St. PAUL INSTITUTE
FOR REPRODUCTIVE HEALTH AND RIGHTS

Bivariate analysis

Mekitie Wondafrash , MD, PhD
St. Paul Institute for Reproductive Health and Rights

Learning objectives

By the end of this session, learners should be able to

- Select variables for bivariate analysis
- Analyzing relationships between variables
    - The Chi-square test
    - Generating cross-tabulations
    - The ‘means’ procedure / comparing means
    - The ‘correlations’ procedure

Bivariate analysis

Bivariate analysis is the second step in the analysis

- It is an analysis made to test the presence of a relationship between two variables
- It also could assess the presence of differences between two variables.
- Answers the question:
  - Is there a relationship or difference between the two variables?
- It is the initial step in hypothesis testing

Chi-square test

Chi-square test is based on the simple idea of comparing the frequencies you observe in certain categories to the frequencies you might expect to get in those categories by chance.

There are two different forms of the chi-square test:

- The multidimensional chi-square test, and
- The goodness of fit chi-square test.

- **The multidimensional chi-square:** test assesses whether there is a relationship between two categorical variables/test of independence
- **This goodness-of-fit test** compares the observed and expected frequencies in each category to test whether all categories contain the same proportion of values or whether each category contains a user-specified proportion of values.

Chi-square test

Assumptions

- Two or more unrelated categories in both variables
- Both variables should be categorical (i.e. nominal or ordinal) and consist of two or more groups
- Groups should be mutually exclusive

Chi-square test

Multidimensional

Sex of the child * Stunting Crosstabulation

			Stunting		Total
			Stunted	Not stunted	
Sex of the child	Male	Count	313	99	412
		% within Sex of the child	76.0%	24.0%	100.0%
		% within Stunting	42.4%	32.7%	39.6%
		% of Total	30.1%	9.5%	39.6%
	Female	Count	425	204	629
		% within Sex of the child	67.6%	32.4%	100.0%
		% within Stunting	57.6%	67.3%	60.4%
		% of Total	40.8%	19.6%	60.4%
Total		Count	738	303	1041
		% within Sex of the child	70.9%	29.1%	100.0%
		% within Stunting	100.0%	100.0%	100.0%
		% of Total	70.9%	29.1%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.519[a]	1	.004		
Continuity Correction[b]	8.117	1	.004		
Likelihood Ratio	8.647	1	.003		
Fisher's Exact Test				.003	.002
Linear-by-Linear Association	8.511	1	.004		
N of Valid Cases	1041				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 119.92.

b. Computed only for a 2x2 table

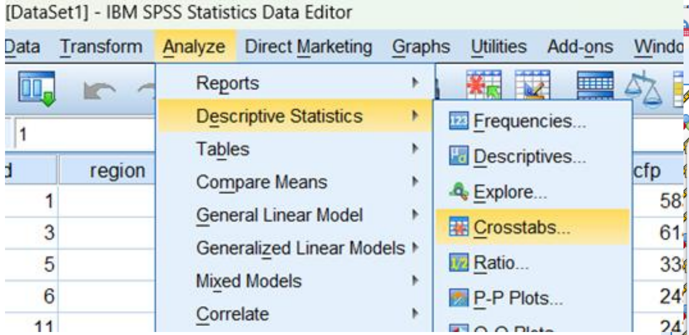
Interpretation-which p-value to take

- If the variables are of 2X2 table format, take the x^2 under the continuity correction
- If it is of 2X(>2) take the x^2 under the Pearson chi-Square
- If any cell in the table has <5 expected count, choose likelihood ratio or Fisher's Ex.
- If the independent variable is of ordinal type, choose linear by linear association.

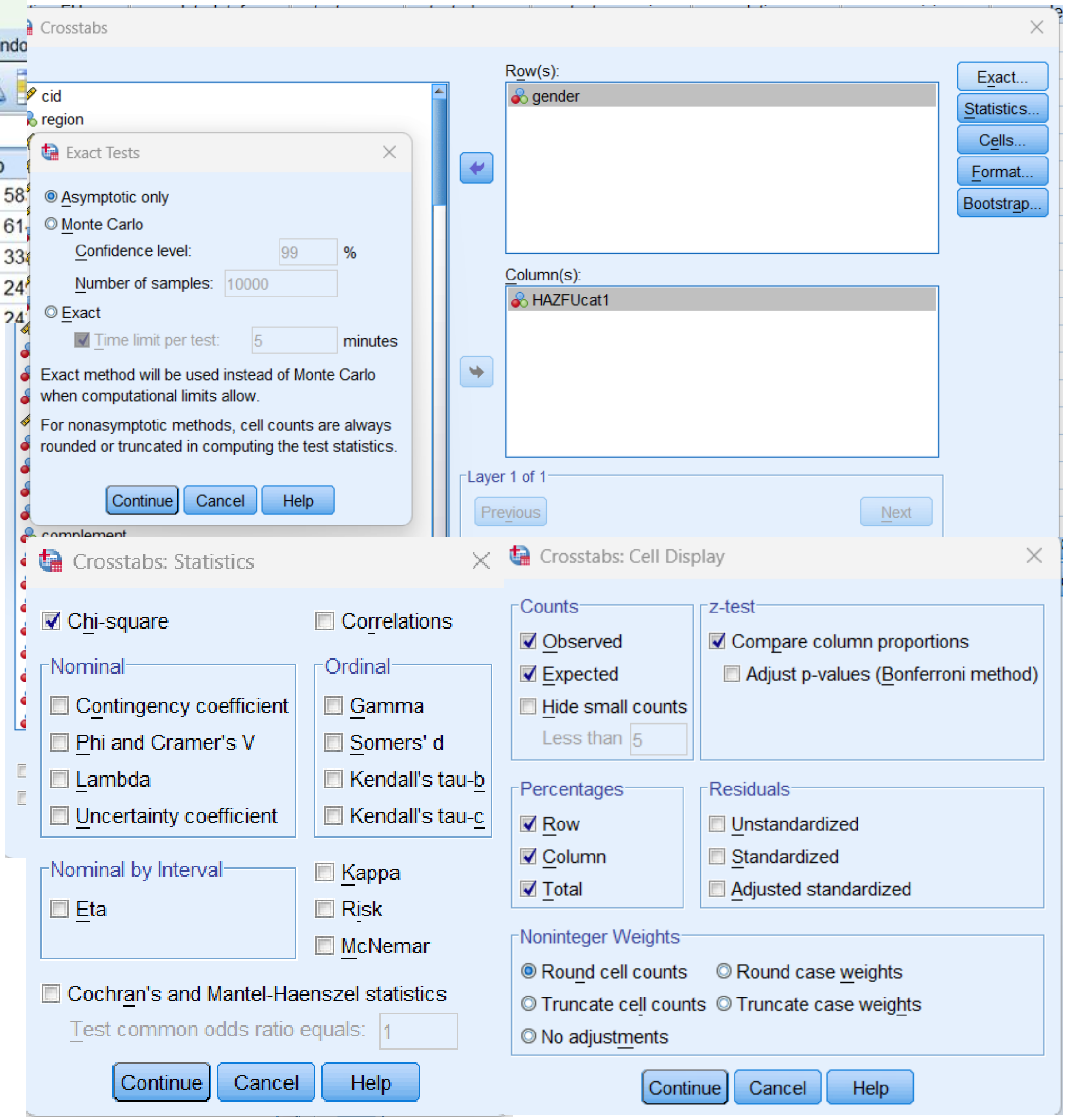
Crosstabulation (contingency table)

What does it do?

It shows :



- A wealth of information on the relationship between two or more variables
- No formula is needed
- The Chi-squared test often is used to accompany a crosstabulation



Crosstabulation (contingency table)

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex of the child * Stunting	1041	94.7%	58	5.3%	1099	100.0%

Sex of the child * Stunting Crosstabulation

			Stunting		Total
			Not stunted	stunted	
Sex of the child	Female	Count	204	425	629
		Expected Count	183.1	445.9	629.0
		% within Sex of the child	32.4%	67.6%	100.0%
		% within Stunting	67.3%	57.6%	60.4%
		% of Total	19.6%	40.8%	60.4%
	Male	Count	99	313	412
		Expected Count	119.9	292.1	412.0
		% within Sex of the child	24.0%	76.0%	100.0%
		% within Stunting	32.7%	42.4%	39.6%
		% of Total	9.5%	30.1%	39.6%
Total		Count	303	738	1041
		Expected Count	303.0	738.0	1041.0
		% within Sex of the child	29.1%	70.9%	100.0%
		% within Stunting	100.0%	100.0%	100.0%
		% of Total	29.1%	70.9%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.519[a]	1	.004		
Continuity Correction[b]	8.117	1	.004		
Likelihood Ratio	8.647	1	.003		
Fisher's Exact Test				.003	.002
Linear-by-Linear Association	8.511	1	.004		
McNemar Test				.000[c]	
N of Valid Cases	1041				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 119.92.

b. Computed only for a 2x2 table

c. Binomial distribution used.

Interpretation

- Asymptotic (2-sided); p=0.004
- There is a statistically significant relationship between gender and stunting

Correlation analysis

- A correlation analysis tests the relationship between two continuous variables in terms of:
  - How strong is the relationship?
  - In what direction the relationship goes?
- The strength of the relationship is given as a coefficient (r)- Pearson’s r
- “r” lies between -1 and 1.

