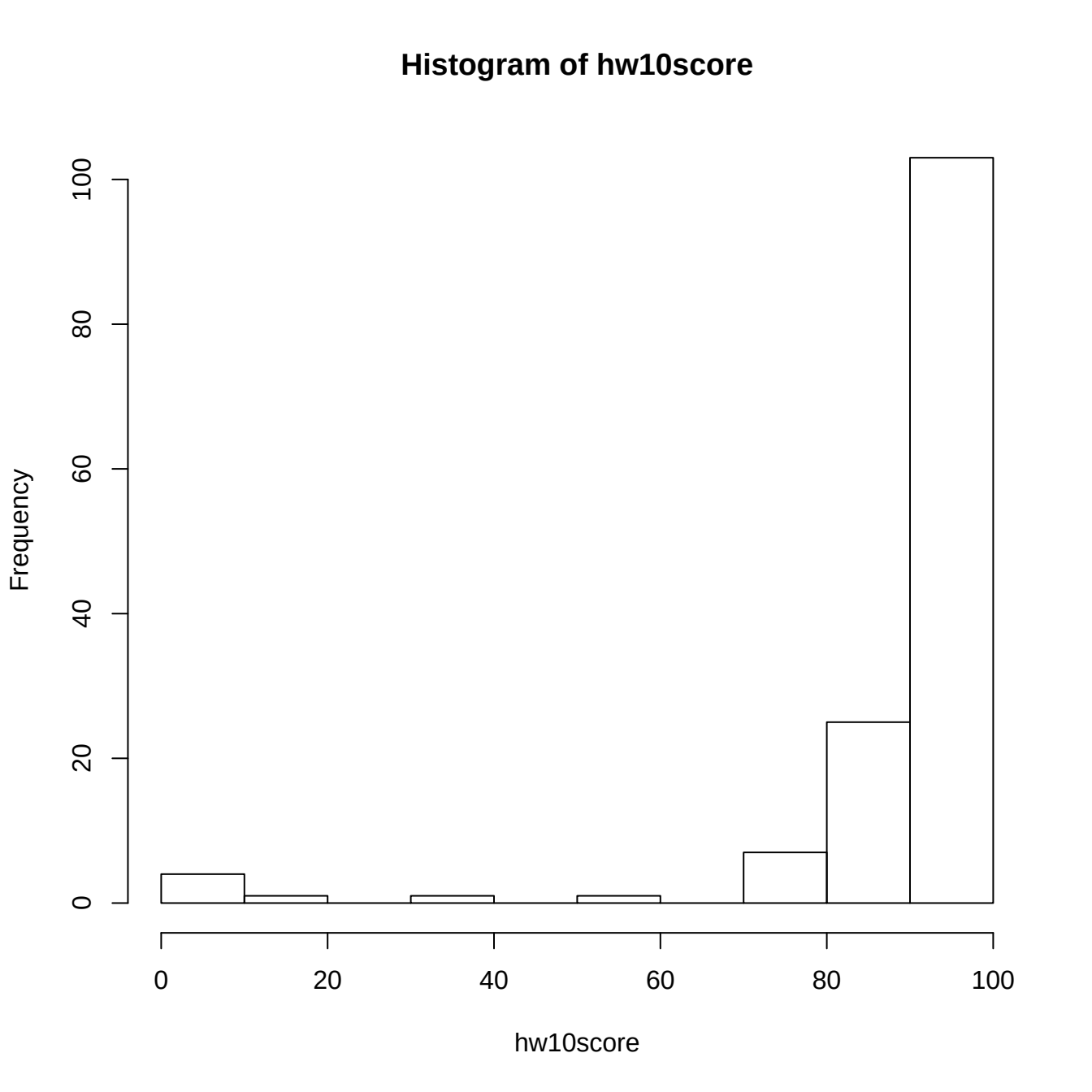

Histogram of hw10score
Frequency
0
20
40
60
80
100
hw10score
0
20
40
60
80
100

$\alpha = 0.05$

4. $\bar{x}_{..} = \frac{1}{n_T} \sum_{i=1}^{k} n_i \cdot \bar{x}_{i.}$

$SSTr = \sum_{i=1}^{k} n_i (\bar{x}_{i.} - \bar{x}_{..})^2$ (+10)

$SSE = SST - SSTr.$

Using the above, ANOVA Table is :

	df	SS	MS	F	p
Tr	6	7.6138	1.269	0.7782	0.5895.
E	27	125.562	1.6307		
Total	33	133.18			

(+10)

8. $n_T = 33$, $k = 3$, $q_{0.05,3,30} = 3.49$, $MSE = 4.96$.

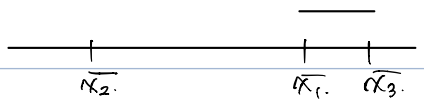
Formula for 95% CI of $\mu_i - \mu_j$ is

$$\bar{x}_{i.} - \bar{x}_{j.} \pm \sqrt{MSE} \frac{q_{0.05,3,30}}{\sqrt{2}} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

a) 1,2 : (0.9665, 5.6535)

1,3 : (−3.4035, 1.2835) $\ni 0$ ← 1,3 : not distinguishable.

2,3 : (−6.7135, −2.0265) (+8)

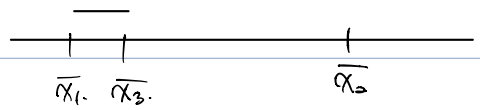
b) 

(+6)

c) $L^2 = \frac{4 MSE\, q_{0.05,3,30}^2}{n} \leq 4 \Rightarrow n \geq MSE \cdot q_{0.05,3,30}^2 = 60.xx$

$\therefore n \geq 61$, so 50 more. (+6)

16. See the Table below. $p < 0.05$. (reject the null)

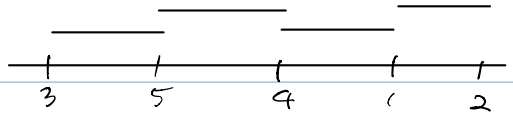


⇒ There is an evidence that layout 2 is the slowest. (+5)

ANOVA Table (+5), pairwise CI (+10).

