When assumptions are violated:

Frequency distributions often aren't normal and standard deviations aren't always equal. So, what do you do when nature won't cooperate?

You can:

1) **Ignore the violations of assumptions** ***(check how robust your test is)***
2) **Transform the data** ***(we'll talk about this here)***
3) **Use a nonparametric method**

Data transformations:

A data tranformation changes each measurement by the same mathematical formula. Then, you can test your data for the assumptions. If they no longer violate the assumptions, then go ahead and use that test. The most frequently used transformations are the log transformation, the arcsine transformation, and the square-root transformation.

Log Transformation: is the most common data transformation in biology. The data are converted by taking the natural log of each measurement:

$$Y'=\ln[Y]$$

This transformation can only be applied when all the values are greater than zero. (If the data include zero, you can try Y'=ln[Y+1]). This transformation is likely to be useful when

- measurements are ratios or products of variables
- the frequency distribution of the data is skewed to the right
- the group with the larger mean also has a larger standard deviation
- the data span several orders of magnitude

Arcsine Transformation: is used almost exclusively on data that are proportions:

$$p'=\text{arcsine}[\sqrt{p}]$$

Proportions tend not to be normally distributed, so this transformation is useful for those instances, especially when the mean is close to zero or to one. Also, groups differing in their mean proportions tend to have different standard deviations.

Square-root Transformation: is often used when the data are counts:

$$Y'=\sqrt{Y+1/2}$$

This transformation helps to equalize standard deviations between groups when the group with the higher mean has a higher standard deviation. This transformation often has similar results to the log transformation.

Other Transformations: when the frequency distribution of the data is skewed left, try the square transformation:

$$Y'=Y^2$$

If that doesn't work, try the antilog transformation on left-skewed data:

$$Y'=e^Y$$

When the data are skewed right, try the reciprocal transformation:

$$Y' = 1/Y$$

The square and reciprocal transformation can only be used if all the data points in all samples have the same sign. If all of the data points are negative, try multiplying each number by -1 before transforming.

While it's okay to try multiple transformations to see which one best meets the assumptions of your test, it is not okay to try multiple transformations to see which one gives you the most significant result. This inflates Type I error.