Continuous data analysis and interpretation

Rohali Keesari, PharmD, MPH

Biostatistician

Pediatrics Biostatistics Core

10/29/2021

Outline

- Describing continuous data
- Summarizing continuous data using the measures of location and dispersion
- Summarizing continuous data using graphs
- Understanding normal distribution
- Checking the assumption of normality
- Understanding the measures of associations
- Interpretation of measures of associations

Continuous data

- Continuous data can have values on a continuum.
- It can be measured on a scale or continuum and can take any numeric value (within a range)

Examples:

- A person's height: could be any value (within the range of human heights), not just certain fixed heights
- Temperature
- A child's birth weight
- Age of a person

Summary statistics

- Measures of location (also called central tendency)
- Measures of spread (also called dispersion)
- Graphs/charts

Measures of location or central tendency

- Mean
- Median
- Mode

Measures of Central tendency

Mean= Sum of observations divided by the number of observations

Specifically, this is the arithmetic mean

6 8 10 12 14

Mean= (6+8+10+12+14)/5

= 50/5

= 10 years

Measures of Central tendency

Median= Middle value in the series of numbers

6 8 10 12 14

4 6 10 12 14 16

Median = 11 years

To find the median:

- Arrange the data points from smallest to largest.
- If the number of data points is odd, the median is the middle data point in the list.
- If the number of data points is even, the median is the average of the two middle data points in the list.

Measures of Central tendency

Mode= Value that occurs most frequently in the dataset

Example: The mode of {2, 4, 4, 3, 6, 4} is 4 because it occurs three times, which is more than any other number.

To find the mode:

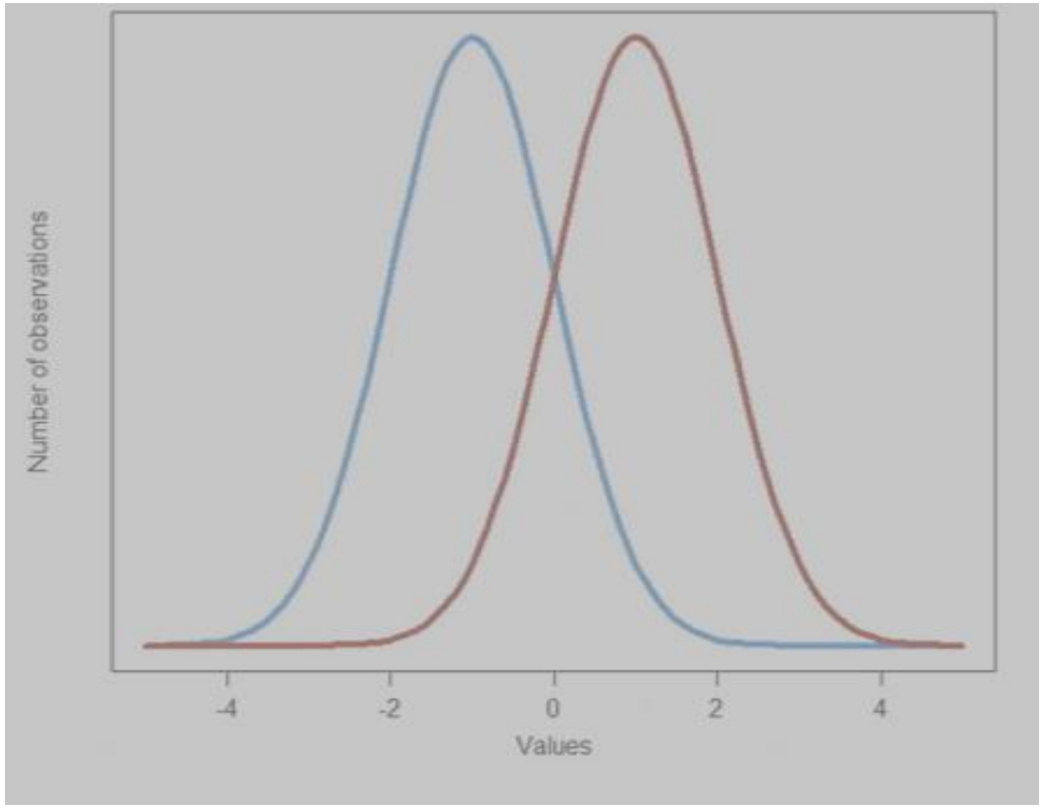
- Arrange the data points from smallest to largest.
- Find the number that appears most often

Hints to remember the difference

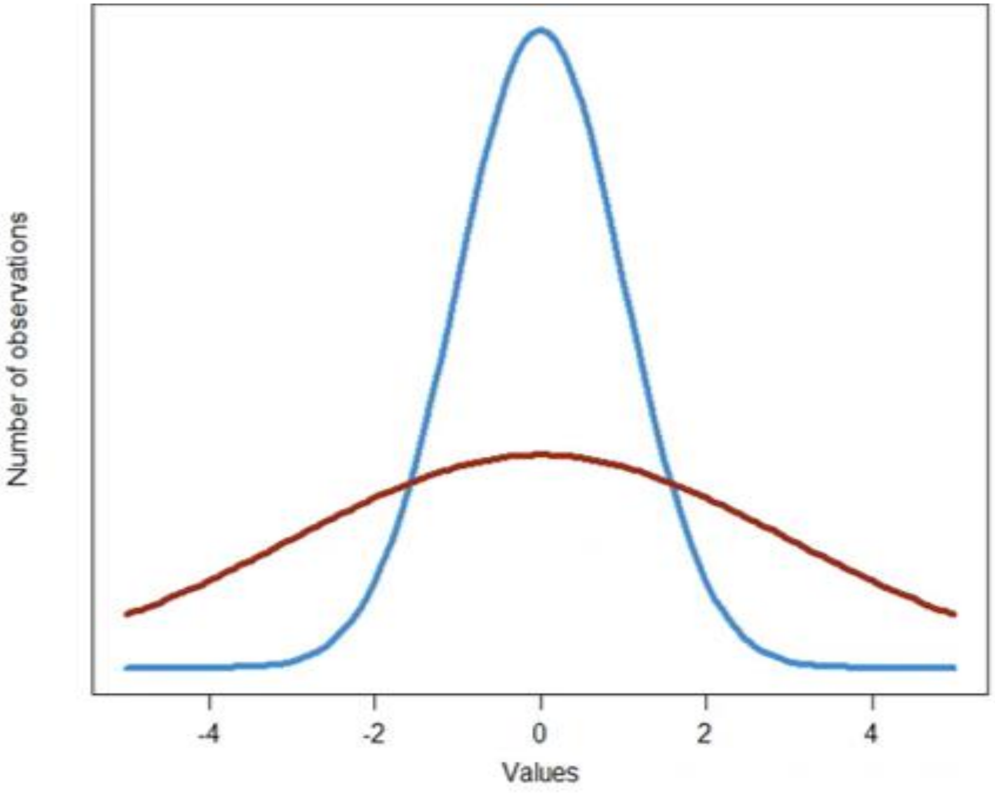
- “A la mode” is a French word that means *fashionable* ; It also refers to a popular way of serving ice cream. So “Mode” is the most popular or fashionable member of a set of numbers.
- The word MOde is also like MOst.
- The “Mean” requires you to do arithmetic (adding all the numbers and dividing) so that’s the “mean” one.
- “Median” has the same number of letters as “Middle”.

Measures of Dispersion of data

Central tendency:
"Typical" or "average" value



Dispersion:
Variability in values



Measures of Dispersion

- Range
- Interquartile range
- Standard deviation

Measures of Dispersion of data

Range= Tells you how spread out the data is. It’s the difference between the largest and the smallest observation in the data.