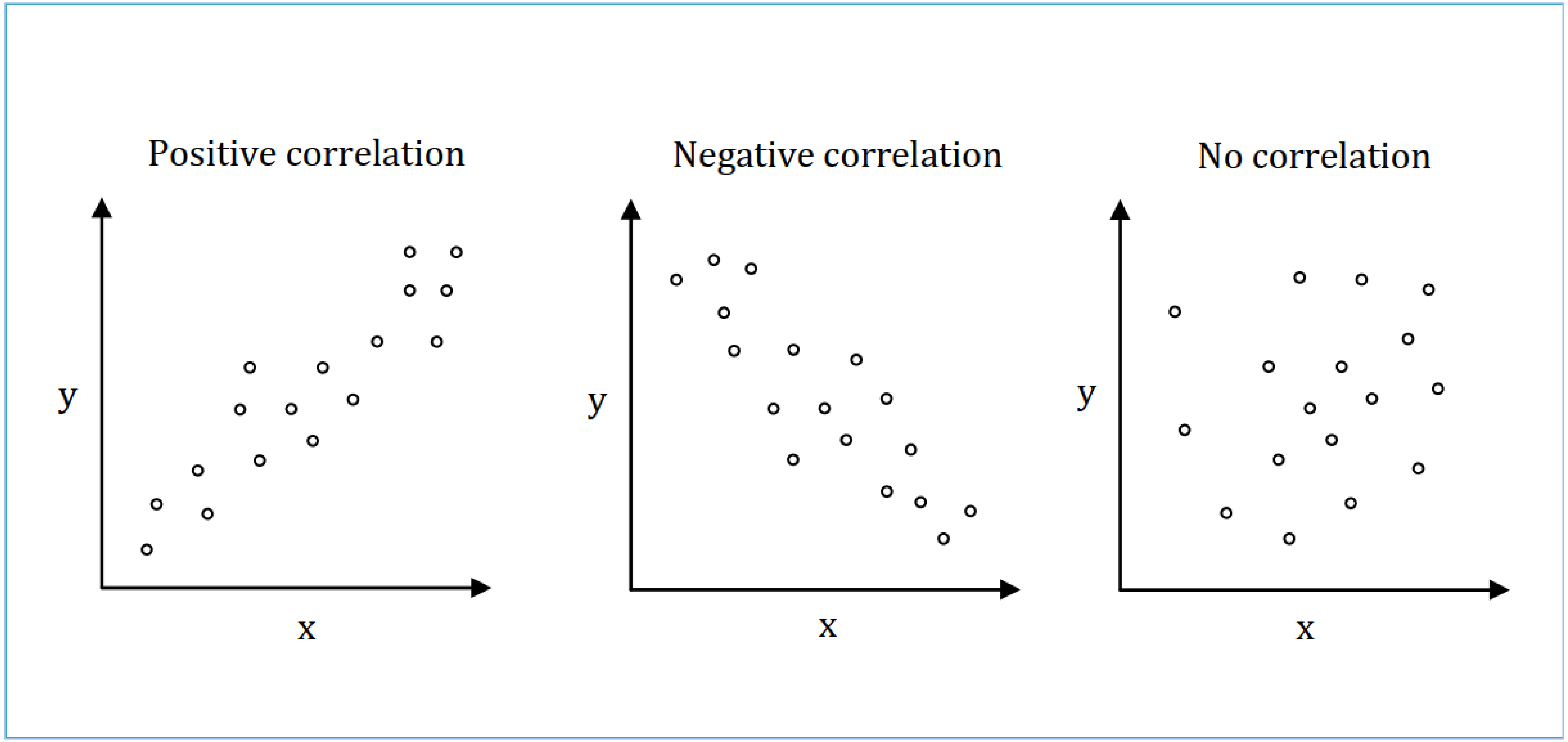
Number of variables

- Two or more

Scale of variable(s)

- Continuous (ratio/interval)

Correlation analysis



Correlation analysis

Assumption

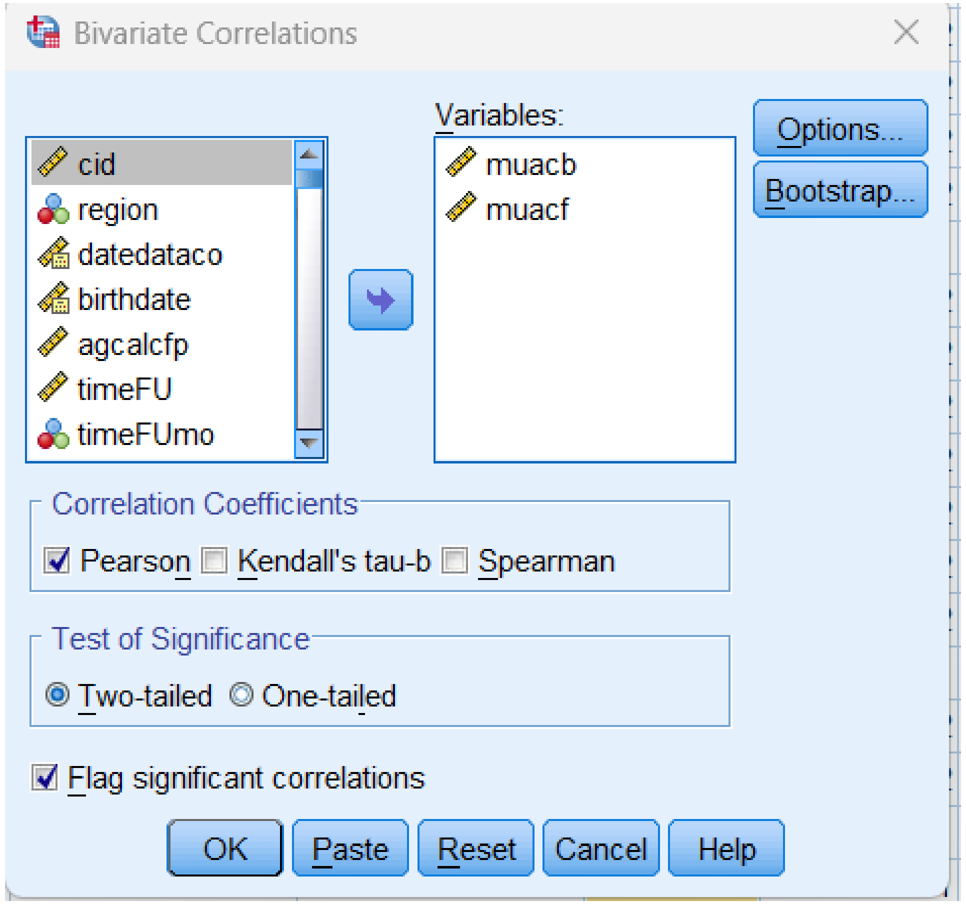
- Continuous variable
- Linear relationship between the two variables
- No outliers

Pearson’s correlation

- Pearson’s correlation is used to examine associations between variables (represented by continuous data) by looking at the direction and strength of the associations

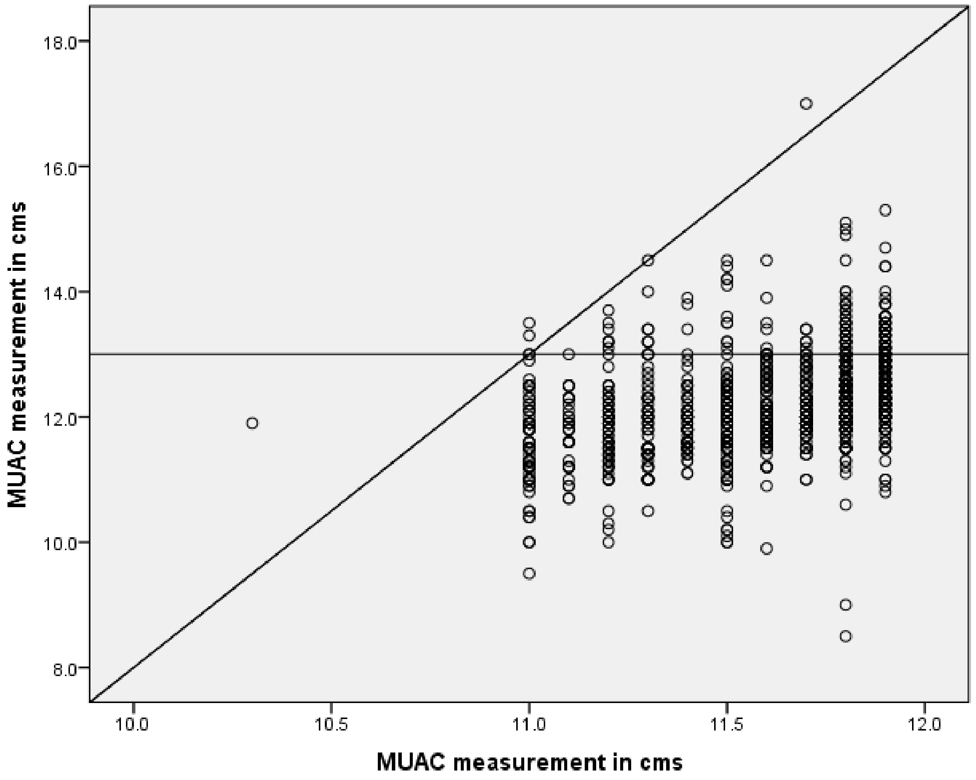
Correlation analysis

Analyze > correlate > bivariate …



- **Checking outlier**

→ "Graphs > Legacy Dialogs > Scatter/Dot…"



Correlation analysis

Correlations

		MUAC measurement in cms	MUAC measurement in cms
MUAC measurement in cms	Pearson Correlation	1	.373**
	Sig. (2-tailed)		.000
	N	1099	1099
MUAC measurement in cms	Pearson Correlation	.373**	1
	Sig. (2-tailed)	.000	
	N	1099	1099

**. Correlation is significant at the 0.01 level (2-tailed).

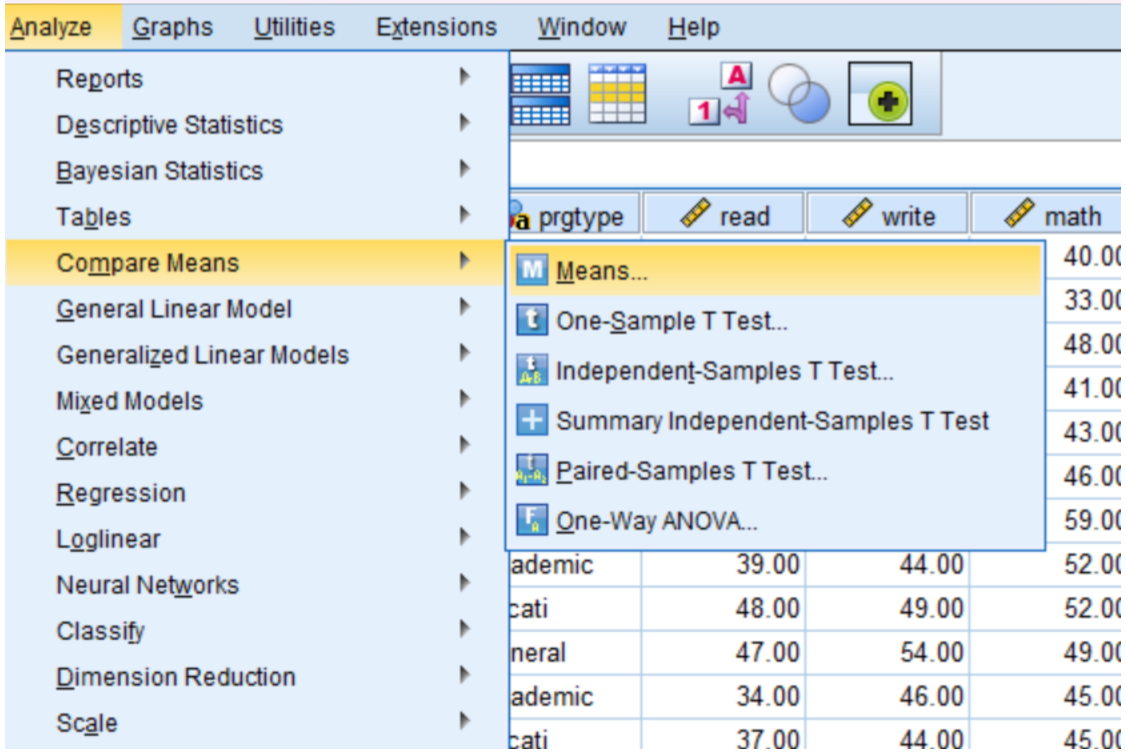
Strength		
Negative	**Positive**	
-1	1	**Perfect**
-0.9 to -0.7	0.7 to 0.9	**Strong**
-0.6 to -0.4	0.4 to 0.6	**Moderate**
-0.3 to -0.1	0.1 to 0.3	**Weak**
0	0	**Zero**

- There were small statistically significant positive correlations between MUACb and MUACf ($r = 0.373$; $p < 0.001$)
- Moderate strength

