FES 524: Natural Resources Data Analysis

Reading 6.2

**Contents**

[Other readings to do before class 1](#_Toc63682264)

[Correlation matrix 1](#_Toc63682265)

[Matrix showing no correlation 2](#_Toc63682266)

[Matrix showing correlation 2](#_Toc63682267)

[Correlation matrix with symbols 3](#_Toc63682268)

[Covariance matrix 3](#_Toc63682269)

[Specific correlation structures 4](#_Toc63682270)

[Mixed model induced correlation 4](#_Toc63682271)

[Compound symmetry 5](#_Toc63682272)

[AR1 5](#_Toc63682273)

[General correlation 6](#_Toc63682274)

[Correlations available in nlme 7](#_Toc63682275)

[Choosing a correlation structure 7](#_Toc63682276)

[For longer series 8](#_Toc63682277)

[For short series 8](#_Toc63682278)

[Information criteria 8](#_Toc63682279)

[Algorithm for choosing a correlation structure for a short series 9](#_Toc63682280)

[Written description of algorithm 9](#_Toc63682281)

[Model description and analysis 10](#_Toc63682282)

[Describe correlation in words 10](#_Toc63682283)

[R details 10](#_Toc63682284)

[Equivalence testing 11](#_Toc63682285)

# Other readings to do before class

Read the description of the study that assignment 6 is based on in Handout 6.2 and come prepared to discuss the research question, comparisons, and aspects of the study design.

# Correlation matrix

We’ll be talking more formally about correlations today. Like with statistical models, the mathematical representation of correlations can be an efficient way to describe and understand correlation structures. You will only be asked to describe any correlations on assignments in words, not with mathematical notation, but the goal of introducing correlations as symbols is to help in your overall understanding of correlations.

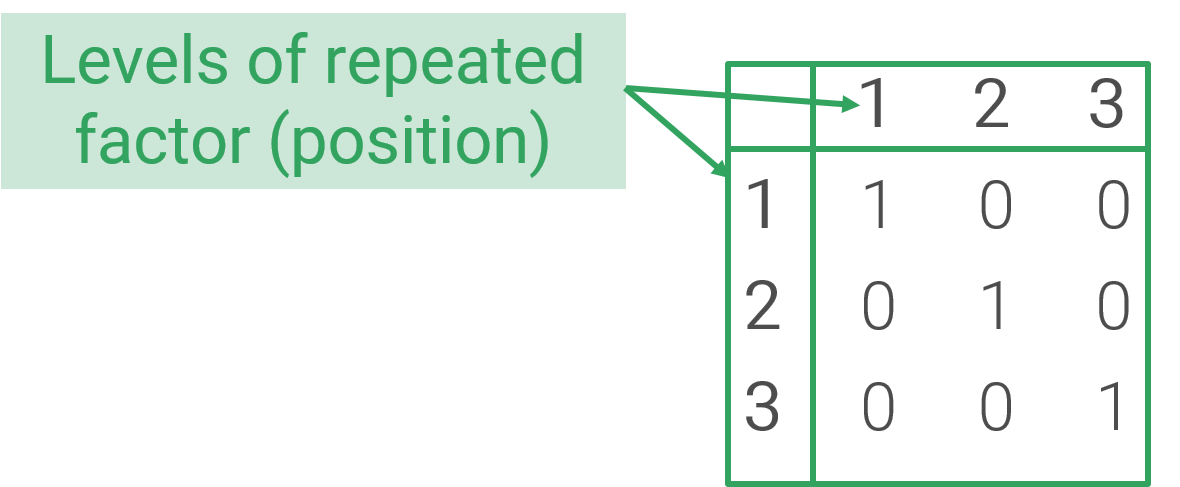
A *correlation matrix* is one way to describe the within-subject correlations among errors. A correlation matrix is a square matrix. The levels of the repeated factor make up both the rows and columns. We’ll go through several examples.

## Matrix showing no correlation

Below is an example of a correlation matrix using the study design from the motivating example. Oak position is the repeated factor, which has levels 1, 2, and 3, measured on transects (the repeated subject). You can see that these levels make up both the rows and columns.

