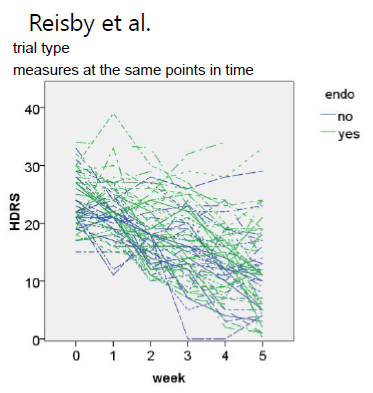
**Longitudinal data**

= repeated measures of individuals over several time points

* Observations on the same individual will not be independent
* Expecting measures closer to each other to be more similar than the ones further away

**Example lecture**:

* Endogenous & exdogenous depression
* Depression scores (HDRS) measures weekly over 5 weeks (start at 0 before treatment)
* Question: Is the pattern in the depression scores over time different for patients with endogenous compared to exdogenous depression?



**Should time be looked at as continuous or categorial?**

* Fixed effect for intercept, time, group & group\*time
* Main effect for time

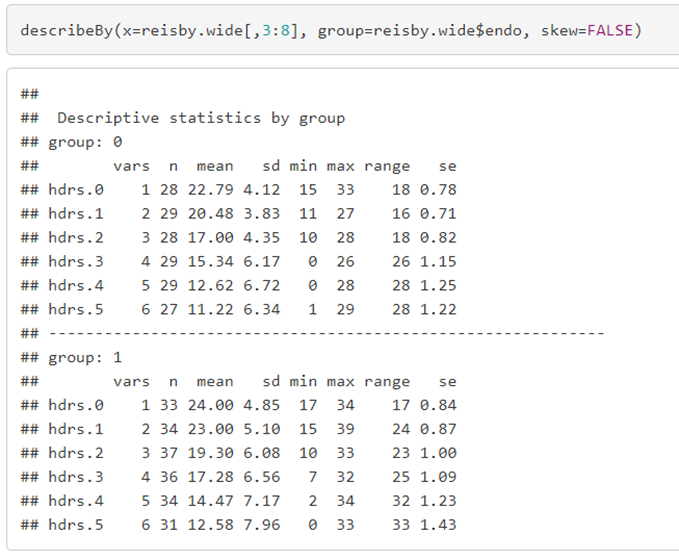
**Multiple measurements**

* random effects
  + intercept? (each patient has different starting point)
  + slope of time? (variation in the slopes – some increase/decrease slower/quicker than others) – Is there random effects for slope over time?

Adding randomness because we are expecting correlation of repeated measures within the patients

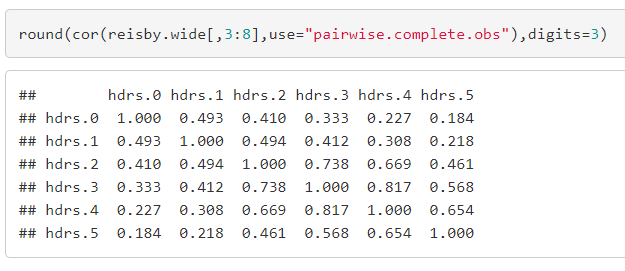
1st level = time measurements

2nd level=individual (repeated measurements)

