11. Brief introduction to multi-way ANOVA, multiple regression and general linear models

Multiple regression and interaction

In regression, multiple predictors may be used in the model:

$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n + \varepsilon$

predictors may be both quantitative and categorical variables. This is based on the fact, that categorical variables may be decomposed into k-1 binary (0-1) variables (where k is number of categories/levels). In general, the maximum number of predictors is limited by degrees of freedom in the model. Complexity of the model measured by the model number of degrees of freedom may never exceed total df (i.e. number of observations – 1).

Models containing two or more predictors may also contain interaction terms:

$Y = a + b_1X_1 + b_2X_2 + \mathbf{c_1X_1X_2} + \ldots + \varepsilon$

interaction means that the dependence of the response variable on one predictor (X_1) depends on the value of second predictor (X_2). Interaction it typically tested in multi-way ANOVA, where even higher-order interactions can be considered. Interaction may be positive (i.e. the value of response is higher than expected from additive sums of main effects; in such case $c_1 > 0$; Fig. 11.1) or negative (response value is lower that the additive sum; $c_1 < 0$).

The interaction is formally notated by × (Alt + 0215), i.e. $Y \sim X_1 + X_2 + X_1 \times X_2$. In R, interaction may be represented by "*" which indicates both additive and interaction effects or by ":" which indicates just the interaction term.

No that 1. testing the interaction is very common in manipulative experiments and 2. interaction does not mean correlation between predictors. As you will see later, correlation between predictors is a serious problem which among other issues prevents from reasonable assessment of interactive effects.

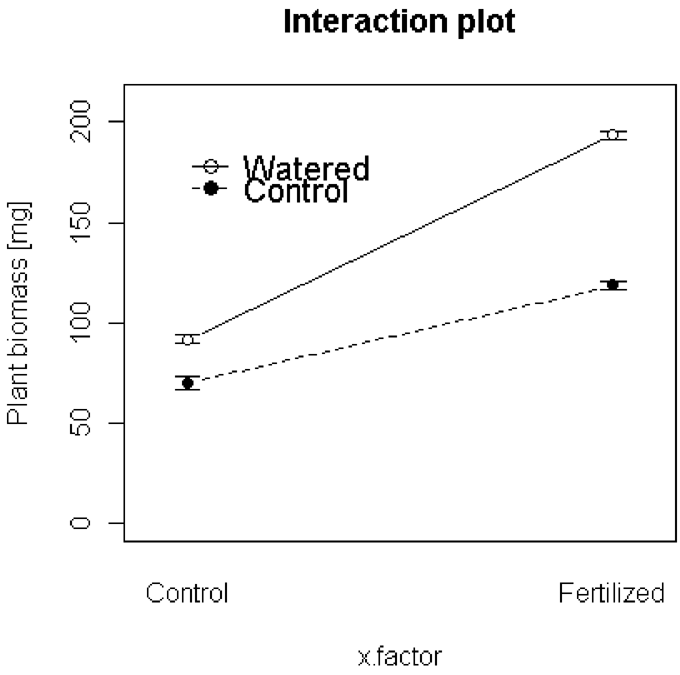

Fig. 11.1. Interaction plot showing positive interactive effects of fertilizer application and watering on plant growth. The interaction is directly visible from the graph as non-parallel lines connecting the mean values.

Testing of linear models and their terms

Statistical significance testing of linear models (as whole predictor structures) using e.g. an F-test is easy and largely follows the same principles as in simple regression. It is more difficult to decide which predictors to include in the model and which not. Finding the best model is done by a model selection procedure which aims at finding the model which contains only the predictors, which have significant effect on the response while those effect of which is non-significant (i.e. do not contribute to the predictive power of the model) are left out. Such models are called *minimum adequate models*. Philosophically, they are based upon the principle of Occam's razor or parsimony (https://en.wikipedia.org/wiki/Occam's_razor).

Statistical methods are very efficient if applied on model testing and/or comparisons between models. However, there are no universal guidelines, which could be used for model/predictor selection in all cases. In models with few candidate predictors, it is possible to fit all possible models and select the one with the highest explanatory power. Frequently (but certainly not always), simple testing of significance of individual predictors (which is based on statistical comparison between models excluding and including given predictor) can also be used.

For efficient model selection, we need 1. a measure of model quality (or quality comparison between models) and 2. a strategy how to build the model.

Measures of model quality

There are several measures of model quality or parameters for model comparison.

F-test: the *F*-test may not only be used to test the significance of a model but also to test whether one model is significantly better that another. Works generally well for models with up to moderate number of observations (~200). With large *n* almost all predictors tend to be significant even if explaining very little variability in response.

R^2: proportion of explained variation is a property of a model itself. It is easy to interpret. For model comparisons, its main disadvantage is, that addition of more predictors *always* increases R^2 even if the predictor added has little effect. As such, it is not suitable to compare models of different complexity.

