Do not use the Two sample-\(t\)-test any more!

Dieter Rasch\(^1\); Klaus, D. Kubinger\(^2\) & Karl Moder\(^1\)


\(^1\)University of Natural Resources and Applied Life Sciences, Vienna, Institute of Applied Statistics and Computing, E-mail: karl.moder@boku.ac.at, dieter.rasch@boku.ac.at

\(^2\)University of Vienna, Faculty of Psychology, Division of Psychological Assessment and Applied Psychometrics, E-mail: klaus.kubinger@univie.ac.at
Do not use the Two sample-t-test any more!

Given two random variables with expectations $E(y_1) = \mu_1; E(y_2) = \mu_2$ and existing fourth moments. We have to test the null hypothesis:

$$H_0: \mu_1 = \mu_2$$

against a two-sided alternative hypothesis:

$$H_A: \mu_1 \neq \mu_2$$
Problem

Given two random variables with expectations

\[ E(y_1) = \mu_1; \quad E(y_2) = \mu_2 \]

and existing fourth moments.
Given two random variables with expectations

\[ E(y_1) = \mu_1; \quad E(y_2) = \mu_2 \]

and existing fourth moments.

We have to test the null hypothesis:

\[ H_0 : \mu_1 = \mu_2 \]
Problem

Given two random variables with expectations

\[ E(y_1) = \mu_1; \quad E(y_2) = \mu_2 \]

and existing fourth moments.

We have to test the null hypothesis:

\[ H_0 : \mu_1 = \mu_2 \]

against a two-sided alternative hypothesis:

\[ H_A : \mu_1 \neq \mu_2 \]
Given 2 independent random samples

\[ y_i^T = (y_{i1}, y_{i2}, \ldots, y_{in_i}), \quad n_i; \quad i = 1, 2 \]

distributed as \( N(\vec{\mu}_i; \sigma^2 I_{n_i}); \quad \sigma^2 > 0 \)

Then

\[ s^2 = \frac{SQ_{y1} + SQ_{y2}}{n_1 + n_2 - 2} \]

is the pooled estimator of \( \sigma^2 \).
From this it follows that

\[ t = \frac{\bar{y}_1 - \bar{y}_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \]

is distributed as

\[ t \left( n_1 + n_2 - 2; \lambda = \frac{\mu_1 - \mu_2}{\sigma} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \right). \]
From this it follows that

\[ t = \frac{\bar{y}_1 - \bar{y}_2}{s} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \]

is distributed as

\[ t \left( n_1 + n_2 - 2; \lambda = \frac{\mu_1 - \mu_2}{\sigma} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \right). \]

A uniformly best unbiased \( \alpha \)-test for all \( 0 < \alpha < 1 \) is

\[ k \left( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) = \begin{cases} 1, & \text{for } |t| > t(n_1 + n_2 - 2; 1 - \frac{\alpha}{2}) \\ 0, & \text{otherwise} \end{cases} \]
In the case of unequal variances, Welch (1947) and Trickett and Welch (1954) proposed an approximate test based on

\[ t^* = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \]

The test is:

\[ k(y_1, y_2) = 1, \quad \text{for } |t^*| > t_{f_0 - \alpha/2} \]

with:

\[ f = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{s_1^4(n_1 - 1) + s_2^4(n_2 - 1)} \]
In the case of unequal variances Welch (1947) and Trickett and Welch (1954) proposed an approximate test based on

\[ t^* = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]
Welch-Test

In the case of unequal variances Welch (1947) and Trickett and Welch (1954) proposed an approximate test based on

\[ t^* = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \]

The test is:

\[ k \left( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) = \begin{cases} 1, & \text{for } |t^*| > (f; 1 - \frac{\alpha}{2}) \\ 0, & \text{otherwise} \end{cases} \]
Welch-Test

In the case of unequal variances Welch(1947) and Trickett and Welch(1954) proposed an approximate test based on

\[ t^* = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]

