**Data Analytics & Visualisations**

**Instructions for the Assignment 2 – Statistical Report Context**

Micheál Martin has been impressed by your Tableau dashboard comparing the situation of Ireland and 9 other countries on the COVID-19 pandemic. The Taoiseach all the National Public Health Emergency Team (NPHET) has decided to promote your position as Senior Data Analyst.

**Task**

NPHET asked you to conduct an academic research with statistical analyses to understand the factors behind the situation of Ireland and 9 other countries regarding the COVID-19 data using the data included in the document “df\_country\_data.csv”1 that has been sent to you by email for the previous assignment. NPHET expects you to publish an academic research paper presenting your results.

For your statistical analyses, you can also use the files “df\_country\_metadata.csv” and “df\_global.csv” that has also been send to you for the previous assignment.

Your academic research paper should investigate **between 3 and 4 hypotheses**. Write an academic research paper, no more than **5 pages long**, including:

* -  A short **abstract**,
* -  A short **literature review** with no more than 5 references which ends with your hypotheses,
* -  A **method section** presenting your data. Your method section should present how the

hypotheses are tested with the corresponding model representation and equation,

* -  A **result section** including descriptive statistics and inferential statistics,
* -  A **discussion and conclusion** section which examines the hypotheses and their results,
* -  A **reference section** listing the academic papers used in your document.

The format of your academic research paper should **follow the style from one of three templates** proposed (see “template\_1.pdf”, “template\_2.pdf”, and “template\_3.pdf” attached).

This research paper can be based on the **same story that the one used in your previous tableau dashboard or be based on a completely different story**.

**Assessment Criteria**Obtaining statistically significant results (i.e., *p* < 0.05), rather than not significant results, **will not lead**

**to a higher mark**. The only assessment criteria are the following:

* **Accuracy of variables, hypotheses, model, and equation description,**
* **Conformity of statistical analyses and results’ interpretations,**
* **Relevance of Tables and/or Figures,**
* **Overall presentation style.**

**Report Submission**

1. **On the MT5000 Loop page, section Assignment.**
2. **In a .pdf file.**
3. **By Wednesday April 14 , 2021.**

**Note, any page after the 5th page will not be assessed (whatever it is in the appendix or in the main document).** Each data attributed to students are unique and will prevent copy-paste of results from another dataset attributed to another student.

Notes provided:

Hypothesis Testing with Jamovi

A guide for academic research

In academic research paper all sections are linked:

**Introduction ➡️ Literature Review ➡️ Method ➡️ Results ➡️ Discussion & Conclusion**

To understand the statistics in the results section it is essential to identify the concepts presented in each section:

1 Variables

A variable …

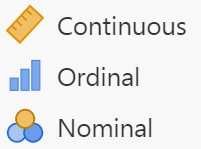
* Is way of assigning values (numbers or characters) to labels
* Corresponds to a column in a spreadsheet

Challenge: Identify the **Role** and the **Type** of each variable

1.1 Type of Variables

There are many type for variables but the only that we should care are … - **Continuous**: If values are numbers - **Categorical**: If values are characters

Note: Distinguish **Categorical Nominal** variables (*e.g.*, Irish, French) vs. **Categorical Ordinal** variables (*e.g.*, XS, S, M, L, XL)



Icons used in Jamovi to indicate the variable types

1.2 Role of Variables

A variable can have one or the other of these roles (no other role exist):

* **Outcome**: “to be explained” variable as Y (also called Dependent Variable or DV)
* **Predictor**: “doing the explaining” as X (also called Independent Variable or IV)

Note: A variable can be also both but in different hypotheses

2 Hypotheses

2.1 Correct Hypothesis Formulation

Hypotheses are:

* Predictions supported by theory/literature
* Affirmations designed to precisely describe the relationships between variables

A hypothesis test consists of a test between two competing hypotheses:

* An alternative hypothesis Ha (also called H1) formulated by the researcher
* A null hypothesis H0 (pronounced “H-naught”) counter to every alternative hypothesis

Note: For H0, there is no relationship between the variables. Ha is the “challenger” hypothesis, it claims the existence of a relationship.

