# PSY 653 Module 4: Categorical Predictors & Nonlinear Regression

## Part 1: Categorical predictors in regression

### Diet data set

This data set contains information on 78 people using one of three diets. The dataset is primarily used for ANOVA.

| Variable name | Variable | Data type |
|---|---|---|
| **Person** | Participant number | |
| **gender** | Gender, 1 = male, 0 = female | Categorical |
| **Age** | Age (years) | Continuous |
| **Height** | Height (cm) | Continuous |
| **preweight** | Weight before the diet (kg) | Continuous |
| **Diet** | Diet: 1 = control, 2 = Atkins, 3 = Nutrisystem | Categorical |
| **weight10weeks** | Weight after 10 weeks (kg) | Continuous |
| **weightLOST** | Weight lost after 10 weeks (kg) | Continuous |

1) Create a new project file and R notebook.
2) Create a first level header: "Load libraries"
    a) In an R chunk, Load up the following packages: psych, olsrr, & tidyverse
3) Create a first level header: "Import data"
    a) In an R chunk, Read in the datafile "Diet.csv"
4) Create a first level header: "Get dataset descriptives"
    a) In an R chunk, Get the dataset descriptives
5) Create a first level header: "Dummy code the Diet variable"
    a) Dummy code the Diet variable so that "control" (Diet == 1) is the reference category.
6) Create a first level header: "Run regression with dummy coded variables"
    a) In an R-chunk, Run a regression analysis so that weightLOST is regressed on your newly dummy coded Diet variable.
        i) In this model, what does the intercept represent?
        ii) Interpret the model
7) Create a first level header: "Run regression with effect coded variables"
8) Create a second level header: "Effect code the Diet variable"
    a) In a new R chunk, Effect code the Diet variable so that "control" (Diet == 1) is always coded as -1.
9) Create a second level header: "Run regression on effect coded diet variable"
    a) Run a regression analysis so that weightLOST is regressed on your newly effect coded Diet variable.
        i) In this model, what does the intercept represent?
        ii) Interpret the model

**Part 2: Nonlinear regression**

**DATASET:**

*We will use data compiled by The New Economics Foundation to compute the Happy Planet Index. The dataset represents 151 countries. We will explore a subset of data, all of the available data are here: http://www.happyplanetindex.org/. Dataset courtesy of Dr. Kim Henry.*

*We will model 3 variables:*

> *1. **gdp** – gross domestic product per capita—this is a measure of a country's wealth per person. Measured in 1000's (i.e. a value of 1.20 = 1,200)*
>
> *2. **wellbe** —using the 'Ladder of Life' from the Gallup World Poll, a random sample of residents in each country was asked to imagine a ladder, where 0 represents the worst possible life and 10 the best possible life, and report the step of the ladder they are currently standing on. The wellbe score for each country is the average of the responses for surveyed residents.*
>
> *3. **footp** —the ecological footprint of the country, expressed as global hectacres (gha) per capita.*

1) Create a new R notebook file and house it within your project file.
2) Create a first level header: "Load libraries"
    a. In an R chunk, Load up the following packages: psych, olsrr, & tidyverse
3) Create a first level header: "Import data"
    a. In an R chunk, Read in the datafile "happyplanet2.csv"
4) Create a first level header: "Get dataset descriptives"
    a. In an R chunk, Get the dataset descriptives
5) For this dataset, *wellbe* will be the outcome variable and *gdp* will be your predictor variable.
6) Plot the linear relationship between gdp and wellbe (Slide 32!)
7) Next plot the graph with a quadratic polynomial function (Slide 33!)
8) In a sentence or two, explain: Which plot seems to be a better fit to the data? Why?
9) Mutate the gdp variable so that you may assess the quadratic, and cubic effects (Slide 34!)
10) Run 3 regression models assessing the relationship of gdp and wellbe
    a. One assessing the linear effect (Slide 35!)
    b. Another assessing the quadratic effect (Slide 36!)
    c. And lastly, one assessing the cubic effect (Slide 37!)
11) Which model fits best? How do you know?