The analysis of continuous and categorical IVs

Transforming a continuous independent variable into a categorical variable with two levels – a procedure called "dichotomization" – leads to a decrease in statistical power. Assuming a normal distribution, the decrease of statistical power is equivalent to the exclusion of 37% of the participants. Said differently, the variance explained by a dichotomized independent variable corresponds to 64.7% of the variance explained by that same variable if it is treated as a continuous variable.

Mean deviation form

The first step in all analyses described below is to center the independent variables. Continuous variables are transformed in "mean deviation form".

\[ nv_i = ov_i - \overline{ov} \]

[where \( nv \) = new variable, and \( ov \) = old variable]

Categorical IVs are recoded in centered, orthogonal contrasts (so-called "contrast codes"). For example, a categorical IV with two levels is coded –1 and +1.

A continuous IV in the presence of a categorical between-participant IV:

Question: Is the beneficial effect of being in school stronger for kids who have already taken a trip outside the country? In other words, do the kids who have already traveled abroad progress faster in school than the kids who have never traveled abroad?

IV1: Years (number of years spent in school, continuous, between), from 0 to 8 ("YEARS")

IV2: Travel (categorical: has traveled abroad vs. not), coded 1 and 2 ("TRAVEL")

DV: Number of questions answers correctly ("CORRECT")

Hypothesis: An interaction between the two IVs.

Step 1: Center all independent variables
(a) Transform "YEARS" in mean deviation form (e.g., "YEARSC")
(b) Recode "TRAVEL" into –1 and +1 (e.g., "TRAVELC")

Step 2: Create the interaction variable by multiplying the two centered IVs (e.g., "INTERC" where INTERC = YEARSC * TRAVELC)

Step 3: Run a regression analysis

\[ CORRECT = b_0 + b_1\text{(YEARSC)} + b_2\text{(TRAVELC)} + b_3\text{(INTERC)} = 3.64 + .48\text{(YEARSC)} + .36\text{(TRAVELC)} + .27 \text{(INTERC)} \]

Interpretation: The coefficient \( b_1 \) tests what we call the main effect of "Years" in an ANOVA. The coefficient \( b_2 \) tests what we call the main effect of "Travel" in an ANOVA. The coefficient \( b_3 \) tests the interaction between the two IVs.

Step 4: To obtain the two regression equations (and the simple effects) do a "split file" and then "Organize output by groups" that are defined by the categorical variable (here: Travel). Then run a regression analysis in which the DV is regressed on the continuous IV in raw form.
Have not travelled:  \[ \text{CORRECT} = b_0 + b_1(\text{YEARS}) = 2.43 + .21 (\text{YEARS}) ; \]

Have travelled:  \[ \text{CORRECT} = b_0 + b_1(\text{YEARS}) = 1.00 + .75 (\text{YEARS}) ; \]

**Step 5:** To obtain a graphic representation of the data, do a scatter plot where the y-axis is the DV, the x-axis the continuous IV in raw form (here: YEARS), and colors are set by the categorical IV (here: TRAVEL). Then fit a line for each subgroup. Finally, it is suggested to add a vertical reference line at the mean of the continuous IV.

**A continuous IV in the presence of a categorical within-participant IV**

**Question:** Is the tendency to help a stranger more than a friend weaker for highly allocentric individuals?

**IV1:** Target of the help (categorical: friend vs. stranger, within)

**IV2:** Allocentrism (continuous, from low to high, between),

**DV:** Usefulness of the hints sent to the two targets

**Hypothesis:** An interaction between the two IVs.

**Step 1:** Transform "Allocentrism" in mean deviation from (e.g., "ALLOC")

**Step 2:** Create a new variable that corresponds to the difference between the two variables that constitute the DV (e.g., "DIFF" where DIFF = STRANGER – FRIEND)

**Step 3:** Run a regression analysis in which the difference score is regressed on the continuous variable in mean deviation form.

\[ \text{DIFF} = b_0 + b_1(\text{ALLOC}) = .06 + .06 (\text{ALLOC}) ; \]

**Interpretation:** The coefficient \( b_0 \) tests what we call the main effect of "Target" in an ANOVA. The coefficient \( b_1 \) tests the interaction between the two IVs.

**Step 4:** Create a new variable that corresponds to the average of the two variables that constitute the DV (e.g., "AVER" where \( \text{AVER} = (\text{STRANGER} + \text{FRIEND}) / 2 \))

**Step 5:** Run a regression analysis in which the average score is regressed on the continuous variable in mean deviation form.

\[ \text{AVER} = b_0 + b_1(\text{ALLOC}) = .45 + .05 (\text{ALLOC}) ; \]

**Interpretation:** The coefficient \( b_1 \) what we call the main effect of "Allocentrism" in an ANOVA.

**Step 6:** To obtain a graphic representation of the data, do an overlay scatter plot where each of the two DVs (STRANGER and FRIEND) is shown as a function of the continuous IV (here: ALLO) Then, fit a line for each of the DVs and add a vertical reference line at the mean of the continuous IV.