Only 2 kind of alternative hypotheses can be formulated:

* **Main Effect Hypothesis**: Relationship between 1 Predictor and 1 Outcome
* **Interaction Effect Hypothesis**: Relationship between 2+ Predictors and 1 Outcome

Challenge: **Appropriate Formulation** the hypothesis according to the type of the Predictor

2.2 Main Effect Hypothesis Formulation

The **Outcome has to be Continuous** but …

* Case 1: Predictor is Continuous

The {**outcome**} increases when {**predictor**} {*increases/decreases/changes*}

* Case 2: Predictor is Categorical (2 Categories)

The {**outcome**} of {**predictor category 1**} is {*higher/lower/different*} than the {**outcome**} of {**predictor category 2**}

* Case 3: Predictor is Categorical (3 or more Categories)

The {**outcome**} of at least one {**predictor**} category is {*higher/lower/different*} than the other {**predictor**} categories

2.3 Interaction Effect Hypothesis Formulation

The **Outcome has to be Continuous** and **whatever the Predictor 1 is** …

* Case 1: Predictor 2 is Continuous

The effect of {**predictor 1**} on {**outcome**} is {*higher/lower/different*} when {**predictor 2**} increases

* Case 2: Predictor 2 is Categorical (2 Categories)

The effect of {**predictor 1**} on {**outcome**} is {*higher/lower/different*} for {**predictor 2 category 1**} than for {**category 2**}

* Case 3: Predictor 2 is Categorical (3 or more Categories)

The effect of {**predictor 1**} on {**outcome**} is {*higher/lower/different*} for at least one of {**predictor 2**}

Notes:

1. Predictor 1 and 2 are commutable (can be inverted and produce the same hypothesis)
2. An interaction effect hypothesis is also called moderation effect
3. By default, an interaction effect involves the test of the main effect hypotheses of all Predictors involved

3 Model & Equation

The basic structure of a statistical model is:

Outcome=Model+Error

where the Model is a series of predictors that are expressed in hypotheses related to the same outcome:

* Main effect hypotheses are indicated with the predictor name only
* Interaction effect hypotheses are indicated with all predictor names separated by ∗

Example:

Outcome=Pred1+Pred2+Pred1∗Pred2+Error

To evaluate their relationship with the outcome, each effect hypothesis is related with a coefficient called **Estimate** and represented with β as follow:

Outcome=β0+β1Pred1+β2Pred2+β3Pred1∗Pred2+Error

Note: β0 is the estimate related to the intercept. It is always included, always tested but has no interest in the analysis

3.1 Evaluation of the Significance

Testing for the significance of the effect means evaluating if this estimate β value is significantly **different, higher or lower than 0** as hypothesised in Ha:

* β≠0 means our hypothesis doesn’t precise the direction of the change, just that there is a change
* β>0 means our hypothesis indicates that the relationship increases or a group is higher than another group
* β<0 means our hypothesis indicates that the relationship decreases or a group is lower than another group

Note: H0 will always predict that β=0

The significance, called p-value, is the probability to consider H0 as True. This probability is between 0% and 100% which corresponds to a value between 0.0 and 1.0.

If the p-value:

* Is **higher** than 5% or 0.05, then H0 is **accepted**
* Is **lower** than 5% or 0.05, then H0 is **rejected** and Ha is considered as plausible

3.2 Graphic Representation of a Model

A graphic representation of the model’s hypothesised effects can be done: - All the arrows correspond to an hypothesis to be tested - All the tested hypotheses have to be represented with an arrow

**A simple arrow is a main effect**

PredictorOutcomeβ1

**A crossing arrow is an interaction effect**

Predictor 1β1OutcomePredictor 2β2β3

Note: By default, an interaction effect involves the test of the main effect hypotheses of all Predictors involved

