Comparing two groups: many faces of the t-test

Tuan V. Nguyen
Genetics Epidemiology of Osteoporosis Lab
Garvan Institute of Medical Research
A close look at body temperature

- Normal human temperature: 37°C (98.6°F)
- Credited to Wunderlich (19th Century)
- Analysis of temp from 25000 patients
  - Mean: 37°C (98.6°F)
  - Range around: 36.2°C (97.2°F) to 37.5°C (99.5°F)
- Women > Men
- Temp >38°C are always “suspicous”
A Critical Appraisal of 98.6°F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich

Philip A. Mackowiak, MD; Steven S. Wasserman, PhD; Myron M. Levine, MD

Setting.—Inpatient clinical research unit.

Participants.—One hundred forty-eight healthy men and women aged 18 through 40 years.

Main Measurements.—Oral temperatures were measured one to four times daily for 3 consecutive days using an electronic digital thermometer.

Results.—Our findings conflicted with Wunderlich's in that 36.8°C (98.2°F) rather than 37.0°C (98.6°F) was the mean oral temperature of our subjects; 37.7°C (99.9°F) rather than 38.0°C (100.4°F) was the upper limit of the normal temperature range; maximum temperatures, like mean temperatures, varied with time of day; and men and women exhibited comparable thermal variability. Our data corroborated Wunderlich's in that mean temperature varied diurnally, with a 6 AM nadir, a 4 to 6 PM zenith, and a mean amplitude of variability of 0.5°C (0.9°F); women had slightly higher normal temperatures than men; and there was a trend toward higher temperatures among black than among white subjects.

Conclusions.—Thirty-seven degrees centigrade (98.6°F) should be abandoned as a concept relevant to clinical thermometry; 37.2°C (99.9°F) in the early morning and 37.7°C (99.9°F) overall should be regarded as the upper limit of the normal oral temperature range in healthy adults aged 40 years or younger, and several of Wunderlich's other cherished dictums should be revised.

(JAMA. 1992;268:1578-1580)
# Temperature dataset

<table>
<thead>
<tr>
<th>temp</th>
<th>gender</th>
<th>hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>96.3</td>
<td>1</td>
<td>70</td>
</tr>
<tr>
<td>96.7</td>
<td>1</td>
<td>71</td>
</tr>
<tr>
<td>96.9</td>
<td>1</td>
<td>74</td>
</tr>
<tr>
<td>97.0</td>
<td>1</td>
<td>80</td>
</tr>
<tr>
<td>97.1</td>
<td>1</td>
<td>73</td>
</tr>
<tr>
<td>97.1</td>
<td>1</td>
<td>75</td>
</tr>
<tr>
<td>97.1</td>
<td>1</td>
<td>82</td>
</tr>
<tr>
<td>97.2</td>
<td>1</td>
<td>64</td>
</tr>
<tr>
<td>97.3</td>
<td>1</td>
<td>69</td>
</tr>
<tr>
<td>97.4</td>
<td>1</td>
<td>70</td>
</tr>
<tr>
<td>97.4</td>
<td>1</td>
<td>68</td>
</tr>
<tr>
<td>97.4</td>
<td>1</td>
<td>72</td>
</tr>
<tr>
<td>97.4</td>
<td>1</td>
<td>78</td>
</tr>
<tr>
<td>97.5</td>
<td>1</td>
<td>70</td>
</tr>
<tr>
<td>97.5</td>
<td>1</td>
<td>75</td>
</tr>
<tr>
<td>97.6</td>
<td>1</td>
<td>74</td>
</tr>
<tr>
<td>97.6</td>
<td>1</td>
<td>69</td>
</tr>
<tr>
<td>97.6</td>
<td>1</td>
<td>73</td>
</tr>
<tr>
<td>97.7</td>
<td>1</td>
<td>77</td>
</tr>
<tr>
<td>97.8</td>
<td>1</td>
<td>58</td>
</tr>
<tr>
<td>97.8</td>
<td>1</td>
<td>73</td>
</tr>
<tr>
<td>97.8</td>
<td>1</td>
<td>65</td>
</tr>
<tr>
<td>97.8</td>
<td>1</td>
<td>74</td>
</tr>
</tbody>
</table>

