Customizing the Analysis

Contrasts, Polynomials, and Generalized Estimation Equations

MAASTRICHT UNIVERSITY

Welcome

This Session

- What are Contrasts (adding them in the GLM)
- What are Polynomials (adding them in the GLM)
- How to add Contrasts and Polynomials in Mixed Models
- Generalized Estimation Equations (GEE)

01

Adding Contrasts

Contrasts

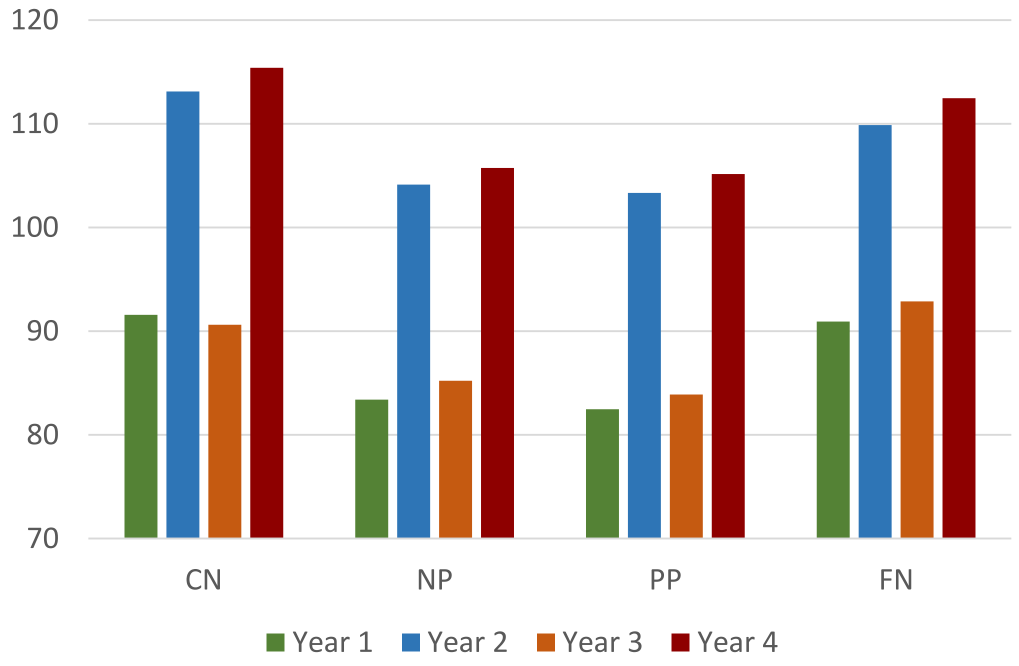
A contrast is a linear combination of variables

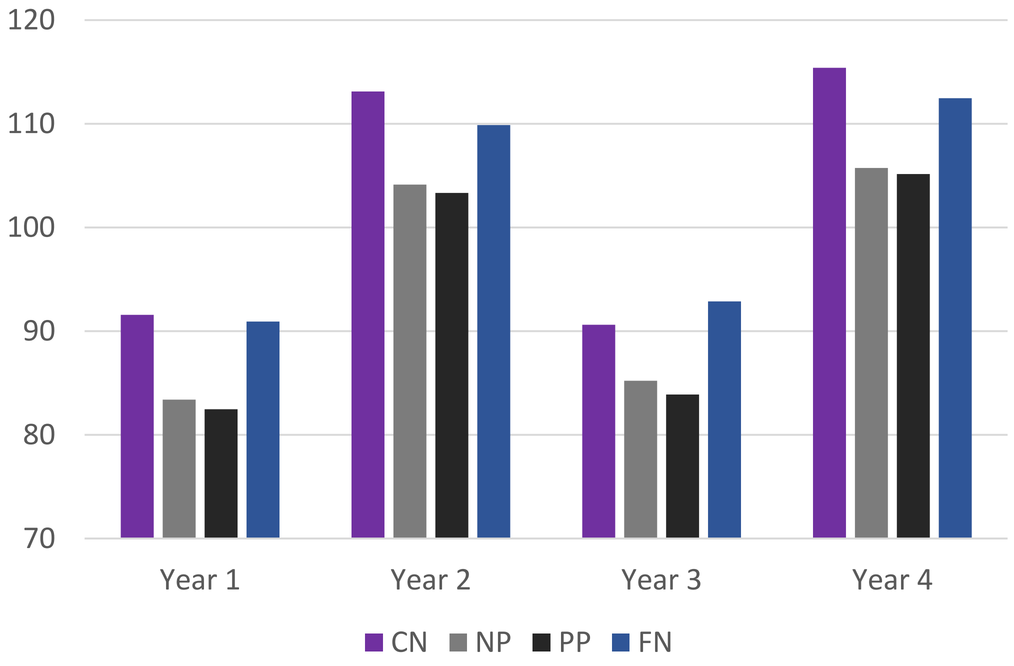
It's like adding weights to variables. We can use this to **compare** them or see if they follow a certain **shape**

The Data

Estimates

Track	Year 1	Year 2	Year 3	Year 4
Cognitive Neuroscience	91.57	113.09	90.61	115.39
Neuropsychology	83.40	104.14	85.21	105.74
Psychopathology	82.45	103.33	83.87	105.14
Fundamental Neuroscience	90.91	109.86	92.85	112.47

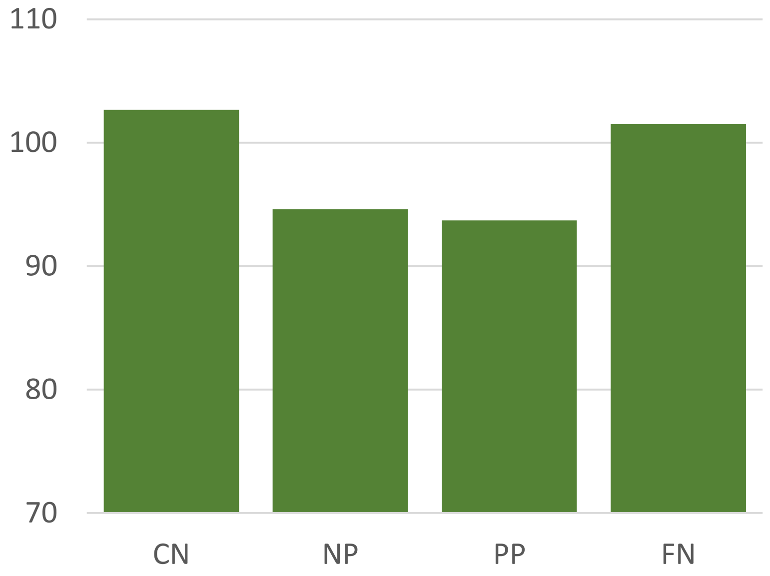




Unplanned Contrasts

Estimates

Track	Mean
Cognitive Neuroscience	102.67
Neuropsychology	94.62
Psychopathology	93.70
Fundamental Neuroscience	101.53



Pairwise Comparisons

(I) Track	(J) Track	Mean Difference (I-J)	Std. Error	Sig.
Cognitive Neuroscience	Neuropsychology	8,045*	,628	,000
	Psychopathology	8,969*	,628	,000
	Fundamental Neuroscience	1,141	,628	,364
Neuropsychology	Psychopathology	,924	,628	,608
	Fundamental Neuroscience	-6,905*	,628	,000
Psychopathology	Fundamental Neuroscience	-7,828*	,628	,000