AIC (*Akaike information criterion*; https://en.wikipedia.org/wiki/Akaike_information_criterion): This measure is derived from information theory and allows straightforward comparisons of model quality. Models with lower AIC value are better. The AIC is computed using the following formula:

$\text{AIC} = 2\text{k} - 2\log(L)$

where, k is number of parameters of the model and log(*L*) is log-likelihood of the model.

Likelihood-ratio. Likelihood ratio is a very general approach, which can be used to compare many types of models. It is based on the principle that the logarithm of likelihood ratio (which numerically equals the difference between log-likelihoods) multiplied by 2 follows the $\chi 2$ distribution; thus the goodness-of-fit test may be used for testing of models differing in numbers of df.

Model building strategies

There are several options how to build a model. Theoretically, the best way would be to fit all possible models and choose the best fitting one based e.g. on AIC. However, number of possible models could be very large (increases with numbers of predictors and complexity of interaction terms) and fitting of models may be demanding for computer power (with increasing availability of big data even with current fast computers). Therefore, it may be useful to use a pragmatic approach to model building. There are two reasonable approaches each of which has its advantages and disadvantages – forward and backward selection.

Forward selection starts with the null (intercept-only) model. Next step includes testing every model containing single predictor against the null model. Such comparisons are indicative of individual predictor explanatory power and on this basis the best fitting predictor (using *R*2 or AIC) can be added to the model. In the next step the model containing the selected predictor is used as the null against with the other predictors are tested and so on until there is no significant candidate predictor left. With two or more predictors in the model, interactions between the predictors may also be tested to be included in the model. An advantage of this approach is its intuitiveness and possibility to use a large number of candidate predictors (though multiple testing issue should be considered here). However, there are also disadvantages related with this approach including often high risk of selecting of non-optimal model due to constraints related to the procedure. Still forward selection is a reasonable choice for observational data, in particular when large number of predictors is available.

Backward selection uses an opposite strategy – first a saturated model is fitted (i.e. model containing all candidate predictors together with all their interactions – these may be limited up to a specified order). Non-significant terms are then removed from the model one-by-one starting with poorest predictors (again measures by AIC). Note that in the case of a significant interaction, main effects are retained in the model even if they are not significant themselves (if the same model was built-up in a forward manner, such interactions would never be tested).

Correlation between predictors

Correlation between predictors is a serious issue in multiple regression analysis. This issue concerns observational data because in experimental studies, we should use an experimental design which ensures independence of tested predictors. The problem is, that if there are two inter-correlated candidate predictors to be included in the model, one of them may be included just by chance (because it may look slightly better with given data). The other predictor will then never be included in the model, because its effect is already accounted for by the first predictor. Depending on the actual data, either one or the other predictor may be included while the other left out. Such inconsistency may lead to very different conclusions even if the relationships between the variables are the same and the data are just slightly different. Such cases are quite common in nature, e.g. soil pH and Ca concentration represent a common case in ecological studies. Unfortunately, none of the model building strategies or model quality measures can control this. However, a detailed exploration of the associations between the predictors themselves and between individual predictors and the response may be useful.

As a part of this exploration, we may first analyze *marginal* or *simple effects* – i.e. effects of given predictor on the response which ignore the effects of other variables. These are simple linear regressions (or one-way ANOVAs) and are indicative of the correlation structure in the study system. Conversely, *partial* or *conditional effects* can be computed (i.e. unique effects of individual predictors), which are computed by testing a given predictor against a model containing all other predictors. Such effects are greatly affected by predictor inter-correlation but if significant, they may really point to mechanisms underlying the correlations.

Computing marginal and partial effects is then a part of a more general approach called *variation partitioning*. With this approach, you can describe the correlation structure among the predictors (or frequently groups of predictors) and quantify their unique or shared effects on the response variable.

How to do in R

1. Fitting a model – function **lm** (see chapter 9 for basics). Individual predictors are included in the formula on the predictor side separated either by + (additive effects) or by * (additive and interactive effects)
2. Testing candidate predictors to be included in the model – function **add1**; e.g. add1(lm.model, scope =~predictor1*predictor2,…). Parameter test is then used for specification of the model quality criterion. AIC is displayed always; for ordinary linear models, it makes sense to ask for an F-test by setting test="F".
3. Testing predictors to be removed from the model – function **drop1**. The use is similar to add1, just the parameter scope is not specified.
4. Changing model structure – function **update**; adding a predictor: new.model<-update(old.model, .~.+added.predictor), removing a predictor new.model<-update(old.model, .~.-removed.predictor). Update can be used to change also other parameters of a model.
5. Comparison of model quality - function anova; e.g. anova(model1, model2) compares
6. Testing individual terms – anova(lm.model) displays sequential F-tests for individual terms. Sequential testing means, that order of the predictors affects the results (unless the predictors are perfectly independent - orthogonal). summary(lm.model) displays detailed model statistics – F-test of the whole model and t-tests of individual regression coefficients. These t-tests are not sequential and thus are independent term order in the model.
7. Model coefficients may be called by function coef – i.e. coef(lm.model)
8. Model residuals may be called by function resid – i.e. resid(lm.model)