Filename: **normtemp.csv**

Variables: **temp, gender, hr**

n = 130
Study 1

Transit times (hr) of marker pellets through the alimentary canal of patients with diverticulosis on 2 treatments

**Treatment A:** 44, 51, 52, 55, 60, 62, 66, 68, 69, 71, 71, 76, 82, 91, 108

**Treatment B:** 52, 64, 68, 74, 79, 83, 84, 88, 95, 97, 101, 116

*Is there difference between the two treatments beyond chance fluctuation?*
Study 2

10 patients, each was on 2 treatments for varicose ulcer. The outcome is the number of days from start of treatment to healing of ulcer.

**Standard Rx:** 35, 104, 27, 53, 72, 64, 97, 121, 86, 41

**New Rx:** 27, 52, 46, 33, 37, 82, 51, 92, 68, 62 days

*Is there difference between the two treatments?*
The answer to the two questions: t-test

- Also known as “Student’s t test”
- Applicable in
  - Comparison between two groups (independent or paired)
  - Outcome must be a continuous variable
Student’s t-test: a little history

- Invented by William Sealy Gosset
- Chemist, statistician, brewer (Guinness); keen on quality control and experimental design
- Attended Karl Pearson’s lectures at UCL in 1906
- Wrote the 2\textsuperscript{nd} paper (now called t-test) in 1908
- R. A. Fisher modified the test and gave us the modern form of the t-test

William S. Gosset (1876 – 1937)
This plaque in honor of William Sealy Gosset aka “Student” – he of Student's test of significance – is now displayed near Student's family home.
A century of $t$-tests

One hundred years ago, an author under the pseudonym of “Student” published a paper which was to become famous. It was entitled “The probable error of a mean”. But what we now know as Student’s $t$-test attracted little attention. It took another statistician of genius, R. A. Fisher, to amend, publicise and make it ubiquitous. But both Student’s and Fisher’s published versions were based upon faulty data. Stephen Senn reminds us of the third dedicated researcher and the quarter of a century delay before the story behind Student’s $t$-test emerged.
Comparing two samples: outlook

- Principle behind the Student’s t-test
- R solutions
  - One sample problem
  - Two independent samples problem
  - Paired sample
  - Non-normal data
  - Non-parametric test
Principle of t-test
### Two independent samples

**Table 1 Basic characteristics of participants**

<table>
<thead>
<tr>
<th>Variable</th>
<th>US white (n = 419)</th>
<th>Vietnamese (n = 210)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>71.5 (8.1)</td>
<td>61.7 (9.6)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>66.7 (12.9)</td>
<td>53.3 (7.9)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>160.8 (6.1)</td>
<td>148.9 (5.7)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td>25.8 (4.8)</td>
<td>24.1 (3.2)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Femoral neck BMD (g/cm(^2))</td>
<td>0.69 (0.12)</td>
<td>0.63 (0.11)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Lumbar spine BMD (g/cm(^2))</td>
<td>0.98 (0.19)</td>
<td>0.76 (0.14)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Whole body BMD (g/cm(^2))</td>
<td>1.05 (0.13)</td>
<td>0.89 (0.11)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Lean mass (kg)</td>
<td>38.6 (5.4)</td>
<td>32.3 (4.1)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Lean mass index (kg/m(^2))</td>
<td>14.8 (1.8)</td>
<td>14.6 (1.5)</td>
<td>0.0730</td>
</tr>
<tr>
<td>Fat mass (kg)</td>
<td>24.8 (8.1)</td>
<td>18.8 (4.9)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Percent body fat (%)</td>
<td>36.4 (6.5)</td>
<td>35.0 (6.2)</td>
<td>0.0122</td>
</tr>
</tbody>
</table>
## Transit times by treatment group

Transit times (hr)

<table>
<thead>
<tr>
<th></th>
<th>Treatment A</th>
<th>Treatment B</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Mean</td>
<td>68.4</td>
<td>83.4</td>
</tr>
<tr>
<td>SD (standard deviation)</td>
<td>16.5</td>
<td>17.6</td>
</tr>
</tbody>
</table>

