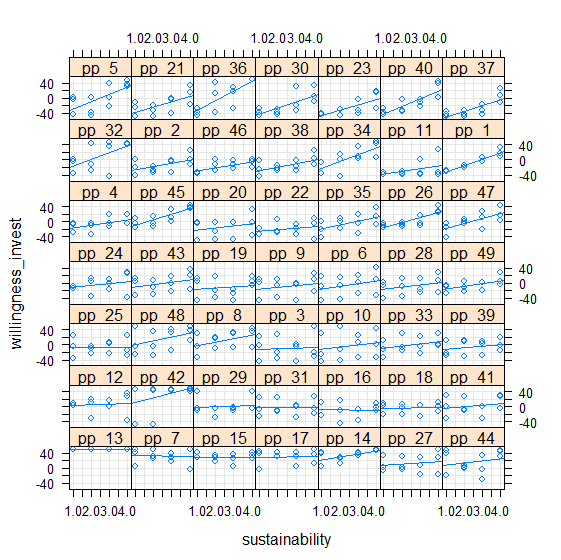
**Home assignment ("take-home exam"): Stats III Mixed-Models 2016**

**André Vargas**

**Introduction**

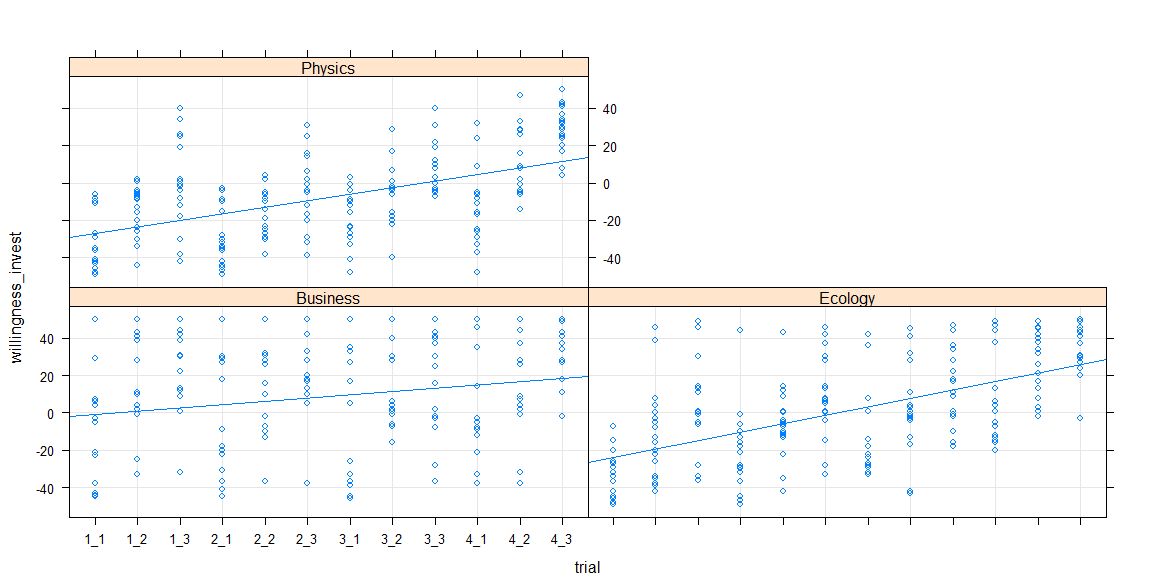
The willingness to invest data was analyzed with a mixed effects model, using the lmer function of the lme4 package (version 1.1.-10; Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2015). The data consists of five variables, sustainability of the investment, expected payoff, participant code, willingness to invest, and major which consists of 3 groups of different major programs. The amount of participants in the business major is 14, in the ecology major is 18, and in the physics major is 17, making a total of 49 participants.

In order to answer several research questions about the relationship of these variables to the willingness to invest of participants a mixed model effects was created. The model included all the possible per-participant random adjustment to the fixed intercept or “random intercepts”, as well as all the per-participant random adjustments to the slopes or “random slopes”, and all possible random correlation terms among the random effects. This is advice was taken from Barr, Levy, Scheepers, and Tily's (2013) to use a maximal random-effects structure. Specifically the analyzed maximal model contains two fixed intercepts: the participant code (pp\_code) to control for individual differences between sustainability and expected payoffs (e.g. see *Figure 1*), and the variable trial (trial) which was created using the two conditions (level of sustainability and expected payoff as it was tested in the experiment).



