Dummy Coding in Regression

**Description:**

* Dummy coding is factorial coding that creates pairwise comparisons for categorical variables in regression.
* Dummy coded variables are created by creating individual comparison variables from the group that is coded first.
* You will get Levels – 1 comparisons for a dummy coded variable.
  + For example, if you had four levels: Catholic, Jewish, Protestant, Other.
  + You would get THREE predictors (X variables):
    - Catholic v Jewish
    - Catholic v Protestant
    - Catholic v Other
  + These variables are the same thing as pairwise post hoc t-test comparison.
  + What if you wanted Jewish versus Protestant?
    - You would need to recode the variables manually to get those comparisons.

**Dataset:** Dummy Coding

**IV:** Group health – excellent, fair, poor

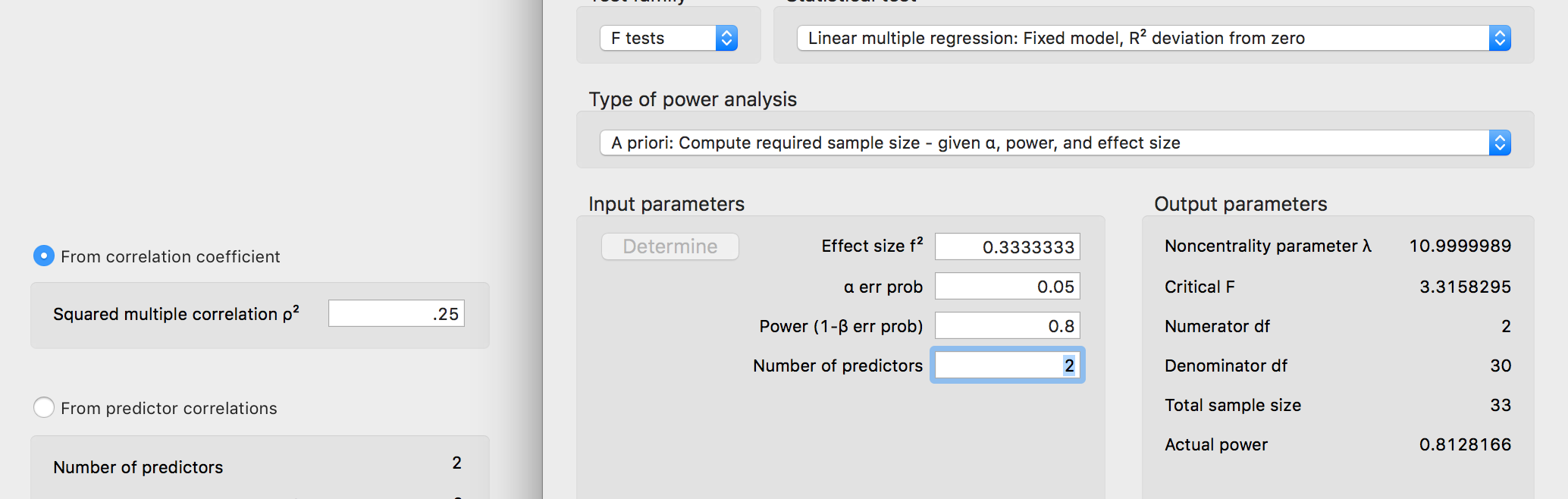
**DV:** Number of Friends

**Research question:** Does a participant’s health predict the number of friends they have?

**Regression type:** In theory, we only have one predictor (group health), but because it has multiple levels, we actually have a **set** of predictors (2 pairwise comparisons). Therefore, this analysis is a multiple linear regression – simultaneous.

**Power:**

1. Open Gpower!
   1. Test family: F-test
   2. Statistical Test: Linear multiple regression: fixed model, R2 deviation from zero.
      1. We are using multiple regression because we have more than one predictor.
      2. R2 deviation from zero indicates that we are interested in the overall model, rather than asking if the addition of more predictors to previous model are useful.
   3. Estimate an effect size: click determine 🡪 use R square sizes you think might be accurate, remember small, medium, and large estimates from the notes.
   4. Alpha = .05
   5. Power (1-beta .20) = .80
   6. Number of predictors: number of IVs/X variables.
2. Let’s estimate the following:
   1. Large effect size (*R2* = .25)
   2. Number of predictors: 2
3. Says we needed to run 33 people to find a significant effect with a large effect size.



**Assumptions are discussed in more detail in the multiple linear regression notes. This guide is for interpreting dummy coded variables more specifically.**

## Linear Regression

| **Model Summary** | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | | **R** | | **R²** | | **Adjusted R²** | | **RMSE** | | | **R² Change** | | **F Change** | | | **df1** | | **df2** | | **p** | |
| 1 |  | 0.750 |  | 0.562 |  | 0.453 |  | 1.568 | |  | 0.562 |  | 5.134 | |  | 2 |  | 8 |  | 0.037 |  |
|  | | | | | | | | | | | | | | | | | | | | | |
| *F*(2,8) = 5.13, *p* = .04, *R²* = .56  The predictors (group health, dummy code) significantly predicted number of friends. | | | | | | | | | | | | | | | | | | | | | |
| **ANOVA** | | | | | | | | | | | | | | | | | | |
| **Model** | |  | | | **Sum of Squares** | | | | **df** | | **Mean Square** | | | **F** | | **p** | | |
| 1 |  | Regression | |  | 25.24 | | |  | 2 |  | 12.621 | |  | 5.134 |  | 0.037 | |  |
|  |  | Residual | |  | 19.67 | | |  | 8 |  | 2.458 | |  |  |  |  | |  |
|  |  | Total | |  | 44.91 | | |  | 10 |  |  | |  |  |  |  | |  |
|  | | | | | | | | | | | | | | | | | | |

| **Coefficients** | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | |  | | **Unstandardized** | | | | | | **Standard Error** | | **Standardized** | | **t** | | **p** | |
| 1 |  | intercept |  | 1.500 | | | | |  | 0.784 |  |  |  | 1.913 |  | 0.092 |  |
|  |  | poorVfair |  | 3.833 | | | | |  | 1.198 |  | 0.845 |  | 3.201 |  | 0.013 |  |
|  |  | poorVexcel |  | 1.500 | | | | |  | 1.109 |  | 0.357 |  | 1.353 |  | 0.213 |  |
| Poor versus fair, *t*(8) = 3.20, *p* = .01  Poor versus excellent, *t*(8) = 1.35, *p* = .21 | | | | | | | | | | | | | | | | | |
| **Part And Partial Correlations** | | | | | | | | |
| **Model** | |  | | | **Partial** | | **Part** | |
| 1 |  | poorVfair | |  | 0.749 |  | 0.749 |  |
|  |  | poorVexcel | |  | 0.432 |  | 0.317 |  |
|  | | | | | | | | |