**Descriptive statistics before modeling**

**average depression scores are decreasing in both groups over time**

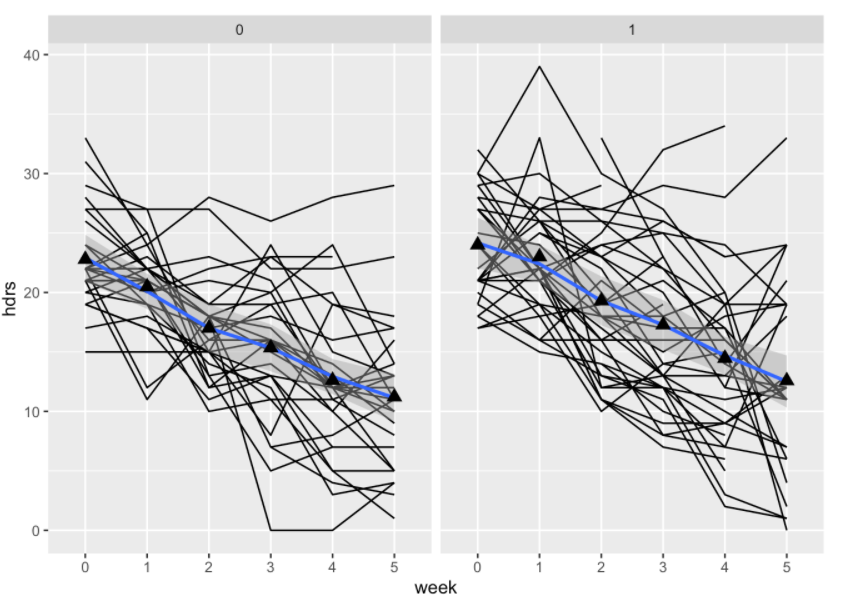
**less variation in the first measurements than later = patients tend to be more similar to one another in the beginning of the study**

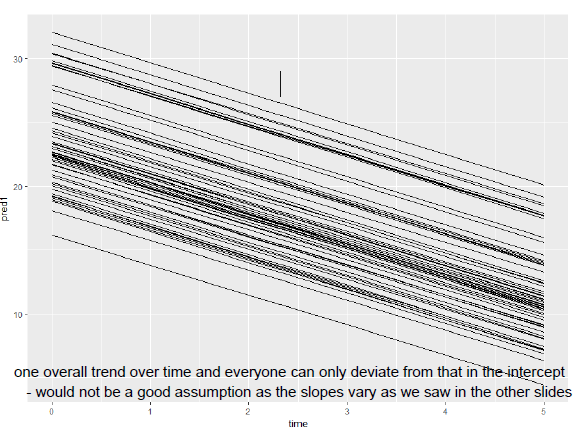


Correlation between the depression score measurements are decreasing over time = less correlated the further away from diagonal

**Linear time effect**

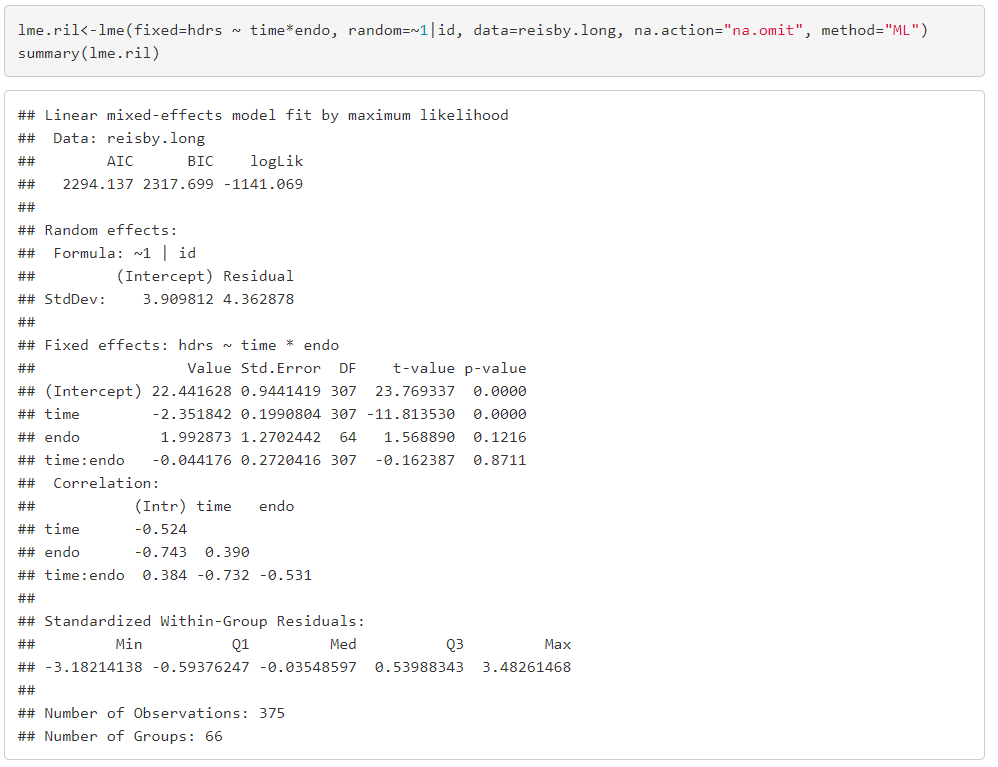
* Question: Would it be reasonable to measure time as a continuous (linear) parameter?