*Figure 1.* Individual differences on the willingness to invest as a function of sustainability levels.

The random slopes were selected according to their variability in their respect random intercept. Specifically, the participant code intercept (pp\_code) the slopes were the expected payoff (eprof10 and eprof20). These variables where transformed using the poly function in order to generate two orthogonal polynomial variables to test the possibility of a limit or non-linear effect in the model. Furthermore the variables where multiplied by ten in order to avoid convergence problems in the model. Additionally, the sustainability levels (sustain10 and sustain20) underwent the same transformation process as the expected payoff variable. For the second random intercept (trial) only the factor variable separating the majors in three levels (f\_major) was necessary (see *Figure 2*). This variable was coded using sum-to-zero contrasts, with physics participants coded as -1, and business and ecology participants as +1.



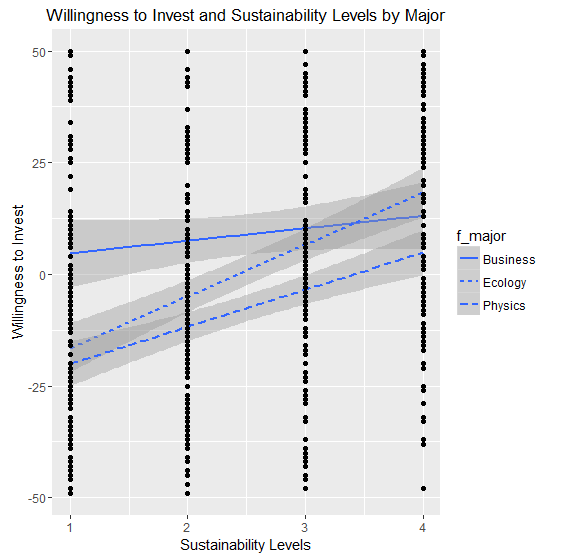
*Figure 2.* Variation of willingness to invest as a function of the different trials and participant’s major.

Finally the fixed variables of the model or independent variables where: the factor major variable, the modified version of both sustainability levels and expected payoff, and the interaction between the linear version of both and the participant’s major. As it can be observe in *Figure 3* a possible effect might appear form this interaction.

The maximal model was written as follows:

willingness\_invest ~ f\_major \* (sustain10 + sustain20) + f\_major \* (eprof10 + eprof20) + f\_major \* sustain10 \* eprof10 + (1 + eprof10\*sustain10 + sustain20 + eprof20| pp\_code) + (1 + f\_major|trial)

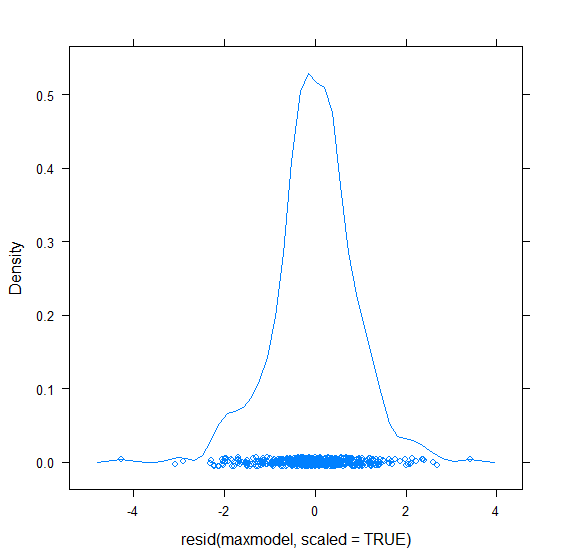
The “ \* ” means all the possible combinations of two variables e.g. A\*B = A + B + AxB.



*Figure 3.* Possible effect of sustainability levels and willingness to invest as a function of the participant’s Major.

