STAT 420

1. Six samples of each of four types of cereal grain grown in a certain region were analyzed to determine thiamin content, resulting in the following data (µg/g):

Wheat	5.2	4.5	6.0	6.1	6.7	5.7
Barley	6.5	8.0	6.1	7.5	5.9	5.6
Maize	5.8	4.7	6.4	4.9	6.0	5.2
Oats	8.3	6.1	7.8	7.0	5.6	7.2

	n_i	$\overline{x}_i$	s_i	s_i^2
Wheat	6	5.7	0.7668	0.588
Barley	6	6.6	0.9508	0.904
Maize	6	5.5	0.6693	0.448
Oats	6	7.0	1.0139	1.028

Source	**SS**	**DF**	**MS**	**F**
Between	9.24	3	3.08	4.151
Within	14.84	20	0.742	
Total	24.08	23		

$F_{0.05}(3, 20) = 3.10$

$t_{0.025}(20) = 2.086$

A $100 \times (1 - \gamma)$-percent confidence interval the difference $\mu_i - \mu_j$, $i \neq j$, is given by

$$\overline{Y}_i - \overline{Y}_j \pm t_{\gamma/2}(N - J \text{ d.f.}) \cdot s_{pooled} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

where $s_{pooled} = \sqrt{\text{MSW}}$.

a) Construct a 95% confidence interval for the difference between the average thiamin content for Oats and Maize.

$$(7.0 - 5.5) \pm 2.086 \cdot \sqrt{0.742} \cdot \sqrt{\frac{1}{6} + \frac{1}{6}}$$

$\mathbf{1.5 \pm 1.0374}$

<u>Tukey's pairwise comparison</u>:

With $100\times(1-\gamma)$-percent confidence *all* pairwise differences $\mu_i - \mu_j$ are bracketed by the bounds

$$\left(\overline{Y}_i - \overline{Y}_j\right) \pm \frac{q_{\gamma,J,N-J}}{\sqrt{2}} \cdot s_{pooled} \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

where $s_{pooled} = \sqrt{\text{MSW}}$,

$q_{\gamma,J,N-J}$ = values from Studentized Range table.

b) Use a 95% confidence level and Tukey's pairwise comparison procedure to compare the average thiamin content for Oats with the average thiamin content for Maize.

$q_{0.05,4,20} = 3.96.$

$$(7.0 - 5.5) \pm \frac{3.96}{\sqrt{2}} \cdot \sqrt{0.742} \cdot \sqrt{\frac{1}{6} + \frac{1}{6}}$$

1.5 ± 1.3926

```
> qtukey(0.95,4,20)
[1] 3.958293
>
> results = glm(Thiamin ~ factor(Grain))
>
> TukeyHSD(aov(results))
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = results)

$`factor(Grain)`
             diff        lwr        upr     p adj
Maize-Barley -1.1 -2.4919842 0.29198423 0.1541415
Oats-Barley   0.4 -0.9919842 1.79198423 0.8516188
Wheat-Barley -0.9 -2.2919842 0.49198423 0.2980981
Oats-Maize    1.5  0.1080158 2.89198423 0.0318296
Wheat-Maize   0.2 -1.1919842 1.59198423 0.9773911
Wheat-Oats   -1.3 -2.6919842 0.09198423 0.0724920
```

`TukeyHSD` – Tukey's Honest Significant Difference

Contrast in the means $\mu_1, \mu_2, \ldots, \mu_J$

$$c_1 \mu_1 + c_2 \mu_2 + \ldots + c_J \mu_J = \sum_{j=1}^{J} c_j \mu_j \qquad \text{where } \sum_{j=1}^{J} c_j = 0$$

<u>Scheffé's multiple comparison</u>:

With $100 \times (1 - \gamma)$-percent confidence *all* contrasts in the J population means of the form $\sum_{j=1}^{J} c_j \mu_j$ are bracketed by the bounds

$$\sum_{j=1}^{J} c_j \overline{Y}_j \pm \sqrt{F_\gamma (J-1, N-J)} \cdot s_{pooled} \cdot \sqrt{(J-1) \cdot \sum_{j=1}^{J} \frac{c_j^2}{n_j}}$$

where $s_{pooled} = \sqrt{\text{MSW}}$.

c) Use a 95% confidence level and Scheffé's multiple comparison procedure to compare …

(i) the average thiamin content for Oats with the average thiamin content for Maize;

$c_W = 0, \qquad c_B = 0, \qquad c_M = -1, \qquad c_O = 1.$

$$(7.0 - 5.5) \pm \sqrt{3.10} \cdot \sqrt{0.742} \cdot \sqrt{3 \cdot \left(\frac{1}{6} + \frac{1}{6} \right)} \qquad \mathbf{1.5 \pm 1.51664}$$

(ii) the average thiamin content for Oats and Barley with the average thiamin content for Maize;

$c_W = 0, \qquad c_B = \mathbf{\tfrac{1}{2}}, \qquad c_M = -1, \qquad c_O = \mathbf{\tfrac{1}{2}}.$

$$\left(\frac{6.6 + 7.0}{2} - 5.5 \right) \pm \sqrt{3.10} \cdot \sqrt{0.742} \cdot \sqrt{3 \cdot \left(\frac{1}{24} + \frac{1}{6} + \frac{1}{24} \right)} \qquad \mathbf{1.3 \pm 1.31345}$$

(iii) the average thiamin content for Oats and Barley with the average thiamin content for Maize and Wheat;

$c_W = -\tfrac{1}{2}, \quad c_B = \tfrac{1}{2}, \qquad c_M = -\tfrac{1}{2}, \quad c_O = \tfrac{1}{2}.$

$$\left(\frac{6.6+7.0}{2} - \frac{5.7+5.5}{2}\right) \pm \sqrt{3.10}\cdot\sqrt{0.742}\cdot\sqrt{3\cdot\left(\frac{1}{24}+\frac{1}{24}+\frac{1}{24}+\frac{1}{24}\right)} \qquad \mathbf{1.2 \pm 1.0724}$$

==

Bonferroni method:

To make m confidence intervals with *simultaneous* confidence level of at least $(1-\gamma)\times 100\,\%$, use $(1-\frac{\gamma}{m})\times 100\,\%$ for the individual confidence level.

We have 6 pairwise differences:

$$\mu_1-\mu_2 \qquad \mu_1-\mu_3 \qquad \mu_1-\mu_4 \qquad \mu_2-\mu_3 \qquad \mu_2-\mu_4 \qquad \mu_3-\mu_4$$

To make 6 confidence intervals with *simultaneous* confidence level of at least 95%,

$$\frac{\gamma}{m} = \frac{0.05}{6}, \qquad \frac{\gamma/m}{2} = \frac{0.05}{12}.$$

```
> qt(1-0.05/12,20)
[1] 2.927119
```

EXCEL: =TINV(0.05/6,20)

Margin of error: $\pm 2.927119\cdot\sqrt{0.742}\cdot\sqrt{\frac{1}{6}+\frac{1}{6}}$ $\quad$ $\mathbf{\pm 1.4557}$

```
Maize-Barley    -1.1  ±  1.4557
Oats-Barley      0.4  ±  1.4557
Wheat-Barley    -0.9  ±  1.4557
Oats-Maize       1.5  ±  1.4557
Wheat-Maize      0.2  ±  1.4557
Wheat-Oats      -1.3  ±  1.4557
```

Kruskal-Wallis test for equivalence of means:

Let $f(x)$ be a density of a continuous random variable with mean 0.

Assume Y_{ij}, $i = 1, 2, \ldots, n_j$, $j = 1, 2, \ldots, J$, are independent random variables with density $f(x - \mu_j)$.

(The J populations have no parametric assumptions, they are assumed to have densities with a common shape, but perhaps different centers.)