4 Statistical Test

4.1 JAMOVI: Stats. Open. Now.

Jamovi an be downloaded or used online on <https://www.jamovi.org/>

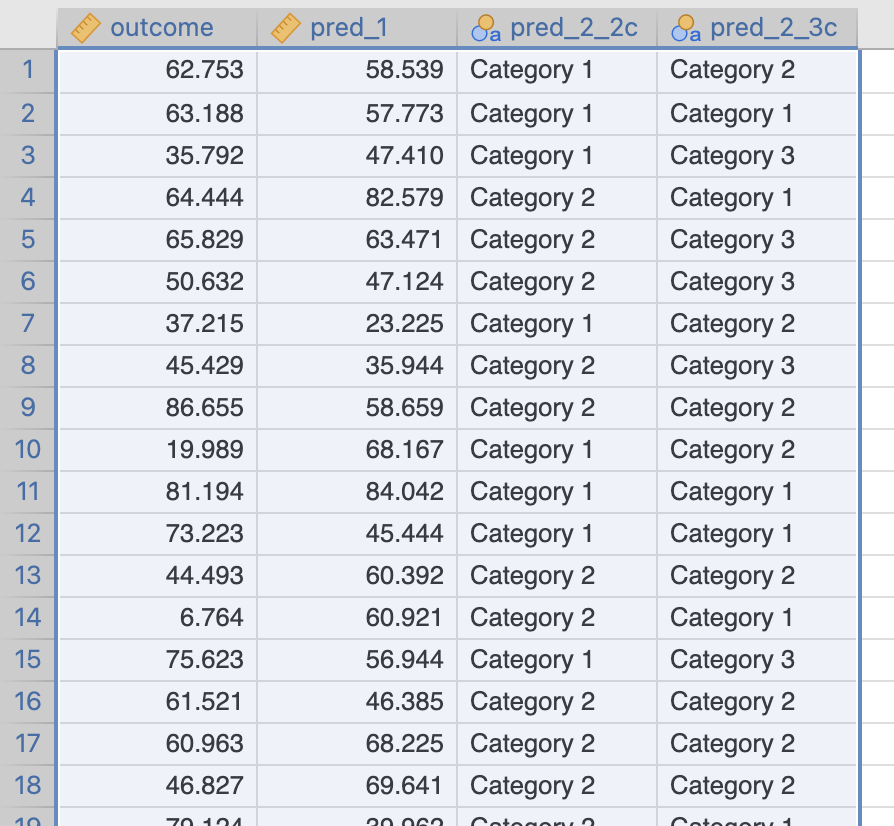
A free book “Learning Statistics with Jamovi” by Navarro and Foxcroft (2019) is available online here: <https://www.learnstatswithjamovi.com/>

Advantages:

1. Free
2. Simple Interface
3. No Missing Values to Declare
4. No Variable to Recode by Default
5. Ready to Publish Tables and Figures
6. Free Modules for Advanced Statistics (Mediation, Generalized LM, Linear Mixed Model)

Note: In Jamovi …

* The outcome is called Dependent Variable
* A continuous predictor is a covariate
* A categorical predictor is a factor



Example of data visualisation in Jamovi with continuous and categorical variables

4.2 Hypotheses with Continuous Predictors and with Categorical Predictors Having 2 Categories

Steps:

1. Open your file
2. Check the type of your variables
3. **Analyses** > **Regression** > **Linear Regression**
4. Set the Outcome as DV and

* **To test the main effect hypotheses**: set the Predictors as Covariates/Factors
* **To test interaction effect hypotheses**: In Model Builder, select all predictor with CTRL (win) or Command (mac) and bring them as interaction in the model