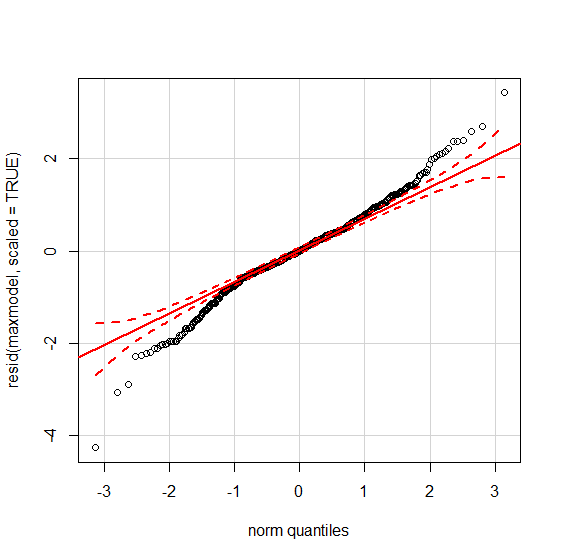
**Diagnostics**

Before looking at the model’s results some diagnostics where performed. First the proportion of scaled residuals was checked. 4.25 % of the residuals bigger or smaller than 2 standard deviations where found, 1.02% of the residuals bigger or smaller than 2.5 standard deviations where found, and finally 0.51% of the residuals bigger or smaller than 3 standard deviations. This means that there are a small amount of outliers bigger or smaller than 3 standard deviations (see *Figure 4*).



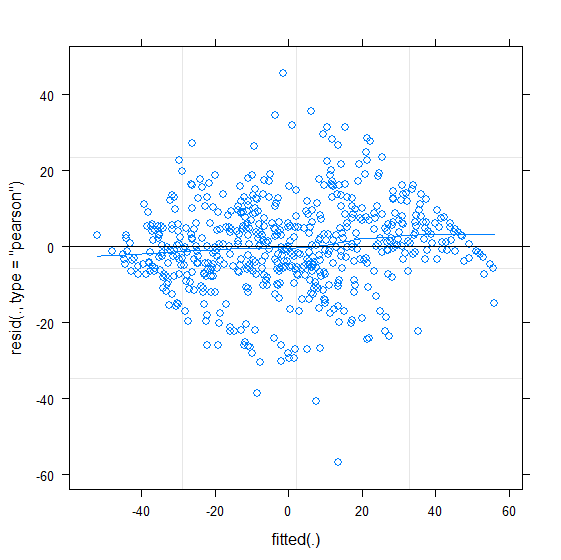
*Figure 4.* Density plot of scaled residuals.

Second, normality of residuals assumption was checked using a qqplot from “car” package (Fox & Weisberg, 2011). In *Figure 5* we can observe that there is non-normality which might become a problem because of the skewedness at both ends of the line.



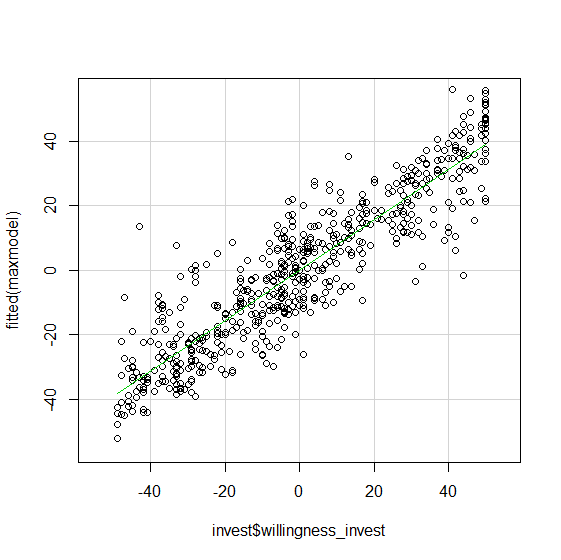
*Figure 5.* Normality of residuals plot.

Third, heteroscedasticity was checked, because in the maximal model it is proposed that there is the possibility of a non-linearity, some variables are transformed versions that represent a polynomial variable. As it is observed in *Figure 6*, the fitted vs. the residual values seem proportionally distributed and the fitted line is very close to zero. So we can assume that there is homoscedasticity.



*Figure 6.* Residual versus fitted values.

Finally, the observed vs. fitted values is plotted to see how well the model estimated the effects. *Figure 7* shows that mostly all the points are near the main regression line, which shows a good prediction of the variables.