* + Only a slope of depression score in time

**Then patients only deviate in their intercept but would all have the same slope for their depression scores – compare to initial plot**

* Incorporate a random (slope) linear time effect as we assume that they differ for each patient

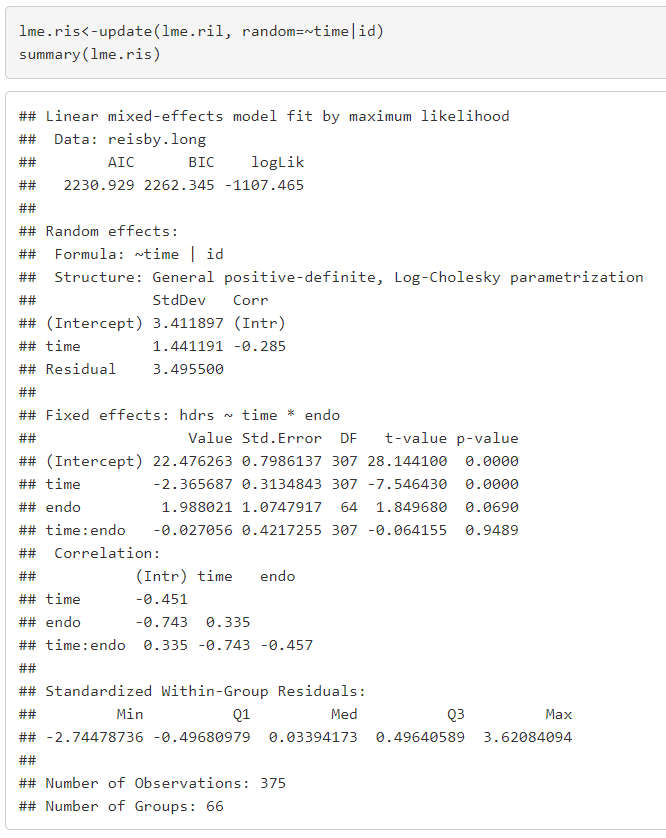
**mixed model with fixed effects for time (time as continuous and linear), endo and the interaction of time\*endo, and just a random intercept per person**

**negative slope for time= -2.352**

**Mixed model with**

fixed effect = time, endo & interaction

Random effect= random intercept & random slope for time per subject

in addition to previous one it shows that interaction between endo and exdo over time is not significant

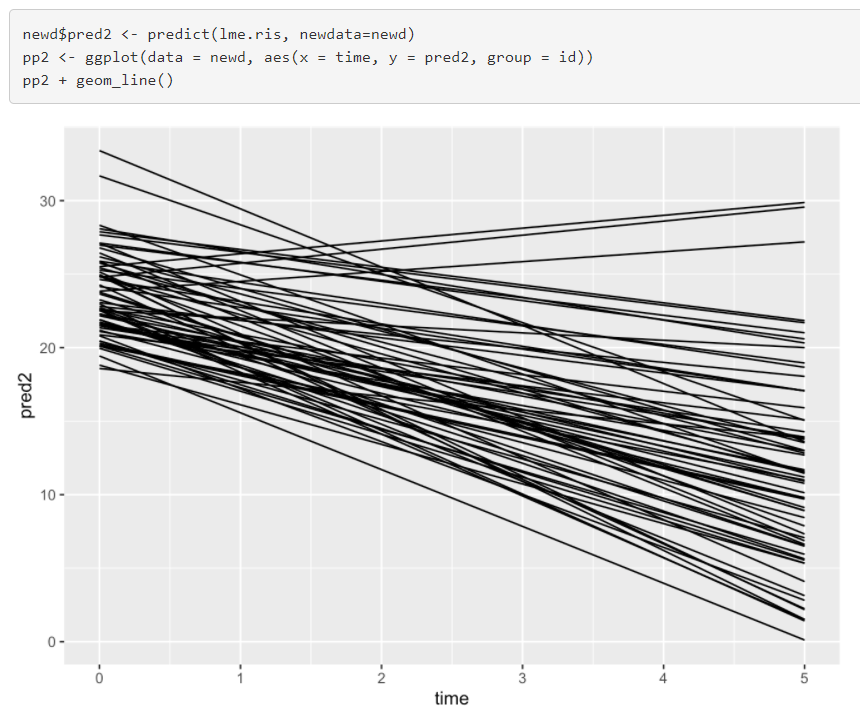
**intercept =22.48** = average depression score when all variables =0