Compare means

Comparing means : parametric tests

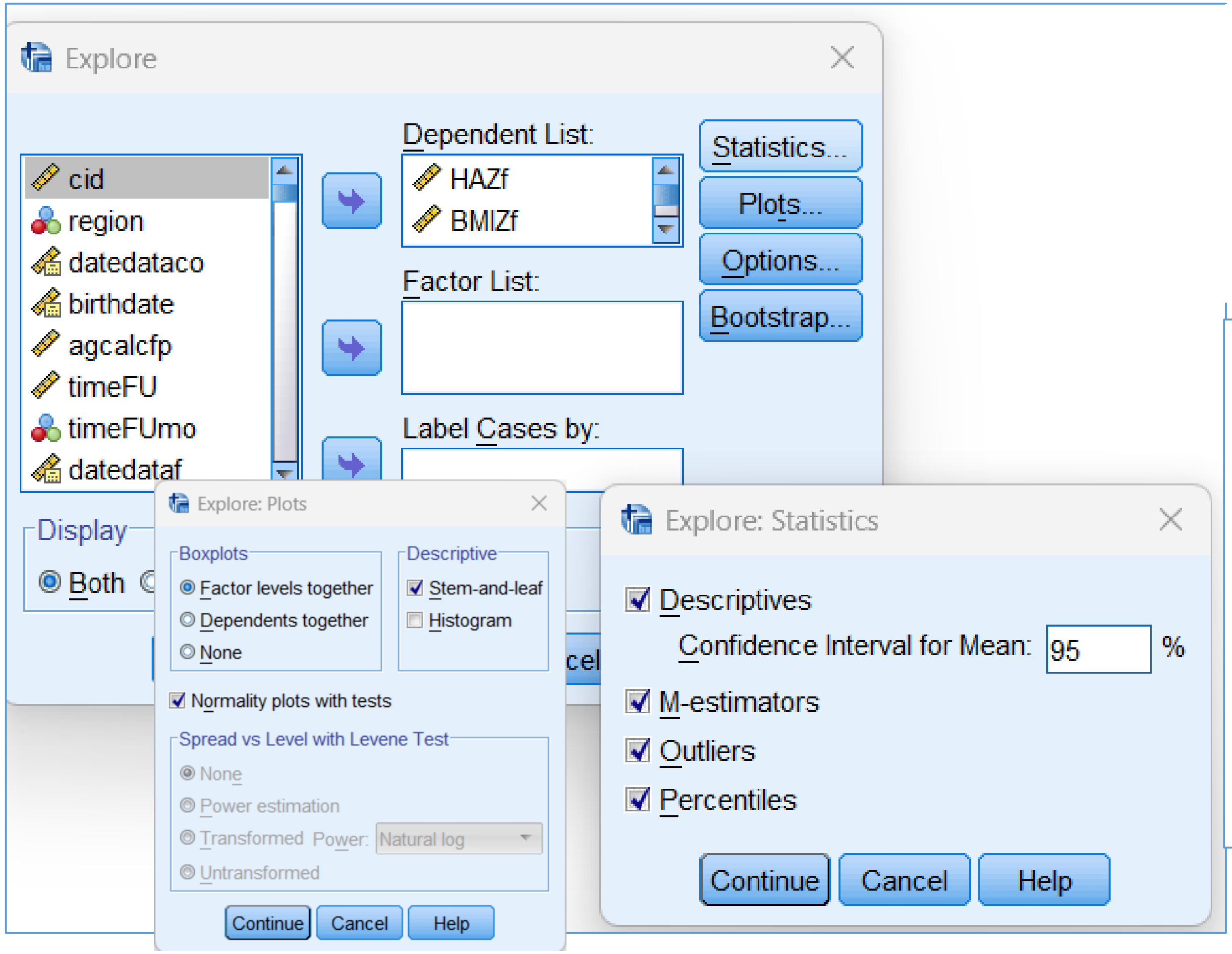
- T-test: One sample
- T-test: independent samples
- T-test: paired samples
- One-way ANOVA
- Repeated measures ANOVA



Test of normality

- Histogram
  - Bars should approximate the bell curve if it is normally distributed
  - Doesn't have to be perfect
- QQ plot
  - In this plot, the normal distribution is a straight line
  - If normally distributed, the points should cluster around the straight line
  - Should not have 'tails'
- Statistical test
  - Kolmogorov-Smirnov standard
  - Shapiro-Wilk for a small sample size
  - Sig. column (p-value) interpreted as :
    - If >0.05, from the normal distribution
    - If <0.05, not from a normal distribution

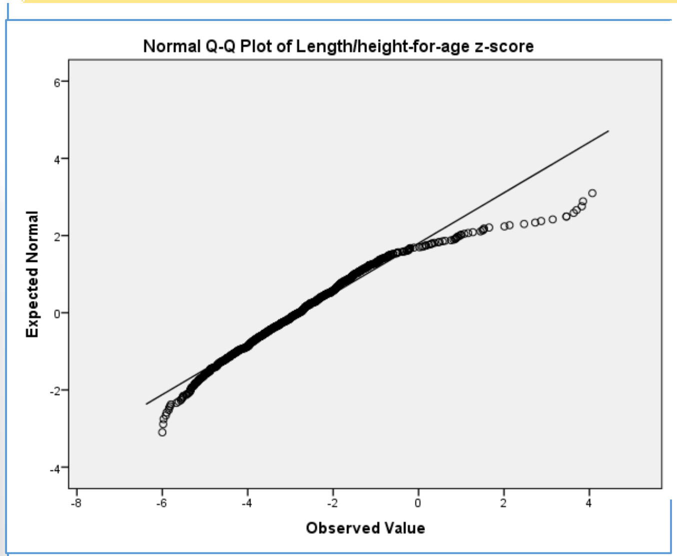
Tests of normality

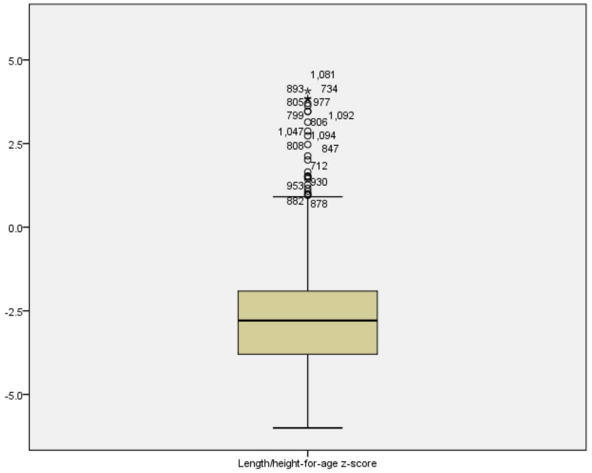


Tests of Normality

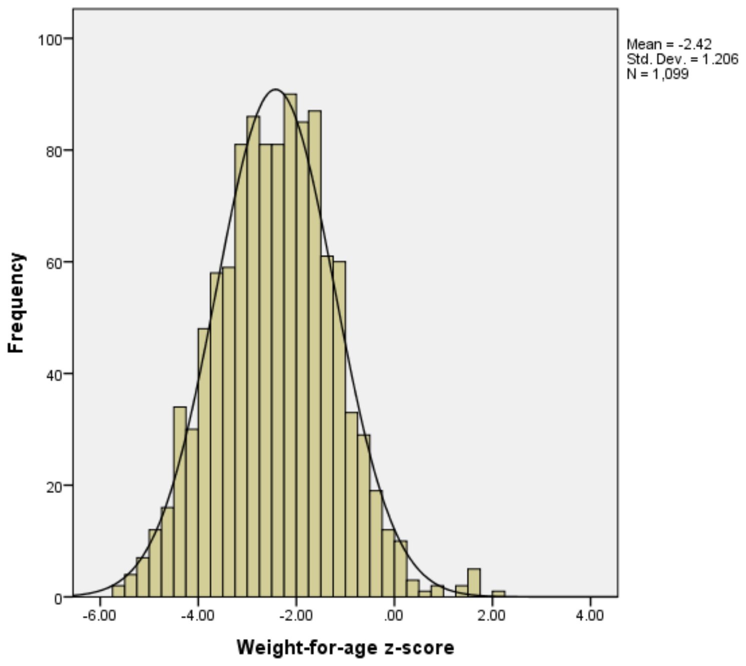
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Length/height-for-age z-score	.050	1025	.000	.958	1025	.000
BMI-for-age z-score	.036	1025	.003	.996	1025	.018
Weight-for-length/height z-score	.039	1025	.001	.996	1025	.015

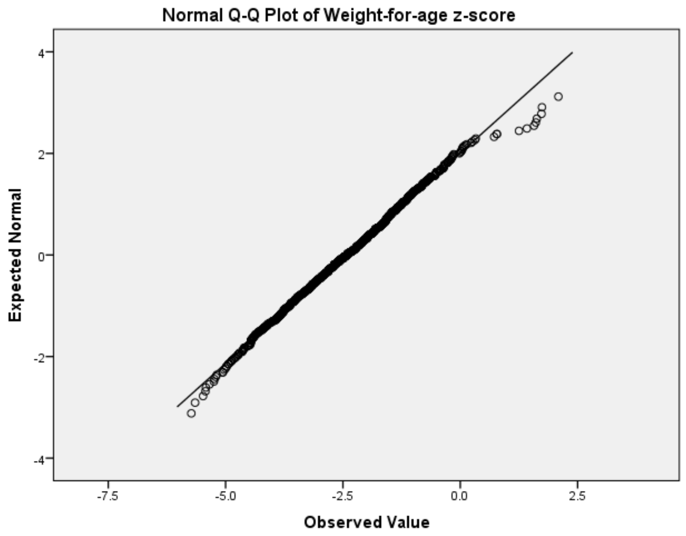
a. Lilliefors Significance Correction

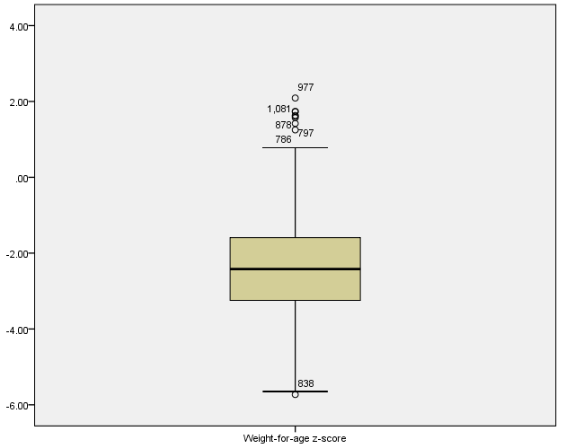




Test of normality







Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weight-for-age z-score	.016	1099	.200*	.996	1099	.003

