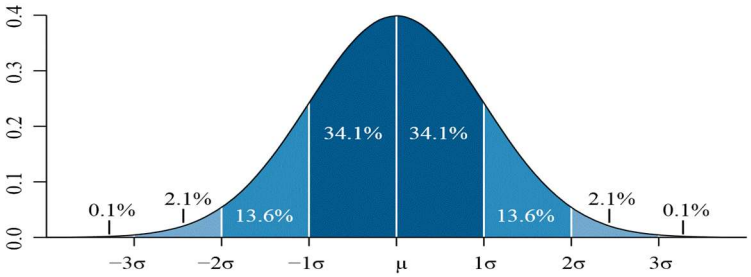
Tests Summary

Type	Test Name	When to use
Normality	**Shapiro-Wilk** (best)	+ Straight line, Slope = 1 Test to see if data deviates from a straight line.
	Kolmogorov-Smirnov (general)	- p<0.05 or p<0.1 Compares the cdf (cumulative density function) of the variable tested vs. cdf of a gaussian variable and determines the probability of observing a deviation as large or larger.
	Lilliefors	
	Anderson-Darling	
Homoscedasticity	**Levene (best)**	Best for non-normal distributions
	Bartlett	Best for normal distributions
	Fisher	Least robust if normality is not present
Bivariate (2 discrete variables)	**Cramer's V**	
	Chi-Square	
Bivariate (1 continuous, 1 discrete)	**Parametric ANOVA**	Requires normality & homoscedasticity
	Wilcoxon-Mann-Whitney	Non-parametric, 2 groups
	Kruskal-Wallis	Non-parametric, >2 groups

Mean comparison tests		
Form of distribution	Two samples	Three or more samples
Normality & Homoscedasticity	Student's *t* test	ANOVA
Normality & Heteroscedasticity	Welch's *t* test	Welch – ANOVA
Non-normality & Heteroscedasticity	Wilcoxon–Mann–Whitney	Kruskal–Wallis
	Median test	Median test
		Jonckheere–Terpstra test (ordered samples)

Tests

Tests for Normality



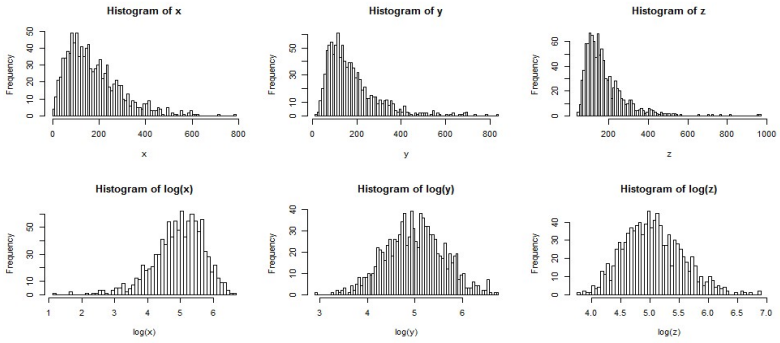
The normality of a variable can be verified by the following tests:

1. **Shapiro–Wilk test (the best)**
    a. Normal distribution: Straight Line with **Slope = 1**
2. **Kolmogorov–Smirnov test (the most general)**
    a. NOT a Normal distribution: If probability **<0.05** or **<0.10** we reject H_0
3. Lilliefors test
4. Anderson–Darling test

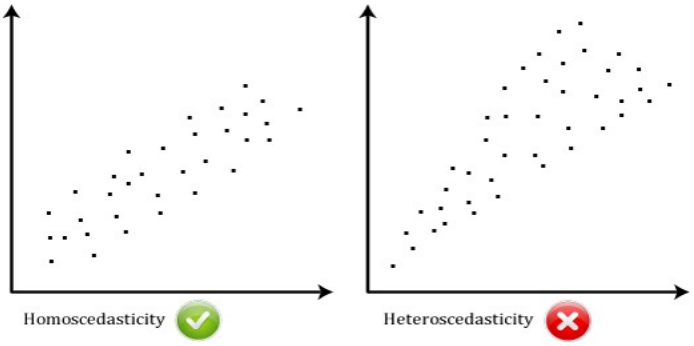
Tests for Normality				
Test	**Statistic**		**P-Value**	
Shapiro-Wilk	W	0.992264	Pr < W	0.9845
Kolmogorov-Smirnov	D	0.051999	Pr > D	> 0.1500
Cramer-von Mises	W-Sq	0.02621	Pr > W-Sq	> 0.2500
Anderson-Darling	A-Sq	0.162071	Pr > A-Sq	> 0.2500

Transformations Used to Reduce Skewness

1. **Log transformation** (has to be adjusted for 0 and negative values)
    a. If data is skewed, log transformation will transform the data to a form where we see a more normal distribution.
2. Square (left skew)
3. Square root (right skew)
4. Cube (right skew)



Test for Homoscedasticity and Heteroscedasticity



1. **Levene Test (best – low sensitivity to non-normality)**
2. **Bartlett Test (best if distribution is equal)**
3. Fisher Test (least robust if normality is not present)

Test of Homogeneity of Variance			
Levene Statistic	**Df1**	**Df2**	**Significance**
50.448	2	396	.000

Bivariate Tests

Measuring Links between Two Discrete Variables

e.g. gender and smoking

1. **Cramer's V**
2. **Chi-Square**

Measuring Links between One Discrete & One Continuous Variable

e.g. dosage of a medicine and recovery time

1. **Parametric ANOVA Test**
    a. Requires normality/homoscedasticity assumption
2. **Non-Parametric Approaches**
    a. **Wilcoxon-Mann-Whitney (2 groups)**
    b. **Kruskal-Wallis (>2 groups)**