Example: 1,2,3,4,5,6,7

Range = Highest value – Lowest value

= (7 – 1) = 6

Measures of Dispersion of data

Inter-quartile range (IQR)= It is just the range of the quartiles: the distance from the largest quartile to the smallest quartile, which is

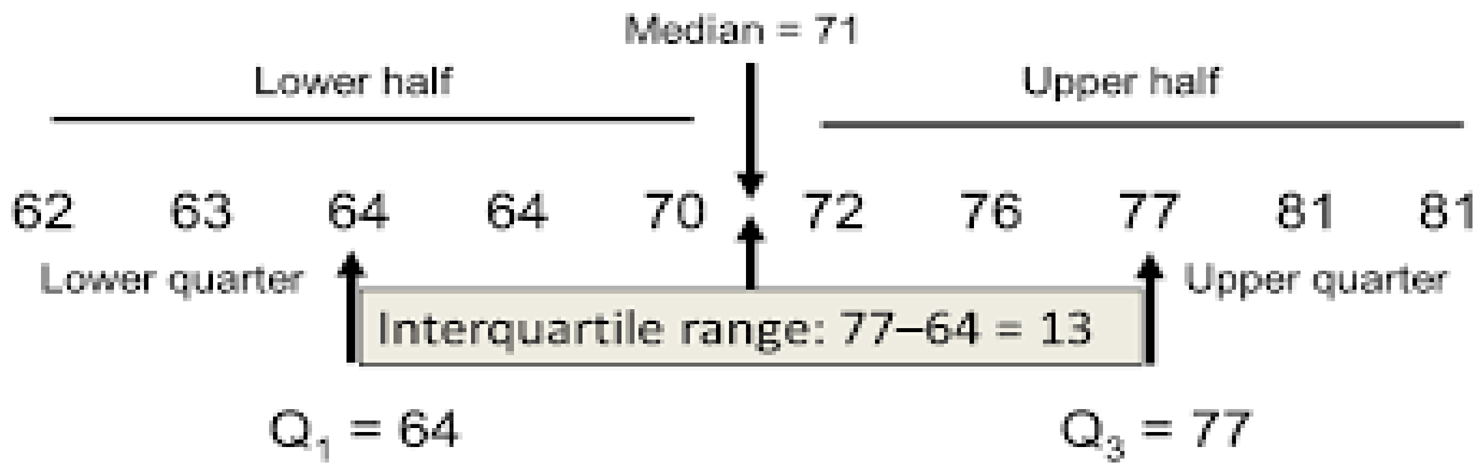
$$IQR = Q3 - Q1$$

Measures of Dispersion of data

What are quartiles?

Quartiles divide a data set into quarters, or four parts.

- Q3 is the upper quartile. It is the median of the upper half of data
- Q2 is the middle quartile. It is the same as the median of the data set.
- Q1 is the lower quartile. It is the median of the lower half of data.

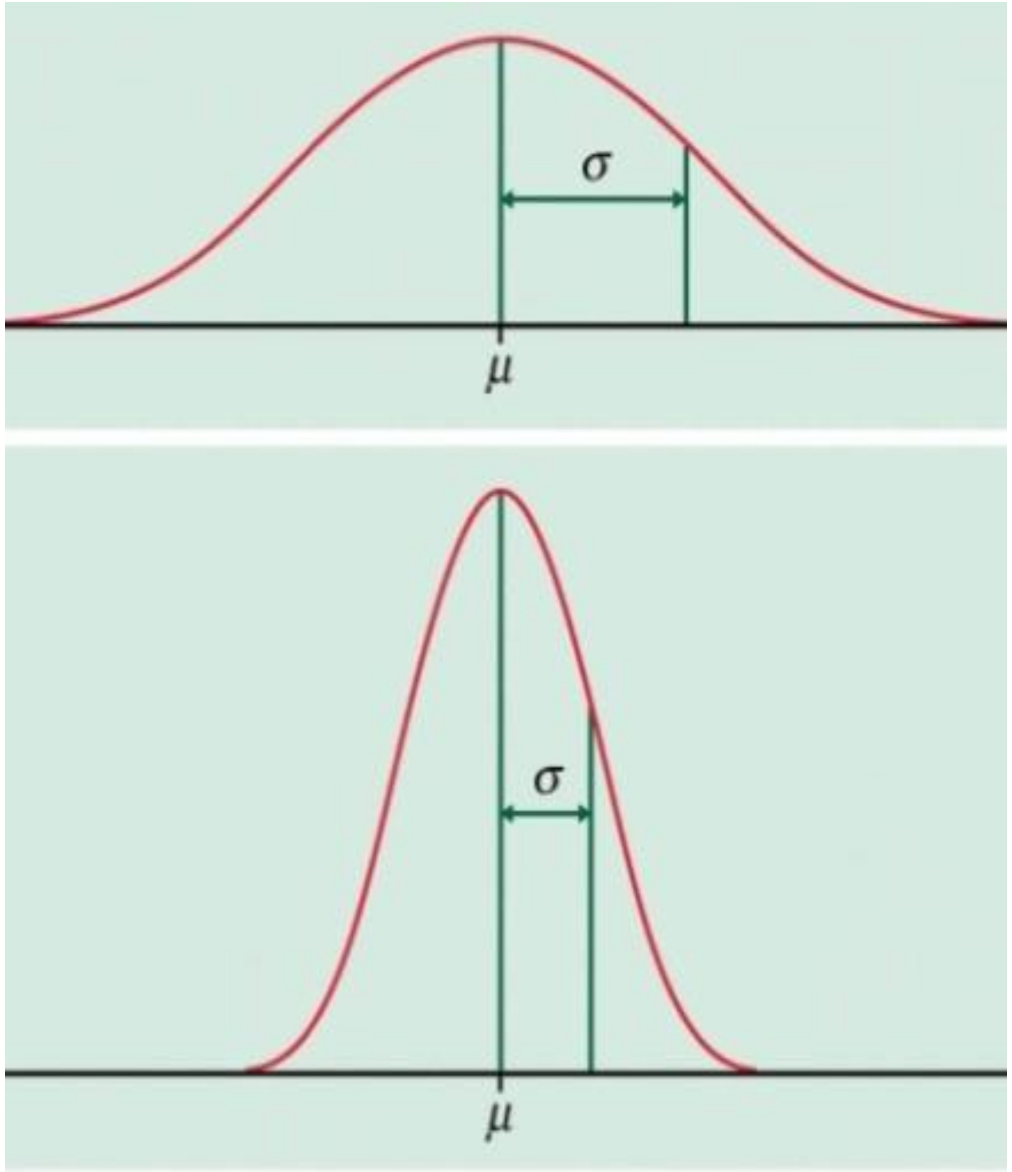


To find the interquartile range

1. Order the data from least to greatest.
2. Split the data into a lower half and an upper half.
3. Find Q_1 and Q_3.
4. Subtract: IQR=**Q3-Q1**

Measures of Dispersion of data

- Standard deviation= measure of how dispersed the data is in relation to the mean
- Its symbol is **σ** (the greek letter sigma)



To calculate the standard deviation

- Work out the Mean (the simple average of the numbers)
- Subtract the mean individually from each of the numbers given and square the result.
- Now add up these results
- Divide by N
- And finally, square that and we are done!