H_0: $\mu_1 = \mu_2 = \ldots = \mu_J$ $\qquad$ H_1: not all of the μ_j are equal.
$\mu_i \neq \mu_j$ for at least one pair i and j.

Let r_{ij} be the respective rank of a data point when all the data is ranked from smallest to largest.

Let $\bar{r}_j$ be the mean of the ranks for each group. Let $\bar{r} = \dfrac{N+1}{2}$ be the grand mean of the ranks.

Test statistic:

$$K = \frac{12}{N(N+1)} \sum_{j=1}^{J} n_j \left(\bar{r}_j - \bar{r}\right)^2 = \frac{12}{N(N+1)} \sum_{j=1}^{J} n_j \left(\bar{r}_j - \frac{N+1}{2}\right)^2 = \frac{12}{N(N+1)} \sum_{j=1}^{J} n_j \bar{r}_j^{\,2} - 3(N+1).$$

Reject H_0 if $K > \chi^2_\alpha (J - 1)$.

1. Six samples of each of four types of cereal grain grown in a certain region were analyzed to determine thiamin content, resulting in the following data (μg/g):

Wheat	5.2	4.5	6.0	6.1	6.7	5.7
Barley	6.5	8.0	6.1	7.5	5.9	5.6
Maize	5.8	4.7	6.4	4.9	6.0	5.2
Oats	8.3	6.1	7.8	7.0	5.6	7.2

Does this data suggest that at least two of the grains differ with respect to true average thiamin content? Use $\alpha = 0.05$.

W	M	M	W	M	B	O	W	M	B	W	M
4.5	4.7	4.9	5.2	5.2	5.6	5.6	5.7	5.8	5.9	6.0	6.0
1	2	3	4.5	4.5	6.5	6.5	8	9	10	11.5	11.5

W	B	O	M	B	W	O	O	B	O	B	O
6.1	6.1	6.1	6.4	6.5	6.7	7.0	7.2	7.5	7.8	8.0	8.3
14	14	14	16	17	18	19	20	21	22	23	24

Wheat $\bar{r}_{W} = 9.5$ **Maize** $\bar{r}_{M} = 7.6666\overline{6}$

Barley $\bar{r}_{B} = 15.25$ Oats $\bar{r}_{O} = 17.5833\overline{3}$ $\bar{r} = 12.5$

$$K = \frac{12}{24 \cdot 25}\left[6\cdot(9.5-12.5)^2 + 6\cdot(15.25-12.5)^2 + 6\cdot\left(7.6666\overline{6}-12.5\right)^2 + 6\cdot\left(17.5833\overline{3}-12.5\right)^2 \right] = 7.89166\overline{6}.$$

A correction for ties can be made, but this correction usually makes little difference in the value of K unless there are a large number of ties.

$\chi^2_{\alpha}(J-1) = \chi^2_{0.05}(3) = 7.815.$ $K > 7.815.$ **Reject $\mathbf{H_0}$** at $\alpha = 0.05$.

```
> 1-pchisq(7.891667,3)
[1] 0.04830449
```

=CHIDIST(7.891667,3) p-value $\approx 0.0483 < 0.05$.

```
> Wheat <- c(5.2,4.5,6.0,6.1,6.7,5.7)
> Barley <- c(6.5,8.0,6.1,7.5,5.9,5.6)
> Maize <- c(5.8,4.7,6.4,4.9,6.0,5.2)
> Oats <- c(8.3,6.1,7.8,7.0,5.6,7.2)
>
> Grain <- c(rep("Wheat",6),rep("Barley",6),rep("Maize",6),rep("Oats",6))
> Thiamin <- c(Wheat,Barley,Maize,Oats)
>
> kruskal.test(Thiamin ~ factor(Grain))

        Kruskal-Wallis rank sum test

data:  Thiamin by factor(Grain)
Kruskal-Wallis chi-squared = 7.9158, df = 3, p-value = 0.04779
```