It looks like we have two “pairs” of CN+FN and NP+PP

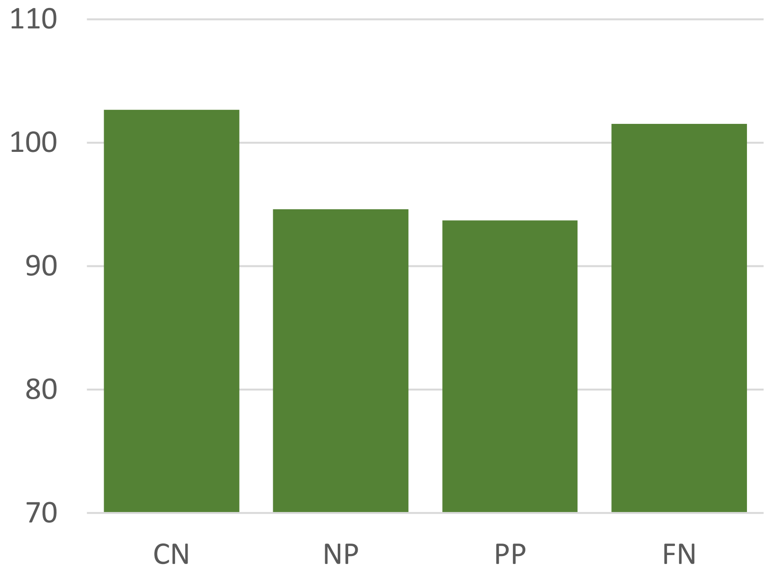
Comparing a combination of groups

Sometimes you want to do planned comparisons or compare one group to a combination of the other groups. We can use contrasts for that.

Planned Contrasts GLM

Estimates

Track	Mean	L1	L2	L3
Cognitive Neuroscience	102.67	3	1	0
Neuropsychology	94.62	-1	-1	0
Psychopathology	93.70	-1	-1	0
Fundamental Neuroscience	101.53	-1	1	1



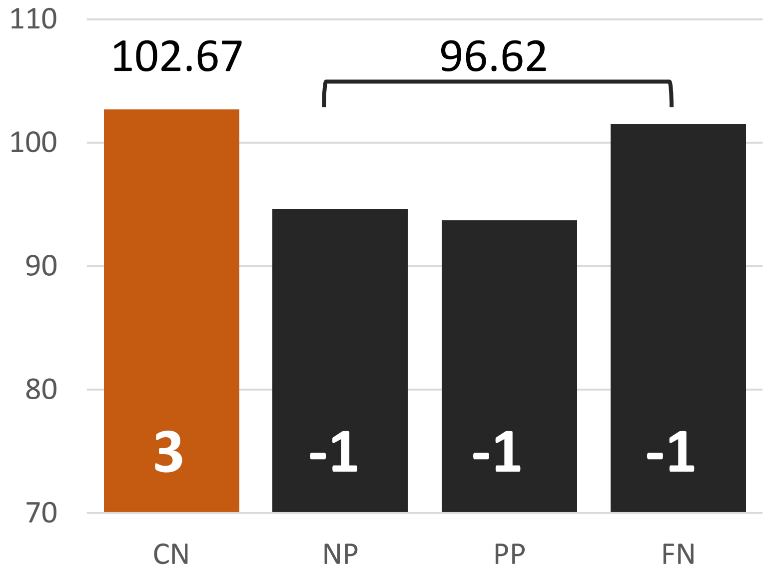
L1 compares CN to the average of the other groups
L2 compares CN+FN to NP+PP
L3 compares FN to the intercept, which should give the average of FN

(as an example of how these contrasts are computed)

Planned Contrasts GLM

Estimates

Track	Mean	L1
Cognitive Neuroscience	102.67	3
Neuropsychology	94.62	-1
Psychopathology	93.70	-1
Fundamental Neuroscience	101.53	-1



L1 compares CN to the average of the other groups
We can do this using multiple commands in the GLM (RM ANOVA):

- The built-in **Helmert** will compare each level to everything above it
- The second is **SPECIAL** and is a bit more advanced, it allows us to specify the contrasts we want, but we wont be able to use fractions
- Lastly **LMATRIX** allows us to specify coefficients and allows fractions

```
/CONTRAST(Track)=Helmert
/CONTRAST(Track) SPECIAL(3 -1 -1 -1)
/LMATRIX = "CN vs All" Track 1/2 -1/6 -1/6 -1/6
```

	Helmert	SPECIAL	LMATRIX
Estimate	6.052	18.155	6.052
SE	.513	1.539	.513
Hypothesis	0	0	0
Sig.	.000	.000	.000

Helmert Coding

Contrast Results (K Matrix)

Track Helmert Contrast			Averaged Variable MEASURE_1
Level 1 vs. Later	Contrast Estimate		6,052
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		6,052
	Std. Error		,513
	Sig.		,000
	95% Confidence Interval for Difference	Lower Bound	5,034
		Upper Bound	7,070
Level 2 vs. Later	Contrast Estimate		-2,991
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-2,991
	Std. Error		,544
	Sig.		,000
	95% Confidence Interval for Difference	Lower Bound	-4,070
		Upper Bound	-1,911
Level 3 vs. Level 4	Contrast Estimate		-7,828
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-7,828
	Std. Error		,628
	Sig.		,000
	95% Confidence Interval for Difference	Lower Bound	-9,075
		Upper Bound	-6,582

SPECIAL Command

Contrast Results (K Matrix)

Track Special Contrast			Averaged Variable MEASURE_1
L1	Contrast Estimate		18,155
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		18,155
	Std. Error		1,539
	Sig.		,000
	95% Confidence Interval for Difference	Lower Bound	15,101
		Upper Bound	21,209

LMATRIX Command

Contrast Results (K Matrix)

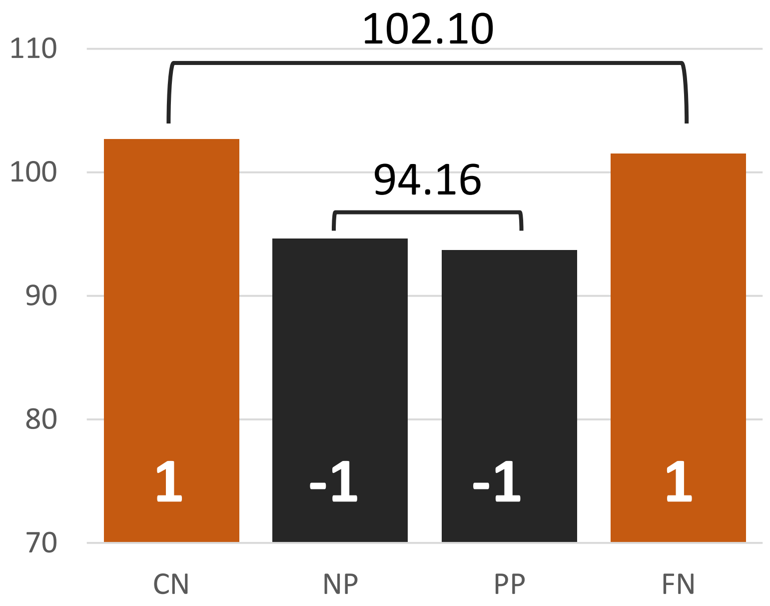
Contrast			Averaged Variable MEASURE_1
L1	Contrast Estimate		6,052
	Hypothesized Value		,000
	Difference (Estimate - Hypothesized)		6,052
	Std. Error		,513
	Sig.		,000
	95% Confidence Interval for Difference	Lower Bound	5,034
		Upper Bound	7,070

a. Based on the user-specified contrast coefficients (L') matrix: CN vs All

Planned Contrasts GLM

Estimates

Track	Mean	L2
Cognitive Neuroscience	102.67	1
Neuropsychology	94.62	-1
Psychopathology	93.70	-1
Fundamental Neuroscience	101.53	1



L2 Compares CN/FN with NP/PP

We will just use **LMATRIX** because we like doing stuff the hard way.