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

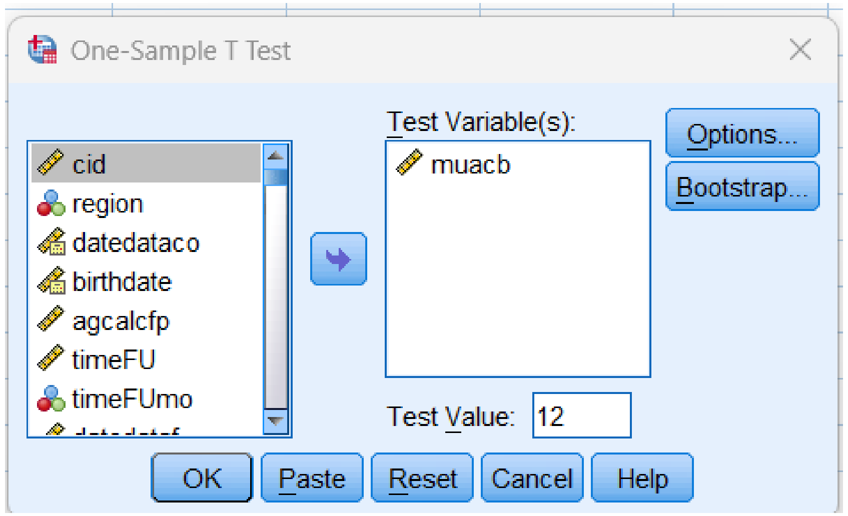
a. Lilliefors Significance Correction

One-Sample T Test

- Evaluates whether the mean on a test variable is significantly different from a constant (test value).
- Test value typically represents a neutral point. (e.g. midpoint on the test variable, the average value of the test variable based on past research)

One sample T-test

- A one-sample t-test is used to test differences between a sample mean and a hypothesized (null) value.
- Assume the mean of MUAC in the population is 12



- This table presents results of the one-sample t-test with the t value=47.5, df(n-1) =1098 .
- This is significant at both p=0.05 and p=0.01.
- We reject our null hypothesis of no difference

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
MUAC measurement in cms	1099	11.593	.2838	.0086

One-Sample Test

	Test Value = 12					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
MUAC measurement in cms	-47.496	1098	.000	-.4066	-.423	-.390

Independent-Sample T test

- Evaluates the difference between the means of two independent (unrelated) groups.
- Also called "Between groups T test"
- Hypothesis
  - Ho: $\mu_1 = \mu_2$
  - H1: $\mu_1 \neq \mu_2$

- Number of variables
  - One independent (x)
  - One dependent (y)
- Scale of variable(s)
  - Independent: categorical with two values (binary)
  - Dependent: continuous (ratio/interval)

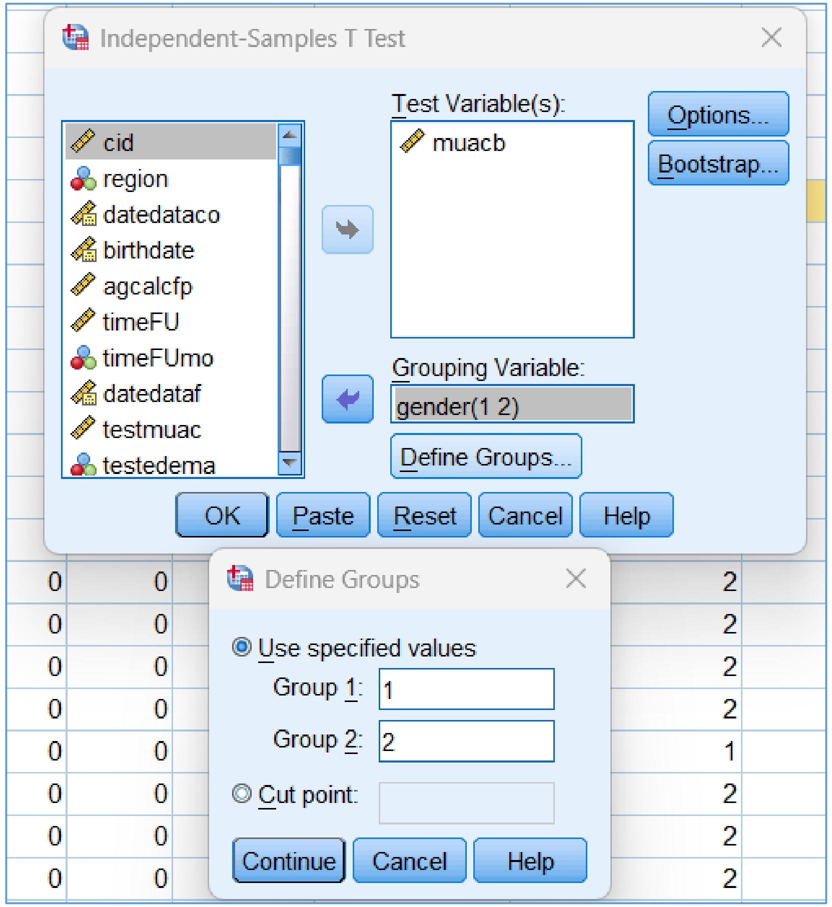
Independent Samples T-Test

Assumption

- Continuous dependent variable
- Two unrelated categories in the independent variable
- No outliers (normally distributed)

Independent Samples T-Test

Analyze → Compare Means → Independent-Samples T Test



Group Statistics

	Sex of the child	N	Mean	Std. Deviation	Std. Error Mean
MUAC measurement in cms	Male	445	11.598	.2813	.0133
	Female	654	11.590	.2856	.0112

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
MUAC measurement in cms	Equal variances assumed	.670	.413	.458	1097	.647	.0080	.0174	-.0262	.0422
	Equal variances not assumed			.459	963.026	.646	.0080	.0174	-.0261	.0421

Interpretation

- Equal variance assumed.
- The t value is 0.458 with df =1097(n1+n2-2).
- This is not significant (p= 0.647). Given alpha level = 0.05
- Accept the null hypothesis of no difference between the means

Paired-Sample T test

- Evaluates whether the mean of the difference between the paired variables is significantly different than zero.
- Applicable to
  - Repeated measures and
  - Matched subjects.
- Aka "Within subject T test" "Repeated measures T test".
- Hypothesis
  - Ho: $\mu_d = 0$
  - H1: $\mu_d \neq 0$

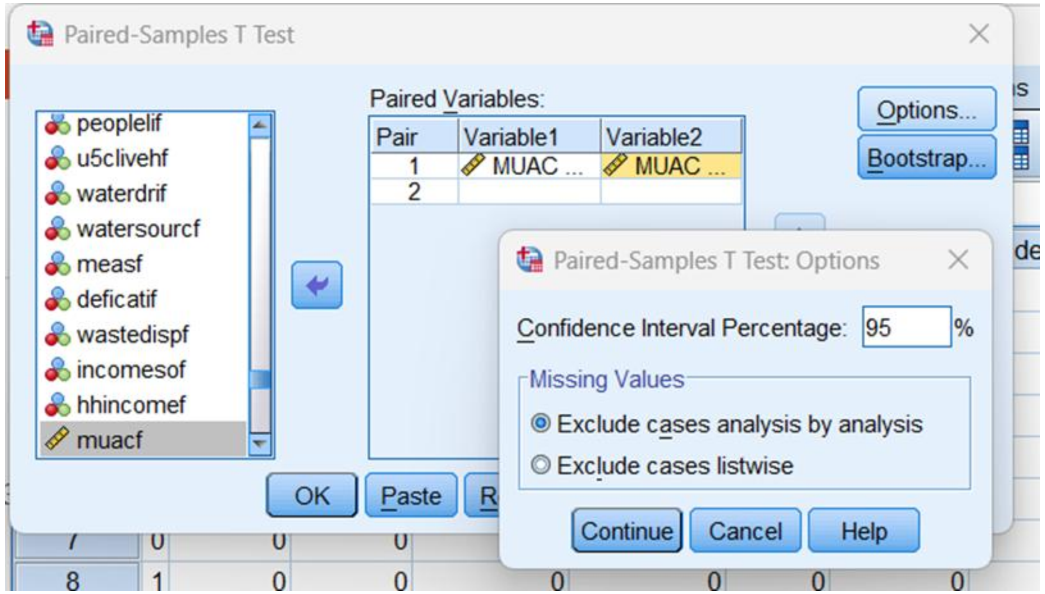
- Number of variables
  - Two (reflecting repeated measurement points)
- Scale of variable(s)
  - Continuous (ratio/interval)

Paired sample T-test

Assumption

- Continuous variable
- Two measurement points
- Normal distribution
- No outliers in the comparison between the two measurement points

Analyze → Compare Means → Paired-Samples T Test



Paired sample T-test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	MUAC measurement in cms	11.593	1099	.2838	.0086
	MUAC measurement in cms	12.185	1099	.7676	.0232

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	MUAC measurement in cms & MUAC measurement in cms	1099	.373	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	MUAC measurement in cms - MUAC measurement in cms	-.5919	.7121	.0215	-.6340	-.5498	-27.557	1098	.000

Interpretation

- The 95% confidence interval for the mean difference in presented (does not include zero)
- The t value for this test is -27.55 with 15 df(np-1)=1098.
- P is .000 (or reported as <.001)
- We can conclude that MUAC at baseline < at the end line

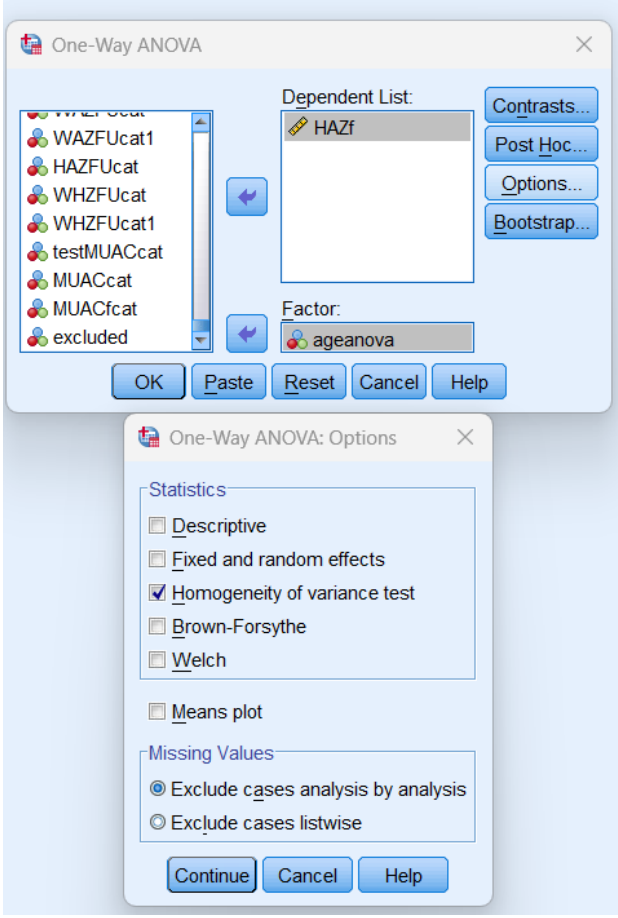
One-Way Analysis of Variance (one-way ANOVA)