The test is:

\[ k\left(\begin{array}{c} y_1 \\ y_2 \end{array}\right) = \begin{cases} 1, & \text{for } |t^*| > (f; 1 - \frac{\alpha}{2}) \\ 0, & \text{otherwise} \end{cases} \]

with:

\[ f = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{(n_1-1)n_1^2} + \frac{s_2^4}{(n_2-1)n_2^2}} \]
Hence, many theoretical statisticians nowadays do not recommend pre-testing (see Moser & Stevens, 1992), as concerns testing variance homogeneity, Easterling & Anderson, 1978, and Schucany & Ng, 2006, for testing normal distribution. Nevertheless in applied statistics pre-testing is often applied.
Unfortunately, statistical program packages, lecture notes and applied statistical text books still recommend a pre-test at least on variance homogeneity in the two-sample location problem. If we google for "variance homogeneity test" (24th Sept, 2008) a note is as follows:

changingminds.org/explanations/research/analysis/variance_homogeneity.htm

At the latter link, note that the $F$-test (later in the text alternatively the Levene-test and Mauchly’s test) is recommended as a pre-test in the package XLSTAT. If the $F$-value is small enough (a table of critical values is given), then it is considered safe to use the $t$-test.
Un fortunately, statistical program packages, lecture notes and applied statistical text books still recommend a pre-test at least on variance homogeneity in the two-sample location problem. If we google for ”variance homogeneity test” (24th Sept, 2008) a note is as follows:

**Variance homogeneity test** Here is a simple test for the homogeneity of variances, as required in several statistical tests.
changingminds.org/explanations/research/analysis/variance_homogeneity.htm
Unfortunately, statistical program packages, lecture notes and applied statistical text books still recommend a pre-test at least on variance homogeneity in the two-sample location problem. If we google for "variance homogeneity test" (24th Sept, 2008) a note is as follows:

**Variance homogeneity test** Here is a simple test for the homogeneity of variances, as required in several statistical tests.
changingminds.org/explanations/research/analysis/variance_homogeneity.htm

At the latter link, note that the $F$-test (later in the text alternatively the Levene-test and Mauchly´s test) is recommended as a pre-test in the package XLSTAT. If the $F$-value is small enough (a table of critical values is given), then it is considered safe to use the $t$-test.
Lecture notes and text books are recommended for this topic. Nothing is said about what to do if variances are not equal. But this is done under http://www.sam.sdu.dk/~nks/St2006uk/Variansanalyse-UK.pdf.

There N.K. Sørensen writes: "If these assumptions not are fulfilled we can conduct a Kruskal-Wallis test".
Lecture notes and text books are recommended for this topic. Nothing is said about what to do if variances are not equal. But this is done under
http://www.sam.sdu.dk/~nks/St2006uk/Variansanalyse-UK.pdf.

There N.K. Sørensen writes:
"If these assumptions not are fulfilled we can conduct a Kruskal-Wallis test".

The equivalent of the latter test in our two-sample problem is the Wilcoxon-(Mann-Whitney-) $U$-test (Wilcoxon, 1945; Mann & Whitney, 1947) which, as a matter of fact, assumes equal moments higher than the first one if the location parameters are to be compared.
Lecture notes and text books are recommended for this topic. Nothing is said about what to do if variances are not equal. But this is done under http://www.sam.sdu.dk/~nks/St2006uk/Variansanalyse-UK.pdf.

There N.K. Sørensen writes:
"If these assumptions not are fulfilled we can conduct a Kruskal-Wallis test".

The equivalent of the latter test in our two-sample problem is the Wilcoxon-(Mann-Whitney-) $U$-test (Wilcoxon, 1945; Mann & Whitney, 1947) which, as a matter of fact, assumes equal moments higher than the first one if the location parameters are to be compared.

We found in our Google-search more than 500 notes, and most of them recommend pre-tests as concerns the assumption of variance homogeneity.
Pre-testing means that before the decision between the two hypotheses is made by the test, a researcher tests the assumptions about the distribution using the observations of the random sample(s). Doing so, the overall risk of erroneous decisions is difficult to specify that concerns the tested assumptions and the tested null-hypothesis in question.
Pre-testing means that before the decision between the two hypotheses is made by the test, a researcher tests the assumptions about the distribution using the observations of the random sample(s). Doing so, the overall risk of erroneous decisions is difficult to specify that concerns the tested assumptions and the tested null-hypothesis in question.