Along the diagonal of the matrix are the *within-position* correlations of errors. The errors are always perfectly correlated (100%) with themselves, which is why you see the 1’s along the diagonal.

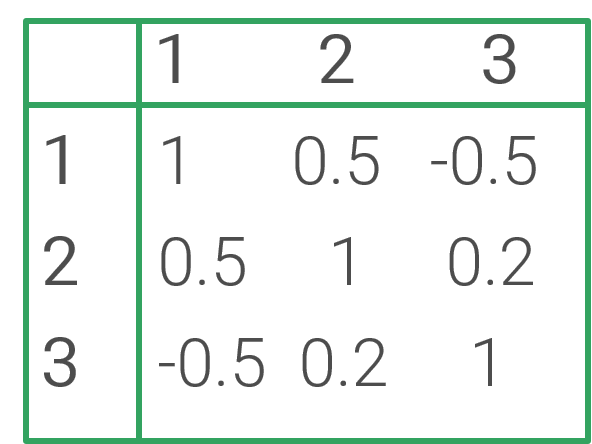
In this example matrix there are no correlations among errors from different positions. Therefore, all off-diagonal values are 0.



## Matrix showing correlation

The example above shows correlations from a model that has met the assumption of independence of errors. Now let’s take a look at a correlation matrix that has within-subject correlations of errors. These are correlations of the errors among the three positions.

Let’s look at an example of a correlation matrix with non-0 correlations of errors among positions within transects. We can see there are now non-0 values in cells off the diagonal. These values can range from -1 to 1 (perfect negative correlation to perfect positive correlation).



The matrix is showing us the correlation of errors within transects for different positions. Each correlation is listed twice, once in the *upper triangle* of the matrix and once in the *lower triangle* of the matrix. Be sure you are able to identify the upper and lower triangle here; each contains 0.5, -0.5, and 0.2.

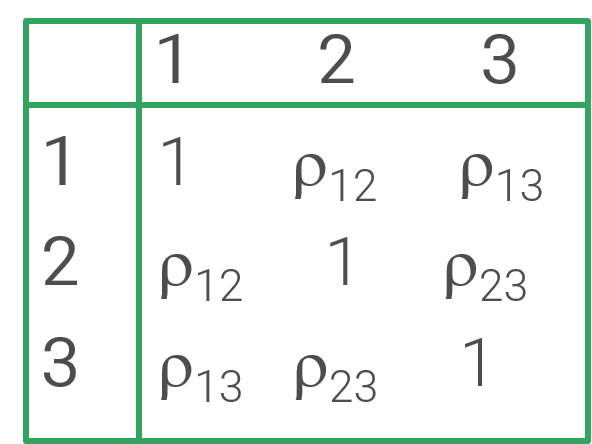
You can see that there is a positive correlation of 0.5 between position 1 and position 2 errors. This is the value in the first column (position 1) and second row (position 2) as well as the second column (position 2) and first row (position 1).

Similarly, you can see that there is a negative correlation between positions 1 and 3 of -0.5 and a positive correlation of 0.2 between positions 2 and 3. Make sure you know how to find the values and the positions each correlation relates to in both the upper and lower triangle.

## Correlation matrix with symbols

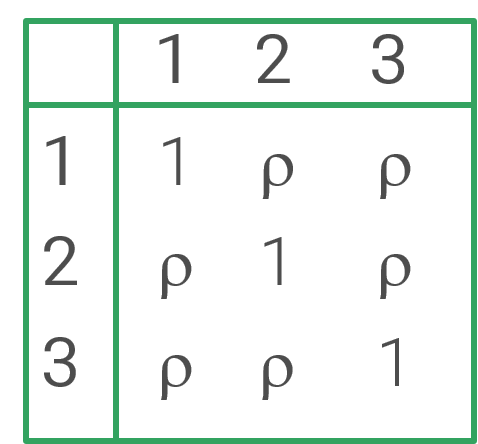
So far I’ve shown you examples using actual numbers in the correlation matrix. However, it is common to write correlation matrices using symbols to represent different correlations.

Here is a matrix showing different correlations among errors for different positions. This is like the correlation example above but uses symbols instead of numbers.



The correlations are represented by the symbol ρ. The subscript on the symbol tells us which positions the correlation is for. For example, the correlation between positions 1 and 2 is represented by ρ12, which is again in column 1 row 2 as well as column 2 row 1.