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

σ = population standard deviation

N = the size of the population

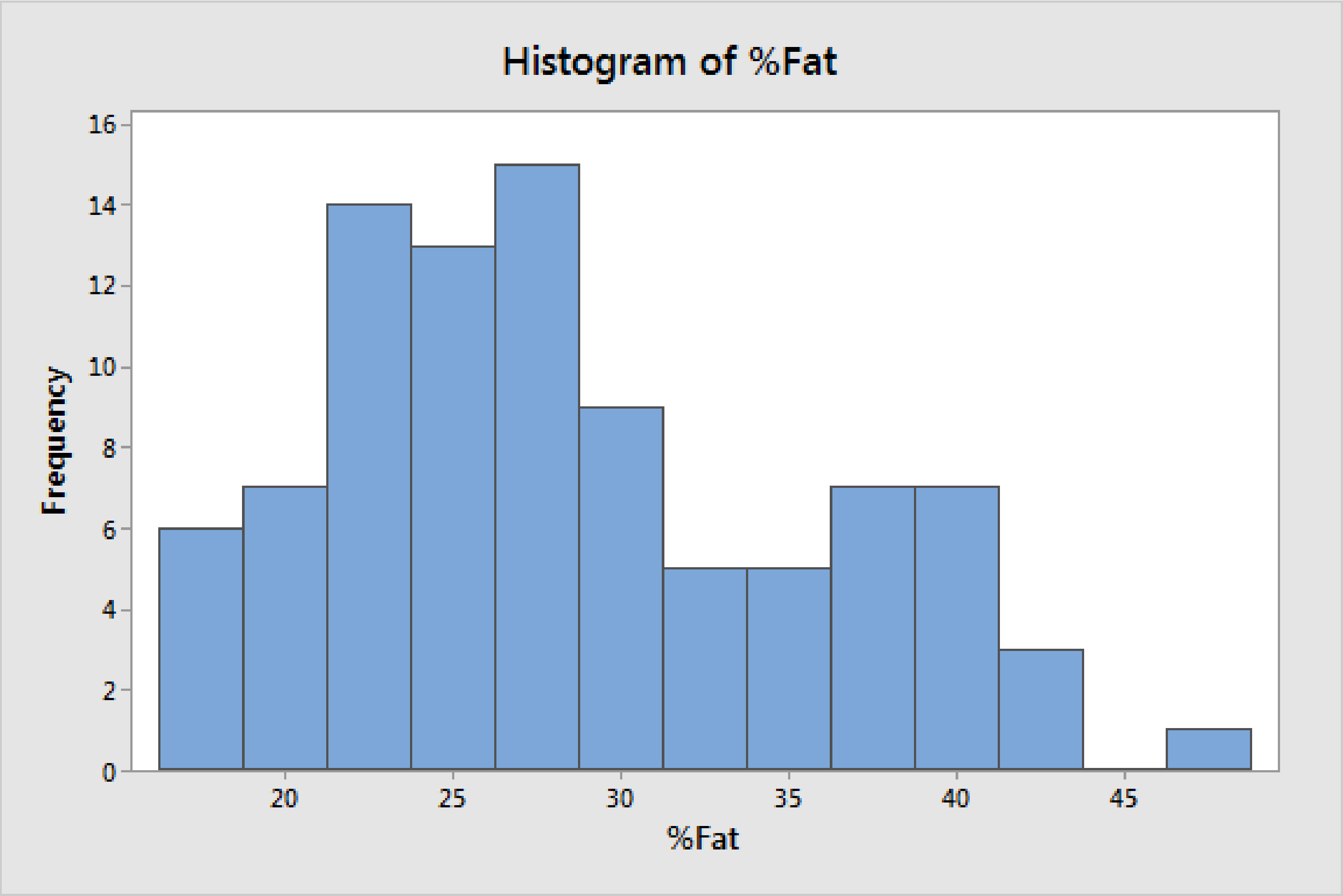
x_i = each value from the population

μ = the population mean

Summarize continuous data: Graphs

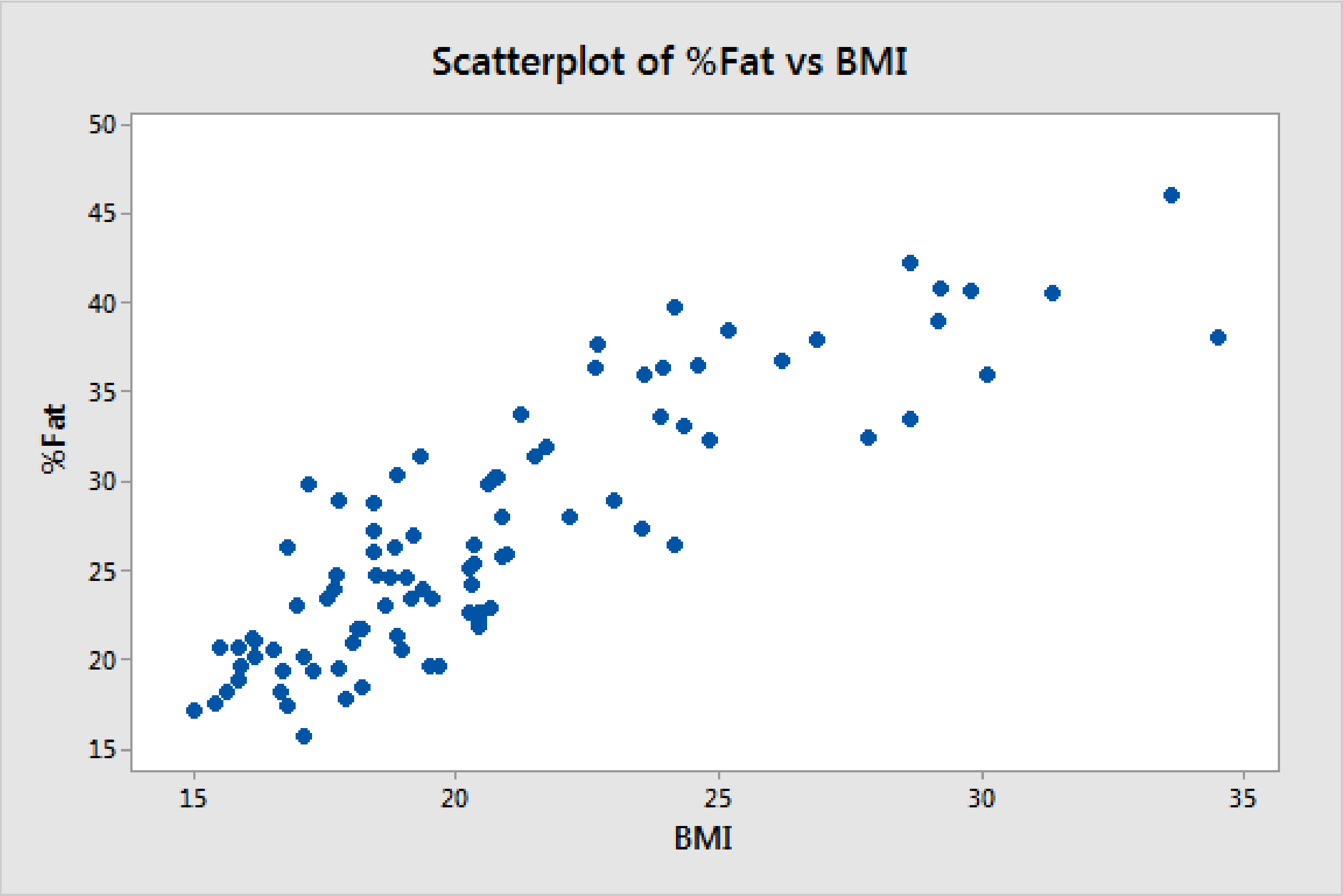
- Histograms
- Scatterplots
- Box plots

Histogram

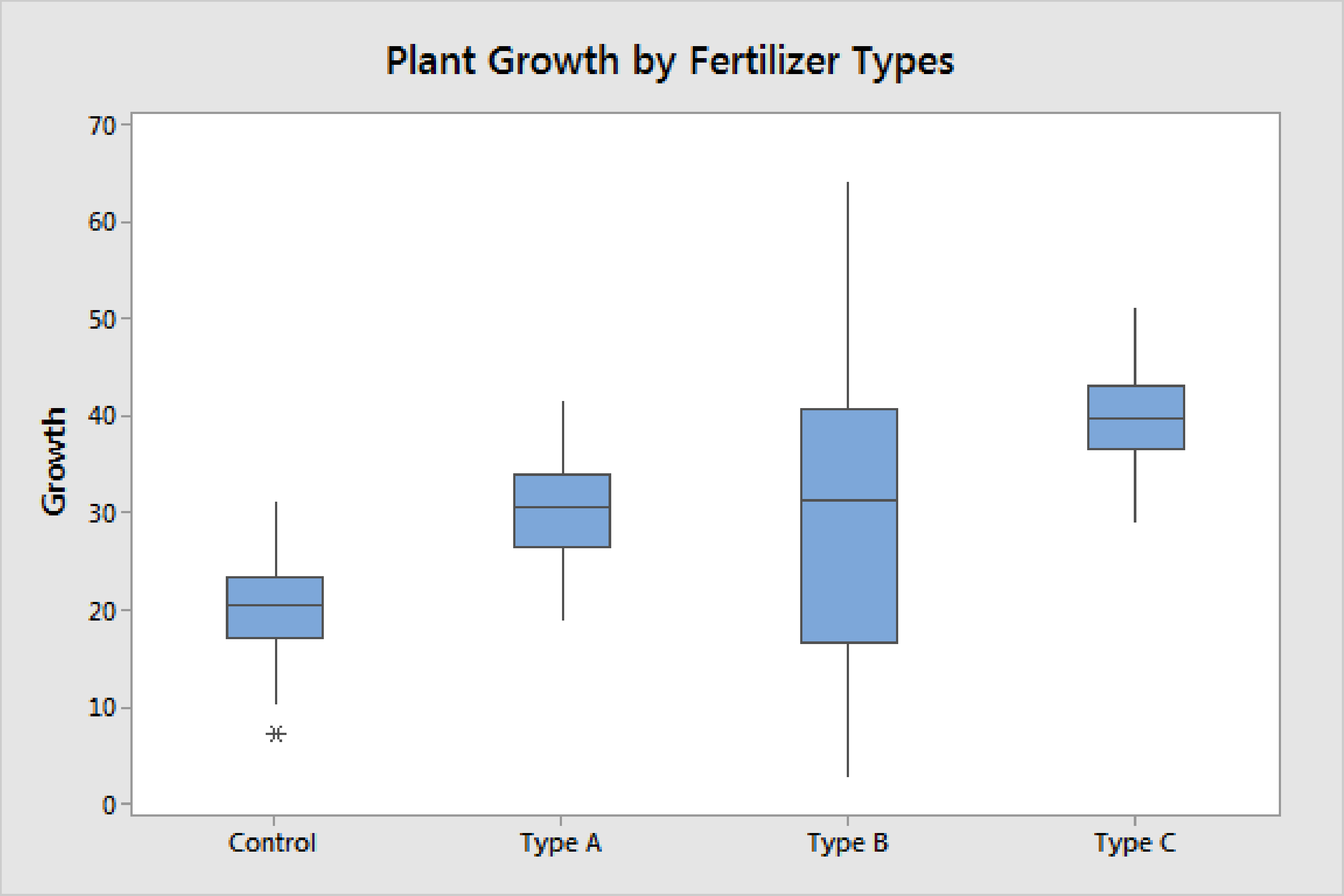


Histogram of %Fat
Frequency
16
14
12
10
8
6
4
2
0
20
25
30
35
40
45
%Fat

Scatterplot

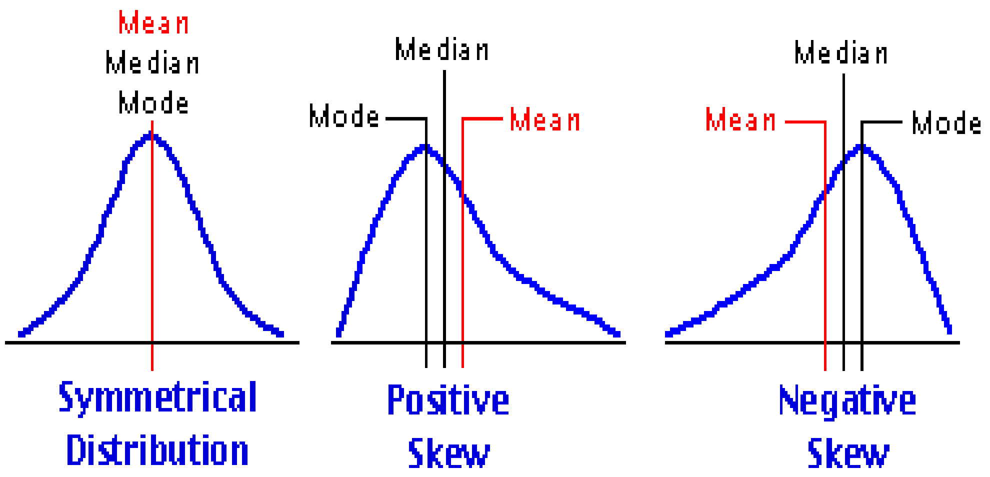


Boxplot


Plant Growth by Fertilizer Types
Growth
70
60
50
40
30
20
10
0
Control
Type A
Type B
Type C

Distributions

- A **positively skewed distribution** means that it has a long tail in the positive direction (a long right tail).
- For a positively skewed distribution, the mode is less than the median, which is less than the mean.
