GAMLj Models

Marcello Gallucci

2023

Contents

Chapter 1

Introduction

1.1 Preface

Draft version, mistakes may be around

This is a book about linear models in jamovi, using GAMLj module. This book is halfway between a software manual and a statistical how-to. Thus, what regards the statistical reasoning and the models interpretation may be applied to analyses carried out also with other software. The practical steps to obtain the results are specific to GAMLj.

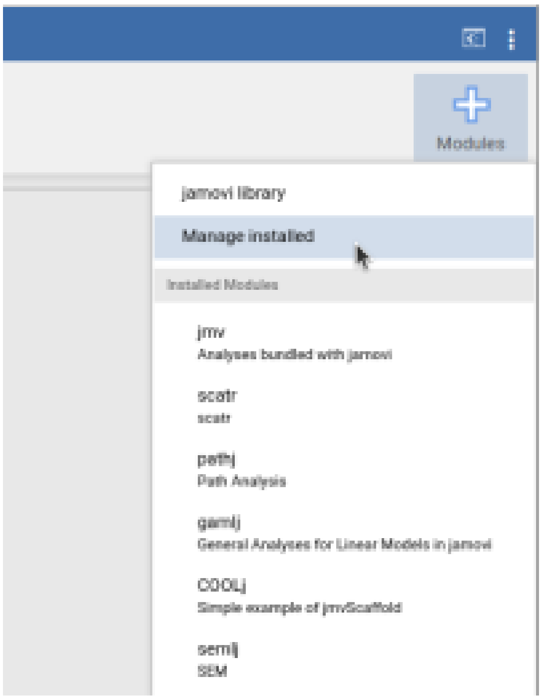
This book, so far, covers the following models:

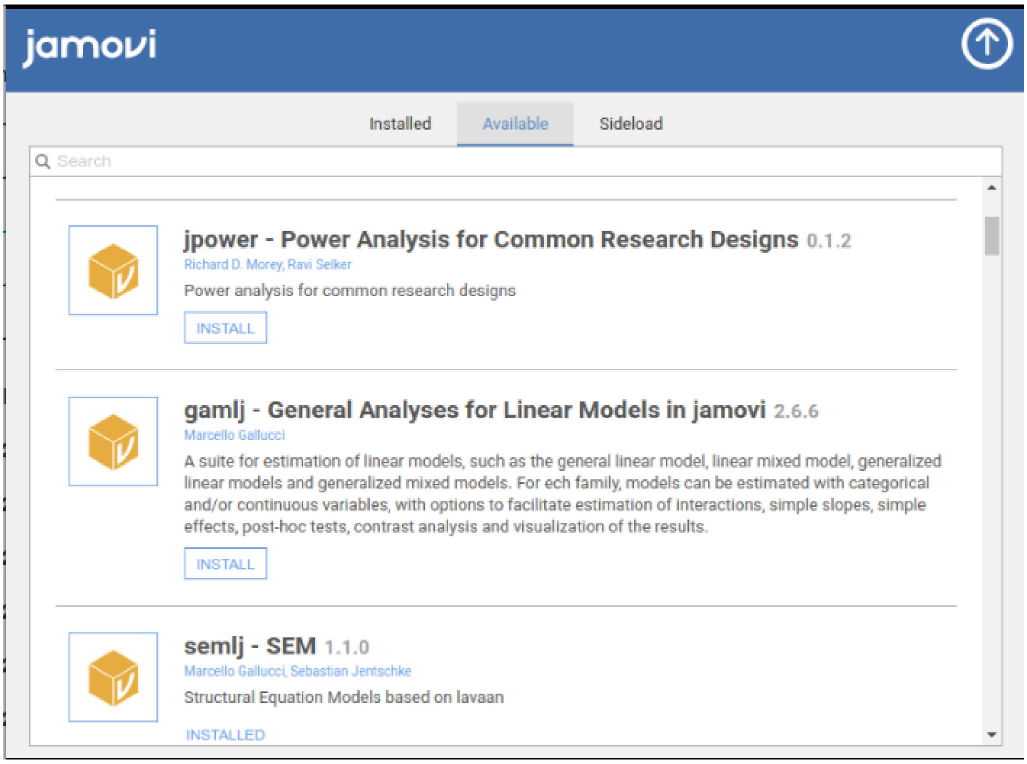
- The general linear model (2)
  - Regression (2.2, 2.3)
  - ANOVA (2.7)
  - ANCOVA (2.11)
  - Moderation (2.4)
- The Generalized linear model (3)
  - Logistic model (3.2)
  - Probit model (3.4)
  - Multinomial model (3.5)

1.2 Getting Started

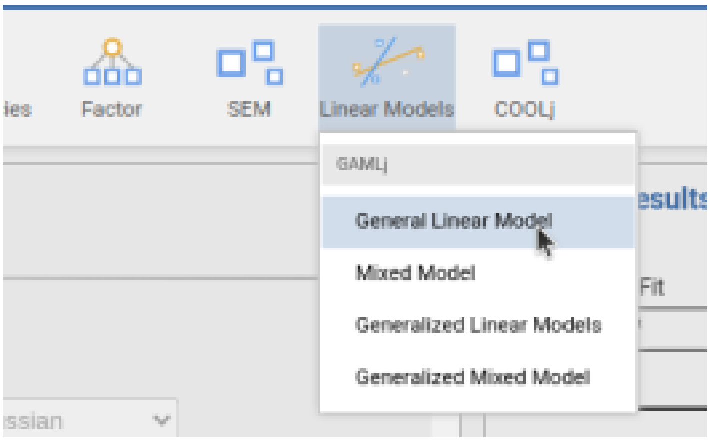
First, we need to install jamovi, and within it, install GAMLj.

- To install jamovi, download it from the jamovi website.
- Within jamovi, access the library and install GAMLj





Installing the module produces a new icon in the icons bar, and the new icon gives access to the list of the module available analyses.



Here we always refer to GAMLj version 3.0.0 . If you need to work with previous versions, you can refer to GAMLj legacy help We are ready to go.

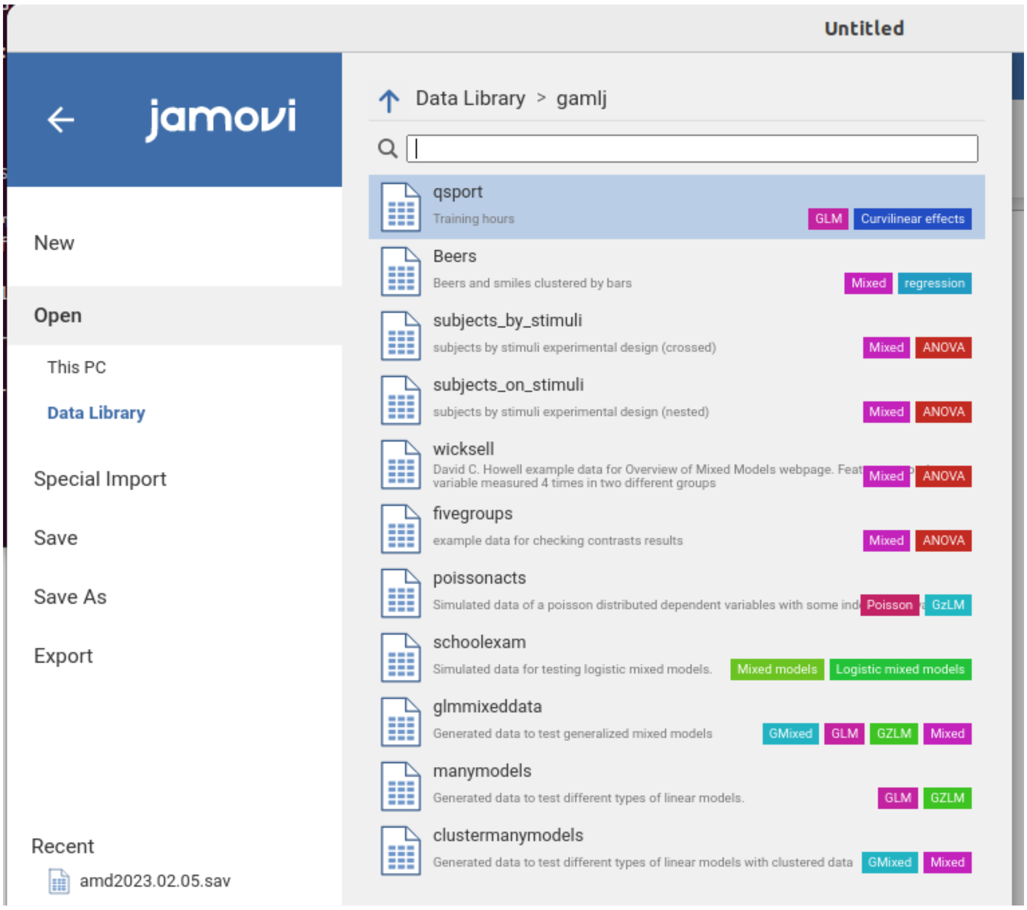
1.3 Data

Throughout this book, we are going to use mostly two simulated datasets, containing variables that allow estimating different types of effects for different types of models. Both datasets contain two continuous independent variables, named `x` and `z`, two categorical independent variables, named `cat2` and `cat3`, with two and three groups respectively. The dependent variable `y` appears in different forms, or types, so we can apply to it different models. We have a continuous normally distributed dependent variable `ycont`, a dichotomous version of it `ybin`, a count version `ypoi` simulating a Poisson distribution of counts, a ordinal version `yord` with 5 ordinal levels, and `ycat`, a three-level categorical variable.

The first dataset containing these variables is called `manymodels` and simulates a sample of 120 cases drawn randomly for a population. We are going to use this dataset for the general linear model (Chapter 2) and the generalized linear models. The second dataset, named `clustermanymodels` has the same variables, but the data a simulated as drawn from 30 different clusters. We are going to use this dataset for the mixed model and the generalized mixed models.

I agree with the idea that x, y, and z datasets are boring, and real data examples are more engaging. Nonetheless, x, y, and z datasets allow exploring many different situations that real data often do not (at least not with one or two datasets), and do not take away the reader's attention from the stats. We will invent some cover story to make the examples more engaging and easier to follow.

Both datasets can be opened from jamovi data library.



1.4 What's in a name

All the fundamental analyses presented in this book are referred to as *models*. With the term *model* I mean a concise and efficient way to represent and quantify the relationships among variables. A simple regression, an ANOVA, or a multi-level random coefficients logistic regression are all models of the data. To put it in a more elegant way, a model provides an approximate and idealized representation of the process generating the data (Neyman, 1957). In several statistical and methodological sources these *models* are called *statistical techniques*, but I do not find this term helpful, because almost everything one does can be a technique. To avoid confusion, we call *models* the linear representation of the dependent variable(s) as a function of the terms of the independent variables. Thus, we have different models when the linear representation is different, or the estimation method is different. A logistic model, for instance, is different from a Poisson model because the first predicts the logit of a dichotomous dependent variable, the second the log of it (more on this later).

In this book we cover four model categories:

- The general linear model (Chapter 2)
- The generalized linear model
- The mixed linear model
- The generalized mixed model

The main differences among them, at least the ones we are interested in, can be classified depending on the type of dependent variable they model and the way the sample has been drawn:

Clustered vs independent cases differ in the way the data are collected. We talk about it in Chapter (XX). In general, we will see for each model how and why one of these macro-categories applies.

1.4.1 Statistical techniques vs points of view

With a model, we can do many things. First, we evaluate the results. A linear model can always be evaluated from two different angles: The model fit (variances explained or deviance reduced) and the size and direction of the effects (the coefficients). Thus, whenever we estimate a model, we can look at it from each of these angles or both. The model fit, often broken down by variables and effects unique contribution, informs us about the ability of the independent variables to explain the variability of the dependent variable. This angle offers also effect size indices, like η^2 or ω^2. This angle is useful, among other things, to evaluate the strength of the effects. In the linear model, this angle is called *Analysis of Variance*, shorten in *ANOVA*, in the generalized linear model it is called *Analysis of Deviance* (not shorten).

The second angle looks at the coefficients, that represent expected changes in the dependent variable as one compares different levels of the independent variable(s). They answer questions like "What is the average increase in salary for every year worked?", or "what is the group with the largest salary among some employees groups?". The coefficients angle also provides effect size indices, such as the B and the β coefficients, the correlation, and several variations of standardized mean differences (Cohen's d). This angle is useful, among other things, to evaluate both the intensity and the direction of the effects.

In GAMLj, both points of view are available for all linear models handled by the module, no matter how complex they are. But this creates often confusion in users acquainted with old-fashion terminology. Why is there an ANOVA table in the results section of a regression? Why do I get regression coefficients for categorical independent variables? The reason is that in the linear model's realm, terms such as regression or ANOVA are not *analyses* or *statistical techniques*, they are angles from which one looks at the results. If one is interested in the direction of the effects, one looks at the coefficients, if one is interested in the variability accounted for by the effects, one looks at the variances (or deviance). Thus, what people usually call *ANOVA*, meaning the analysis of a design with one continuous dependent variable and one or more categorical independent variables (factors), is just a general linear model evaluated only from the point of view of the explained variances. What people call *regression* is a general linear model with continuous independent variables for which the analyst focuses on the coefficients. ANCOVA? A general linear model with at least one focal categorical variable and at least one continuous variable, for which the analyst focuses on the effect of the categorical variable knowing that the continuous variable(s) is (are) kept constant. The same applies to any other model, general, generalized, mixed, or generalized mixed.

In this book we keep using terms such as *regression* or *ANOVA*, keeping in mind that we can do well without them.

1.4.2 Statistical techniques vs Analyses

Alright, so what do we mean by *statistical techniques* in this book? Things we do with the model. If we find a main effect of a categorical independent variable, we usually want to probe it and check which group is different from any other group. We employ a *posthoc test* technique. Along the way, we also want to see the *estimated marginal means* for each level of the independent variable. If an interaction appears solid in our results, we often probe it to estimate and test the effect of one variable at different levels of another, so we do a *simple effects* technique. In other cases, we want to compare one big model with a smaller one, in which some terms are absent. We do a *model-comparison* technique, so we can evaluate the overall contribution of the effects that are in the big model and not in the smaller (nested) one.

GAMLj provides these techniques (and many others) for all models estimable within the module, with the same user interface and results tables. The basic principle that GAMLj tries to follow is that *if we can do something in one model, we can do it with any other model.*

1.4.3 Terms that we need to generalize

In the methodological literature, statistical techniques emerge often in one field for one application, and then emerge again in other fields or for other applications and get different names. This may create confusion. To clarify, we are going to use some simplification. The most important ones are the following:

- Any *ANOVA like* results, not matter how they are tested (F-test, LRT, χ^2), are referred to as *Omnibus tests*.
- *Moderation* is an interaction in which the analyst focuses on one effect and desires to evaluated it at different levels of the other variable involved in the interaction, no matter whether the variable is continuous or categorical. A *moderator* is a variable that the analyst believes can change the effect of an independent variable, no matter the types of variable involved in the analysis.
- Estimating the effect of one variable at different levels of a moderator is called *Simple Effects*, no matter the types of variable involved. So, slicing of an interaction in the (classical) ANOVA or a simple slopes analysis are all refereed to as *Simple Effects* technique (they are indeed the same technique).
- *Estimated marginal means* are the average predicted values of a model for some level of the independent variables. No matter which model we have at hand, and what kind of variables we have, we always mean this.

1.4.4 Terms we need to cope with

In linear models, independent variables are of two kinds: *categorical* or *continuous*. The former defines nominal levels (groups or conditions) the latter defines quantities. Despite the simplicity of this definition, commonly used statistical software calls a categorical independent variable as *factor* and a continuous independent variable as *coviariate*. jamovi and GAMLj follows this tradition. We should be aware, however, that these terms do not mean anything else that categorical and continuous variables. Every independent variable in a linear model is covariated (partialed out) when the other variable effect is computed, no matter whether the variable is a categorical or continuous one. How it is covariated, however, depends on the model, so we need to pay attention to that.

1.5 General References

Obviously, I did not invented any of the statistical ideas or methods that this book mentions. Much of this material comes from statistical common knowledge and logical necessity. However, when not otherwise specified, the fundamental concepts treated here can be found in the seminal work of Cohen et al. (2014), Searle and Gruber (2016), Raudenbush and Bryk (2002), Agresti (2012), Aiken et al. (1991). When needed, specific references are provided for more novel or critical issues.

Chapter 2

The General Linear Model

Draft version, mistakes may be around

KEYWORDS: General Linear Model, Regression, ANOVA, Interaction, Moderation

2.1 Introduction

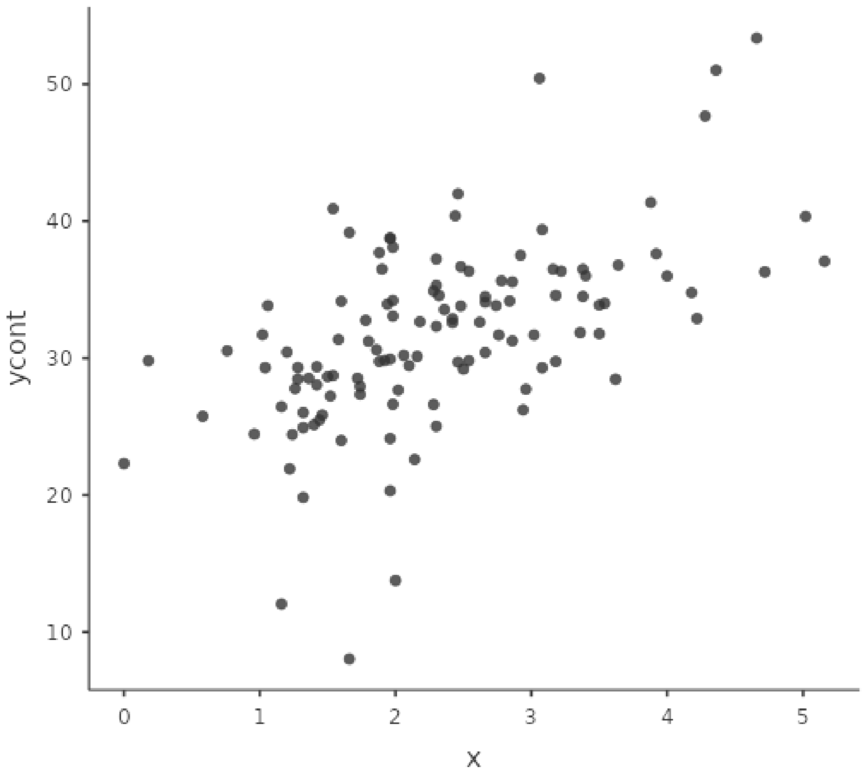
The general linear model (GLM) encompasses the majority of the analyses that are commonly part of any practitioner's statistical toolbox. There are good reasons for that: with a good knowledge of the GLM one can go a long way. Within the GLM one finds analyses such as simple and multiple regression, Pearson correlation, the independent-samples t-test, the ANOVA, the ANCOVA, and many of their derivations, such as mediation analysis, planned comparisons, etc (cf. 1.4). The common theme of all these applications is that the dependent variable is continuous, hopefully normally distributed, and that the sample is composed of non-related, independent cases. The most basic yet very important application of the GLM is the simple regression, a GLM with one continuous independent variable (IV).

2.2 One continuous IV

AKA: Simple Regression

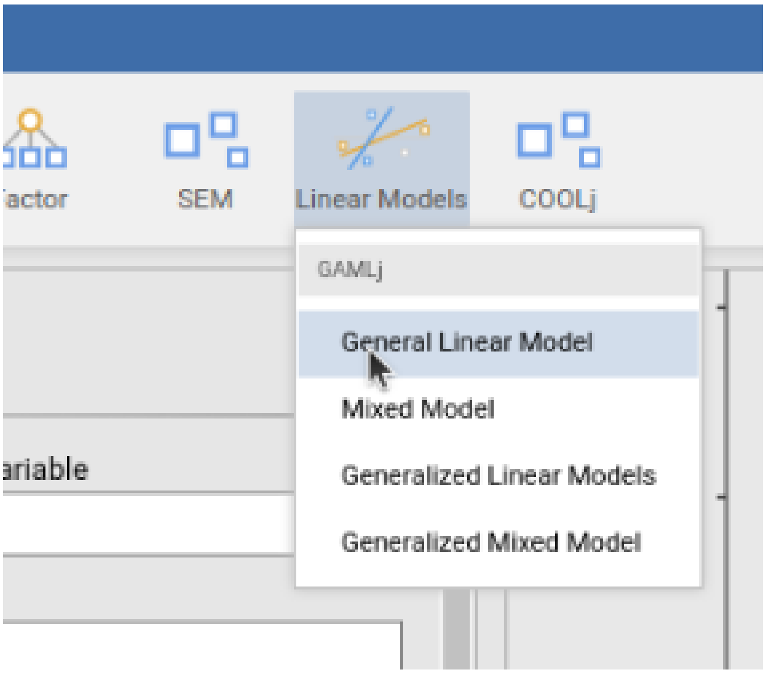
Consider the dataset `manymodels` (cf. 1.3). The dependent variable is a continuous variable named `ycont`, and we want to estimate its linear relation with a continuous variable named `x`. The extensive relation between the two variables can be appreciated in a scatterplot. It is clear that `ycont` and `x` can be any variable, as long as we can consider them as continuous. For the sake of the argument, let us imagine that we went to a bar and measured `ycont` as the average number of smiles smiled by each customer in a given time and x as the number of beers drunk for the same period.

Scatterplot

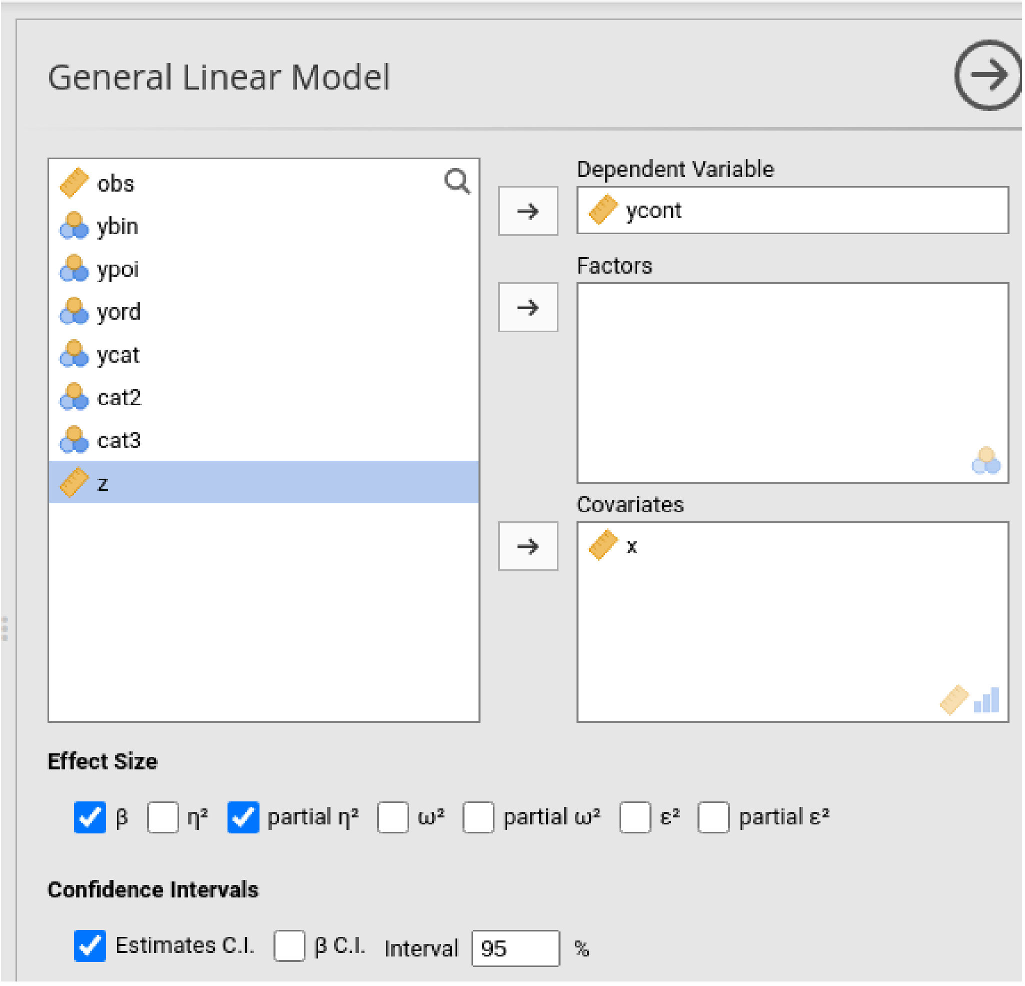


What we want to know is the average increase (or decrease) of the dependent variable as the independent variable increases. Thus, how many smiles on average one should expect for one more beer? We ran a GLM to get the answer.

2.2.1 Input



We set the `ycont` variable as the dependent variable and the `x` variable as the independent continuous variable (see 1.4.4), and look at the results.



2.2.2 Model Recap

First, we check out the Model Info Table.

Model Info

Info		
Model Type	Linear Model	OLS Model for continuous y
Model	lm	ycont ~ 1 + x
Distribution	Gaussian	Normal distribution of residuals
Omnibus Tests	F	
Sample size	120	
Converged	yes	
Y transform	none	
C.I. method	Wald	

[3]

>

This is a recap table that says that we did what we wanted to do, and how we did it. The second table we get is the Model Fit Table, where the R^2, the adjusted R^2, and their inferential test are presented.

2.2.3 Model Fit

Model Results

Model Fit

R^2	Adj. R^2	df	df (res)	F	p
0.323	0.317	1	118	56.2	< .001

[4]

The R^2 gives us the first glance of the model from the *variance angle* (cf. 1.4.1). The short story says that our model (in this case the independent variable x) explains, or accounts for, .323*100 percent of the variance. So, if all differences in the smiles (`ycont`) are set to 100, 32% of them can be associated with the number of beers drunk (x). The `Adj.`R^2 is the estimation of the variance explained by the model in the population, and the `df` is the number of parameters estimated by the model apart from the intercept: Here is one because we have one independent variable that requires only one coefficient. The `F` column gives the F-test testing the null hypothesis that R^2 is zero, and `p` is the probability of obtaining the observed R^2 under the null hypothesis. If you find this story a bit dull, you might want to read the full story in Appendix A.

2.2.4 Omnibus Tests

ANOVA Omnibus tests

	SS	df	F	p	$\eta^2 p$
Model	1722.837	1	56.206	< .001	0.323
x	1722.837	1	56.206	< .001	0.323
Residuals	3616.958	118			
Total	5339.794	119			

With only one continuous variable this table is not very useful, but we comment on it anyway to get us familiar with the ideas of the two points of view always available in a linear model (cf 1.4.1). The first line is there only for legacy compatibility purposes. It reports the inferential test of the model as a whole, which we already saw in 2.2.3. The second line tells us the amount of variance of the dependent variable explained by the independent variable. In this case, the $p\eta^2$ is equal to the R^2, because there is nothing to partial out (there is only one independent variable). Instructive, however, is to select the option ϵ^2.

Model Results

Model Fit

R^2	Adj. R^2	df	df (res)	F	p
0.323	0.317	1	118	56.2	< .001

[4]

ANOVA Omnibus tests

	SS	df	F	p	η^2p	ω^2	ε^2
Model	1722.837	1	56.206	< .001	0.323	0.315	0.317
x	1722.837	1	56.206	< .001	0.323	0.315	0.317
Residuals	3616.958	118					
Total	5339.794	119					

we can notice that the ϵ^2 is equal to the adjusted R^2. Yes, that is going to stay: R^2 and η^2 indices (partial or not) are the sample estimates of variance explained, whereas R^2_{adj} and ϵ^2 effect size indices (partial or not) are the population version (ω^2 is population too). People tend to use the sample version of these indices (η^2 and R^2) when they should use the population version (R^2_{adj} and ϵ^2). The same goes for ω^2 index, but you want to read this about why you want to use them, and this how they are computed . In a nutshell, R^2 and η^2 tell what happened in the sample, R^2_{adj} and ϵ^2 tell what should happen in the population.

2.2.5 Coefficients

Parameter Estimates (Coefficients)

			95% Confidence Intervals					
Names	Estimate	SE	Lower	Upper	β	df	t	p
(Intercept)	31.738	0.505	30.737	32.739	0.000	118	62.798	< .001
x	3.805	0.508	2.800	4.810	0.568	118	7.497	< .001

[5]

The regression coefficients table, here called the Parameters Estimates (Coefficients) table, informs us about the size and the direction of the effect. The interesting coefficient is the one associated with the independent variable (the `Estimate` column). Here it is 3.808. This means that for every unit increase in the independent variable the dependent variable increases, on average, of 3.808 units. In our toy example, for each beer one drinks, on average, one smiles 3.808 smiles more. This is the **regression coefficient**, the very first and most solid pillar of the linear model. This interpretation is going to stick, so keep it in mind, because when models get more complex, we are going to amend it, but only to make it more precise, never to betray it.

The intercept, which is not focal here (*nobody looks at the intercept*), is worth mentioning for the sake of comparison with other software. If you run the same analysis in SPSS, R, Jasp, Stata, etc, you get the same `estimate` for the `x` variable, but a different `(Intercept)`. Recall that in any linear model the intercept is the expected value of the dependent variable for $x = 0$. In GAMLj, however, the independent variables are centered to their means by default, so $x = 0$ means $x = \bar{x}$. So, in GAMLj the intercept is the expected value of the dependent variable for the average value of x.

Why centering? First, centering does not change the results of the regression coefficients of simple and multiple regression, so it is harmless in many situations. However, when an interaction is in the model, centering guarantees that the linear effects are the *main effects* one expects, and not some weird effects computed for the moderator equal to (possibly non-existing) zero. Furthermore, I believe that very few variables have a real and meaningful zero, so their mean is a more sensible value than zero [1]. If your variables really have a meaningful zero (which you care about), you can always "unscale" your independent variables setting them to `Original` in the Covariates scaling panel.

Model Results

Model Fit

R^2	Adj. R^2	df	df (res)	F	p
0.323	0.317	1	118	56.2	< .001

[4]

2.2.6 Pearson Correlation

Across sciences, the most used index of association between two variables is the Pearson Correlation, r, otherwise named zero-order correlation, bivariate correlation, standardized covariance index, product-moment correlation, etc (pick any name, the Pearson correlation is a case of the Stigler law of eponymy anyway).

What is important here is that the Pearson correlation is just the standardized regression coefficient of a GLM with only two continuous variables (one DV Dependent Variable , on IV Independent Variable). In GLM terminology, it takes the name of β. In our example, the correlation between `ycont` and `x` is .568. We can verify this by asking jamovi to produce the correlation between the two variables in the Regression->Correlation Matrix menu.

[1] I have nothing against zero: my favorite number is 610, which in Italian literally translates to *you are a zero*, where "you" is meant to be "we all"

Correlation Matrix

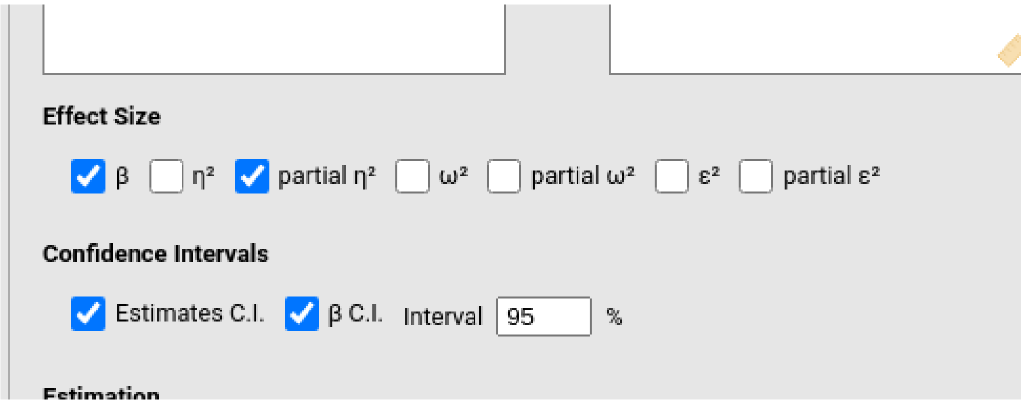
Correlation Matrix

		ycont	x
ycont	Pearson's r	—	
	df	—	
	p-value	—	
x	Pearson's r	0.568	—
	df	118	—
	p-value	< .001	—

As expected, the correlation and the β are the same. More specifically, the Pearson correlation is the regression coefficient that one obtains if the GLM is run after standardizing (computing the z-scores) both dependent and independent variable. This gives us a key to interpret the Pearson correlation in a precise way: Remembering that any standardized variable has 0 mean and standard deviation equal to 1, we can interpret the r (and therefore the β) as the number of standard deviations the dependent variable moves as we move the independent variable of one standard deviation. It varies from -1 to 1, with 0 meaning no relation.

When we deal with GLM with more than one independent variable, the link between the β and the Pearson correlation is lost, but β's remain the coefficients obtained after standardizing the variables, so they remain the standardized coefficients.

If the user decides to report the β coefficients, they would likely want to report the β confidence intervals. They can be asked for in the input by flagging the "β C.I. option".



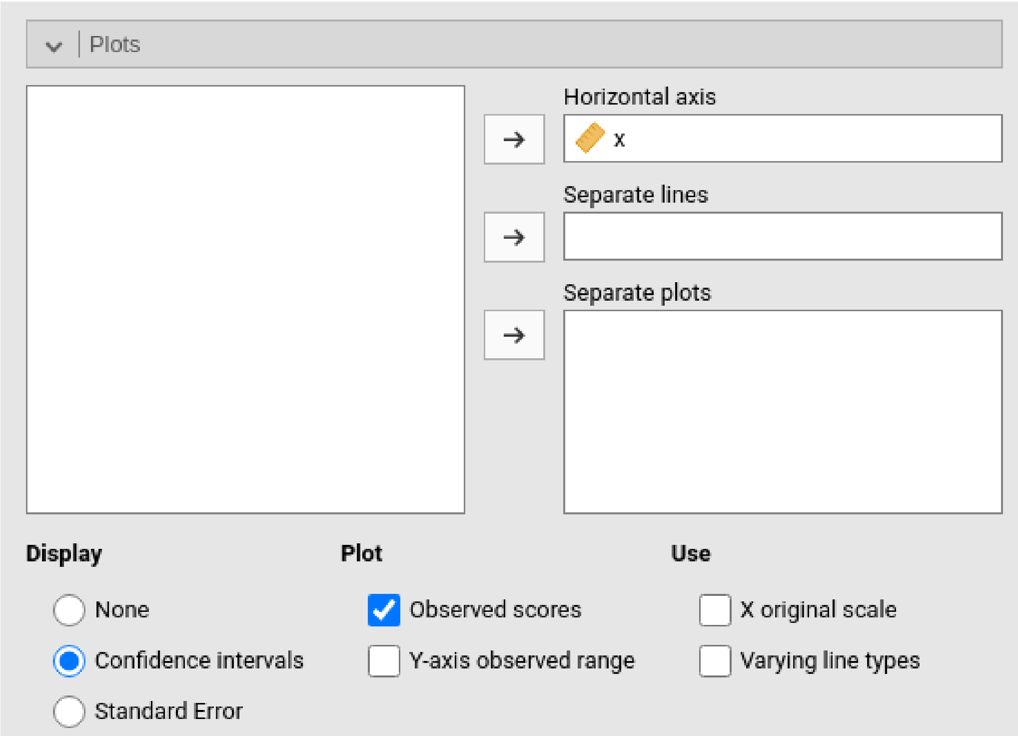
Parameter Estimates (Coefficients)

			95% Confidence Intervals			β 95% Confidence Intervals				
Names	Estimate	SE	Lower	Upper	β	Lower	Upper	df	t	p
(Intercept)	31.738	0.505	30.737	32.739	0.000	−0.149	0.149	118	62.798	< .001
x	3.805	0.508	2.800	4.810	0.568	0.418	0.718	118	7.497	< .001

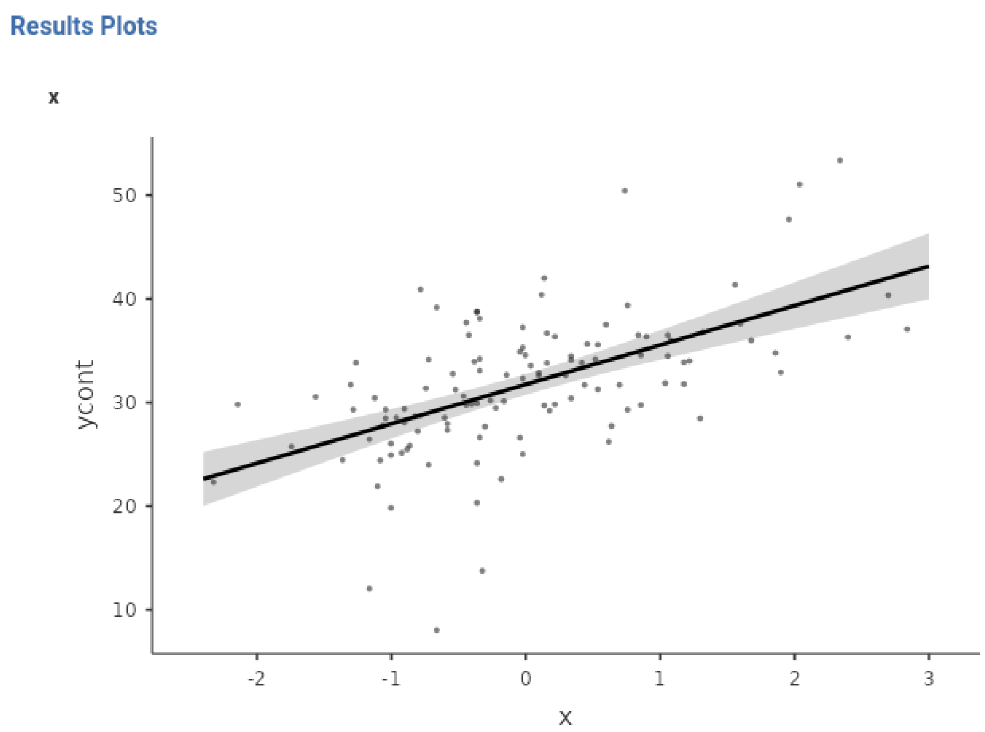
[5]

2.2.7 Plots

It is always a good idea to look at your model results with a plot. A plot shows your fitted model (predicted values). Because a model is always an approximation of the real data, we want to show our predicted values against, or together, the actual data. In Plots panel, we can set up a plot as follows:



and see the results:



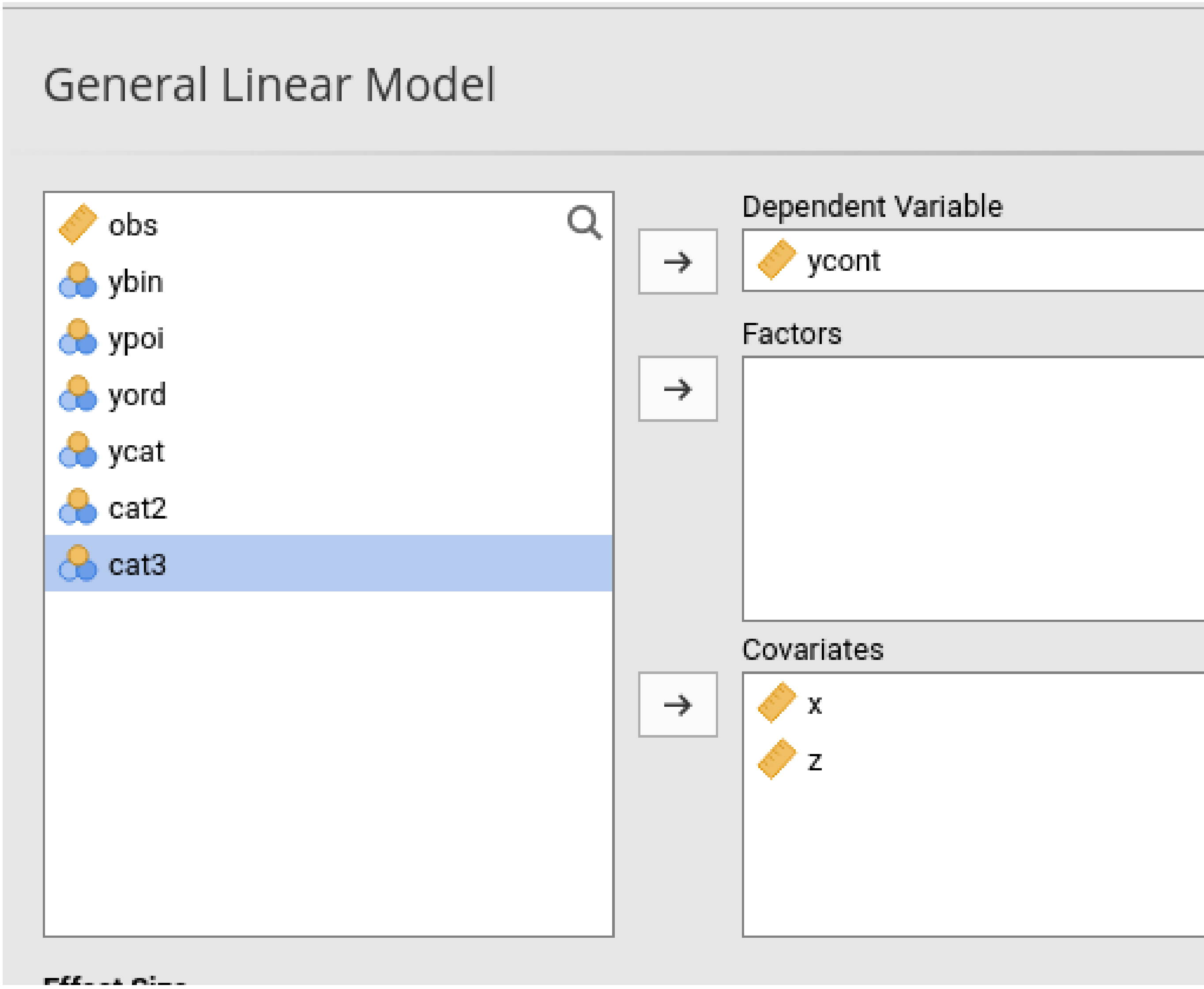
By default, the plot shows also the confidence bands around the regression line. The bands are the continuous version of the confidence intervals, and indicate the range of values where the predicted value are expected to lay.

Notice that the independent variable scale is centered to 0. This is because GAMLj centers continuous variable by default (cf 2.2.5). If a plot with the original scale is preferred, one can flag the option X original scale in the Plots panel.

2.3 More continuous IVs

AKA: Multiple Regression

Let us add to our model the `z` variable, again a continuous variable. To keep up with our toy example, let's assume that `z` was a measure of participants' extroversion.



The results are now updated. Let's go through the most important tables.

2.3.1 Model Fit

Model Results

Model Fit

R^2	Adj. R^2	df	df (res)	F	p
0.346	0.335	2	117	31.0	< .001

[4]

The R^2 gives us the variance explained, or accounted for, of the dependent variable by the whole model. This means by both `x` and `z`, alone and together. The overall variance explained is statistically different from zero, so we can say we do explain some variance of *smiles* (`ycont`) thanks to the variability of *beers* (`x`) and *extroversion* (`z`). The question is now how each independent variable contributes to this variance explained. We need to look at the individual contributions, so the Omnibus Tests.

2.3.2 Omnibus Tests

ANOVA Omnibus tests

	SS	df	F	p	η^2	η^2p	ω^2
Model	1849.387	2	30.996	< .001	0.3463	0.346	0.3
x	1696.744	1	56.876	< .001	0.3178	0.327	0.3
z	126.550	1	4.242	0.042	0.0237	0.035	0.0
Residuals	3490.408	117					
Total	5339.794	119					

Please notice that I selected η^2, `partial` η^2, ϵ^2, and `partial` ϵ^2, so we can interpret these indices. Before that, however, we can mention that the `x` variable effect is statistically different from zero, F(1,117)=58.87, p.<.001, whereas the effect of `z` reaches the predefined level of significance by a very tiny margin, F(1,117)=4.242, p.=.042. So we can say that there is enough evidence to believe that both effects are different from zero, although the former seems more solid than the latter. Statistical significance, however, is only a part of the story: effects should be evaluated on at least three dimensions: **significance**, **size** and **direction**. We now want to evaluate their size.

Effect size indexes are good tools to evaluate effect sizes (*nomen omen*). We start with the `partial` η^2, mainly because it is the most used and reported one effect size index in the literature (I always thought that is the case because it is the only ES produced by SPSS GLM). The `partial` η^2 is the proportion of variance uniquely accounted for by the independent variable, expressed as the proportion of the variance of the dependent variable not explained by the other independent variables. In short, for `x` (*beers*) is the proportion of variance of *smiles* not explained by *extroversion* that is explained by *beers*. In other words, it answers the question: if everybody had the same level of *extroversion*, how much variance would *beers* explain?

Notice that the unique variance explained by *beers* (`x`), namely 31.2%, is computed after removing the variance explained by *extroversion* (`z`). Often you want to know how much variance a variable explains of the total variance, so of all the variance of the dependent variable. That is the η^2, the variance uniquely explained by a variable as a proportion of the total variance of the dependent variable. In other words, it answers the question: how much variance of *smiles* would *beers* uniquely explain?

The ϵ^2 and `partial` ϵ^2 indexes can be interpreted as the η's, but they are adjusted to represent the population variances, not the sample variances. So, they are better estimations of the "real" effect size indexes.

A detailed description of the computation of these indexes can be found in GAMLj help page

2.3.3 Household Chores

Effect size indexes (partial vs not partial ones) are usually explained with Venn Diagrams. We try here without them. Assume you leave in a house with another person. There are 100 chores to do in your household, such us washing the dishes, cleaning the windows, and taking the dog out for a walk. You do 20 chores, 5 of which you did together with your companion. Your companion did 40 chores (they are always better), including the ones you did together. So, altogether you people did 55 chores, your companion did 35 alone, you did 15 alone, and 5 chores were done together.