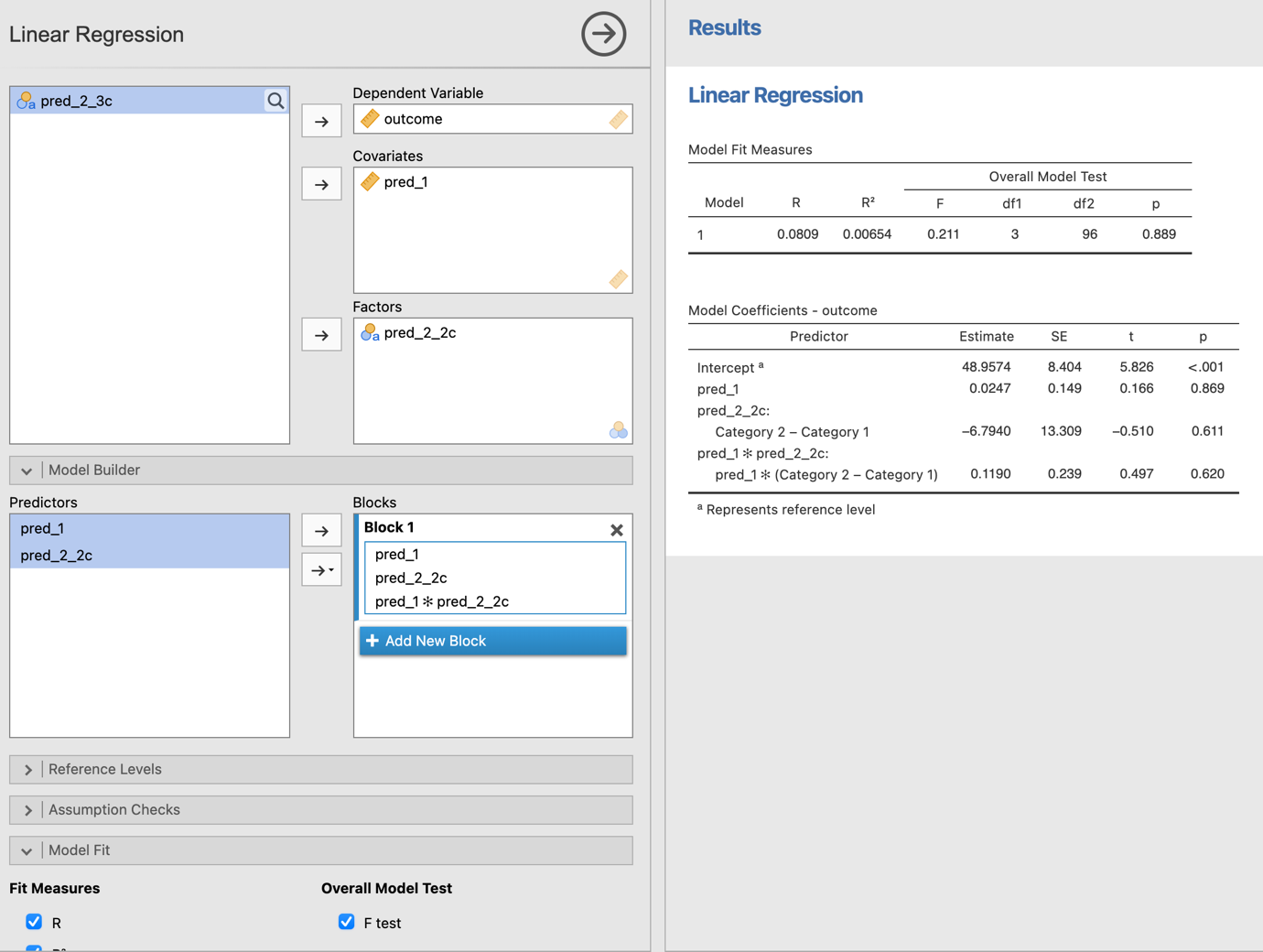
Communicate the Results about the full model and each hypothesis:

* Use **Model Fit Measure Table** to evaluate the accuracy of the full model

The predictions from a model including all effects are significant/not-significant better than without these effects ( R2=valueR2, F(df1,df2)=valueF, p=valuep)

* Use **Model Coefficients Table** to conclude about each hypothesis

The effect of Predictor on Outcome is statistically significant/not-significant, therefore H0 can be rejected/accepted ( b=valueestimate,95%CI[lowerCI,upperCI], t(df)=valuet, p=valuep).