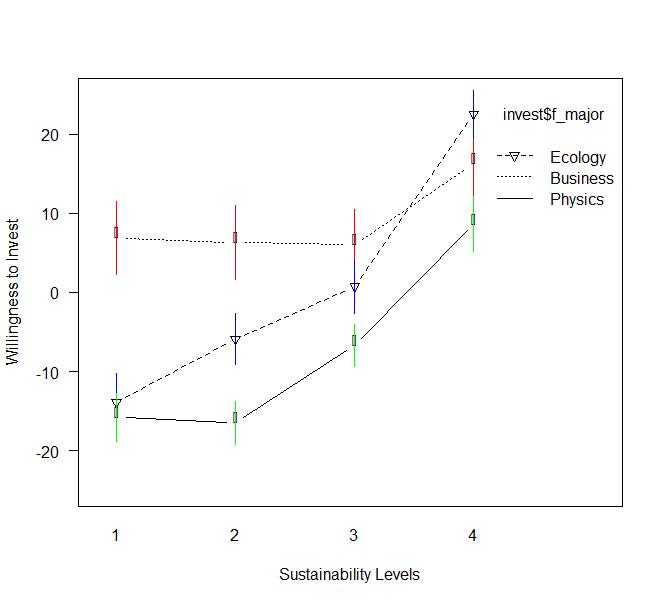


*Figure 7.* Observed versus Fitted Values.

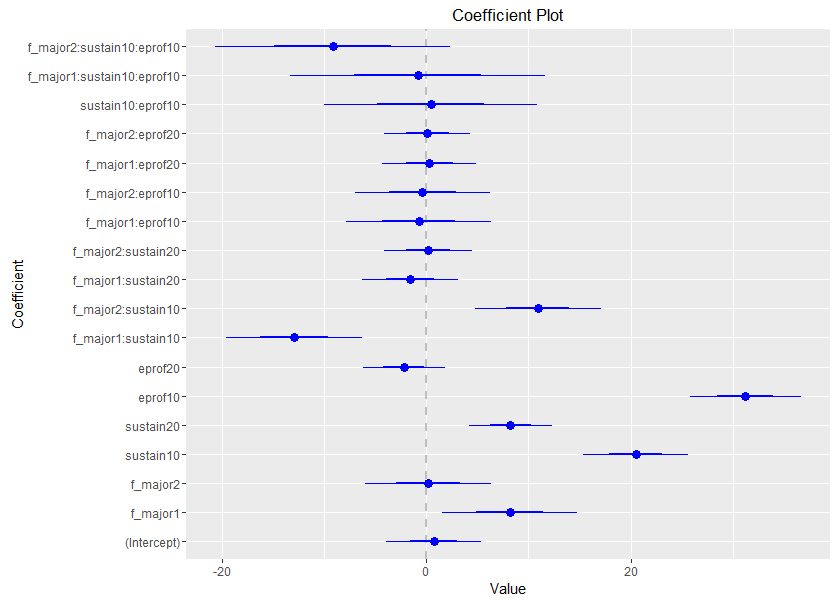
**Results**

The poly function generates variables that have small value and that may cause convergence problems. Therefore in order to solve convergence issues in the maximal model the polynomial variables where multiplied by ten after the poly function was applied to the variables. Moreover the number of iterations was maxed to increase the chances to find a convergent model.

To obtain p-values, we computed Type 3 Likelihood Ratio Tests as implemented in the mixed function of the package afex (Singmann, Bolker, & Westfall, 2015). Sustainability and willingness to invest showed a positive linear relationship, which means that willingness to invest increases as a function of the sustainability levels (Estimate = 20.49(2.23), χ2 (46) = 30.86, p < .001). The quadratic sustainability levels was also positive and significant, indicating an increasing effect and relationship with willingness to invest (Estimate = 8.22(1.61), χ2 (46) = 13.88, p < .001). Besides, the linear version of the variable expected payoff was positively related to the willingness to invest of participants and was significant (Estimate = 31.22(2.38), χ2 (46) = 40.72, p < .001). This means that willingness to invest increases as the expected payoff increases. Moreover, the quadratic version of expected payoff was found not significant (Estimate = -2.19(1.6), χ2 (46) = 1.75, p = .18). This means that there is only a linear relationship between these two variables. The interaction term between the linear relationship between sustainability levels and willingness to invest changes depending on the participant’s major (χ2 (46) = 16.62, p < .001) (see *Figure 8*), but the quadratic relationship between sustainability levels and willingness to invest did not differ as a function of the participant’s major (χ2 (46) = .75, p = .68). Besides, the interaction term between the participant’s major and the linear expected profit (χ2 (46) = .13, p = .93) does not affect the participant’s willingness to invest. Likewise, the quadratic version of expected profit and its interaction with the participant’s major (χ2 (46) = .03, p = .98) was not significant for the participant’s willingness to invest. The next interaction term between the linear version of both the sustainability levels and expected profit was not significant (Estimate = .41(4.2), χ2 (46) = .009, p = .92). This means that the relationship between willingness to invest and sustainability is not moderated by the expected profit. Furthermore, a three way interaction term between participant’s major, sustainability levels (linear) and expected profit (linear) was found not to be significant (χ2 (46) = 4.4, p = .11). This means that, as mentioned, there is no moderation effect and it doesn’t differ across majors. Finally, participants’ willingness to invest differ across the participants’ major (χ2 (46) = 8.38, p = .015). This significance and the significant interactions related to the factor variable lead to follow up models (to visualize all the estimated coefficients from the maximal model and their standard errors see *Figure 9*).