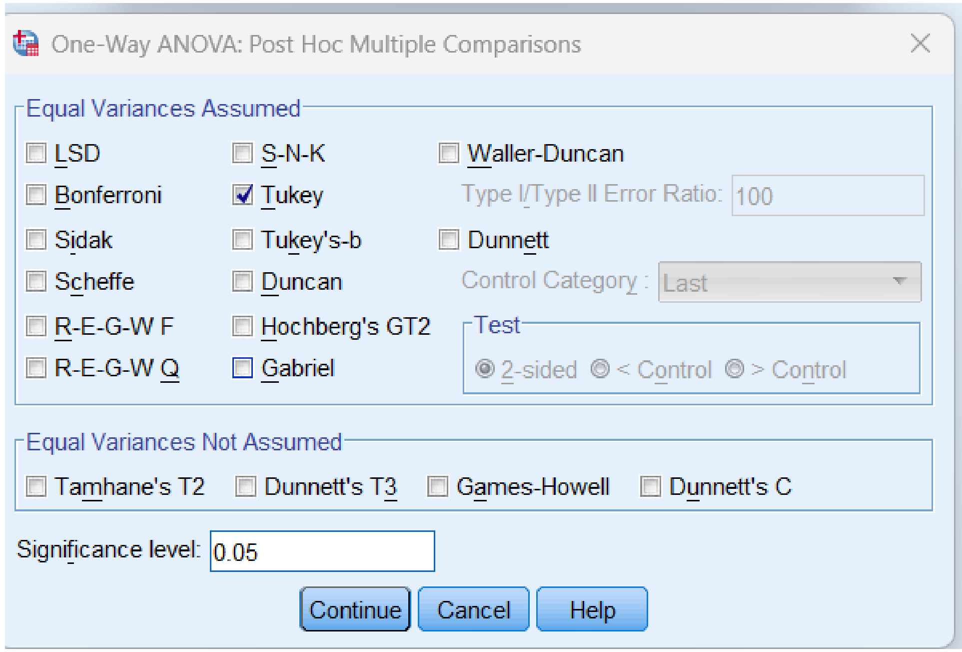
- It is very similar to the independent samples t-test but with more than two categories in the independent variable
- Number of variables
  - One independent(x)
  - One dependent (y)
- Scale of variable(s)
  - Independent: categorical (nominal/ordinal)
  - Dependent: continuous (ratio/interval)

Assumptions

- Continuous dependent variable
- Two or more unrelated categories in the independent variable
- No outliers

One-Way Analysis of Variance (one-way ANOVA)





One-Way Analysis of Variance (one-way ANOVA)

Analyze > Compare Means > One-Way ANOVA …

Test of Homogeneity of Variances

Length/height-for-age z-score

Levene Statistic	df1	df2	Sig.
1.868	2	1031	.155

Variance of the three group is not different (p>0.05)

ANOVA

Length/height-for-age z-score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	81.718	2	40.859	16.260	.000
Within Groups	2590.765	1031	2.513		
Total	2672.483	1033			

There was statistically significant group differences in level of LFA among the age groups, p <0.001.

Homogeneous Subsets

Length/height-for-age z-score

Tukey HSD[a,b]

ageanova	N	Subset for alpha = 0.05: 1	2
>24 months	253	-3.0075	
12-23 months	398	-2.8597	
<12 months	383		-2.3472
Sig.		.454	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 330.539.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Length/height-for-age z-score

Tukey HSD

(I) ageanova	(J) ageanova	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval: Lower Bound	Upper Bound
<12 months	12-23 months	.51252*	.11347	.000	.2462	.7788
	>24 months	.66036*	.12843	.000	.3589	.9618
12-23 months	<12 months	-.51252*	.11347	.000	-.7788	-.2462
	>24 months	.14784	.12746	.478	-.1513	.4470
>24 months	<12 months	-.66036*	.12843	.000	-.9618	-.3589
	12-23 months	-.14784	.12746	.478	-.4470	.1513

*. The mean difference is significant at the 0.05 level.

- The level of LFA for <12 months was higher than 12-23 and >24 months (p<0.001)
- The level of LFA for 12-23 months was lower than <12 months (p<0.001)
- The level of LFA for >24 months was lower than <12 months (p<0.001)

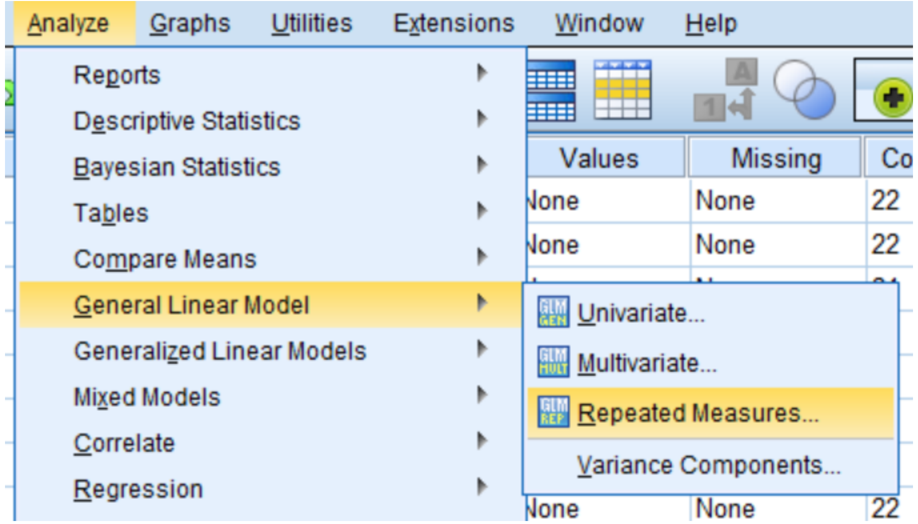
One-way ANOVA -post hoc comparisons

Post hoc comparisons

- Comparisons of group means are made after data have been collected.
- They do not assume any prior hypotheses.
- The more number of group means comparison, the higher the overall type-I error
- Control for multiple pairwise comparisons is needed like using Tukey's test
- Requires a larger difference to declare significance compared to if no adjustment was used

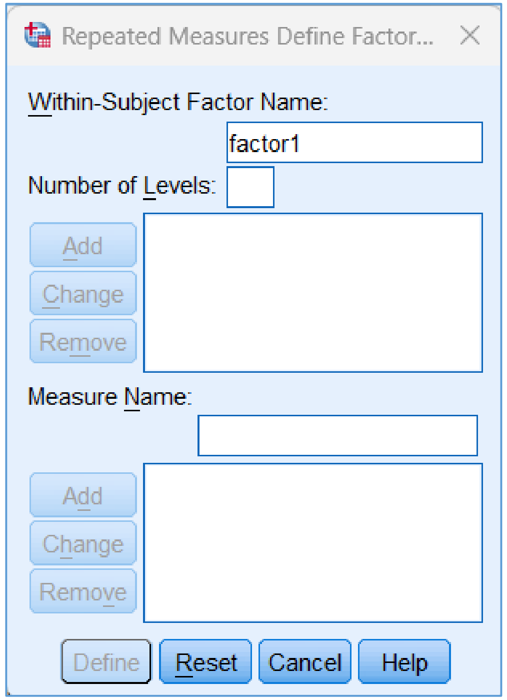
Repeated measures ANOVA

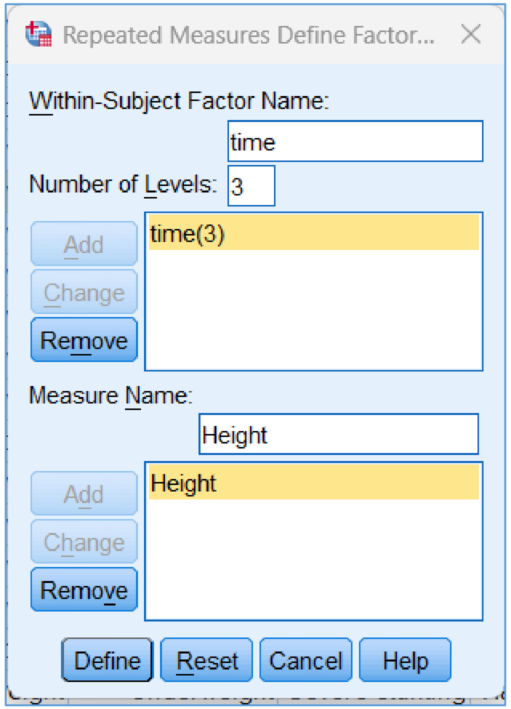
- A repeated measures ANOVA is also referred to as a within-subjects ANOVA or ANOVA for correlated samples
- Used to compare three or more group means where the participants are the same in each group.
- This usually occurs in two situations:
  - When participants are measured multiple times
  - When participants are subjected to more than one condition/trial

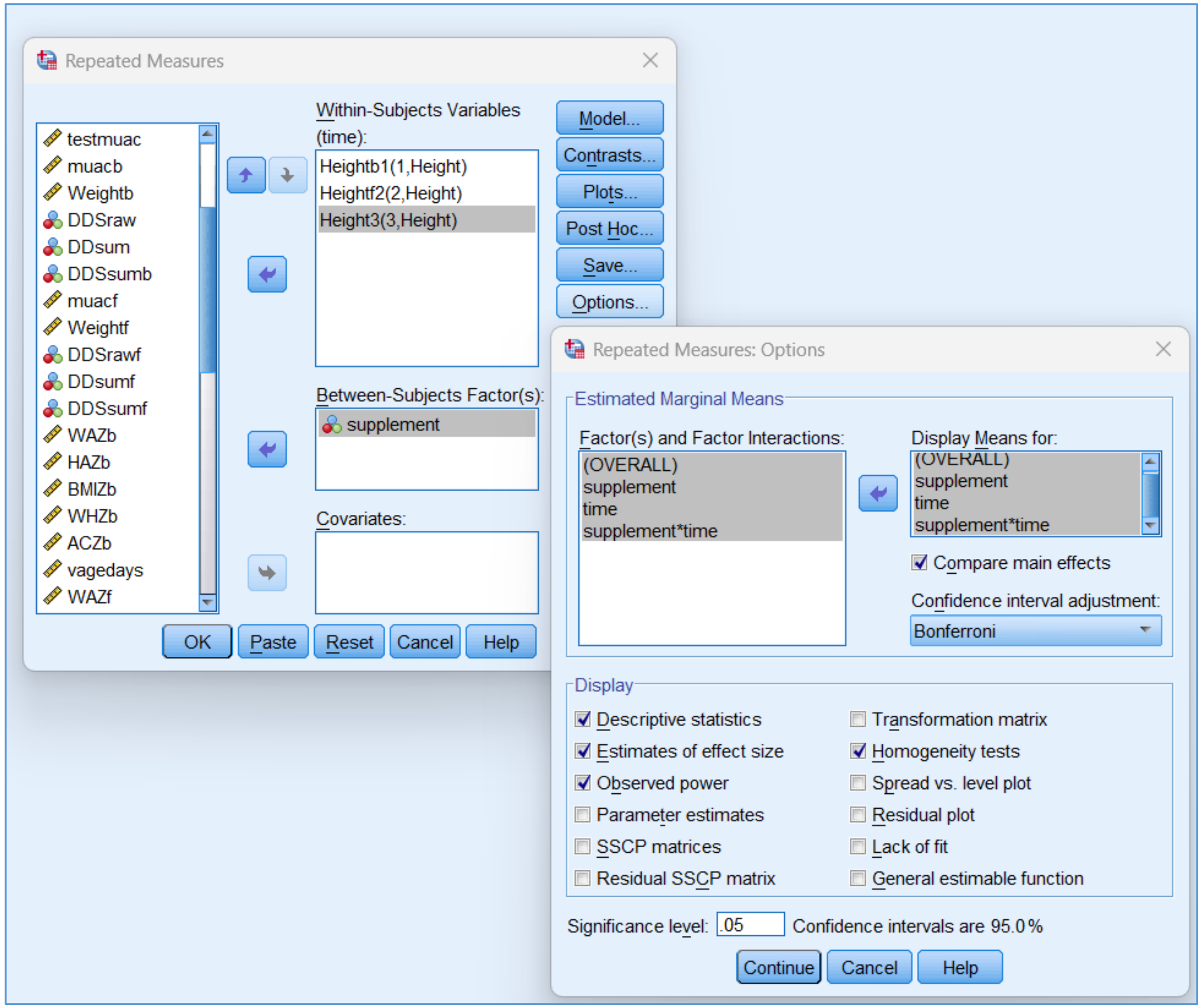


Use = “RM_anova.sav”