As a couple, your overall contribution is 55/100, so your $R^2 = .55$. You alone did 15 chores, so your unique contribution is $\eta^2 = 15/100 = .15$ of the total amount of chores to be done. However, of the chores left to do by your companion (60), you did alone 15, so your *partial* contribution is $p\eta^2 = 15/60 = .25$. So the difference between η^2 (or any non-partial ES) and its partial version is the denominator. Non-partial indexes are proportions of the total, partial ones are proportions of the total minus the part accounted for by the others.

In any household, we would use the η^2, but many authors are still using the $p\eta^2$. What is important is to know the difference between the two computation methods, so we can feel free to use which one we prefer. A deep discussion of this matter can be found in Olejnik and Algina (2003), which I recommend reading.

2.3.4 Coefficients

Now we look at the direction and intensity of the effects, interpreting the Parameters Estimates (Coefficients) Table.

Parameter Estimates (Coefficients)

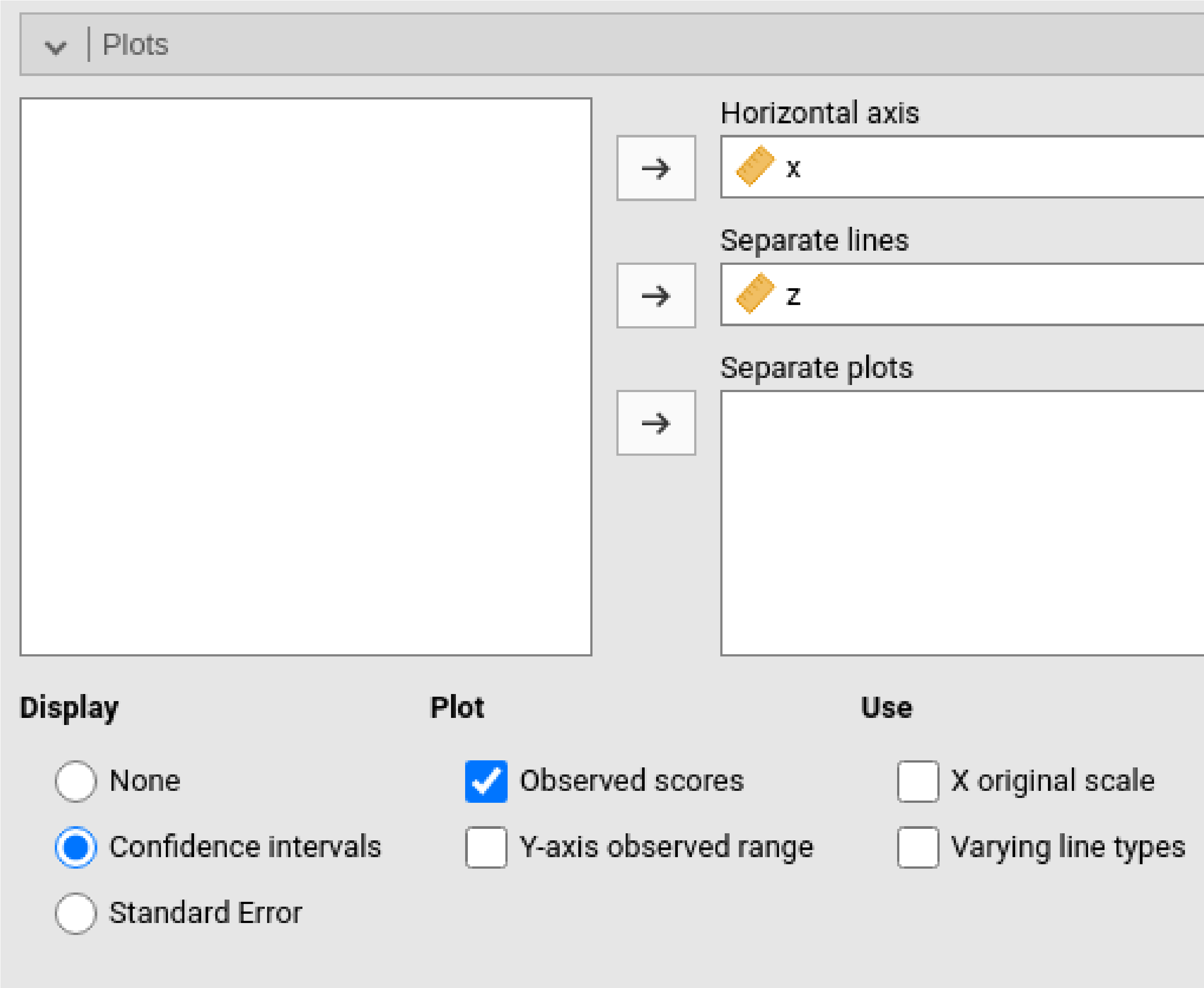
Names	Estimate	SE	95% Confidence Intervals Lower	Upper	β	df
(Intercept)	31.738	0.499	30.751	32.726	0.000	117
x	3.777	0.501	2.785	4.769	0.564	117
z	0.945	0.459	0.036	1.854	0.154	117

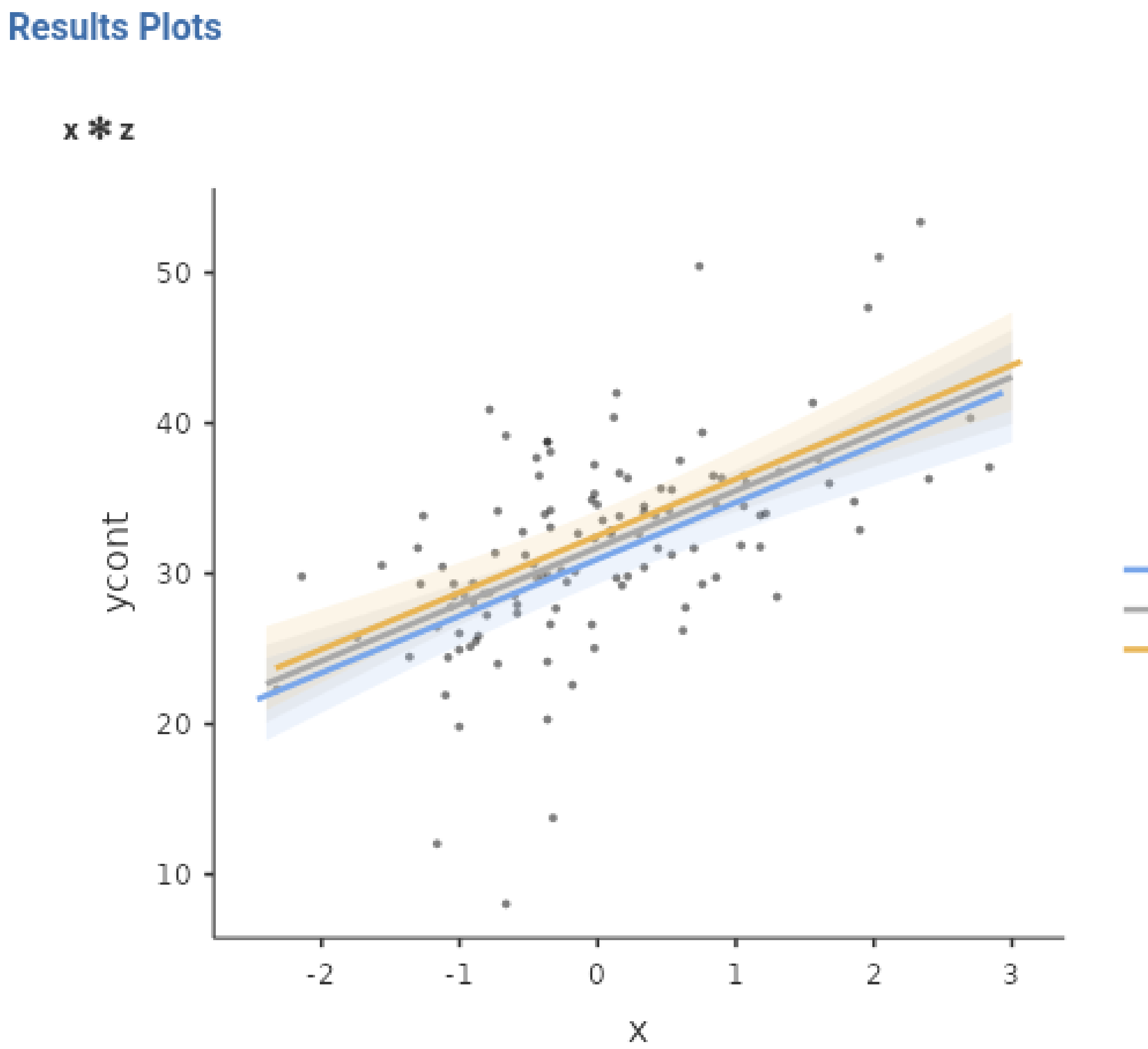
For each variable effect, we see the `Estimate` (usually called the b coefficients), the standard error (`SE`), the confidence intervals, the β, the degrees of freedom (`df`), the t-test (`t`) and the p-values (`p`). So, the full set of estimations (b and β) and the inferential tests (t-test, C.I. df and p-values). Let's interpret them for x:

- $b:$ keeping constant z, for one unit more in x we expect the dependent variable average score to increase of 3.777 units.
- $\beta:$ the effect corresponds to a standardized coefficient of .564, indicating that for one standard deviation more in x, *ycont* increases of .564 standard deviations, keeping constant the effect of `z`.
- $C.I.:$ "Were this procedure to be repeated on numerous samples, the proportion of calculated 95% confidence intervals that encompassed the true value of the population parameter would tend toward 95%" (cf. Wikipedia). Weird? Yes, that's what confidence intervals are. But what about 2.785 and 4.769? Well, they are the two values that include 95% of the cases of the distribution of sample estimates if, and only if, we got exactly the right population parameter in our sample. But we have no way to assess whether we did, so there is not much to learn from these two numbers. This book is not the right place to discuss this issue, but interested readers may enjoy reading Mayo (1981).

2.4 Continuous IVs and interaction

In the previous model, we can make a plot of the effect of beers (`x`) on smiles (`ycont`), at different levels of extroversion (`z`). This can be achieved in GAMLj by asking for a plot as follows:


Plots
Horizontal axis
x
Separate lines
z
Separate plots
Display
None
Confidence intervals
Standard Error
Plot
Observed scores
Y-axis observed range
Use
X original scale
Varying line types

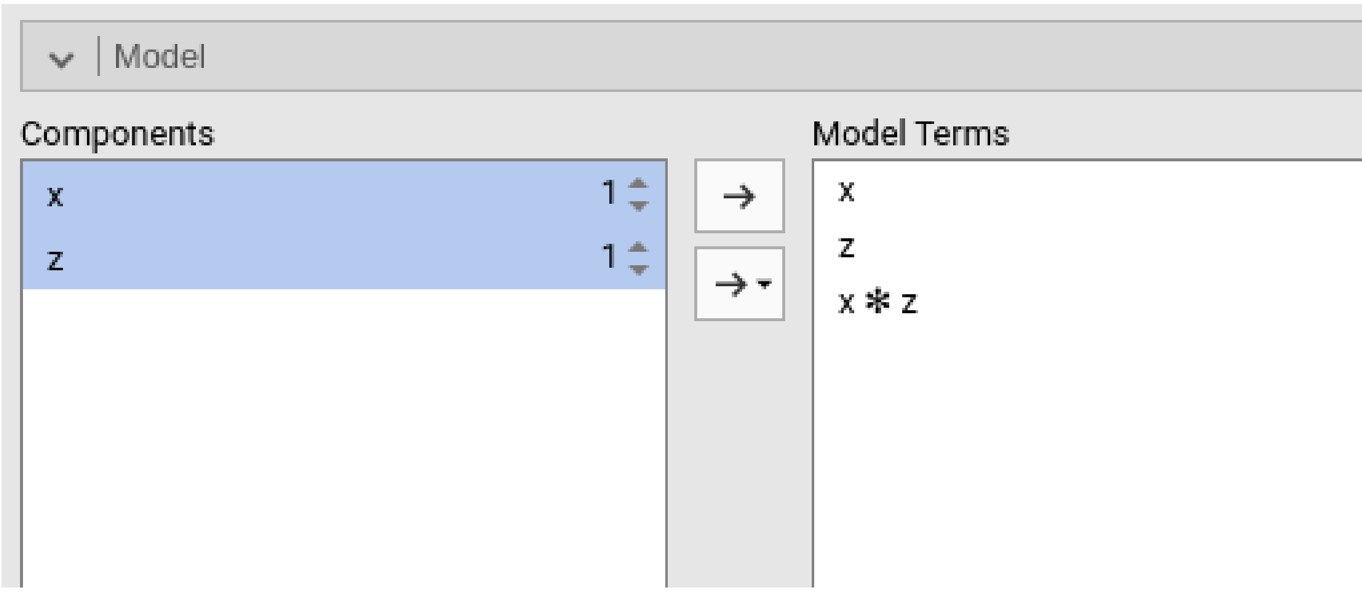


The three lines depicted in the plot are the effects of `x` on `ycont`, estimated at different levels of `z`. Because `z` is continuous, GAMLj automatically sets the focal levels of `z` equal to the mean, one SD above the mean, and one SD below the mean (this can be changed, see below).

As expected, the three effects depicted are perfectly parallel. This is because in multiple regression, the effect of each variable is computed keeping constant the other variable, which is equivalent to saying that the effect is computed as if it were the same at each level of the other variable. In other words, the effects at different levels of the other variable are forced to be the same, no matter the data. But this constrain can be removed by adding an interaction in the model.

An interaction allows the effect one variable to change at different levels of the other. In other words, an interaction allows the effect of `x` to depend on the levels of `z`. The interaction coefficient tells us how much the effect of one variable changes at different levels of the other. The value of the interaction coefficient is the expected change in the effect of one independent variable associated with one unit increase in the other independent variable.

Let's do it. In the Model panel, we select both variables on the left field and move them together to the right field, defining an interaction $x * z$.



The results updated accordingly showing now also the interaction effect.

ANOVA Omnibus tests

	SS	df	F	p	$\eta^2 p$
Model	2005.033	3	23.248	< .001	0.375
x	1605.173	1	55.836	< .001	0.325
z	129.023	1	4.488	0.036	0.037
x ✻ z	155.647	1	5.414	0.022	0.045
Residuals	3334.761	116			
Total	5339.794	119			

Parameter Estimates (Coefficients)

			95% Confidence Intervals			
Names	Estimate	SE	Lower	Upper	β	df
(Intercept)	31.709	0.490	30.739	32.678	−0.004	116
x	3.686	0.493	2.709	4.663	0.550	116
z	0.954	0.450	0.062	1.846	0.156	116
x ✻ z	1.028	0.442	0.153	1.903	0.168	116

Let's focus on the Parameter Estimates (Coefficients) Table. When there is an interaction in a linear model, the effect associated with the independent variables are the effect of the independent variable computed for the other variable equal to zero. This is not a software choice, that is the way a linear model is (Aiken et al., 1991). However, GAMLj, by default, centers the continuous variables to their means, so the linear effects can be interpreted as *main effects*, namely, the effect of the variable computed on average, at the average level of the other variable. Thus, we can say that, *on average*, beers (x) has a positive effect on smiles ($ycont$) of 3.686, corresponding to a standardized effect of .550. For the average level of beers (x), the effect of extroversion (z) is .954, corresponding to a standardized effect of .156.

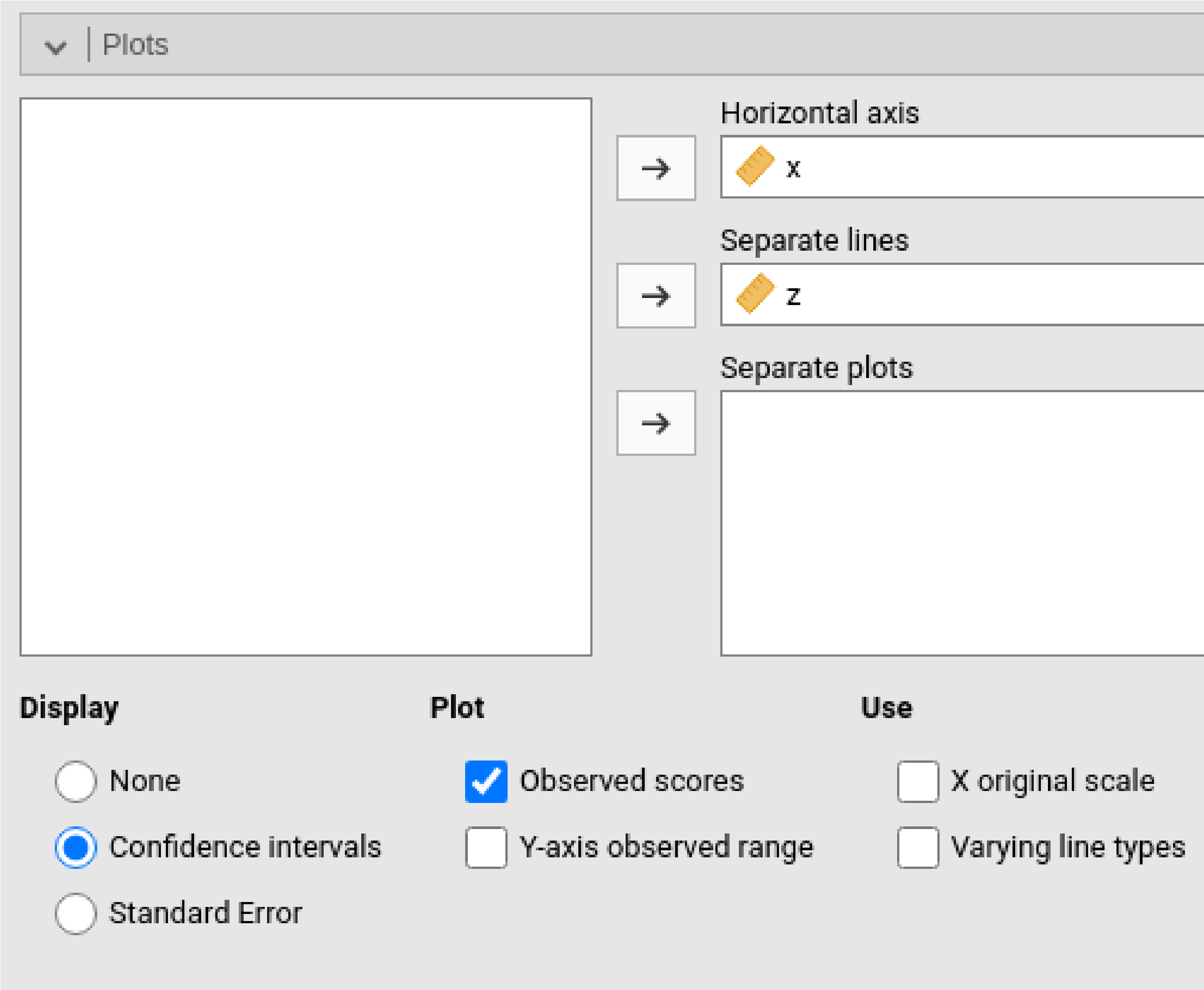
As regards the interaction effect, $x * z$, it tells us that the effect of beers (x) increases of 1.028 units for each unit more of extroversion (z). We see that this effect is statistically significant, so we can say that the effect of beers changes at different levels of extroversion. We should make clear that the same interpretation is valid if we invert z with x. The effect of z is different at different levels of x.

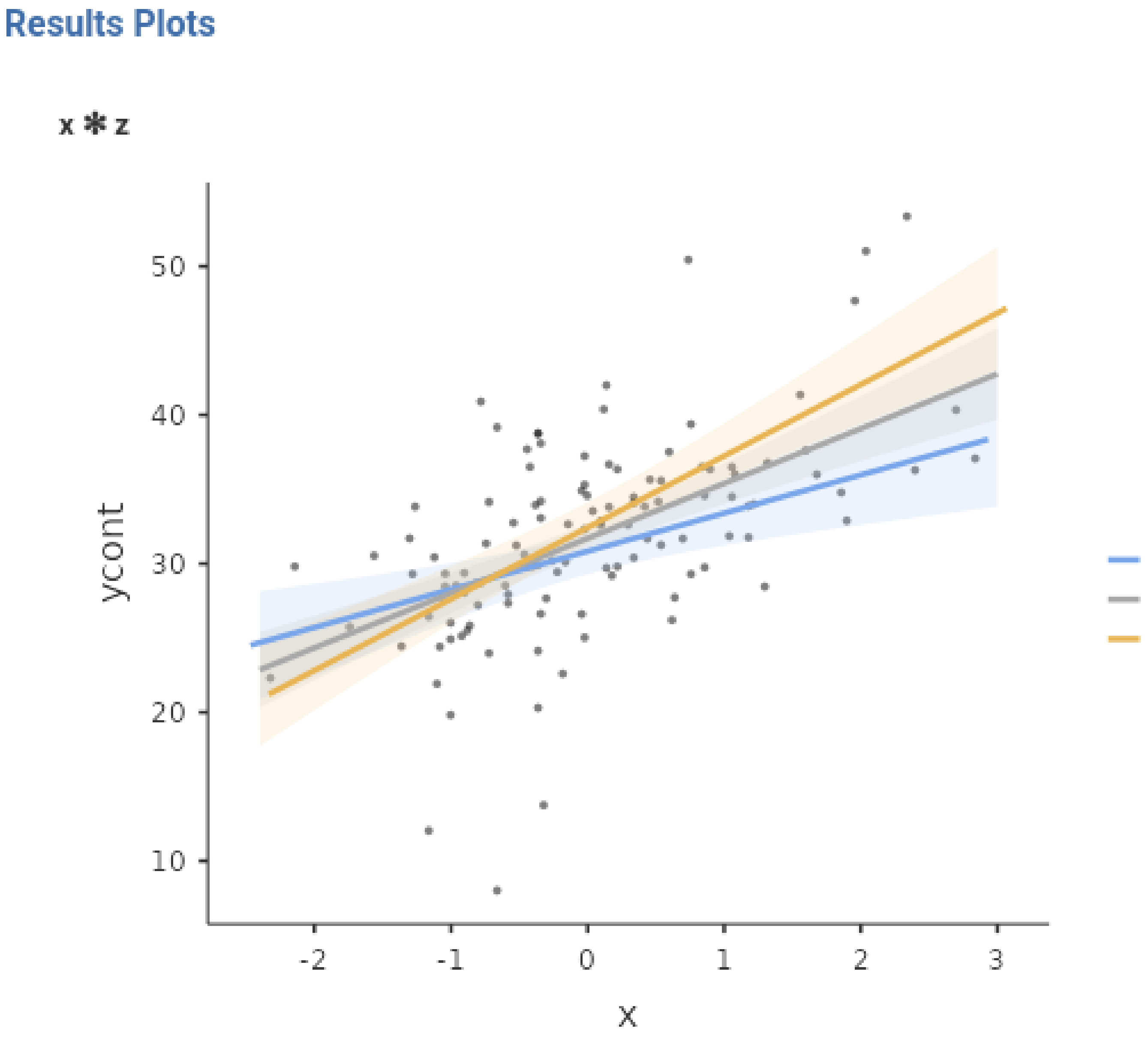
2.5 Moderation=interaction

Many authors calls this type of effect a moderation effect. They are correct, but there is nothing special about interactions between continuous variables. Any interaction effect, no matter if the independent variables are continuous or categorical, tests a moderation model. The only difference between moderation and interaction is theoretical. When we lay out a theoretical model, we declare which variable is the hypothesized moderator and which is the independent variable. Statistically, they are equivalent, and the strength of moderation is tested with an interaction. We will see that moderation models are tested also with categorical variables (within what people calls *ANOVA*) or when one variable is categorical and the other is continuous (within what people used to call *ANCOVA*, but not anymore). We have seen that naming techniques rather than models is cumbersome (1.4). The same goes for *moderation.* Just keep in mind that moderation refers to a theoretical model, its statistical counterpart is the *interaction*, for all kinds of variables.

2.6 Simple Slopes

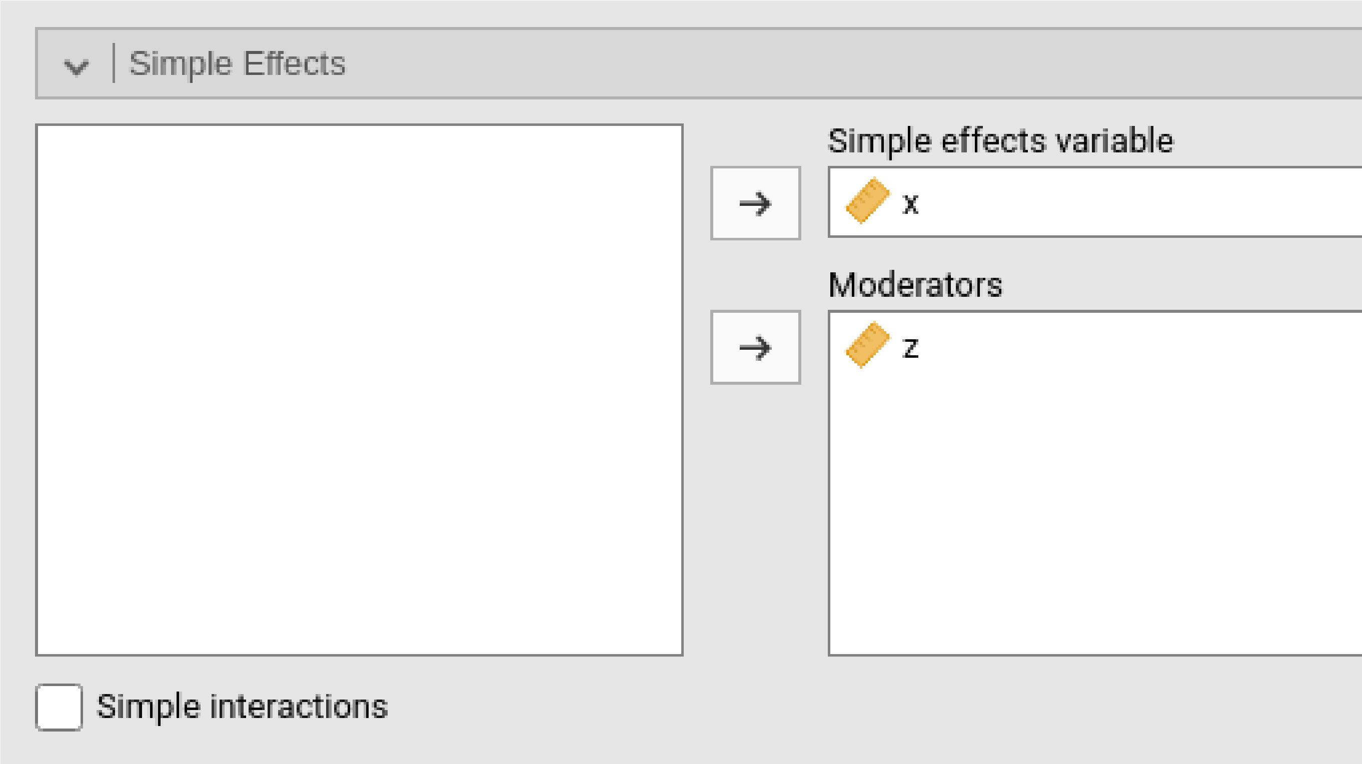
When an interaction is present (let's say it is significantly different from zero), we can probe it (Aiken et al., 1991). Probing means to estimate, test, and visualize the effect of one independent variable at different levels of the other. The "other variable" is usually called the *moderator*: The variable that is supposed to change the effect of the independent variable. In our example, we focus on the effect of beers (x), and different levels of the moderator extroversion (z). First, let's look at the plot.


Plots
Horizontal axis
x
Separate lines
z
Separate plots
Display
None
Confidence intervals
Standard Error
Plot
Observed scores
Y-axis observed range
Use
X original scale
Varying line types



We can see now that the lines representing the effect of x at different levels of z are no longer parallel, they have *different slopes*. Looking at the plot, we can see that the effect of x is stronger for high levels of z (`Mean+1*SD`), than for the average level of z (`Mean`) than for low levels of z (`Mean-1*SD`).

The plot is very useful to visualize how the effect of one variable changes at different levels of the moderator. Often, however, we also want to estimate those slopes and maybe test them. We can do that in the Simple Effects panel. Recall that *simple slopes* is just a name for *simple effects* applied to continuous variables, so GAMLj uses the term *simple effects* because it generalizes to any type of independent variable.


Simple Effects
Simple effects variable
x
Moderators
z
Simple interactions

Simple Effects

ANOVA for Simple Effects of x

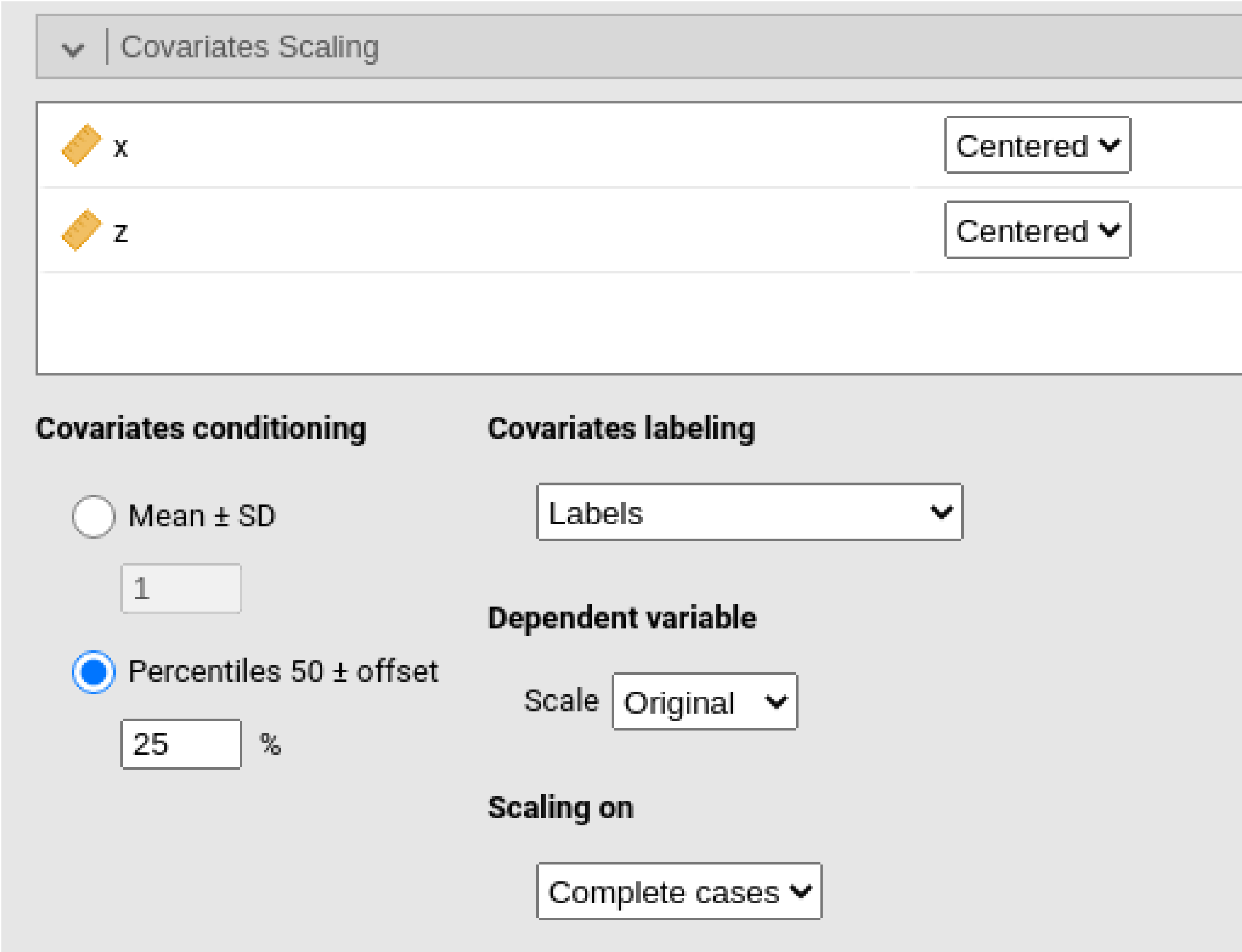
Moderator					
z	F	Num df	Den df	p	η²p
Mean-1·SD	12.8	1	116	< .001	0.099
Mean	55.8	1	116	< .001	0.325
Mean+1·SD	52.8	1	116	< .001	0.313

Parameter Estimates for simple effects of x

Moderator				95% Confidence Intervals		
z	Effect	Estimate	SE	Lower	Upper	β
Mean-1·SD	x	2.563	0.717	1.143	3.983	0.38
Mean	x	3.686	0.493	2.709	4.663	0.55
Mean+1·SD	x	4.808	0.662	3.497	6.119	0.71

The first table reports the F-tests and indices of explained variance (2.3.2). The coefficients table reports the slopes of the effect of the `x` variable at different levels of the `z` variable. Practically, the tables report the effect sizes and the inferential tests associated with the lines depicted in the plot.

If one wants to change the levels of the moderator at which the effects are estimated and plotted, one can go to the Covariate Scaling panel and change the `Covariate conditioning` setup. For instance, one may want to explore the effect of `x` at 2 SD away from the mean, so one sets the field under Mean $\pm$ SD to 2. One can also decide to condition the slopes to specific percentiles of the `z` variable, by selecting Percentiles $\pm$ SD, which conditions to 25th, 50th, 75th percentile (cf GAMLj help page).


Covariates Scaling
x
Centered
z
Centered
Covariates conditioning
Mean ± SD
1
Percentiles 50 ± offset
25
%
Covariates labeling
Labels
Dependent variable
Scale
Original
Scaling on
Complete cases

Simple Effects

ANOVA for Simple Effects of x

Moderator z	F	Num df	Den df	p	η^2p
50-25%	30.2	1	116	< .001	0.207
50%	56.1	1	116	< .001	0.326
50+25%	63.6	1	116	< .001	0.354

Parameter Estimates for simple effects of x

Moderator				95% Confidence Intervals		
z	Effect	Estimate	SE	Lower	Upper	β
50-25%	x	3.118	0.567	1.994	4.242	0.466
50%	x	3.692	0.493	2.715	4.668	0.551
50+25%	x	4.271	0.535	3.210	5.331	0.638

We can be happy with the analysis, as we have estimated, tested, and quantified all the interesting effects of our independent variables on the dependent variables. We discuss simple effects again in 2.13. An equivalent example, with different data, can be found in GAMLj help page: GLM example 1.

2.7 Categorical IVs

AKA: ANOVA

We can now work with a model with categorical independent variables. The dataset provides `cat2` and `cat3` variables, with two groups and three groups respectively. Their combination produces the following groups.

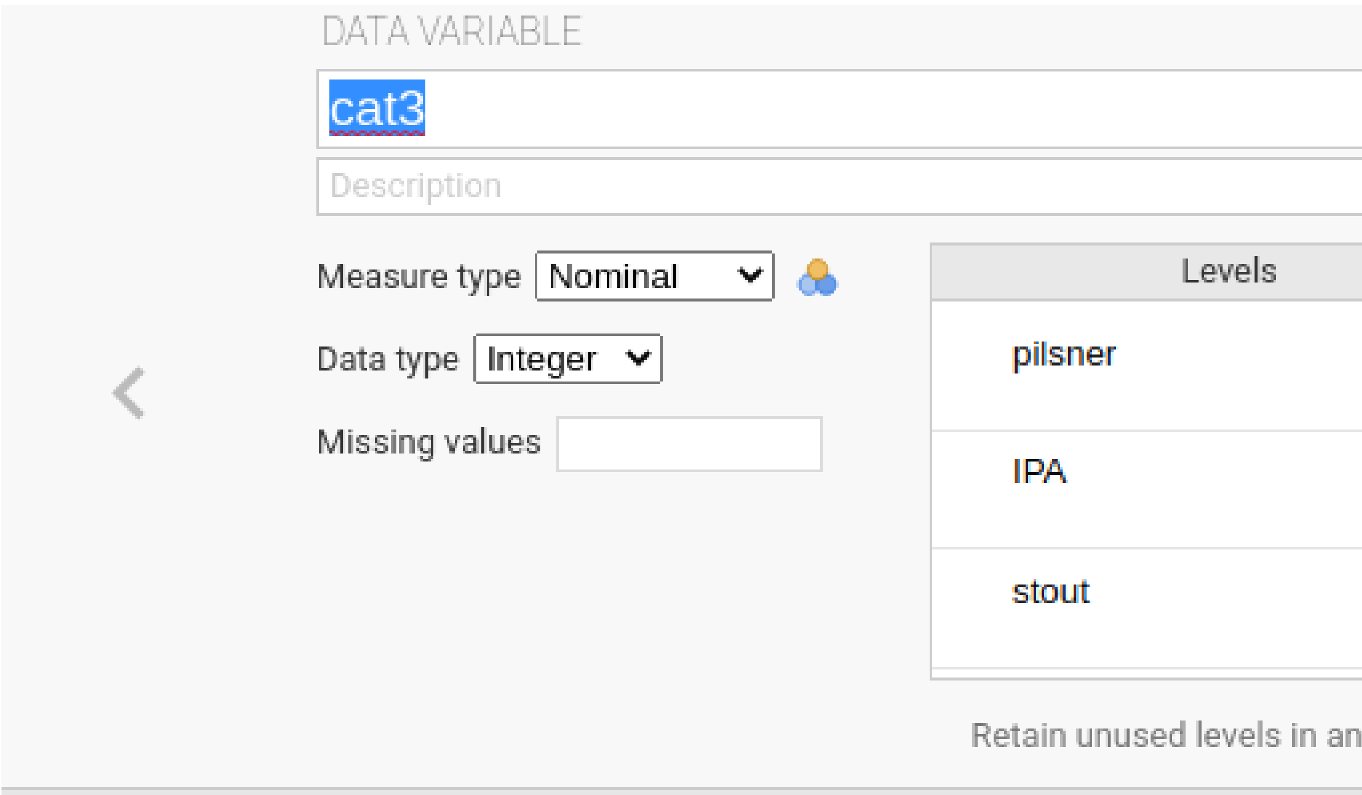
Contingency Tables

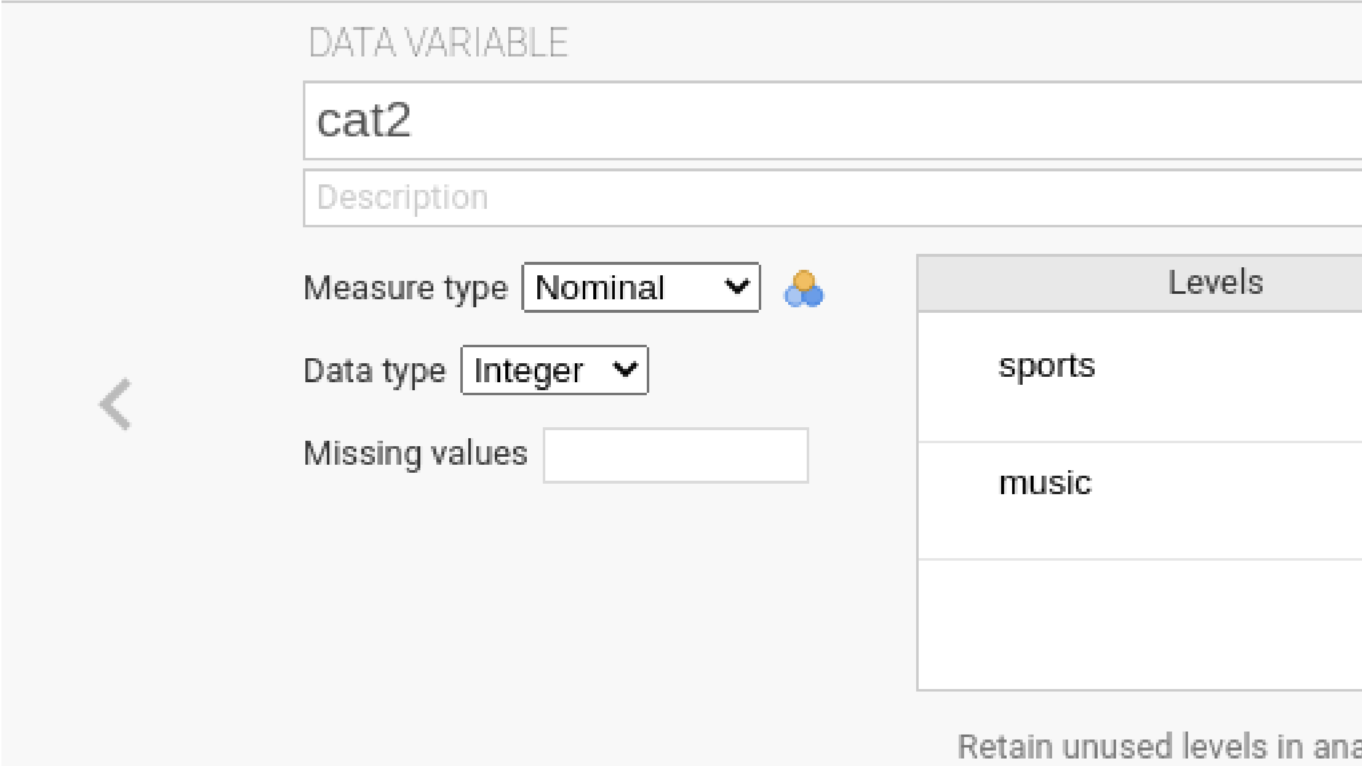
Contingency Tables

cat2	cat3 -1	cat3 0	cat3 1	Total
-1	20	20	20	60
1	20	20	20	60
Total	40	40	40	120

>

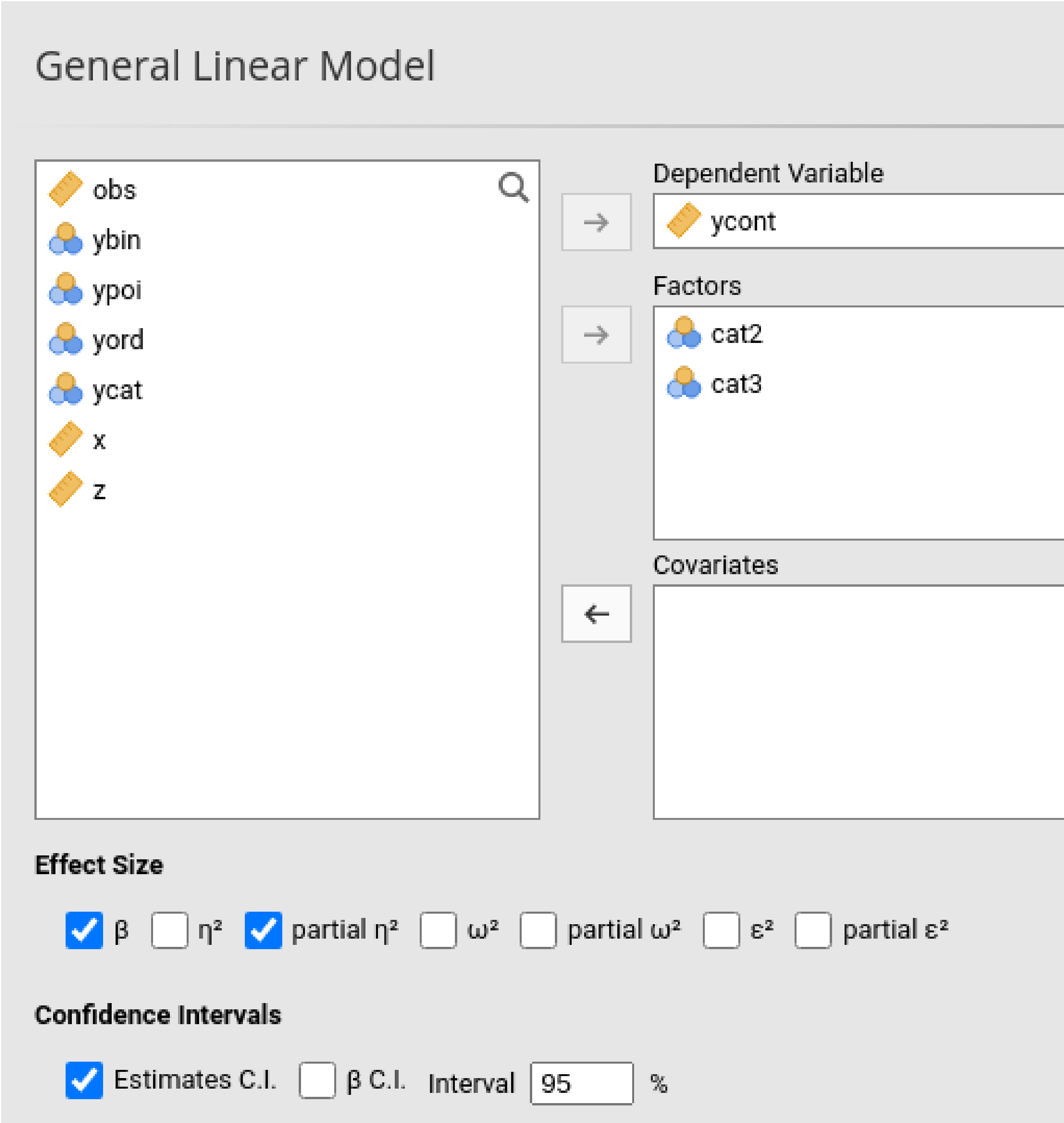
To put some flesh on the bones, let's imagine that cat3 be the type of beer one drinks, with levels *stout*, *IPA* and *pilsner*. Assume `cat2` be type of bar, *music bar* vs *sports bar*. To remember, let's put some labels on the variables levels.





Take a note about the fact that the specific values present in the dataset of a categorical variable have no bearing on the results of the analysis. The categorical variables are coded by GAMLj independently of their values: It applies a coding system to cast the categorical IV Independent Variable into the model. We can also change the coding system with the module options (see below).

We can now run a new model, with `ycont` as dependent variable, and the two categorical variables as the independent ones.



The tables we have seen for the first model are the same produced now, we only need to adjust some interpretation to reflect the categorical nature of the variables.