If the correlations of errors among positions are all the same, a single symbol will be used on the off-diagonal of the matrix. Since correlations are all identical in the example below, the symbol ρ is used with no subscripts.

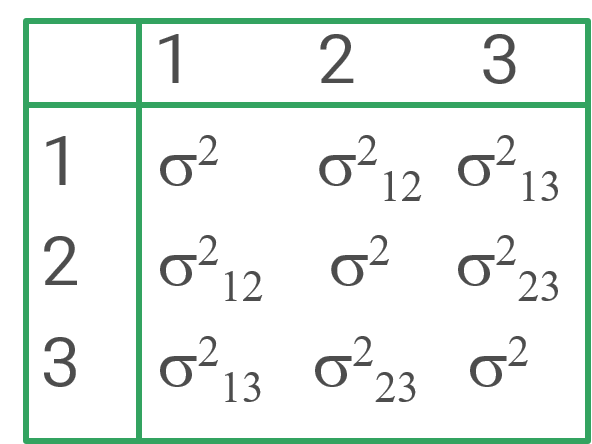


## Covariance matrix

Another type of matrix you may see instead of a correlation matrix is a *covariance matrix,* showing the *within-subject* variances and covariances. Correlations are calculated using variances and covariances. We can discuss the relationship between correlations and covariances/variances more in class, as needed.

Below is an example covariance matrix using the motivating example study. Now the diagonal of the matrix contains the variances for each group. Since these values have no subscripts on them, we are assuming constant variance of errors among positions.

The off-diagonal of the matrix shows the covariances among positions. This particular example shows a different covariance for each pair of positions, which you can see from the subscripts.



# Specific correlation structures

There are many different correlation structures. We will go over four of them today.

## Mixed model induced correlation

We’ll start by formally talking about the kind of correlation that is implied or *induced* when used random effects in a mixed model. We will need to use more formal language than we have so far to talk about this, which I’ll introduce below.

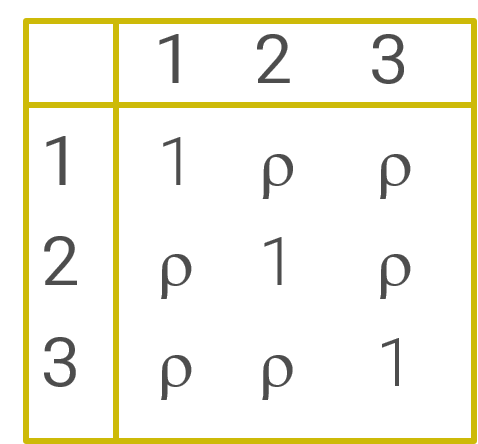
You already know that after accounting for the fixed and random effects we assume the model errors are independent in a linear mixed model. Another way to say “accounting for” is *conditioning on*. The errors are independent *conditional on* the variables included in the model. We can therefore refer to these errors as *conditional errors*. The conditional errors from a linear mixed model are assumed to be independent.

As we’ve discussed, we believe that observations taken from the same physical unit are correlated with each other. This is one of the reasons we’ve been fitting mixed models. Such correlations are called *marginal* correlations; these are the correlations among errors without conditioning on the random effects. The marginal errors are the errors after taking only the fixed effects into account.

The *marginal* errors from a linear mixed model are correlated. There is an overall within-subject correlation of errors for all pairs of the repeated measurements. This makes sense, since we believe there is some systematic effect of the variable we are using as a random effect on the response variable.

These within-subject marginal correlations from a linear mixed model must be strictly positive. We cannot model negative correlation among errors using a mixed model. This is usually fine; since we expect errors to be more alike within groups than between groups we expect to have positive correlation.

Here is a correlation matrix the represents the marginal correlation *induced* by the mixed model fit to the motivating example dataset. We estimate a single correlation that is the same among all pairs of positions.



There is no need to have equal spacing among with-subject observations when fitting a mixed model.