I used fractions so the estimate matches the mean difference.

```
/LMATRIX = "CN/FN vs. NP/PP" Track 1/4 -1/4 -1/4 1/4
```

If you have a difference hypothesis for the difference (by default it is set to 0, you can change it using the **/KMATRIX** command.

```
/KMATRIX = 8
```

	LMATRIX	LMATRIX
Estimate	7.937	7.937
SE	.444	.444
Hypothesis	0	8
Sig.	.000	.887

Polynomial Contrast

If you have **repeated measures** you might want to do a **trend analysis**

how the values change over time

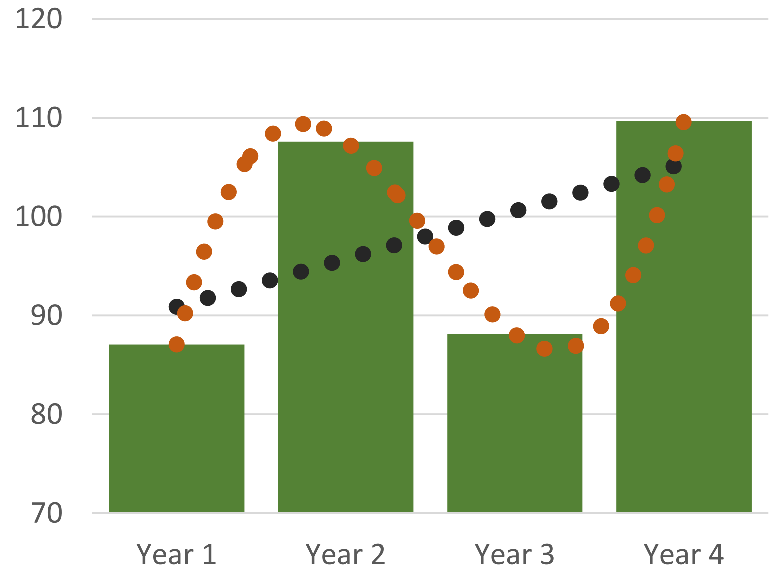
Polynomial Contrasts GLM

Estimates

Year	Mean
Year 1	87.08
Year 2	107.60
Year 3	88.13
Year 4	109.68

Pairwise Comparisons

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.
Year 1	Year 2	-20,520*	,725	,000
	Year 3	-1,049	,736	,642
	Year 4	-22,601*	,698	,000
Year 2	Year 3	19,471*	,797	,000
	Year 4	-2,081*	,752	,040
Year 3	Year 4	-21,552*	,666	,000



With 4 levels it's possible to have three polynomials: Linear, Quadratic, and Cubic.

We have a rise, then a drop, then a rise again, which looks like a cubic trend (I am 99.99% sure it is).

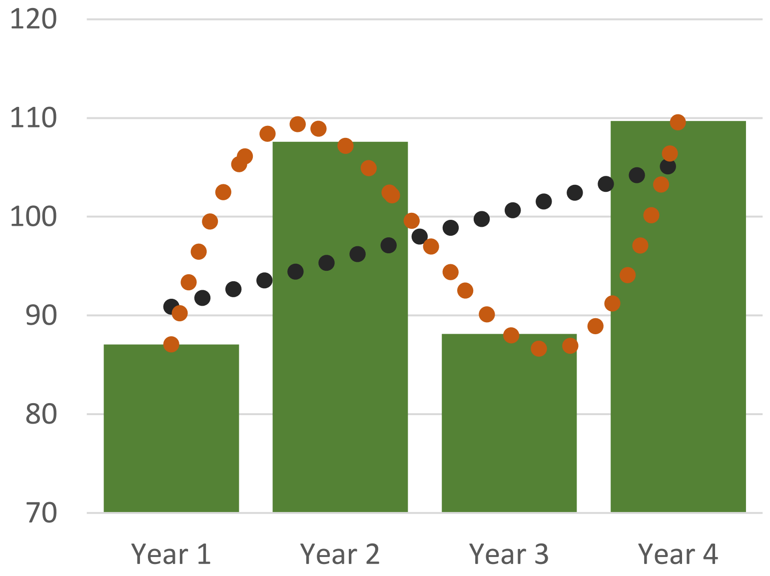
We could conclude this from the pairwise comparisons and the picture, but where's the fun in that?

Polynomial Contrasts GLM

Estimates

Year	Mean	Linear	Quadratic	Cubic
Year 1	87.08	-3	1	-1
Year 2	107.60	-1	-1	3
Year 3	88.13	1	-1	-3
Year 4	109.68	3	1	1

- **Linear** weighs the extremes, we assume that the change over time is constant
- **Quadratic** assumes there is a trough (think the famous inverted-U shape)
- **Cubic** fits one of those squiggly lines, where there are is a peak and a trough



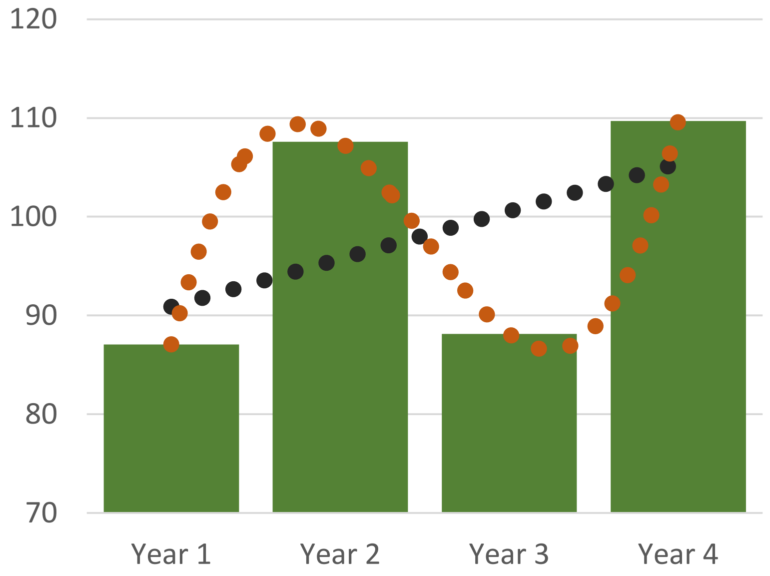
	Linear	Quadratic	Cubic
F	430.064	1.116	1134.184
Sig.	.000	.293	.000

Polynomial Contrasts GLM

Estimates

Year	Mean	Linear	Quadratic	Cubic
Year 1	87.08	-3	1	-1
Year 2	107.60	-1	-1	3
Year 3	88.13	1	-1	-3
Year 4	109.68	3	1	1

- **Linear** weighs the extremes, we assume that the change over time is constant
- **Quadratic** assumes there is a trough (think the famous inverted-U shape)
- **Cubic** fits one of those squiggly lines, where there are is a peak and a trough



Tests of Within-Subjects Contrasts

Source	Year	Type III Sum of Squares	df	Mean Square	F	Sig.
Year	Linear	11680,367	1	11680,367	430,064	,000
	Quadratic	26,619	1	26,619	1,116	,293
	Cubic	32816,392	1	32816,392	1134,184	,000
Error(Year)	Linear	2607,322	96	27,160		
	Quadratic	2289,828	96	23,852		
	Cubic	2777,657	96	28,934		

Rewind Time

We can do all this using the Mixed Procedure…
but we'll have to do it the hard way

Planned Contrasts

The Mixed Procedure has its own **/LMATRIX**, it uses the **/TEST** command and works the same way. We will have to specify our own contrasts.

```
/TEST="Contrast CN vs All" Track 3 -1 -1 -1 DIVISOR=3.
/LMATRIX = "CN vs All" Track 3 -1 -1 -1.
```

We can specify a name, just like the `/LMATRIX` command, followed by the variable we want to compare. Lastly the `/TEST` command gives us the **`DIVISOR`** option, which is most convenient. This allows us to use regular numbers as coefficients and then have the computer divide them to get the properly scaled fractions.