*Figure 8. This Figure shows the interaction effect from the linear relationship between sustainability and willingness to invest, and how it differs from majors. Standard errors are shown with the bars.*



*Figure 9.Estimated coefficients from the Maximal Model. This figure shows all the estimated coefficients and uncertainty (standard errors) from the maximal model variables.*

*Note: (intercept) = constant (grand mean), f\_major1= business major, f\_major2 = ecology major, sustain10 = polynomial orthogonal linear sustainability levels multiplied by ten, sustain20 = polynomial orthogonal quadratic sustainability levels multiplied by ten, eprof10 = polynomial orthogonal linear expected profit multiplied by ten, eprof20 = polynomial orthogonal quadratic expected profit multiplied by ten.*

In order to simulate pairwise comparisons six follow up models were made. The follow up models where used to assess how willingness to invest linear (significant) relationship with sustainability levels differs across majors.

The first model was made to compare business and ecology students, convergence problems were found. In order to solve them first we scaled the polynomial variables, which didn’t solve the problem, then we multiplied the variables by 10, unfortunately it didn’t help to resolve the issue. Furthermore the number of iterations was maxed and the optimizer “bobyqa” was selected given that it was previously tested that this optimizer can increase the chances to converge the max model presented above, but neither of those possible “fixes” resolved the convergence issue. Finally, following Bolker’s (2015) advice, by eliminating the highest random slope (linear expected profit and linear sustainability levels interaction slope) solved the convergence issue.

In this model the contrasts of the factor variable were also sum to zero (1 for business and -1 for ecology). The p-values were computed with a Type 3 Likelihood Ratio Tests as implemented in the mixed function of the package afex (Singmann, Bolker, & Westfall, 2015). The participant’s major was not significant, meaning that business (*M* = 8.8, *SD* = 29.65) and ecology (*M* = .88, *SD* = 27.77) students do not differ in willingness to invest (χ2 (31) = 1.6, p = .20). Furthermore, the linear relationship between sustainability and willingness to invest does differ across business and ecology students (χ2 (31) = 17.95, p < .001). The linear relationship of sustainability with willingness to invest level was positive and significant (Estimate = 15.74(1.94), χ2 (31) = 26.52, p < .001), and the quadratic (negative effect) also remained significant (Estimate = -1.63(1.54), χ2 (31) = 10.33, p = .001). This means that willingness to invest increases as sustainability levels, until it reaches an inflexion point where it increases even more. Furthermore the linear relationship between expected profit and willingness to invest remained significant (Estimate = 24.76(2.35), χ2 (31) = 36.38, p < .001), and the quadratic remained not significant (Estimate = -1.63(1.54), χ2 (31) = 1.09, p = .29). All the other variables and interactions remained non-significant.

