- <u>Simplified procedure</u>.
  - Rank all obs from the smallest to the largest
  - $H = \frac{12}{N(N+1)} \sum_{i=1}^{g} n_i (\bar{R}_i - \frac{N+1}{2})^2$, where $\bar{R}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} R_{ij}$

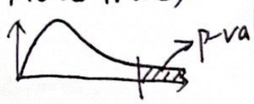
    If no ties, this is the T.S.; otherwise, proceed.
  - $D = 1 - \frac{\sum (t^3 - t)}{(N-1)N(N+1)}$, where t is the number of ties for each rank value.

    $\Rightarrow H_{adj} = \frac{H}{D}$

- p-value: Under the null, the T.S. $\sim \chi^2_{g-1}$, i.e., χ^2 distribution w/ df ~~#gro~~ = #group - 1.

  (Recall in ANOVA, T.S. $\sim F_{g-1, N-g}$ if H_0 is true)

  so $P(Y \geq T.S.)$ where $Y \sim \chi^2_{g-1}$.



- conclusions: (As usual)

  Not enough / some / (pretty) strong / very strong evidence to say <u>Ha is true</u> ("actual")

  (Replace it w/ the desired statement in English)

$N = \cancel{}\ 15$

$g = 3$ groups   Quant var = wt loss

Example: A pharmaceuticals company is developing a new appetite suppressant. 15 random selected lab mice are given 3 different treatments (Phentermine, Cathine, and Benfluorex) and their weight losses are recorded as follows:

Phentermine	Cathine	Benfluorex
2.2 (4)	1.3 (1)	3.2 (12)
1.6 (2)	2.7 (7)	2.9 (9)
3.6 (13.5)	2.9 (9)	3.6 (13.5)
2.4 (5)	1.8 (3)	2.5 (6)
2.9 (9)	3.1 (11)	6.2 (15)

Ordered ranks:

$1,2,3,4,5,6,7,\underbrace{9,9,9}_{t=3},11,12,\underbrace{13.5,13.5}_{t=2},15$

Check if $W_P + W_C + W_B = \frac{(N+1)N}{2}$.

- Which statistical inference procedures could we use on this data? Which one should we use?

Could: ANOVA, K-W

Should: Examine assumptions.

K-W preferred

KW: • SRS (both need it)

• min 5 obs per group ✓

ANOVA: • SRS

• normal distribution of res var per group

• equal variances

6.2 is an outlier, so not normal. and ~~~~ prolly no equal variances.

- Using the output below, conduct the test and interpret the results.

△ H_0: All three trts have the same median (or distribution) of weight loss in population.

H_a: Not all of the population medians are the same.

△ TS: • Rank from the smallest to the largest (shown in the table)

• $\bar{R}_1 = \frac{4+2+13.5+5+9}{5} = 6.7$, $\bar{R}_2 = 6.2$, $\bar{R}_3 = 11.1$, $\bar{R} = \frac{N+1}{2} = 8$

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{g} n_i(\bar{R}_i - \bar{R})^2 = \frac{12}{15 \times 16} \times 5 \times \left[(6.7-8)^2 + (6.2-8)^2 + (11.1-8)^2\right]$$

$= 3.635$

• Now adjust for ties. $\Sigma(t^3 - t) = (3^3 - 3) + (2^3 - 2) = 24 + 6 = 30$

$$D = 1 - \frac{\Sigma(t^3-t)}{(N-1)N(N+1)} = 1 - \frac{30}{14 \times 15 \times 16} = 0.9911$$

Kruskal-Wallis Test: Weight Loss versus Drug

```
Kruskal-Wallis Test on Weight Loss

Drug          N  Median  Ave Rank      Z
Benfluorex    5   3.200      11.1   1.90
Cathine       5   2.700       6.2  -1.10
Phentermine   5   2.400       6.7  -0.80
Overall      15               8.0

H = 3.63  DF = 2  P = 0.162
H = 3.67  DF = 2  P = 0.160  (adjusted for ties)
```

$$H_{adj} = \frac{H}{D} = \frac{3.635}{0.9911} = 3.67$$

$Y \sim \chi^2_{3-1} = \chi^2_2$

△ p-value $= P(Y \geq 3.67) = 0.16$

△ Conclusions: We do not have enough evidence to say that there is some difference between the 3 trts in terms of the median/distribution of the weight loss in population (of lab mice).