Only if consecutive, independent sampling were applied for both kinds of statistical tests (the pre-test(s) on the one side and the test of $H_0$ on the other side), could this overall risk of erroneous decisions be calculated using the multiplication rule of probability theory.
However, for reasons of feasibility, just a single sampling of data occurs, meaning that the pre-test(s) and the main test are applied at the same observations. As a consequence, the over-all risk can - due to the dependency of the different test statistics - difficult be calculated in closed form.
However, for reasons of feasibility, just a single sampling of data occurs, meaning that the pre-test(s) and the main test are applied at the same observations. As a consequence, the over-all risk can - due to the dependency of the different test statistics - difficult be calculated in closed form.

In this paper, we now will estimate the overall risk of erroneous decisions in the two-sample *t*-test problem using simulation experiments.
However, for reasons of feasibility, just a single sampling of data occurs, meaning that the pre-test(s) and the main test are applied at the same observations. As a consequence, the over-all risk can - due to the dependency of the different test statistics - difficult be calculated in closed form.

In this paper, we now will estimate the overall risk of erroneous decisions in the two-sample \( t \)-test problem using simulation experiments.

As a pre-test of normality, we use the Kolmogorov-Smirnov-test (Kolmogorov, 1930; Smirnov, 1939) and as a pre-test of variance homogeneity the Levene-test (Levene, 1960; this because according to Rasch & Guiard, 2004, the \( F \)-test is very sensitive against non-normality and has already been replaced in SPSS).
Considerations about problems with Pre-test

If the same sample (as usual) is used as well for the pre-test as also for the final test we have a sample size problem.

We go back to our two-sample problem.

We have to test the null hypothesis:

\[ H_0 : \mu_1 = \mu_2 \]

against a two-sided alternative hypothesis:

\[ H_A : \mu_1 \neq \mu_2 \]

Do not use the Two sample-\(t\)-test any more!
Considerations about problems with Pre-test

If the same sample (as usual) is used as well for the pre-test as also for the final test we have a sample size problem. We go back to our two-sample problem.
Considerations about problems with Pre-test

If the same sample (as usual) is used as well for the pre-test as also for the final test we have a sample size problem. We go back to our two-sample problem.

We have to test the null hypothesis:

$$H_0 : \mu_1 = \mu_2$$
Considerations about problems with Pre-test

If the same sample (as usual) is used as well for the pre-test as also for the final test we have a sample size problem. We go back to our two-sample problem.

We have to test the null hypothesis:

\[ H_0 : \mu_1 = \mu_2 \]

against a two-sided alternative hypothesis:

\[ H_A : \mu_1 \neq \mu_2 \]
Which sample size for pre-tests?
Which sample size for pre-tests?

Let us assume we like to test all hypotheses with a first kind risk of 0.05 and a second kind risk of 0.1.
Which sample size for pre-tests?

Let us assume we like to test all hypotheses with a first kind risk of 0.05 and a second kind risk of 0.1.

For the Kolmogorov-Smirnov-tests (normality of each distribution) the sample size is difficult to calculate. But we know that it is relatively large → 500.
Sample size for comparing 2 variances

The variances of two normally distributed characters have to be compared.
The experimental design is specified by the risk of first kind (alpha), second kind (beta) and Q for the F-test.
(Q : limits for the ratio of the two variances)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha</td>
<td>0.050</td>
</tr>
<tr>
<td>beta</td>
<td>0.100</td>
</tr>
<tr>
<td>Q</td>
<td>1.500</td>
</tr>
</tbody>
</table>

Test:
- One-sided
- Two-sided

Do not use the Two sample-t-test any more!
Results (sample size for comparing 2 variances)

Decision:
- Comparison of Two Variances
- Both Character Values are Normally Distributed

For the entered precision parameters
\[ \alpha = 0.050 \quad \text{(Two-sided)} \]
\[ \beta = 0.100 \]
and \[ Q = 1.500000 \]
minimal sample sizes of
\[ n(1) = 258 \]
\[ n(2) = 258 \]
are obtained.
Which sample size for pre-tests?