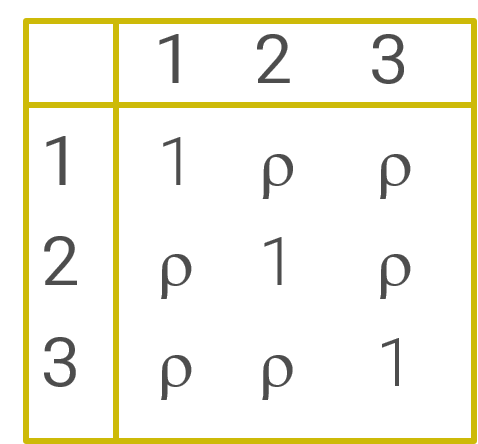
We can model correlations of errors when we don’t have any random effect using the gls() function from package nlme in R. We will not see these models in this class.

## Compound symmetry

All other correlation structures we’ll talk about are correlations of the *conditional* errors, the leftover correlation after we take into account both the fixed and random effects from a mixed model. We’ll start with compound symmetry. Compound symmetry is the simplest kind of correlation structure. When using a compound symmetry structure we calculate a single correlation, which is shared among all within-subject repeated measurements.

This correlation from compound symmetry is rarely different than the correlation induced by a mixed model. The main instance where compound symmetry adds something more than what we got from having a random effect in the model is if the within-subject correlation is negative.

Below is the correlation matrix the represents the correlation from a compound symmetry correlation structure fit to the motivating example dataset. This is the conditional correlation among errors; the correlation among errors after accounting for the fixed and random effects in the mixed model. You should recognize that this correlation matrix is identical to the marginal correlation matrix of the mixed model. The difference is that these correlations could be negative where the induced correlation from a mixed model must be positive.

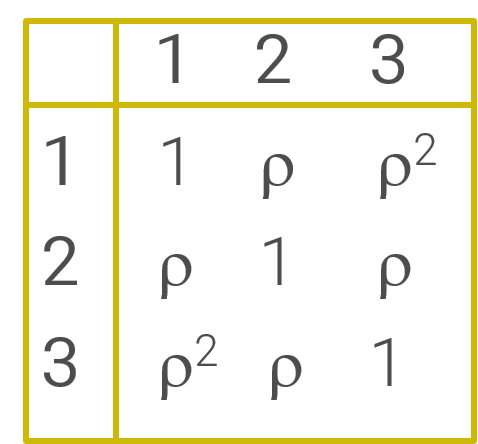


There is no need to have equal spacing among with-subject observations when using the compound symmetry correlation structure.

## AR1

Autoregressive lag 1 correlation, more commonly known as AR1, is commonly used for time series but can be used for some repeated measures in space. In AR1, correlations systematically diminish as observations get farther apart in space or time. When using an AR1 correlation structure, we estimate only a single correlation. This correlation is used to describe how correlations diminish with increasing space or time between measurements.

Below is the correlation matrix that represents AR1 correlation for a model fit to the motivating example dataset. There is a single correlation, ρ.



In the correlation matrix you can see that correlations of errors 1 unit apart are correlated ρ. Positions 1 and 2 are 1 unit apart and positions 2 and 3 are 1 unit apart. This correlation can be negative or positive, ranging from -1 to 1.

Correlations of errors 2 units apart are ρ2. In this example, only positions 1 and 3 are 2 units apart. Make sure you understand why ρ2 is smaller than ρ.

If relevant, correlations of errors 3 units apart are correlated ρ3, correlations 4 units apart ρ4, and so on. The amount of correlation diminishes with increasing distance between positions.

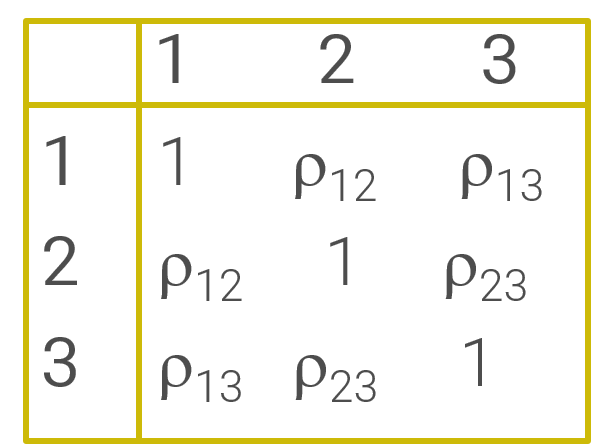
AR1 correlations only make sense if the repeated measures within a subject are equally spaced. This is most common with time series (e.g., observations every day or every year) but can be relevant for repeated measures in space. Since oak trees in the plantation in the motivating example are all equally spaced, we can use AR1 to model the correlation of the conditional within-transect errors.