Is there a real difference between the two group?
**Inference for two-sample problems**

- *Estimation and test of hypothesis*

Assumptions:

- Both groups were chosen randomly
- Two groups are *independent*
- Data are *normally* distributed
- Two groups have *equal variance* (homogeneity)
## Estimation: *sample vs population*

<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15 ((n_1))</td>
<td>Infinite</td>
</tr>
<tr>
<td>B</td>
<td>12 ((n_2))</td>
<td>Infinite</td>
</tr>
<tr>
<td>Mean</td>
<td>68.4 ((x_1))</td>
<td>(\mu_1 = ?)</td>
</tr>
<tr>
<td></td>
<td>83.4 ((x_2))</td>
<td>(\mu_2 = ?)</td>
</tr>
<tr>
<td>SD</td>
<td>16.5 ((s_1))</td>
<td>(\sigma_1 = ?)</td>
</tr>
<tr>
<td></td>
<td>17.6 ((s_2))</td>
<td>(\sigma_2 = ?)</td>
</tr>
</tbody>
</table>
## Estimation: *sample vs population*

<table>
<thead>
<tr>
<th></th>
<th>Sample</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>15 (n_1)</td>
<td>12 (n_2)</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>68.4 (x_1)</td>
<td>83.4 (x_2)</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>16.5 (s_1)</td>
<td>17.6 (s_2)</td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>(d = x_1 - x_1)</td>
<td></td>
</tr>
<tr>
<td><strong>Status</strong></td>
<td>Known</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sample</td>
<td>Population</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>N</td>
<td>15 ((n_1))</td>
<td>12 ((n_2))</td>
</tr>
<tr>
<td>Mean</td>
<td>68.4 ((x_1))</td>
<td>83.4 ((x_2))</td>
</tr>
<tr>
<td>SD (standard deviation)</td>
<td>16.5 ((s_1))</td>
<td>17.6 ((s_2))</td>
</tr>
<tr>
<td>Difference</td>
<td>(d = x_1 - x_1)</td>
<td></td>
</tr>
<tr>
<td>Status</td>
<td>Known</td>
<td></td>
</tr>
</tbody>
</table>

- “Is there real difference between A and B” means \(d = 0\).
- We need to work out the sampling variability of \(d\).
Estimation of standard deviation of $d$

Note that

$$d = \bar{x}_1 - \bar{x}_2$$

$$\text{var}(d) = \text{var}(\bar{x}_1) - \text{var}(\bar{x}_2)$$

Where

$$\text{var}(\bar{x}_1) = \frac{s_1^2}{n_1} \quad \text{and} \quad \text{var}(\bar{x}_2) = \frac{s_2^2}{n_2}$$
Estimation of standard deviation of $d$

- Variance of $d$
  \[ s^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \]

- Standard error of $d$
  \[ s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \]

- 95% confidence interval of $d$:
  \[ d \pm 1.96s \]
Null hypothesis

\[ H_0 : \mu_1 = \mu_2 \ (\delta = 0) \]

Alternative hypothesis

\[ H_1 : \mu_1 \neq \mu_2 \ (\delta \neq 0) \]

Question: IF \( Ho \ is \ true \), what is probability that we observed the actual data? ➔ \( P\)-value
Test of hypothesis

- Set $\alpha = 0.05$ or $\alpha = 0.01$
- Calculate the t statistic
- Compare the t statistic with what we would expect if $H_0$ is true

$t = -2.575$
$t = 2.575$

$\mu_1 - \mu_2 = 0$

or $t = 0$
Test statistic

\[ t = \frac{\text{Difference}}{\text{SD of difference}} = \frac{d}{s} \]

\[ = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = \frac{d}{s} \]
T-test using R
Function \texttt{t.test()}

Four forms of \texttt{t.test}

- **One-sample t-test**
  \begin{align*}
  \texttt{t.test}(Y, \texttt{mu}=y)
  \end{align*}
  \(Y\) is the variable, \(y\) is hypothetical value

- **Two-sample t-test**
  \begin{align*}
  \texttt{t.test}(Y \sim G)
  \end{align*}
  \(Y\) is the variable, \(G\) is the group variable
  \begin{align*}
  \texttt{t.test}(X1, X2)
  \end{align*}
  \(X1\) and \(X1\) are data for group 1 and group 2

- **Paired sample t-test**
  \begin{align*}
  \texttt{t.test}(X1, X2, \texttt{paired}=T)
  \end{align*}
One-sample problem:
Is the body temperature 37°C?