2.7.1 Model Fit and Omnibus tests

Model Fit

R^2	Adj. R^2	df	df (res)	F	p
0.162	0.125	5	114	4.40	0.001

[4]

ANOVA Omnibus tests

	SS	df	F	p	η^2p
Model	864.577	5	4.405	0.001	0.162
cat2	565.747	1	14.412	< .001	0.112
cat3	263.137	2	3.352	0.039	0.056
cat2 ✻ cat3	35.694	2	0.455	0.636	0.008
Residuals	4475.217	114			
Total	5339.794	119			

>

Model Fit and ANOVA Omnibus tests tables do not require adjustments of the interpretation. Here we see that our model explains $.162 * 100$ percent (R^2) of the dependent variable variance, $.125 * 100$ percent as population estimate (Adj. R^2), and that the type of bar (`cat2`) has a main effect, type of beer (`cat3`) has a main effect, and there is no indication of an interaction. Effects are small, with `cat2` main effect explaing $.112 * 100$ percent of the variance not explained by the other effects, `cat3` main effect explaining $.056 * 100$ percent, and the interaction explaining only around 2% of the partial variance.

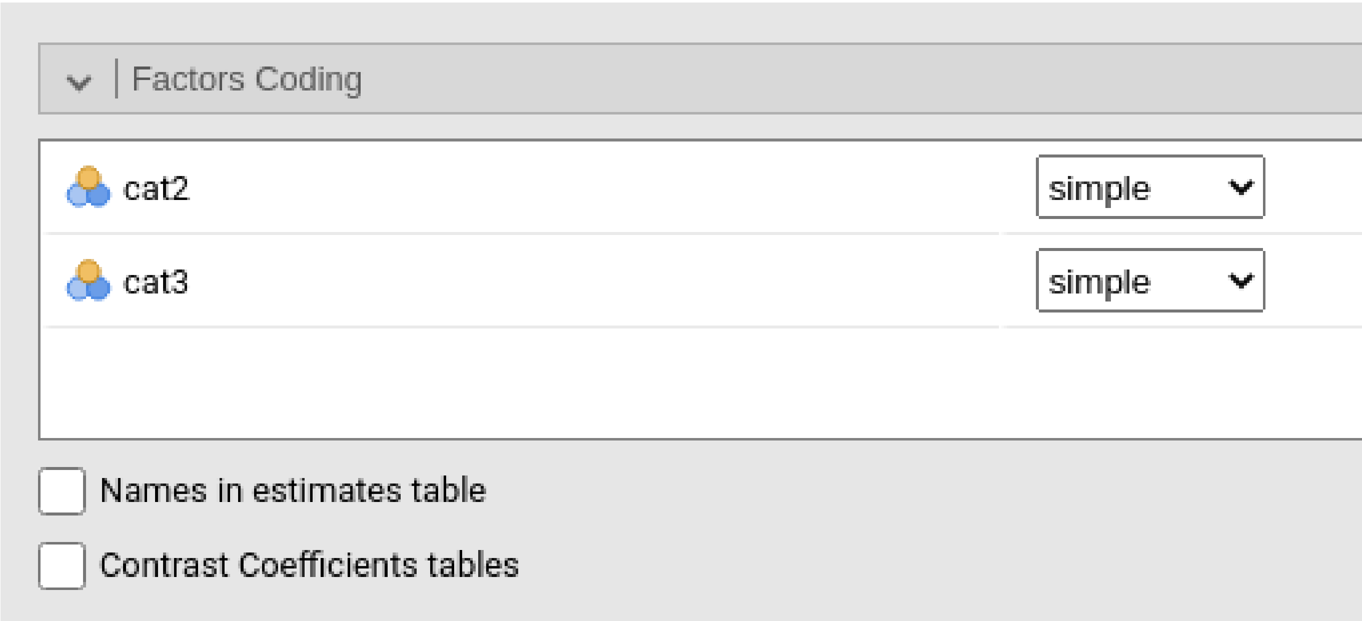
2.7.2 Coefficients

When dealing with categorical independent variables, one usually does not look at the coefficients, but one goes straight to the plots to interpret the results. Nonetheless, the coefficients are present and they can be interpreted.

Parameter Estimates (Coefficients)

				95% Confi
Names	Effect	Estimate	SE	Lower
(Intercept)	(Intercept)	31.738	0.572	30.605
cat21	music - sports	4.343	1.144	2.077
cat31	IPA - pilsner	2.788	1.401	0.013
cat32	stout - pilsner	3.403	1.401	0.628
cat21 ✻ cat31	(music - sports) ✻ (IPA - pilsner)	-2.129	2.802	-7.680
cat21 ✻ cat32	(music - sports) ✻ (stout - pilsner)	0.334	2.802	-5.217

Their interpretation depends on the way the levels (groups) of the variables are coded. In fact, to cast a categorical variable into a linear model (any linear model), it must be coded with weights (numbers) that represent specific contrasts. We need $K - 1$ contrasts to represent K groups (see appendix B for more details). These contrasts are commonly called *dummy variables*, but it is more correct to call them *contrast variables*. GAMLj default uses the *simple* coding system, as it is evident in the Factor Coding tab.



Simple coding contrasts estimate the mean difference between one reference group and each of the other groups. The first level present in the dataset is used as reference group. *Simple* coding yields the same comparisons as the more classical *dummy* system, but *simple* coding centers the contrasts to zero, so in the presence of an interaction in the model, the main effects are correctly estimated as *average effects*. So, for `cat2`, we need only one *contrast* which compares *music* vs *sports*. The coefficient 4.343 is the mean difference (in the dep variable) between the two groups. So, people in the *music* bar smile 4.343 smiles more than people in the *sports* bar. For type of beer (`cat3`), we need two contrast variables, because we have three groups: `cat31` compares the reference group *pilsner* against *IPA*, the second contrast `cat32` compares *pilsner* with *stout*. The remaining coefficients are required to estimate the interaction `cat2 X cat3`.

We should notice that the model does not estimate all possible contrasts, for instance *stout* is not compared with *IPA*. The reason is that those contrasts are redundant in order to capture the whole explained variance (cf. Appendix B). If one needs all possible comparisons, one can use Post Hoc Tests panel to obtain the comparisons (cf. 2.8).

GAMLj offers several coding systems to code the categorical variables. If you want to take a look at what the contrasts are comparing, you can ask for Contrast Coefficients tables, so a table visualizing the actual coding is produced.

Contrast Coefficients

Factor: cat2

Level=sports	Level=music	Contrast
−0.500	0.500	music - sports

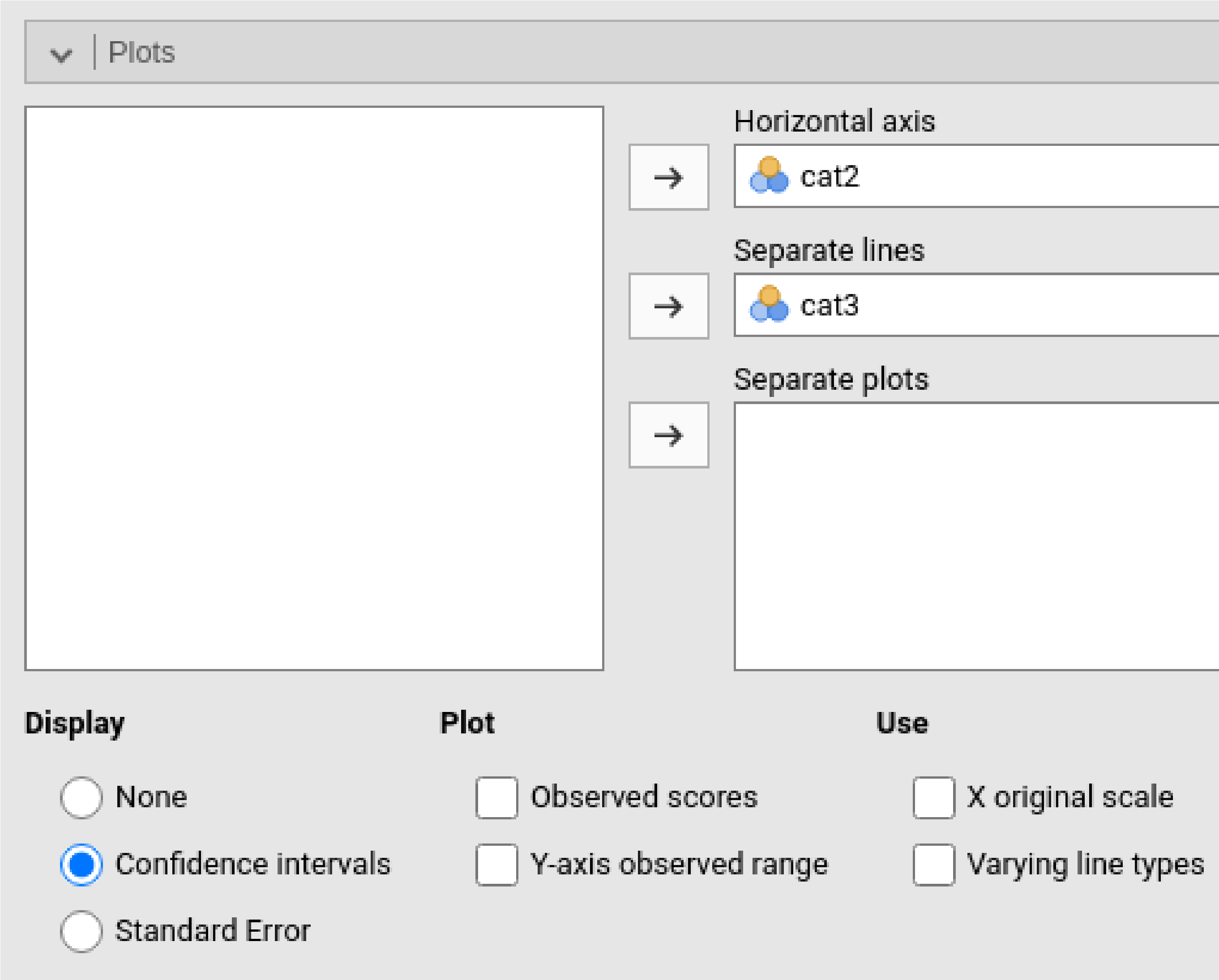
Factor: cat3

Level=pilsner	Level=IPA	Level=stout	Contrast
−0.333	0.667	−0.333	IPA - pilsner
−0.333	−0.333	0.667	stout - pilsner

All coding system used in GAMLj are explained in details in the GAMLj help page: contrasts.

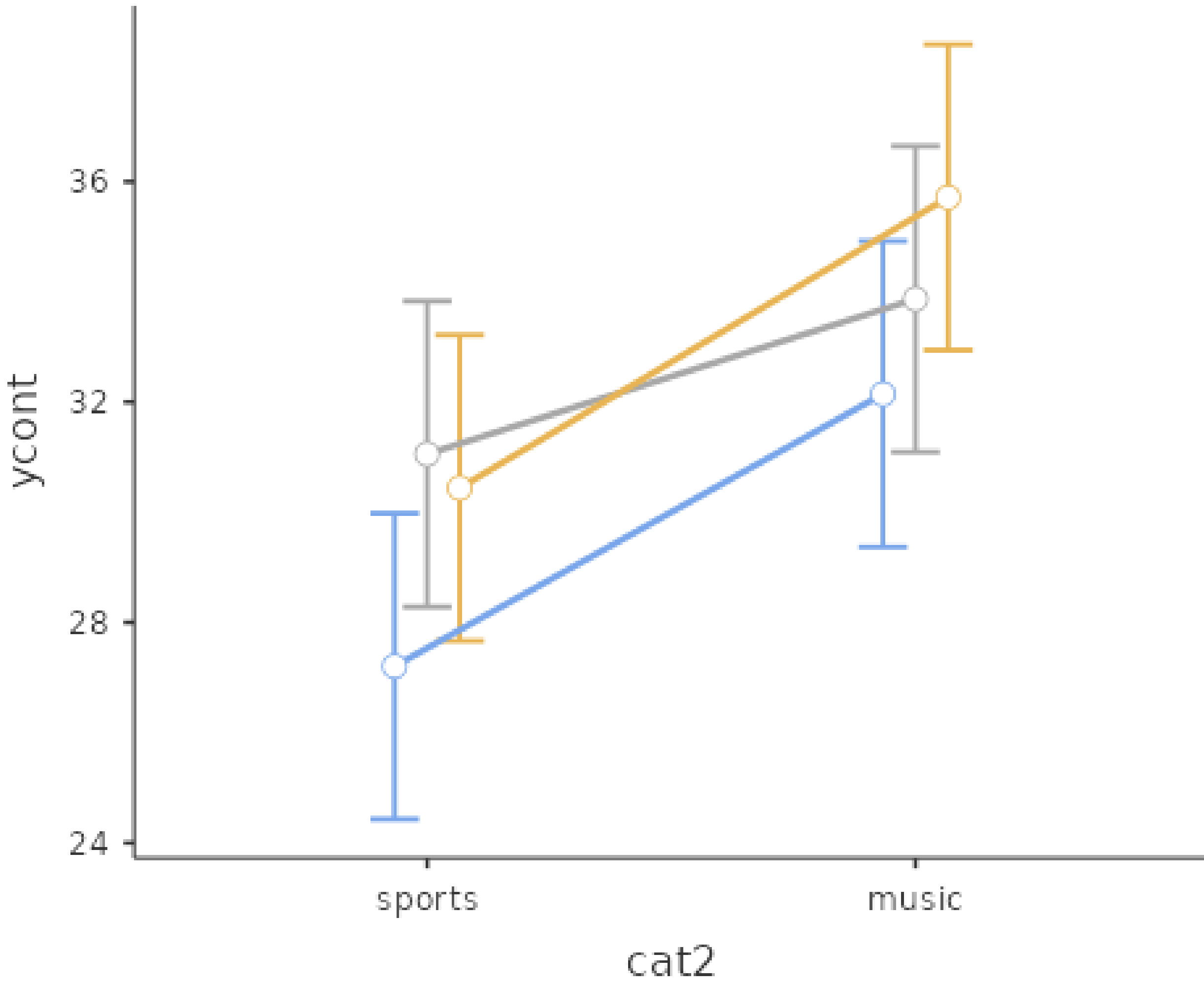
2.7.3 Plots

As for the continuous IVs Independent Variables case, we can plot the results. When the IVs Independent Variables are categorical, we obtain the plot of the means.


Plots
Horizontal axis
cat2
Separate lines
cat3
Separate plots
Display
None
Confidence intervals
Standard Error
Plot
Observed scores
Y-axis observed range
Use
X original scale
Varying line types

Results Plots

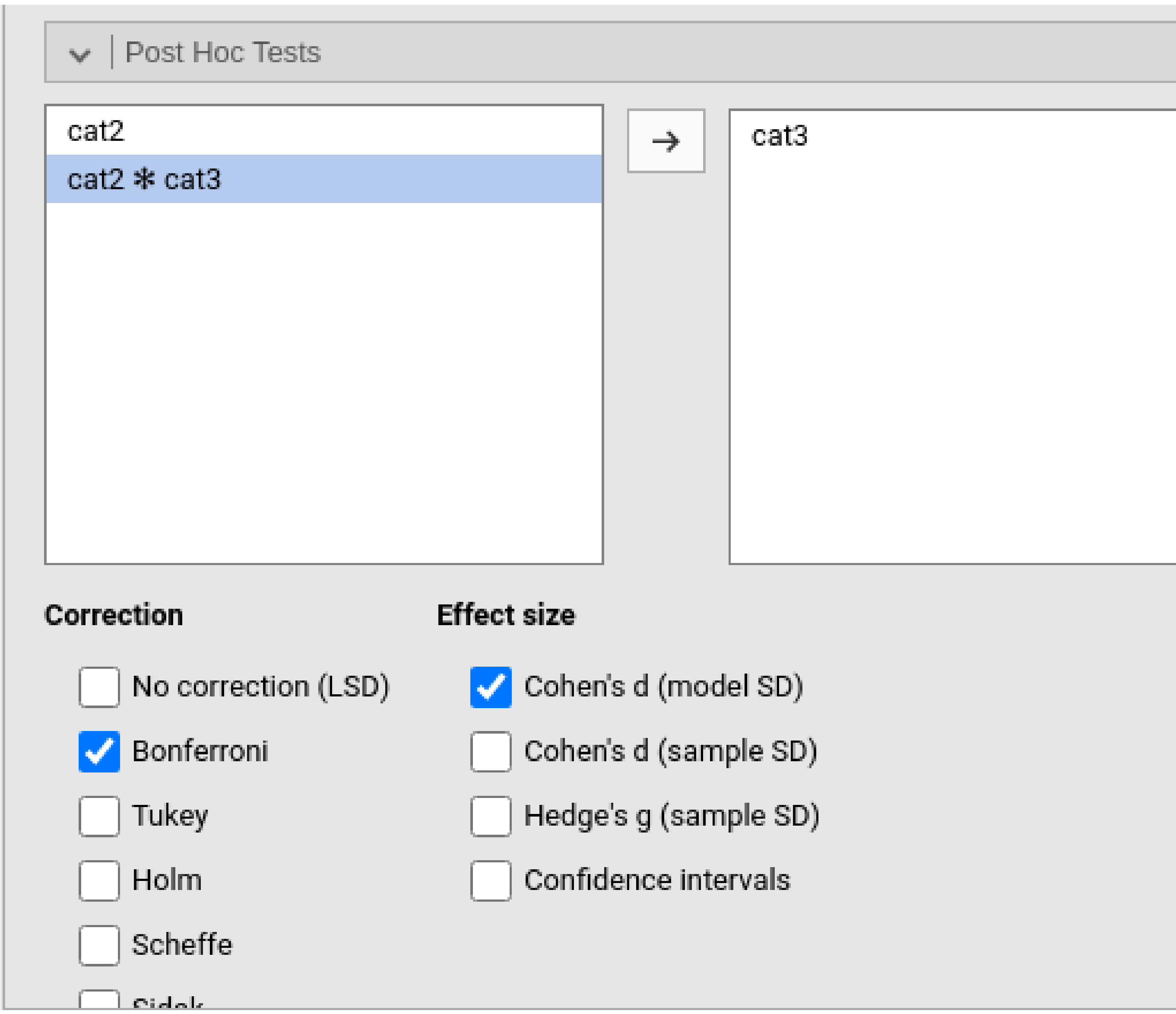
cat2 ✻ cat3



From the results is evident the main effect of `cat2`, with *music* bar showing a higher average than *sports* bar, and a small main effect of type of beer, with *stout* vaguely larger than *IPA* and larger than *pilsner*.

2.8 Post-hoc tests

Sometimes people want to probe main effects or interactions of categorical variables to test all possible comparisons among means. It should not be a habit to do so, because the coefficients table already provides comparisons that may be enough to explain the results. One can also use a simple effects analysis to test specific comparisons. Nonetheless, if one really needs all possible comparisons, one can use the Post Hoc Tests panel. Here we ask for the *post hoc* tests of the variable `cat3`, because `cat2` features only two levels, so probing is useless (we have already its main effect). In this example we do not probe the interaction, because it is very small and not significant, but the module allows probing all possible effects.



Post Hoc Tests

Post Hoc comparison: cat3

Comparison					95% Confidence Intervals	
cat3	vs	cat3	Difference	SE	Lower	Upper
pilsner	-	IPA	−2.788	1.401	−5.564	−0.013
pilsner	-	stout	−3.403	1.401	−6.179	−0.628
IPA	-	stout	−0.615	1.401	−3.390	2.161

Comparisons effect size

Comparisons: cat3

Comparison					
cat3	vs	cat3	Difference	SE	d_{mod}
pilsner	-	IPA	−2.788	1.401	−0.445
pilsner	-	stout	−3.403	1.401	−0.543
IPA	-	stout	−0.615	1.401	−0.098

Post hoc tests are basically t-tests comparing any pair of levels of the independent variables in their dependent variable means. So, here we have the mean for the *pilsner* group compared with the *IPA* group, the *pilsner* group compared with the *stout* group, and the *IPA* group compared with the *stout* group. For each comparison we have the difference in means, the confidence intervals, the t-test, df, and p-value. All columns report what a simple t-test would yield, but the p-value column is different. The p-value is adjusted for multiple comparisons, meaning that the p-value is calculated to counteract the higher probability of finding something singificant when multiple tests are run. The adjustment is proportional to the number of comparisons that are tested.

Why adjusting? Well, adjustment is required when the researcher does not have an *a priori* hypothesis regarding which comparison should be *significant* and which should not. If one does not have a clear hypothesis, any comparison that comes out as significant will be considered as *real*, so different from zero. However, when several comparisons are tested, the probability of obtaining at least one comparison as significant is not .05 (α), as one expects, but higher: the more comparisons one tests, the higher the probability.

Recall that any inferential test (*frequentist* tests, I should add) lays out a null-hypothesis, say $\Delta = 0$ (difference equal to zero). The t-test returns the probability of obtaining the observed result (here -2.788 for *pilsner* vs *stout*) if we were sampling from a distribution in which $\Delta = 0$. When the p-value is low, we say that it is very unlikely that our result comes from a distribution where $\Delta = 0$, so we reject the null-hypothesis. Unlikely, however, does not mean impossible, so there is always a chance to be wrong in rejecting the null-hypothesis. If we use a significance cut-off of $\alpha = .05$, we accept the risk of being wrong 5% of the time, if the population difference is indeed zero. The good news is that we'll be right $1 - \alpha = .95$ (*100) of the times. However, this reasoning is valid for one test. If we run two tests, we want to take the right decision for boths, so the probability of being right in both tests is $(1 - \alpha)^2 = 0.9025$. If we run three tests, we will be right with probability $(1 - \alpha)^3 = 0.857375$. So, we capitalize on chance, meaning that it becomes more and more likely to get at least one test as significant, even if they all come from a population where no difference is present.

Multiple comparisons adjustment corrects the p-value in order to make a significant result more difficult to obtain. Practically, the p-value is increased proportionally to the number of comparisons that are tested. There are different methods to adjust the p-value, and they are listed in the panel as options: *Bonferroni*, *Tukey*, *Holm*, *Scheffe*, *Sidak*. They are all alternative ways to adjust the p-value. The interesting dimension along which they differ is liberalism-conservativism. In statistics, a *liberal* test is more likely to yield a significant result than a conservative test, *ceteris paribus*. Liberal tests are more powerful but their correction of the p-value may not be enough, whereas conservative tests correct for multiplicity but may result under-powered. Bonferroni and Sidak adjustment tend to be more conservative than the others. *Tukey* correction seems the more reasonable choice in most circumstances (Midway et al., 2020).

Recall, post hoc tests are needed when the researcher is willing to accept any significant result as worth mentioning and interpreting. On the other hand, if one has a clear pattern of means hypothesized, adjustment may not be needed and the specific comparisons may be evaluated without correction.

2.9 Cohen's d

Cohen's d is probably the most used effect size index to quantify a mean difference. It expresses the mean difference in a standardized scale: It is the ratio of the mean difference over the within groups standard deviation, or residual standard deviation. Unfortunately, Cohen (Cohen, 2013) defined the d index for the population, and thus it is not clear how to compute it in the sample. GAMLj offers three options.

- Cohen's d (model SD) d_{mod}: the means difference is divided by the estimated standard deviation computed based on the model residual variance.
- Cohen's d (sample SD) d_{sample}: the means difference is divided by the pooled standard deviation computed within each group.
- Hedge's g g_{sample}: the means difference is divided by the pooled standard deviation computed within each group, corrected for sample bias. The correction is the one describe by Hedges and Olkin (2014) based on the Gamma function.

Comparisons effect size

Comparisons: cat3

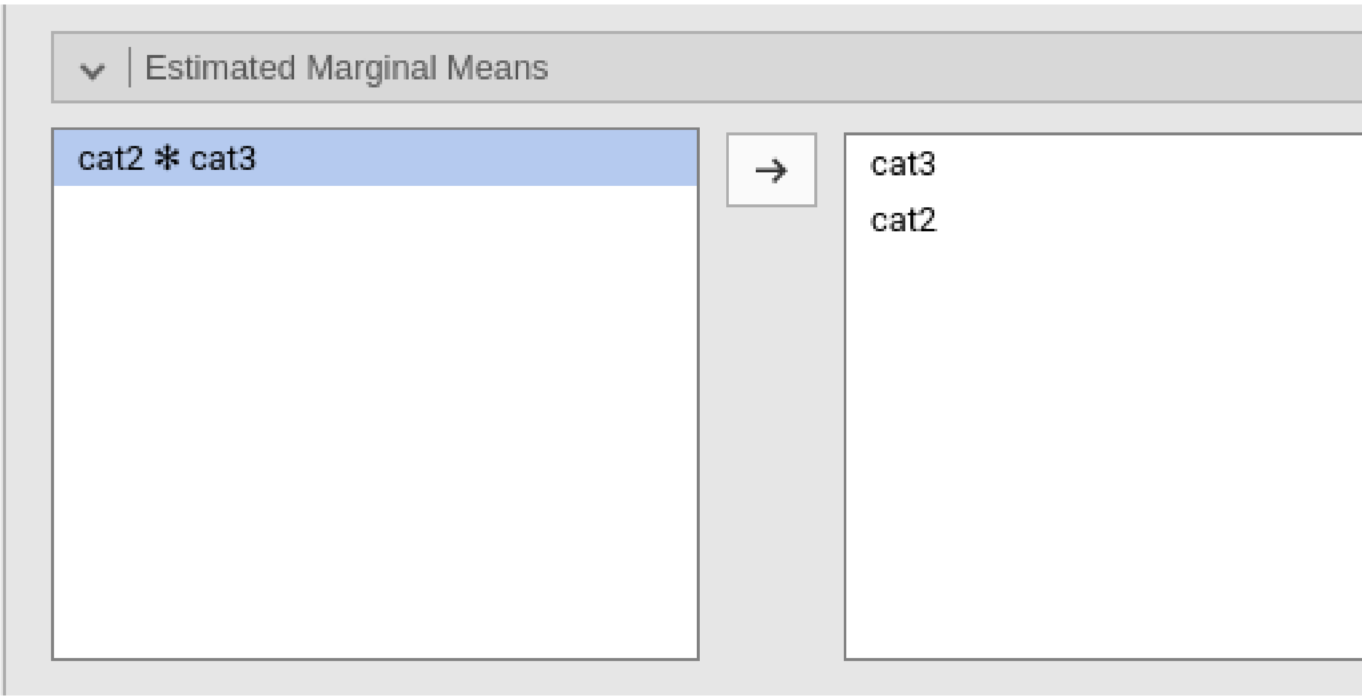
Comparison						95% Confide
cat3	vs	cat3	Difference	SE	d_{mod}	Lower
-1	-	0	−2.788	1.401	−0.445	−0.816
-1	-	1	−3.403	1.401	−0.543	−0.916
0	-	1	−0.615	1.401	−0.098	−0.465

The two d's differ in their adherence with the model being estimated. The *model SD* version, gives the estimated `d` after removing the variance explained by the other effects in the model, so it is the actual effect size of the comparison of the estimated marginal means. The *sample SD* gives the crude standardized difference, as if the model was not estimated at all, but only the two groups were considered. The *model SD* should be preferred as default index to report in a model, the *sample SD* version can be useful to compare effects in the literature obtained with a different model or without a model.

Hedge's g gives a population estimate of the sample *d*.

2.10 Estimated marginal means

The means that are plotted in the plot can be visualized, with their standard error and confidence intervals, by defining the variables for which we want them in the Estimated Marginal Means.



Estimated Marginal Means

Estimate Marginal Means - cat3

				95% Confidence Intervals	
cat3	Mean	SE	df	Lower	Upper
pilsner	29.674	0.991	114	27.712	31.637
IPA	32.463	0.991	114	30.500	34.425
stout	33.078	0.991	114	31.115	35.040

Estimate Marginal Means - cat2

				95% Confidence Intervals	
cat2	Mean	SE	df	Lower	Upper
sports	29.567	0.809	114	27.965	31.169
music	33.910	0.809	114	32.307	35.512

In balanced designs with only categorical IVs Independent Variables , they are the means of the groups (or combinations of groups). When there are also continuous IVs Independent Variables , they are adjusted for the continuous variables: They are the means estimated after keeping constant the continuous independent variables.

If marginal means are requested for a continuous variable, they represent the expected value of the dependent variable for three *interesting* levels of the independent variable, where the *interesting values* are defined as for the *simple slope* technique (cf 2.6)

For instance, in model 2.4, if we ask for the estimated marginal means for x, we obtain the following estimates:

Estimated Marginal Means

Estimate Marginal Means - x

				95% Confidence Intervals	
x	Mean	SE	df	Lower	Upper
Mean-1·SD	28.023	0.694	116	26.648	29.398
Mean	31.709	0.490	116	30.739	32.678
Mean+1·SD	35.394	0.696	116	34.016	36.772

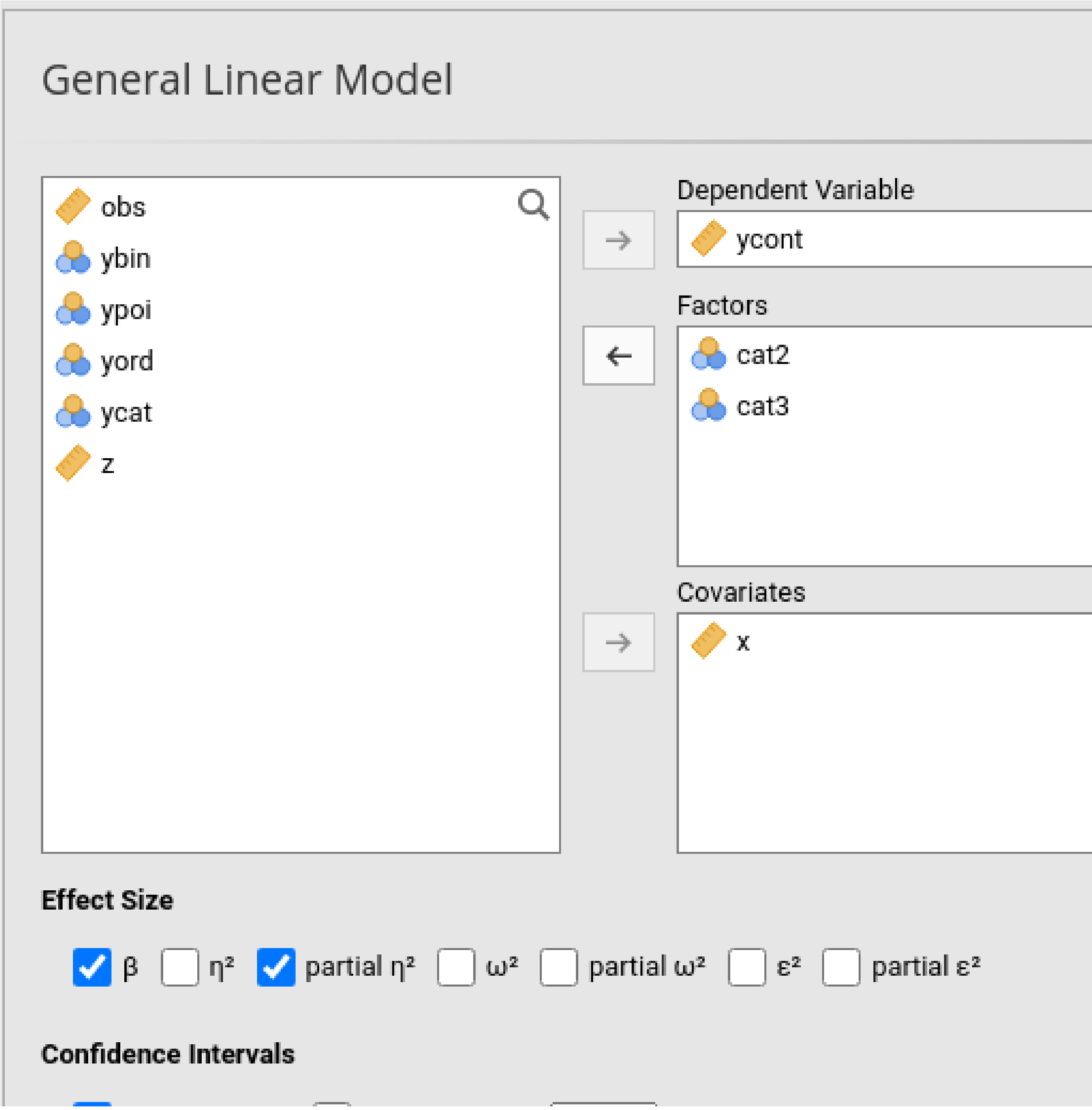
>

meaning that, based on the model, we expect the number of smiles (`ycont`) to be 28.03 for low level of beers (1 SD below average of `x`), 31.7 for the average level of beers (average of `x`), and 35.9 for high levels of beers (1 SD above average of `x`).

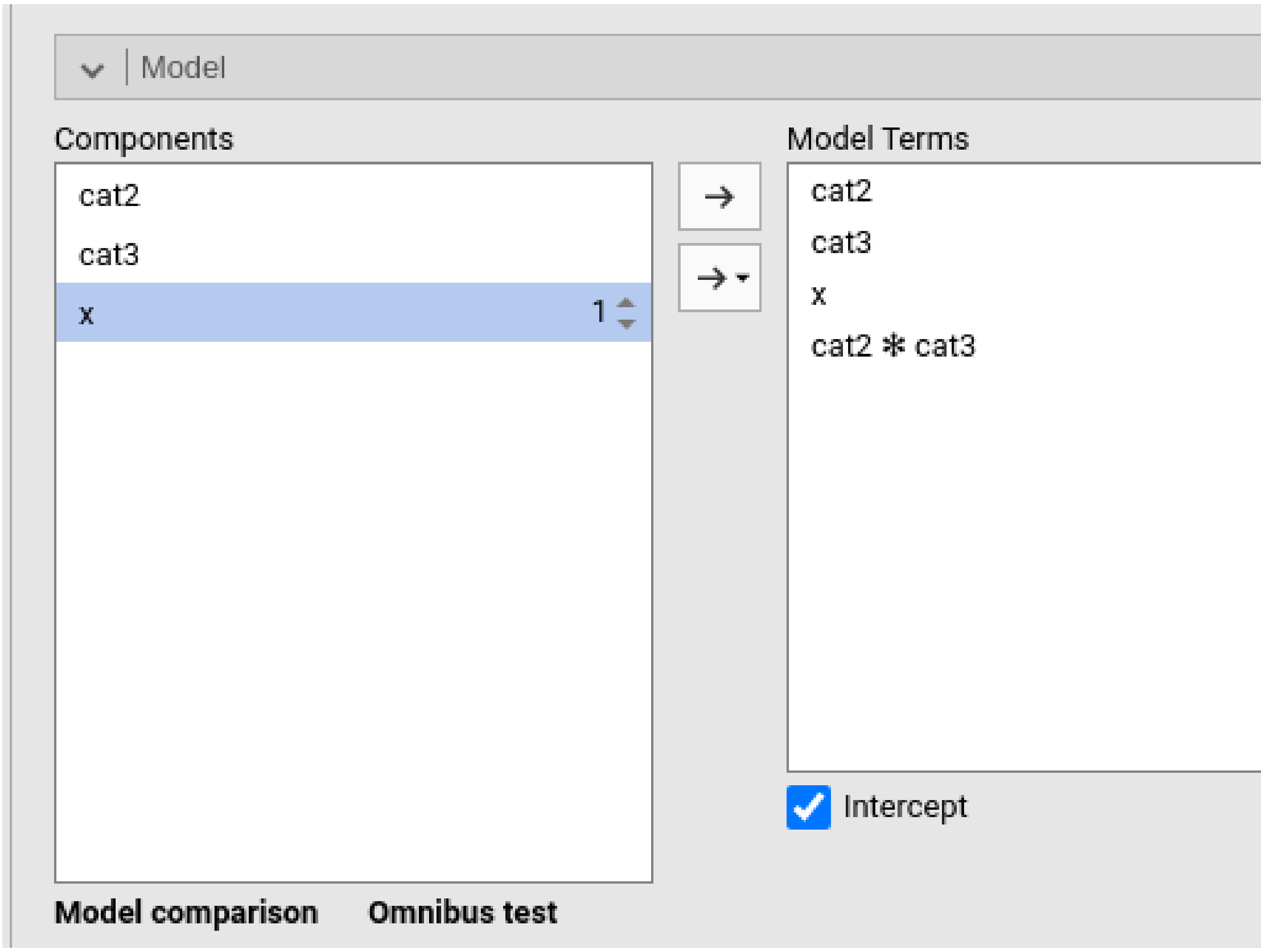
2.11 Categorical and Continuous IVs

AKA: ANCOVA

We now insert in the model also x, so we have both categorical and continuous IVs Independent Variables . This model was once called *ANCOVA*, but it did not allow for interactions. We simply call it a GLM.



Keeping up with an old (SPSS?) tradition, GAMLj does not automatically insert into the model interactions involving a continuous variable, so if one needs them, they should be added manually (see below). For the moment, here is our model.


Model
Components
cat2
cat3
x
1
Model Terms
cat2
cat3
x
cat2 ✻ cat3
Intercept
Model comparison
Omnibus test

2.11.1 Results tables

Model Results

Model Fit

R^2	Adj. R^2	df	df (res)	F	p
0.376	0.343	6	113	11.4	< .001

[4]

ANOVA Omnibus tests

	SS	df	F	p	$\eta^2 p$
Model	2009.509	6	11.364	< .001	0.376
cat2	41.725	1	1.416	0.237	0.012
cat3	127.780	2	2.168	0.119	0.037
x	1144.932	1	38.849	< .001	0.256
cat2 ✻ cat3	9.778	2	0.166	0.847	0.003
Residuals	3330.286	113			
Total	5339.794	119			

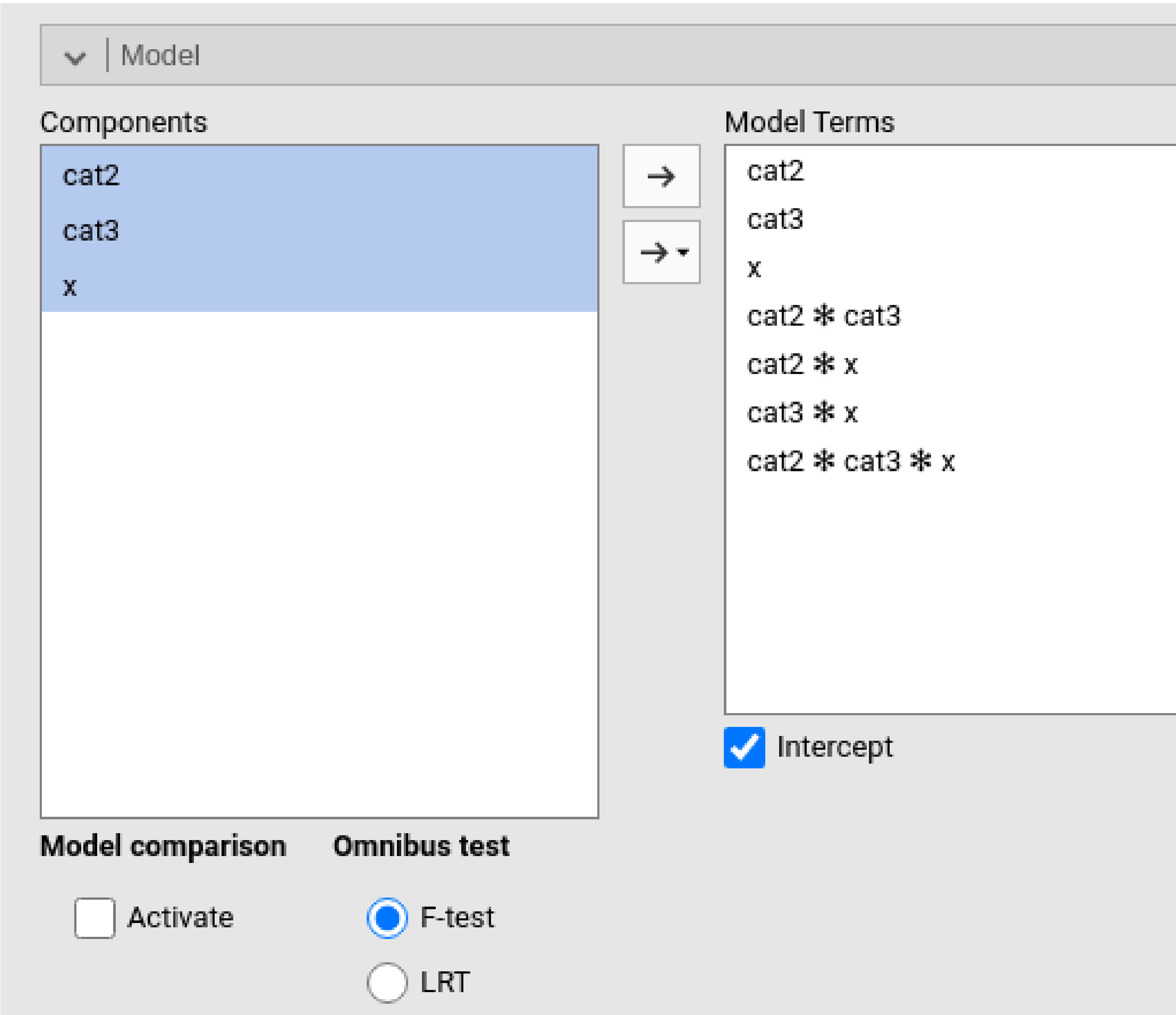
Parameter Estimates (Coefficients)

				95% Confidence Intervals	
Names	Effect	Estimate	SE	Lower	Upper
(Intercept)	(Intercept)	22.657	1.416	19.852	25.462
cat21	1 - -1	2.113	1.776	-1.405	5.631
cat31	0 - -1	1.695	1.227	-0.735	4.125
cat32	1 - -1	2.499	1.223	0.076	4.921
x	x	3.308	0.531	2.256	4.359
cat21 ✻ cat31	(1 - -1) ✻ (0 - -1)	-0.181	2.448	-5.030	4.669
cat21 ✻ cat32	(1 - -1) ✻ (1 - -1)	1.111	2.431	-3.705	5.928

Combining categorical and continuous IVs Independent Variables does not really change the way we interpret the results. We interpret the continuous variable effect like we did for the continuous variables GLM (2.3) and the categorical independent variables effects as we did for the GLM with categorical variables (2.7). So, in the Model Fit Table we found the variance explained by all the effects combined, in the ANOVA Omnibus Tests Table we find the explained variances and their tests, and in the Parameter Estimates (Coefficients) Table we find the coefficients. We just need to keep in mind that all the effects are computed keeping constant the other variables, so we can use this kind of model to *covariate* variables that may have spurious effects. That is why in the last century this model was called *ANalysis of COVAriance*. At that time, one assumption of this analysis was that there was no interaction between the categorical variables and the continuous variables. Nowadays, we can release the assumption, and just test the interaction, in case is there.

2.12 Categorical and Continuous Interactions

There is nothing special about interactions between continuous and categorical variables, they just test a moderation model. To obtain the interactions we select all variables in the Model panel and click the arrow to bring them in the `Model Terms` field.



Upon updating the model, the results update as well, and now we can check the main effects, the 2-way interactions and the 3-way interaction. We focus on the variances explained and F-tests.

Model Fit

R²	Adj. R²	df	df (res)	F	p
0.505	0.454	11	108	10.0	< .001

[4]

ANOVA Omnibus tests

	SS	df	F	p	η²p
Model	2694.506	11	10.001	< .001	0.505
cat2	165.015	1	6.737	0.011	0.059
cat3	96.462	2	1.969	0.145	0.035
x	880.953	1	35.967	< .001	0.250
cat2 ✻ cat3	19.142	2	0.391	0.677	0.007
cat2 ✻ x	97.908	1	3.997	0.048	0.036
cat3 ✻ x	245.889	2	5.019	0.008	0.085
cat2 ✻ cat3 ✻ x	310.577	2	6.340	0.002	0.105
Residuals	2645.289	108			
Total	5339.794	119			

>

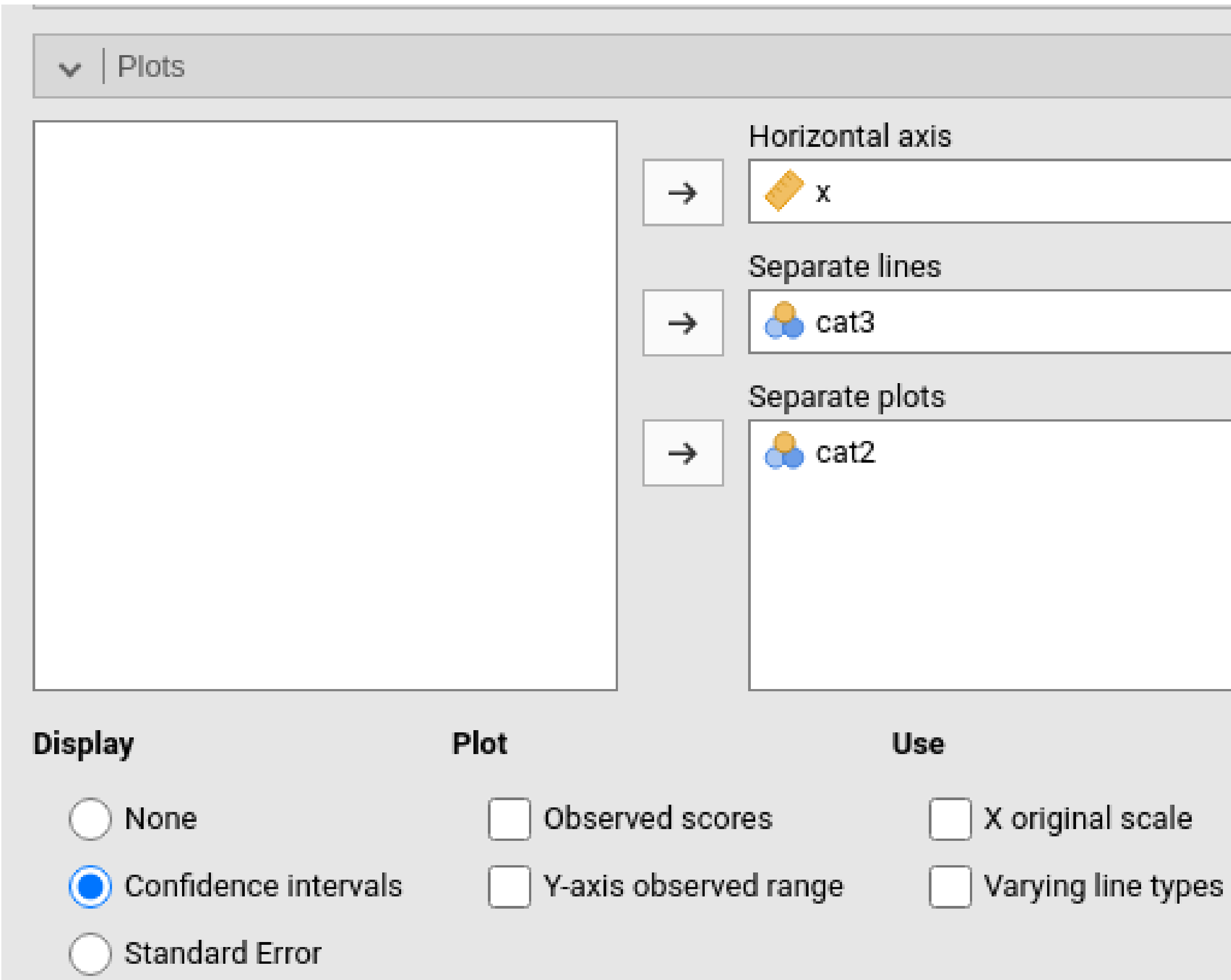
When the model features interactions of different orders, it is a good idea to start the interpretation from the highest order interaction, in our case the 3-way interaction. Here it seems to be different from zero, F(2,108)=6.340, p=.002, $p\eta^2 = .105$. This means that the 2-way interaction `x * cat3` is different at different levels of `cat2`. In general, a 3-way interaction can be interpreted by picking a moderator (any will do, the interaction is symmetrical), and saying that the other two variables interaction changes at different levels of the moderator.