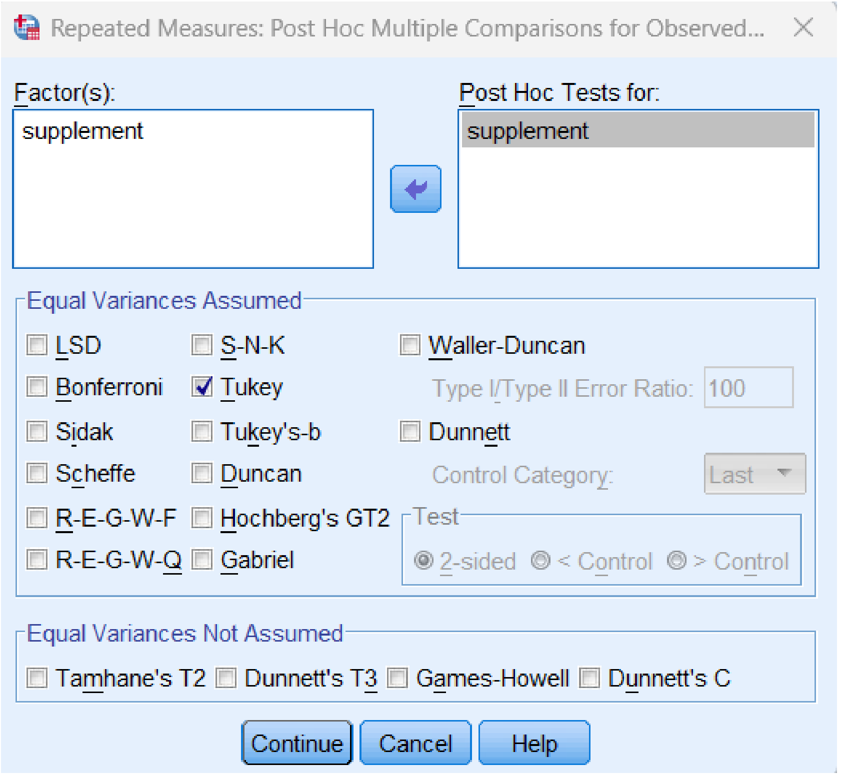
Repeated measures ANOVA

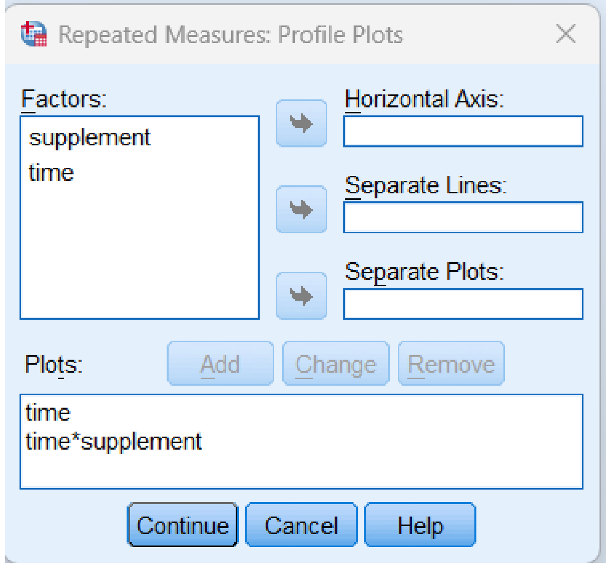






Repeated measures ANOVA





Descriptive Statistics

	Intervention supplement	Mean	Std. Deviation	N
Child's height, cm	Not fortified	72.050	8.0596	544
	Fortified	72.843	8.2386	547
	Total	72.448	8.1558	1091
Child's height, cm	Not fortified	74.133	7.9539	544
	Fortified	75.303	8.5404	547
	Total	74.720	8.2701	1091
Child's height, cm	Not fortified	74.188	7.9714	544
	Fortified	75.651	8.5555	547
	Total	74.922	8.2979	1091

Box's Test of Equality of Covariance Matrices[a]

Box's M	11.521
F	1.914
df1	6
df2	8591634.503
Sig.	.074

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + supplement
Within Subjects Design: time

Box's test is not significant, then you have evidence of no violation

Repeated measures ANOVA (cont…)

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[c]
time	Pillai's Trace	.299	231.631[b]	2.000	1088.000	.000	.299	463.262	1.000
	Wilks' Lambda	.701	231.631[b]	2.000	1088.000	.000	.299	463.262	1.000
	Hotelling's Trace	.426	231.631[b]	2.000	1088.000	.000	.299	463.262	1.000
	Roy's Largest Root	.426	231.631[b]	2.000	1088.000	.000	.299	463.262	1.000
time * supplement	Pillai's Trace	.097	58.577[b]	2.000	1088.000	.000	.097	117.154	1.000
	Wilks' Lambda	.903	58.577[b]	2.000	1088.000	.000	.097	117.154	1.000
	Hotelling's Trace	.108	58.577[b]	2.000	1088.000	.000	.097	117.154	1.000
	Roy's Largest Root	.108	58.577[b]	2.000	1088.000	.000	.097	117.154	1.000

a. Design: Intercept + supplement
Within Subjects Design: time

b. Exact statistic

Mauchly's Test of Sphericity[a]

Measure: Height

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
time	.025	4016.358	2	.000	.506	.507	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + supplement
Within Subjects Design: time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Here, we have the multivariate test results for time (the within-subjects factor) and the time X supplement interaction.

- The main effect of time is statistically significant, Wilks' lambda=.701, F(2,1088)=231.63, p<.001.
- This effect is qualified by a significant time X supplement interaction, Wilks' lambda = .903, F(2,1088)=58.577, p<.001.

Repeated measures ANOVA

Tests of Within-Subjects Effects

Measure: Height

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
time	Sphericity Assumed	4115.646	2	2057.823	254.400	.000	.189	508.801	1.000
	Greenhouse-Geisser	4115.646	1.013	4064.336	254.400	.000	.189	257.612	1.000
	Huynh-Feldt	4115.646	1.014	4060.463	254.400	.000	.189	257.858	1.000
	Lower-bound	4115.646	1.000	4115.646	254.400	.000	.189	254.400	1.000
time * supplement	Sphericity Assumed	61.407	2	30.704	3.796	.023	.003	7.592	.693
	Greenhouse-Geisser	61.407	1.013	60.642	3.796	.051	.003	3.844	.498
	Huynh-Feldt	61.407	1.014	60.584	3.796	.051	.003	3.847	.498
	Lower-bound	61.407	1.000	61.407	3.796	.052	.003	3.796	.495
Error(time)	Sphericity Assumed	17617.665	2178	8.089					
	Greenhouse-Geisser	17617.665	1102.748	15.976					
	Huynh-Feldt	17617.665	1103.800	15.961					
	Lower-bound	17617.665	1089.000	16.178					

a. Computed using alpha = .05

The main effect of time on height is statistically significant, sphericity assumed F(2,2178)=254.400, p<.001

This effect was qualified by a significant time X supplement interaction effect, sphericity assumed **F(2, 2178)=3.796, p=023.**

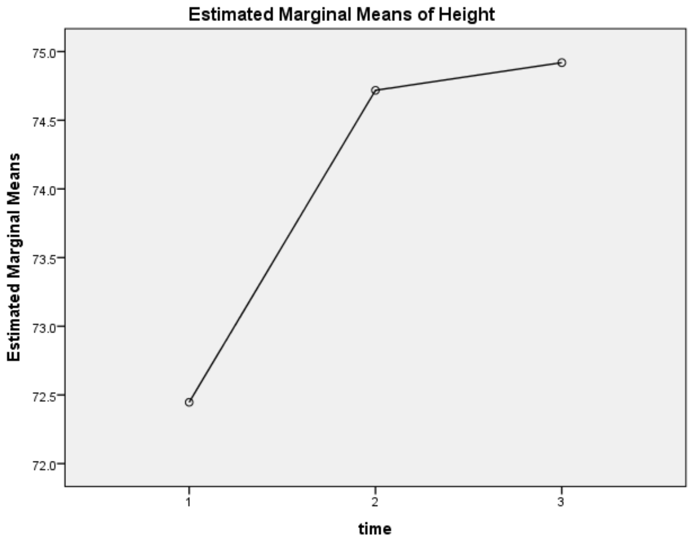
.

Tests of Within-Subjects Contrasts

Measure: Height

Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
time	Linear	3336.236	1	3336.236	275.640	.000	.202	275.640	1.000
	Quadratic	779.410	1	779.410	191.302	.000	.149	191.302	1.000
time * supplement	Linear	61.083	1	61.083	5.047	.025	.005	5.047	.612
	Quadratic	.324	1	.324	.080	.778	.000	.080	.059
Error(time)	Linear	13180.817	1089	12.104					
	Quadratic	4436.847	1089	4.074					

a. Computed using alpha = .05



- Although the test of the linear component of the trend is significant (p<.001), the higher-order quadratic component was also significant (p<.001).
- Suggests the mean level of height exhibited a quadratic trend over the three measurement occasions.

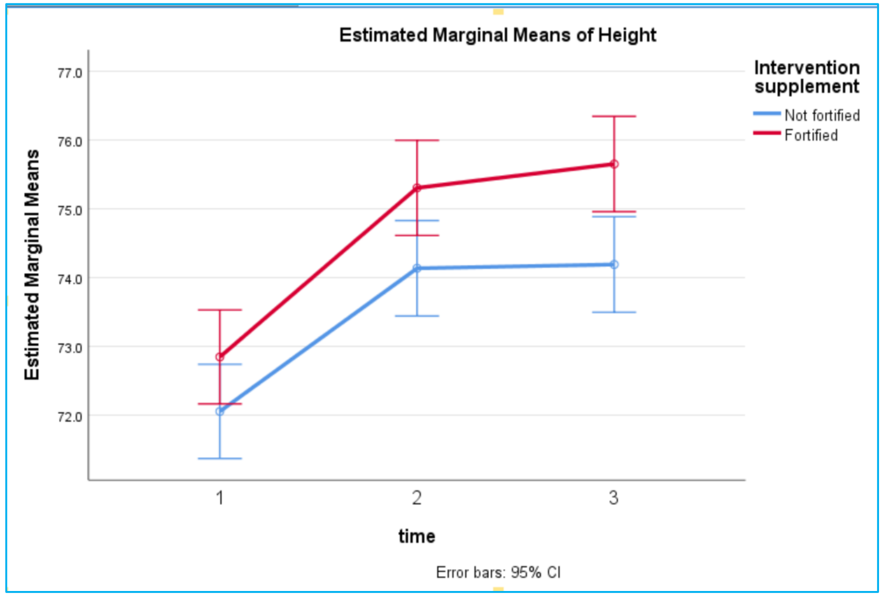
Repeated measures ANOVA

Tests of Within-Subjects Contrasts

Measure: Height

Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
time	Linear	3336.236	1	3336.236	275.640	.000	.202	275.640	1.000
	Quadratic	779.418	1	779.418	191.302	.000	.149	191.302	1.000
time * supplement	Linear	61.083	1	61.083	5.047	.025	.005	5.047	.612
	Quadratic	.324	1	.324	.080	.778	.000	.080	.059
Error(time)	Linear	13180.817	1089	12.104					
	Quadratic	4436.847	1089	4.074					

a. Computed using alpha = .05



- The test of the interaction between the linear component of the trend and supplement group is significant (p=0.005)
- The interaction between treatment group and the higher-order quadratic component was also significant (p<0.001)

Repeated measures ANOVA

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
Child's height, cm	.306	1	1089	.580
Child's height, cm	1.176	1	1089	.278
Child's height, cm	1.228	1	1089	.268

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + supplement
Within Subjects Design: time

- The Levene's test results involve tests of differences in variances at each time point, an assumption of the univariate ANOVA
- All are not significant and assumption fulfilled ($p>0.05$)

Tests of Between-Subjects Effects

Measure: Height

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	17936454.90	1	17936454.90	96050.255	.000	.989	96050.255	1.000
supplement	1066.854	1	1066.854	5.713	.017	.005	5.713	.666
Error	203360.203	1089	186.740					

a. Computed using alpha = .05

- The main effect of supplement on the average height across time is statistically significant, $F(1, 1089)=5.713$, $p=0.017$.