- Null hypothesis $\mu = 37$
- Let the observed (sample) mean be $X$
- $t$-test is defined as:

$$ t = \frac{X - \mu}{s / \sqrt{n}} $$

Where $s$ is the standard deviation, $n$ is sample size
(Note that $s / \sqrt{n} = \text{standard error}$)
Is the body temperature 37°C?

- Read the data (normtemp.csv)
- Convert F to C
- Use function t.test()

```r
# Read the data
read.csv(normtemp.csv, header=T)

t=temp=5*(t-temp-32)/9

# Convert
attach(t)

# Use function t.test()
hist(celsius, col="blue", border="white")

t.test(celsius, mu=37)
```
Output (one-sample problem)

> t.test(celsius, mu=37)

One Sample t-test

data:  celsius

  t = -5.4548, df = 129, p-value = 2.411e-07
alternative hypothesis: true mean is not equal to 37

95 percent confidence interval:
  36.73445 36.87581

sample estimates:
  mean of x
  36.80513
Two-sample problem
Two ways of organising data

**Long format:** each row/obs is identified by group

<table>
<thead>
<tr>
<th>X</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>1</td>
</tr>
<tr>
<td>68</td>
<td>2</td>
</tr>
<tr>
<td>62</td>
<td>1</td>
</tr>
<tr>
<td>71</td>
<td>1</td>
</tr>
<tr>
<td>76</td>
<td>1</td>
</tr>
<tr>
<td>95</td>
<td>2</td>
</tr>
<tr>
<td>101</td>
<td>2</td>
</tr>
</tbody>
</table>

**Wide format:** each group is a column/variable with actual value

<table>
<thead>
<tr>
<th>x1</th>
<th>x2</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>68</td>
</tr>
<tr>
<td>62</td>
<td>95</td>
</tr>
<tr>
<td>71</td>
<td>101</td>
</tr>
<tr>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{t.test}(X \sim \text{group}) \quad \text{t.test}(x1, x2) \]
Transit time study

TreatA = c(44, 51, 52, 55, 60, 62, 66, 68, 69, 71, 71, 76, 82, 91, 108)
TreatB = c(52, 64, 68, 74, 79, 83, 84, 88, 95, 97, 101, 116)

t.test(TreatA, TreatB)

Time = c(TreatA, TreatB)
Treatment = c(rep("A", 15), rep("B", 12))

t.test(Time ~ Treatment)
> t.test(Time ~ Treatment)

Welch Two Sample t-test

data:  Time by Treatment
t = -2.2636, df = 22.937, p-value = 0.03337
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:  
  -28.742142   -1.291191
sample estimates:  
mean in group A mean in group B  
  68.40000       83.41667
R output – t-test

t = -2.2636, df = 22.937, p-value = 0.03337
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
-28.742142  -1.291191
sample estimates:
mean in group A mean in group B
68.40000        83.41667

<table>
<thead>
<tr>
<th>Treatment</th>
<th>N</th>
<th>Treatment</th>
<th>N</th>
<th>Difference and 95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15</td>
<td>B</td>
<td>12</td>
<td>-15.0 (-28.7, -1.3)</td>
<td>0.033</td>
</tr>
</tbody>
</table>
The transit time in patients on treatment A was lower than that in those on treatment B, with average difference being -15 hours (95% CI: -28.7 to -1.3; P = 0.03).
se = (-1.3+28.7)/(2*1.96)
d = rnorm(10000, mean=-15, sd=se)
hist(d, breaks=30, xlim=c(-40, 30))
Paired design
Paired design

• “Before – After” design

• Each individual is measured twice
  – before and after treatment
  – on two different treatments
  – Etc

• Interested in the **difference** in EACH individual (δ)

• Hypothesis: δ = 0?
Study 2

10 patients, each was on 2 treatments for varicose ulcer. The outcome is the number of days from start of treatment to healing of ulcer.