Upon finding a higher order interaction, one wants to plot it and interpret its direction. This is because the higher order interaction shows a pattern of results that is more specific then lower order effects. In fact, lower order effects are always interpreted *on average*, so they are less specific than the higher order effect. Practically, the `cat2 * cat3` significant interaction in our results is not very informative given these results, because it says that *on average*, meaning averaging across levels of `x`, there is an interaction between type of bar and type of beer. But we know from the significant 3-way interaction that the 2-way interaction changes at different levels of `x`, so it is not really important its value on average.

If the 3-way interaction was not significant, or minuscule for our standards, we would simply start probing the lower order effects.

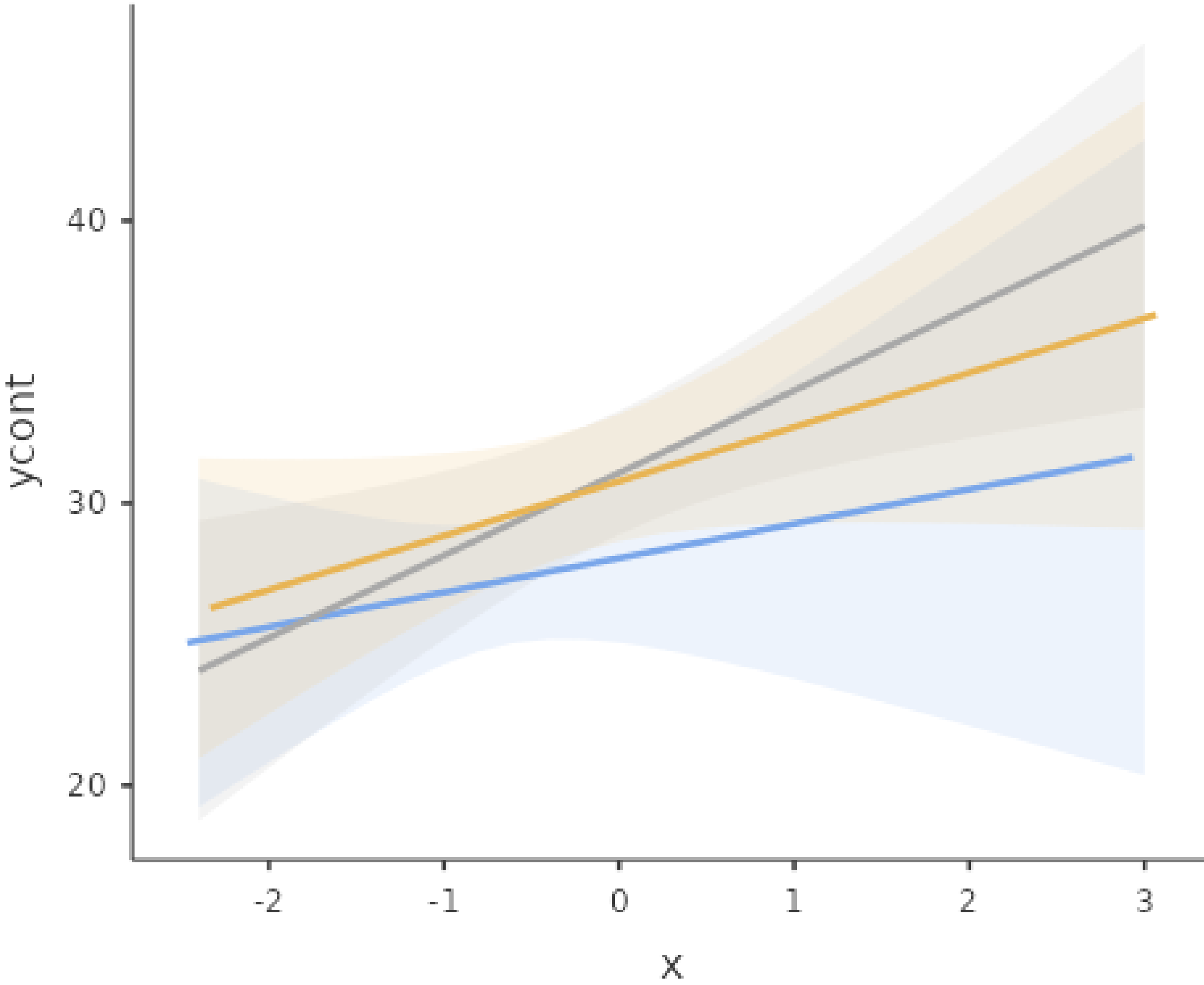
2.12.1 Plot

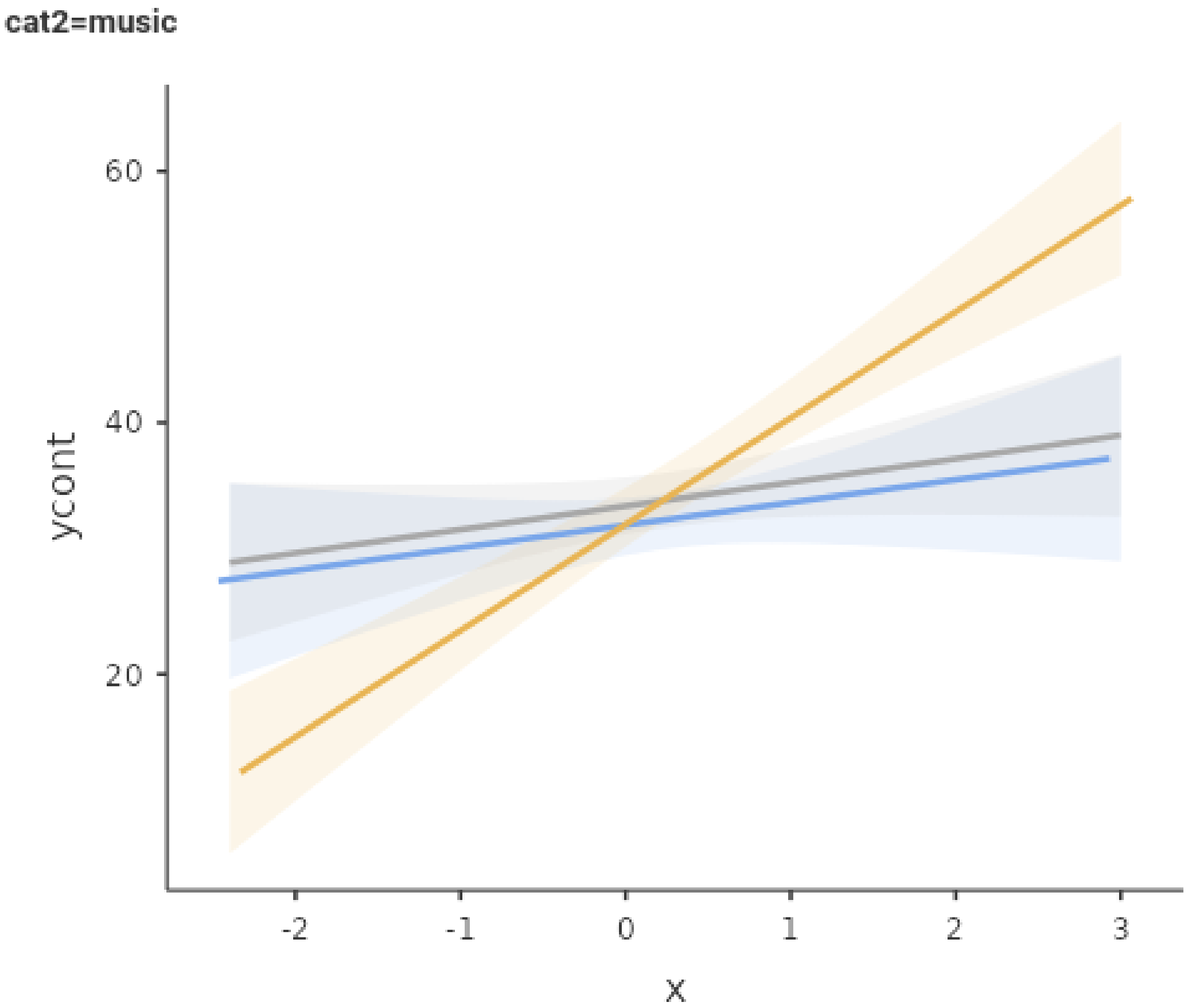
To visualize the 3-way interaction, we should pick the two variables that create the 2-way interaction we want to explore, and a moderator: the 2-way interaction is displayed for different levels of the moderator. Here we want to see the effect of beers (`x`) by type of beers (`cat3`), displayed at different levels of type of bar `cat2`.



Results Plots

cat2=sports

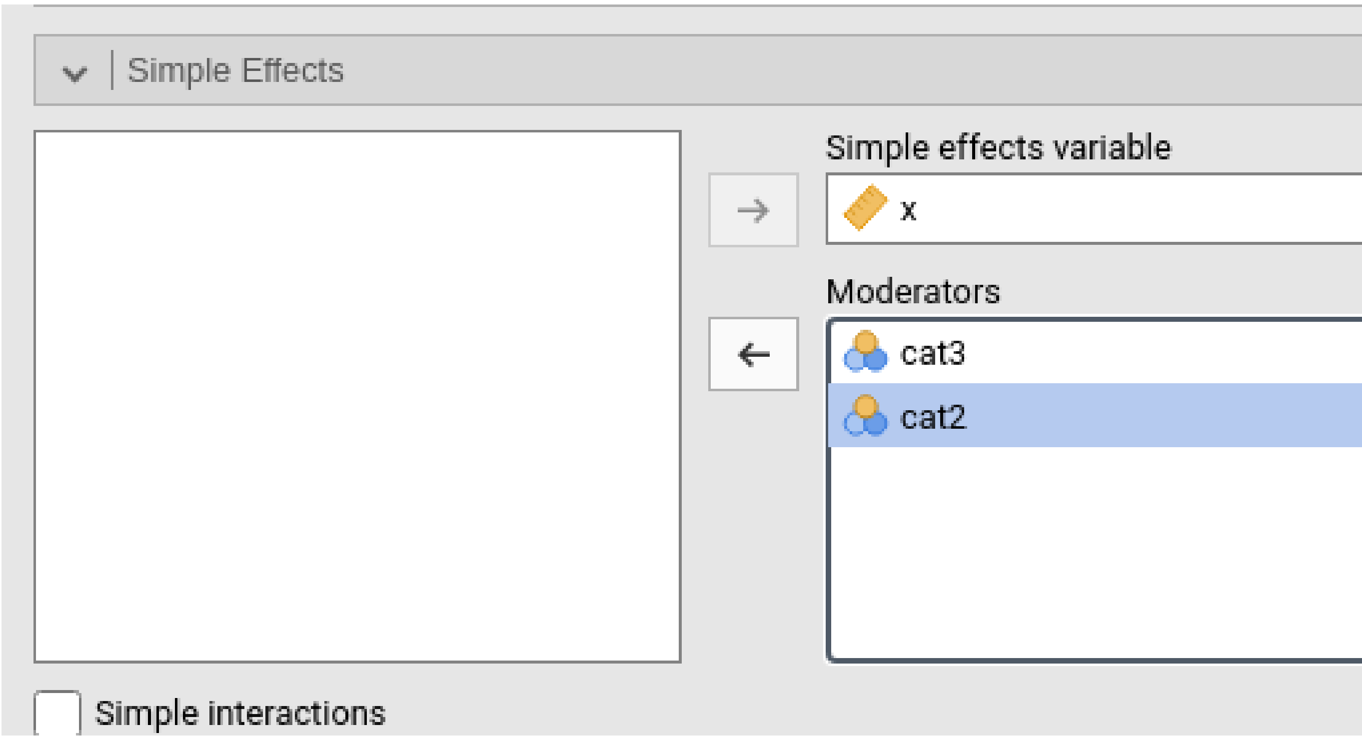




The interpretation can follow these lines: In *sports bar*, the effect of beers on smiles is generally positive, it is stronger for *stout* and IPA beers, and weaker for *pilsner*. In *music bar*, the effect of beers on smiles is positive as well, but stronger for *stout* and weaker and very similar for *pilsner* and *IPA*. The fact that the two patterns can be described differently across types of bar is justified by the significant 3-way interaction.

2.12.2 Simple Effects

We can quantify and tests all the effects depicted in the plots by asking for *simple effects* analysis. We just need to pick the variable for which we want to study the effects, and select the moderators: the focal variable effect will be estimated and tested for each combination of the moderators values.


Simple Effects
Simple effects variable
x
Moderators
cat3
cat2
Simple interactions

Simple Effects

ANOVA for Simple Effects of x

Moderator						
cat2	cat3	F	Num df	Den df	p	η²p
sports	pilsner	0.629	1	108	0.430	0.006
	IPA	8.169	1	108	0.005	0.070
	stout	2.895	1	108	0.092	0.026
music	pilsner	1.587	1	108	0.210	0.014
	IPA	2.797	1	108	0.097	0.025
	stout	58.633	1	108	< .001	0.352

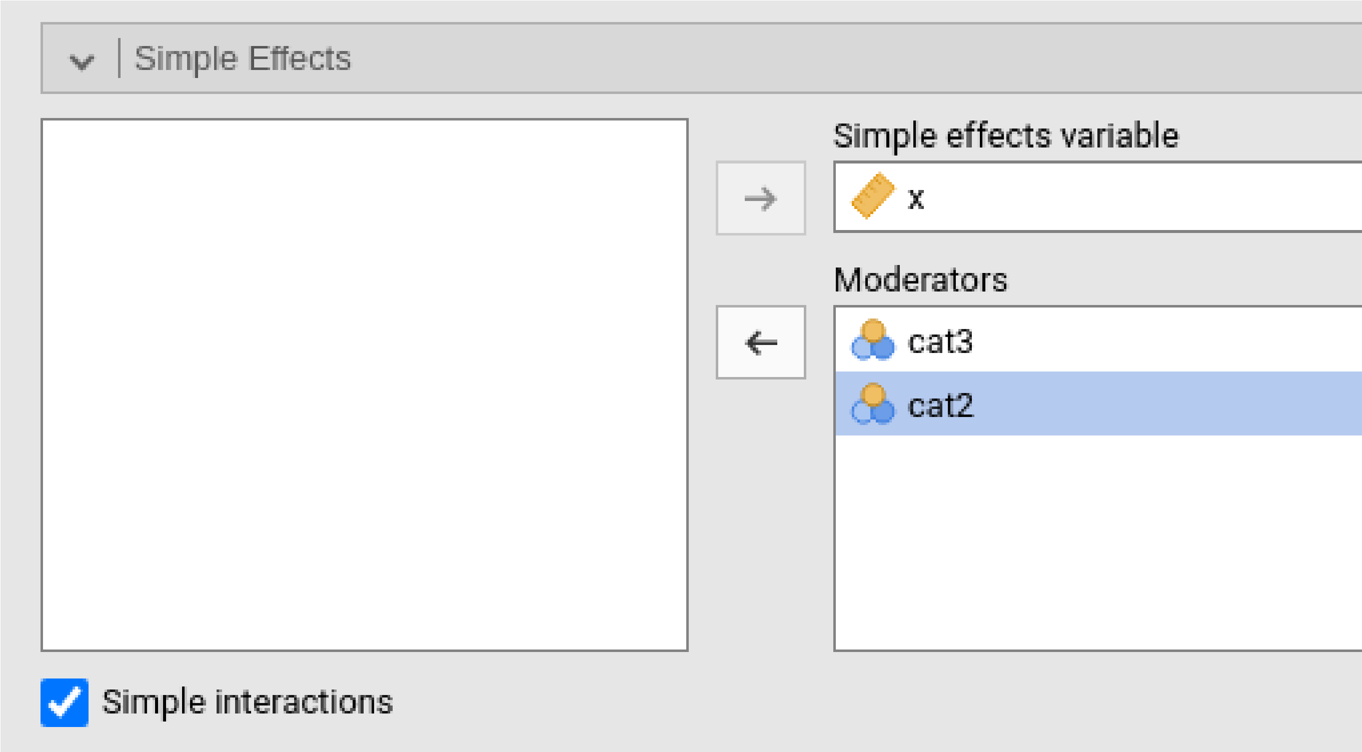
Parameter Estimates for simple effects of x

Moderator					95% Confidence Intervals	
cat2	cat3	Effect	Estimate	SE	Lower	Upper
sports	pilsner	x	1.218	1.536	−1.826	4.261
	IPA	x	2.920	1.021	0.895	4.944
	stout	x	1.925	1.131	−0.317	4.167
music	pilsner	x	1.801	1.430	−1.033	4.635
	IPA	x	1.872	1.119	−0.347	4.090
	stout	x	8.451	1.104	6.264	10.639

The interpretation of these effects follows what we have done for the 2-way interaction before (2.6). However, because we have a 3-way interaction, we can also probe for *simple interactions*.

2.13 Simple Interactions

Simple interactions are simple effects, in which the focal effect is an interaction. We can ask for this analysis by selecting Simple Interaction option.


Simple Effects
Simple effects variable
x
Moderators
cat3
cat2
Simple interactions

Simple Interactions

Interaction: x ✻ cat3

ANOVA

Moderator						
cat2	Effect	F	df1	df2	p	η^2p
sports	x ✻ cat3	0.482	2	108	0.619	0.009
music	x ✻ cat3	10.939	2	108	< .001	0.168

Parameter Estimates

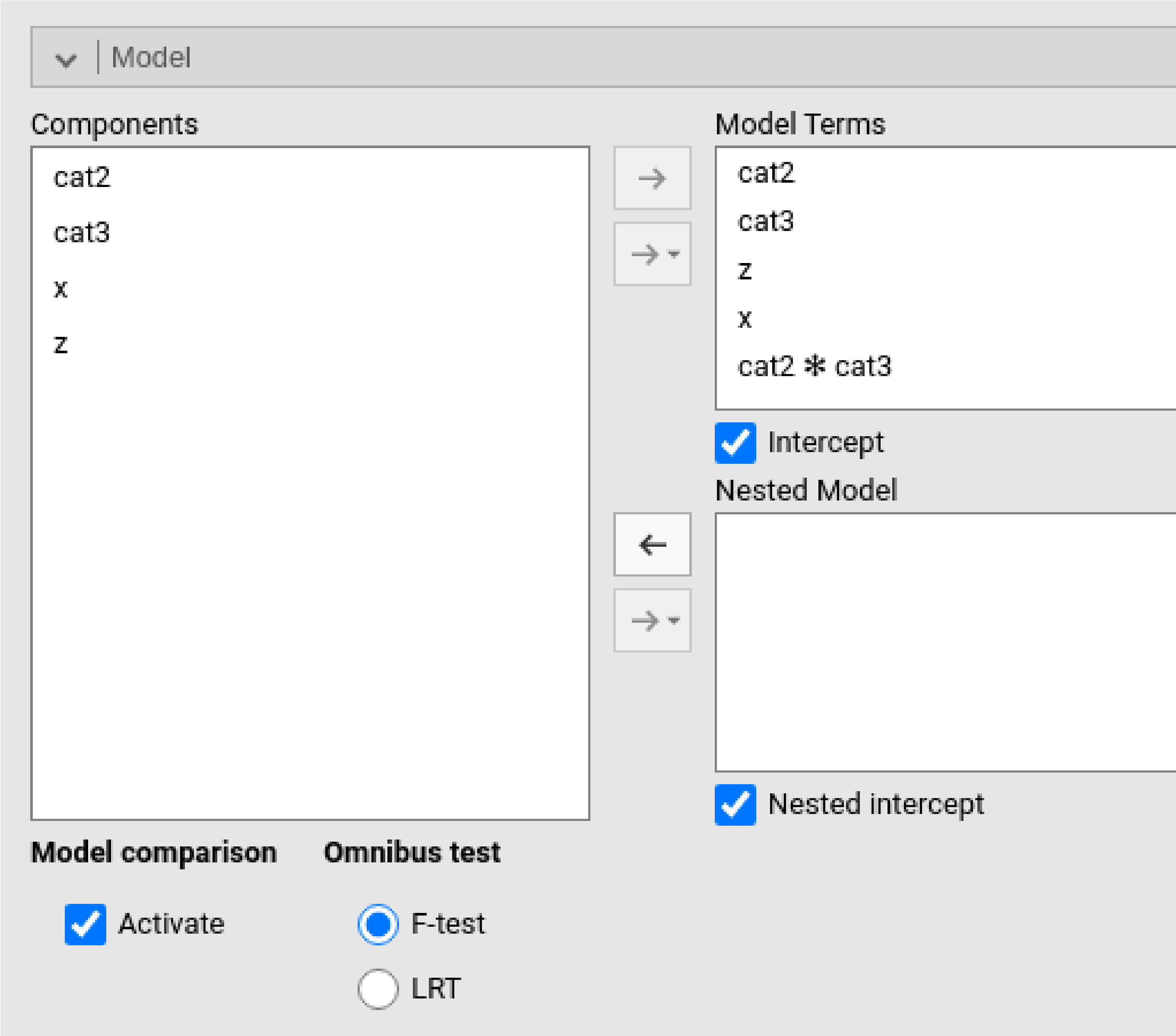
Moderator					95% Confide
cat2	Effect	Estimate	SE	df	Lower
sports	x ✻ (IPA-pilsner)	1.702	1.844	108	-1.954
sports	x ✻ (stout-pilsner)	0.707	1.907	108	-3.073
music	x ✻ (IPA-pilsner)	0.071	1.816	108	-3.529
music	x ✻ (stout-pilsner)	6.650	1.806	108	3.070

This analysis is useful to probe high order interaction. Here we see the `x * cat3` interaction computed at the two levels of `cat2`. So, we can say that for the group *music*, the interaction is present, whereas for group *sports*, the `x * cat3` interaction is not statistically significant.

2.14 Model-comparison approach

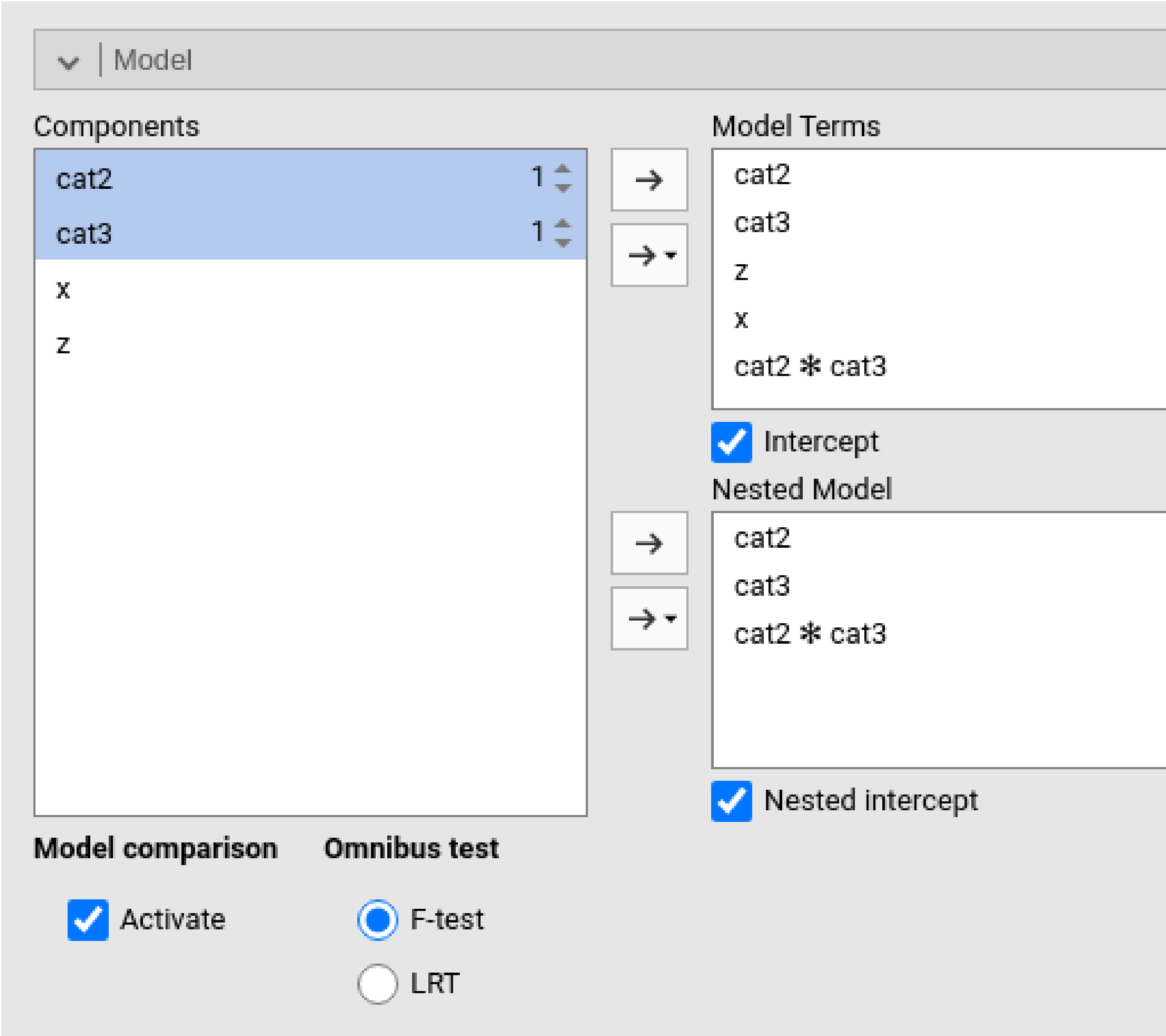
2.14.1 The Method

We have touched upon the fact that many tests and indices in the linear model can be derived by the comparison between two models, one being our full model, the other being a nested model (Judd et al., 2017), (cf. Appendix A). A nested model is simply a model that, compared with the full model, lacks some terms, and does not have any term not present in the full model. GAMLj allows custom model comparisons by flagging the option Activate under Model Comparison.



Comparing models is useful when we want to estimate the variance explained by a set of effects (Cohen et al., 2014), and test this variance. For instance, one may have an experiment with a series of possible confounding variables, and the aim is to estimate the variance explained by the experimental factors (all together) over and beyond the confounding variables explained variances. Another case may be a model in which socio-economical variables (e.g income, real-estate properties, etc) and psychological variables (e.g self-esteem, emotional regulations) are compared in their ability to explain an outcome (say *happyness*). Besides each individual variable effect size, the researcher may be interested in estimating the variance explained by *economics* vs the variance explained by *psychology*. A model-comparison approach may be helpful.

In our running example, we would like to estimate and test the impact of *beer* (x) and *extraversion* (z) over and beyond the effect of *type of bar* ($cat2$) and *type of beer* ($cat3$). Our full model involves all mentioned variables, plus the interaction between the factors (see Figure above). We now need to specify a smaller model, which includes only the categorical variables and their interaction. We set this model in the Nested Model field.



What we are saying to the module is to estimate a full model, then estimate a model with only the categorical variables, and compare the fit (R^2). The difference between the two R^2's is the variance uniquely explained by the terms present only in the full model, in our case x and z. The output is the following:

Model Fit

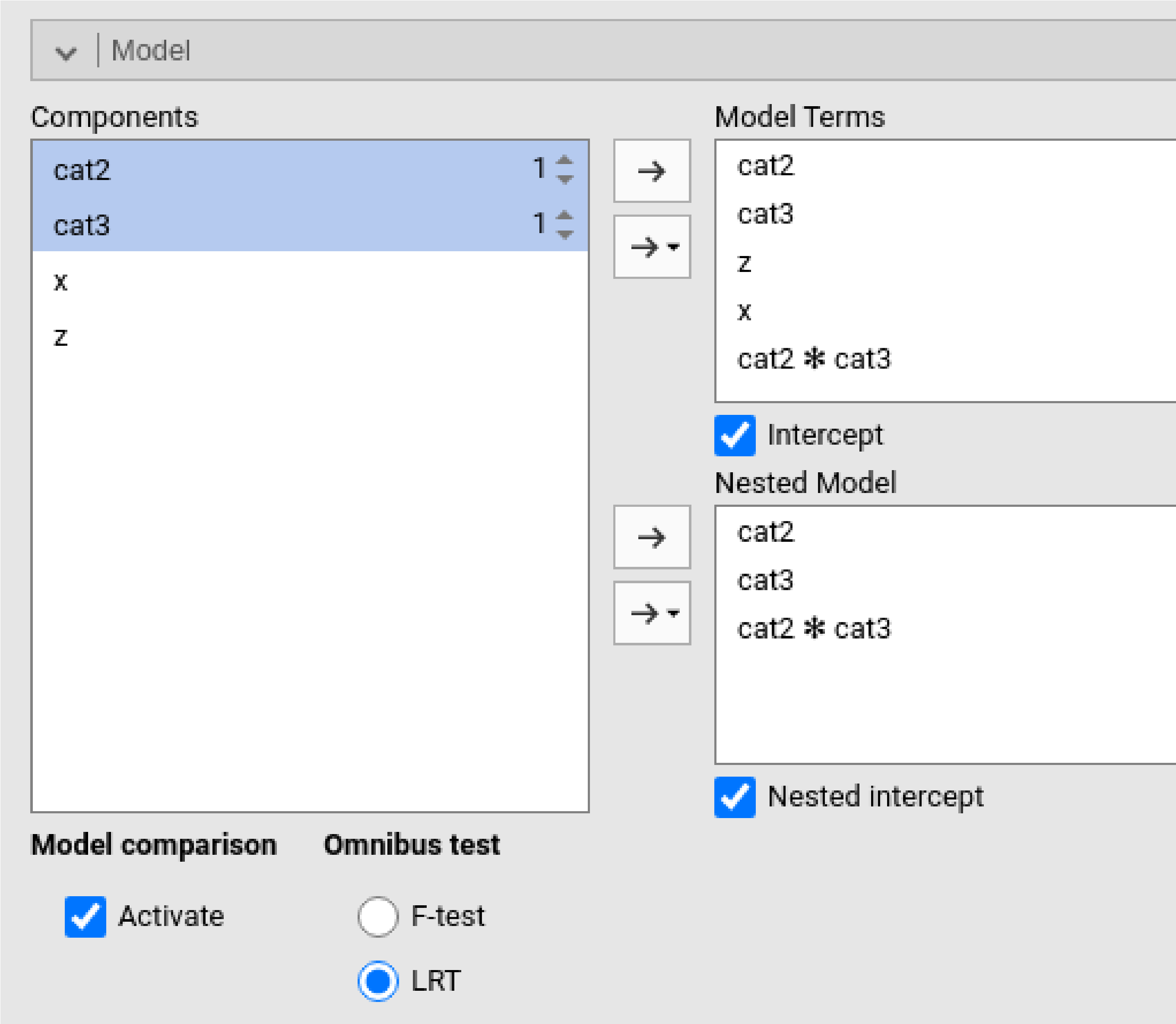
Model	R^2	Adj. R^2	df	df (res)	F	p
Full	0.406	0.369	7	112	10.95	< .001
Nested	0.162	0.125	5	114	4.40	0.001
ΔR^2	0.245	0.244	2	112	23.07	< .001

[4]

Thus, the full model explains $R^2 = .406$ (*100) of the variance of the dependent variable, the Nested model, without x and z explains $R^2 = .162$ (*100) of the variance. Their difference, $\Delta R^2 = .245$ is the variance uniquely due to x and z together.

2.14.2 Types of tests

In the General Linear Model, all tests are usually performed with F-test. The F-test works fine and has a solid tradition in statistics. Therefore, it is perfectly fine to use it also to compare models. After all, model-comparison entails comparing variances explained, which is the F-test job since more than 100 years. Recently, the statistical literature has made another test popular in model-comparison methods: The LRT, log-likelihood ratio test. This test is very useful for models estimated maximizing the log-likelihood of the data given the model, such as the generalized linear model or (with some caveats), the mixed model. For those passionate about the LRT, GAMLj offers the option to obtain the LRT also for the GLM, by selecting LRT under Omnibus test. In the GLM, the p-values based on the F-test and the LRT are usually very similar.



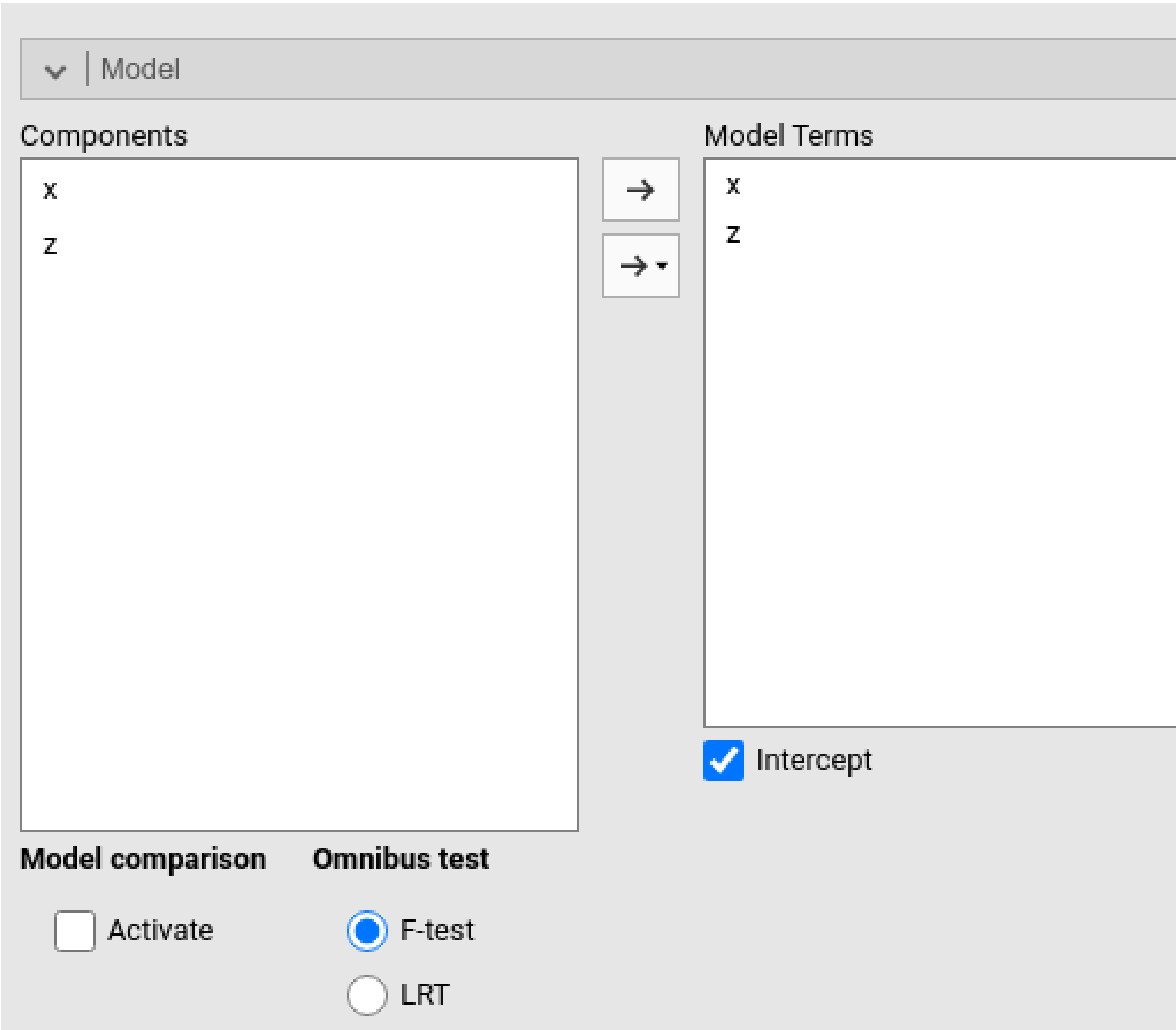
Model Fit

Model	R^2	Adj. R^2	df	LRT X^2	p
Full	0.406	0.369	7	62.6	< .001
Nested	0.162	0.125	5	21.2	< .001
ΔR^2	0.245	0.244	9	41.4	< .001

[4]

2.14.3 Not necessary model-comparisons

One should not be carried away by model-comparisons (the custom version of the method), because it is seldom useful and the *usual* estimates and tests of the linear model are already some sort of model-comparisons. Recall, for instance, the model with x and z as independent variables.



Model Results

Model Fit

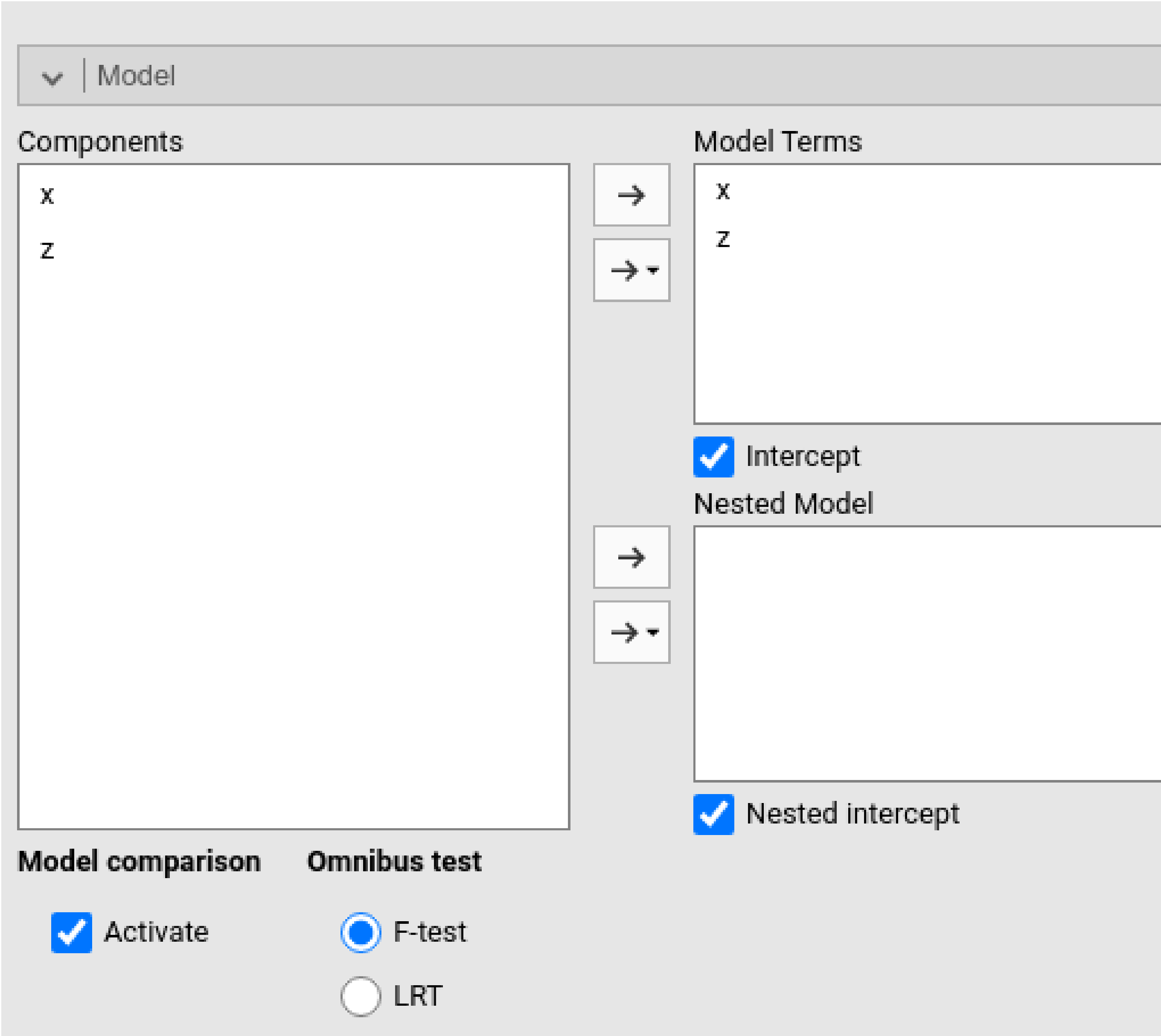
R^2	Adj. R^2	df	df (res)	F	p
0.346	0.335	2	117	31.0	< .001

[4]

ANOVA Omnibus tests

	SS	df	F	p	η^2	η^2p
Model	1849.387	2	30.996	< .001	0.3463	0.346
x	1696.744	1	56.876	< .001	0.3178	0.327
z	126.550	1	4.242	0.042	0.0237	0.035
Residuals	3490.408	117				
Total	5339.794	119				

Notice that we selected η^2 as an additional effect size index. First, if we make a model-comparison with an intercept only model, we get exactly the R^2 we obtained in the standard analysis. This is because the *usual* R^2 is already a fit comparison between our full model and an intercept-only model (cf appendix A).



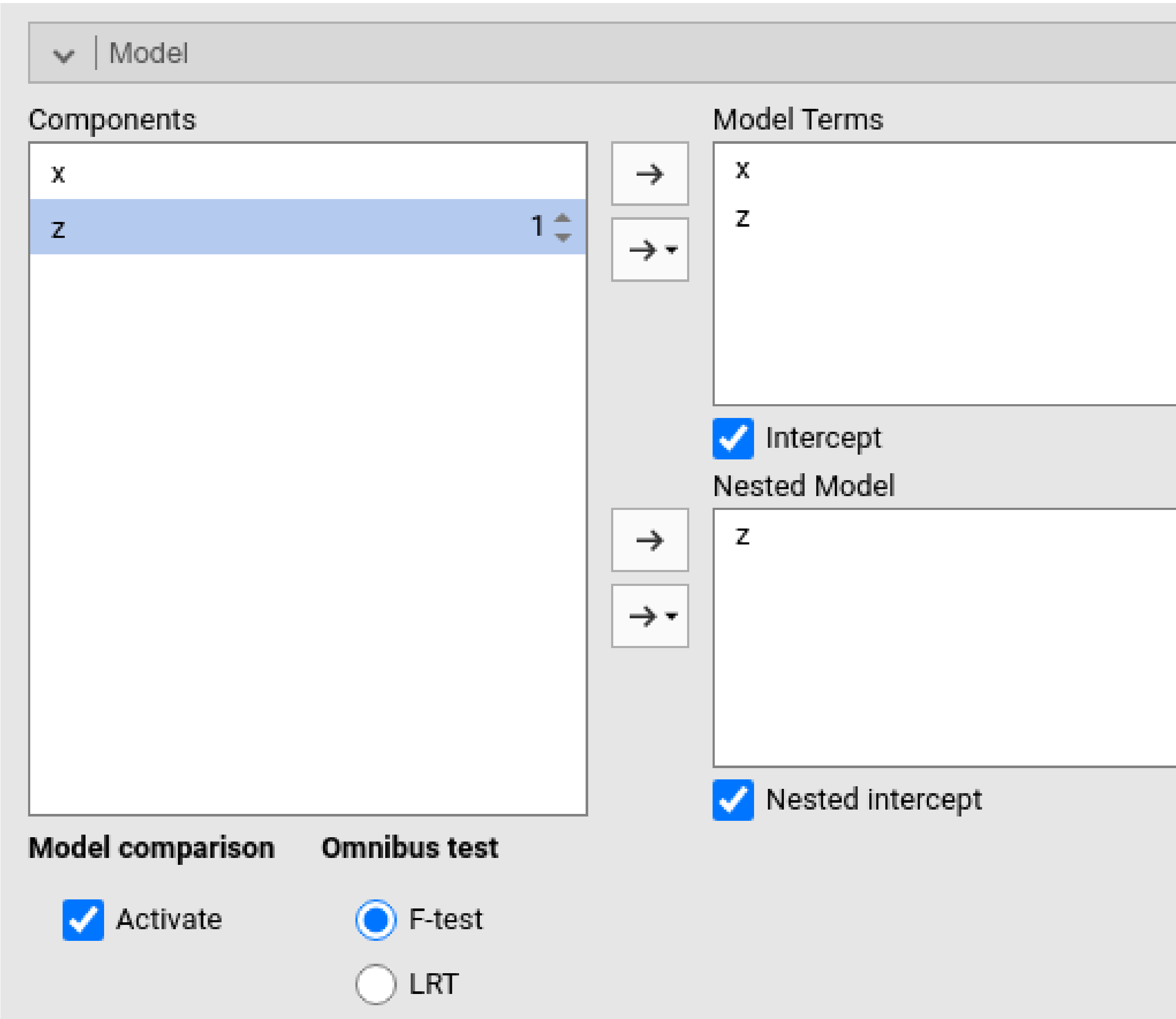
Model Fit

Model	R^2	Adj. R^2	df	df (res)	F	p
Full	0.346	0.335	2	117	31.0	< .001
Nested	0.000	0.000	.	.	.	.
ΔR^2	0.346	0.335	2	117	31.0	< .001

[4]

Second, if we now add to the nested model the variable z, we are comparing a model with x and

z as terms with a model with only z, so we are estimating the contribution of x to the explained variance.