Planned Contrasts

The Mixed Procedure has its own **/LMATRIX**, it uses the **/TEST** command and works the same way. We will have to specify our own contrasts.

```
/TEST="Contrast CN vs All" Track 3 -1 -1 -1 DIVISOR=3.
/LMATRIX = "CN vs All" Track 3 -1 -1 -1.
```

We can specify a name, just like the `/LMATRIX` command, followed by the variable we want to compare. Lastly the `/TEST` command gives us the **`DIVISOR`** option, which is most convenient. This allows us to use regular numbers as coefficients and then have the computer divide them to get the properly scaled fractions.

Contrast Estimates

Contrast	Estimate	Std. Error	df	Test Value	t	Sig.
L1	6,244923	,504375	96,000	0	12,382	,000

The Mixed Procedure seems to be off, the actual difference should be 6.052-ish. The reason for the difference is that we didn't specify the entire model, **we didn't add the interaction** (GLM does it automatically)

Planned Contrasts

We’re gonna have to take interactions into account

The previous model had no interactions. Main effects are easier to imagine: coefficients for the track and done. The interactions are a bit trickier, which is why the GLM does them for you. The L-Matrix has to be specified correctly, so in the right order.

Since we specified `/FIXED= Track Year Track*Year` it will change **Year** and then **Track**.

The order of effects is shown below (by printing the L-Matrix you can see this)

[Track=1; Year=1]
[Track=1; Year=2]
[Track=1; Year=3]
[Track=1; Year=4]
[Track=2; Year=1]
Etc.

Planned Contrasts

We're gonna have to take interactions into account

The best way to see the interaction and the order for the L-Matrix is to make a nice table.

We want to compare CN to all the other track, so we plug that into the table…

		Year 1	*Year 2*	*Year 3*	*Year 4*	
CN	**3**					
NP	**-1**					
PP	**-1**					
FN	**-1**					

Planned Contrasts

We're gonna have to take interactions into account

The best way to see the interaction and the order for the L-Matrix is to make a nice table.

We don't care about Year, so we take the average of all the years by adding $1/4^{th}$ of each year

		Year 1	*Year 2*	*Year 3*	*Year 4*	
		1/4	**1/4**	**1/4**	**1/4**	
CN	**3**					
NP	**-1**					
PP	**-1**					
FN	**-1**					

Planned Contrasts

We're gonna have to take interactions into account

The best way to see the interaction and the order for the L-Matrix is to make a nice table.

To get the coefficients for the **Interaction** fill in the inner cells (multiply the row by the column)

		Year 1	*Year 2*	*Year 3*	*Year 4*
		1/4	**1/4**	**1/4**	**1/4**
CN	**3**	**3/4**	**3/4**	**3/4**	**3/4**
NP	**-1**	**-1/4**	**-1/4**	**-1/4**	**-1/4**
PP	**-1**	**-1/4**	**-1/4**	**-1/4**	**-1/4**
FN	**-1**	**-1/4**	**-1/4**	**-1/4**	**-1/4**

Planned Contrasts

We're gonna have to take interactions into account

The best way to see the interaction and the order for the L-Matrix is to make a nice table.

Finally, to get the coefficients for the Main Effects of **Track** and **Year** sum the rows and columns

		Year 1	*Year 2*	*Year 3*	*Year 4*	
		1/4	**1/4**	**1/4**	**1/4**	
CN	**3**	**3/4**	**3/4**	**3/4**	**3/4**	***3***
NP	**-1**	**-1/4**	**-1/4**	**-1/4**	**-1/4**	***-1***
PP	**-1**	**-1/4**	**-1/4**	**-1/4**	**-1/4**	***-1***
FN	**-1**	**-1/4**	**-1/4**	**-1/4**	**-1/4**	***-1***
		0	***0***	***0***	***0***	

Planned Contrasts

We're gonna have to take interactions into account

The best way to see the interaction and the order for the L-Matrix is to make a nice table.

But we don't like fractions, so let's fix this to only contain whole numbers

		Year 1	*Year 2*	*Year 3*	*Year 4*	
		1/4	**1/4**	**1/4**	**1/4**	
CN	**3**	**3**	**3**	**3**	**3**	***12***
NP	**-1**	**-1**	**-1**	**-1**	**-1**	***-4***
PP	**-1**	**-1**	**-1**	**-1**	**-1**	***-4***
FN	**-1**	**-1**	**-1**	**-1**	**-1**	***-4***
		0	***0***	***0***	***0***	

Planned Contrasts

We're gonna have to take interactions into account

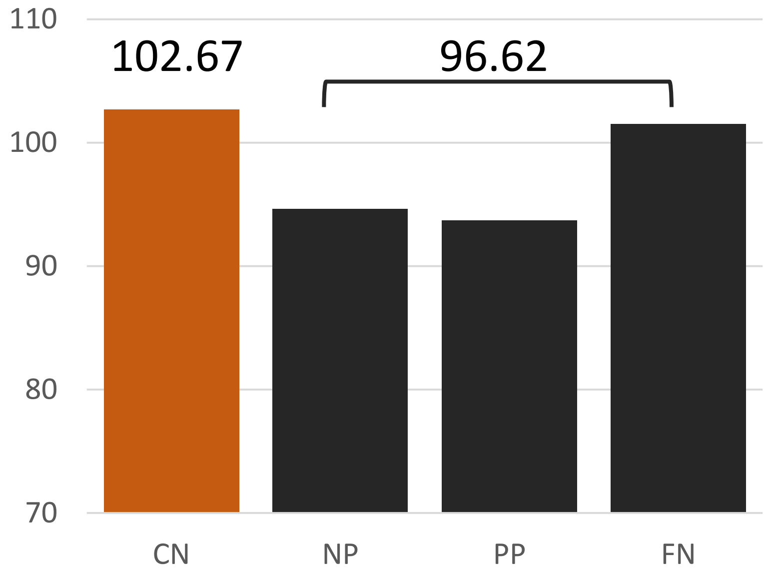
Now to translate the table to the SPSS Syntax. I used a divisor of 12 because my largest number is 12 and I want that to be 1.

```
/TEST="Contrast CN vs all" Track 12 -4 -4 -4
                            Year 0 0 0 0
                      Track*Year 3  3  3  3
                                 -1 -1 -1 -1
                                 -1 -1 -1 -1
                                 -1 -1 -1 -1
                                 DIVISOR=12.
```

Planned Contrasts

Estimates

Track	Mean	L1
Cognitive Neuroscience	102.67	3
Neuropsychology	94.62	-1
Psychopathology	93.70	-1
Fundamental Neuroscience	101.53	-1



L1 compares CN to the average of the other groups

- We used **/LMATRIX** in the GLM
- We use **/TEST** in the Mixed Procedure (random effects use a | before the coefficients)