Test of a model including 2 main effect hypotheses and 1 interaction effect hypothesis with a continuous predictor and a categorical predictor having 2 categories

4.3 Hypotheses with Categorical Predictors Having 3 or more Categories

1. Open your file
2. Check the type of your variables
3. **Analyses** > **Regression** > **Linear Regression**
4. Set the Outcome as DV and

* **To test the main effect hypotheses**: set the Predictors as Factors
* **To test interaction effect hypotheses**: In Model Builder options, select all predictor with CTRL (win) or Command (mac) and bring them as interaction in the model

1. Tick **ANOVA Test** in Model Coefficient options

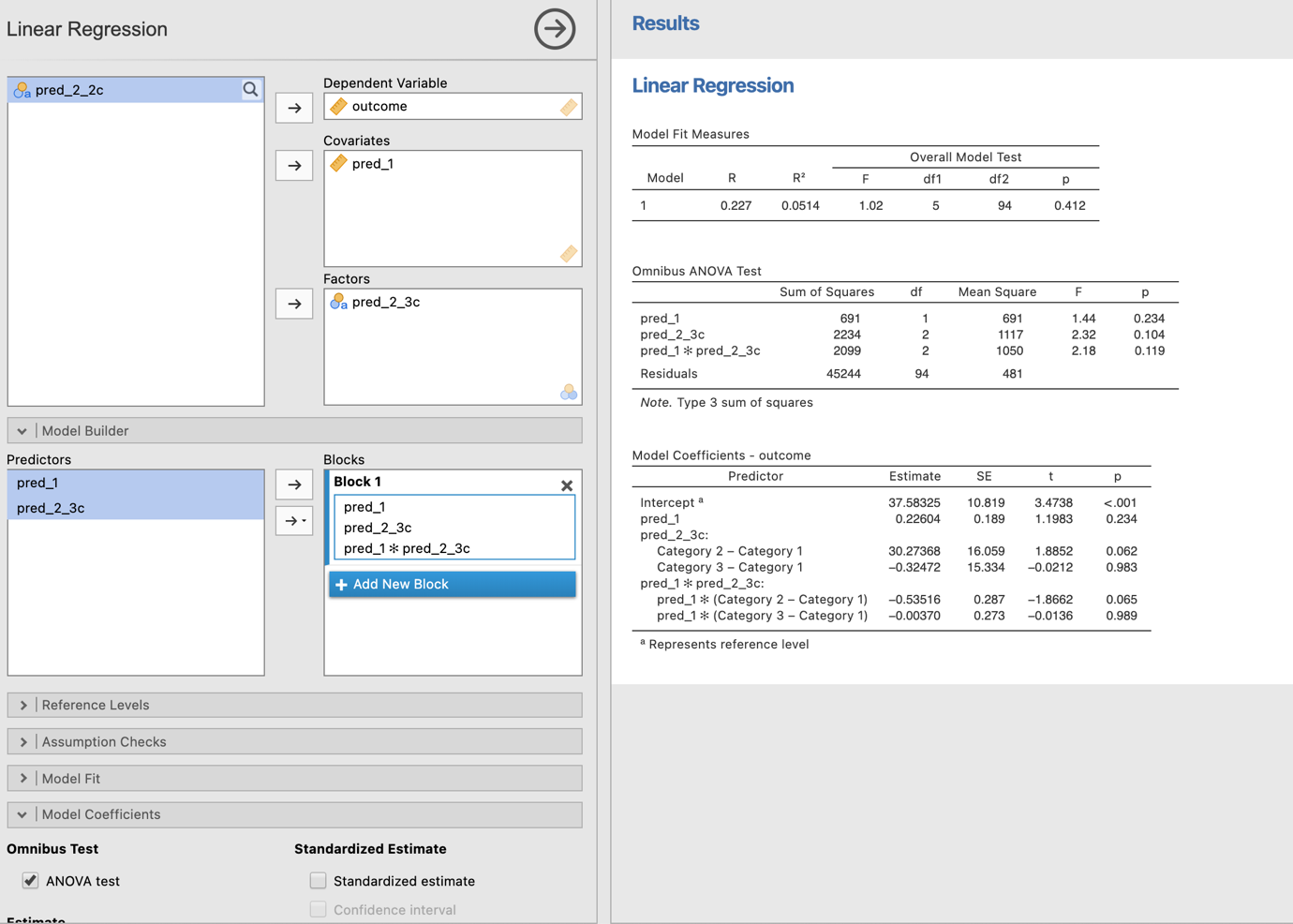
Communicate the Results about the full model and each hypothesis:

* Use **Model Fit Measure Table** to evaluate the accuracy of the full model

The predictions from a model including all effects are significant/not-significant better than without these effects ( R2=valueR2, F(df1,df2)=valueF, p=valuep)

* Use **Omnibus ANOVA Test Table** to conclude about each hypothesis

The effect of Predictor on Outcome is statistically significant/not-significant, therefore H0 can be rejected/accepted ( F(dfpredictor,dfresidual)=valueF, p=valuep).