Model Results

Model Fit

Model	R²	Adj. R²	df	df (res)	F	p
Full	0.3463	0.3352	2	117	31.00	< .001
Nested	0.0286	0.0204	1	118	3.47	0.065
ΔR²	0.3178	0.3148	1	117	56.88	< .001

[4]

ANOVA Omnibus tests

	SS	df	F	p	η²	η²p
Model	1849.387	2	30.996	< .001	0.3463	0.346
x	1696.744	1	56.876	< .001	0.3178	0.327
z	126.550	1	4.242	0.042	0.0237	0.035
Residuals	3490.408	117				
Total	5339.794	119				

The ΔR^2 is now .317, with `F(1,117)=56.68`, which is exactly the η^2 of the x in the full model, with the same F, df, and p-value. The effect of x is already a result of a model-comparison test, so we do not need to test it explicitly.

2.14.4 Hierarchical regression

The model-comparison approach allows estimating what many people call *hierarchical regression*. Hierarchical regression is an analytic strategy, it is not a statistical model. By *hierarchical regression* one means the estimation of the coefficients of different IV Independent Variable in different models, containing different sets of IV Independent Variable . For instance, one may want to estimate the effects of $cat2$ and $cat3$ independently of x and z, but the effects of x and z keeping constant $cat2$ and $cat3$. In a hierachical regression software, one specify a first block with $cat2$ and $cat3$, and a second block with all variables. The software would estimate two models, and produce a recap table with the coefficients obtained in the first model for $cat2$ and $cat3$, and the results obtained in the second model for x and z. Possibly, the ΔR^2 is also produced.

This analytic strategy is rarely useful, but if this is really the intent of the analyst, one can use GAMLj to estimate two different models to obtain the coefficients, and then obtain the ΔR^2 with a model-comparison approach. Alternatively, one can use jamovi `regression` command, which allows to specify blocks. Results will be identical, only the tables will be organized in different ways.

2.15 Assumptions checks

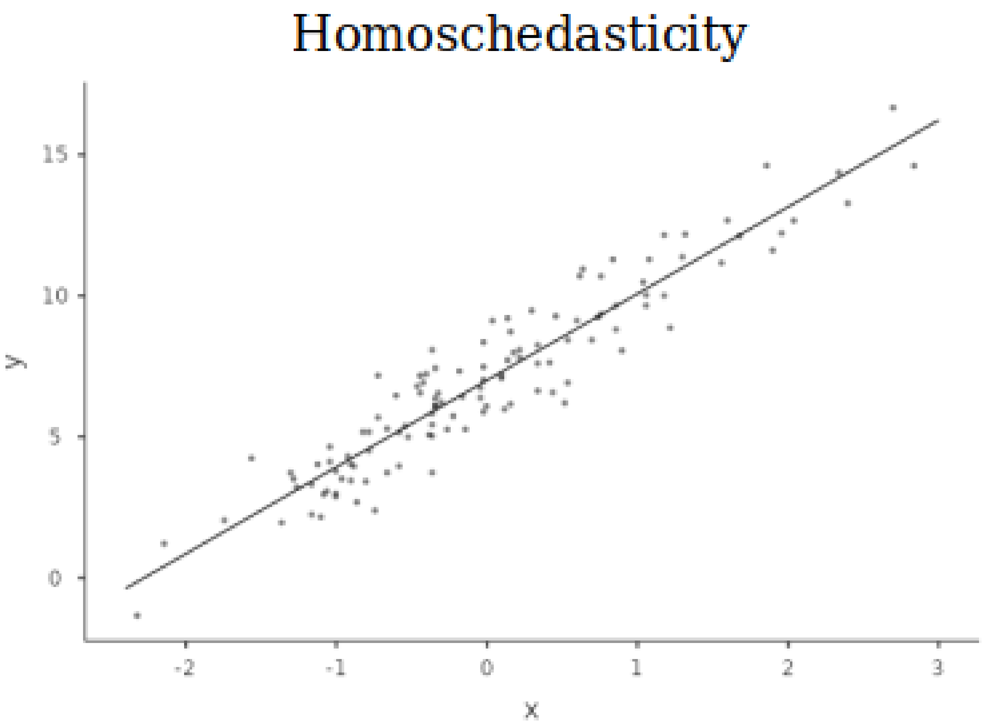
The GLM is based on several assumptions, a few of which are popular and they are regularly tested or at least checked. We should start saying that assumptions are idealized scenarios in which data show a required property. They are needed to unsure that the expected results (F-test, p-values, etc.) posses the expected properties (unbiasedness, consistency, etc). In other words, the assumptions are required so we can trust the results, cf. Nimon (2012) and Glass et al. (1972) for details.

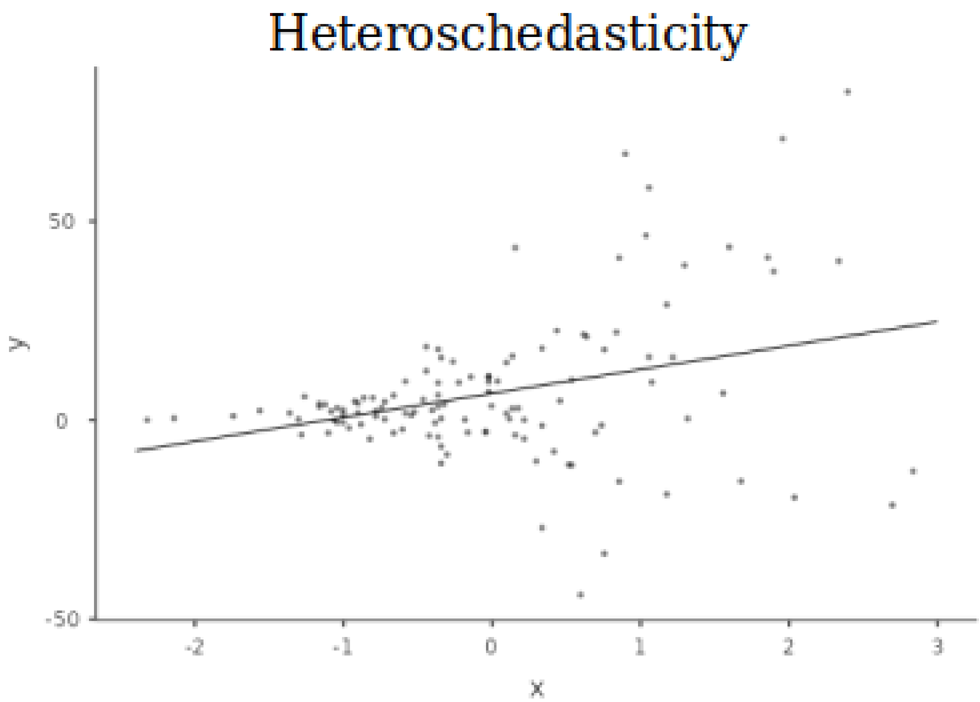
Being idealized scenarios, the observed data are never perfectly abiding by the assumptions, but they approximate the required property with different degrees. The better a property is approximated, the more we can trust the results. The worse is approximated, the more doubts we should cast on our results. Because the properties required by the assumptions are only approximated to a certain degree, assumptions cannot be evaluated only using an inferential test. Inferential tests tend to be interpret as "significant" vs "not significant", and such a dichotomy is not always useful when evaluating assumptions. For this reason, GAMLj provides both inferential tests and graphical methods to assess the appropriateness of the data with respect of the assumptions.

Here we focus on the homoschedasticity and normality of residuals assumptions.

2.15.1 Homoschedasticity

By *Homoschedasticity* we mean that the variance of the residuals is constant around the predicted values, that this the residuals spread around the predicted values at more or less the same distance along the model predictions. In other words, the spread of the clouds of points representing the DV Dependent Variable as a function of the IV Independent Variable is constant. When the spread is not constant, we have *Heteroschedasticity*. Here are two exemplifications:

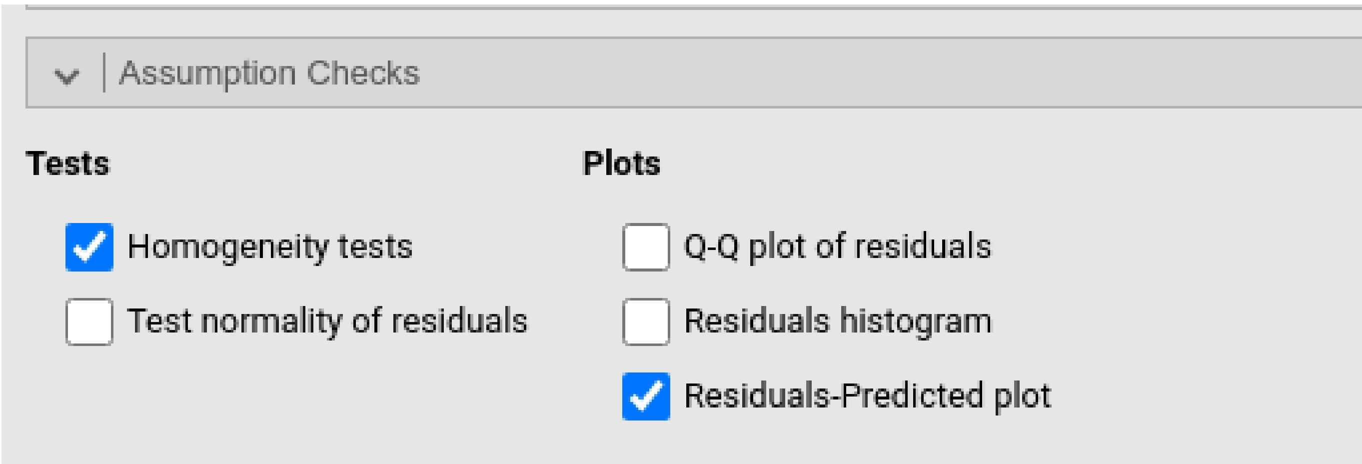




In cases where the independent variables are categorical, the assumption requires that the variances within groups are more or less the same across groups.

The idea is that the error term, which is the residuals variance, should be representative of the residual variance across all values of the predicted values. This is *Homoschedasticity*. On the contrary, the error term is not representative of the whole model if the variance of the resifuals is different for different predicted values, because it would be larger or smaller in different parts of the model. That is *Heteroschedasticity*. *Homoschedasticity* ensures that the standard errors associated with the estimates are correct, and thus are the inferential tests and the p-values one obtains along the coefficients estimates.

To evaluate the *Homoschedasticity* we can use two inferential tests and one graphical method.



The tests are the Breusch-Pagan test and the Levene's test.

Assumption Checks

Test for Homogeneity of Residual Variance

Test	Statistics	df1	df2	p
Breusch-Pagan Test	12.12	7		0.097
Levene's Test	4.25	5	114	0.001

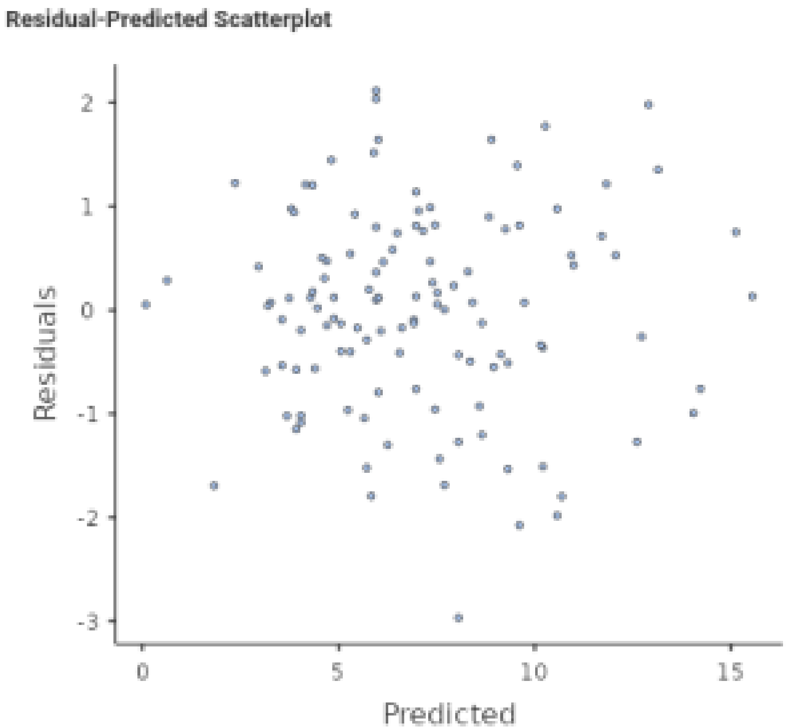
Note. Levene's test is done only for factors.

Both tests test the null-hypothesis that the variance of the residuals do not change along the predicted values, so the assumption is met when they are not significant. A significant test, on the contrary, indicates that the residuals variance departs from the assumption of homogeneity, and thus we should cast some doubt on the validity of the results.

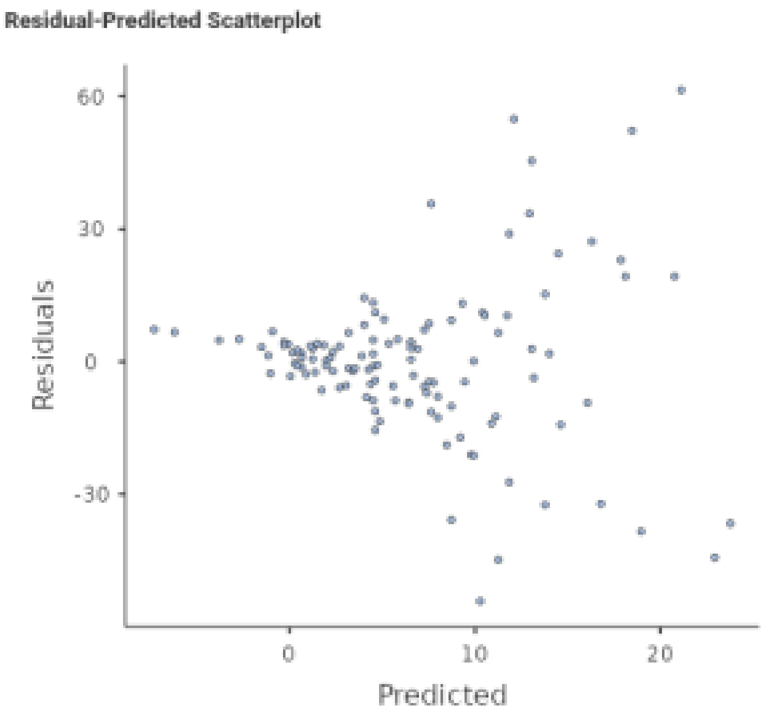
The Breusch-Pagan is defined for any GLM, so it is estimated whatever our IVs Independent Variables are. The Levene's test is defined only for categorical IVs Independent Variables , so it is not produced when the IVs Independent Variables are all continuous.

Graphically, we can check the Residuals-Predicted plot. This plot depicts the residuals (in the Y-axis) as a function of the predicted (on the X-axis), so it makes it easy to see whether the spread of the cloud of points changes along the X-axis. The assumption is approximated well when the depicted cloud of points has more or less the same spread along the whole plot. Examples may be:

Homoschedasticity

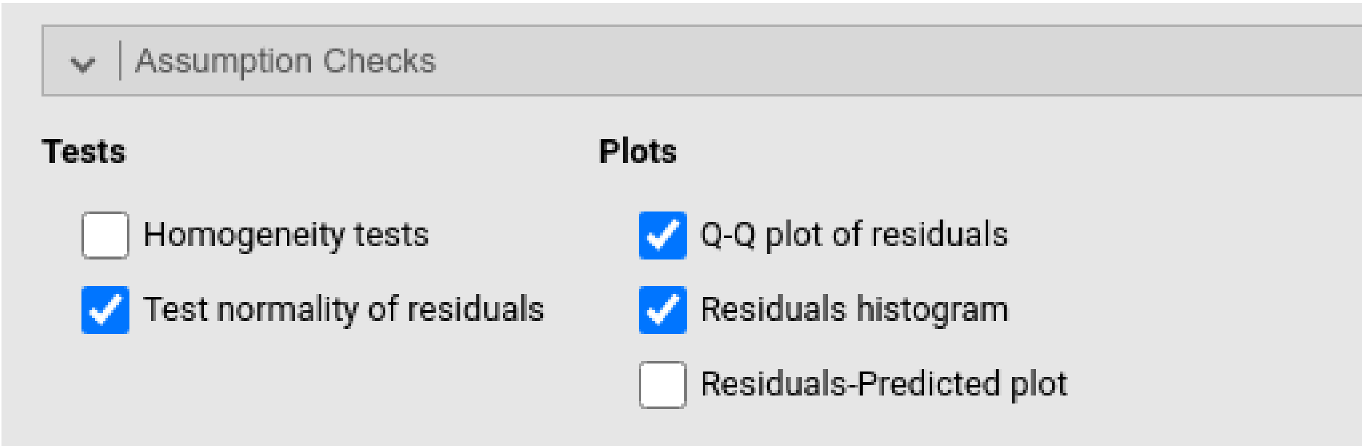


Heteroschedasticity



2.15.2 Normality of residuals

To guarantee unbiased inferential tests, that is valid p-values, residuals should be normally distributed (Gaussian). GAMLj provides the Kolmogorov-Smirnov test, the Shapiro-Wilk test, the histogram plot and the Q-Q plot to assess this assumption.



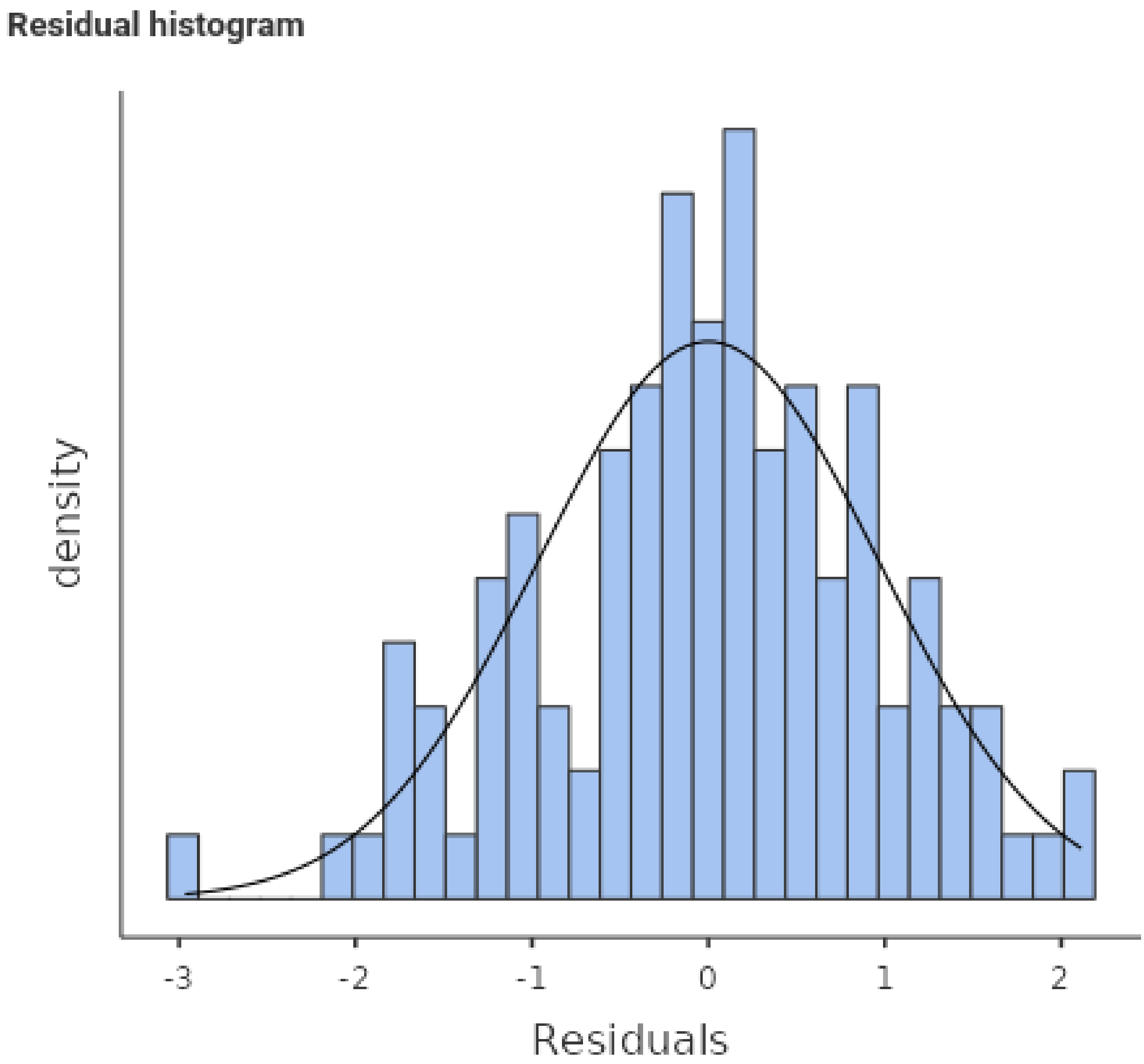
Assumption Checks

Test for Normality of residuals

Test	Statistics	p
Kolmogorov-Smirnov	0.0586	0.804
Shapiro-Wilk	0.9912	0.645

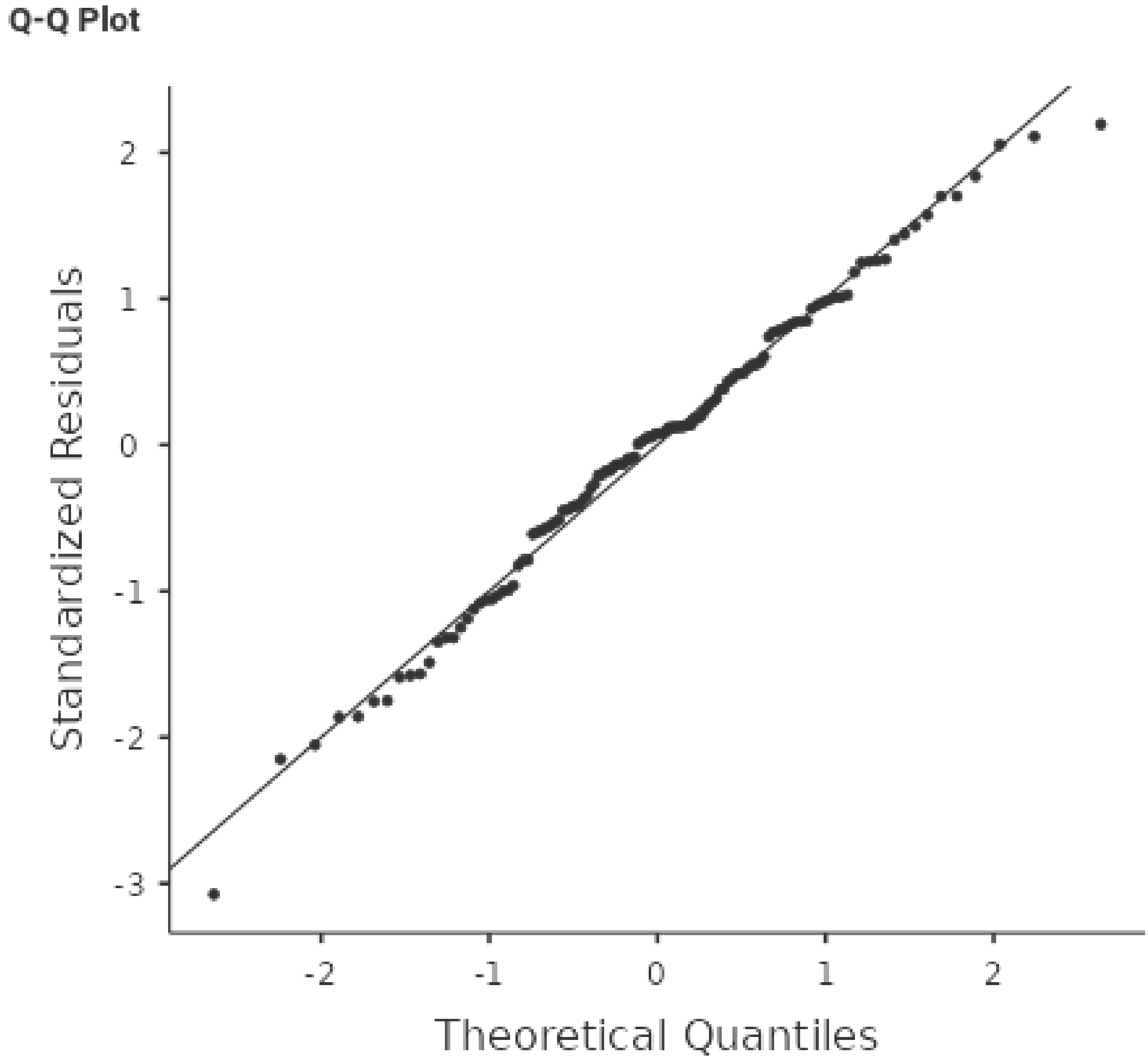
Both inferential tests share the null-hypothesis that the distribution is normal (Gaussian), thus a non-significant test indicates lack of evidence against the assumption, whereas a significant test indicates departure from the assumptions. Shapiro-Wilk seems more powerful, so it should be preferred in small samples, whereas the Kolmogorov-Smirnov may be more appropriate in large samples.

The first graphical method is a simple histogram of the residuals, on which the module overlays a perfectly normal curve with the same mean and standard deviation of the observed one, so the comparison becomes easier.



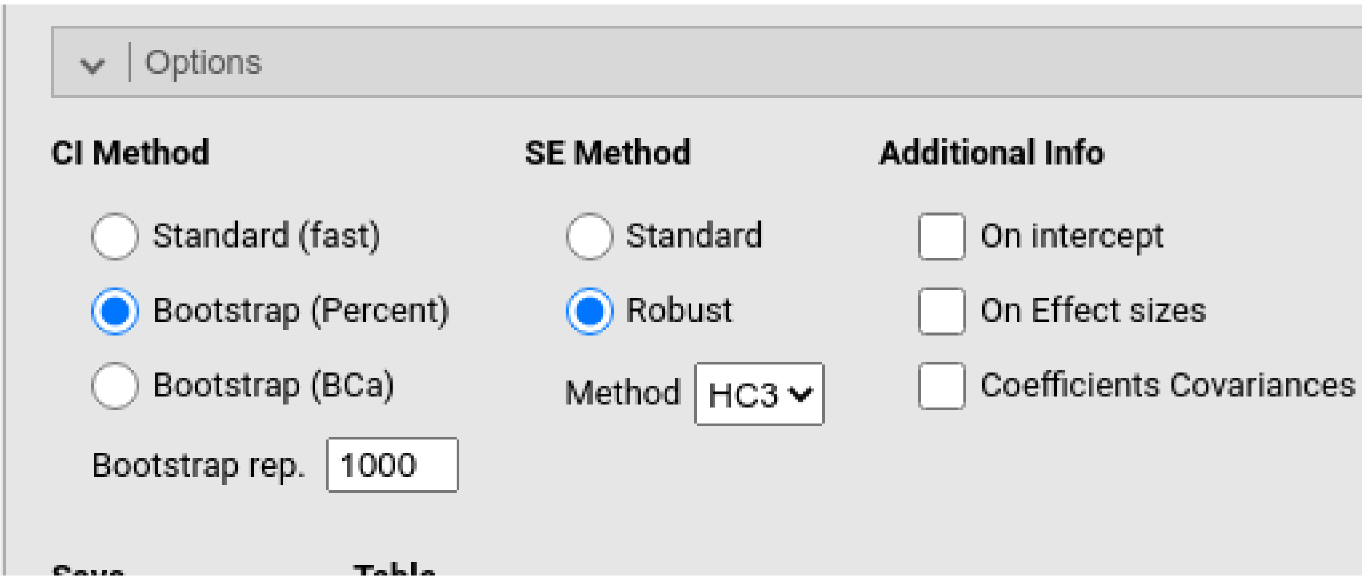
Here we want to check that our distribution is not too far away from a normal distribution, especially with regards of the main properties of the normal one: symmetry and increasing density (frequency of cases) closer to the mean.

Another popular graphical method is the Q-Q plot.



The Q-Q plot plots the theoretical quantiles (percentiles) of a perfectly normally distributed variable with the quantiles of our observed distribution. The closer our distribution to the normal one, the closer the scattered points will be to the 45 degrees line.

2.16 Violations Remedies



2.16.1 Robust Standard Error

Whereas it is often the case that both assumptions are violated, the violation of *homoschedasticity* and the violation of *normality of the residuals* have different remedies.

Lack of *homoschedasticity*, that is *heteroschedasticity*, can be counteracted by using a robust method to compute the standard errors. Whereas *robust* estimations is a very general term, in the context of the GLM it usually means *robust against heteroschedasticity*. GAMLj offers *robust* standard error in the Options panel, under SE method. When `Robust` is selected, one can choose the algorithm to compute the HC (**H**eteschadasticity-**C**onsistent) standard errors. GAMLj implements these algorithms as implemented in the sandwich R package, setting `HC3` as default as recommended by the package authors.

Setting the SE method to robust updates all results related with inferential tests, and thus the results will be different, and more accurate, proportionally to the strength of the assumption violation.

2.16.2 Bootstrap Confidence Intervals

When the residuals are not normally distributed, or when other assumptions are suspicious, one can rely on Bootstrap Confidence Intervals. The advantage of the bootstrap C.I. is that they do not assume any shape for the residual distribution, because their boundaries are estimated by resampling the observed distribution (cf. Bootstrapping, and Efron and Tibshirani (1994)).

GAMLj offers two methods to compute bootstrap confidence intervals. The `percent` and the `BCa` method. The `percent` method entails building a bootstrap distribution of estimates and select the $100 \cdot (\alpha/2)$th and the $100 \cdot (1-\alpha/2)$th percentile of the distribution as the boundary of the interval. In other words, for the 95% C.I (where $\alpha = .05$), it selects the 2.5th and 97.5th percentile of the bootstrap distribution. The `BCa` method stands for *Bias corrected and accelerated* method. The method corrects the percentile values depending on the skewness of the bootstrap estimates distribution and the offset of the distribution mean compared with the observed estimate. The specialized literature seems to indicate that for generic situations, the *percent* method gives more accurate results (cf., for instance, Jung et al. (2019)).

The number of bootstrap resampling is set to 1000, but for reliable and replicable results one can set it to 5000 or 10000. Just keep in mind that resampling means to estimate a full model for every bootstrap sample, so the process might be quite time-consuming for large models.

Chapter 3

The Generalized Linear Model

Draft version, mistakes may be around

KEYWORDS: Generalized Linear Model, Logistic Regression, logit

3.1 Introduction

In the general linear model (cf. 2) the dependent variable must be a quantitative variable: It should express quantities, such that its values can retain all properties of the numbers: 1 is less than 2, 4 is 2*2. In this way, the mean, the variances and all the estimates can make sense. We have also seen that the GLM assumes the residuals being normally distributed, which is akin to saying that the dependent variable is normally distribute (a part from the IVs Independent Variables influence). There are many research designs that require the dependent variable to be not-normally distributed or not even quantitative. One may want to predict the vote in a referendum, the number of smartphone owned by individuals, the sex of the offspring of a herd, a choice between two or more product colors in a marketing study. In all these circumstances, the GLM cannot be applied.

We still want to use a linear model, though, because we know very well how to estimate it and interpret it. So, instead of forgoing the linear model, we change the way the linear model predicts the dependent variable, such that the estimates are unbiased and reasonable even when the dependent variable values are not quantities, or they are clearly not normally distributed.

Here comes the *generalized linear model.* Consider the standard regression model.

$$\hat{y_i} = a + bx_i$$

The $\hat{y_i}$ are the predicted values, meaning the points lying on the regression line, that correspond to the expected (average) values of y for any possible value of x. In the GLM, the predicted values have the same scale of the observed values: This is because $\hat{y_i}$ can take any value (the straight line is defined for any possible value in the Y-axis), and so can y_i, the observed values. When the variable is not quantitative, or it has a weird distribution, we cannot be sure that the predicted values will make sense. If one is predicting a probability, for instance, the observed values vary from 0 to 1, and thus the predicted values of a GLM will surely be nonsense, because the line will overcome the 0-1 boundaries and predicts probabilities of, say, 10 or -30, that are not admissible. If the predicted values make no sense, the model make no sense.

If we, however, express the predicted values as a transformation of the dependent variable, we can choose the right transformation to be sure that the predicted values will make sense. The transformation is called the *link function* ($f()$), and one can pick different link functions to accommodate different types of dependent variables. The *generalized linear model* is a linear model in which the predicted values are expressed as a transformation of the dependent variable:

$$f(\hat{y}_i) = a + bx_i$$

In addition to predicting a transformation of the dependent variable, the generalized linear model does not assume the dependent variable to be normally distributed, but allows assuming different families of distribution: Binomial, multinomial, Poisson, etc.

Combining a particular link function with a distribution makes a particular application of the generalized linear model. The combination of link function and distribution makes a particular application a model suitable for predicting a particular type of dependent variable. Here is a brief recap of the generalized linear models that can be estimated with GAMLj.

Model

Link Function

Distribution

DV type

Logistic Model

Logit

Binomial

Dichotomous

Probit Model

Inverse cumulative normal

Binomial

Dichotomous

Multinomial Model

Logit

Multinomial

Categorical

Ordinal Model

Cumulative Logit

Logit

Ordinal

Poisson Model

Log

Poisson

Count (frequencies)

Negative Binomial Model

Log

Negative binomial

Count (frequencies)

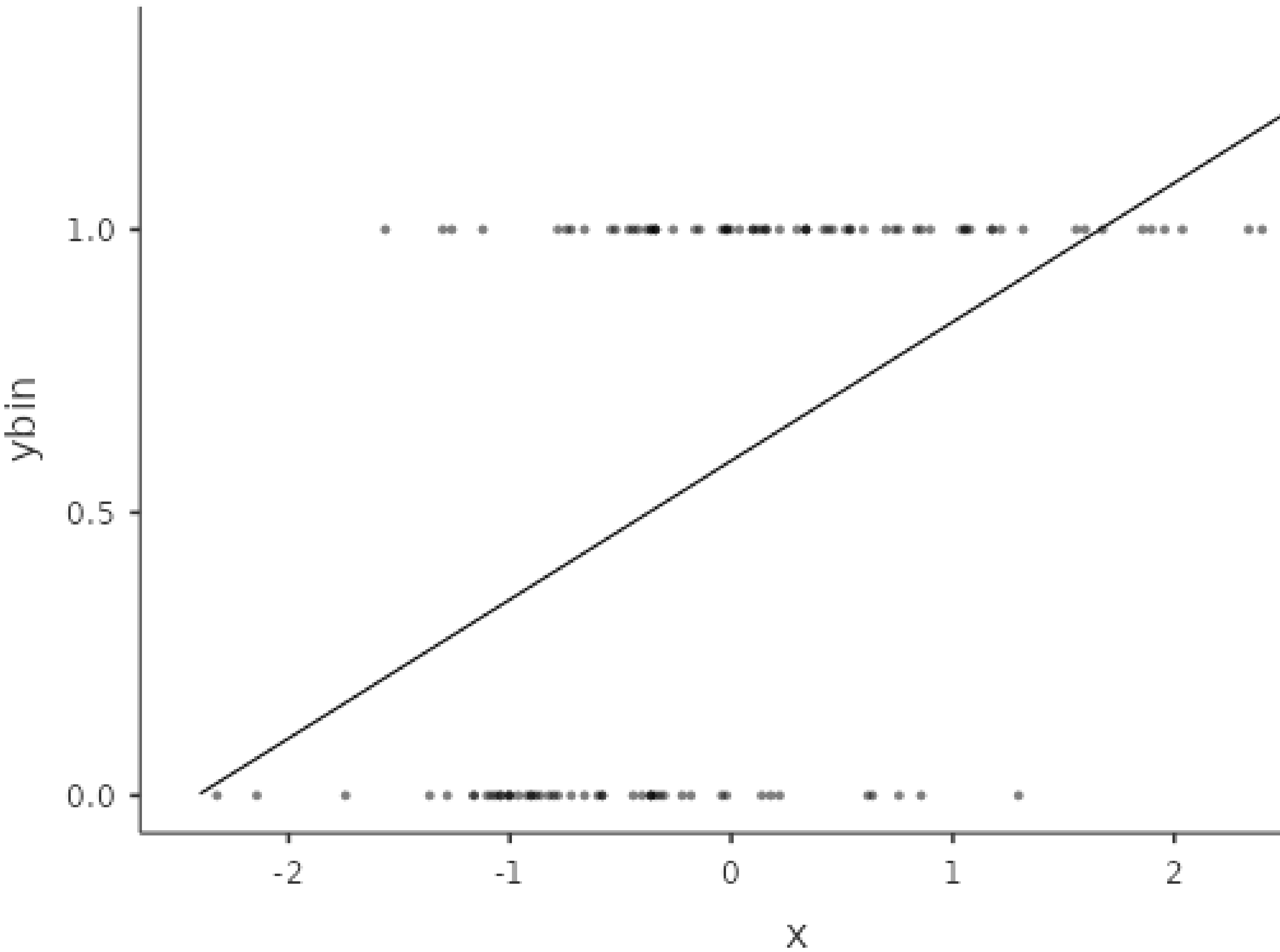
We are going to explore all these models, highlighting their specificity but keeping in mind that all techiques and methods doable to the GLM (cf. 2) can be applied also to the models within the generalized linear model.

As a general reference for the material discussed in this chapter, the book Agresti (2012) is a great source of information and a precise guide to the statistical details.

3.2 Logistic Model

3.2.1 The rationale

A logistic model can be estimated when the dependent variable features two groups, or two levels. The necessity to change the GLM into a different linear model becomes clear by inspecting a scatterplot between a continuous independent variable and a dichotomous dependent variable.



It is clear that the dependent variable scores can only be 0 or 1, and that the scatterplot will always present two horizontal stripes points, forming a cloud that cannot be represented well by a straight line. A straight line that will feature predicted values surely above 1 and below 0, providing nonsensical predicted values. The residuals, furthermore, will depend on the predicted values (cf heteroschedasticity in 2.15.1), because the line will cross a strip once (so residual is zero) and depart from it everafter (increasing the residual variance).

The solution is to change the form of predicted values such that any values can be a sensible predicted value. To achieve this, we should first notice that when one has a dichotomous dependent variable, what one is predicting is the probability of being in the group scoring 1. Indeed, the predicted values are the expected (average) values of y for each given x_i. The average value of a dichotomous variable is

$$E(y) = \frac{n_1}{N}$$

where n_1 is the number of cases in group 1, and N is the total sample. $E(y)$, however, is the probability of being in group 1. Thus, $p(x = 1) = E(y)$, which we simply call p.

So, the aim of the logistic model is to estimate how the probability of being in group 1 rather than 0 depends on the IVs Independent Variables scores. The probability scale, varying from 0 to 1, is not suitable to be fit by a straight line. We can change this by predicting the *odd* of the probability, namely:

$$odd = \frac{p}{1-p}$$

The odd frees us from the upper boundary of 1, because any positive value expressed in odd can be sensible predicted value. For each positive value, one can always transform it back and get back a probability. The problems are negative values, that a straight line will always yield. Since the odd cannot be negative, we need to apply to it another transformation, namely the logarithm transformation. A logarithmic transformation transforms a positively-valued variable into a variable with all possible values, positive and negative. The function that expresses a probability into a variable admitting all possible values is the logit function:

$$logit(y) = log\left(\frac{p}{1-p}\right)$$

The logistic model is a linear model predicted the logit

$$logit(\hat{y}) = a + b_1x_1 + b_2x_2 + ...$$

Everything we can do with a linear model can be done with the logistic model, we just need to keep in mind that the interpretation of the results must consider the fact that the predicted values have a logistic scale, and not the original dependent variable scale.

3.2.2 Model Estimation

We can use our *manymodels* dataset to see an example of a logistic model. The dataset contains a dichotomous dependent variables called `ybin`, which features two groups. To keep up with our cover story, we can imaging it to represent visiting the toilet behavior. $ybin = 1$ means that the customer has visited the bar restroom, $ybin = 0$ means that the client has not visited the restroom that evening.

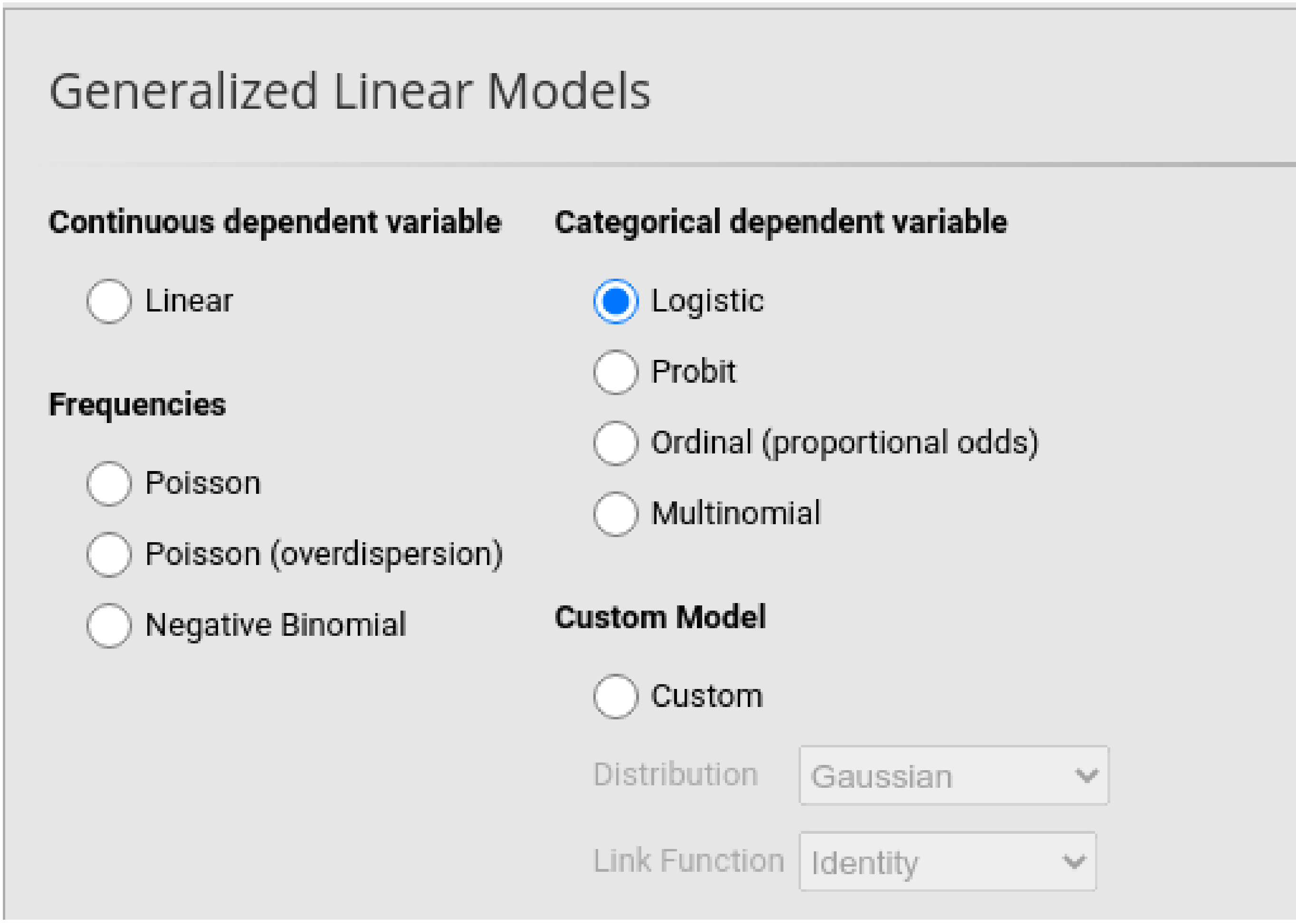
Frequencies

Frequencies of ybin

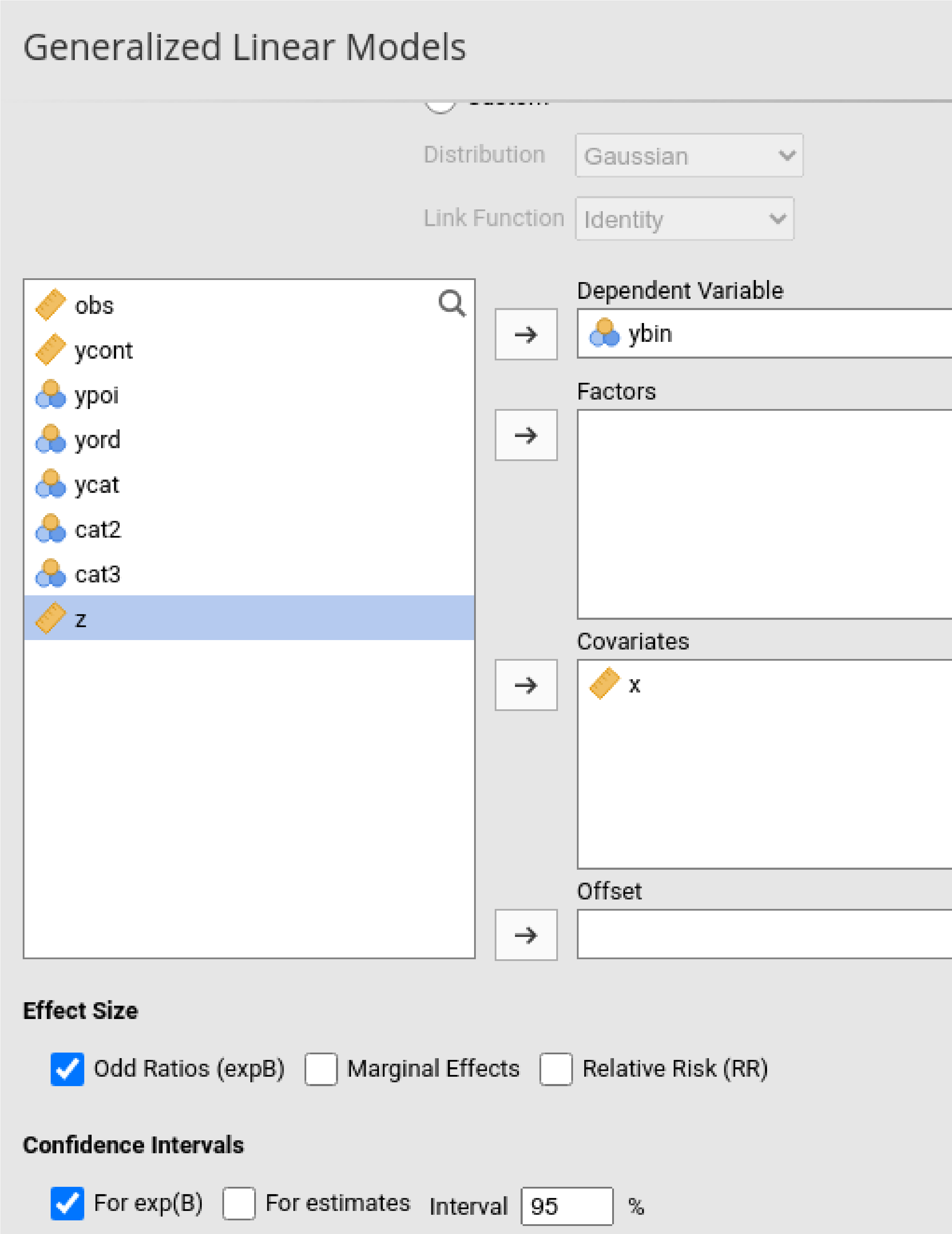
ybin	Counts	% of Total	Cumulative %
0	49	40.8 %	40.8 %
1	71	59.2 %	100.0 %

The aim of the model is to estimate the relationship between number of beers (x) and the probability of visiting the restroom ($ybin$).