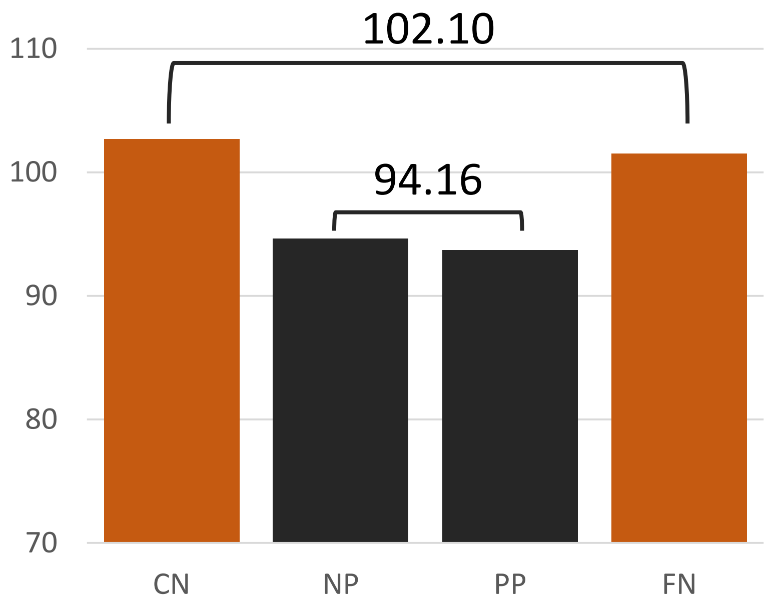
```
/LMATRIX = "CN vs All" Track 1/2 -1/6 -1/6 -1/6
/TEST    = "CN vs All" Track 12 -4 -4 -4
                       Year 0 0 0 0
                       Track*Year  3  3  3  3
                                  -1 -1 -1 -1
                                  -1 -1 -1 -1
                                  -1 -1 -1 -1
                                   DIVISOR=12.
```

	GLM	MIXED
Estimate	6.052	6.0516
SE	.513	.5129
Hypothesis	0	0
Sig.	.000	.000

Planned Contrasts

Estimates

Track	Mean	L2
Cognitive Neuroscience	102.67	1
Neuropsychology	94.62	-1
Psychopathology	93.70	-1
Fundamental Neuroscience	101.53	1



L2 Compares CN/FN with NP/PP

```
/TEST(0)  = "Contrast CN/FN vs NP/PP"  Track 4 -4 -4 4
                               Track*Year 1  1  1  1
                                         -1 -1 -1 -1
                                         -1 -1 -1 -1
                                          1  1  1  1
                                          DIVISOR=8
```

If you have a difference hypothesis for the difference (by default it is set to 0, you can change it using brackets, like `/TEST(8)`.

	TEST	TEST(8)
Estimate	7.9369	7.9369
SE	.44418	.44418
Hypothesis	0	8
Sig.	.000	.887

Polynomial Contrast

Polynomials in Mixed Models are more of the same; a matter of using the right coefficients on the right variables

Polynomial Contrasts

We can use the same table method we used for the planned contrasts, except now we plug in a polynomial contrast instead. The contrast for a Linear polynomial for a variable with 4 levels is: -3 -1 1 3

		Year 1	*Year 2*	*Year 3*	*Year 4*	
		-3	**-1**	**1**	**3**	
CN	**1/4**	**-3/4**	**-1/4**	**1/4**	**3/4**	***0***
NP	**1/4**	**-3/4**	**-1/4**	**1/4**	**3/4**	***0***
PP	**1/4**	**-3/4**	**-1/4**	**1/4**	**3/4**	***0***
FN	**1/4**	**-3/4**	**-1/4**	**1/4**	**3/4**	***0***
		-3	***-1***	***1***	***3***	

Polynomial Contrasts

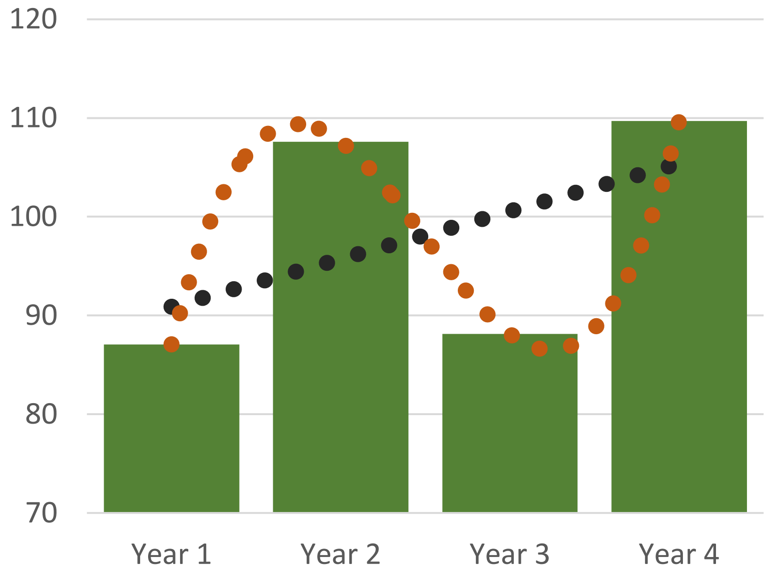
We can use the same table method we used for the planned contrasts, except now we plug in a polynomial contrast instead. The contrast for a Linear polynomial for a variable with 4 levels is: -3 -1 1 3

		Year 1	*Year 2*	*Year 3*	*Year 4*	
		-3	**-1**	**1**	**3**	
CN	**1/4**	**-3**	**-1**	**1**	**3**	***0***
NP	**1/4**	**-3**	**-1**	**1**	**3**	***0***
PP	**1/4**	**-3**	**-1**	**1**	**3**	***0***
FN	**1/4**	**-3**	**-1**	**1**	**3**	***0***
		12	***-4***	***4***	***12***	

Polynomial Contrasts GLM

Estimates

Year	Mean	Linear	Quadratic	Cubic
Year 1	87.08	-3	1	-1
Year 2	107.60	-1	-1	3
Year 3	88.13	1	-1	-3
Year 4	109.68	3	1	1




```
/TEST = "Linear Year"          Year  -12 -4 4 12
                         Track*Year   -3 -1 1 3
                                      -3 -1 1 3
                                      -3 -1 1 3
                                      -3 -1 1 3 DIVISOR=12
/TEST = "Quadratic Year“      Year    4 -4 -4 4
                        Track*Year    1 -1 -1 1
                                      1 -1 -1 1
                                      1 -1 -1 1
                                      1 -1 -1 1 DIVISOR=4
/TEST = "Cubic Year“           Year -4 12 -12 4
                         Track*Year -1 3 -3 1
                                    -1 3 -3 1
                                    -1 3 -3 1
                                    -1 3 -3 1 DIVISOR=12
```

	Linear	Quadratic	Cubic
Estimate	16.11098	1.031864	27.004688
SE	.776883	.976778	.801858
Hypothesis	0	0	0
Sig.	.000	.293	.000

Polynomial Interactions

Polynomials can interact with each other and to check for this we'll need to specify such an interaction