- In a positively skewed distribution, there are large positive outliers which will tend to "pull" the mean upward
- A **negatively skewed distribution** has a long tail in the negative direction (a long left tail)
- For the negatively skewed distribution, the mean is less than the median, which is less than the mode.
- In this case, there are large negative outliers which tend to "pull" the mean downward.

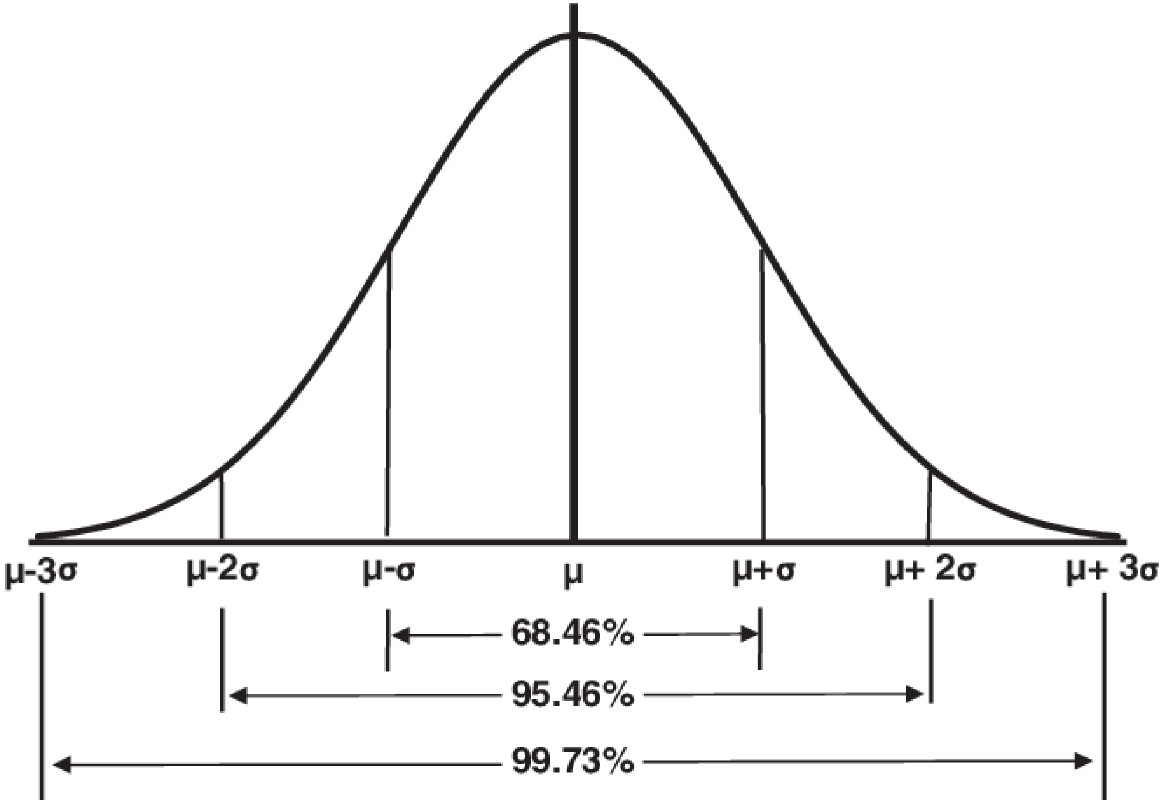


Normal distribution

1. Bell-shaped curve
2. Total area under the curve is equal to 1
3. Symmetrical

"Standard" Normal distribution

4. Population mean= 0
5. Population standard deviation= 1



Examples of normally distributed data

- Heights of people
- Size of things produced by machines
- Errors in measurements
- Blood pressure
- Marks on a test

Normal

Always report mean & standard deviation

- **Use parametric tests**
  - Two sample t-test
  - Paired t-test
  - One-way ANOVA