In GAMLj we launch `Generalized Linear Models` menu of the `Linear Models` icon. The first part of the user interface allows selecting the appropriated model. In our case, we selected Logistic because our dependent variable is a dichotomous variable.



Once we have selected the model, we can set up the variables in the variables role fields, as we did in the GLM (cf. 2.2.1).


Generalized Linear Models
Distribution
Gaussian
Link Function
Identity
obs
ycont
ypoi
yord
ycat
cat2
cat3
z
Dependent Variable
ybin
Factors
Covariates
x
Offset
Effect Size
Odd Ratios (expB)
Marginal Effects
Relative Risk (RR)
Confidence Intervals
For exp(B)
For estimates
Interval
95
%

3.2.3 Model Recap

Generalized Linear Models

Model Info

Info		
Model Type	Logistic Model	Model for binary y
Model	stats::glm	ybin ~ 1 + x
Distribution	Binomial	Dichotomous event distribution of y
Link function	Logit	Log of the odd of y=1 over y=0
Direction	P(y=1)/P(y=0)	P(ybin = 1) / P(ybin = 0)
Sample size	120	
Converged	yes	
C.I. method	Wald	

[3]

In the Model Info table we find a set of properties of the estimated model. The most important one to check is described in the `direction` row. When the dependent variable is dichotomous, in fact, the direction of the probability is arbitrary, so we need to know which group is predicted. GAMLj models the probability of being in the group with the “largest” label value, after ordering the value labels. In our case, it models the probability of being in group `ybin=1` over the probability of being in group `ybin=0`. This is indicated in the `direction` row of the table.

3.2.4 Model Fit

Model Fit

R^2	Adj. R^2	df	X^2	p
0.224	0.199	1	36.4	< .001

[4]

Additional indices

Info	Model Value	Comment
LogLikelihood	−62.97	
AIC	129.94	Less is better
BIC	135.51	Less is better
Deviance	125.94	Less is better
Residual DF	118	
Chi-squared/DF	1.00	Overdispersion indicator

First, the model R-squared is produced with its inferential test, in this case a Chi-square test. This provides a test of the ability of the model to predict the dependent variable better than chances. The R-squared is the McFadden's pseudo R squared (cf GAMLj help). We can interpret it as the proportion of reduction of error, or the proportion of increased accuracy in predicting the dependent variable using our model as compared with a model without independent variables (cf. Appendix A). The adjusted R^2 is the population R^2 estimate.

The additional indices reported in Additional indices reports other indices useful for model comparisons or evaluation of models, especially for other types of generalized linear models. They are rather technical, and we'll not discuss them here.

3.2.5 Omnibus Tests

As for the GLM, we have omnibus tests for the effects of our IVs Independent Variables . They are expecially useful when dealing with categorical independent variables, because with continuous independent variables they are redundant as compare with the coefficients tests. With one IV Independent Variable , the omnibus test is equivalent to the model fit test.

Omnibus tests

	X^2	df	p
x	36.4	1	< .001

3.2.6 Coefficients

Parameter Estimates (Coefficients)

				Exp(B) 95% Confidence Intervals		
Name	Estimate	SE	Exp(B)	Lower	Upper	z
(Intercept)	0.599	0.233	1.82	1.15	2.88	2.
x	1.531	0.324	4.62	2.45	8.72	4.

With continuous IVs Independent Variables , the coefficients `Estimates` are the most interesting parameters. The `Estimate` column reports the regression coefficients. Their interpretation should be based on the fact that the predicted values scale is the logit scale. Thus, as regard the effect of x, we can say that for one unit more in x, the logit of the probability of being in the group $ybin = 1$ increases of one 1.53 units. In our cover story, for one more beer the logit of visiting the restroom increases of one unit. Being positive, we can conclude that the more you drink, the higher the probability of visiting the restroom. Being statistically significant (z-test=4.73, p.<.001), we can conclude tha the effect is different from zero.

3.2.7 Odd ratios exp(B)

The issue here is that it is very difficult to fathom the practical size of the effect. Is 1.53 units increase in the logit scale a big or small increase? Honestly, nobody knows, because the logarithm scale is difficult to master, and even if one could, the readers of the results would probably not. So, we interpret the `exp(B)` parameter, which is the logit after removing the logarithm scale. The logarithm scale is removed by simply passing the logit to the exponential function, hence the notation `exp(B)`. If we remove the logarithm scale, we are left with the odd scale. However, we should pay attention to what happens to the coefficients when the scale is changed from the logit to the odd. Two pieces of information are important here. First, recall that the b coefficient in a linear model (any linear model) is the difference in the predicted values for two consecutive values of the independent variable. That is

$$b = \hat{y}_{(x=1)} - \hat{y}_{(x=0)}$$

In the logistic model, we have

$$b = log(odd_{(x=1)}) - log(odd_{(x=0)})$$

The second piece of information is that when you take the exponential function of a difference between two logarithms, the result is the ratio between the logarithms arguments. That is

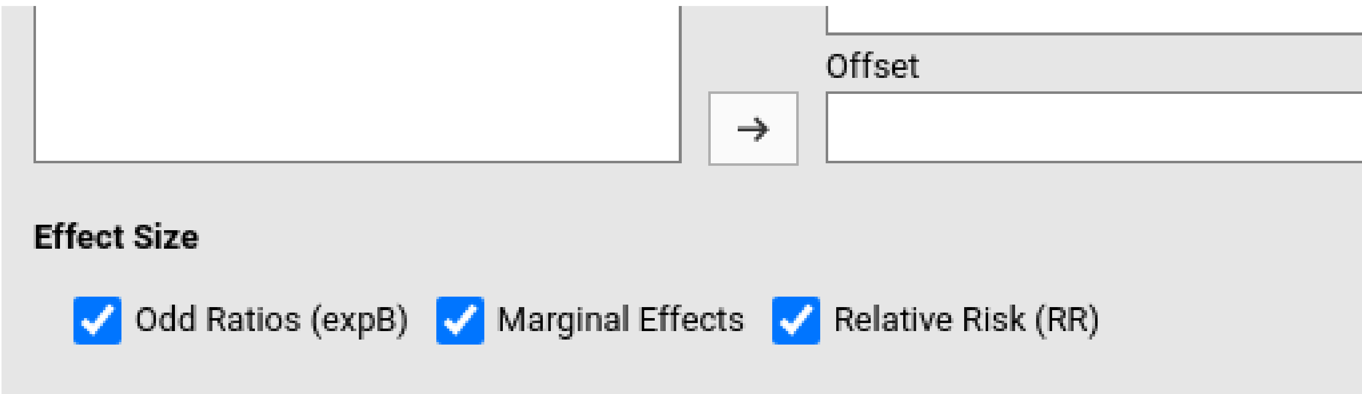
$$exp(log(a) - log(b)) = \frac{a}{b}$$

Thus, if we take the exponential function of the logit B, we obtain the ratio between two odds

$$exp(b) = \frac{odd_{(x=1)}}{odd_{(x=0)}}$$

Therefore, `exp(B)` is a ratio between two odds, computed at two consecutive values of the independent variable. That is why it is called the *odd ratio*. It is *the rate of change of the odd as you increase the independent variable of one unit*. In other words, it indicates how many times the odd changes as one increases the exponential function of one unit. In our example `exp(B)=4.62`, so, for every beer more, the odd of going to the restroom increases of 4.62 times.

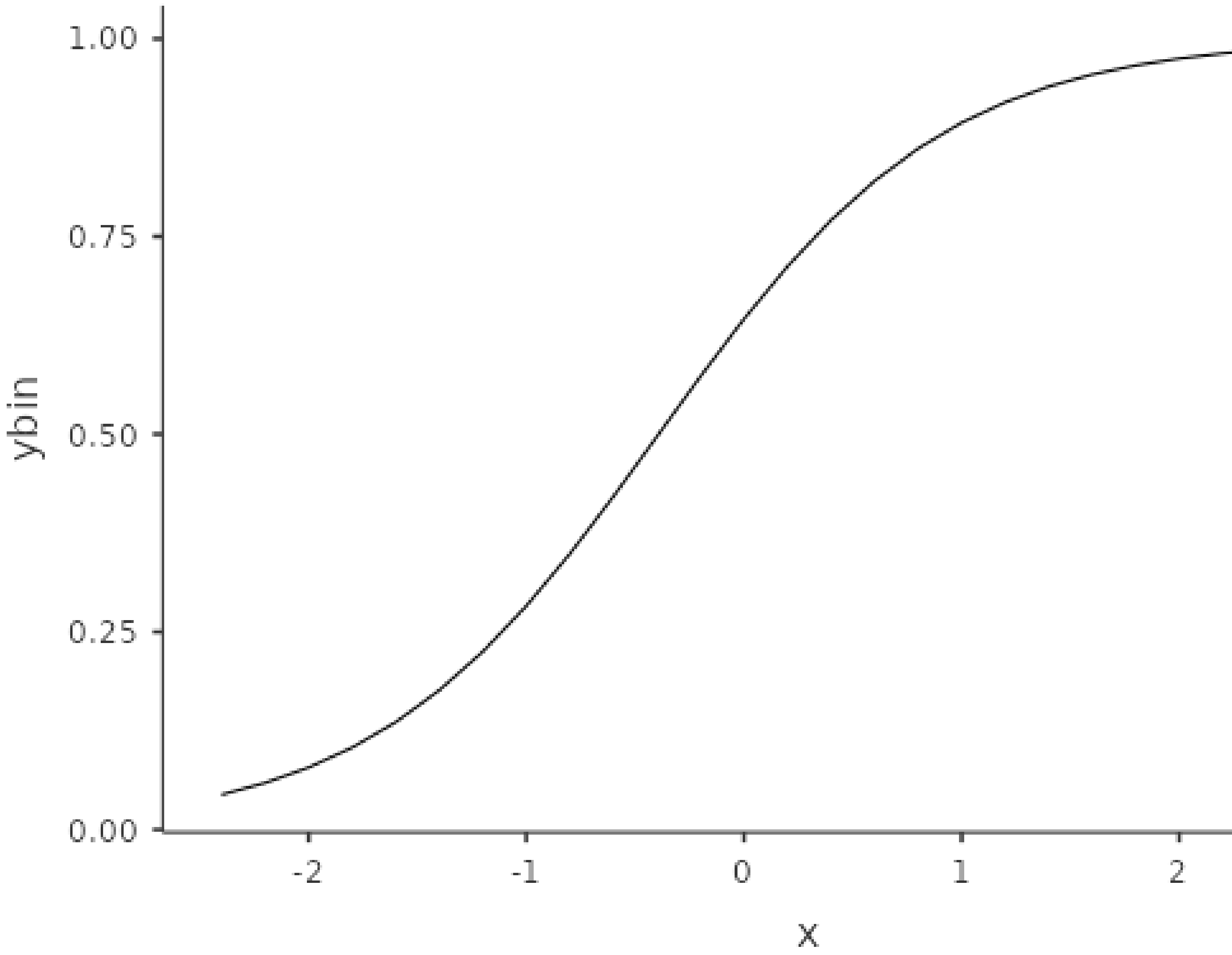
The odd ratio is the standard effect size used in logistic regression, but it is not the only one. In different disciplines other ways to quantify the logistic effects are used. GAMLj provides the most common ones: *Marginal effects* and *Relative Risk*. Now we discuss the former because it is more appropriate with continuous IVs Independent Variables . We discuss the RR when we deal with categorical IV Independent Variable (cf 3.3)



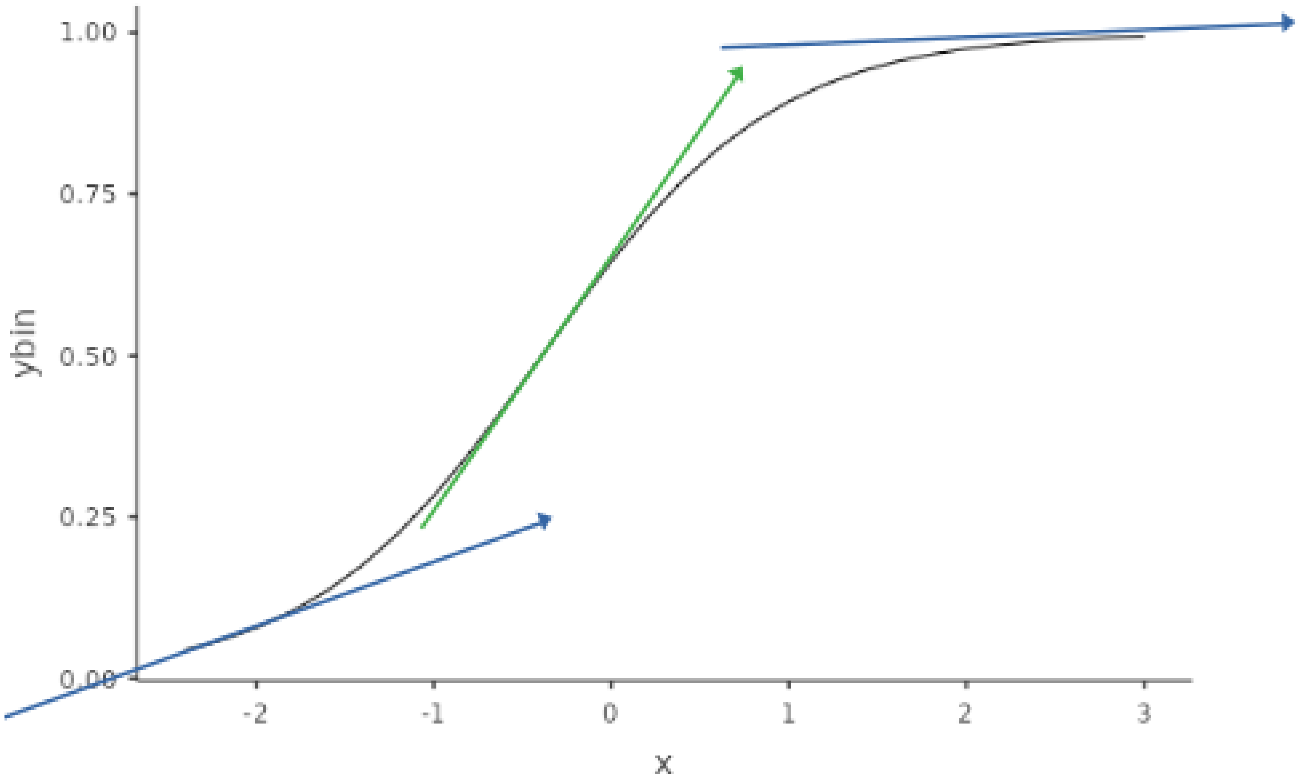
3.2.8 Marginal effects

In the same way that the `exp(B)` gets rid of the log scale, marginal effects get rid of the odd scale (Agresti and Tarantola, 2018). If we get rid of the odd scale, we are back in the probability scale. Let's see our model in probability scale, by asking the plot of the predicted values in the Plots panel (as we did in GLM 2.2.7).

x



First, notice that our model is not linear, because the logistic model is linear for the logit outcome. The plots represents the predicted values in probabilities, so the linearity is lost, but the predicted values make sense. Second, recall that the coefficient of a regression tells us in which direction and how steep is the change in the predicted values as one increases the independent variables. The problem with probability-scaled predicted values is that the direction and size of the change is not constant along the independent variable scores.



In the model above, for instance, for $x = -2$ we can see a mildly positive increase, whereas for $x = 0$ the increase is very steep, which becomes almost flat for $x = 2$. Each of this possible increases (or change in probability) is a marginal effect. However, if we want to quantify the increase (or change) in probability due to the increase in the IV Independent Variable , we have a different marginal effect for each value of x. But we can compute them all (for all observed values of x) and take the average: This is the average marginal effects (AME) produced by GAMLj.

Marginal Effects

				95% Confidence Intervals		
Name	Effect	AME	SE	Lower	Upper	z
x	x	0.271	0.0338	0.205	0.337	8.02

Thus, we can say that on average, the probability of visiting the restroom ($ybin = 1$) increases of .271 as you increase beer (x) of one unit. Please consult Leeper (2021) for more details and technical information.

3.2.9 Multiple IVs

Adding independent variables in the logistic regression, as well as interaction terms, follows the same principles used for the GLM (2). The coefficients are interpreted as partial coefficients, keeping constant the other independent variables. If interactions are included, the linear effects are interpreted as *main effects* (averaged across leves of the moderator). Simple effects analysis and simple slopes plots can be obtained as we did in the GLM applications.

3.3 Logistic with Catecorical IVs

We know that the GLM can accommodate categorical IVs Independent Variables , and so does the logistic model. Categorical IVs Independent Variables are cast into the model using contrast variables (cf. 2.7.2), their coefficients represent group comparisons, and their omnibus tests inform us on the effect of the categorical variable on the probability of being in the group $y = 1$ rather than the group $y = 0$.

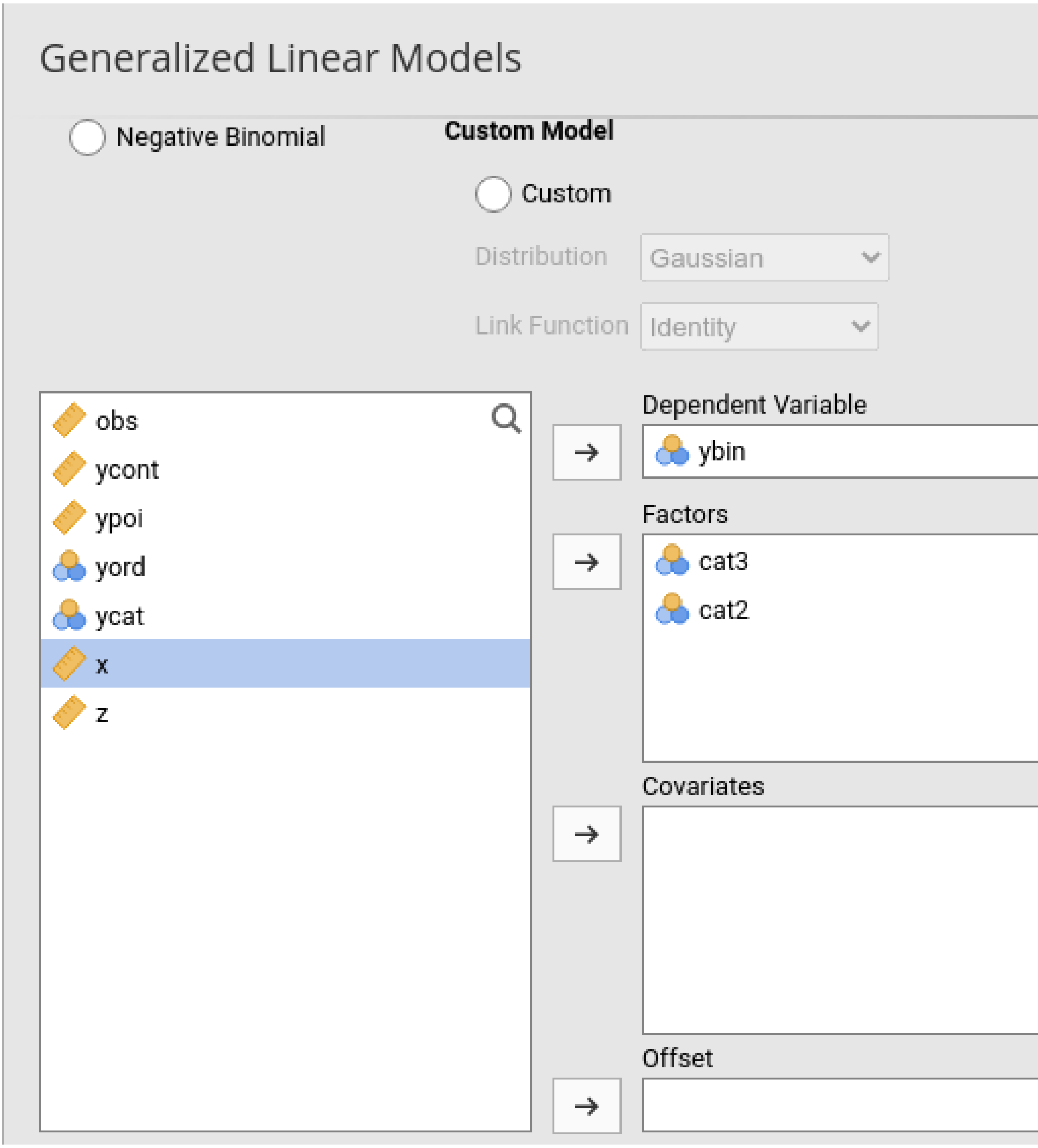
We are going to exemplify this model using the `manymodels` dataset, using `cat2` and `cat3` as our categorical IVs Independent Variables . Recall we use as a cover story the type of beer drunk for `cat3` and the type of bar for `cat2`.

Contingency Tables

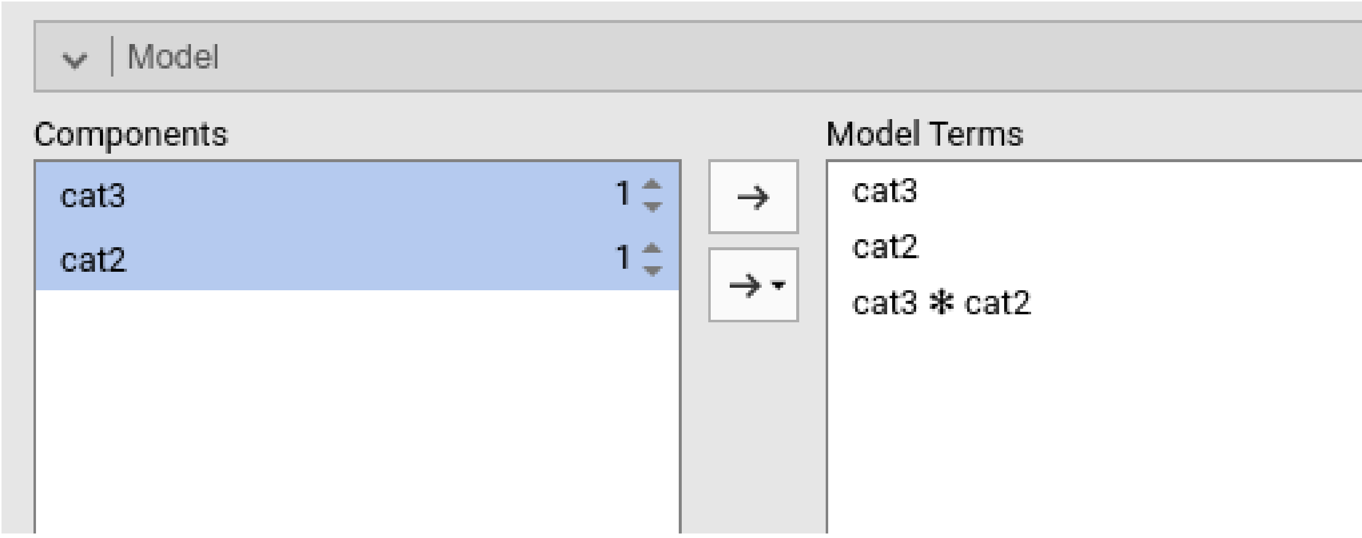
	cat3			
cat2	pilsner	IPA	stout	Total
sports	20	20	20	60
music	20	20	20	60
Total	40	40	40	120

3.3.1 Model Estimation

Now we want to establish possible differences among these groups in the probability of visiting the restroom ($ybin = 1$). We set up a logistic model in which `cat2` and `cat3` are factors, meaning categorical IVs Independent Variables .



As usual in GAMLj, in the presence of categorical IVs Independent Variables the model is automatically set up with main effects and interactions.



3.3.2 Model Fit

Model Results

Model Fit

R^2	Adj. R^2	df	X^2	p
0.122	0.0478	5	19.8	0.001

[4]

Additional indices

Info	Model Value	Comment
LogLikelihood	-71.27	
AIC	154.55	Less is better
BIC	171.27	Less is better
Deviance	142.55	Less is better
Residual DF	114	
Chi-squared/DF	1.05	Overdispersion indicator

The output tables concerning the fit of the model do not really change when the independent variables are categorical. The R^2 indicates the advantage of fit of our model as compared with a intercept-only model, the R^2_{adj} estimate the quantity in the population, and the inferential test (χ^2) indicates whether our model predicts the dependent variable better than chances. More precisely, the model fit indicates if and how much our model predicts the probability of group membership better than just saying that the probability of being in group 1 is equal for every case and it is the number of cases in group 1 divided by the total number of cases.

3.3.3 Omnibus Tests

With categorical IVs Independent Variables , the crucial table is the Omnibus Tests table. Here we find the inferential tests for the main effects and the interactions.

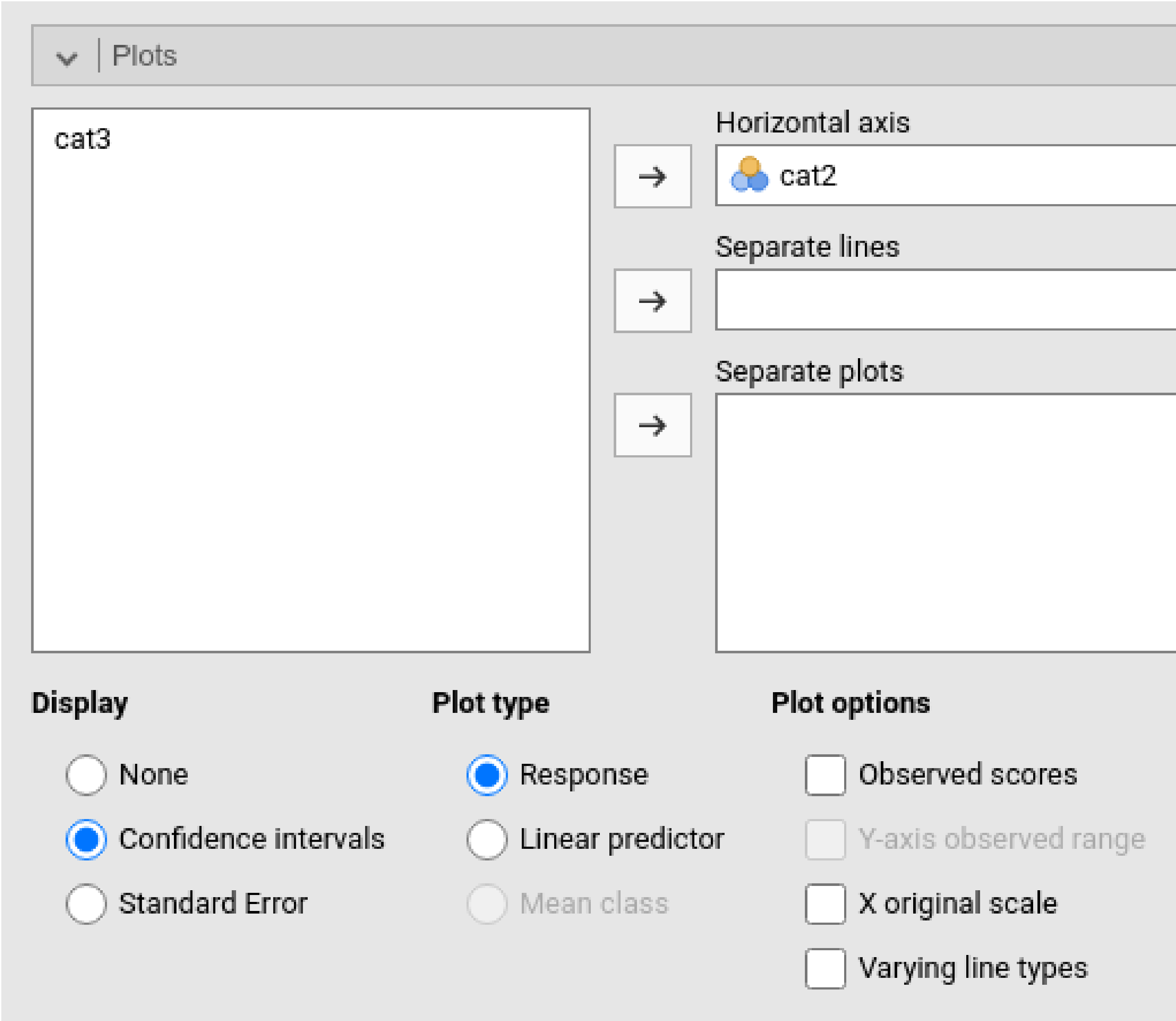
Omnibus tests

	χ^2	df	p
cat3	2.45	2	0.293
cat2	15.73	1	< .001
cat3 ✻ cat2	1.21	2	0.547

We can see that we obtain a non significant main effect for `cat3` indicating that there is not enough evidence to show that the three groups defined by `cat3` have different probabilities to go to the restroom ($y = 1$). The lack of interaction indicates that the effect of `cat3` is not different across levels of `cat2`. We do find a main effects of `cat2`, with $\chi^2(1) = 15.73$, p.<.001. This means that the probability of being in $ybin = 1$ group is different in the two groups defined by `cat2`. People in the two types of bar visit the restroom with different probability. To interpret the direction of the effect, we can look at the plot

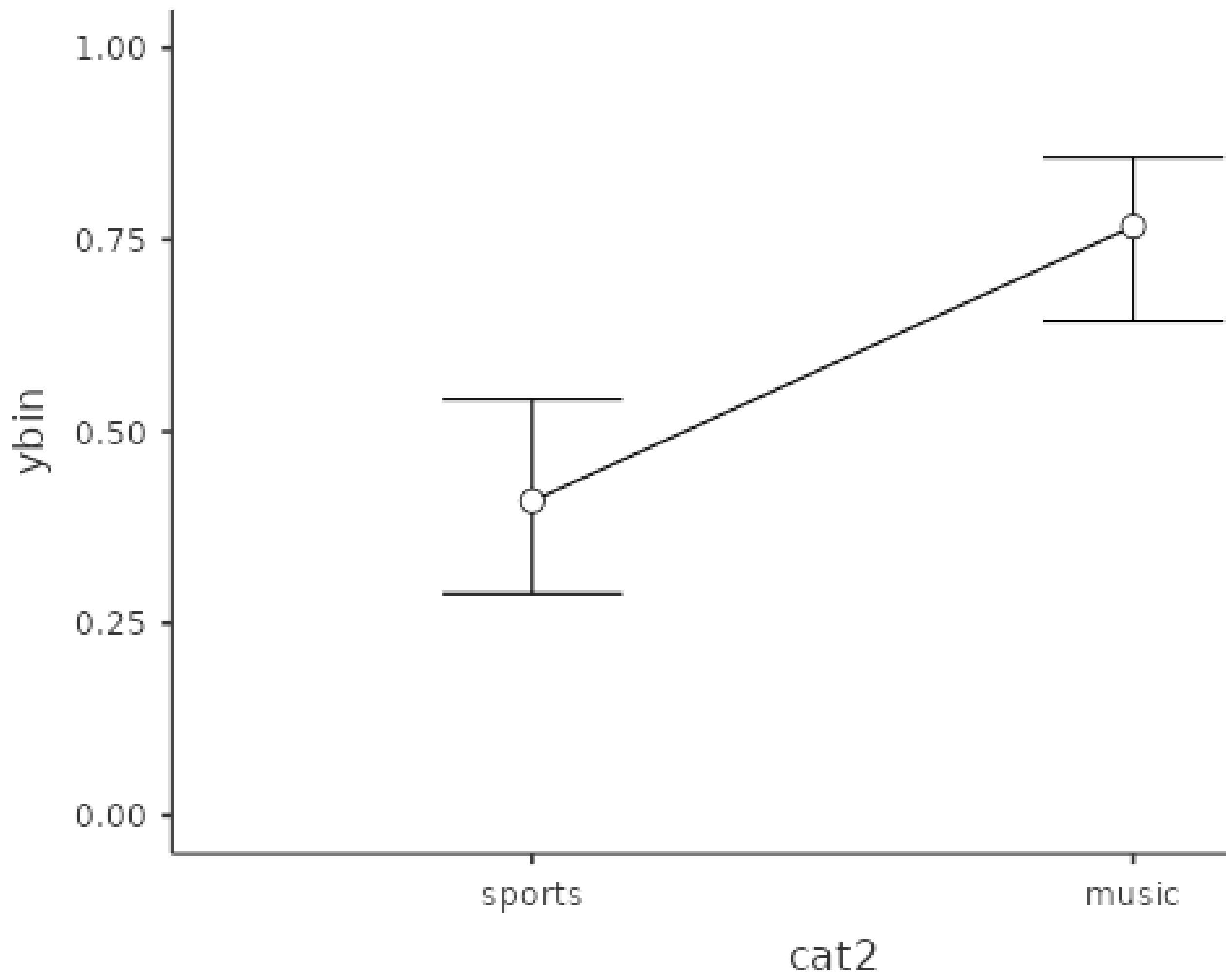
3.3.4 Plots

The plot depicts the average probability of $ybin = 1$ for the groups defined by the variables we ask the plot for. In our case, we ask the plot for `cat2`.


Plots
cat3
Horizontal axis
cat2
Separate lines
Separate plots
Display
None
Confidence intervals
Standard Error
Plot type
Response
Linear predictor
Mean class
Plot options
Observed scores
Y-axis observed range
X original scale
Varying line types

Results Plots

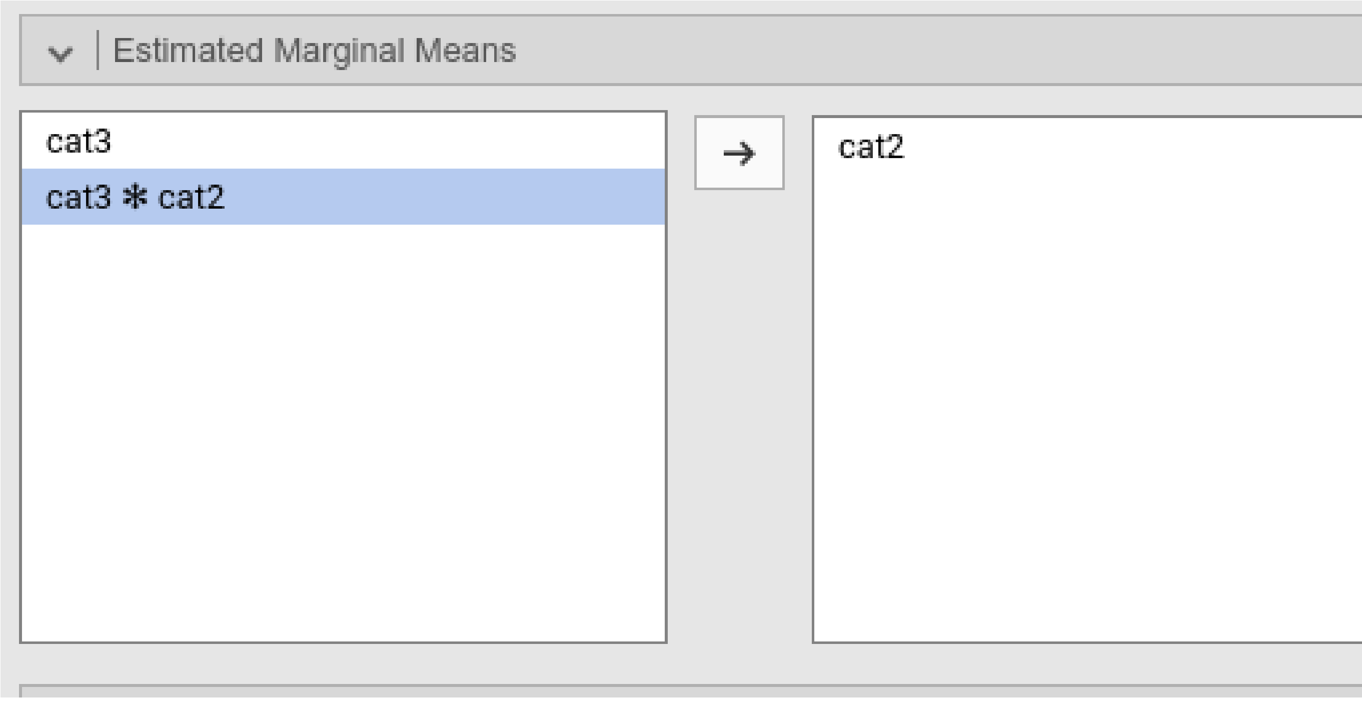
cat2



We can see that the group `music` has a higher probability of visiting the restroom than the group `sport`. We can see the same information in the *Estimated Marginal Means*

3.3.5 Estimated Marginal Means

Estimated marginal means gives us the average probabilities of $ybin = 1$ for the groups. They are expressed in probabilities.



Estimated Marginal Means

Estimate Marginal Means - cat2

cat2	Mean	SE	95% Confidence Intervals Lower	Upper
sports	0.409	0.0660	0.289	0.542
music	0.768	0.0547	0.644	0.858

Note. Expected means are expressed as probabilities

3.3.6 Relative Risk

The relative risk (RR) indices are often used when the IVs Independent Variables are categorical. Set aside some technical details (cf. Zou (2004)), you can think of the relative risk as the rate of change in the probability when comparing two groups.

Relative risk

Effect	RR	SE	95% Confidence Intervals Lower	Upper
Probability	0.551	0.0709	0.428	0.709
IPA - pilsner	1.532	0.4964	0.812	2.891
stout - pilsner	1.342	0.4472	0.698	2.578
music - sports	1.937	0.4993	1.169	3.211
(IPA - pilsner) ✻ (music - sports)	0.485	0.3142	0.136	1.727
(stout - pilsner) ✻ (music - sports)	0.556	0.3704	0.150	2.052

When comparing `IPA` with `pilsner` group, we have that the probability of visiting the restroom is 1.532 times larger in `IPA` than in `pilsner`. The probability is 1.342 times larger in `stout` than in `pilsner`. And so on. In `music` group, the probability is 1.937 times larger than in `sports` group.

A caveat is in order here. If one computes these values based on the estimated marginal means, they do not correspond exactly: $.768/.409 = 1.89$, whereas the RR of cat2 is 1.937. Close, but not equal. The reason is the presence of the interactions, so it has to do with the way probabilities are averaged across the levels of other variables. The difference, however, is always rather small. For models with only one IV Independent Variable , the computations correspond exactly.

3.3.7 Marginal Effects

Marginal Effects

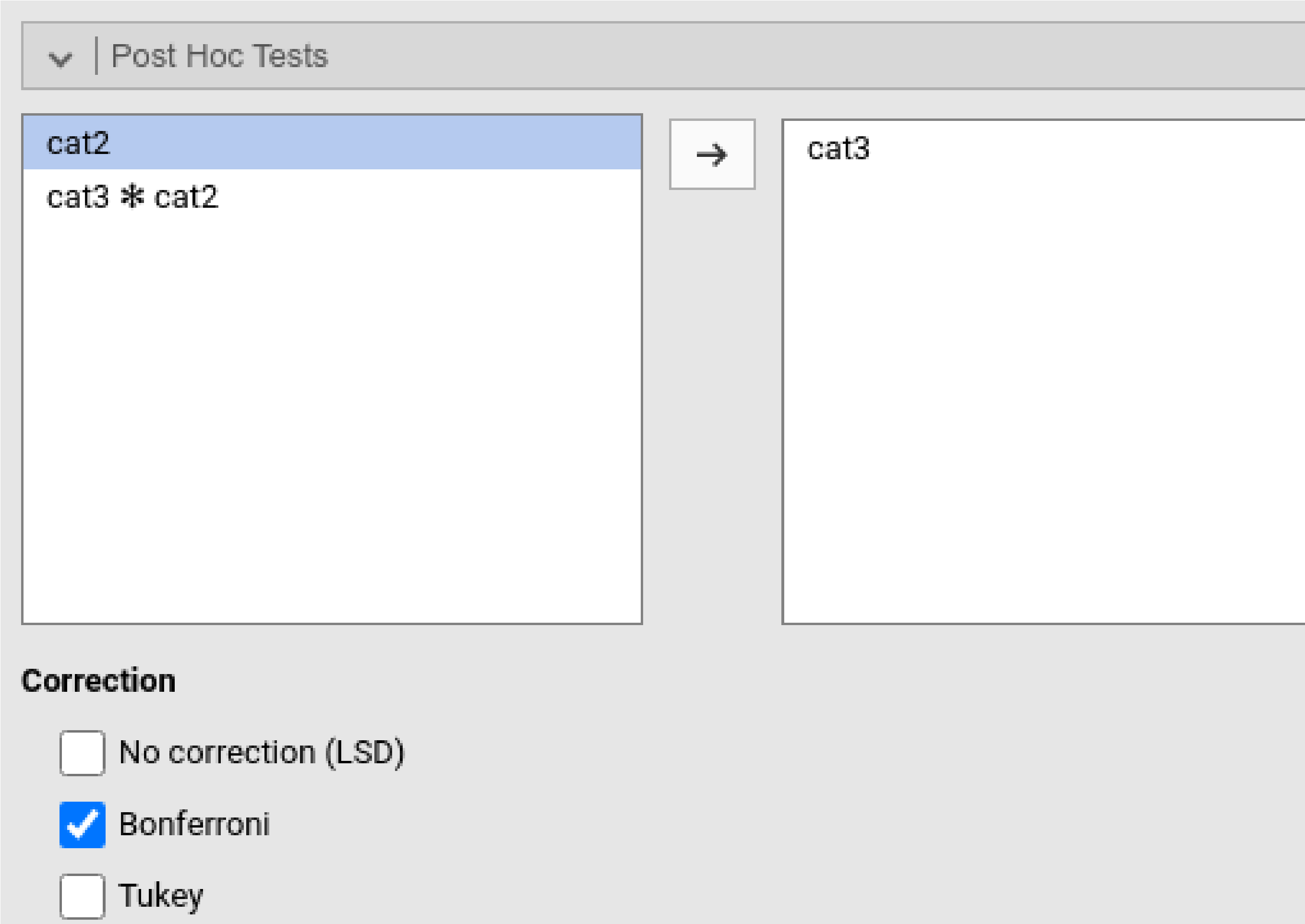
Name	Effect	AME	SE	95% Confidence Intervals Lower	Upper	z
cat2	music - sports	0.3500	0.0822	0.1888	0.511	4.256
cat3	stout - pilsner	0.1000	0.1006	−0.0972	0.297	0.994
cat3	IPA - pilsner	0.1750	0.0989	−0.0188	0.369	1.769

For categorical variables, the marginal means are the expected differences between groups probabilities. As for the RR, in presence of interactions the estimated difference may be slightly different as compared with the EMM difference. For models with only one IV Independent Variable , the computations correspond exactly.

3.3.8 Post-hoc tests

As for the GLM (cf 2.8), one can perform a series of groups comparisons using a post-hoc tests technique. The method is equivalent to the one discussed in the GLM, so we do not need to add much here. The only noticible point here is that the comparisons are estimated and tested as *odd ratios*, so the comparison is based on the odd in one group over the odd in the other group.

As an example, here we ask for the post-hoc comparisons within `cat3`.



Post Hoc Tests

Comparison						
cat3	vs	cat3	OR	SE	z	$p_{bonferroni}$
pilsner	-	IPA	0.452	0.231	−1.550	0.363
pilsner	-	stout	0.638	0.319	−0.897	1.000
IPA	-	stout	1.411	0.700	0.695	1.000

The first comparison shows an $OR = .452$, meaning that the ratio of the pilsner group odd and the IPA group odd is .452. In other words, the pilsner group odd is .452 times the odd of the IPA group. Significance and inferential tests are interpreted as usual, keeping in mind that the p-values are adjusted for the number of comparisons carried out.

3.3.9 Other options

All other options, Simple Effects, Factor Coding, Covariates Scaling, Bootstrap confidence intervals, etc. are the same as in the GLM.

