What is a Student’s t test?

- A Student’s t-test is a statistical hypothesis test in which the test statistic follows a Student’s t distribution (or t distribution) if the null hypothesis is true.

- The Student’s t-statistic was introduced in 1908 by William Sealy Gosset, using his pen name “Student”, when he worked for the Guinness brewery in Dublin, Ireland. He devised the t-test as a way to cheaply monitor the quality of stout. He published the test in Biometrika in 1908.
Student’s distribution

- Assume $\sigma^2$ is known. If $X \sim N(\mu, \sigma^2)$, then
  \[ \frac{X - \mu}{\sqrt{\sigma^2}} \sim N(0,1) \]

- When $\sigma^2$ is unknown, and estimated by the sample variance, then
  \[ \frac{\bar{X} - \mu}{\sqrt{\sum_i (X_i - \bar{X})^2 / (n-1)}} \sim t_{n-1} \]

Uses of t-test

- Assuming the null hypothesis (H0) is true:
  - One-sample location test: whether the mean of a normally distributed population has a value specified.
  - Two sample location test: whether the means of two normally distributed populations are equal.
  - Paired t test: whether the difference between two responses measured on the same statistical unit has a zero mean.
  - Test of trend: whether the slope of a regression line differs significantly from 0.
Assumptions

- Let \( t = \frac{Z}{s} \) be a variable following a t-distribution. Then, the underlying assumptions are as follows:
  - \( Z \) follows a standard normal distribution under the null hypothesis (H0)
  - \( ps^2 \) follows a Chi-square distribution with \( p \) degrees of freedom under H0, where \( p \) is a positive constant.
  - \( Z \) and \( s \) are independent.

Example data

- Daily energy intake in kJ for 11 women (Altman, 1991, p. 183)

```r
> energy.pre <- c(5260, 5470, 5640, 6180, 6390, 
>                   6515, 6805, 7515, 7515, 8230, 8770) 
> energy.post <- c(3910, 4220, 3885, 5160, 5645, 
>                   4680, 5265, 5975, 6790, 6900, 7335)
```
One-sample $t$ test

```r
> t.test(energy.pre,mu=7725)

    data:  energy.pre
     t = -2.8208, df = 10, p-value = 0.01814
  alternative hypothesis: true mean is not equal to 7725
95 percent confidence interval:
       5986.348 7520.925
  sample estimates:
     mean of x
          6753.636
```

Computing the $t$ value

```r
> diff<-mean(energy.pre)-7725
> se<-sd(energy.pre)/sqrt(11)
> t<-diff/se
> t

[1] -2.820754
t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}

> (1-pt(abs(t),df=10))*2
[1] 0.01813724
```
Two-sample t test

- **Equal sample sizes, equal variance**
  \[ t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} \]
  \[ S_{\bar{X}_1 - \bar{X}_2} = \frac{1}{n} \left( S^2_{\bar{X}_1} + S^2_{\bar{X}_2} \right) \]

- **Unequal sample sizes, equal variance**
  \[ S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(n_1 - 1) S^2_{\bar{X}_1} + (n_2 - 1) S^2_{\bar{X}_2}}{n_1 + n_2 - 2} \times \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \]

- **Unequal sample sizes, unequal variance (Welch t-test)**
  \[ S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{1}{n_1} s^2_1 + \frac{1}{n_2} s^2_2} \]

Two-sample t test: equal variance

```r
> t.test(energy.post, energy.pre, var.equal=T)
```

```
Two Sample t-test

data:  energy.post and energy.pre
t = -2.6242, df = 20, p-value = 0.01625 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -2370.0763 -270.8328 sample estimates: mean of x mean of y 5433.182 6753.636
```
T-test for paired samples

```r
> t.test(energy.post, energy.pre, paired=T)

Paired t-test

t = -11.9414, df = 10, p-value = 3.059e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1566.838 -1074.072
sample estimates:
mean of the differences
-1320.455
```

The formula for the Paired t-test is:

\[
t = \frac{\bar{X}_D - \mu_0}{s_D / \sqrt{n}}
\]

data: energy.post and energy.pre

t = -11.9414, df = 10, p-value = 3.059e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1566.838 -1074.072
sample estimates:
mean of the differences
-1320.455