Repeated measures ANOVA

Estimated Marginal Means

1. Grand Mean

Measure: Height

Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
74.028	.239	73.559	74.497

2. Intervention supplement

Estimates

Measure: Height

Intervention supplement	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Not fortified	73.457	.338	72.794	74.121
Fortified	74.599	.337	73.937	75.261

Pairwise Comparisons

Measure: Height

(I) Intervention supplement	(J) Intervention supplement	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
Not fortified	Fortified	-1.142*	.478	.017	-2.079	-.204
Fortified	Not fortified	1.142*	.478	.017	.204	2.079

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Univariate Tests

Measure: Height

	Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Contrast	355.618	1	355.618	5.713	.017	.005	5.713	.666
Error	67786.734	1089	62.247					

The F tests the effect of Intervention supplement. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Computed using alpha = .05

Bonferroni-adjusted paired t-tests. Here we see all pairwise differences on height are statistically significant ($p<0.05$)).

Repeated measures ANOVA

3. time

Estimates

Measure: Height

time	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	72.447	.247	71.962	72.931
2	74.718	.250	74.228	75.209
3	74.920	.250	74.428	75.411

Pairwise Comparisons

Measure: Height

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
1	2	-2.272*	.149	.000	-2.628	-1.915
	3	-2.473*	.149	.000	-2.830	-2.116
2	1	2.272*	.149	.000	1.915	2.628
	3	-.201*	.014	.000	-.234	-.169
3	1	2.473*	.149	.000	2.116	2.830
	2	.201*	.014	.000	.169	.234

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Pillai's trace	.299	231.631[a]	2.000	1088.000	.000	.299	463.262	1.000
Wilks' lambda	.701	231.631[a]	2.000	1088.000	.000	.299	463.262	1.000
Hotelling's trace	.426	231.631[a]	2.000	1088.000	.000	.299	463.262	1.000
Roy's largest root	.426	231.631[a]	2.000	1088.000	.000	.299	463.262	1.000

Each F tests the multivariate effect of time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

b. Computed using alpha = .05

4. Intervention supplement * time

Measure: Height

Intervention supplement	time	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Not fortified	1	72.050	.349	71.364	72.736
	2	74.133	.354	73.439	74.828
	3	74.188	.355	73.493	74.884
Fortified	1	72.843	.348	72.159	73.527
	2	75.303	.353	74.611	75.996
	3	75.651	.354	74.957	76.345

- These are Bonferroni-adjusted pairwise comparisons
- These are pairwise comparisons on the average height (averaged over time) for each group.
- All pairwise differences were significant (as all p<0.001).

Group means on height at each measurement occasion.

Questions?

Non-parametric tests

Non-parametric tests

Overview

- Used when assumptions of outliers, linearity, normality, and homoscedasticity are violated in common statistical analysis
- Non-parametric tests make less assumption, not no assumption at all
- The compromise is that non-parametric tests are lower in power

Types of Non-parametric Tests

The non-parametric version of *t*-tests and ANOVAs

Parametric Test	Non-parametric Test
Independent t-test	Mann-Whitney *U* Test
Paired t-test	Wilcoxon Signed Ranked Test
One-way ANOVA	Kruskal-Wallis One-way ANOVA
Repeated measures ANOVA	Friedman's ANOVA

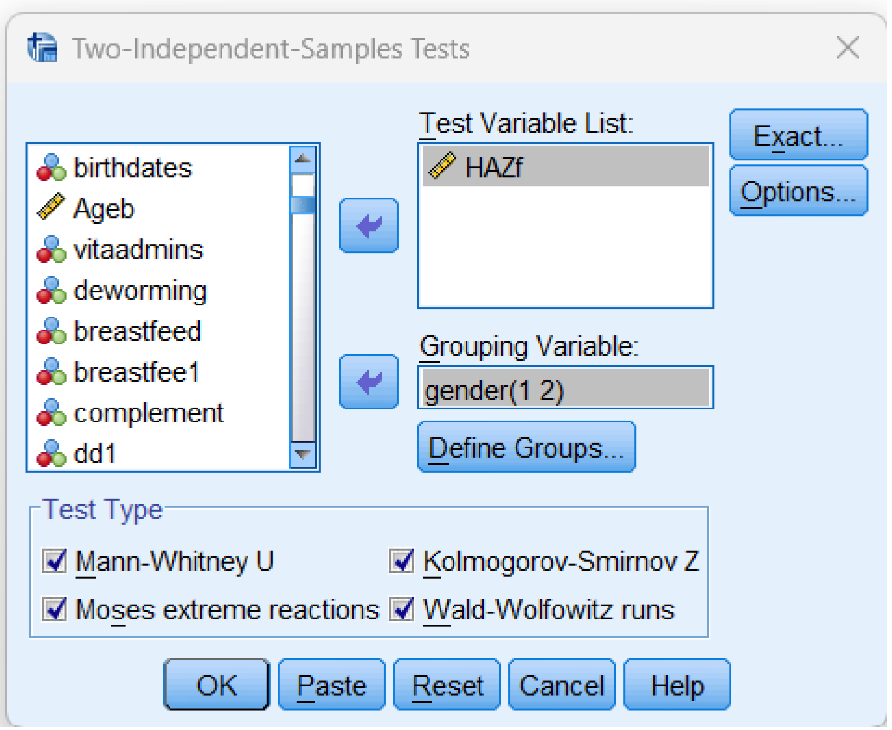
Two-Independent-Samples Tests

- Useful for determining whether or not the values of a particular variable differ between two groups.
- This is especially true when the assumptions of the t-test are not met.

- **Mann-Whitney U test:** To test for differences between two groups.
- **The two-sample Kolmogorov-Smirnov test:** To test the null hypothesis that two samples have the same distribution
- **Wlad-Walfowitz Run:** **U**sed to examine whether two random samples come from populations having same distribution

Mann-Whitney *U* Test

Analyze > Non-parametric test > legacy dialogs>2-Independent- Sample Tests …



Mann-Whitney Test

Ranks

	Sex of the child	N	Mean Rank	Sum of Ranks
Length/height-for-age z-score	Male	412	467.52	192619.50
	Female	629	556.03	349741.50
	Total	1041		

Test Statistics[a]

	Length/height -for-age z-score
Mann-Whitney U	107541.500
Wilcoxon W	192619.500
Z	-4.645
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Sex of the child

Interpretation

- Normally we have two p-values (Asympt, Exact)
- Asymptotic is appropriate for large sample
- Exact is independent of sample size.
- p <0.001 = there a significant difference in LAZ between males and females

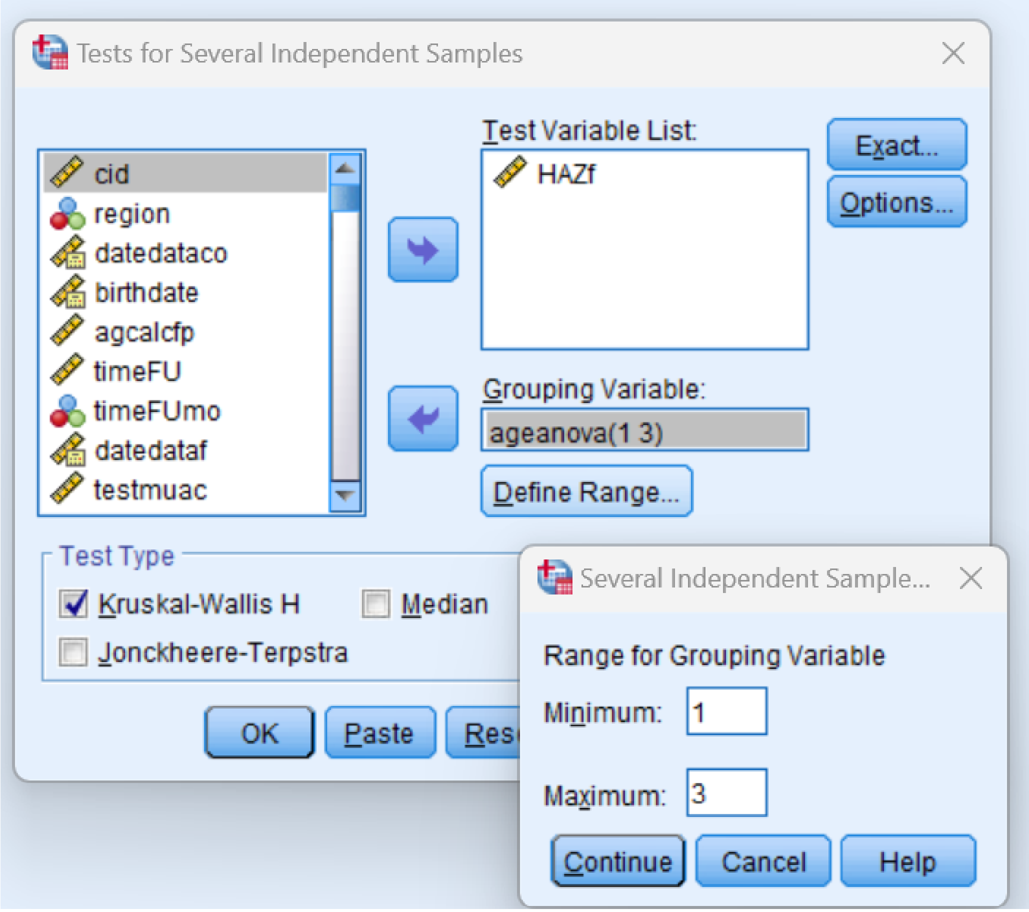
Multiple Independent Samples Tests

- Useful for determining whether or not the values of a particular variable differ between two or more groups.
- Applied when the assumptions of ANOVA are not met.