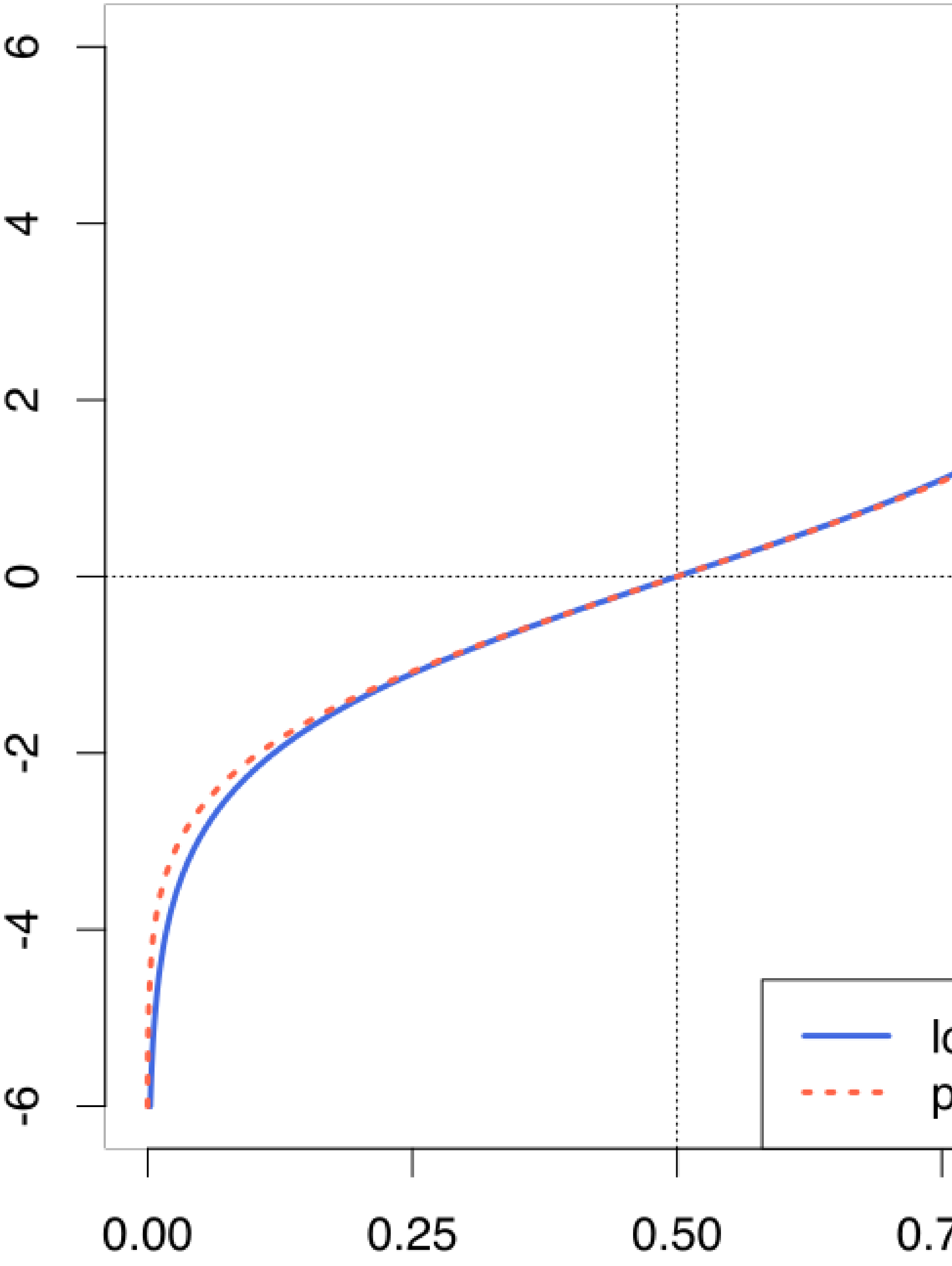
3.4 Probit Model

The probit model (cf Wiki) is practically equivalent to the logistic model. All examples, options and interpretations are the same, so we are not going to explore it in details. The reason GAMLj offers the probit model is because there are several disciplines in which this model is more commonly used than the logistic model. The aim of the two models is the same: Predicting a dichotomous dependent by its relations with a set of IVs Independent Variables .

The only difference between the logistic and the probit model is the link function (3.1). Rather than predicting the *logit* of the probability, the probit model uses the *probit* of the probability.

The *probit* function uses the inverse of the cumulative distribution function of the normal distribution. In a nutshell, imagine a histogram: The cumulative distribution function of the normal distribution associates a probability to any possible value of the X-axis. Its inverse return the X-axis value for any probability value, yielding the predicted values in a scale that admits any positive or negative value. In other words, it does what the logit does, with a different function.

The fact that the logit and the probit models give almost identical results can be easily understood by looking at the way the two link functions transform probabilities in real values: practically in the same way.


6
4
2
0
-2
-4
-6
0.00
0.25
0.50

3.5 The Multinomial Model

3.5.1 Rationale

Sooner or later a dependent variable with more than two groups will cross our path. A choice may be added to the experimental outcome, a third group may be necessary to exhaust a classification, a set of products needs to be tested. When the dependent variable features more than two groups, the logistic or probit model cannot be used. It must be generalized into the *Multinomial Model*. A multinomial model is logically equivalent to estimating a set of logistic regressions, one for each dummy variable (cf. 2.7.2 and B) representing the categorical dependent variable.

Consider a three group variable, with levels A, B and C. To represent it with a set of dummy variables we need 2 dummies.

Levels

D1

D2

A

0

0

B

1

0

C

0

1

`D1` compares level B with level A, and `D2` compares level C with level A. We do not need any other comparison to exhaust the variability in the dependent variable (cf Appendix B). If now we use a logistic model to predict each of these dummies with the independent variables, we can estimate the effects of the independent variables on the probability of belonging to a group rather than another. Thus, a set of logistic regressions would do the job.

$$
\begin{aligned}
logit(D1) &= a_1 + b_1 \cdot x \\
logit(D2) &= a_2 + b_2 \cdot x
\end{aligned}
$$

or, equivalently

$$
\begin{aligned}
log(\frac{p(B)}{p(A)}) &= a_1 + b_1 \cdot x \\
log(\frac{p(C)}{p(A)}) &= a_2 + b_2 \cdot x
\end{aligned}
$$

The overall model fit will be given by the cumulative fit of the two logistic models; the omnibus test of x will be given by testing that both b_1 and b_2 are zero, and the specific effects of x on the comparisons is given by the b_1 and b_2 coefficients. This will be repeated for $K-1$ logistic models, where K is the number of levels of the dependent variable.

3.5.2 Model Estimation

To exemplify, we use our `manymodels` dataset, which has a variable named `ycat`. This variable has three levels. To give some names to its levels and keep up with the bar cover story, imagine the three levels are the choice of an activity to do in the bar: `1=play darts`, `2=chatting`, `3=dancing`.

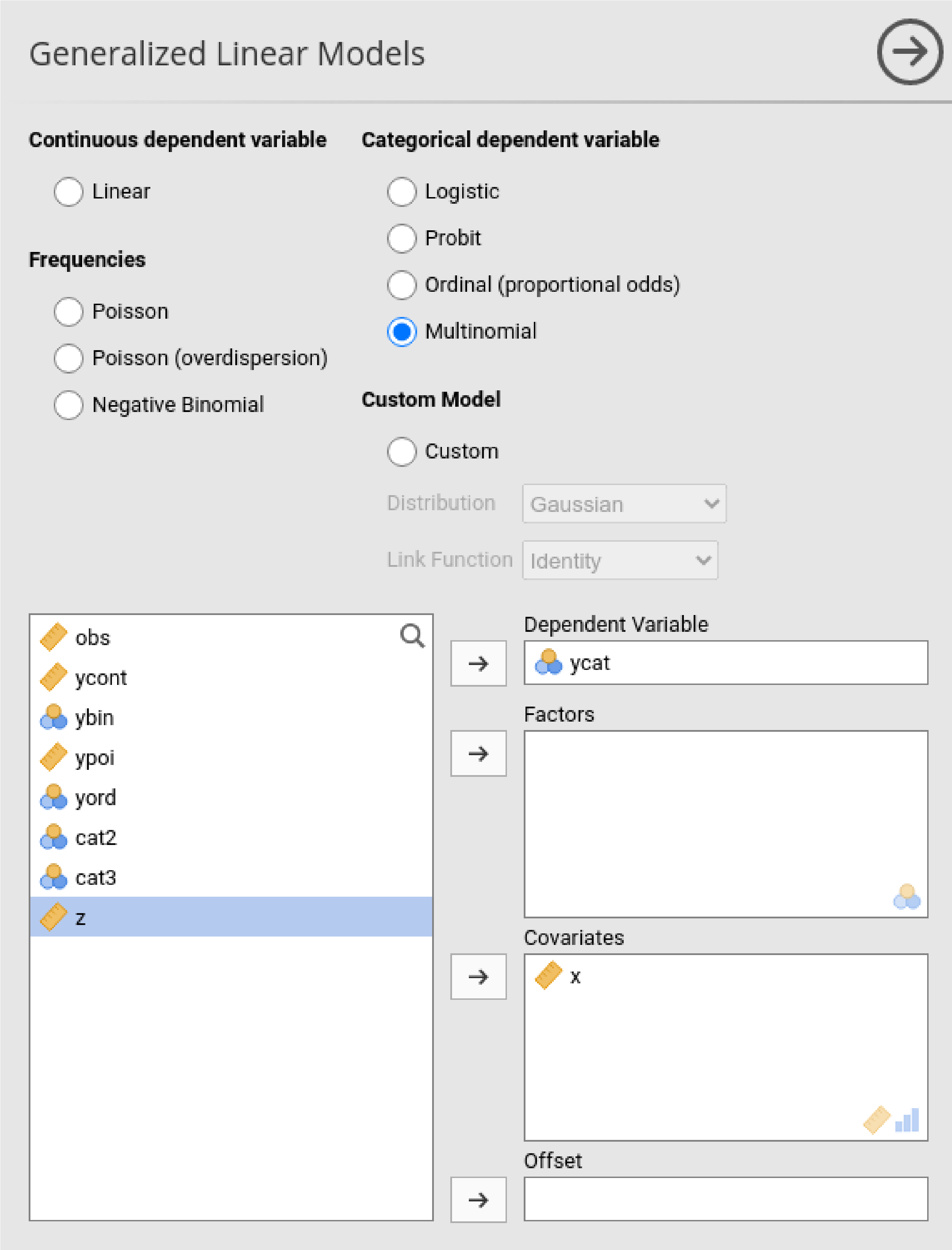
Frequencies

Frequencies of ycat

ycat	Counts	% of Total	Cumulative %
darts	8	6.7 %	6.7 %
chat	99	82.5 %	89.2 %
dance	13	10.8 %	100.0 %

Thus, we want to estimate how the number of beer drunk (x) influences the probability of being in any of these three groups ($ycat$).

We first select Multinomial in the model selection tab, and then set up the dependent variable and the independent variable.


Generalized Linear Models
Continuous dependent variable
Linear
Frequencies
Poisson
Poisson (overdispersion)
Negative Binomial
Categorical dependent variable
Logistic
Probit
Ordinal (proportional odds)
Multinomial
Custom Model
Custom
Distribution
Gaussian
Link Function
Identity
obs
ycont
ybin
ypoi
yord
cat2
cat3
z
Dependent Variable
ycat
Factors
Covariates
x
Offset

3.5.3 Model Recap

Model Info

Info		
Model Type	Multinomial Model	Model for categorical y
Model	nnet::multinom	ycat ~ 1 + x
Distribution	Multinomial	Multi-event distribution of y
Link function	Logit	Log of the odd of y=1 over y=0
Direction	P(Y=j)/P(Y=0)	P(ycat=chat)/P(ycat=darts) , P(ycat=dance)/P(ycat=da
Sample size	120	
Converged	yes	
C.I. method	Wald	

>

This table is useful to remind us what we are estimating in particular, so our interpretation will be correct. In the row `direction`, we can see

$$P(Y = j)/P(Y = 0)$$

This means that we are predicting the (log of) the odd of each level j against level 0. The specification of the levels is in the adjacent column. Here we see

$$P(ycat = chat)/P(ycat = darts), P(ycat = dance)/P(ycat = darts)$$

meaning that the first logistic we meet would predict the odd of being in group `chat` rather than `darts`, the second predicts the odd of being in `dance` rather than `darts`.

Before looking at the specific comparisons, we have the overall fit and tests.

3.5.4 Overal Fit

Model Results

Model Fit

R^2	df	X^2	p
0.114	2	15.9	< .001

[4]

Additional indices

Info	Model Value	Comment
LogLikelihood	-61.7	
AIC	131.3	Less is better
BIC	142.5	Less is better
Deviance	123.3	Less is better

Omnibus tests

	X^2	df	p
x	15.9	2	< .001

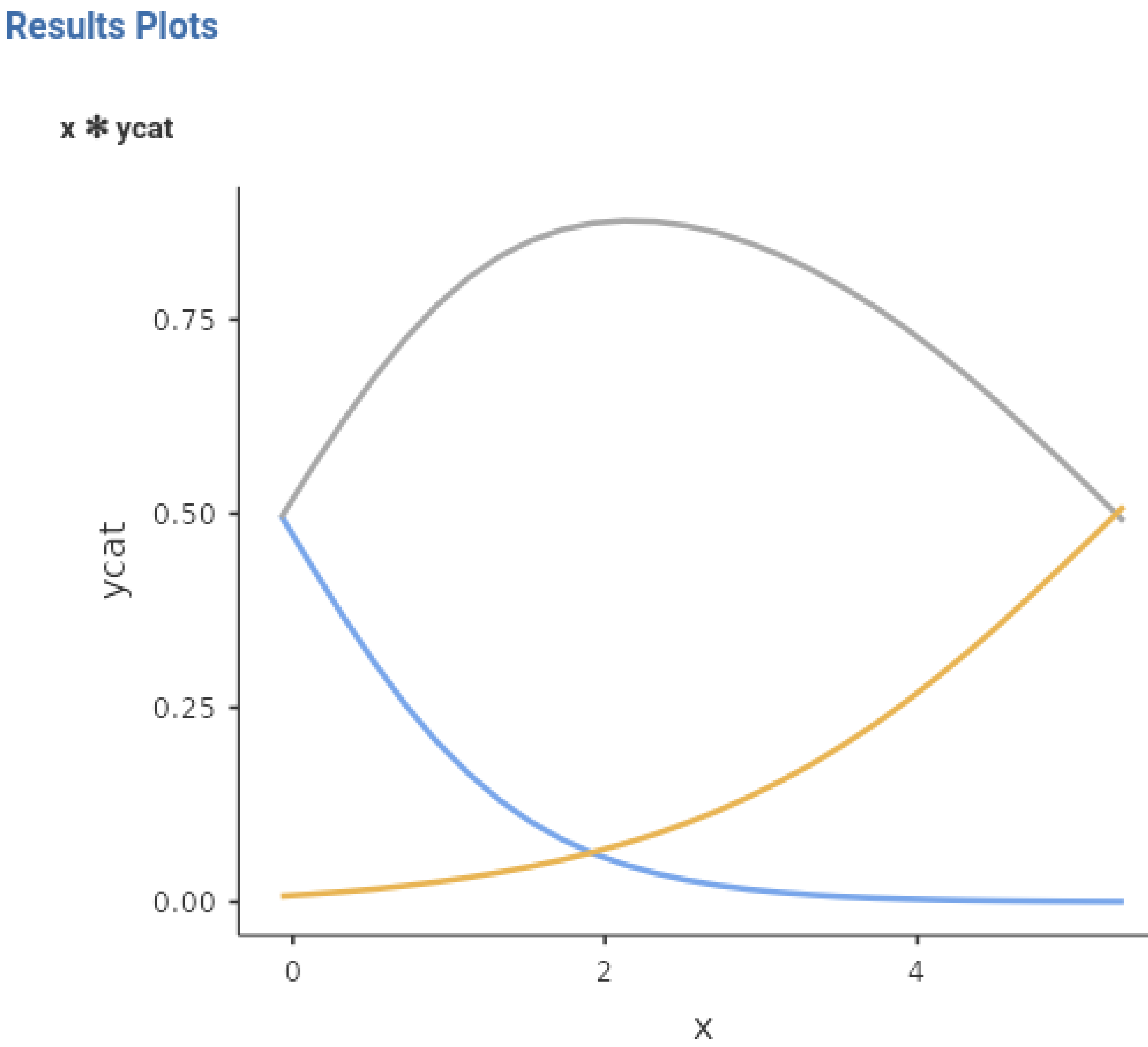
>

As usual, the R^2 tells us the improvement in fit due to our independent variables, and its inferential test provides a test of the hypothesis that our model predicts the dependent variable better than chances. The Ominibus tests are interesting here: They test the null hypothesis that the independent variable(s) has no effect on the probabilities of belonging to the three groups, thus they provide an overall test across the logistic models predicting the dummies. In our case, we can say that *beers* (x) influences the choice of the activity ($ycat$), with $\chi^2(2) = 15.9$, $p. < .001$.

How the independent variable influences the group probabilities can be seen with a plot and by inspecting the coefficients.

3.5.5 Plots

Plot of probabilities are very useful to interpret multinomial models. For multinomial models, the plot depicts one line for each level of the dependent variable. Each line depicts the expected probability of being in that group as a function of the independent variables (plots are produced like for any other model, so it is not shown here. `original X-scale` option is selected as well).



Here we see how the effect of *beers* (x) unfolds into probabilities differences. For low level of beers, it is very unlikely to dance, but this group becomes more likely as *beers* (x) increases. `chats` and `darts` start at the same level of probability, they diverge as *beers* (x) increases: `darts` becomes less and less likely, whereas `chats` increases to decrease again for high levels of *beers* (x). With the plot, a full picture of the effect can be obtained and a clear interpretation of the results can be given.

3.5.6 Coefficients

One can also examine the specific effects of the independent variable(s) on the groups comparisons in the Parameter Estimates (Coefficients) table. Here there are the individual logistic models predicting the dummies representing the dependent variable.

Parameter Estimates (Coefficients)

Response	Name	Estimate	SE	Exp(B)	Exp(B) 95% Confidenc
					Lower
chat - darts	(Intercept)	3.171	0.580	23.83	7.558
	x	1.324	0.547	3.76	1.273
dance - darts	(Intercept)	0.865	0.665	2.37	0.636
	x	2.106	0.618	8.21	2.415

Focusing on the effect of x, we can say that as x increases, the odd of `chatting` rather than `playing darts` increases of 3.76 times, $exp(B) = 3.75$, a significant increase ($z = 2.42$,$p = .017$). Even stronger is the effect of x on the odd of being `dancing` rather than `playing darts`. The odd increases of 8.21 times for each unit more of x.

The other options of the multinomial models are logically equivalent to the options one can use with the GLM (2) or the logistic model (3.2). However, there are some peculiar features that are worth mentioning.

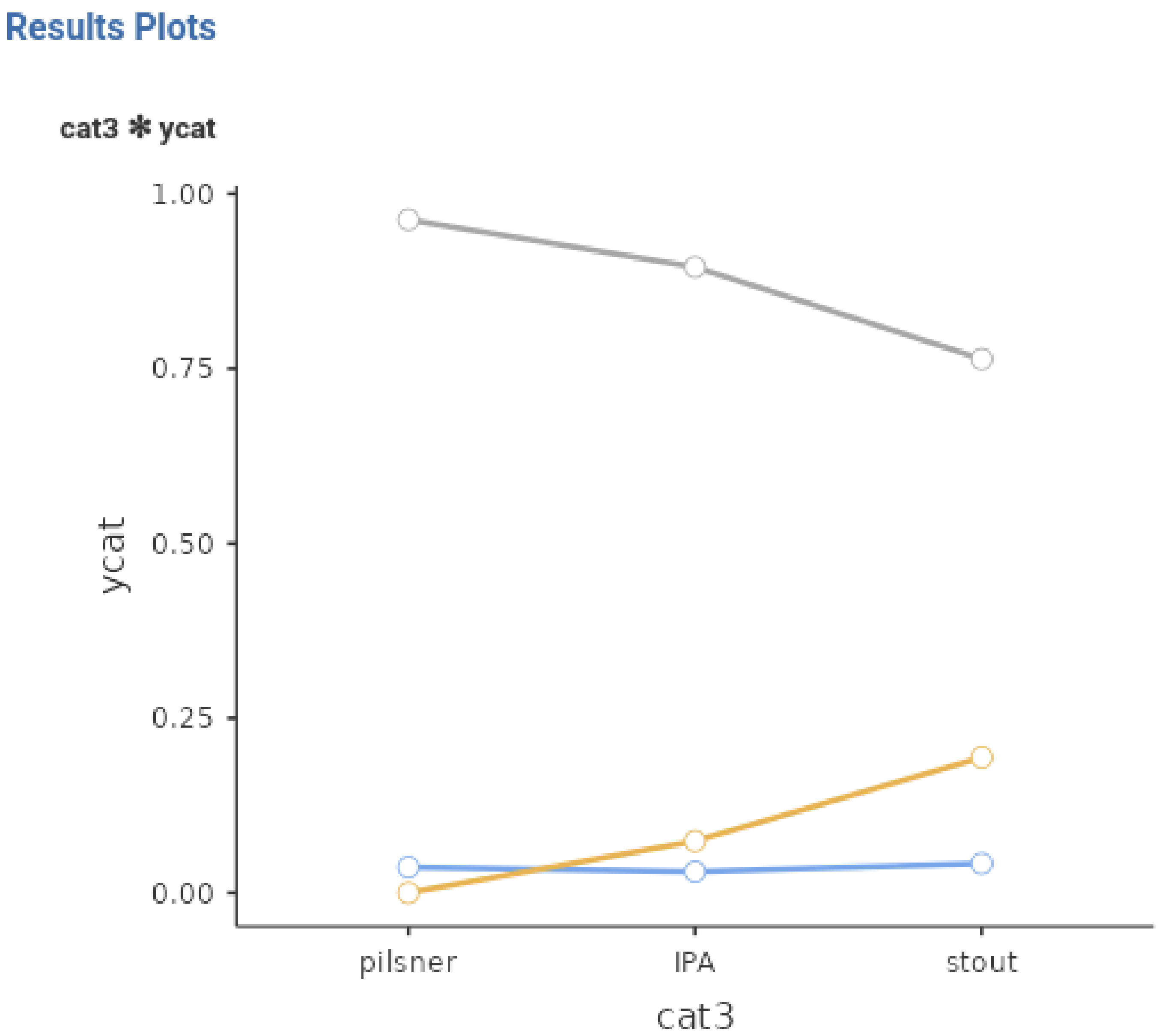
3.5.7 Post Hoc Tests

Let's introduce a categorical variable `cat3` (the type of beer in the story), so we can see the post-hoc tests.

Omnibus tests

	X^2	df	p
x	14.2	2	< .001
cat3	12.4	4	0.015

>



The omnibus test suggest a main effect of `cat3` on the probability of `ycat`, the plot suggests that for the `darts` group there is not much of a difference due to `cat3`, which is a little stronger for the `dance` group and for the `chat` group. This is the logic of the post hoc tests in multinomial models: the probability of each group of the dependent variable is compared between each pair of groups of the independent variable (for input, see 2.8).

Post Hoc Tests

	Comparison						
Response	cat3	vs	cat3	Difference	SE	z	p
darts	pilsner	-	IPA	0.00640	0.0320	0.200	
	pilsner	-	stout	−0.00529	0.0340	−0.156	
	IPA	-	stout	−0.01170	0.0340	−0.344	
chat	pilsner	-	IPA	0.06778	0.0511	1.325	
	pilsner	-	stout	0.19934	0.0727	2.743	
	IPA	-	stout	0.13157	0.0778	1.690	
dance	pilsner	-	IPA	−0.07418	0.0411	−1.804	
	pilsner	-	stout	−0.19405	0.0671	−2.891	
	IPA	-	stout	−0.11987	0.0736	−1.629	

The `cat3` groups do not differ in the probability of being in the `darts` group. The `cat3` groups do not differ in the probability of being in the `chat` group, although some difference can be seen for the comparison `pilsner-stout`. The `cat3` groups do not differ in the probability of being in the `dance` group, although some difference can be seen for the comparison `pilsner-stout`, again.

We noticed that `cat3` had a significant effect on the dependent variable (Omnibus Test), but no post hoc test reaches a significant level. That is not an error, it could happen because of the correction for multiplicity. Because in multinomial models the comparisons are usually many, the adjustment may result in very under-powered comparisons. The indication is to use the post-hoc only when strictly necessary, namely when one has really no idea of what to expect from our data.

3.5.8 Marginal Effects

Recall the marginal effects in the logistic model (cf. 3.2.8 and 3.3.7). They are the average change in probability (probabilities differences) along a continuous IV Independent Variable or between two groups defined by a categorical IV Independent Variable . In multinomial models, they have the same interpretation, but they are produced for each comparison (dummy) between the dependent variable groups. In our example with `x` and `cat3` IVs Independent Variables , we have the following marginal effects.

Marginal Effects

Response	Name	Effect	AME	SE	95% Confidence In	
					Lower	Up
chat - darts	x	x	0.08263	0.0365	0.0112	0.
	cat3	stout - pilsner	0.67794	0.1167	0.4493	0.
	cat3	IPA - pilsner	0.00509	0.0563	-0.1053	0.
dance - darts	x	x	0.27741	0.0693	0.1415	0.
	cat3	stout - pilsner	-0.02225	0.0606	-0.1410	0.
	cat3	IPA - pilsner	0.66664	0.1374	0.3973	0.

The first row presents the average marginal effect, $AME = .082$ of x on the comparison between `chat` and `darts`: that is, the average change in probability of being in the `chat` group rather than the `darts` group along the values of `x`. The second row ($AME = .677$) is the difference in the probability of being in group `chat` rather than `darts` between the group `stout` and group `pilsner`. The third row indicates the difference in the probability of being in group `chat` rather than `darts` between the group `stout` and group `pilsner`.

The following three rows are the same comparisons, but operated on the probability of being in group `dance` rather than in group `darts`.

3.5.9 Simple Effects

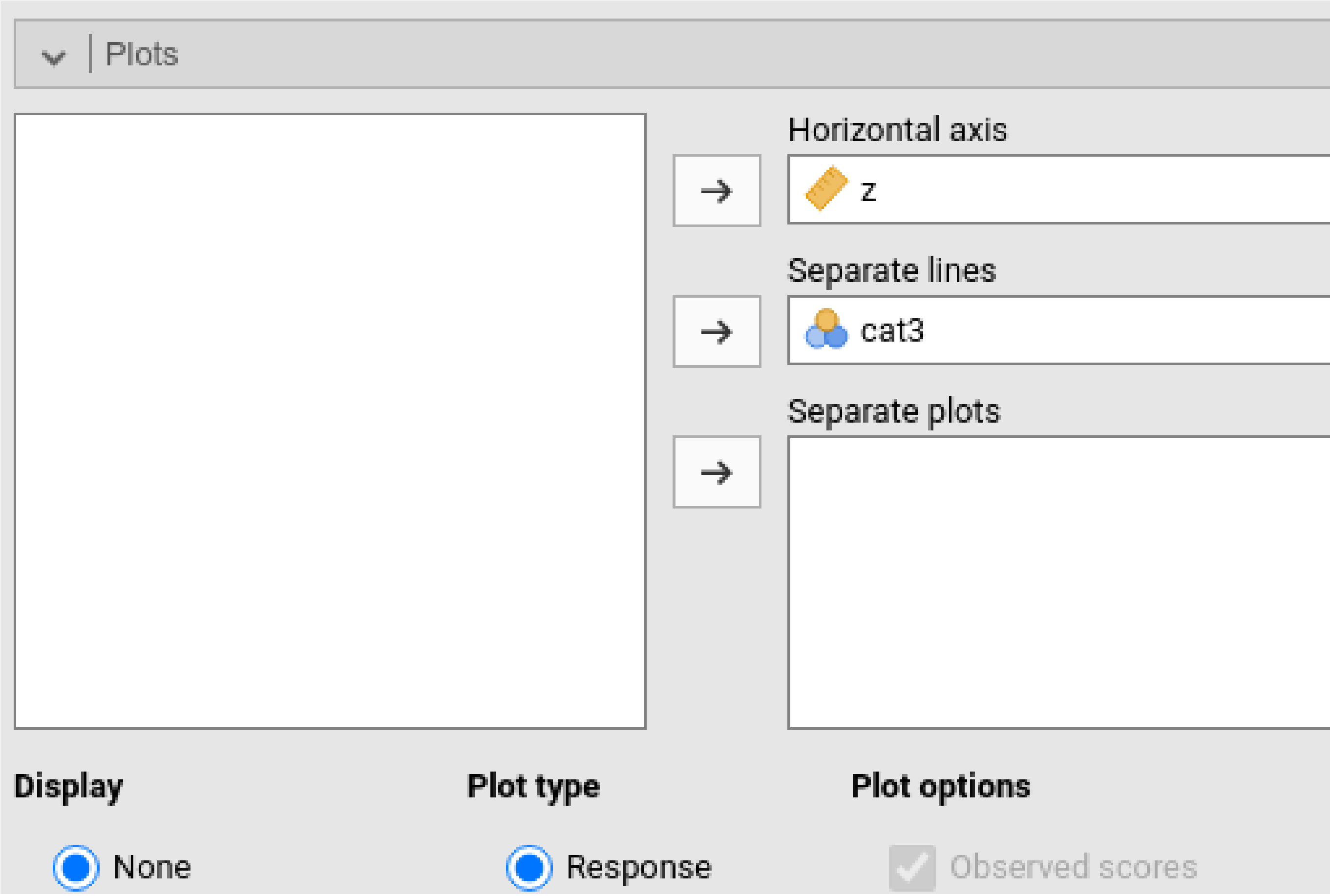
We now examine a multinomial model with `z` (remember *extraversion*) and `cat3` as independent variables, with the addition of their interaction as a term in the model.

The omnibus tests are the following:

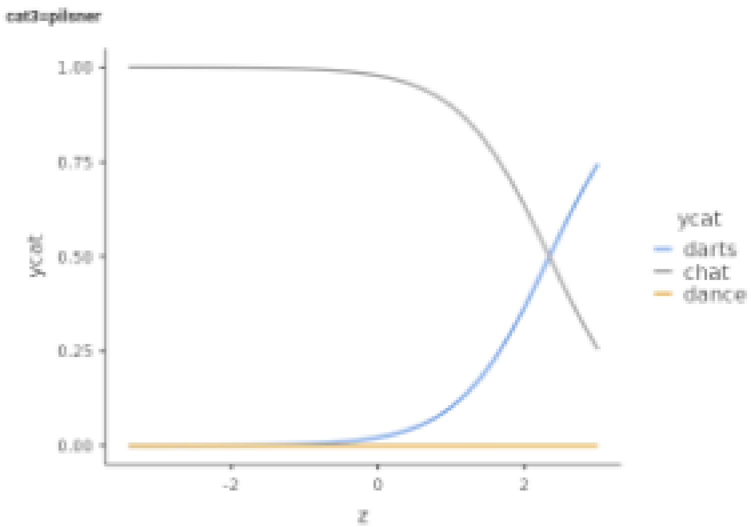
Omnibus tests

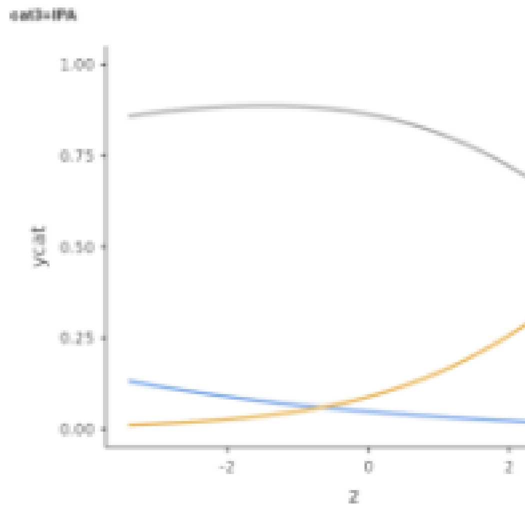
	X^2	df	p
cat3	11.658	4	0.020
z	0.456	2	0.796
cat3 ✻ z	17.488	4	0.002

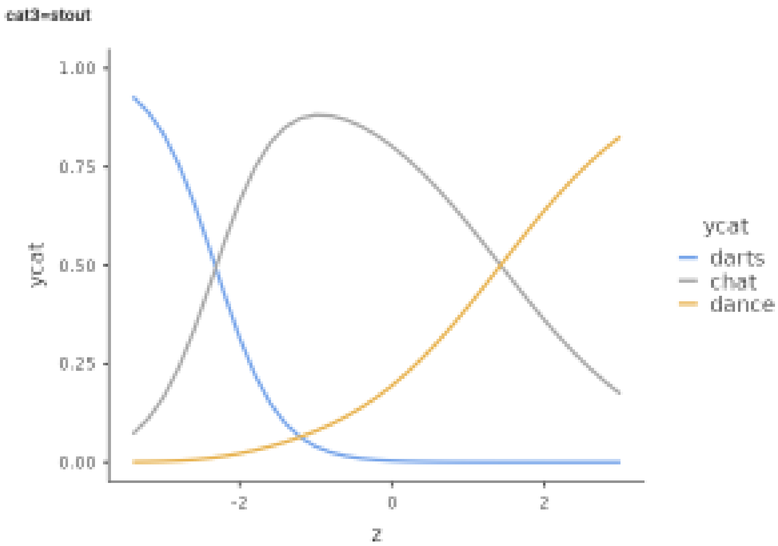
We find an interaction between the continuous variable `z` and the categorical variable `cat3`. To explore this interaction we can estimate and test the simple slopes of `z` at different levels of `cat3`. This estimation provides the numerical version of a simple *slopes* plot, that we can obtain in Plots as usual



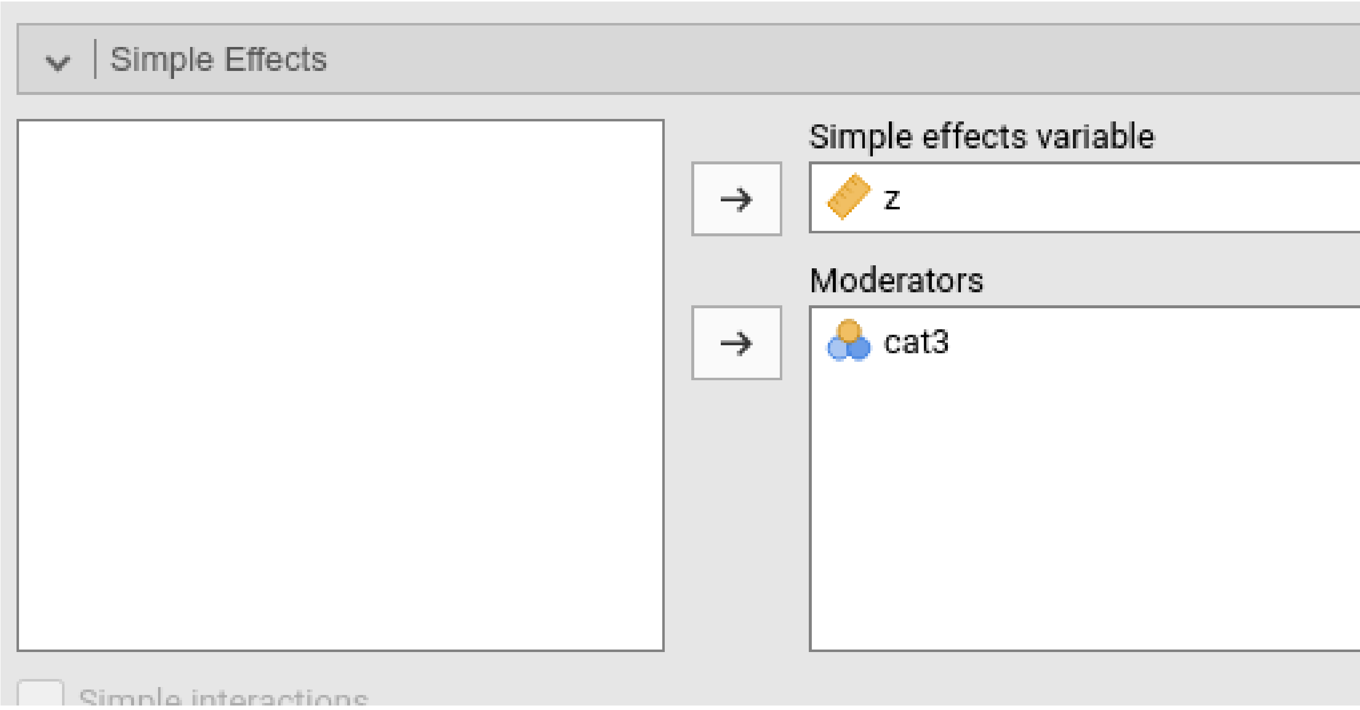
The plots (rearanged) looks like this:







These are the effects (difference in probability) of `z` on `ycat` estimated for the thee groups defined by `cat3`. To know where the effects are or are not present, we can ask for the simple effects.



Simple Effects

ANOVA for Simple Effects of z

Moderator			
cat3	X^2	df	p
pilsner	6.55	2	0.038
IPA	1.59	2	0.450
stout	18.44	2	< .001

It is clear that the effect of `z` is different from zero for the group `pilsner`, $\chi^2(2) = 6.55, p. = .038$, it is very weak and not significant for the group `IPA`, $\chi^2(2) = 1.59, p. = .450$, and is significant and strong for the group `stout`, $\chi^2(2) = 18.44, p. < .001$.

3.5.10 Other options

All other options, Simple Interactions, Factor Coding, Covariates Scaling, Bootstrap confidence intervals, etc. are the same as in the GLM.

3.6 The Ordinal Model

Consider now a different type of variable: immagine we asked the people at the bar to express their appreciation for the bar. We gave them 5 options

- I will never come back
- I may come back sometime
- I will come back once in while
- I will come back often
- I would like to be here every day

Silly as it may seems, this variable represents a very common type of variable in science. It produces 5 possible values, that are clearly ordered in terms of preferences for the bar. Despite that, we cannot honestly assume that this is a continuous variable, because the distance between, say, *I will never come back* and *I may come back sometime* cannot confidently be assumed to be the same as the distance between *I will come back often* and *I would like to be here every day.* Nevertheless, It is plain to see that *I will never come back* conveys much less appreciation than *I will come back often*, which is in turn less appreciative than *I would like to be here every day* . So, this is an *ordinal* variable (Stevens, 1946). The argument become more serious if we focus on the Likert scale (Likert, 1932), one of the most frequently used measurement instrument in social science. Different authors have claimed that a Likert scale is an ordinal variable and it cannot be considered a continuous scale, whereas other have claimed that the assumption of continuity does not really bias the results of parametric analyses. See for instance Wu and Leung (2017) and the references there. We are not going to solve this conundrum here. We assume one has decided that the dependent variable is of the ordinal type.

3.6.1 Rationale

Assume we treat an ordinal variable as a multinomial variable,featuring K levels. If we do that, we treat the levels as completely unordered, and thus the only information that we are using to estimate the generalized model is that the K levels are different. As we have seen in the discussion of multinomial regression (3.5), we would need K-1 logistic models to obtain our results. In doing that, however, we over-parameterize the model, because we do not need so many coefficients to describe the effect of an IV Independent Variable on the dependent variable. We can take advantage of the order of the levels to simplify the model (its results, not really its logic).

Consider our ordinal $y = \{1, 2, 3, 4, 5\}$, and say that each i level (the choice made by the participant in the example) has probability π_i. We can ask what is the probability of choosing *I will never come back* , or the probability of choosing either *I will never come back* or *I may come back sometime*, or the probability of choosing *I will come back once in while* or lover levels, etc. In other words, we can focus on the probability of choosing *up to a level*, or equivalently, a level or any other below it. This is called *cumulative probability*, and it can be written like this:

$$p(y \leq k) = \sum_{i=1}^{k} \pi_i$$

meaning: the probability to choose any level up to k is the sum of the probabilities of the levels less or equal to k. As we have seen above (3.2), the linear model does not work well in predicting probabilities, so let express these probabilities as logits:

$$logit_k = log\left(\frac{p(y \leq k)}{p(y > k)}\right) = log\left(\frac{\sum_{i=1}^{k} \pi_i}{\sum_{i=k+1}^{K} \pi_i}\right)$$

which translates into predicting the log of the odd of choosing up to one level over choosing any other higher levels. That can be done with a linear model

$$logit_k = a_k + b_k x$$

In this set of models, each b_k coefficient would tell us the effect of x on the (logit) probability of choosing up to one level over choosing any other higher level. However, that would not be much of simplification, because we still have $K-1$ linear models, one for each levels, apart from the last one. But we can assume, and check, that the effect of x on the logit is the same for each level k, so we end up with only one regression coefficient:

$$logit_k = a_k + bx$$

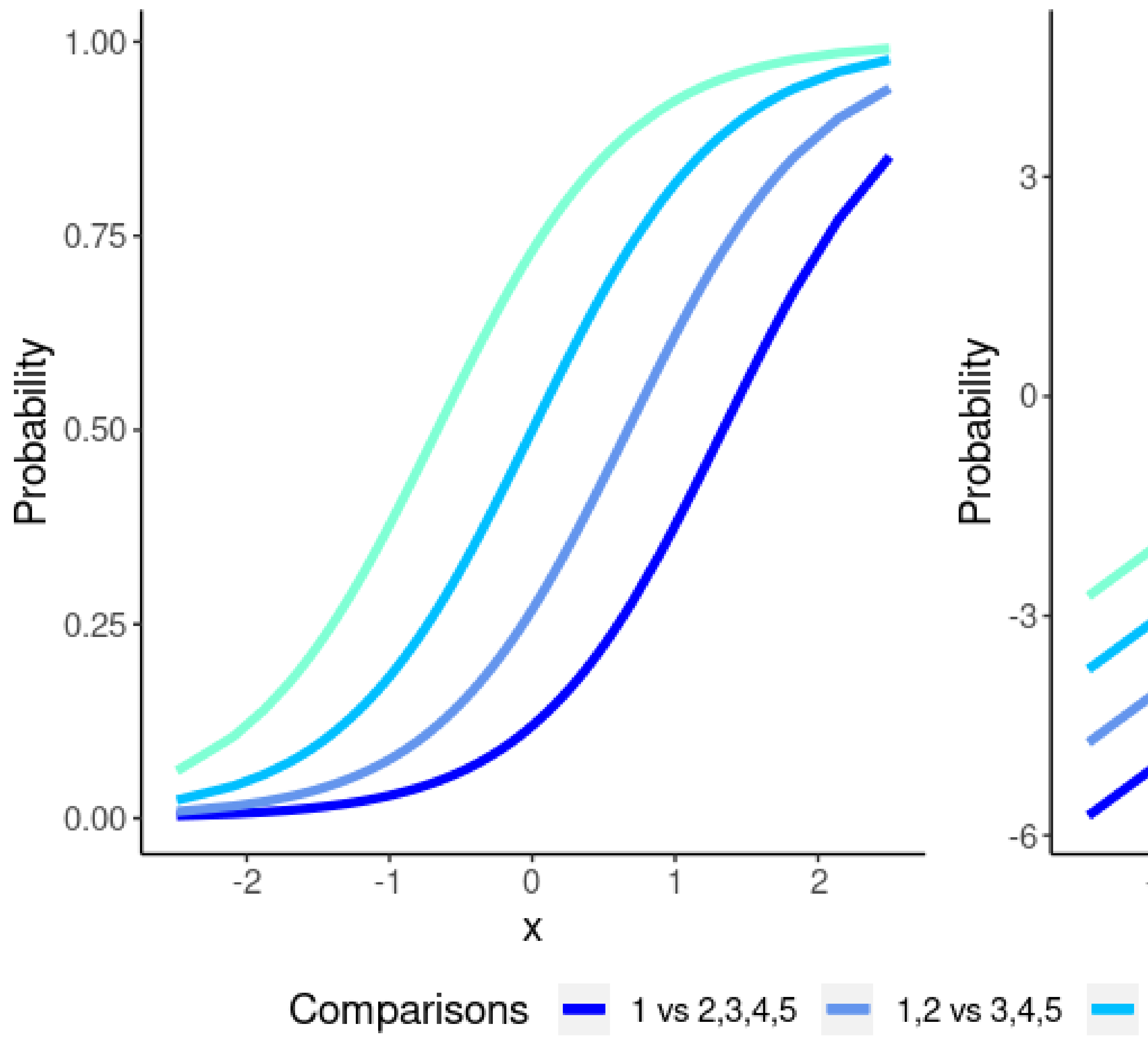
This is the *proportional odds* assumption, which often gives its name to the model: *the propotional odds model.* Thus, the ordinal regression is a generalized linear model that uses the *cumulative-logit* as link function and assumes proportional odds.

To interpret the coefficients, however, we need an extra step. So far , the b is the influence of x in choosing k or below, so a positive value means that as you increase x it is more likely to choose a lower level. That is counter-intuitive, therefore the model is estimated reversing the sign of b, so that the interpretation comes more natural.

$$logit_k = a_k - bx$$

In this model, intercepts are the log-odds of choosing a level or below that level. The regression coefficient describes the increase in log-odds of choosing a level or above associated with a one unit increase in x. In other words, the b coefficient indicates how much the independent variable increases the probability of choosing a higher level, so taking a step up on the ordinal scale.

Another way to see how the model works, is to consider the following plots.