For the $F$-test (equality of variances) the sample size is calculated by CADEMO for a variance ratio of at least 1.5. We need 258 observations in each sample.
Sample size for comparing 2 means

Test of Two Means for Independent Samples and Unknown Variances

Relationship of the Variances
- $\sigma^2(1) = \sigma^2(2)$
- $\sigma^2(1) \neq \sigma^2(2)$
- No Information about the Relationship

Sample Sizes
- Equal
- Unequal

Risks
- $\alpha : 0.05$
- $\beta : 0.1$

Precision Requirement
- $d : 1$

Variances
- Estimate
- Smallest + Largest Value known
- Max. Value
- No Information about the Variances

$s^2(1)$

$s^2(2)$
Results (sample size for comparing 2 means)

Decision:
Two Means Test for Normal Distributions
Independent Sampling + Two-sided Test
Variances have the same Estimates.

For the given risk of first kind \( \alpha = 0.05 \),
risk of second kind \( \beta = 0.100 \),
minimal difference \( d = 1.000 \)
and the estimate for the common variance \( \sigma^2 = 1.000 \),
a minimal sample size of
\[ n = 23 \]
is obtained for each of the two samples.
Summarizing:

per sample we need:
- 500 units for testing normality
- 258 units for testing homogeneity of variances
- 23 units for testing equality of means

What a nonsense!!

Do not use the Two sample-$t$-test any more!

June 24, 2009
Which sample size?

Summarizing:
per sample we need:

---

Do not use the Two sample-\(t\)-test any more!
Which sample size?

Summarizing:
per sample we need:

- **500** units for testing normality
Which sample size?

Summarizing:

per sample we need:

- **500** units for testing normality
- **258** units for testing homogeneity of variances

What a nonsense!!
Which sample size?

Summarizing:
per sample we need:

- **500** units for testing normality
- **258** units for testing homogeneity of variances
- **23** units for testing equality of means

What a nonsense!!
Normal distribution in both populations?
Test-algorithm

- Normal distribution in both populations?
  - no
    - Wilcoxon-Test

Do not use the Two sample-t-test any more!

June 24, 2009
Test-algorithm

1. Normal distribution in both populations?
   - yes
   - no

2. Variances equal?
   - no
      - Wilcoxon-Test
   - yes

Do not use the Two sample-t-test any more!

June 24, 2009 20 / 28
Test-algorithm

Normal distribution in both populations?

- Yes
  - Variances equal?
    - Yes: $t$-Test
    - No: Wilcoxon-Test

- No: Welch-Test

Do not use the Two sample-$t$-test any more!

June 24, 2009
Normal distribution in both populations?

- Yes
  - Variances equal?
    - Yes: t-Test
    - No: approximative Welch-Test
- No: Wilcoxon-Test

Dieter Rasch, Klaus, D. Kubinger & Karl M

Do not use the Two sample-\( t \)-test any more!

June 24, 2009
The risk of the second kind is the most important one in pre-testing. Accepting a wrong null hypothesis leads to the wrong final test.
The risk of the second kind is the most important one in pre-testing. Accepting a wrong null hypothesis leads to the wrong final test.

Simulation experiment:

In the case ”always $t$ – Test” is the actual first kind risk $\alpha_{act}$
The risk of the second kind

The risk of the second kind is the most important one in pre-testing. Accepting a wrong null hypothesis leads to the wrong final test.

Simulation experiment:

In the case "always \( t \) – Test" is the actual first kind risk \( \alpha_{act} \)

- "near to" the nominal one \( \alpha_{nom} \)?

Dieter Rasch\(^1\); Klaus, D. Kubinger\(^2\) & Karl M.

Do not use the Two sample-\( t \)-test any more! June 24, 2009 21 / 28
The risk of the second kind is the most important one in pre-testing. Accepting a wrong null hypothesis leads to the wrong final test.