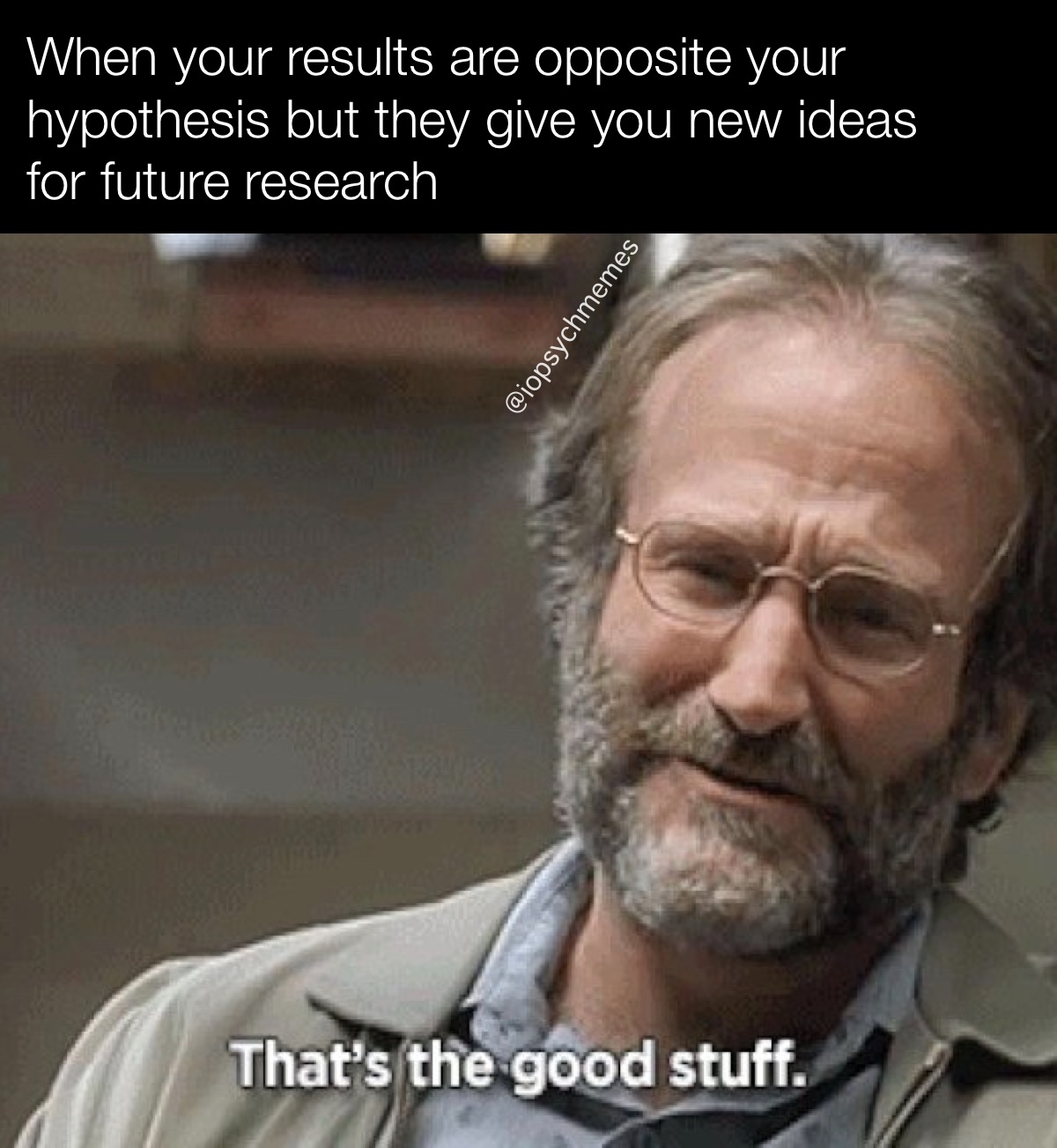


Test of a model including 2 main effect hypotheses and 1 interaction effect hypothesis with a continuous predictor and a categorical predictor having 3 categories