## General correlation

The general correlation structure, similar to what SAS calls *unstructured* correlation, is the most complex correlation structure we have for package nlme.

The general correlation structure estimates a separate correlation for every pair of levels from the repeated measurement factor. This means we end up estimating a lot of correlations if we have a lot of repeated measurements. This structure can lead to overly complex models when we have relatively little data and many repeated measurements. Because of this, the general correlation structure is rarely advised for situations with many repeated measurements. In many cases we simply can’t justify estimating many separate correlations given our sample size.

The general correlation structure has no pattern, as we can see below for a correlation matrix representing the motivating example. Every pair of positions has a unique correlation, unrelated to the other correlations.



The number of correlations we end up estimating when using the general correlation structure is the number of levels in the repeated factor choose 2. In the motivating example we have three levels of position. 3 choose 2 equals 3, so we estimate 3 unique correlations. You can see this in the matrix above.

Within-subject observations do not need to be equally spaced to use the general correlation structure.

## Correlations available in nlme

There are more correlation structures available in the nlme package than we will see in this class. Below is the complete list. See the documentation for more details, ?corStruct.

corCompSymm() compound symmetry

corSymm() general

corAR1() \* autoregressive order 1

corARMA() \* autoregressive-moving average

corCAR1() continuous AR1

corExp() exponential spatial

corGaus() Gaussian spatial

corLin() linear spatial

corRatio() rational quadratic spatial

corSpher() spherical spatial

\*These two “time series” correlation structures require equal spacing among within-subject observations. If you do not have equal spacing you should not consider using these correlation structures.

# Choosing a correlation structure

Overall, choosing a correlation structure is easier if we have some idea of what sort of correlation we expect. This expectation might be due to previous research or logical justification based on scientific knowledge of your field along with an understanding of different correlation structures.

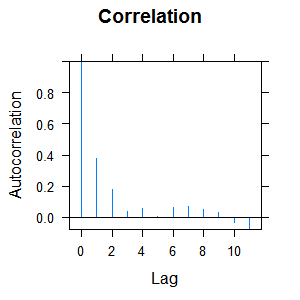
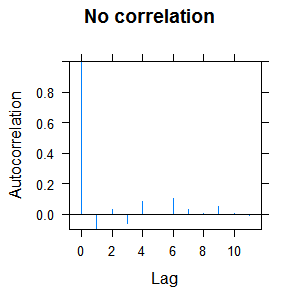
We may have known limitations when choosing a correlations structure. For example, we might be limited to relatively simple structures where we don’t estimate many different correlations because of a limited sample size. Alternatively, we could have limited structures to choose from because our repeated measurements are not equally spaced.

Keep in mind that in many cases we are not interested in exactly what kind of within-subject correlation of errors we have. Instead, we need to reasonably model the correlation in order to have valid model for inference; otherwise the model assumptions would not be met.

## For longer series

In a lot of ways, the job of choosing a correlation structure is easiest when we have a longer series of repeated measurements. As you saw in reading 6.1, we can check for the presence of correlation in the residuals using ACF plots or semivariograms. These can help us get an idea of what sort of correlation we might have if we don’t already have a good idea of what it might be.

If we can identify a reasonable correlation structure, we can then use that correlation structure to model the within-subject correlations of errors. We can use ACF plots/semivariograms of normalized residuals from our new model to see if that has improved things. Below are example ACF plots of the residuals showing correlation in the initial model and then no correlation among the normalized residuals after allowing for a correlation structure. The ACF plot on the left is an example of AR1 correlation.

## For short series

When working with a short series of repeated measurements we are much more limited in what we can do. There aren’t any useful residual plots that I have ever seen to help us figure out what kind of correlation we might have.