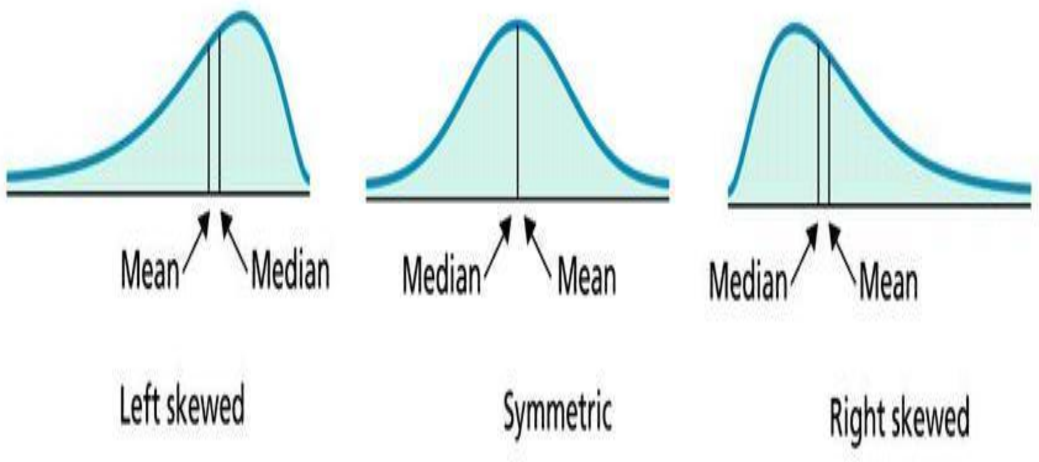
Non-normal Data

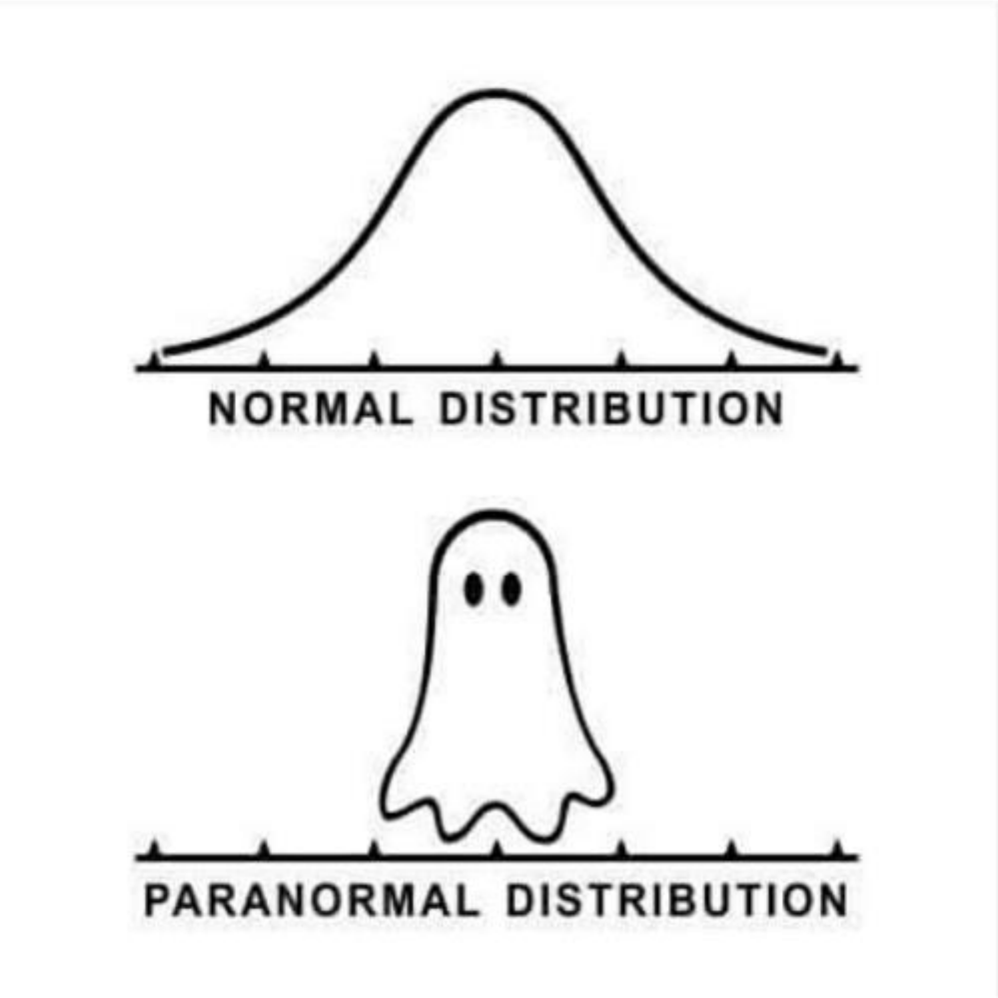
- **Lots of variables are non-normally distributed**
  - Cost
  - Length of stay
  - Duration of mechanical ventilation
  - Weight
  - Lab values (ALT, AST, serum glucose)

- **May not be reasonable to assume normality; when in doubt, go with a non-parametric test**
  - Conservative
  - Less powerful when data are normally distributed
  - More powerful if data are skewed (less influence of outliers)