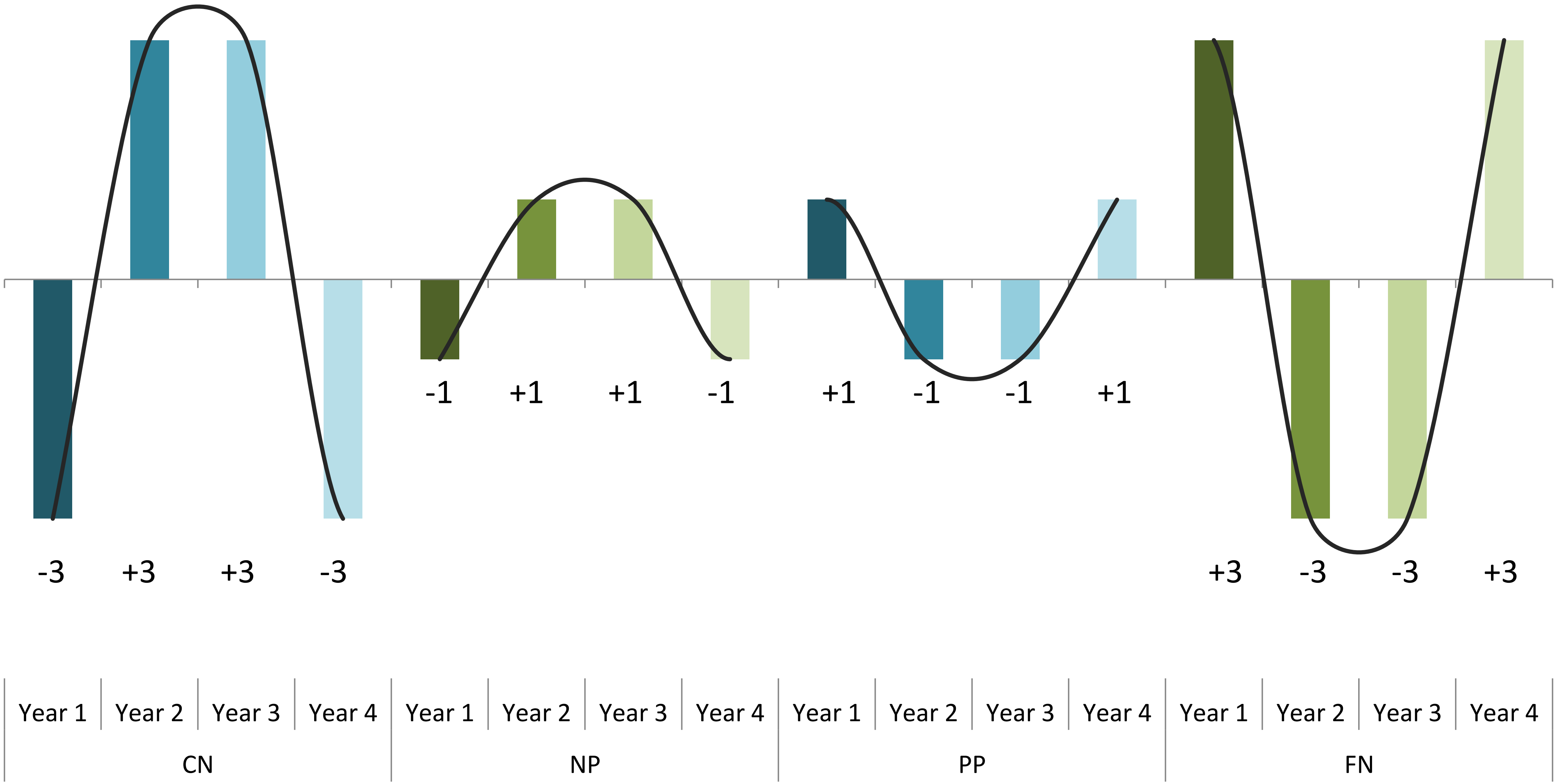
Polynomial Contrasts

An interaction between polynomials sounds complicated, but it really isn't all that different from a regular interaction. Let's specify such an interaction by saying that **Year** is **Linear** and **Track** is **Quadratic**, resulting in a **Linear*Quadratic** interaction.

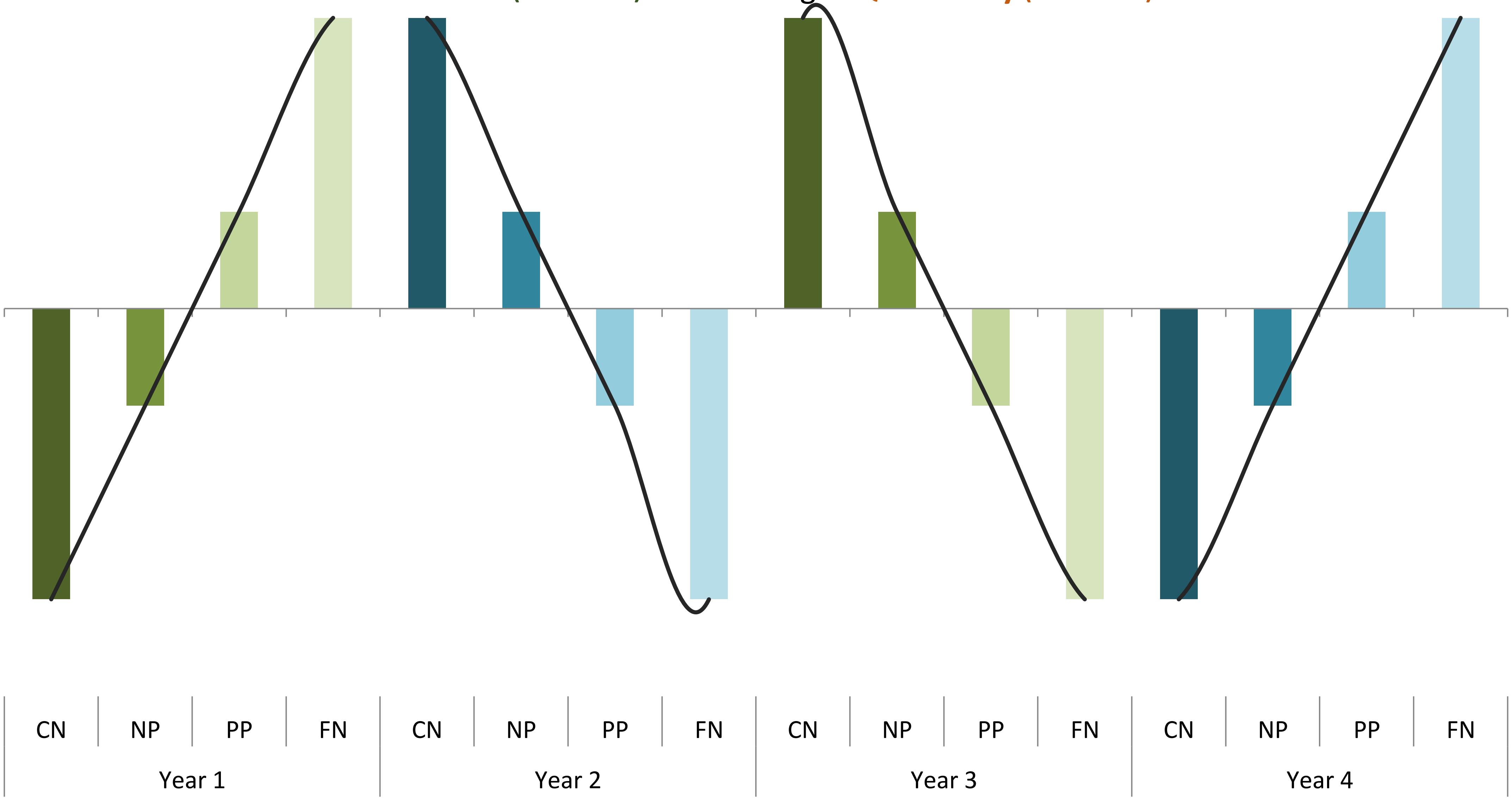
		Year 1	*Year 2*	*Year 3*	*Year 4*	
		1	**-1**	**-1**	**1**	
CN	**-3**	-3	3	3	-3	*0*
NP	**-1**	-1	1	1	-1	*0*
PP	**1**	1	-1	-1	1	*0*
FN	**3**	3	-3	-3	3	*0*
		0	*0*	*0*	*0*	

```
/TEST = "Linear*Quadratic Interaction"
                        Track    0  0  0  0
                         Year    0  0  0  0
                   Track*Year   -3  3  3 -3
                                -1  1  1 -1
                                 1 -1 -1  1
                                 3 -3 -3  3 DIVISOR=3
```