Example: Compare 4 teaching techniques - say, Method 1, Method 2, Method 3, and Method 4. Different students are assigned at random to different methods. The exam scores for the students are given below, with the ranks in parentheses:

Method 1	Method 2	Method 3	Method 4
65 (3)	72 (7.5)	59 (1)	94 (23)
87 (19)	69 (5.5)	78 (11)	89 (21)
73 (9)	83 (17.5)	67 (4)	80 (14)
79 (12.5)	81 (15.5)	62 (2)	88 (20)
81 (15.5)	72 (7.5)	83 (17.5)	
69 (5.5)	79 (12.5)	76 (10)	
	90 (22)		

Resp var – exam scores (Quant)

$g=4$ groups, indep

$n_1=6,\ n_2=7,\ n_3=6,\ n_4=4$

$N=6+7+6+4=23$

- Which statistical inference procedures could we use on this data? Which one should we use?

ANOVA
- SRS
- normal distribution } looks
- equal variances } OK → can compute s.d. for each group.

K-W
- SRS
- 5 obs per group ⊗ Method 4 has 4 ppl.

- Using the output below, conduct the test and interpret the results.

K-W: △ Ho: No diff in median/distribution of exam scores for these 4 teaching methods.

Ha: some diff.

OR Ho: $\eta_1=\eta_2=\eta_3=\eta_4$ Ha: Not all of the η_i's are the same.

△ TS: • $\bar{R}_1=\frac{3+19+9+12.5+15.5+5.5}{6}=10.75,\ \bar{R}_2=12.57,\ \bar{R}_3=7.58,\ \bar{R}_4=19.5$

$\bar{R}=\frac{23+1}{2}=12$

$H=\frac{12}{23\times24}\times\left[6\times(10.75-12)^2+7\times(12.57-12)^2+6\times(7.58-12)^2+4\times(19.5-12)^2\right]$

$=7.69$

• Ordered ranks: 1, 2, 3, 4, 5.5, 5.5, 7.5, 7.5, 9, 10, 11, 12.5, 12.5, 14, 15.5, 15.5, 17.5, 17.5, 19, 20, 21, 22, 23

$\sum(t^3-t)=5\times(2^3-2)=30$

$D=1-\frac{30}{22\times23\times24}=0.9975$

$H_{adj}=\frac{H}{D}=\frac{7.69}{0.9975}=7.71$

△ p-val $=P(Y\geq7.71)=0.052$

$Y\sim\chi^2_{4-1}=\chi^2_3$

Kruskal-Wallis Test: Grade versus Method

```
Method     N  Median  Ave Rank      Z
1          6   76.00      10.6  -0.60
2          7   79.00      12.7   0.33
3          6   71.50       7.6  -1.86
4          4   88.50      19.5   2.43
Overall   23              12.0

H = 7.78  DF = 3  P = 0.051
H = 7.79  DF = 3  P = 0.051  (adjusted for ties)

NOTE * One or more small samples
```

See canvas for correct output

△ Conclusions: some evidence of some difference . . .

△ Follow-up: Wilcoxon rank sum tests to compare methods (1-2, 1-3, 1-4, 2-3, 2-4, 3-4)

- Using the output below, conduct the One-Way ANOVA test and interpret the results.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ H_a: some diff

TS: $F = 3.77$

p-val: 0.028

Conclusions: Pretty strong evidence of some diff in avg exam scores

Follow-up: Bonferroni, Tukey, Fisher, Individual CIs.

One-way ANOVA: Grade versus Method

```
Source  DF      SS     MS     F      P
Method   3   712.6  237.5  3.77  0.028
Error   19  1196.6   63.0
Total   22  1909.2

S = 7.936   R-Sq = 37.32%   R-Sq(adj) = 27.43%

                               Individual 95% CIs For Mean Based on
                               Pooled StDev
Level  N    Mean  StDev  ------+---------+---------+---------+---
1      6  75.667  8.165       (------*-----)
2      7  78.429  7.115          (-----*------)
3      6  70.833  9.579  (------*------)
4      4  87.750  5.795                 (--------*-------)
                         ------+---------+---------+---------+---
                              70        80        90       100
```

- Compare the Kruskal-Wallis *H*-test and the one-way ANOVA test. Which one is preferred in this scenario?

ANOVA – assumptions were ALL satisfied

also ANOVA is more powerful