Simulation experiment:

In the case "always $t$ – Test" is the actual first kind risk $\alpha_{act}$

- "near to" the nominal one $\alpha_{nom}$?
- "near to" means

$$|\alpha_{act} - \alpha_{nom}| \leq 0.2\alpha_{nom}$$
Skewness - Kurtosis

For the standardised 3. (Skewness) and 4. moment -3 (Kurtosis) of any distribution we have:

\[ \gamma_2 \geq \gamma_4 \]
For the standardised 3. (Skewness) and 4. moment -3 (Kurtosis) of any distribution we have:

$$\gamma_2 \geq \gamma_1^2 - 2$$
Skewness - Kurtosis

For the standardised 3. (Skewness) and 4. moment -3 (Kurtosis) of any distribution we have:

\[ \gamma_2 \geq \gamma_1^2 - 2 \]

In the by \( \gamma_2 \geq \gamma_1^2 - 2 \) defined parabola we find all theoretical (and empirical) distributions with a fourth moment.
Skewness - Kurtosis

For the standardised 3. (Skewness) and 4. moment -3 (Kurtosis) of any distribution we have:

$$\gamma_2 \geq \gamma_1^2 - 2$$

In the by defined parabola we find all theoretical (and empirical) distributions with a fourth moment.
Simulation

Simulated distributions
Simulated distributions

- Let $u \sim N(0; 1)$.
Simulated distributions

- Let \( u \sim N(0; 1) \).
- By the transformation

\[
y = a + bu + cu^2 + du^3
\]

we obtain a distribution at each point within the parabola. Fleishman-System, Fleishman, 1978
Simulation

Simulated distributions

- Let $u \sim N(0; 1)$.
- By the transformation

$$y = a + bu + cu^2 + du^3$$

we obtain a distribution at each point within the parabola. Fleishman-System, Fleishman, 1978

Simulation experiment

- Each test was done 100 000 times with simulated data. The relative frequency of rejecting $H_0$ is an estimate of the power function.
We selected:

<table>
<thead>
<tr>
<th>Type</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0,5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>
Simulation - parameter

We selected: and further:

<table>
<thead>
<tr>
<th>Type</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>

\[ \alpha_{nom} = 0.01; 0.05 \text{ and } 0.10 \]
**Simulation - parameter**

We selected:

<table>
<thead>
<tr>
<th>Type</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0,5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>

and further:

- \( \alpha_{nom} = 0,01; 0,05 \)  and  \( 0,10 \)
- \( \delta/\sigma = (\mu_1 - \mu_2)/\sigma = 0; 1; 2; 3; 4 \)  and  \( 5; \sigma_1/\sigma_2 = 1, 2, \ldots, 10. \)
Simulation - parameter

We selected:

<table>
<thead>
<tr>
<th>Type</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>

and further:

\[ \alpha_{nom} = 0.01; 0.05 \text{ and } 0.10 \]
\[ \frac{\delta}{\sigma} = \frac{\mu_1 - \mu_2}{\sigma} = 0; 1; 2; 3; 4 \text{ and } 5; \]
\[ \frac{\sigma_1}{\sigma_2} = 1, 2, \ldots, 10. \]
\[ n_1 = n_2 = 10; 30; 50; 100; \]
## Simulation - parameter

We selected:

<table>
<thead>
<tr>
<th>Type</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>

and further:

- $\alpha_{nom} = 0.01; 0.05$ and $0.10$
- $\delta/\sigma = (\mu_1 - \mu_2)/\sigma = 0; 1; 2; 3; 4$ and $5$;
- $\sigma_1/\sigma_2 = 1, 2, \ldots, 10$;
- $n_1 = n_2 = 10; 30; 50; 100$;
- $n_1 = 10; n_2 = 30$;
Simulation - parameter

We selected: and further:

<table>
<thead>
<tr>
<th>Type</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>( \alpha_{nom} = 0,01; 0,05 ) and 0,10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \delta / \sigma = (\mu_1 - \mu_2) / \sigma = 0; 1; 2; 3; 4 ) and 5;</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>15</td>
<td>( \sigma_1 / \sigma_2 = 1, 2, \ldots, 10. )</td>
</tr>
<tr>
<td>2</td>
<td>0,5</td>
<td>15</td>
<td>( n_1 = n_2 = 10; 30; 50; 100; )</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
<td>( n_1 = 10; n_2 = 30; )</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
<td>( n_1 = 30; n_2 = 10; )</td>
</tr>
</tbody>
</table>
We selected:

<table>
<thead>
<tr>
<th>Type</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>

and further:

\[ \alpha_{nom} = 0.01; 0.05 \text{ and } 0.10 \]

\[ \delta/\sigma = (\mu_1 - \mu_2)/\sigma = 0; 1; 2; 3; 4 \text{ and } 5; \]

\[ \sigma_1/\sigma_2 = 1, 2, \ldots, 10. \]

\[ n_1 = n_2 = 10; 30; 50; 100; \]

\[ n_1 = 10; n_2 = 30; \quad n_1 = 30; n_2 = 10; \]

\[ n_1 = 30; n_2 = 100; \]
We selected:

<table>
<thead>
<tr>
<th>Type</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>

and further:

\[ \alpha_{nom} = 0.01; 0.05 \text{ and } 0.10 \]

\[ \delta / \sigma = (\mu_1 - \mu_2) / \sigma = 0; 1; 2; 3; 4 \text{ and } 5; \]

\[ \sigma_1 / \sigma_2 = 1, 2, \ldots, 10. \]

\[ n_1 = n_2 = 10; 30; 50; 100; \]

\[ n_1 = 10; n_2 = 30; \quad n_1 = 30; n_2 = 10; \]

\[ n_1 = 30; n_2 = 100; \quad n_1 = 100; n_2 = 30; \]
Empirical risk of the 1. kind for $t-$, Wilcoxon-, and Welch-test if $H_0$ is true and $\alpha_{nom} = 0.05$. The ratio of standard deviations for the first and second sample $(\sigma_1, \sigma_2)$ are in ratio 1:2 ($n_1=30$, $n_2=10$).
Empirical risk of the 1. kind for $t$— and Wilcoxon-test with pre-testing if $H_0$ is true and $\alpha_{nom} = 0.05$. The ratio of standard deviations for the first and second sample ($\sigma_1, \sigma_2$) are in ratio 1:2 ($n_1=30$, $n_2=10$).
Conclusions

Results - conclusions

Never do a pre-test.

Choose always the approximate Welch-Test.

The Wilcoxon – Test is useless, if higher moments differ in both populations,

The \textit{t}-test can also not be recommended.

Dieter Rasch\textsuperscript{1}; Klaus, D. Kubinger\textsuperscript{2} & Karl Moder

Do not use the Two sample-\textit{t}-test any more!

June 24, 2009

27 / 28
Never do a pre-test.
Results - conclusions

- Never do a pre-test.
- Choose always the approximate Welch-Test.
Results - conclusions

- Never do a pre-test.
- Choose always the approximate Welch-Test.
- The Wilcoxon – Test is useless, if higher moments differ in both populations,
Results - conclusions

- Never do a pre-test.
- Choose always the approximate Welch-Test.
- The Wilcoxon – Test is useless, if higher moments differ in both populations,
- The t-test can also not be recommended.
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Publication Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspin, A.A.</td>
<td>Tables for use in comparisons whose accuracy involves two variances, separately estimated.</td>
<td>Biometrika 36, 290-296</td>
</tr>
<tr>
<td>Rasch, D. and Guiard, V.</td>
<td>The Robustness of Parametric Statistical Methods.</td>
<td>Psychology Science, 46; 175-208</td>
</tr>
<tr>
<td>Rasch, D.; Teuscher, F. and Guiard, V.</td>
<td>How Robust are tests for two independent samples in case of ordered categorical data?</td>
<td>Journal of Statistical Planning and Inference 133, 2706-2720</td>
</tr>
<tr>
<td>Rasch, D.; Kubinger, K.D. and Moder, K.</td>
<td>The two-sample t-test: pre-testing its assumptions does not pay.</td>
<td>Journal of Statistical Theory and Practice, accepted</td>
</tr>
<tr>
<td>Welch, B.L.</td>
<td>The generalization of Students problem when several different population variances are involved.</td>
<td>Biometrika 34, 28-35</td>
</tr>
</tbody>
</table>