Comparisons Among Several Samples

KRUSKAL-WALLIS “NONPARAMETRIC ANOVA”

WELCH’S ANOVA

Rank-Sum Test: Review

Rank-Sum Test: Discussion and Assumptions

- No distributional assumptions and resistant to outliers
- When t-test assumptions are met, the rank-sum test performs about 95.49% as well
- Performs arbitrarily better if the t-test assumptions are not (approximately) met
- Works well with **ORDINAL** data

(Realistically required for t-tools)

(**NOMINAL:** order is arbitrary. **ORDINAL:** order matters. **INTERVAL:** subtraction is meaningful. **RATIO:** multiplication is meaningful)

- Works with censored values

(Censored means that the actual value was too large/small to

be accurately recorded)

- It still requires some assumptions:
    1. All observations are independent
    2. The *Y* values are ordinal

59 patients with arthritis who participated in a clinical trial were assigned to two groups, active and placebo. The response status:
(excellent=5, good=4, moderate=3, fair=2, poor=1)
of each patient was recorded.

Rank-Sum Test: Hypotheses

For the rank-sum test, our null hypothesis is in terms of **DISTRIBUTIONS** instead of means

H_0: The **DISTRIBUTION** of the "new" method scores is the same as the **DISTRIBUTION** of the "traditional" method scores

The Alternative Hypotheses:

H_A: The **DISTRIBUTION** of the "new" method scores is different from the **DISTRIBUTION** of the "traditional" method scores **(TWO SIDED)**

H_A: The **DISTRIBUTION** of the "new" method scores is larger than the **DISTRIBUTION** of the "traditional" method scores **(ONE SIDED)**

Note: "larger than" can be interpreted as "systematically higher than" in the sense that the probability of getting any value from one distribution is larger than for the other distribution

Kruskal-Wallis Test

Kruskal-Wallis test

- When there are/is…
  - extreme values/outliers
  - Small or very unequal sample sizes + lack of normality
  - strong evidence of unequal variances between groups

… a nonparametric version of ANOVA is preferable

Though there are various tests for detecting unequal variances, it is best practice to evaluate this assumption visually

Kruskal-Wallis test: Assumptions

- Resistant to outliers
- Works well with **ORDINAL** data (this is the main application in practice)

(**NOMINAL:** order is arbitrary. **ORDINAL:** order matters. **INTERVAL:** subtraction is meaningful. **RATIO:** multiplication is meaningful)

- Can work with censored values

(Censored means that the actual value was too large/small to be accurately recorded)

- It still requires some assumptions:
  1. All observations are independent
  2. The groups are independent
  3. The Y values are ordinal

- Under these assumptions, we can perform the following hypothesis tests

Kruskal-Wallis test: Hypotheses

For the Kruskal-Wallis test, the null hypothesis is in terms of **DISTRIBUTIONS** instead of means

H_0: The **DISTRIBUTION** of the groups are all the same

Under the null, we can immediately replace the observations with ranks from 1, …, n

(Here, n = $n_1 + \ n_2 + \cdots + n_I$)

The sum of the ranks for each of the I groups divided by n_i is a natural test statistic

Like the rank-sum test, the alternative hypothesis is a bit nuanced

Kruskal-Wallis test: P-values

For the Kruskal-Wallis test, the null hypothesis is in terms of **DISTRIBUTIONS** instead of means

H_0: The **DISTRIBUTION** of the groups are all the same

H_A: The **DISTRIBUTIONS** of the groups are not all the same

There are two main ways in which p-values are commonly reported:

- A normal approximation (which actually goes by the name of a 'Chi-Squared' distribution)
  - (We will return to what the 'Chi-Squared' distribution is next semester)
- An exact p-value
  - (The actual sampling distribution of the ranks is every combination of assigning observations to the groups)

Kruskal-Wallis test: Spock data

```
PROC GLM DATA = spock2;
    CLASS judge;
    MODEL percFemale = judge;
RUN;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1927.080865	321.180144	6.72	<.0001
Error	39	1864.445222	47.806288		
Corrected Total	45	3791.526087			

Kruskal-Wallis Test	
Chi-Square	21.9647
DF	6
Pr > Chi-Square	0.0012

(Normal approximation:
Requires that $n_i > 5$ for each i)