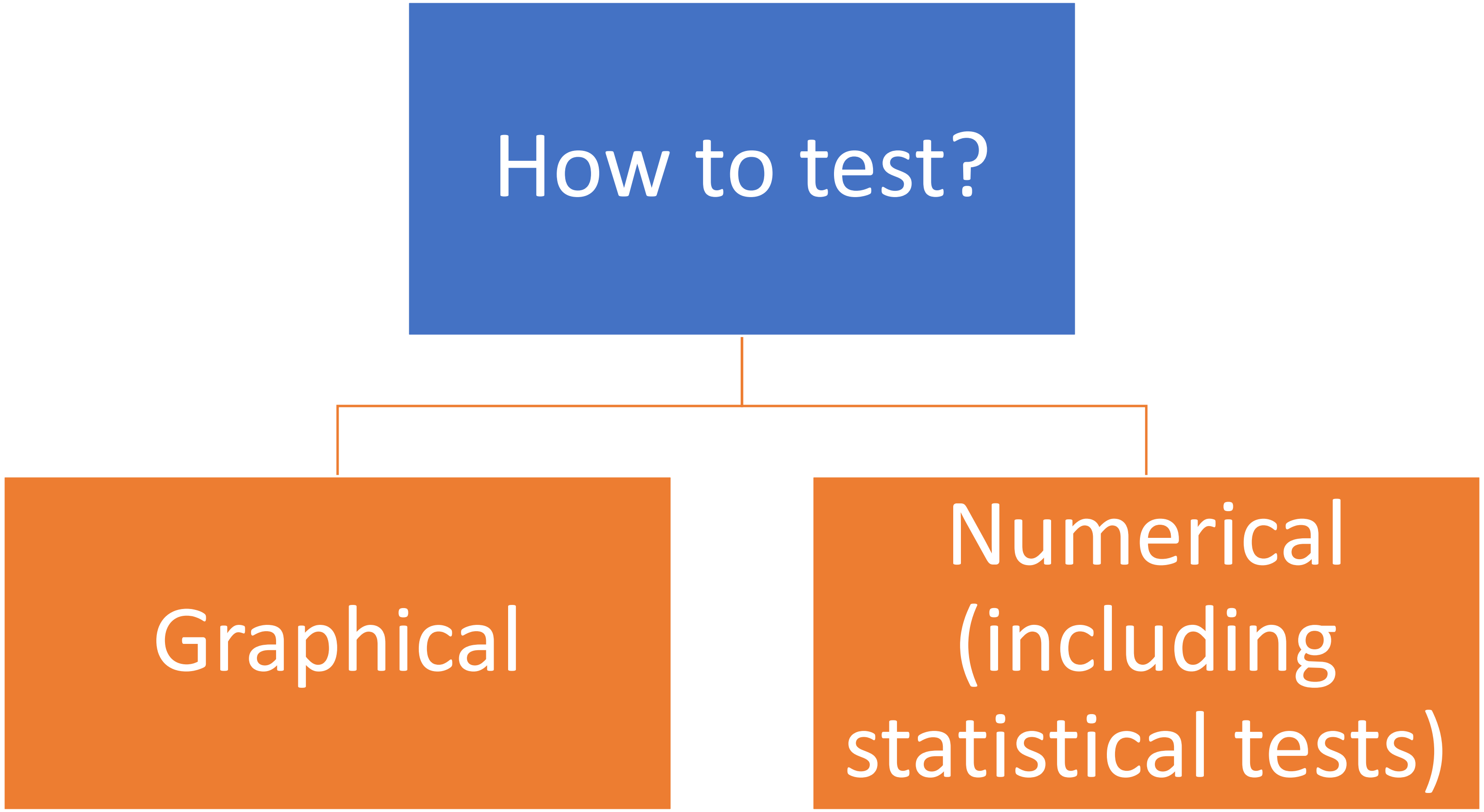
Non-normal

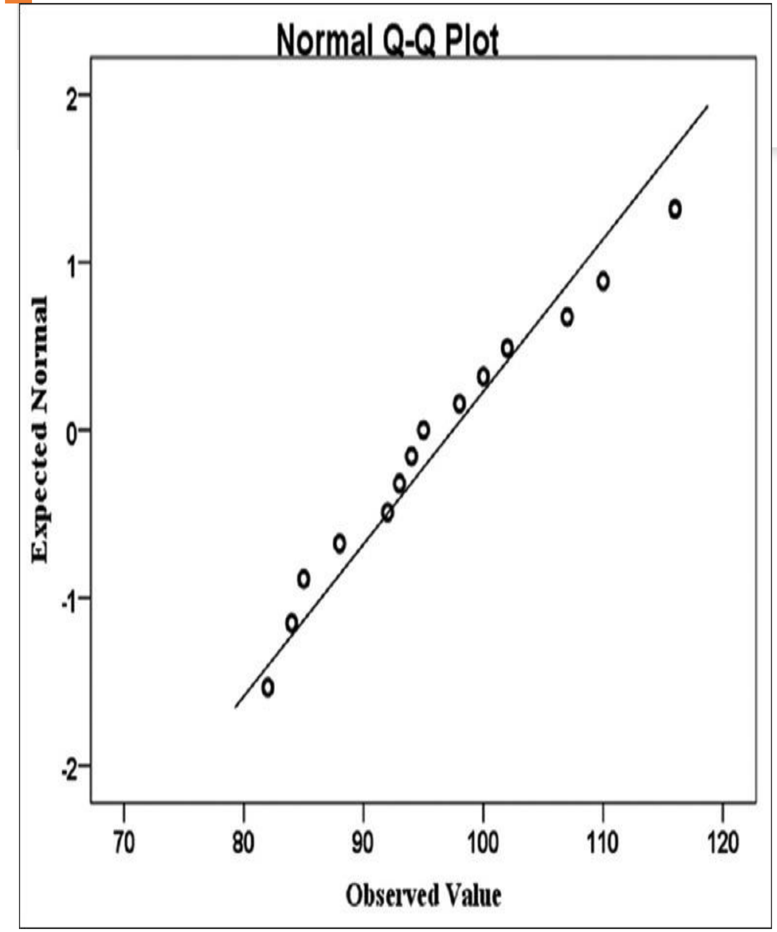
- **Skewed data**
  - Report median & interquartile range
  - Transform data (i.e., take log)
- **Use non-parametric tests**
  - Wilcoxon-Rank Sum test (Two sample t-test)
  - Wilcoxon-Signed rank test (Paired t-test)
  - Kruskal-Wallis test (One-way ANOVA)

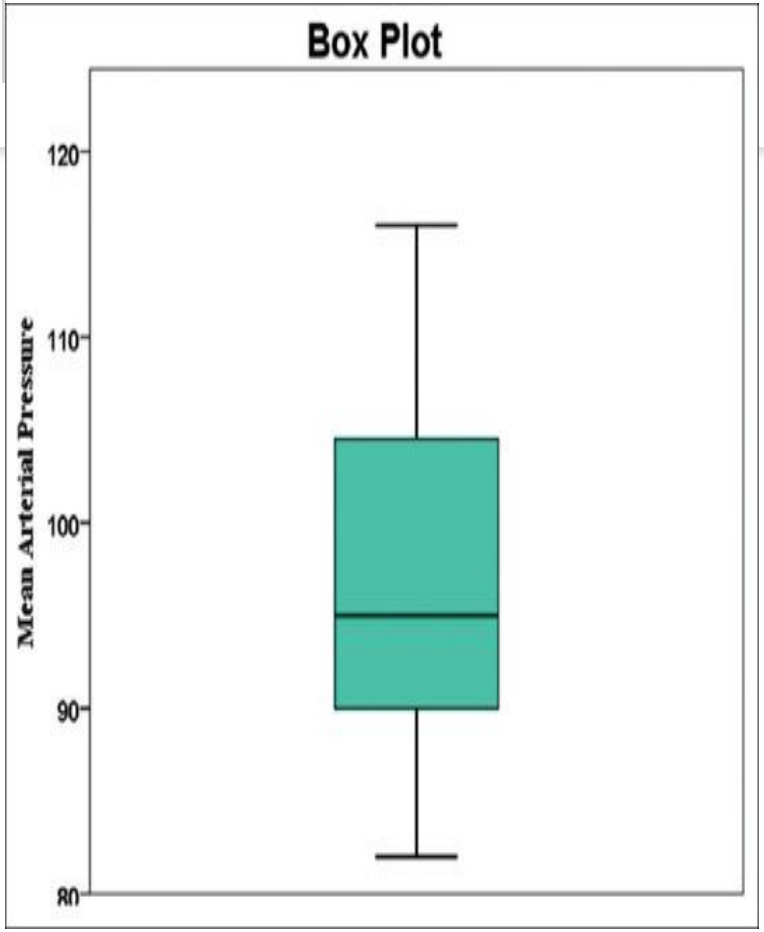


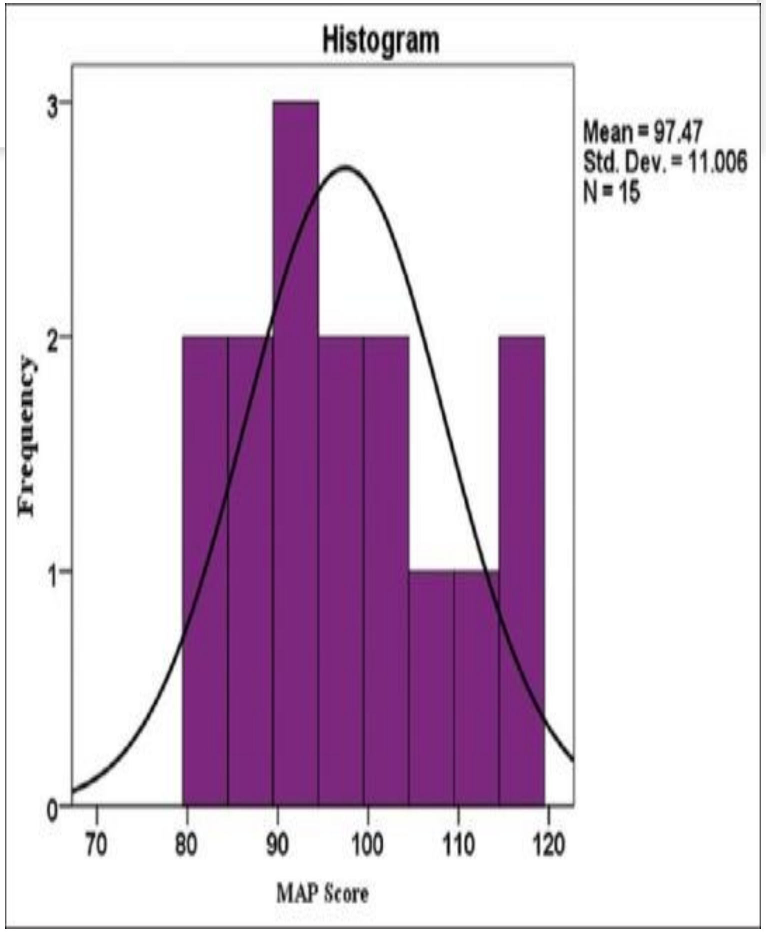

NORMAL DISTRIBUTION
PARANORMAL DISTRIBUTION

Testing the normality of data




Normal Q-Q Plot
Expected Normal
Observed Value
2
1
0
-1
-2
70
80
90
100
110
120


Box Plot
Mean Arterial Pressure
120
110
100
90
80


Histogram
Frequency
MAP Score
Mean = 97.47
Std. Dev. = 11.006
N = 15
3
2
1
0
70
80
90
100
110
120

Testing normality assumption

Null hypothesis (P> 0.05)

The values are sampled from a population that follows a normal distribution

Alternative hypothesis (P< 0.05)

The values are not sampled from a population that follow a normal distribution

Tests for normality

- Shapiro- Wilk
- Kolmogorov-Smirnov

Testing normality assumption

We can see from the table that for the "Beginner", "Intermediate" and "Advanced" Course Group the dependent variable, "Time", was normally distributed. How do we know this? If the **Sig.** value of the Shapiro-Wilk Test is greater than 0.05, the data is normal. If it is below 0.05, the data significantly deviate from a normal distribution.