**negatively correlated** = intercept is higher and slope is more negative because the higher the patients started (the worse the depression) the more likely the values will go down

* Often true in longitudinal data- the higher the intercept the more negative the slope

**Estimate for endo =1.99** = average difference in depression scores between endo and exo patient at time=0

**sd of intercept and slope**

3.41 = considerable fluctuation around fixed intercepts (starting point)

Random intercept + random linear time effect explains it better but

**What if unreasonable to use time as continuous?**

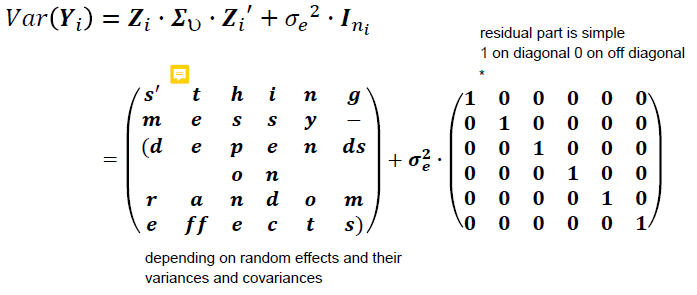
* Categorial time as a fixed effect in the model

Linear Mixed Model (LMM):



= fixed intercept + fixed time slope + random intercept + random time slope + vector of residuals

p.20

**LMM matrix formulation & var-covar matrix**

**CPM (Covariance pattern model)s & var-covar matrix**

* Fixed effects and just adding residual variation (matrix) which would have to explain all “messiness” from the original data

no random effects!

To take the correlation between measurements of the same person into account we use variance-covariance matrix

it has to be complicated enough to reflect the data

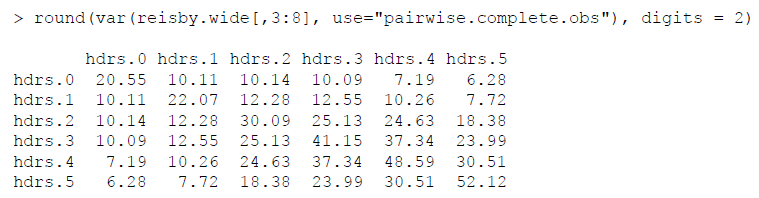
assume correlation for **∑** because residuals are not independent

**Possibilities for modelling correlated measures**

**LMM:**

* Model correlation of the measurements **implicitly**
* **Repeated observations (Level 1) are nested within patient (Level 2)**
* Handle with:
  + Random intercept per patient
  + random intercept per patient + random slope
  + etc.

**CPM:**

* Model correlation of the measurements **explicitly**
* Covariance structure of the residuals is incorporated into the model
* Assumes equal time between measurements

Shows that variances increase over time & measurements differ more the more they are apart