**Standard Rx:** 35, 104, 27, 53, 72, 64, 97, 121, 86, 41

**New Rx:** 27, 52, 46, 33, 37, 82, 51, 92, 68, 62 days

<table>
<thead>
<tr>
<th>Standard Rx</th>
<th>New Rx</th>
<th>Difference (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>27</td>
<td>-8</td>
</tr>
<tr>
<td>104</td>
<td>52</td>
<td>-52</td>
</tr>
<tr>
<td>27</td>
<td>46</td>
<td>19</td>
</tr>
<tr>
<td>53</td>
<td>33</td>
<td>-20</td>
</tr>
<tr>
<td>72</td>
<td>37</td>
<td>-35</td>
</tr>
<tr>
<td>64</td>
<td>82</td>
<td>18</td>
</tr>
<tr>
<td>97</td>
<td>51</td>
<td>-46</td>
</tr>
<tr>
<td>121</td>
<td>92</td>
<td>-29</td>
</tr>
<tr>
<td>86</td>
<td>68</td>
<td>-18</td>
</tr>
<tr>
<td>41</td>
<td>62</td>
<td>21</td>
</tr>
</tbody>
</table>

**Mean SD**

-15

27
R solution for paired t test

Std = c(35, 104, 27, 53, 72, 64, 97, 121, 86, 41)  
New = c(27, 52, 46, 33, 37, 82, 51, 92, 68, 62)  
d = New - Std

t.test(New, Std, \texttt{paired=T})

t.test(d, \texttt{mu=0})
R solution for paired t test

> t.test(New, Std, paired=T)

    Paired t-test

data:  New and Std
t = -1.7583, df = 9, p-value = 0.1126
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -34.298438   4.298438
sample estimates:
mean of the differences
  -15
On average, the new treatment resulted in a shorter duration of healing by 15 days. However, there was considerable uncertainty in the effect as the 95% CI shows that compared with the standard treatment the new treatment could shorten the duration of healing by 34 days, but there is a small probability that the new treatment could also increase the duration by 4 days.
Transformation
Assumptions of t-test

- Data are normally distributed
- Variance of group 1 is equivalent to variance of group 2 (i.e., homogeneity)
- Independent group
- Random sampling
> describe.by(xlap, sex)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th></th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>var</strong></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>n</strong></td>
<td>100</td>
<td>144</td>
<td>1</td>
</tr>
<tr>
<td><strong>mean</strong></td>
<td>0.45</td>
<td>0.34</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>sd</strong></td>
<td>0.31</td>
<td>0.24</td>
<td>0.24</td>
</tr>
<tr>
<td><strong>median</strong></td>
<td>0.35</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>trimmed</strong></td>
<td>0.41</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td><strong>mad</strong></td>
<td>0.26</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>min</strong></td>
<td>0.02</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>max</strong></td>
<td>1.57</td>
<td>1.26</td>
<td>1.26</td>
</tr>
<tr>
<td><strong>range</strong></td>
<td>1.55</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>skew</strong></td>
<td>2.15</td>
<td>1.39</td>
<td>1.39</td>
</tr>
<tr>
<td><strong>kurtosis</strong></td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>se</strong></td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Beta-crosslap data (bone resorption marker)
> library(nortest)
> pearson.test(xlap)
Pearson chi-square normality test
data:  xlap
P = 87.877, p-value = 6.145e-12