20. a) ANOVA Table (+10) $p < 0.05$.

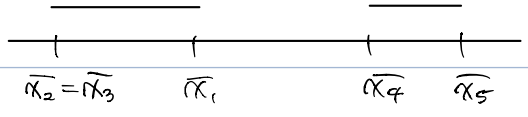
b)



∴ (3,5) : smallest, (1,2) : largest.

by observing pairwise CI (+10)

24. ANOVA Table (+10) $p < 0.05$.



∴ smallest : (1,2,3), largest : (4,5).

by observing pairwise CI (+10)

Importing packages, modules and data

In [1]:
```
import pandas as pd
df1 = pd.read_excel('data/ds11.1.4-keyboard-layout-designs.xls')
df2 = pd.read_excel('data/ds11.1.7-oneway-layout-3.xls')
df3 = pd.read_excel('data/ds11.1.9-ecoli-colonies-in-riverwater.xls
')
```

In [2]:
```
import scipy.stats as stat
from statsmodels.stats.anova import anova_lm
from statsmodels.formula.api import ols
from statsmodels.stats.multicomp import pairwise_tukeyhsd
```

ANOVA Table for 16

In [3]:
```
aq = df1
newaq = pd.melt(aq).dropna()
model = ols('value ~ C(variable)', newaq).fit()
print(anova_lm(model))
```

```
                df      sum_sq    mean_sq          F        PR(>F)
C(variable)    2.0  121.238197  60.619098  52.83944  1.476166e-10
Residual      30.0   34.416961   1.147232       NaN           NaN
```

CI for 16 ¶

In [4]:
```
comp = pairwise_tukeyhsd(newaq['value'], newaq['variable'], alpha=0
.05)
print(comp)
```

```
  Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
 group1   group2  meandiff p-adj   lower   upper  reject
-----------------------------------------------------
Layout 1 Layout 2    3.985  0.001  2.8546  5.1154   True
Layout 1 Layout 3  -0.3614 0.6879 -1.4633  0.7406  False
Layout 2 Layout 3  -4.3464  0.001 -5.4998 -3.1929   True
-----------------------------------------------------
```

ANOVA Table for 20

In [5]:
```
aq = df2
newaq = pd.melt(aq).dropna()
model = ols('value ~ C(variable)', newaq).fit()
print(anova_lm(model))
```

```
                df   sum_sq  mean_sq          F    PR(>F)
C(variable)    4.0  1102.74  275.685  18.50856  0.000002
Residual      20.0   297.90   14.895        NaN       NaN
```

CI for 20

In [6]:
```
comp = pairwise_tukeyhsd(newaq['value'], newaq['variable'], alpha=0
.05)
print(comp)
```

```
      Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================================
   group1      group2   meandiff p-adj   lower    upper   reject
---------------------------------------------------------------------
Treatment 1 Treatment 2      6.4 0.1039  -0.9046  13.7046  False
Treatment 1 Treatment 3    -13.2  0.001 -20.5046  -5.8954   True
Treatment 1 Treatment 4     -2.6  0.834 -10.3477   5.1477  False
Treatment 1 Treatment 5     -7.1 0.0455 -14.0937  -0.1063   True
Treatment 2 Treatment 3    -19.6  0.001 -26.9046 -12.2954   True
Treatment 2 Treatment 4     -9.0  0.018 -16.7477  -1.2523   True
Treatment 2 Treatment 5    -13.5  0.001 -20.4937  -6.5063   True
Treatment 3 Treatment 4     10.6 0.0046   2.8523  18.3477   True
Treatment 3 Treatment 5      6.1 0.1063  -0.8937  13.0937  False
Treatment 4 Treatment 5     -4.5 0.3988 -11.9553   2.9553  False
---------------------------------------------------------------------
```

ANOVA Table for 24

In [7]:
```
aq = df3
newaq = pd.melt(aq).dropna()
model = ols('value ~ C(variable)', newaq).fit()
print(anova_lm(model))
```

```
                df  sum_sq  mean_sq          F    PR(>F)
C(variable)    4.0   596.3  149.075  24.438525  0.000002
Residual      15.0    91.5    6.100        NaN       NaN
```

CI for 24

In [8]:
```
comp = pairwise_tukeyhsd(newaq['value'], newaq['variable'], alpha=0
.05)
print(comp)
```

```
    Multiple Comparison of Means - Tukey HSD, FWER=0.05
=============================================================
  group1     group2   meandiff p-adj   lower   upper  reject
-------------------------------------------------------------
Location 1 Location 2    -0.25    0.9 -5.6435  5.1435  False
Location 1 Location 3    -0.25    0.9 -5.6435  5.1435  False
Location 1 Location 4      8.0 0.0028  2.6065 13.3935   True
Location 1 Location 5     13.0  0.001  7.6065 18.3935   True
Location 2 Location 3      0.0    0.9 -5.3935  5.3935  False
Location 2 Location 4     8.25 0.0022  2.8565 13.6435   True
Location 2 Location 5    13.25  0.001  7.8565 18.6435   True
Location 3 Location 4     8.25 0.0022  2.8565 13.6435   True
Location 3 Location 5    13.25  0.001  7.8565 18.6435   True
Location 4 Location 5      5.0 0.0757 -0.3935 10.3935  False
-------------------------------------------------------------
```