Tests of Normality

	Course	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Time	Beginner	.177	10	.200*	.964	10	.827
	Intermediate	.166	10	.200*	.969	10	.882
	Advanced	.151	10	.200*	.965	10	.837

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Parametric vs Non-parametric tests

Parametric tests

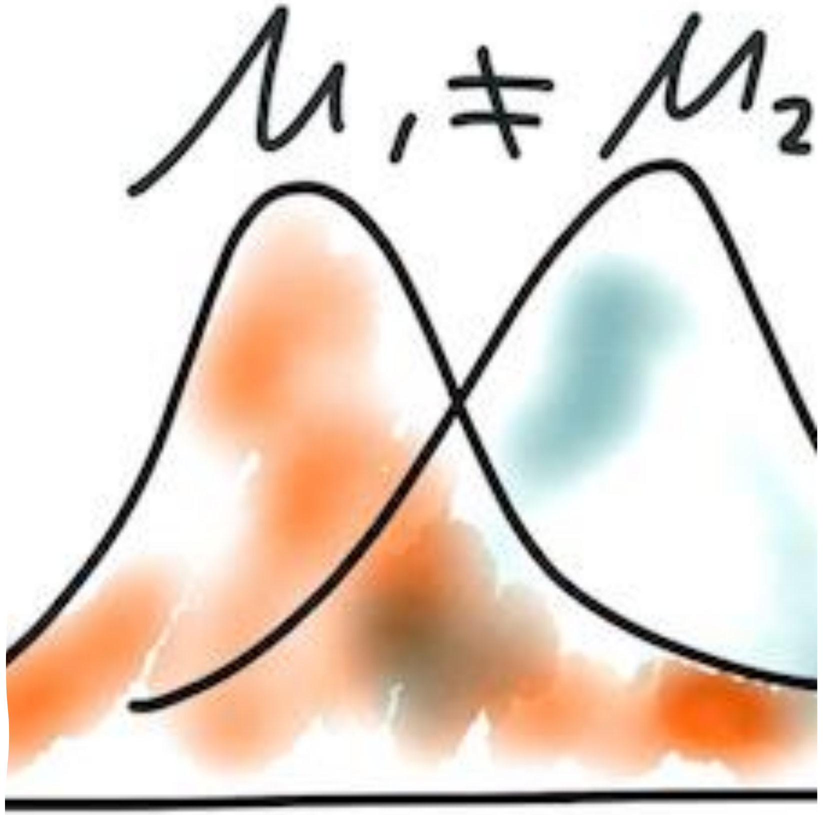
- Two-sample T-test
- Paired T-test
- One-way ANOVA

Non-parametric tests

- Mann-Whitney U test
- Kruskal-Wallis test
- Wilcoxon signed rank test

Two Sample T-test

- **Used to compare continuous outcomes between two groups**
- H_0: the population means of the two groups are equal (i.e., $\mu_1 = \mu_2$)
- H_A: the population means of the two groups are not equal (i.e., $\mu_1 \neq \mu_2$)
- **Outcome must be normally distributed (think bell-shaped curve / empirical rule)**
- **Groups should have similar variance (Levene's test)**
  - If so, use equal variance
  - If not, use unequal variance (*when in doubt, use this one*)



Two-sided vs. One-sided

- **Two sided tests are <u>most common</u>**
  - i.e., don't specify the direction of the alternative hypothesis
    - $\mu 1 \neq \mu 2$ (They are just different)

- **One sided tests are used when you specify the direction of the alternative hypothesis (e.g., > or <)**
  - P-values are smaller (1/2 as small) than two-sided
  - Examples: Superiority, non-inferiority, and non-equivalence trials use one-sided tests
  - Outside of these designs, *be super cautious of someone that reports a one-sided test without major justification*

When to use a t-test

T-test assumes your data:

- are independent (dependent for paired t-test)
- are (approximately) normally distributed.
- have a similar amount of variance within each group being compared (a.k.a. homogeneity of variance)

Reporting the results of a two- sample t-test

1. The means of both samples
2. The type of t test (one-vs. two-tailed)
3. The value of the t statistic
4. The degrees of freedom
5. The p value

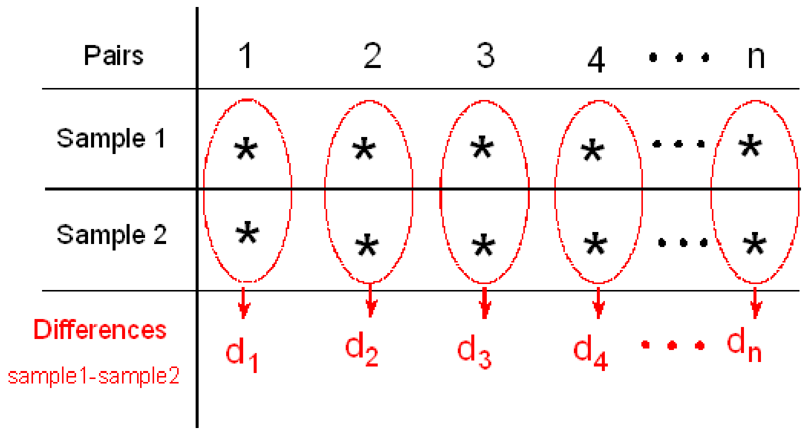
Interpretation of two sample t-test

- Data are mean ± standard deviation, unless otherwise stated. There were 20 male and 20 female participants. A two-sample t-test was run to determine if there were differences in engagement to an advertisement between males and females. The advertisement was more engaging to male viewers (5.56 ± 0.29) than female viewers (5.30 ± 0.39), a statistically significant difference of 0.26 (95% CI, 0.04 to 0.48), $t(35.055) = 2.365$, $p = .024$.

Paired samples t-test

- When the same subject is measured twice (e.g., pre/post, admission/discharge), samples are **<u>NOT</u>** independent
- Some degree of **correlation** exists between measurements from the same person
- Test statistic is based on the observed **differences between** two measurement occasions
- **Paired t-test:** Test whether the difference is equal to zero
  - H_o: Md = 0
  - H_a: Md ≠ 0

Pre	Post	Difference (pre – post)
20	10	10
30	21	9
45	27	18
39	27	12
41	42	-1
44	43	1



Interpretation of paired t-test

- A paired-samples t-test was used to determine whether there was a statistically significant mean difference between the distance ran when participants imbibed a carbohydrate-protein drink compared to a carbohydrate-only drink. Data are mean ± standard deviation, unless otherwise stated. Participants ran further when imbibing the carbohydrate-protein drink (11.302 ± 0.717 km) as opposed to the carbohydrate only drink (11.167 ± 0.726 km), a statistically significant increase of 0.136 (95% CI, 0.091 to 0.180) km, $t(19) = 6.352$, $p < .0005$

ANOVA

- **Analysis of Variance (ANOVA) is the statistical test used to compare more than 2 groups**

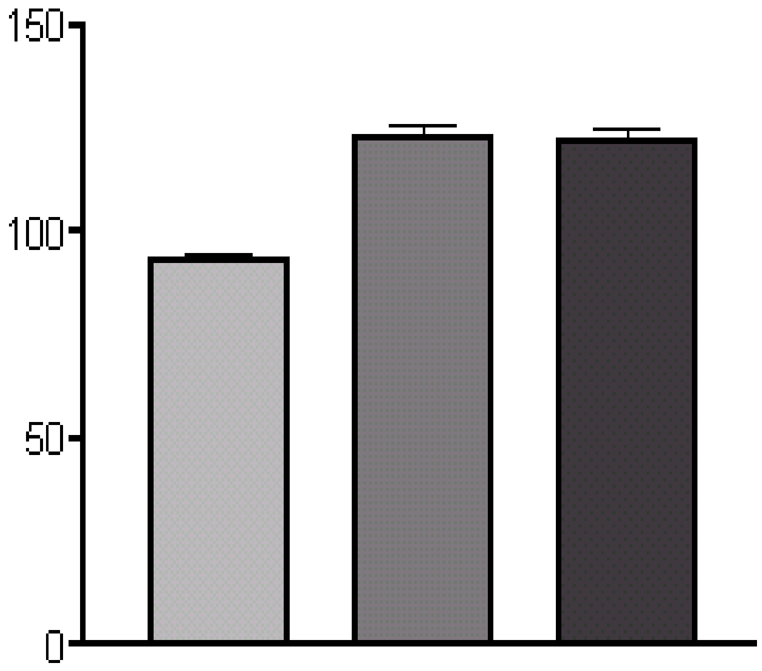
  H_0: μ1 = μ2 = μ3 =….= μn

  H_a: Not all groups are the same

- **Provides an "omnibus" test for <u>any difference</u> between the groups**

  F-test (F- distribution)

  P-value < 0.05 implies some means are different than others



When to use One-way ANOVA

One-way ANOVA assumes your data:

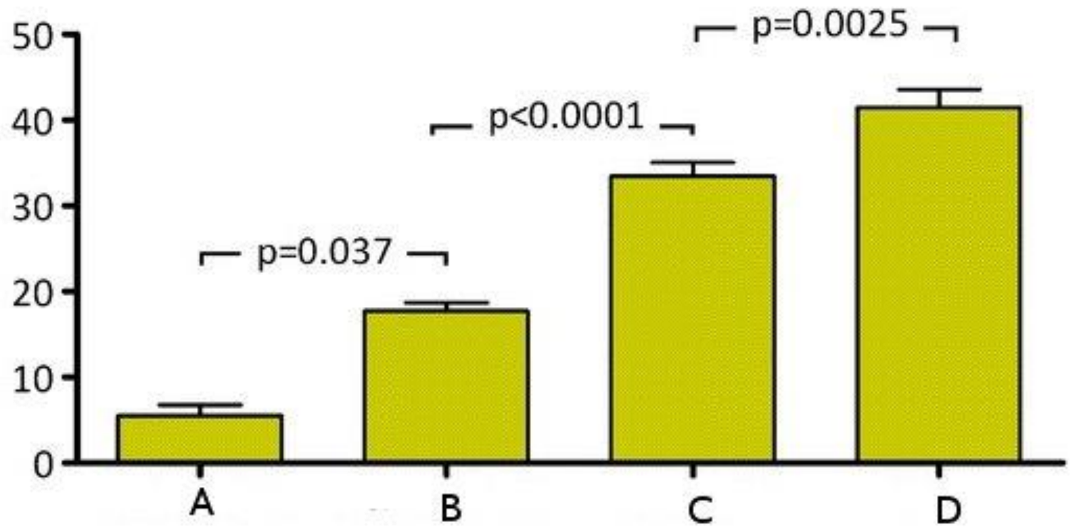
- are independent
- are (approximately) normally distributed.
- have a similar amount of variance within each group being compared (a.k.a. homogeneity of variance)

Post hoc testing

Multiple comparison procedures

- Tukey-Kramer, Dunnet's, Bonferroni, Scheffe's, Holm

- Be conscious of the number of pairwise comparisons



Reporting the Results of ANOVA

1. The means of all samples
2. The test name
3. The value of the F statistic
4. The degrees of freedom (there are two)
5. The p value

Interpretation of One-way ANOVA

A one-way ANOVA was conducted to determine if the ability to cope with workplace-related stress (CWWS score) was different for groups with different physical activity levels. Participants were classified into four groups: sedentary (n = 7), low (n = 9), moderate (n = 8) and high levels of physical activity (n = 7).

CWWS score was statistically significantly different between different physical activity groups, $F(3, 27) = 8.316$, $p < .001$. CWWS score increased from the sedentary (M = 4.15, SD = 0.77) to the low (M = 5.88, SD = 1.69), moderate (M = 7.12, SD = 1.57) and high (M = 7.51, SD = 1.24) physical activity groups, in that order. Tukey post hoc analysis revealed that the mean increase from sedentary to moderate (2.97, 95% CI [0.99, 4.96]) was statistically significant (p = .002), as well as the increase from sedentary to high (3.35, 95% CI [1.30, 5.40], p = .001), but no other group differences were statistically significant.

Non-Parametric Tests

- **Non-parametric**
  (no distribution is assumed)

- **Data must be ranked**

- **Brief 'How To'**
  - Order data from least to greatest
  - Replace actual variate value by its rank order (irrespective of study groups)
  - Perform analysis on the ranks

Parametric	Non-parametric
Two-Sample T-Test →	Wilcoxon Rank Sum Tests (Mann Whitney U)
ANOVA →	Kruskal-Wallis
Paired T-Tests →	Wilcoxon Signed Rank Test

Mann-Whitney U test

- A popular nonparametric test to compare outcomes between two independent groups
- The null and alternative hypotheses for Mann-Whitney U test is as follows:

  H_0: The two populations are equal versus

  H_1: The two populations are not equal.

- The procedure for the test involves pooling the observations from the two samples into one combined sample, keeping track of which sample each observation comes from, and then ranking lowest to highest

Interpretation

A Mann-Whitney U test was run to determine if there were differences in engagement score between males and females. Median engagement score was statistically significantly higher in males (7.12) than in females (5.38), $U = 218$, $z = -3.422$, $p = .001$.

Wilcoxon signed rank test

- A popular nonparametric test to compare outcomes between two paired or matched groups
- The null and alternative hypotheses for the Wilcoxon signed rank test is stated as follows:

  H_0: median difference between two groups = 0

  H_1: median difference between two groups ≠ 0

- It is based on difference scores, but in addition to analyzing the signs of the differences, it also takes into account the magnitude of the observed differences.

Interpretation

- A Wilcoxon signed-rank test was conducted to determine the effect of exercise on C-Reactive Protein (CRP). Ten participants were recruited to the study who each underwent a 6-month exercise-training programme. CRP concentration was measured pre-intervention and immediately post-intervention.
- Of the 10 participants recruited to the study, the exercise-training programme elicited a decrease in CRP concentration in nine participants post-intervention, whereas one participant's CRP concentration increased post-intervention. There was a statistically significant median decrease in CRP concentration (0.68 mg/L) from pre-intervention (4.33 mg/L) to post-intervention (3.65 mg/L), $z = 5.92$, $p = .001$.

Kruskal-Wallis test

- A popular nonparametric test to compare outcomes among more than two independent groups is the Kruskal Wallis test
- The null and alternative hypotheses for the Kruskal Wallis nonparametric test are stated as follows:

H_0: The k population medians are equal versus

H_1: The k population medians are not all equal

- The procedure for the test involves pooling the observations from the k samples into one combined sample, keeping track of which sample each observation comes from, and then ranking lowest to highest from 1 to N

Interpretation

- A Kruskal-Wallis test was conducted to determine if there were differences in coping from work related stress (CWWS) scores between groups that differed in their level of physical activity: the "sedentary" (n = 7), "low" (n = 9), "moderate" (n = 8) and "high" (n = 7) physical activity level groups.
- Median CWWS scores were statistically significantly different between the different levels of physical activity group, $\chi^2(3) = 14.468$, $p = .002$. Subsequently, pairwise comparisons were performed using Dunn's procedure. A Bonferroni correction for multiple comparisons was made with statistical significance accepted at $p < .0083$. This post hoc analysis revealed statistically significant differences in CWWS score between the sedentary (4.12) and moderate (7.10) ($p = .001$) and sedentary and high (7.47) ($p = .001$) physical activity groups, but not between the low physical activity group (5.50) or any other group combination.

Analysis Type	Example	Parametric Procedure	Nonparametric Procedure
Compare means between two distinct/independent groups	Is the mean systolic blood pressure (at baseline) for patients assigned to placebo different from the mean for patients assigned to the treatment group?	Two-sample t-test	Wilcoxon rank-sum test
Compare two quantitative measurements taken from the same individual	Was there a significant change in systolic blood pressure between baseline and the six-month follow-up measurement in the treatment group?	Paired t-test	Wilcoxon signed-rank test
Compare means between three or more distinct/independent groups	If our experiment had three groups (e.g., placebo, new drug #1, new drug #2), we might want to know whether the mean systolic blood pressure at baseline differed among the three groups?	Analysis of variance (ANOVA)	Kruskal-Wallis test

Tests

TABLE 3. The Characteristics of the Groups Along With the Most Relevant Comparative Results

	Group AA (n = 265)	Group Non-AA (n = 46)	*P*
Demographic data			
Age, y	11.6 (3–17)	12 (5–16)	ns
Sex (male/female), %	57/43	55/45	ns
Duration of symptoms, h	29.4 (6–96)	31.9 (2–240)	ns
Preoperative laboratory values			
Leukocytes ($\times 10^9$/L)	15.4 (5.5–34.1)	12 (3–23.9)	<0.001
C-reactive protein, mg/dL	40 (0.5–280)	26 (0.2–243)	<0.001
Neutrophils, %	81 (58–94)	70 (48.5–91)	<0.001
Scoring systems			
Alvarado score, mean ± SD	8.2 ± 1.5	6.7 ± 1.8	<0.001
PAS, mean ± SD	7.8 ± 1.4	6.6 ± 1.6	<0.001

Chi-Square Test

Wilcoxon Rank-Sum

Two-Sample T-Test

CHEAT SHEET

Number of Groups	Independent or Paired	Equality of Variances	Distribution	Test
2	Independent	Yes	Normal	Equal variances 2-sample t-test
2	Independent	No	Normal	Unequal variances 2-sample t-test
2	Independent	Yes	Non-normal	Mann Whitney / Wilcoxon rank sum
2	Paired	NA	Normal	Paired t-test
2	Paired	NA	Non-normal	Wilcoxon signed-rank test
3+	Independent	Yes	Normal	ANOVA
3+	Independent	Yes	Non-normal	Kruskal-Wallis test

Practice, Practice, Practice

http://www.artofstat.com/

https://statistics.laerd.com/

Pediatric Biostatistics Core

https://www.pedsresearch.org/research/cores/biostatistics-core/overview/