The **Quadratic** (+1 -1 -1 +1) effect changes **Linearly** (-3 -1 +1 +3)



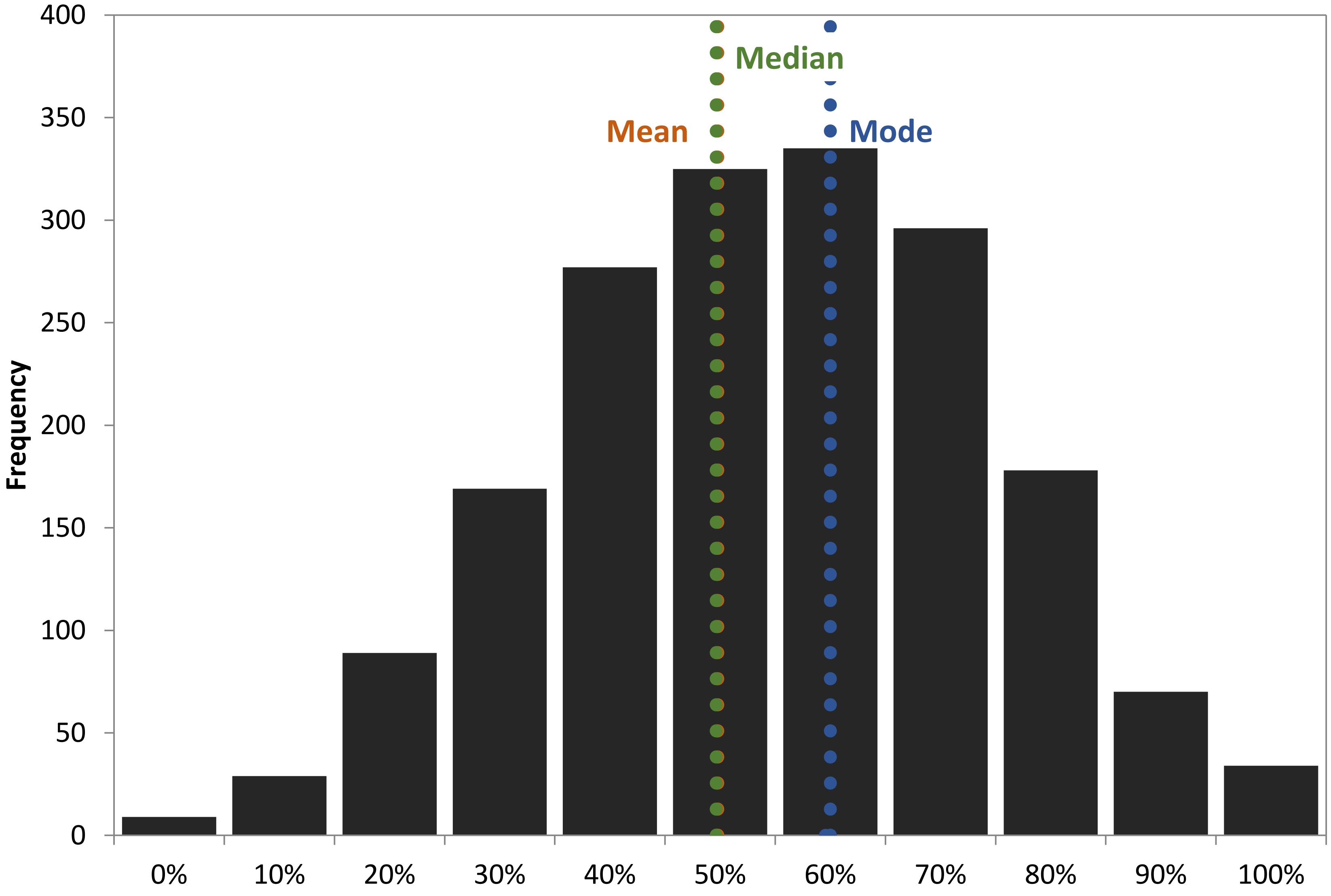
The **Linear** (-3 -1 +1 +3) effect changes **Quadraticly** (+1 -1 -1 +1)



02 Generalized Estimation Equation

Ugly Data

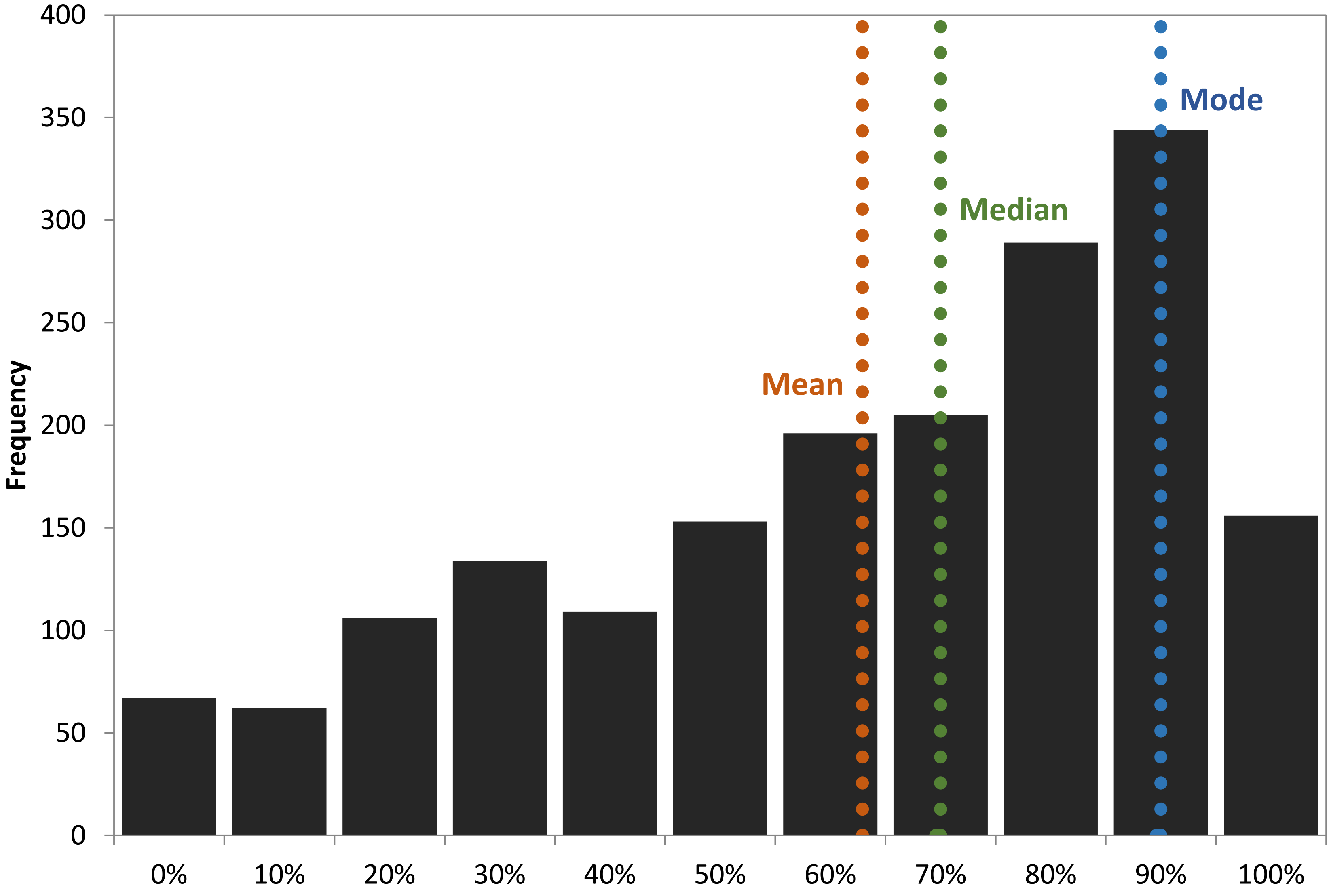
Sometimes you have ugly data, it’s still good data but it doesn’t quite fit into the regular analyses



