

MATH 60604A
Statistical modelling
§ 5b - Example of longitudinal data

HEC Montréal
Department of Decision Sciences

Example: desire for revenge and how it varies over time

- This example concerns the growing internet phenomenon of customer revenge.
- We will only examine the impact of certain variables on the desire for revenge and how this desire varies over time.
- The data in this example are fictional, but in the actual study, the data were taken from people who had made complaints on the websites `ConsumerAffairs.com` and `RipOffReport.com`.
- Five sets of questionnaires were sent to these individuals, one every two weeks.

Description of the measurements

- **Response variable:** desire for revenge, measured once for each questionnaire.
 - Mean of five items on the following scale, ranging from strongly disagree (1) to strongly agree (7).
 - For example, one of the items is "I wanted to take actions to get the firm in trouble".
- **Explanatory variables:** only taken after the first questionnaire — sex, age, and two variables measuring revenge-related behaviour:
 - **vindictive complaining:** based on four items such as "I complained to the service firm to give a hard time to the representatives".
 - **negative word-of-mouth:** based on three items such as "I spread negative word-of-mouth about the firm".

- A sample of 80 people participated in the study.
- The data are in the file `revenge.sas7bdat`.
- The variables are
 - `id`: subject id (from 1 to 80).
 - `t`: time of measurement (1 to 5).
 - `revenge`: desire for revenge (dependent variable).
 - `sex`: male (0) or female (1)
 - `age`: age in years.
 - `vc`: "vindictive complaining"
 - `wom`: "negative word-of-mouth"

Data for the first three individuals

SAS code to print only selected records

```
proc print data=statmod.revenge(where=(id<4));  
run;
```

Obs	sex	age	vc	wom	id	t	revenge
1	1	38	1	5.6666666667	1	1	4.6
2	1	38	1	5.6666666667	1	2	4
3	1	38	1	5.6666666667	1	3	3.6
4	1	38	1	5.6666666667	1	4	2.4
5	1	38	1	5.6666666667	1	5	2.4
6	0	28	1	1.3333333333	2	1	1.2
7	0	28	1	1.3333333333	2	2	1
8	0	28	1	1.3333333333	2	3	1.8
9	0	28	1	1.3333333333	2	4	1
10	0	28	1	1.3333333333	2	5	1
11	1	40	4		3	3 1	5
12	1	40	4		3	3 2	4.6
13	1	40	4		3	3 3	3.6
14	1	40	4		3	3 4	4.2
15	1	40	4		3	3 5	1.2

Evolution of desire for revenge over time

- It's important to understand the **structure** of the data, here "one measurement per line".
- Five lines in the file correspond to each individual.
 - there are no missing values.
 - Each of the five line corresponds to a measurement time t .
- The only variable that changes over time is **revenge**.
 - The variables `sex`, `age`, `vc` and `wom` were only measured at time one, but they are repeated over each time point.
- For longitudinal models, it is often necessary to format the data so that each line corresponds to a measurement (long format).

We need to be careful if we want to get descriptive statistics for variables that do not vary with time, such as sex, age, vc et wom.

- The sample average is correct only because we have the same number of observations $T = 5$ for each individual.
- On the other hand, its estimated standard error will be too small because the effective sample size $N = 80$ is smaller than the number of measurements $NT = 400$.

Example calculation

The average age of the $N = 80$ participants is $\overline{\text{age}} = 42.075$ years

- the standard deviation is $S = 7.49$ and $\text{se}(\overline{\text{age}}) = 0.837$, where

$$S^2 = \frac{\sum_{i=1}^N (\text{age}_i - \overline{\text{age}})^2}{N - 1}, \quad \text{se}(\overline{\text{age}}) = \frac{S}{\sqrt{80}}.$$

Contrast this with the following **incorrect calculation**

$$S_*^2 = \frac{\sum_{i=1}^{NT} (\text{age}_i - \overline{\text{age}})^2}{(NT - 1)} \approx S^2, \quad \text{se}(\overline{\text{age}}) \neq \frac{S_*}{\sqrt{400}} = 0.37$$

If a variable is repeated over time, we can focus on a single time measurement.

SAS code to compute summary statistics

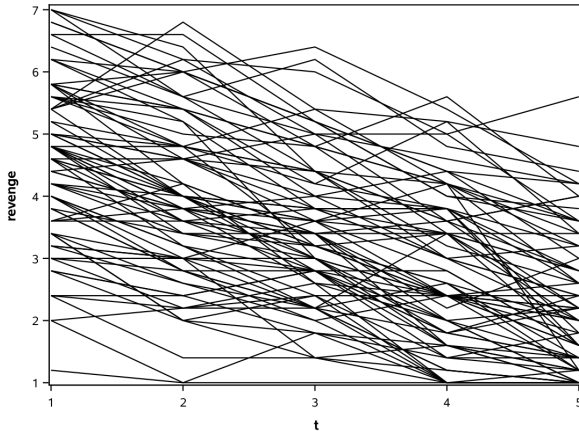
```
proc means data=statmod.revenge(where=(t=1));  
var sex age vc wom;  
run;  
proc corr data=statmod.revenge(where=(t=1));  
var sex age vc wom;  
run;
```

- It's possible that the desire for revenge changes over time. A simple way of visualizing the time evolution of desire for revenge is to plot revenge as a function of t for each individual.

SAS code to draw a spaghetti plot

```
proc sgplot data=statmod.revenge;  
series x=t y=revenge / group=id;  
run;
```

Spaghetti plot



This so-called spaghetti plot shows 80 curves (one for each person). The desire for revenge value seems to decrease with time, on average. The decrease in `revenge` through time could very well be linear.