> pearson.test(log(xlap))
Pearson chi-square normality test
data:  log(xlap)
P = 20.7541, p-value = 0.1882
Analysis based on transformed data

```r
> describe.by(log(xlap), sex)

group: 1

<table>
<thead>
<tr>
<th>var</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>skew</th>
<th>kurtosis</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>-0.73</td>
<td>0.52</td>
<td>-0.8</td>
<td>-0.73</td>
<td>0.54</td>
<td>-2.1</td>
<td>0.51</td>
<td>2.61</td>
<td>-0.34</td>
<td>-0.34</td>
<td>0.05</td>
</tr>
</tbody>
</table>

------------------------------------------------------------

group: 2

<table>
<thead>
<tr>
<th>var</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>skew</th>
<th>kurtosis</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>144</td>
<td>-0.94</td>
<td>0.5</td>
<td>-0.97</td>
<td>-0.97</td>
<td>0.52</td>
<td>-1.82</td>
<td>0.31</td>
<td>2.13</td>
<td>0.3</td>
<td>-0.66</td>
<td>0.04</td>
</tr>
</tbody>
</table>
```
A review of algebra …

\[
\log(x_1 x_2) = \log(x_1) + \log(x_2)
\]

\[
\log\left(\frac{x_1}{x_2}\right) = \log(x_1) - \log(x_2)
\]

\[-0.73 + 0.9447 = 0.2147 = \log(x_1) - \log(x_2)\]

\[
\log\left(\frac{x_1}{x_2}\right) = 0.2147
\]

\[
\frac{x_1}{x_2} = e^{0.2147} = 1.24
\]

Gee! That is great (year 8 math!)
T-test on transformed data

> t.test(log(xlap) ~ sex)
data:  log(xlap) by sex
t = 3.216, df = 206.284, p-value = 0.001509
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  0.08306888 0.34626674
sample estimates:
mean in group 1 mean in group 2
  -0.7300382       -0.9447060

Mean difference: $d = \exp(-0.73 + 0.9447) = 1.24$

Lower 95% CI: $\exp(0.083) = 1.09$

Upper 95% CI: $\exp(0.346) = 1.41$
### Interpretation

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
<th>Percentage difference and 95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>100</td>
<td>144</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.45 (0.31)</td>
<td>0.34 (0.24)</td>
<td>24% (9, 41)</td>
<td>0.0015</td>
</tr>
</tbody>
</table>

Compared with women, beta-crosslap was 24% (95% CI: 8.6 to 41.3%) higher in men, and the difference was statistically significant ($P = 0.001$)
Non-parametric test
Scenario

• Sometimes, it is impossible to
  – “Normalize” the data
  – Homogenize the data
  – Outliers

• Two alternative approaches
  – Non-parametric test
  – Bootstrap method
Non-parametric method

- Makes no assumption of distribution of data
- Works based on median and ranks
- Non-parametric test for two samples (continuous data)
  - Mann-Whitney U test
  - Also Mann-Whitney-Wilcoxon test
Mann-Whitney-Wilcoxon U test

- Rank the values in both groups (together) from highest to lowest
- Sum the ranks for each group
- The sum of ranks for each group are used to make the statistical comparison
Mann-Whitney-Wilcoxon U test

TreatA = c(44, 51, 52, 55, 60, 62, 66, 68, 69, 71, 71, 76, 82, 91, 108)
TreatB = c(52, 64, 68, 74, 79, 83, 84, 88, 95, 97, 101, 116)

Time = c(TreatA, TreatB)

Treatment = c(rep("A", 15), rep("B", 12))

> Time
[1] 44 51 52 55 60 62 66 68 69 71 71 76 82 91 108 52 64
[18] 68 74 79 83 84 88 95 97 101 116

> rank(Time)
[1] 1.0 2.0 3.5 5.0 6.0 7.0 9.0 10.5 12.0 13.5 13.5 16.0 18.0
[14] 22.0 26.0 3.5 8.0 10.5 15.0 17.0 19.0 20.0 21.0 23.0 24.0 25.0
[27] 27.0
Mann-Whitney-Wilcoxon test using R

> wilcox.test(Time ~ Treatment)

    Wilcoxon rank sum test with continuity correction

data:  Time by Treatment
W = 45, p-value = 0.02983
alternative hypothesis: true location shift is not equal to 0
Summary

- T-test: comparing two independent groups
  
  \texttt{t.test(Y ~ group)}
  
  \texttt{t.test(Y1, Y2, paired=T)}

- Assumptions of t-test: normal distribution, similar variance, independence, random samples

- Transforming data to normal distribution, if necessary

- Non-parametric test: Wilcoxon rank sum test