5 Discussion & Conclusion

From here…

* There is no number to be shown and no specific guidelines
* Correct interpretation comes if results have been understood and if reasons for the results to be the ones obtained have been identified



Source @iopsychmemes

Example of how to answer

Retaliation on a voodoo doll symbolizing an abusive supervisor restores justice  
A model paper based on simulation data to explain the principle of academic writing

**Abstract**

The present paper draws from the paper entitled “Righting a wrong: Retaliation on a voodoo doll symbolizing an abusive supervisor restores justice” (Liang et al., 2018). This academic article was published in the journal “The Leadership Quarterly” and has been awarded with the prestigious IG Nobel Price (section Business Studies). The paper looked at how subordinate retaliation moderates the relationship between abusive supervision and subordinate injustice perception. More precisely, the authors asked their participants to imagine an abusive or non-abusive supervision context and they measured how much using a virtual voodoo doll, representing the supervisor, would restore people’s sense of justice. By testing the hypotheses formulated

by its authors on simulated data, the present paper will describe each step of the academic writing.

**Introduction**

Abusive supervision is a problem for the mental well-being of employees. In cases of behaviours charactering abusive supervision such as such as public ridicule, yelling, scapegoating, or other forms of supervisor mistreatment, employees will have to deploy coping-mechanisms in order to preserve enough motivation to work in this negative environment. The premise of the paper Righting a wrong: Retaliation on a voodoo doll symbolizing an abusive supervisor restores justice is that “a natural response for the subordinate is to directly retaliate against the abusive supervisor” (Liang et al., 2018, p. 443). This premise is probably what made the paper published by Liang et al. (2018) an excellent candidate for the IG Nobel. However, this paper has an excellent structure and a very robust method. For these reasons, it will be used as example of how to write an academic paper.

The first step of an academic paper is the Introduction section which is supposed 1) to present the problem that the paper is trying to understand and 2) to end with the author’s research question. Here the research question of Liang et al. (2018) can be summarised as the following: How abusive supervision and subordinate retaliation influence the subordinate injustice perception?

**Literature Review**

The literature review is used to describe each concept included in the research question, independently, and in specific subsections. It aims to give a short and up-to-date presentation of the current state of the research involving these concepts with some example of results obtained in this research. Here, there are three concepts to investigate: abusive supervision, subordinate retaliation and subordinate injustice perception.

**Abusive Supervision**

The goal of the present paper is to describe the process of academic writing and not to investigate any kind of abusive supervision, subordinate retaliation or subordinate injustice perception. This subsection is only an example of structure.

1

**Subordinate Retaliation**

Sections in the literature review should present the state of the art of previous scientific research investigating these variables and their relationship. It is strongly recommended to use theories that have already been published and to describe how they are relevant to you. You can also use references from previous research to give some context to the evidence that has been found. This scientific evidence should support the formulation of your hypotheses.

**Injustice Perception**

The most important content of the literature review is the formulation of hypotheses. These hypotheses can be implicitly formulated in the body of the literature review but a common and efficient way to present the hypotheses is explicitly at the end of the literature review.

**Hypothesis 1**: *The average injustice perception in the condition of abusive supervision is higher than the average injustice perception in the condition of non-abusive supervision.*

**Hypothesis 2**: *The injustice perception decrease when the subordinate retaliation increase.***Hypothesis 3**: *The effect of subordinate retaliation on injustice perception in the condition of abusive*

*supervision is higher than in the condition of non-abusive supervision*

**Method**

The method section usually describes the participants/observations, the material used to collect the data, the procedure followed by the authors and the analyses performed.