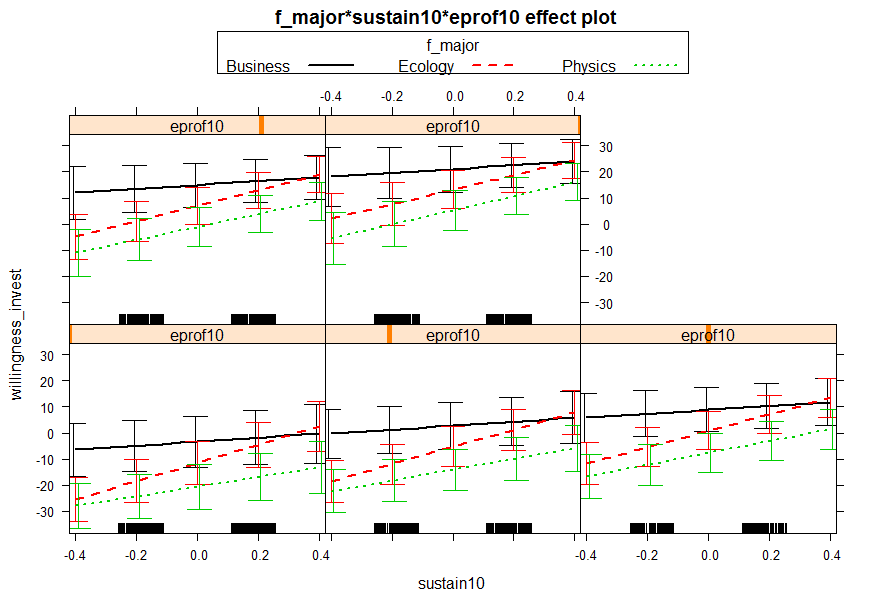
The second model was made to compare business and physics students, convergence problems were found. In order to solve them the same motivation and procedure was performed as in the second follow up model. Moreover, the p-values where obtained with the same method. The participant’s major was significant, meaning that business (*M* = 8.8, *SD* = 29.65) and physics (*M* = -7.59, *SD* = 23.52) students differ in willingness to invest (χ2 (31) = 7.05, p = .007). Furthermore, the linear relationship between sustainability and willingness to invest does differ across business and physics students (χ2 (31) = 8.2, p = .004). The linear relationship of sustainability with willingness to invest level was significant (Estimate = 11.96(2.1), χ2 (31) = 19.04, p < .001), and the quadratic also remained significant (Estimate = 6.47(1.58), χ2 (31) = 10.37, p = .001). Furthermore the linear relationship between expected profit and willingness to invest remained significant (Estimate = 24.98(2.5), χ2 (31) = 35.13, p < .001), and the quadratic remained not significant (Estimate = -1.77(1.54), χ2 (31) = 1.24, p = .26). All the other variables and interactions remained non-significant.

The third model was made to compare ecology and physics students, convergence problems were found. In order to solve them the same motivation and procedure was performed as in the third follow up model. Moreover, the p-values where obtained with the same method. The participant’s major was significant, meaning that ecology (*M* = .88, *SD* = 27.77) and physics (*M* = -7.59, *SD* = 23.52) students differ in willingness to invest (χ2 (35) = 4.57, p = .03). Furthermore, the linear relationship between sustainability and willingness to invest do not differ across ecology and physics students (χ2 (35) = 2.49, p = .11). The linear relationship of sustainability with willingness to invest level was significant (Estimate = 22.77(2.3), χ2 (35) = 41.58, p < .001), and the quadratic also remained significant (Estimate = 7.64(1.37), χ2 (35) = 19.01, p < .001). Furthermore the linear relationship between expected profit and willingness to invest remained significant (Estimate = 26.7(2.04), χ2 (35) = 45.84, p < .001), and the quadratic remained not significant (Estimate = -1.97(1.42), χ2 (35) = 1.89, p = .16). All the other variables remained non-significant except the three way interaction between major linear sustainability levels and linear expected profit (Estimate = -6.86(3.36), χ2 (35) = 3.99, p = .045), meaning that the linear relationship between sustainability and willingness to invest is moderated by the linear predictor of expected profit only between physics and ecology students (see *Figure10*).

The fourth model was made to assess how the continuous predictors behave only for one type of major (business). Convergence problems were found and in order to solve them the same motivation and procedure was performed as in the fourth follow up model. Moreover, the p-values where obtained with the same method. The linear relationship between sustainability and willingness to invest is significant for business students (Estimate = 4.04(1.44), χ2 (23) = 5.9, p = .015), and the quadratic relationship is also significant (Estimate = 3.51(1.49), χ2 (23) = 4.55, p = .03). Furthermore the linear relationship between expected profit and willingness to invest remained significant for business students (Estimate = 16.28(2.7), χ2 (23) = 17.91, p < .001), and the quadratic was not significant (Estimate = -1.02(1.5), χ2 (23) = .43, p = .5). The linear relationship between willingness to invest and the interaction between sustainability and expected profit was not significant for business students (Estimate = --0.12(1.76), χ2 (23) = .005, p = .94).