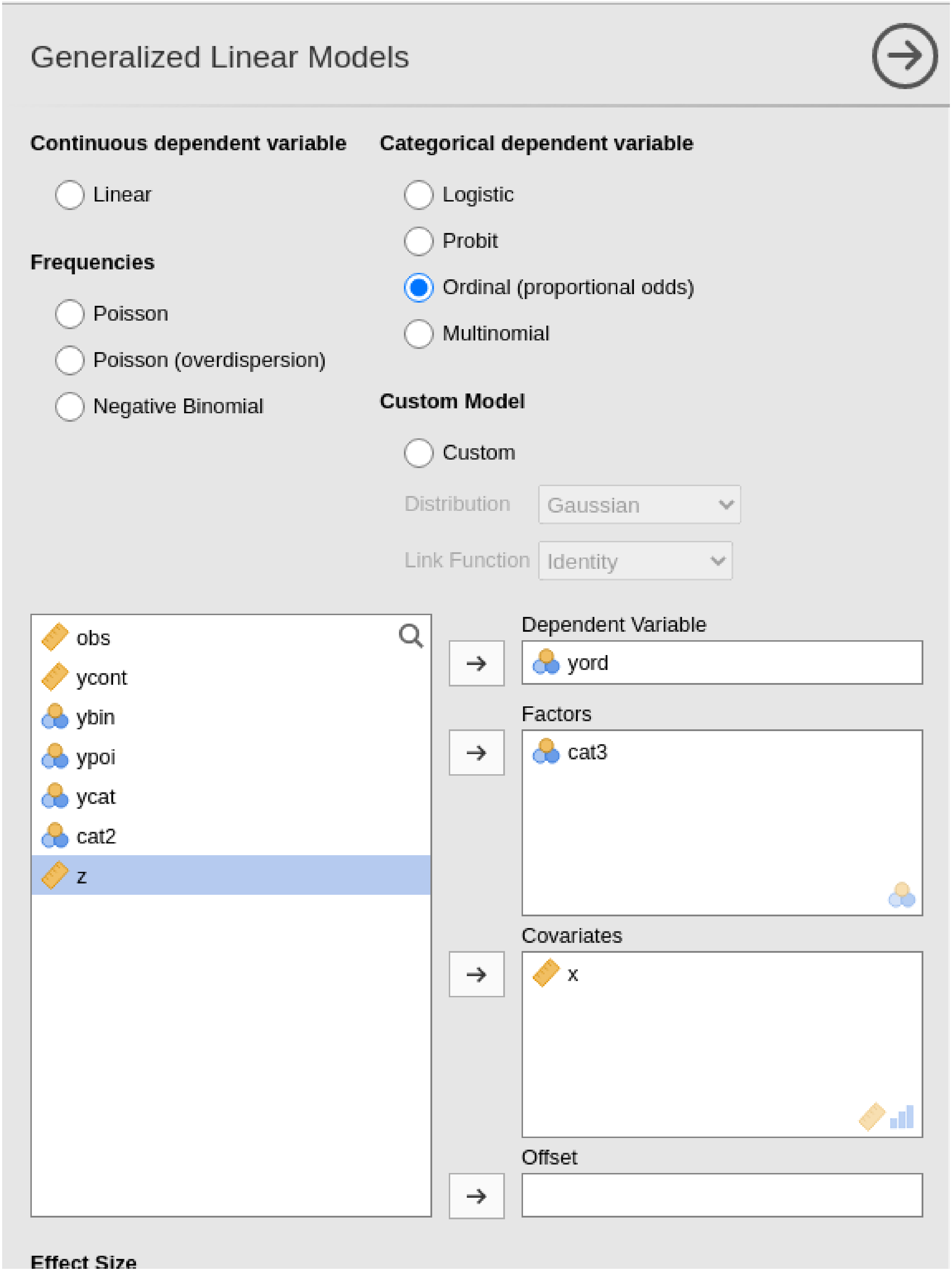


On the left panel, we have the cumulative probability functions for each level compared with any other higher level. When the probability is predicted as a logit, its relationship with the independent variable is linearized (right panel), so we are fitting regression lines, with different intercept but the same slope (lines are parallel). This is the proportional odds model.

GAMLj provides the proportional odds model as implemented by the R package `ordinal`. Details can be found in the ordinal package vignettes.

3.6.2 Model Estimation

We simply set the model as `Ordinal` at the top of the input variable, and set our dependent and independent variables as for any other model. Here we inserted as independent variables both a continuous (x) and a categorical variable ($cat3$), so we can explore more options and techniques within the ordinal model.


Generalized Linear Models
Continuous dependent variable
Linear
Frequencies
Poisson
Poisson (overdispersion)
Negative Binomial
Categorical dependent variable
Logistic
Probit
Ordinal (proportional odds)
Multinomial
Custom Model
Custom
Distribution
Gaussian
Link Function
Identity
obs
ycont
ybin
ypoi
ycat
cat2
z
Dependent Variable
yord
Factors
cat3
Covariates
x
Offset
Effect Size

3.6.3 Model Recap

Generalized Linear Models

Model Info

Info		
Model Type	Cumlative Link Model	Proportional odds logistic
Model	ordinal	yord ~ 1 + x
Distribution	Logistic	
Link function	Logit	Log of the odd of y=1 over y=0
Direction	P(Y ≤ j)/P(Y > j)	j= 1 | 2 | 3 | 4 | 5
Sample size	120	
Converged	yes	
C.I. method	Wald	

[3]

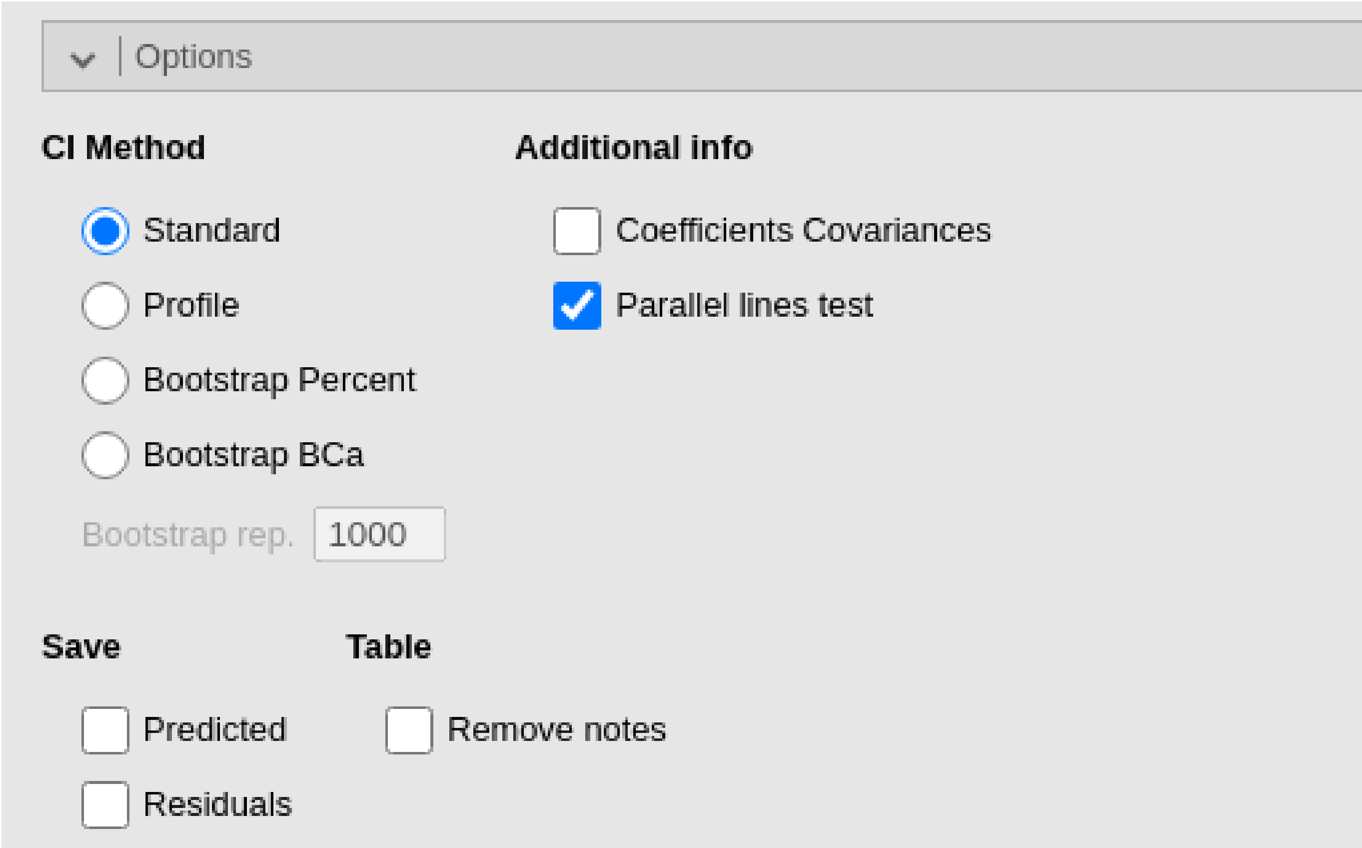
In the Model Info table we find a set of properties of the estimated model, as we have seen for all the other generalized linear models. The `direction` row indicates how the levels of the dependent variables are ordered, in this example as 1|2|3|4|5. This is usually obvious, but in some cases it is important to check that the order of the dependent variable levels is indeed how intended by the user.

3.6.4 Parallel Lines test

Before looking at the results of the model estimation, we should remember that the ordinal model estimated by GAMLj is a *proportional odds model*, in which we assume that the coefficient associated with an independent variable is the same for all logit estimated along the dependent variable scale (cf 3.6). GAMLj provides a test for this assumption, usually named *parallel lines test*. It is named like that because proportional odds functions, when estimated in the logit scale, are equivalent to parallel lines.

The logic of the test is simple: the model with all coefficients of an independent variable set equal is compared with a model in which the coefficients are allowed to vary from logit to logit. The two models are compared with a log-likelihood ratio test. If the test is not significant, we have indication that the assumption of proportional odd is met. When significant, we have indication that a breach to the assumption may be in place.

The test can be found in the input Options panel.



Parallel lines test

	Log-Lik.	AIC	X²	df	p
cat3	−103	232	16.48	6	0.011
x	−110	239	3.30	3	0.347

The test is performed for each independent variable effect. We can see that no problem arises with x, because the χ^2 is clearly not significant. A doubt can be cast on `cat3`, that show that the assumption of proportional odds does not perfectly apply to its effects ($\chi^2 = 16.48$, $p = .011$). It should be said, however, that this tests are quite conservative, so one should be very lenient in their interpretation. One can argue, for instance, that the deviation from the assumption does not seem very strong, so the model can be saved and the results interpreted anyway.

3.6.5 Model Fit

Model Results

Model Fit

R^2	df	X^2	p
0.181	3	49.1	< .001

[4]

Additional indices

Info	Model Value	Comment
LogLikelihood	−111	
AIC	237	Less is better
BIC	256	Less is better
Deviance	910	Less is better

The model R-squared is produced with its inferential test, the Chi-square test. This provides a test of the ability of the model to predict the dependent variable better than chances. The R-squared is the McFadden's pseudo R squared (cf GAMLj help). We can interpret it as the proportion of reduction of error, or the proportion of increased accuracy in predicting the dependent variable using our model as compared with a model without independent variables (cf. Appendix A). The adjusted R^2 is not produced for the ordinal model.

The additional indices reported in Additional indices reports other indices useful for model comparisons or evaluation of models.

3.6.6 Omnibus Tests

As for all the other generalized linear models, we have omnibus tests for the effects of our IVs Independent Variables . They are especially useful when dealing with categorical independent variables, because with continuous independent variables they are redundant as compare with the coefficients tests.

Omnibus tests

	X^2	df	p
x	50.5	1	< .001
cat3	79.4	2	< .001

We noticed in our example that both variables show significant effects. Thus, while keeping the other constant, each variable is able to influence the probability of choosing a higher level of the ordinal variable. We can now examine these effects to understand their size and direction.

3.6.7 Coefficients

Parameter Estimates (Coefficients)

					Exp(B) 95% Confide
Name	Effect	Estimate	SE	Exp(B)	Lower
(Threshold)	1|2	−4.460	0.628	0.0116	0.00338
(Threshold)	2|3	−2.262	0.306	0.1042	0.05716
(Threshold)	3|4	1.337	0.255	3.8063	2.30709
(Threshold)	4|5	4.406	0.612	81.9390	24.70899
cat31	IPA - pilsner	0.630	0.462	1.8774	0.75915
cat32	stout - pilsner	1.087	0.480	2.9653	1.15787
x	x	1.343	0.229	3.8312	2.44433

The first rows of the table report the intercepts, in this context called *thresholds*. Those are the expected logit (under `Estimate` column) or the expected odd (under `exp(b)` column) of the regression fitted for each logit. They are seldom interpreted, but they are basically the odds of choosing a level or lower over choosing any higher level, for all IVs Independent Variables equal to zero.

The x effect $b = 1.343$ and its $exp(B) = 3.831$ indicate the effect of the independent variable on the probability of choosing an higher level: for one unit more in x, the odd of passing from one level to the higher level increases of 3.831 times. Thus, the more beers one drinks (x), the higher is the appreciation of the bar. As regards the effects of `cat3`, as compared with `pilsner` group, the `IPA` group has an odd of increasing the chosen level .630 times smaller, whereas for the `stout` group the odd is 1.097 times larger. The latter comparison is significant.

We can probe the categorical variable effect by displaying and evaluating the expected means.

3.6.8 Estimated marginal means

Obtained as for any other model in the input (cf 2.10), the estimated marginal means for the ordered model are expected mean classes: the expected mean level chosen, broken down by groups defined by the independent variable.

Estimate Marginal Means - cat3

			95% Confidence Intervals	
cat3	Mean	SE	Lower	Upper
pilsner	2.96	0.0868	2.79	3.13
IPA	3.13	0.0888	2.96	3.30
stout	3.26	0.0996	3.06	3.46

Note. Expected means are expressed as expected class
Note. Classes are: 1=1, 2=2, 3=3, 4=4, 5=5

In this example, we see that the group `pilsner`, on average, is choosing level 3, that in our example means, *I will come back once in while*, whereas both group `IPA` and `stout` tend to be between *I will come back once in while* and *I will come back often*. Classes are always marked from 1 to K, where K is the number of levels in the dependent variable. The correspondence between class and mark is illustrated in the footnote of the table.

3.6.9 Post-hoc tests

Post-hoc tests are asked in the input and interpreted as for any other model (cf. 2.8)

Post Hoc Tests

Comparison						
cat3	vs	cat3	Ratio	SE	z	$p_{bonferroni}$
IPA	-	pilsner	1.88	0.867	1.363	0.518
stout	-	pilsner	2.97	1.423	2.265	0.070
stout	-	IPA	1.58	0.737	0.980	0.982

The only remark needed here is that for ordinal models, the group comparisons are conducted using the logit, testing them as comparisons of groups based on the difference in logit. However, in the table, these comparisons are reported as odds ratios, where the difference in logit is exponentiated before being displayed.

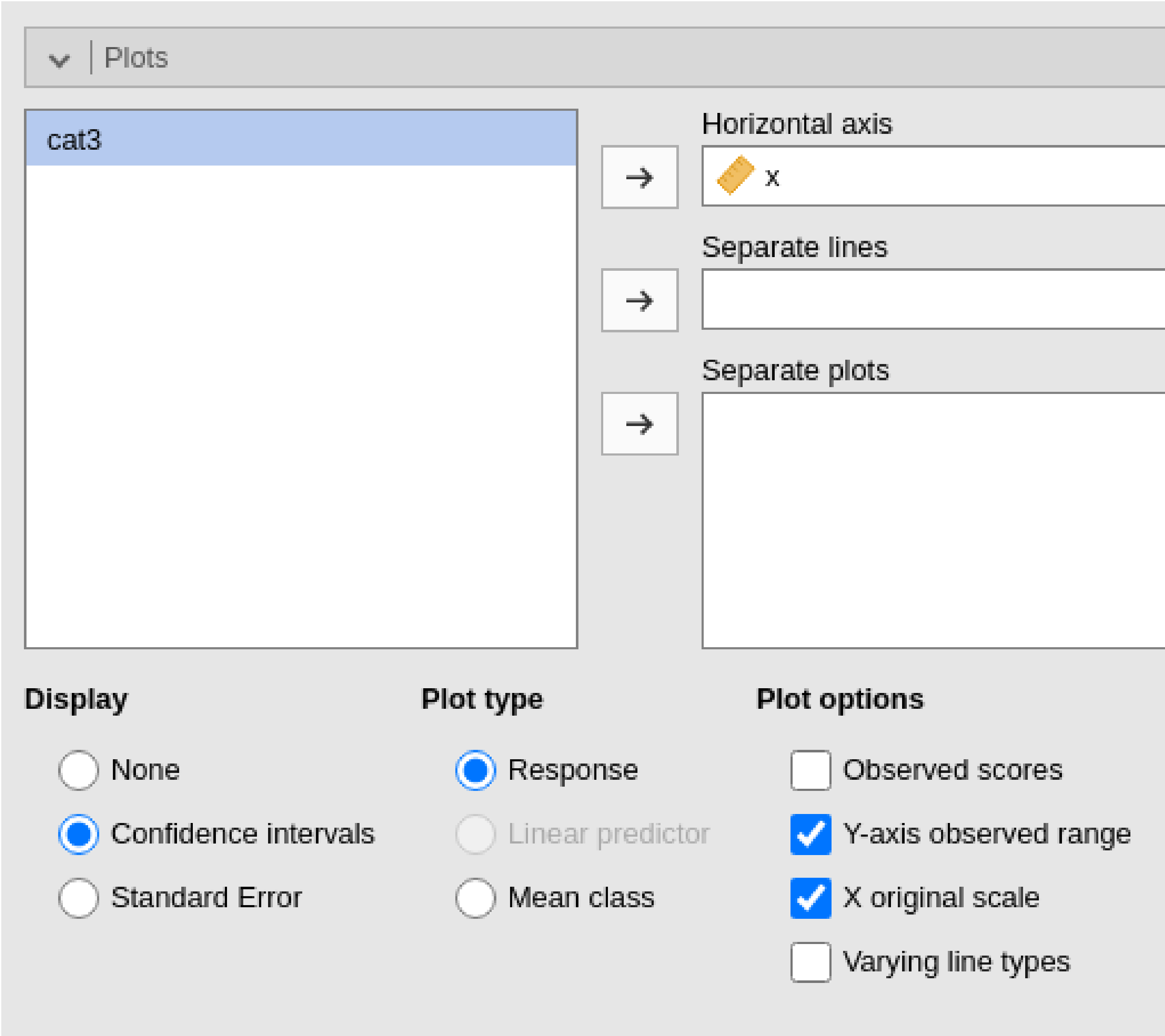
In our example, we can see that `IPA` group shows an odd of increasing the dependent variable level $Ratio = 1.88$ times larger than the `pilsner`, not significantly different, $z = 1.363$, $p. = 518$. The `stout` group shows an odd of increasing the dependent variable level $Ratio = 2.97$ times larger than the `pilsner` group, not significantly different, $z = 2.265$, $p. = 0.070$, whereas `stout` group shows an odd of increasing the dependent variable level $Ratio = 1.58$ times larger than the `pilsner`, again not significantly different, $z = 0.980$, $p. = .982$. We can notice that the same odd ratios were reported in the Parameter Estimates (Coefficients) table, with different p-values because no adjustment was operated there.

3.6.10 Plot

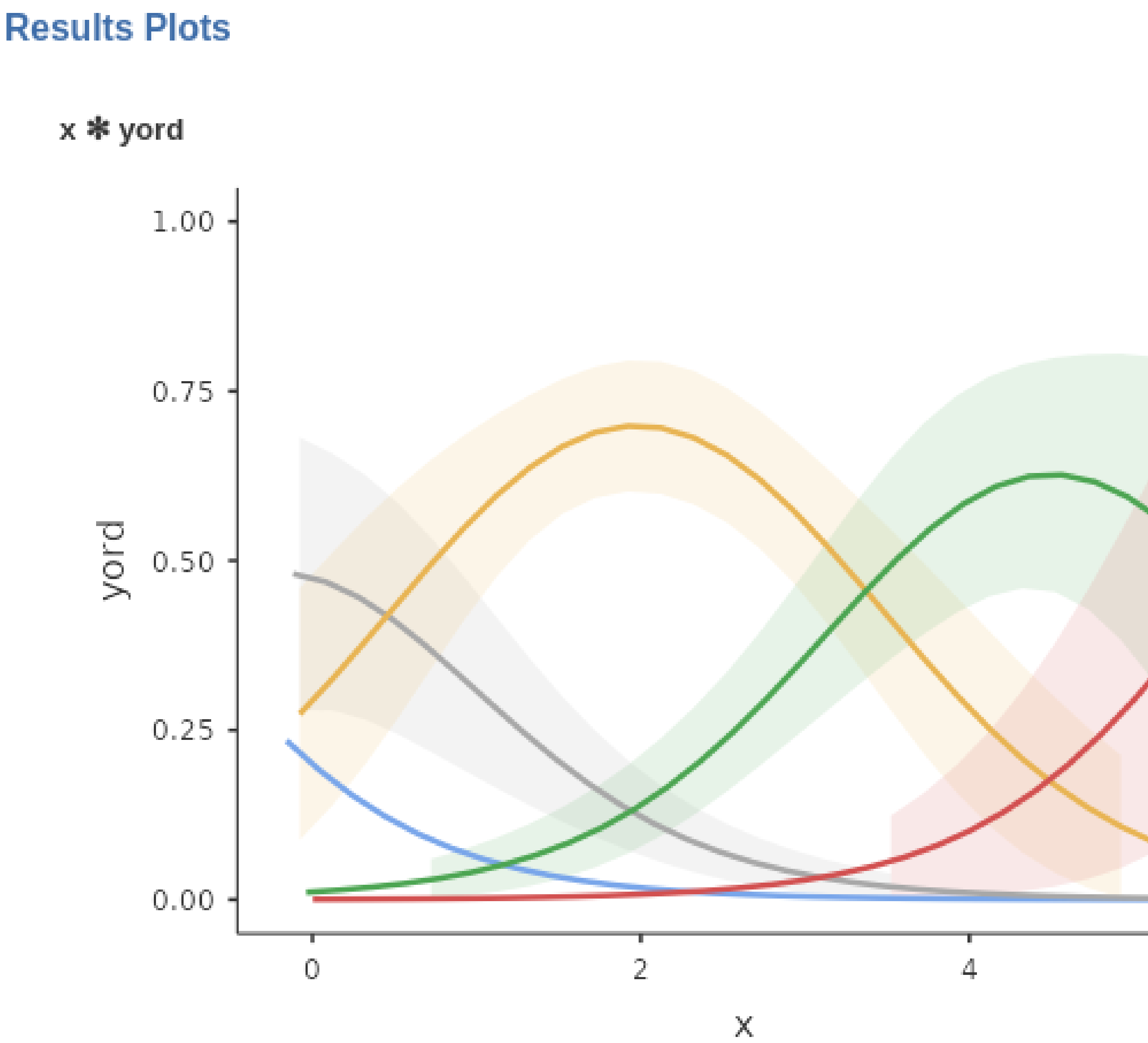
When dealing with generalized linear model, it is always a good idea to have a look at the effects in terms association of the predicted values with the independent variables. For ordinal models, we can choose two types of predicted values:

1) **Response**, the predicted values are expressed in probability of belonging to a level of the dependent variable. In our example, the plot diplays the probability of choosing any of the response level as a function of the independent variables.
2) **Mean class**, the predicted values are the expected classes, coded from 1 to K. In our example it would display the expected average classes as a funtion of the independent variable values. Let us see both versions.

The input is as usual.

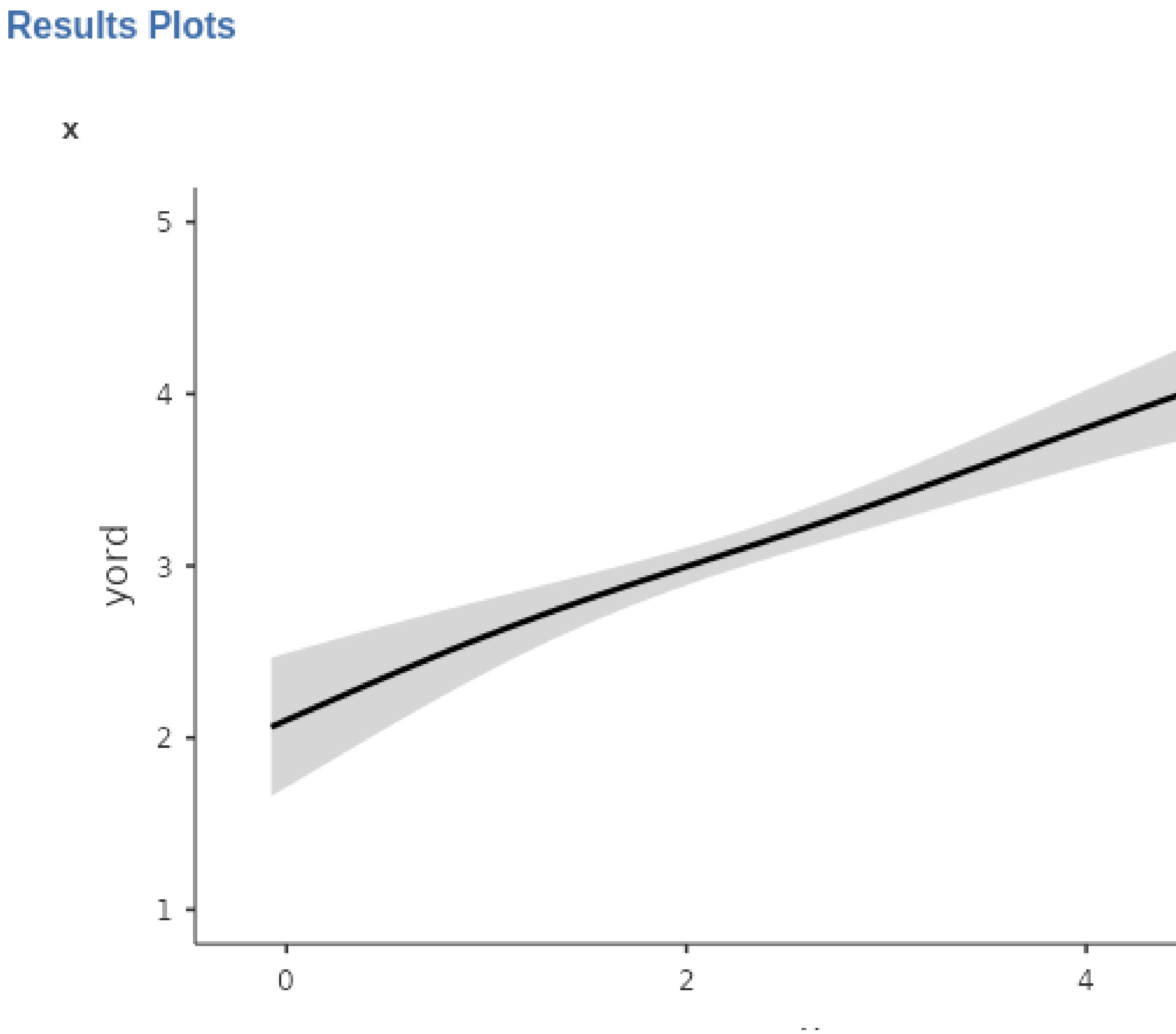


By default we have the Response option, so the type is the probability of being in any of the dependent variable level.



Here, we can track the probability function to observe the likelihood of different levels as we increase the independent variable. It is evident that the higher levels (4 and 5) become increasingly likely as we increase x, while the lower levels (1 and 2) become less probable.

A similar information can be obtained if we select Mean class in Plot type.



Here we see what is the average expected level as a function of x. For low values of x, participants pick level 2 on average, where as you increase x, the level chosen increases such that for high level of x, the average level is between 4 and 5.

Depending on the specific application, one of the two plot types may result useful.

3.6.11 Other options

All other options, Simple Interactions, Factor Coding, Covariates Scaling, Bootstrap confidence intervals, Model comparison etc. are the same as in the GLM.

3.7 Poisson Model

Very often, rather than measuring quantities we count stuff.

Work in progress: incomplete version

```
library(mcdocs)
```

Appendix A

The R^2's

Contrary to what many people think, almost all effect sizes and their corresponding inferential tests in the linear model's realm are based on some sort of model-comparison technique (Judd et al., 2017). The R^2 is one of them.

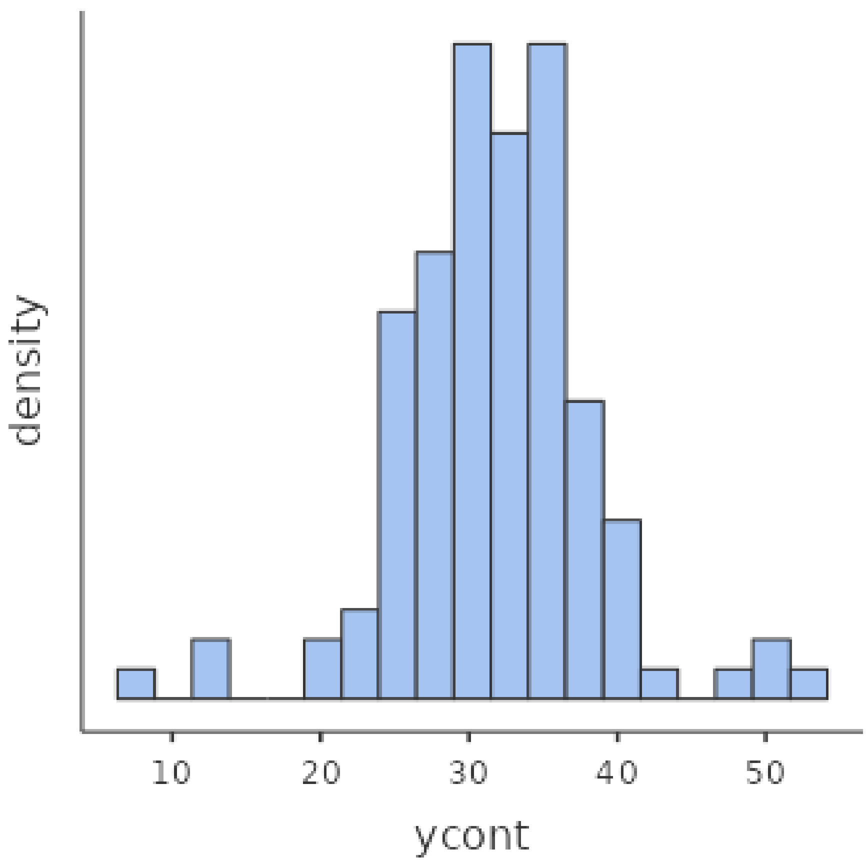
Assume you have a dependent variable that you want to model, meaning that you want to recap its scores by expressing them with some predicted values. For the moment, assume the variable to be continuous. If you only have the variable scores available, without any other information, your best guess is to use the mean as the most likely value. Keeping up with our toy example, assume that `ycont` in the dataset was the number of smiles for a given time done by our participants in a bar. If you only have the `ycont` variable, your best bet would be that the next customer will smile, on average, $\hat{y} = 31.7$ times, because that is the expected value (mean) of the variable distribution. So we say, whoever comes next in the bar, they will smile $\hat{y} = 31.7$ times on average.

Descriptives

	ycont
N	120
Missing	0
Mean	31.7
Median	31.7
Standard deviation	6.70
Variance	44.9
Minimum	8.03
Maximum	53.4

Plots

ycont



What I am saying translates into the most simple linear model, the mean model:

$$\hat{y}_i = \bar{y}$$

If we use the mean as our expected value, that is our model, we have an approximation error (σ^2), which amounts to the discrepancy between the predicted values (the variable mean) and the actual observed values. Because errors larger or smaller than the actual values are the same, we can square the errors, and compute the average error across cases (forget about the minus 1, it is not important now).

$$\sigma^2_{\bar{x}} = \frac{\Sigma(y_i - \hat{y})^2}{N-1}$$

When we associate an independent variable with a dependent variable in a linear model, we are seeking a better account of the differences in the scores of the dependent variable. It is like answering the question "how many smiles would a randomly taken person from our sample do?", with "it depends on how many beers they had". If it was only for the beers, the predicted values would be

$$\hat{y}_i = a + bx_i$$

and the error we make would be:

$$\sigma^2_r = \frac{\Sigma[a + bx_i - y_i]^2}{N-1}$$

How good is this error variance associated with the regression? Well, it depends on how big was the error without the regression, that is using the mean as the only predictor, namely $\sigma^2_{\bar{x}}$.

So we take the difference between these possible error variances, and we know how much the regression *reduced the error*

$$\frac{\sigma^2_{\bar{x}} - \sigma^2_r}{\sigma^2_{\bar{x}}} = R^2$$

The R^2 (and its variants) is the amount of error that we reduce in our predictions thanks to our model as compared to not using our model. In other words, is the comparison of our model error with the error of a model without the independent variable(s).

A.1 Commuting R^2

Let's assume you commute to the university every day and it takes 60 minutes (T_0) to get there from your home, following one route. A friend of yours (probably a know-it-all colleague), suggests an alternative route. You follow the lead, and you got to your department in 50 minutes (T_c). Nice, but what was your colleague's contribution to your happiness (assuming you do not enjoy commuting)? We can say that it was 10', which is given by $60 - 50 = 10$. Is that a lot? Well, it depends on the overall commuting time, because saving 10' from Lakeville to Berwick (they are in Nova Scotia, CA, 16 minutes apart) is different than saving 10' traveling from Milan (Italy) to Manila (Philippines), which takes around 17 hours. Thus, we can compute our colleague's contribution to our happiness as:

$$\frac{(T_0 - T_c)}{T_0} = \frac{10}{60}$$

which simply means that our colleague made us save 1/6 of our journey time. This is our colleague R^2. In statistical terms, we have the error variance without the model ($\sigma^2_{\bar{x}}$), the error variance of the model σ^2_r, and we have:

$$R^2 = \frac{\sigma_{\bar{x}}^2 - \sigma_r^2}{\sigma_{\bar{x}}^2}$$

which is how much our model "saved" of (or reduced) our error. That is called the *Proportion of Reduced Error* (Judd et al., 2017).

A.2 Variance explained

So, why is the R^2 index interpreted as *proportion of variance explained*? The reason is simply that a portion of the dependent variable variance can be associated with the variance of the independent variable(s), and thus we know why is there: because for a certain part (equal to R^2) people are different in the number of smiles because they are different in their number of beers.

You can get a broader view of this topic by consulting Judd et al. (2017). If you get excited by this, you can consult Searle and Gruber (2016), which explains that almost any test we are familiar with can be cast as a model comparison test.

Appendix B

How many contrasts?

B.1 Sufficiency of K-1 dummies

The fact that a linear model does not estimate all possible comparisons among levels of a categorical variable may appear puzzling. We have seen, in fact, that to represent a categorical variable with K levels, we only need $K-1$ contrasts and not $(K(K-1)/2)$, which will be the number of all possible comparisons. Let's see why.

A linear model requires as many coefficients as are necessary to compute the predicted values for all possible levels of the independent variables. This means that if I plug in the model a certain value of the independent variable, a sensible predicted value is produced. In simple regression $\hat{y_i} = 2 + 3 \cdot x_i$, for instance, if I plug $x_i = 4$, I get $\hat{y_i} = 14$, so it works. When you have a dichotomous independent variable, the model should work in the same way. For simplicity, assume the dichotomous IV is coded with the *dummy* coding system, so 0 vs 1, and the groups have the same N. The model is

$$\hat{y_i} = a + b \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The model should be able to produce predicted values for $x = 0$ and for $x = 1$. It is easy to verify that this happens without problems: $\hat{y_0} = a + b \cdot 0 = a$, the dependent variable mean for group 0, and $\hat{y_1} = a + b \cdot 1 = a + b$, the dependent variable mean for group 1. Now, if the IV has three groups, $x = \{1, 2, 3\}$, and I cast it with 2 dummies, I get this new model:

$$\hat{y_i} = a + b_1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Is this model able to produce three predicted values? First, for all dummies equal to zero, we have the expected value for group x=1, so $a = \bar{y}_1$. Therefore $b_1 = \bar{y}_2 - a = \bar{y}_2 - \bar{y}_1$ and $b_2 = \bar{y}_3 - a = \bar{y}_3 - \bar{y}_1$. Thus:

- for $x = 1$, the two dummies have values 0 and 0, so the predicted value is $\hat{y_1} = a + b_1 \cdot 0 + b_2 \cdot 0 = \bar{y}_1$.
- for $x = 2$, the two dummies have values 1 and 0, so the predicted value is $\hat{y_2} = a + b_1 \cdot 1 + b_2 \cdot 0 = a + b_1 = \bar{y}_1 + \bar{y}_2 - \bar{y}_1 = \bar{y}_2$.
- for $x = 3$, the two dummies have values 0 and 1, so the predicted value is $\hat{y_3} = a + b_1 \cdot 0 + b_2 \cdot 1 = a + b_2 = \bar{y}_1 + \bar{y}_3 - \bar{y}_1 = \bar{y}_3$.

No other coefficient or term is needed to account for all differences in the dependent variable due to independent variable. In fact, if a third dummy is inserted here, its coefficient must be necessarily zero, otherwise the predicted values will be biased, so the third dummy is redundant.

Often readers wonder if this works also for other coding systems. Sure it does, it is just a little more complicated to see it. Take the *deviation* method. Here the model for three-group IV is:

$$\hat{y_i} = a + b_1 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

For $x = 1$, the two dummies have values -1 and -1, so the predicted value is $\hat{y_1} = a - b_1 - b_2$. For $x = 2$, the two dummies have values 0 and 1, so the predicted value is $\hat{y_2} = a + b_1 \cdot 1 + b_2 \cdot 0 = a + b_1$. For $x = 3$, the two dummies have values 0 and 1, so the predicted value is $\hat{y_2} = a + b_1 \cdot 0 + b_2 \cdot 1 = a + b_2$. Are they the correct expected values?

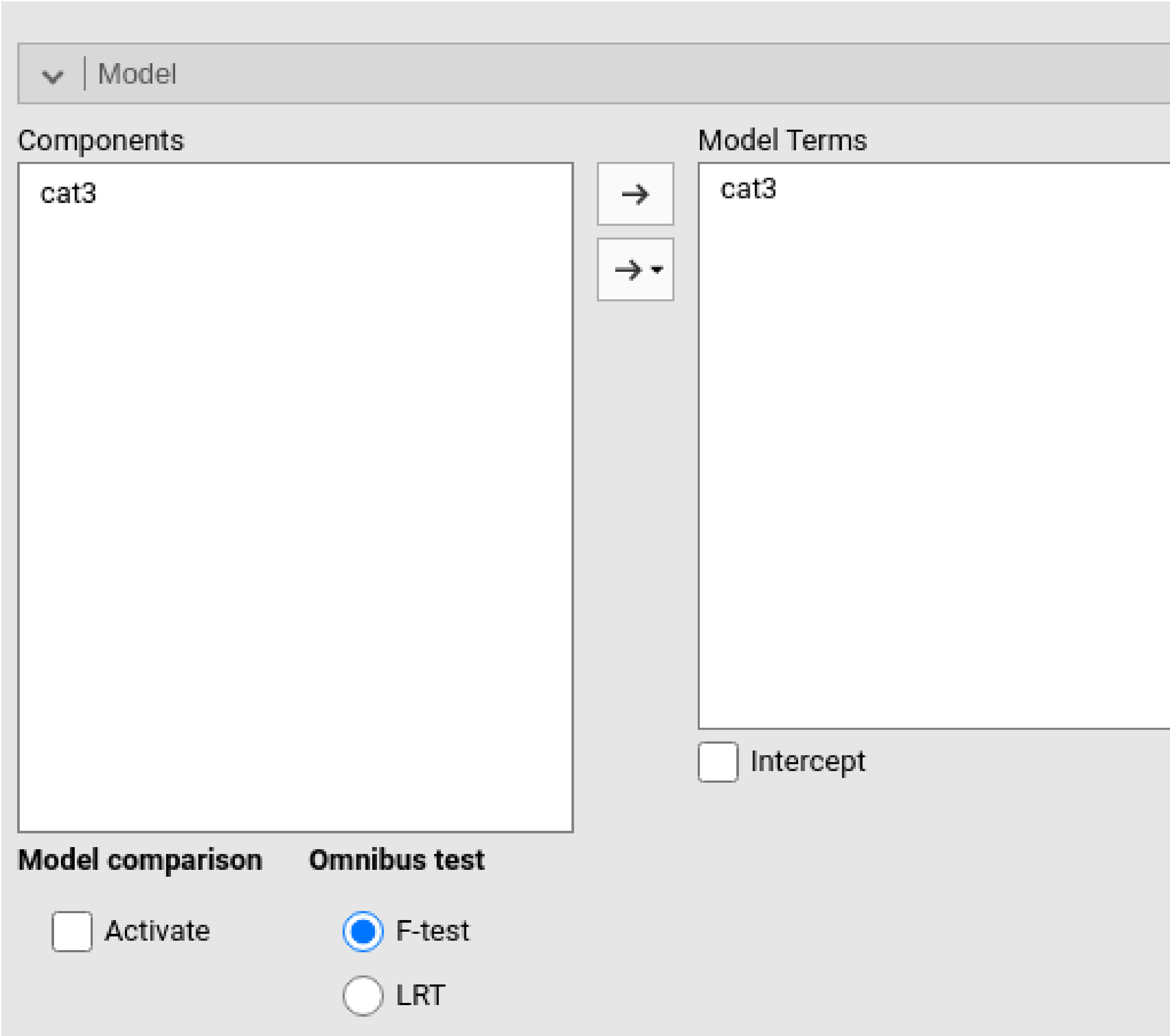
First, here the intercept a is the expected mean of the dependent variable, namely $(\bar{y}_1 + \bar{y}_2 + \bar{y}_3)/3$, because both contrast variables are centered to 0. Therefore, $b_1 = \bar{y}_2 - a$ and $b_1 = \bar{y}_3 - a$. So:

- For $x = 1$, we have $\hat{y_1} = a - b_1 - b_2 = 3a - \bar{y}_2 - \bar{y}_3 = \bar{y}_1 + \bar{y}_2 + \bar{y}_3 - \bar{y}_2 + \bar{y}_3 = \bar{y}_1$
- For $x = 2$, we have $\hat{y_2} = a + b_1 = a + \bar{y}_2 - a = \bar{y}_2$
- For $x = 2$, we have $\hat{y_3} = a + b_2 = a + \bar{y}_3 - a = \bar{y}_3$

One can show that it works for any coding system offered by GAMLj.

B.2 Zero-intercept ANOVA

The demonstration in Appendix B may explain also a curious phenomenon, which often surprises users of the linear model (independently of the software used). If one estimates a GLM with categorical IVs without the intercept, the model consists of K coefficients, where K is the number of groups, and the coefficients values are the means of the dependent variable for the groups. These models are called zero-intercepts models. Using `manymodels` data, a GLM with `cat3` as independent, `ycont` as dependent, and no intercept (we unflag the Intercept option), we get these coefficients:



It is clear that without an intercept, the three predicted values required here can be estimated by simply recasting the model as follows:

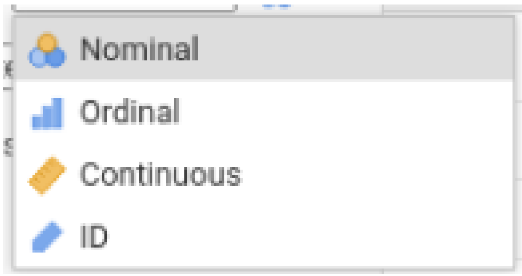
$$\hat{y}_i = \bar{y}_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + \bar{y}_2 \begin{bmatrix}0\\1\\0\end{bmatrix} + \bar{y}_3 \begin{bmatrix}0\\0\\1\end{bmatrix}$$

Almost any software does that automatically, including GAMLj.

Appendix C

Appendix Variable Types

jamovi classifies data variables in four classes:



- **Nominal** : categorical factor, it is passed to the R engine as a factor. Its behavior in jamovi interface depends on the `Data Type` property. We have
  - `Data Type: integer` can be inserted in the input field that permits numerical variable and nominal variables
  - `Data Type: text` can be inserted in the input field that permits nominal variables
  - `Data Type: decimal` it does not exist. Setting `Data Type` to `decimal` makes the variable a `continuous type`
- **Continuous** : numerical variable, it is passed to the R engine as a number. It can be input in the variable field that permits numerical variable. The data type property behaves like this:
  - `Data Type: integer` rounds the values to the closer integer
  - `Data Type: decimal` allows for floating points
  - `Data Type: text` does not exist, setting `Data Type` to `text` transforms the variable into a nominal variable
- **Ordinal** : numerical variable, it is passed to the R engine as an ordered factor. It can be input in the variable field that permits numerical and ordinal variables variable. The data type property behaves like this:
  - `Data Type: integer` can be inserted in the input field that permits numerical variable and nominal variables
  - `Data Type: text` can be inserted in the input field that permits nominal variables
  - `Data Type: decimal` does not exist. Setting `Data Type` to `decimal` makes the variable a `continuous type`
- **ID** : something cool which I do not know about.

Bibliography

Agresti, A. (2012). *Categorical data analysis*, volume 792. John Wiley & Sons.

Agresti, A. and Tarantola, C. (2018). Simple ways to interpret effects in modeling ordinal categorical data. *Statistica Neerlandica*, 72(3):210–223.

Aiken, L. S., West, S. G., and Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions.* sage.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Academic press.

Cohen, P., West, S. G., and Aiken, L. S. (2014). *Applied multiple regression/correlation analysis for the behavioral sciences.* Psychology press.

Efron, B. and Tibshirani, R. J. (1994). *An introduction to the bootstrap.* CRC press.

Glass, G. V., Peckham, P. D., and Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of educational research*, 42(3):237–288.

Hedges, L. V. and Olkin, I. (2014). *Statistical methods for meta-analysis.* Academic press.

Judd, C. M., McClelland, G. H., and Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond.* Routledge.

Jung, K., Lee, J., Gupta, V., and Cho, G. (2019). Comparison of bootstrap confidence interval methods for gsca using a monte carlo simulation. *Frontiers in psychology*, 10:2215.

Leeper, T. J. (2021). *margins: Marginal Effects for Model Objects.* R package version 0.3.26.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology.*

Mayo, D. G. (1981). In defense of the neyman-pearson theory of confidence intervals. *Philosophy of Science*, 48(2):269–280.

Midway, S., Robertson, M., Flinn, S., and Kaller, M. (2020). Comparing multiple comparisons: practical guidance for choosing the best multiple comparisons test. *PeerJ*, 8:e10387.

Neyman, J. (1957). " inductive behavior" as a basic concept of philosophy of science. *Revue De L'Institut International De Statistique*, pages 7–22.

Nimon, K. (2012). Statistical assumptions of substantive analyses across the general linear model: A mini-review. *Frontiers in Psychology*, 3.

Olejnik, S. and Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological methods*, 8(4):434.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, volume 1. sage.

Searle, S. R. and Gruber, M. H. (2016). *Linear models.* John Wiley & Sons.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684):677–680.

Wu, H. and Leung, S.-O. (2017). Can likert scales be treated as interval scales?—a simulation study. *Journal of Social Service Research*, 43(4):527–532.

Zou, G. (2004). A modified poisson regression approach to prospective studies with binary data. *American journal of epidemiology*, 159(7):702–706.