**Correlation structures:**

**Best for longitudinal data**

* Unstructured correlation
* Autoregressive of order 1

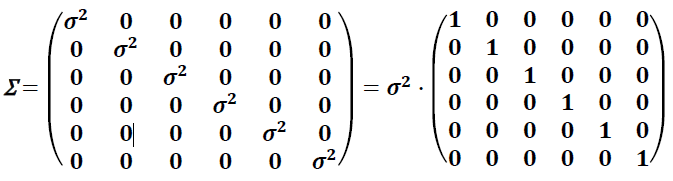
**Bad ideas**

* Uncorrelated/independent
* Compound symmetry

**Independent correlation structure**

Assumptions

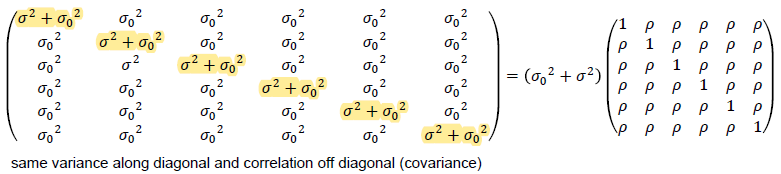
* residuals independent (as though from different subjects)
* All variances equal
* **All correlations within person & between different time points = 0** (not realistic especially for longt. Data)
* Like assumptions of ordinary linear regression/ANOVA



Would not make sense for our example. Even when observations were independent, the design of the study is to take random patients and measure multiple times not random sample for each time point (no matter the same patient) – patients ARE correlated

**Compound symmetry correlation structure p.29**

Assumptions:

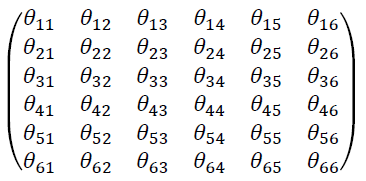
* correlation between the time points to be equal no matter how far apart
* All variances are assumed equal

**In this case**: assuming that time has constant variance and constant correlation over time measurements

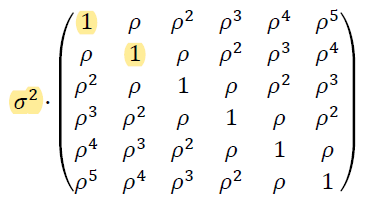
* CPM with compound symmetry pattern for the residuals = linear mixed model with random intercept per patient

**Unstructured correlation**

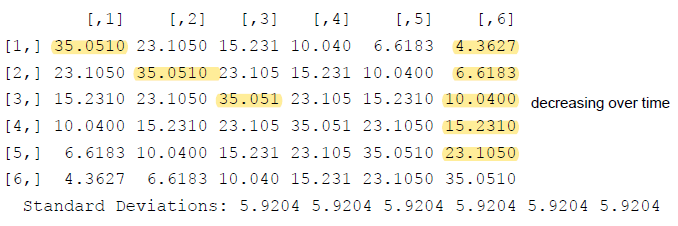
* Most complicated
* Separate variance of the residuals at each time point & different covariance between the time points
* Many df
* Good fit but very costly
* No random effects



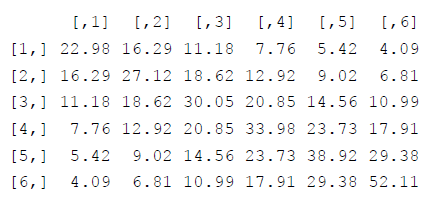
**(Homogeneous) autoregressive of order 1 correlation**



**Assumptions:**

* Equally spaced time intervals
* All observations that are one time unit apart have the same correlation
* Two units apart have the correlation (p) to the power 2, 3 units = observation to the power 3
* Same fixed effects and no random effects
* Pattern for residuals shows that sd and variance are assumed constant over time
* Decreasing variance and covariance over time (the further away from each other)

**AR1 with heterogeneous variances - autoregressive of order 1 correlation**

* Allows for variances to differ over time (diagonal not fixed like above)
* Nice compromise with 6 variance parameters and 1 correlation parameter
  + 7 parameters which explain a nice amount of variance

**Covariance patterns of linear MM p.36**

* Random intercept model implies a compound symmetry structure for all data combines
* Linear mixed model with random intercept and random slope also implies a correlation structure for the data but very complex
  + Structure depends on the estimates but usually the variances increase later in time and correlations decrease the further time points are apart

**Baseline Measurement**

* Usually baseline measure before randomizing the patients **(clinical trials)**
* Is baseline an outcome? Because outcome of a trial is then everything measures after the randomization
  + Yes: baseline = first outcome measurement in MM
  + No: ignore the first measurement
  + No: use as covariate in the model and not as an outcome?
* **Observational study** (no experimental intervention)
  + Usually baseline = first of the measured outcomes
  + Not a randomization issue
  + Should it be used as covariate?

**p.40**