The fifth model was made to assess how the continuous predictors behave only for one type of major (ecology). Convergence problems were found and in order to solve them the same motivation and procedure was performed as in the fourth follow up model. Moreover, the p-values where obtained with the same method. The linear relationship between sustainability and willingness to invest is significant for ecology students (Estimate = 19.02(2.3), χ2 (23) = 35, p < .001), and the quadratic relationship is also significant (Estimate = 5.08(1.5), χ2 (23) = 8.94, p = .002). Furthermore the linear relationship between expected profit and willingness to invest remained significant for ecology students (Estimate = 18.6(2.09), χ2 (23) = 35.18, p < .001), and the quadratic was not significant (Estimate = -1.2(1.7), χ2 (23) = .56, p = .45). The linear relationship between willingness to invest and the interaction between sustainability and expected profit was not significant for ecology students (Estimate = -3.2(2.1), χ2 (23) = .2.18, p = .13).

Finally, the sixth model was made to assess how the continuous predictors behave only for one type of major (physics). Convergence problems were found and in order to solve them the same motivation and procedure was performed as in the fourth follow up model. Moreover, the p-values where obtained with the same method as previous models. The linear relationship between sustainability and willingness to invest is significant for physics students (Estimate = 13.2(2.3), χ2 (23) = 18.07, p < .001), and the quadratic relationship is also significant (Estimate = 5.7(2.3), χ2 (23) = 13.08, p < .001). Furthermore the linear relationship between expected profit and willingness to invest remained significant for physics students (Estimate = 19.07(2), χ2 (23) = 37.36, p < .001), and the quadratic was not significant (Estimate = -1.5(1.12), χ2 (23) = 1.51, p = .21). The linear relationship between willingness to invest and the interaction between sustainability and expected profit was significant for physics students (Estimate = 3.62(1.6), χ2 (23) = 4.66, p = .03). This means that the linear relationship between sustainability and willingness to invest is moderated by the linear predictor of expected profit only for the physics major (see *Figure10*).



*Figure 10*. This graph shows the three way interaction effect as the model-based means and confidence intervals as “error bars”. Only for physics the slope changes as the linear expected profit changes and linear sustainability increases as a function of the willingness to invest. Furthermore this moderation effect is also significant between ecology and physics.

**References**

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language, 68(3), 255-278.  
  
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1 – 48.

Fox, J. (2003). Effect Displays in R for Generalized Linear Models – The R Package effects. Journal of Statistical Software, 8(15), 1-27. Retrieved from: http://www.jstatsoft.org/v08/i15/.

Fox, J., & Weisberg, S. (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage.  
  
Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger Approximantion and Parametric Bootstrap Methods for Tests in Linear Mixed Models – The R Package pbkrtest. Journal of Statistical Software, 59(9), 1 – 30.

Lander, J. P. (2016). Plots Coefficients from Fitted Models– The R Package coefplot. Retrieved April 3, 2016, from https://cran.r project.org/web/packages/coefplot/coefplot.pdf

Lemon, J. (2006) Plotrix: a package in the red light district of R. R-News, 6(4): 8-12.

Lenth, R., V. (2016). Least-Squares Means: The R Package lsmeans. Journal of Statistical Software, 69(1), 1 – 33.   
  
R Core Team (2015). R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013. Retrieved from: [https://www.R-project.org/](https://l.facebook.com/l.php?u=https%3A%2F%2Fwww.R-project.org%2F&h=6AQFRBDqh)

Nieuwenhuis, R., Grotenhuis, M., & Pelzer, B., (2012). Influence.ME: Tools for Detecting Influential Data in Mixed Effects Models. R Journal, 4(2): pp. 38-47.

Revelle, W. (2015) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA. Retrieved from: http://CRAN.R-project.org/package=psych Version = 1.5.8.

Sarkar, D. (2008). *Lattice: Multivariate data visualization with R*. New York: Springer.

Singmann, H., Bolker, B., & Westfall, J. (2015). afex: Analysis of Factorial Experiments. R package version ¬0.15-2. Retrieved from: [https://CRAN.R-project.org/package=afex](https://l.facebook.com/l.php?u=https%3A%2F%2FCRAN.R-project.org%2Fpackage%3Dafex&h=6AQFRBDqh)

Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.