Module 11
One-Sample t-Tests

<u>Module Learning Outcomes</u>

- Define a proper parameter and set up the null hypothesis and alternative hypothesis for one-sample mean.
- Calculate the test statistic.
- Use the t-table to find corresponding probabilities.
- Find p-values using the t-tables.

Assumption and Review of Notation/Symbols

- Assumption #1: All parameters are unknown to us
- Assumption #2: The variable of interest (X) is assumed to have a Normal distribution.
- Here are some key notations or symbols used in this module.

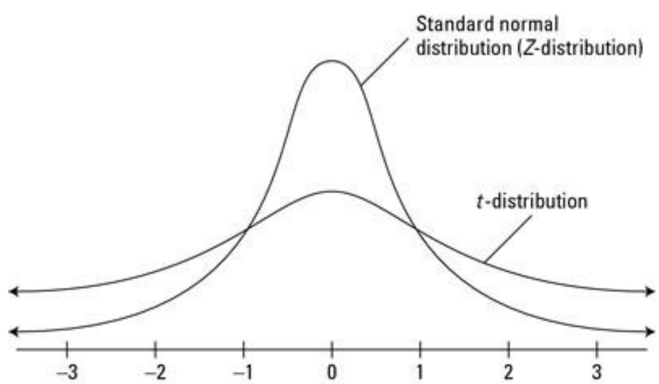
Symbols	Name	Description	
μ	Population Average/Mean	The average/mean of the population	Parameter
$\bar{X}$	Sample Average/Mean	The average/mean of a sample	Statistic

σ	Population Standard Deviation	The standard deviation of the population	Parameter
s	Sample Standard Deviation	The standard deviation of a sample	Statistic

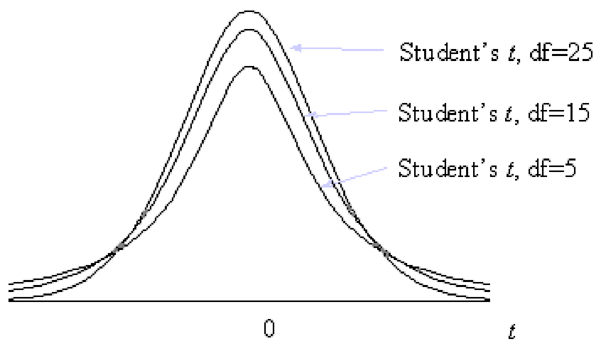
- Note: Statistics are used to estimate unknown parameters.

11.1 Student's t-Distributions

- The family of t-distributions is invented by William Gosset when he was doing studies with behaviour of means in small sample size.
- He chose to use a pseudonym "Student" in his publication of this finding to avoid any conflict of interest as he was working at Guinness Brewery at the time. Hence, it is now commonly known as "***Student's t-distributions***".
- In general, t-distributions look very similar to the Z-distribution.



- Both looks more or less a bell-shape.
- Both distributions are centred at zero.
- The only difference is the variability, where t-distribution is bigger variability or spread.
- From the look point of view, it looks wider or more spread out.
- Another difference is that the Z-distribution (or Normal distribution in general) has two parameters of interest, namely μ and σ. In contrast, t-distributions only have one parameter and it is called the "**degree of freedom**" or simply **DF**.



- There is only one Z-distribution but there are many t-distributions with various degrees of freedom.

11.2 Null Hypothesis and Alternative Hypothesis

- The setup is exactly the same as Z-tests (in the previous module).
- The three sets of hypotheses are listed here for your reference.

	Null Hypothesis	**Alternative Hypothesis**
Lower-tailed test	$H_0: \mu \geq 10$	$H_a: \mu < 10$
Upper-tailed test	$H_0: \mu \leq 10$	$H_a: \mu > 10$
Two-tailed test	$H_0: \mu = 10$	$H_a: \mu \neq 10$

- MLO: Define a proper parameter and set up the null hypothesis and alternative hypothesis for one-sample mean.

11.3 Test Statistic Calculations

- From the last module, we know that the sample mean $\bar{X}$ becomes a Z-score after being standardized, i.e.

$$Z = \frac{value - mean}{SD} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

- The only difference here is that when the population standard deviation (σ) is unknown and the sample equivalent (s) is used to estimate it, the whole expression is no longer a Z-score.
- Instead, it (the expression) follows the t-distribution with ν degrees of freedom as long as the variable of interest (X) follows a Normal distribution.
- Note: This symbol is pronounced as "nu" and is a pretty common Greek letter used for degree of freedom.
- The value of ν varies in different tests. Here, for a one-sample t-tests, $\nu = n - 1$.
- The test statistic becomes $T.S. = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ and it follows the t-distribution with $\nu = n - 1$ degree of freedom.
- MLO: Calculate the test statistic.

11.4 The t-Table

- Contrast: In Z-table, there were only two items – Z-scores and cumulative areas.
- In t-table, there are three items: 1) degrees of freedom, 2) upper-tail areas, and 3) t-scores.
- 1) The degree of freedom usually goes non-stop from 1 to 30. The main reason is because when $n \geq 30$, most of us would use Normal approximation to find p-value for convenience sake.
- 2) Most t-table only shows "upper tails" because t-distributions are symmetrical. So, if you have a negative test statistic, you just need to flip the values.
- 3) The t-scores are like Z-scores. They are numbers on the x-axis.
- For example, when $\nu = 29$, the t-score will be 1.6991 when the tailed area is 0.05. In some textbooks, a more compact way to represent this is $t_{29,0.05} = 1.6991$.
- MLO: Use the t-table to find corresponding probabilities.

11.5 Finding p-values

- Finding p-values in the t-table is not as straight-forward as using the Z-table.
- We will have to find the two adjacent t-scores and guess the range of p-values.
- For example, in a one-tailed tests, the p-value of the $TS = +1.26$ with $v = 20$ will be between 0.10 and 0.20. It is because the $TS = +1.26$ falls between the two adjacent t-scores of $t_{20,0.10} = 1.3253$ and $t_{20,0.20} = 0.8600$.
- Note that the tail area will have to be multiplied by two for the p-value of two tailed tests.
- MLO: Find p-values using the t-tables.

11.6 Conclusion

- The conclusion is made exactly the same way.