**Participants/Observations**

In general, the participant section indicates the average age with standard deviation, the number of male and female participants (and other answers) and their origin. This selection also states how they were recruited or selected. Here, 2000 participants are created randomly.

**Material**

To manipulate abusive supervision, Liang et al. (2018) have asked their participants to imagine themselves in a condition of abusive supervision or non-abusive supervision. Consequently, this variable is categorical with two categories: abusive (N = 1015) and non-abusive (N = 985).

The method that Liang et al. (2018) used to measure subordinate retaliation is original and novel. They asked some of their participants to use a virtual voodoo doll as if they were the subordinate imagining that the doll was the abusive supervisor. Liang et al. (2018) examined whether participants used or not the virtual voodoo doll. However, for the purposes of this paper, it is considered how long participants are using the virtual voodoo doll (measured in seconds). As a result the variable is continuous which explains how Hypothesis 2 is formulated.

Liang et al. (2018) used a word completion task to measure subordinate injustice perception. In this task, five words have missing letters, and have to be completed so that they become either a neutral word or a negative word. The ratio of negative words used among the five words is supposed to reveal the injustice perception still felt by the subordinate. The measurement of subordinate injustice perception is simulated to have the same shape as the data obtained from this task.

**Procedure**

As indicated previous, the data are simulated.

**Data Analysis**

Here comes the main part of the method section. The absence of a Data Analysis section in an academic paper reveals its poor quality. However, by including both the graphic representation of the model tested and its corresponding equation, authors can display the robustness of their analyses.

The model is represented Figure 1 where AS is abusive supervision, IP is injustice perception, and SR is subordinate retaliation. It shows a classic moderation model which is the alternative name for interaction effect. A note of caution; a default moderation model includes not only the interaction-effect hypothesis but also the main effect hypotheses of the two predictors which is not obvious from the model representation.

Diagram

Description automatically generated with medium confidence

**Results**

The results section does not have to be long; it only needs to include some information about the variables, mainly their mean and standard deviation. This information can also include validity or reliability measures if not presented in the method section. Here, the average time spent using the virtual voodoo doll was 100s (SD = 20s), the shortest time spent was 28s and the longest 167s. However, 100 values are missing from the subordinate retaliation variable (voluntarily removed). The average proportion of negative words (from the word completion task) is 49.9% (SD = 28%).

Next, is the inferential statistics information, which has two parts: description of the overall model accuracy and test of each hypothesis. First, the model including both main effects of abusive supervision and subordinate retaliation as well as their interaction effect explains a significant part of subordinate injustice perception (Rˆ2 = 0.498, F(3,1896) = 626, p < 0.001). More precisely, the model explains 49.8% of the variance of injustice perception.

To interpret the statistics behind hypothesis testing, only the p-value is necessary. A p-value higher than indicates that the null hypothesis is true whereas a p-value lower than 0.05 indicates that the null hypothesis is rejected, and the alternative hypothesis considered as plausible. In the current simulated data, the results reveal a significant effect of abusive supervision on injustice perception (b = -0.09956, 95%CI[-0.19535,- 0.00378], t(1896) = -2.039, p = 0.042). This means that being in a non-abusive supervision context decreases, on average, by 9.95% the amount of negative words in the completion task. They also reveal a significant effect of subordinate retaliation on injustice perception (b = -0.00858, 95%CI[-0.00925,-0.00790], t(1896) = -24.969, p < 0.001). More precisely, for every second spent with the virtual voodoo doll, the amount of negative words in the completion task decreases by 0.8%. Finally the results did not reveal a significant interaction effect (or moderation effect) between abusive supervision and subordinate retaliation on injustice perception (b = -0.000213, 95%CI[-0.00115,0.000724], t(1896) = -0.445, p = 0.656) as show in Figure 2.

Chart, line chart

Description automatically generated

**Discussion and Conclusion**

The purpose of the current paper was to display the structure of a research paper rather than actually writing a paper. In this section, authors are presenting potential explanations for the effect/non-effect obtained. Limitations regarding the data collection and data analyses can also been added in these sections. Here, the data were randomly generated. Your task is to conduct your own discussion based on the results you obtain and the hypotheses which you formulate.