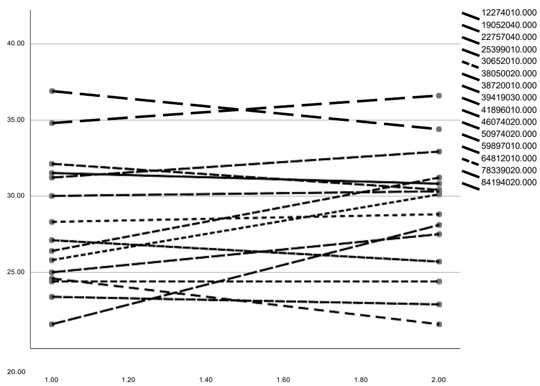
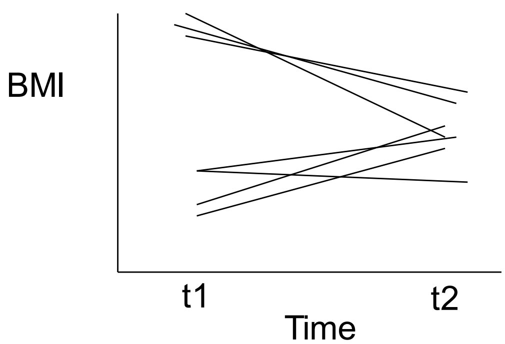
Longitudinal Examples

Below is a graph of the BMI (body mass index) values from the first time point to the second time point for a few individuals (from the Health and Retirement Study of middle-aged and older adults). Notice that those who start out with a higher BMI tend to decline, and those who start out with a lower BMI tend to increase over the two-year period. This is typical of regression to the mean, which occurs whenever there is a less than perfect correlation between measures at two time points. (It is also the case that the middle values of BMI at the beginning tend to move out to either higher or lower BMI scores).



Had we selected the individuals with the highest and lowest weight at the beginning, we would likely see a pattern like this.



If we had designed a study with an intervention in between t1 and t2, we would have assumed that the changes were due to the intervention. This also illustrates that we should expect a variable positively correlated with BMI to be negatively correlated with BMI difference scores—those high on BMI, and hence the correlated variable, tend to decrease in BMI over time; and those low on BMI (or the correlated variable), would tend to increase in BMI over time.

The correlation table below examines the relationship between vigorous activity measured at the first time point and BMI difference scores (`bmidiff = s8bmi - s7bmi`). First, notice that there is a negative correlation between BMI at the first time point (`s7bmi`) and the difference scores (`bmidiff`). This is typical because of the regression to the mean pattern we saw above. If we imagine for a moment that we had measured sedentariness (the opposite of vigorous activity), then the correlation between sedentariness and BMI at the first time point would be positive, .144. The correlation between sedentariness and BMI difference scores would be -.024, mirroring the relationship between BMI at the first time point and BMI difference scores. The result below is not significant, but you can imagine more extreme cases in which there is a higher correlation between the two variables at baseline and, thus, a significant relationship between the predictor variable and the difference scores that occurs simply because of the regression toward the mean phenomenon.

Correlations

S7BMI S7BMI: | S8BMI S8BMI:

				bmidiff	vigact7
S7BMI S7BMI:W7 Body Mass Index=kg/m2	Pearson Correlation	1	.888	-.153	-.144
	Sig. (2-tailed)		.000	.000	.000
	N	5182	4489	4489	5180
S8BMI S8BMI:W8 Body Mass Index=kg/m2	Pearson Correlation	.888	1	.318	-.145
	Sig. (2-tailed)	.000		.000	.000
	N	4489	4693	4489	4534
bmidiff	Pearson Correlation	-.153	.318	1	.024
	Sig. (2-tailed)	.000	.000		.108
	N	4489	4489	4489	4488
vigact7	Pearson Correlation	-.144	-.145	.024	1
	Sig. (2-tailed)	.000	.000	.108	
	N	5180	4534	4488	5277

Difference Score Regression

The simple regression with vigorous activity predicting BMI difference scores shows the same pattern as observed in the correlation matrix above, with the standardized coefficient equal to the correlation coefficient.

Model Summary

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.024 [a]	.001	.000	2.54719	.001	2.578	1	4486	.108

a. Predictors: (Constant), vigact7

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients				Collinearity Statistics	
Model		B	Std. Error	Beta	Std. Error	t	Sig.	Tolerance	VIF
1	(Constant)	.181	.070			2.603	.009		
	vigact7	.046	.028	.024	.015	1.606	.108	1.000	1.000

a. Dependent Variable: bmidiff

Lagged regression

The lagged regression (or ANCOVA) approach examines the relationship between vigorous activity and BMI but taking into account (partialling out) the initial relationship between vigorous activity and BMI.

Descriptive Statistics

	Mean	Std. Deviation	N
S8BMI S8BMI:W8 Body Mass Index=kg/m2	28.0880	5.48532	4488
vigact7	2.06194	1.341656	4488
S7BMI S7BMI:W7 Body Mass Index=kg/m2	27.8127	5.26103	4488

Model Summary

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.888 [a]	.789	.789	2.51852	.789	8399.900	2	4485	.000

a. Predictors: (Constant), S7BMI S7BMI:W7 Body Mass Index=kg/m2, vigact7

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	Standardized Coefficients Std. Error	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	2.329	.222			10.499	.000		
	vigact7	-.001	.028	.000	.007	-.045	.964	.974	1.027
	S7BMI S7BMI:W7 Body Mass Index=kg/m2	.926	.007	.888	.007	127.901	.000	.974	1.027

a. Dependent Variable: S8BMI S8BMI:W8 Body Mass Index=kg/m2

Though neither of the two regression approaches are significant, the difference score model has a larger effect and is much closer to the .05 level than the lagged regression model below. The discrepancy is a (modest) illustration of Lord's paradox. We might conclude different things from the two analyses in many instances. Because the difference score regression does not take into account the association between the predictor and outcome at baseline, we cannot know whether the association between the predictor and the difference scores might stem from regression toward the mean. The difference score regression helps identify who increases or decreases in absolute value over time, whereas the lagged regression attempts to examine temporal precedence and whether X precedes Y causally.