

Longitudinal Analyses in SPSS

Background

Longitudinal analyses look at how variables change across time. You can do different types of longitudinal analyses depending on whether your dataset is in wide or long format (see the handout on “Wide and Long Datasets”).

Wide-Format Analyses

When your dataset is in wide format, each participant has a single row. For example, in the below dataset, each person has “work_satisfaction” measured at three separate timepoints.

	employee_id	work_satisfaction.1	work_satisfaction.2	work_satisfaction.3
1	30490	5	5	5
2	30491	5	4	4
3	30492	5	4	4
4	30493	5	5	5
5	30495	4	.	.
6	30496	5	5	4
7	30497	4	3	4
8	30498	2	2	2
9	30499	4	4	4
10	30500	4	4	4

When your data are in wide format, you can easily perform analyses such as correlating variables across time, e.g.:

```
correlations work_satisfaction.1 work_satisfaction.2.
```

Or even running regressions, e.g.:

```
mixed work_satisfaction.2 with work_satisfaction.1 extraversion.1  
/fixed=work_satisfaction.1 extraversion.1  
/print=solution.
```

These types of analyses tell you how variables correlate with one another across time.

Long-Format Analyses

When your dataset is in long format, each participant has a row *for each measurement occasion*. Thus, if there are three waves in your study, each participant would have three rows. Each row contains the person’s scores on each variable across time. For example, in the below dataset, each person’s “work_satisfaction” is measured at three different timepoints.

	employee_id	time	work_satisfaction
1	30490	1	5
2	30490	2	5
3	30490	3	5
4	30491	1	5
5	30491	2	4
6	30491	3	4
7	30492	1	5
8	30492	2	4
9	30492	3	4
10	30493	1	5
11	30493	2	5
12	30493	3	5
13	30495	1	4
14	30496	1	5
15	30496	2	5
16	30496	3	4

The major benefit to this data format is that you can correlate variables **with time itself**. We call such analyses “growth models” (because they show how a variable *grows* across time). Such analyses simply use regression:

```
mixed work_satisfaction with time
  /fixed=time
  /random=intercept|subject(employee_id)
  /print=solution.
```

Notice that the analyses are super familiar! We’re simply looking at how work_satisfaction changes with time. But do also take note that when we run analyses across time we must (1) include the participant identifier (“employee_id” in this case) and (2) control for it as a “random effect” (which is why we have the `/random=` statement). The reason we need to do this is somewhat technical,¹ but you can think of in simple terms as: some of the data come from the same person (e.g., each person provided three waves of data) and thus we need to “statistically control” for each individual person’s idiosyncrasies.

It’s extremely popular to examine whether growth in a variable is different for different types of people, e.g., does work satisfaction change differently across time for men and women:

```
mixed work_satisfaction with time male
  /fixed=time male time*male
  /random=intercept|subject(employee_id)
  /print=solution.
```

¹ In order to produce accurate results, regression requires that all observations be “independent.” For example, each row in the dataset should come from a different person. When we have multiple rows in the dataset from the same person, the assumptions of regression are violated, and we will get inaccurate results. However, by controlling for each person’s identity as a “random effect,” we explicitly model and control for the dependence in the data, once again allowing regression to produce accurate results.