- **Kruskal-Wallis H:**
  - This test is a one-way analysis of variance by ranks.
  - It tests the null hypothesis that multiple independent samples come from the same population.
- **Median test:**
  - This method tests the null hypothesis that two or more independent samples have the same median.
  - It assumes nothing about the distribution of the test variable, making it a good choice when you suspect that the distribution varies by group
- **Jonckheere-terpstra test:** Exact test

Kruskal-Wallis Test

Analyze > Non-parametric test > legacy dialogs>K Independent- Samples



Kruskal-Wallis Test

Ranks

	age of the child	N	Mean Rank
Length/height-for-age z-score	<12 months	383	579.19
	12-23 months	398	487.69
	>24 months	253	471.00
	Total	1034	

Test Statistics[a,b]

	Length/height-for-age z-score
Kruskal-Wallis H	26.445
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: age of the child

Interpretation:

The p-value is <0.001 which is significant. Therefore we conclude that there is a significant difference between the groups (meaning- at least two groups are different)

Tests for Two Related Samples

Description

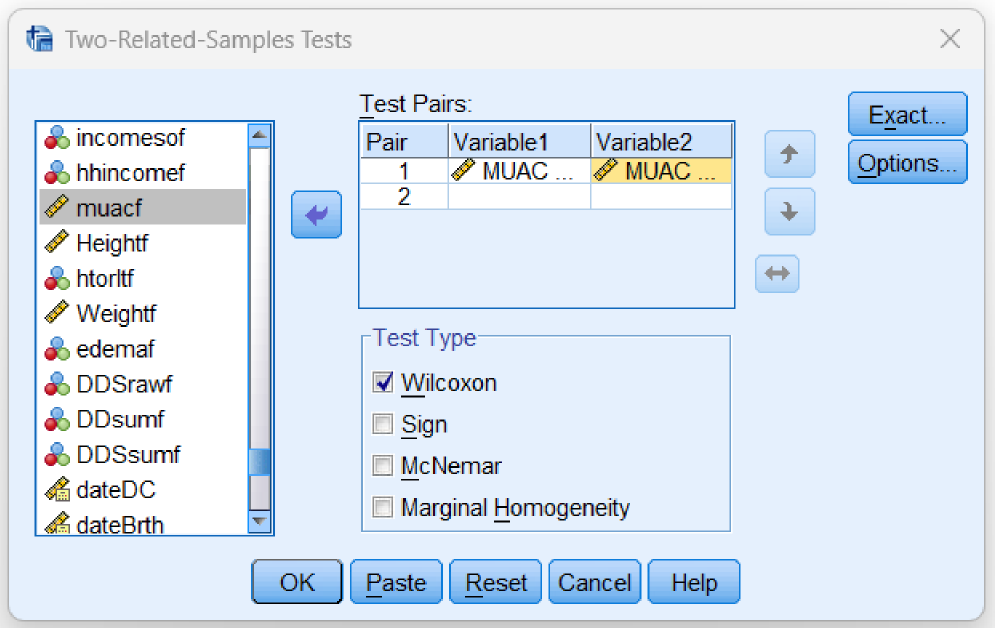
- Allow you to test for differences between paired scores when you cannot (or would rather not) make the assumptions required by the paired-samples t test (within subjects t-test).
- Procedures are available for testing **nominal, ordinal, or scale variables.**

Tests for Two Related Samples /paired

Types

- **Wilcoxon signed-ranks**
  - A nonparametric alternative to the paired-sample t-test.
  - The only assumptions made by the Wilcoxon test are that the test variable is continuous and that the distribution of the difference scores is reasonably symmetric.
- **McNemar**
  - Tests the null hypothesis that binary responses are unchanged.
  - As with the Wilcoxon test, the data may be from a single sample measured twice or from two matched samples.
  - The McNemar test is particularly appropriate with nominal or ordinal test variables for binary data.

Wilcoxon Signed Ranks Test



Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
MUAC measurement in cms - MUAC measurement in cms	Negative Ranks	126[a]	361.61	45563.00
	Positive Ranks	923[b]	547.30	505162.00
	Ties	50[c]		
	Total	1099		

a. MUAC measurement in cms < MUAC measurement in cms

b. MUAC measurement in cms > MUAC measurement in cms

c. MUAC measurement in cms = MUAC measurement in cms

Test Statistics[a]

	MUAC measurement in cms - MUAC measurement in cms
Z	-23.432[b]
Asymp. Sig. (2-tailed)	.000

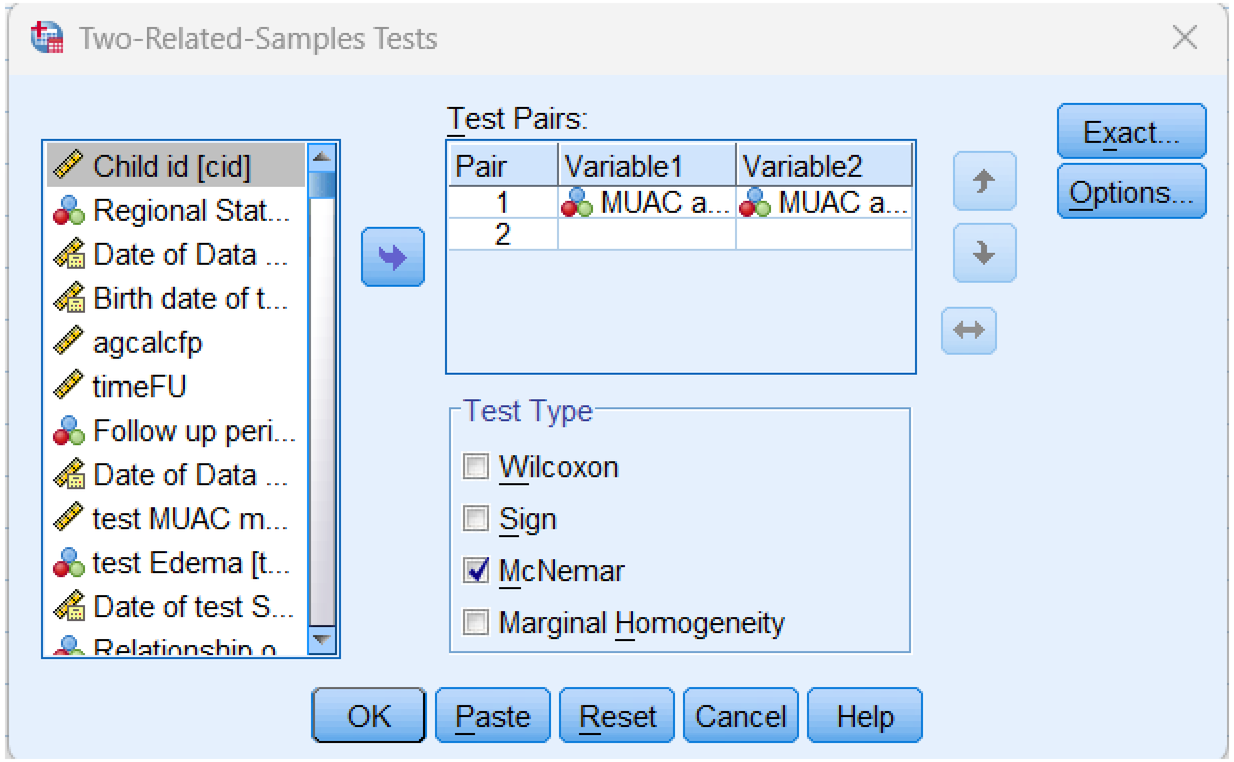
a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

Interpretation

- P<0.001 which is significant.
- This indicates that baseline and follow up MUAC are different

McNemar test



McNemar Test

Crosstabs

MUAC at baseline & MUAC at follow up

MUAC at baseline	MUAC at follow up: 12cm	MUAC at follow up: >=12cm
12cm	398	701
>=12cm	0	0

Test Statistics[a]

	MUAC at baseline & MUAC at follow up
N	1099
Chi-Square[b]	699.001
Asymp. Sig.	.000

a. McNemar Test

b. Continuity Corrected

Interpretation

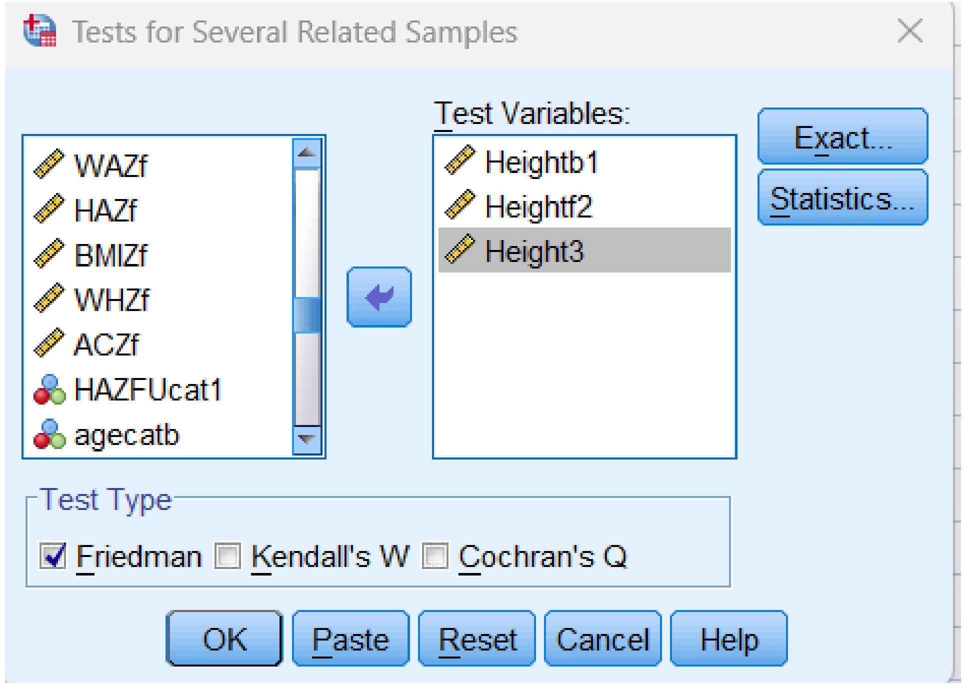
- P<0.001 which is significant.
- Baseline and follow up dichotomized MUAC values are different

Tests for multiple related samples

- Useful alternatives to a repeated measures analysis of variance.
- They are especially appropriate for small samples and can be used with nominal or ordinal test variables.

- **Friedman test** is a nonparametric alternative to the repeated measures ANOVA.
- It tests the null hypothesis that multiple ordinal responses come from the same population.
- The only assumptions made by the Friedman test are that the test variables are at least ordinal and that their distributions are reasonably similar.

Friedman test



Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Child's height, cm	1091	72.448	8.1558	53.0	103.1
Child's height, cm	1091	74.720	8.2701	55.2	120.3
Child's height, cm	1091	74.922	8.2979	55.2	120.6

Friedman Test

Ranks

	Mean Rank
Child's height, cm	1.29
Child's height, cm	2.09
Child's height, cm	2.62

Test Statistics[a]

N	1091
Chi-Square	1124.764
df	2
Asymp. Sig.	.000

a. Friedman Test

Interpretation

- P value < 0.001. Hence there is significant difference among the three groups (meaning- at least two groups are different)

Questions?