We usually default to assuming there is some sort of within-subject correlation based on our scientific knowledge and the study design. Our only option is then to try a few *reasonable* structures and choosing a “best” one. What is reasonable will need to be based on the spacing of the measurements, the number of correlations we feel comfortable estimating based on the sample size, and our *a priori* knowledge of correlations that have been used in the past for a similar response variable.

Since we can’t use graphical checks on this, we are forced into the uncomfortable position where we must rely on some statistical measure to help us choose the correlation structure.

## Information criteria

What statistical metric can we use to help choose a correlation structure?

Since models with different correlation structures are not nested, we cannot use p-values. We’ll want a metric that balances how well the model fits the observed data with not fitting too many parameters given limited sample sizes.

Our primary option, then, is to use some information criterion such as AIC or BIC. We will not be formally discussing how to decide which one of these to use in this class. However, you need to pick the one you will use before doing the analysis and stick with it. In this class we will practice using AIC because of the way it balances model fit and number of parameters. AIC will more often allow models with correlation structures to be chosen.

Recognize that this use of information criteria is a very ad-hoc approach. This is not how these tools were designed to be used. Review some of the information in reading 5.2 and the recommended resources there, especially Burnham et al. 2011 and Burnham and Anderson 2002, if you use an information criterion in your work other than for this ad-hoc approach.

Don’t forget that all other assumptions of the model must also be met before making inference. Choosing a correlations structure for repeated measures data is only one step in your model checking.

## Algorithm for choosing a correlation structure for a short series

1. Select an IC that you are going to use to choose among models. You will use only one, and this needs to be selected before proceeding.
2. Fit a mixed model with all relevant fixed and random effects in it. Make sure you are using REML (restricted maximum likelihood) for the model fitting, which is the default in package nlme.
3. Based on the spacing of the repeated measurements, your understanding of the problem, and any limits on sample size and the number of parameters you can estimate, choose a few correlation structures that make sense for your study design.
4. Fit models with these correlation structures along with the same fixed and random effects as your initial model. Make sure you continue to use REML and that the fixed effects are the same for all models. You may need to address problems with other assumptions at the same time (i.e., if you need to relax the assumption of constant variance then do so for all models).
5. Choose a model from the small set (3-4) models based on the IC you selected. In R, smaller is better for IC. The initial mixed model will be one of the models in the set.
6. Check assumptions of your chosen model, using *normalized* residuals. If things look good, go on to use that model for any inference/estimation. Otherwise pick an alternative modeling approach and a make a new model set.
7. Describe all the models you considered and include a table showing the models and the IC values in your write-up.

## Written description of algorithm

Using this algorithm, here is how I would describe the process I used for choosing a correlation structure for the motivating example. This would fit in a “methods” section:

Four different models were proposed to account for among-position correlations within transects. The best model from this pool was selected based on the minimum AIC value. Every model considered contained the random effect of transect and the same fixed effects. The following models were in the set:

Linear mixed model

Linear mixed model with compound symmetry correlation of errors

Linear mixed model with autoregressive lag 1 correlation of errors

Linear mixed model with general correlation of errors

In a “results” section I would then report:

A model with a general correlation structure among errors had the minimum AIC value (Table 1) and this model was used for inference.

**Table 1:** The number of parameters (df) and AIC values for the set of models used to choose a correlation structure of errors within each transect. The model with the lowest AIC value was selected.

|  |  |  |
| --- | --- | --- |
|  | **df** | **AIC value** |
| Linear mixed model | 8 | 281.5 |
| Linear mixed model with compound symmetry correlation of errors | 9 | 283.5 |
| Linear mixed model with autoregressive lag 1 correlation of errors | 9 | 282.2 |
| Linear mixed model with general correlation of errors | 11 | 271.4 |

# Model description and analysis

## Describe correlation in words

You should be able to describe the correlation structure of the model you use for inference in words. This means you will describe the kind of within-subject correlation structure that you used in your final model in enough detail to make it clear what sort of correlation of errors you allowed for. You will also need to describe any nonconstant variance you allowed for, if relevant. This information will be included as part of your “analysis strategy”.