For parametric tests we want **normally distributed data**

The Mean and Median are more or less the same. And the Mean describes the data well

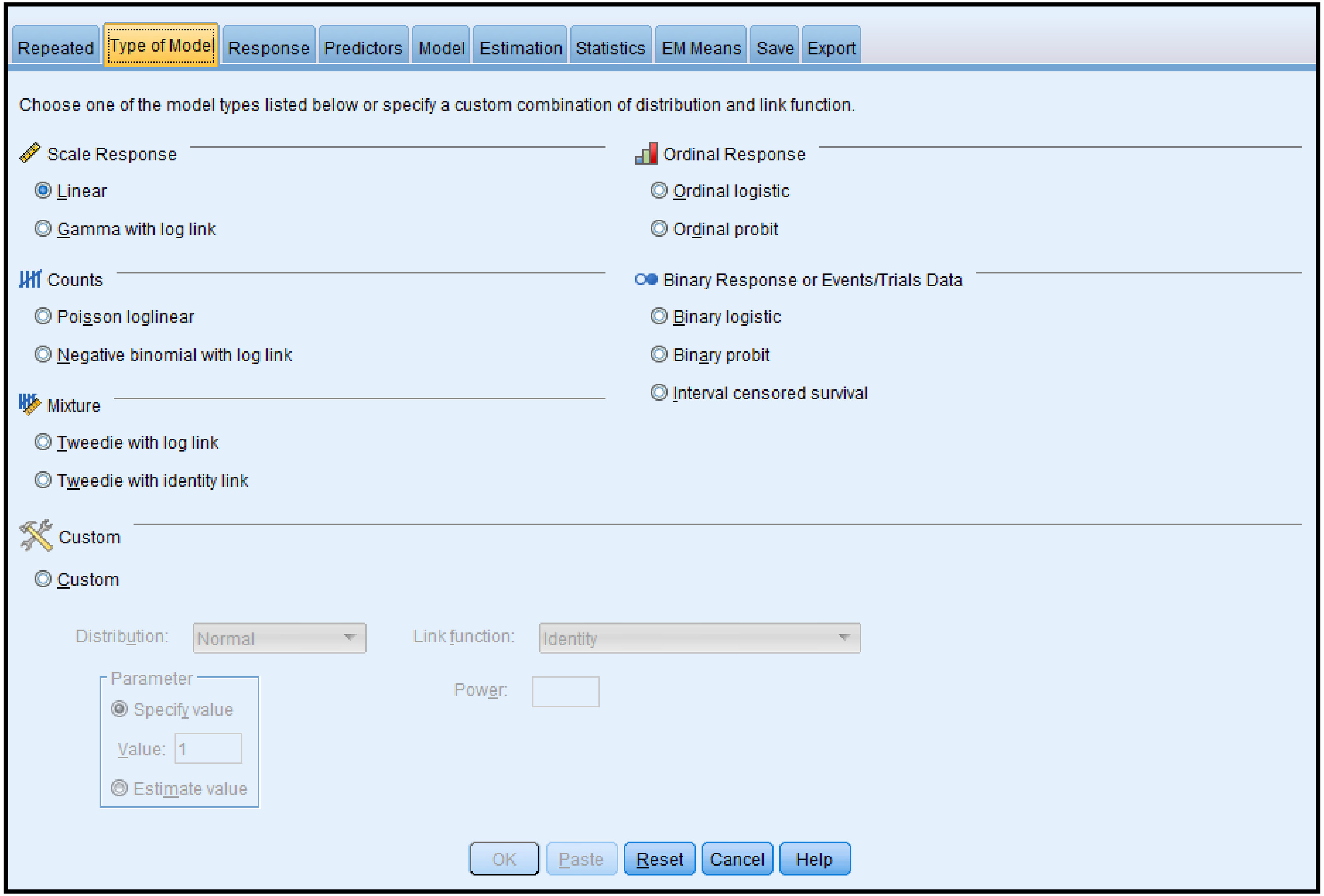
This allows us to **compare means** between distributions



Ugly data doesn't have a normal distribution.

The Mean and Median are different and the Mean no longer describes the data well

Comparing means is now **deceptive** and won't give us the right conclusion


Repeated
Type of Model
Response
Predictors
Model
Estimation
Statistics
EM Means
Save
Export
Choose one of the model types listed below or specify a custom combination of distribution and link function.
Scale Response
Linear
Gamma with log link
Counts
Poisson loglinear
Negative binomial with log link
Mixture
Tweedie with log link
Tweedie with identity link
Ordinal Response
Ordinal logistic
Ordinal probit
Binary Response or Events/Trials Data
Binary logistic
Binary probit
Interval censored survival
Custom
Custom
Distribution:
Normal
Link function:
Identity
Parameter
Specify value
Value:
1
Estimate value
Power:
OK
Paste
Reset
Cancel
Help

Generalized Estimation Equation

- **Binomial**: this is for binary data (yes/no).
- **Gamma or Inverse Gaussian**: this is appropriate for positive data skewed to larger positive values (like accuracy data).
- **Negative Binomial**: *"the number of trials required to get an x-amount of successes"*. It's for positive integer values.
- **Normal**: this is the familiar one and assumes the good old bell-shaped numeric distribution with a central mean.
- **Poisson**: the *"number of occurrences in a fixed period of time"*. This would for example work with cell-cultures and you're counting the number of cells (only positive integers).
- **Tweedie**: this is a mix between continuous and discrete distributions, the data has to be 0 or larger.
- **Multinomial**: the last one is for ordinal data, think Likert-scale values (agreement or even SES).

Generalized Estimation Equation

Model Information

Dependent Variable		Score
Probability Distribution		Normal
Link Function		Identity
Subject Effect	1	PP
Within-Subject Effect	1	Year
Working Correlation Matrix Structure		Unstructured

Goodness of Fit

	Value
Quasi Likelihood under Independence Model Criterion (QIC)	9600,873
Corrected Quasi Likelihood under Independence Model Criterion (QICC)	9600,873

Tests of Model Effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	203347,974	1	,000
Track	332,593	3	,000
Year	1973,460	3	,000
Track * Year	11,571	9	,239

The tables might have different names in them but they show the same information as the Mixed Model output.

You have your summary, goodness of fit and overall test. Based on these you make the same decisions and order the same EMMs

This Session

- What are Contrasts (adding them in the GLM)
- What are Polynomials (adding them in the GLM)
- How to add Contrasts and Polynomials in Mixed Models
- Generalized Estimation Equations (GEE)

This Course

The Logic of Repeated Measures

(why multiple sessions with a few is better than one session with many)

How to build Models and decide on covariance structures

(plus talk about what covariance structures are)

How to add covariates, random intercepts, and random slopes

(Making the Mixed Models truly Mixed)

Adding Planned and Polynomial Contrasts

(Thinking ahead like a good data scientist)

Session Evaluation