```
PROC NPAR1WAY DATA = spock WILCOXON;
    CLASS judge;
    VAR percFemale;
    EXACT WILCOXON / MC;
RUN;
```

Monte Carlo Estimate for the Exact Test	
Pr >= Chi-Square	
Estimate	0.0003
99% Lower Conf Limit	<.0001
99% Upper Conf Limit	0.0007
Number of Samples	10000
Initial Seed	56388143

(Approximates the exact distribution with 10000 re-randomizations)

Kruskal-Wallis test: P-values

If we add the additional assumption that the **SHAPE OF THE DISTRIBUTIONS** are the same, then..

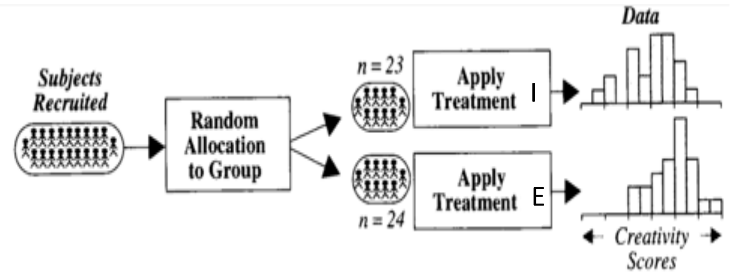
H_0: The **MEDIANS** of the groups are all the same

H_A: The **MEDIANS** of the groups are not all the same

Now, if you reject H_0, you can conclude that at least two groups have different medians

Welch’s T-Tools

Creativity Study: Reminder



→ Population mean: μ_I
→ Population sd: σ_I

→ Population mean: μ_E
→ Population sd: σ_E

- We additionally need to know/estimate the standard deviation of $\bar{Y}_I - \bar{Y}_E$
- There are two ways mentioned in the book
  1. Pooled SD
  2. Welch's SD
- To create the pooled SD, we need to assume that $\sigma_I = \sigma_E$
- Then, we can form an estimate of this common standard deviation via
- $s_p = \sqrt{\frac{(n_I - 1)\, s_I^2 + (n_E - 1)\, s_E^2}{n_I + n_E - 2}}$
- $SE(\bar{Y}_I - \bar{Y}_E) = \sqrt{\frac{\sigma_I^2}{n_I} + \frac{\sigma_E^2}{n_E}} \leftrightarrow SE(\bar{Y}_I - \bar{Y}_E) = s_p\sqrt{\frac{1}{n_I} + \frac{1}{n_E}}$

What if this assumption isn't true?

Welch's t-Test

The only differences between Welch's t-Test and the "pooled" t-test are:

- The standard error ($SE(\bar{Y}_I - \bar{Y}_E)$)
- The degrees of freedom (df)

(The new degrees of freedom are formed via a Satterthwaite approximation)

Luckily, we already know how to get the output from a Welch's t-Test: PROC TTEST

Welch’s ANOVA

Welch’s ANOVA

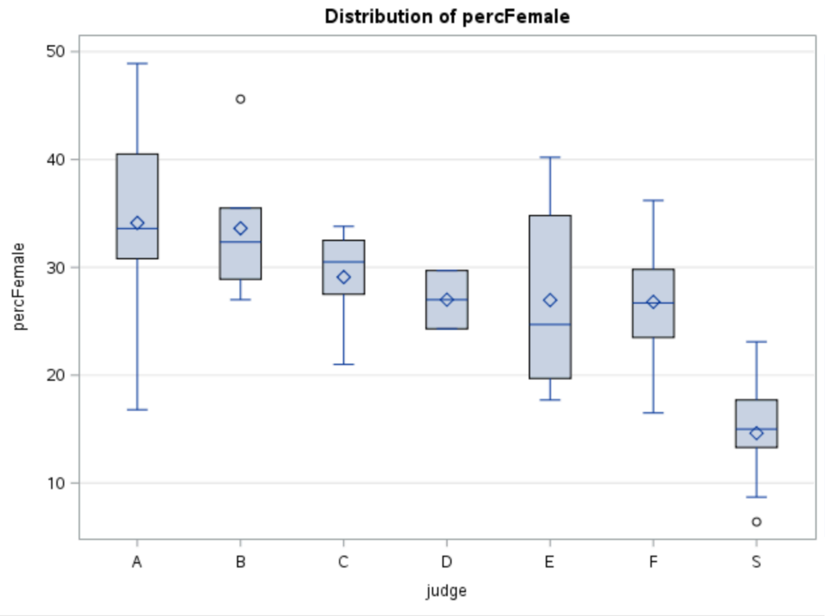
Just like with Welch’s t-test, we can test the hypothesis

$H_0: \mu, \mu, \dots, \mu$
$H_A:\ \mu_S, \mu_A, \mu_B, \dots, \mu_F$

without the assumption of equal variances:

```
PROC GLM DATA = spock;
    CLASS judge;
    MODEL percFemale = judge;
    MEANS judge / WELCH;
RUN;
```

Welch's ANOVA for percFemale			
Source	DF	F Value	Pr > F
judge	6.0000	7.49	0.0031
Error	9.9246		

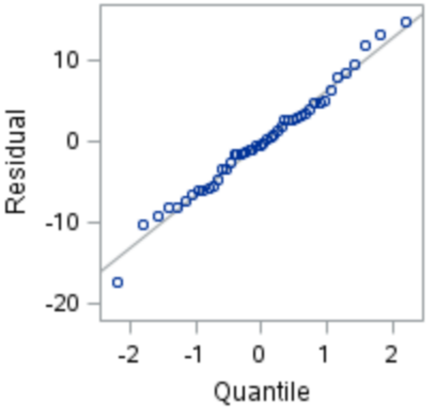


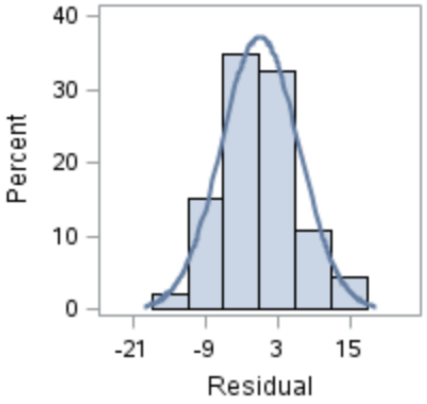
Level of judge	N	percFemale Mean	percFemale Std Dev
A	5	34.1200000	11.9418173
B	6	33.6166667	6.5822235
C	9	29.1000000	4.5929293
D	2	27.0000000	3.8183766
E	6	26.9666667	9.0101424
F	9	26.8000000	5.9688776
S	9	14.6222222	5.0387939

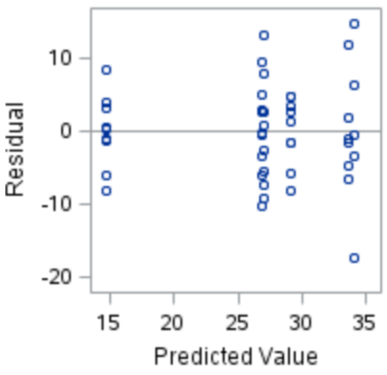
Additional ANOVA assumption checks

Additional ANOVA assumption checks

```
ods graphics on;
PROC GLM DATA = spock plot=diagnostics;
    CLASS judge;
    MODEL percFemale = judge;
RUN;
ods graphics off;
```







(This is only some of the output)

```
PROC GLM DATA = spock;
    CLASS judge;
    MODEL percFemale = judge;
    MEANS judge / HOVTEST;
RUN;
```

H_0: Equal variances

H_A: At least two variances unequal

Levene's Test for Homogeneity of percFemale Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
judge	5	39361.8	7872.4	2.30	0.0641
Error	38	130130	3424.5		

(Note: it considered best practice to use visual checks for the variance assumption instead of hypothesis tests)