Here is an example of a description I use for the correlation structure for a model fit to the example data. This sentence is in addition to the standard description of the type of model used and all effects (fixed and random) in the model:

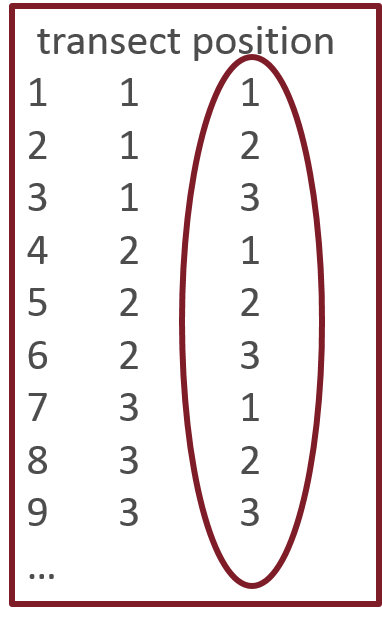
We extended this linear mixed model to allow for among-position correlations of the errors within transects. The correlations among errors within each transect were described by a general correlation structure, where the errors for each position have a unique correlation with the errors of every other position within a transect.

I would use an additional sentence to describe relaxing the assumption of constant variance (if used):

Variances of the errors were allowed to vary among the different positions.

## R details

Correlation structures



To use residual correlation structures in function lme() in R, we’ll need a variable that represents the within-subject spacing of measurements. For our example case, this could simply be 1, 2, 3 for each transect (represented by position).

It is always going to be “safest” to have a variable to represent spacing. R will treat things as if they are spaced equally and in the order they appear in the dataset if we don’t have a variable that represents spacing, at least for the non-spatial correlation structures we are using.

I’ve pasted an example of the variable that represents equal spacing, the integer-valued “position” variable, to the left. This variable must be an integer and not a factor to work with correlation structures.

If we have unequal spacing, we need a variable or variables that can represent the “distance” between observations within each subject. If the oak trees in our plantation weren’t equally spaced, for example, we would want a variable to show how far each tree is from the beginning of the transect. See the hypothetical example to the left that contains a “distance” variable.

Latitude and longitude are common “distance” variables for repeated measures in space. For time series a “time” variable is most often used to represent distance (e.g., doy, Julian day, continuous date-time).

Normalized residuals

Note that if you’ve allowed for a correlation structure when using function lme() in R, you must check the assumptions of that model using the *normalized* residuals. Normalized residuals have had the covariance divided out of them. If you were seeing issues of correlation in the residuals in an initial model and then fit a model allowing for a correlation structure of the errors, you should no longer see an issue in the normalized residuals of the second model if the correlation structure you used was sufficient.

See the documentation, ?residuals.lme, for a somewhat more thorough definition of normalized residuals.

# Equivalence testing

A topic that comes up in this week’s assignment that we haven’t discussed yet has to do with research questions, where the goal is to state there is no difference in means among groups. This is a problem. Fisherian hypothesis testing is set up to only show evidence *against* the null hypothesis. We don’t get to “accept” the null hypothesis when we are assuming the null hypothesis is true as the basis for calculating a p-value. The way you’ve been taught to report p-values and confidence intervals in this class should help avoid some of this problem.

You will see researchers use p-values to conclude “no difference” and use that as evidence to say that two groups are equivalent. This is poor statistical practice. As you know, it is easy to get a big p-value: just make the sample size super small. Because of this, if the goal is to show no difference then researchers must switch to using some sort of *equivalence* testing. Equivalence tests are also called noninferiority trials in some fields.

We are not going to have a chance to learn much about equivalence testing. One main approach to such tests, though, uses confidence intervals a lot like we’ve already been using them. In equivalence testing, two groups are declared to have evidence of equivalence at some alpha level if the confidence interval around the difference contains only values that are smaller in absolute value than some practically important value. The practically important value is one defined by researchers that would indicate the two groups are *meaningfully* different.

If you are working on research where the goal is to declare groups to be equivalent you will want to learn more about equivalence testing. The introduction to equivalence testing on Wikipedia is one place to start, <https://en.wikipedia.org/wiki/Equivalence_test>. The University of Iowa ST 6220 class notes PDF on equivalence testing shows some good examples of what equivalence testing with confidence intervals looks like starting on page 6, <http://homepage.divms.uiowa.edu/~rdecook/stat6220/Class_notes/equivalence_testing.pdf>.