Data Science
HW7
Total: 20 points
Due: February 22nd, 2019, NOON

Assignment Paper: Gino et al., 2015

1. In your own words, specify the research question(s) (2 sentences or less) for **Study 1 through 4** [4]

2. Include and explain IVs, DVs for **Study 1 through 4** [4]

3. R section (please complete the following and **include your script and outputs**** as in-line text below)

   a. Reproduce **the power analyses conducted in Study 1 through 4** (they can all be found in the method section under "participants and design") [6]

4. *Challenge question*: Power Analysis for Chi-Square Tests

   Let's assume that you're looking at the relationship between ethnicity and promotion. You anticipate that 70% of your sample will be Caucasian, 10% will be African-American, and 20% will be Hispanic. Further, you believe that 60% of Caucasians tend to be promoted, compared with 30% for African-Americans and 50% for Hispanics.

   a. Complete the following table. ***Cells should show proportions*** [2]

Ethnicity	Promoted	Not promoted
Caucasian		
African-American		
Hispanic		

   b. What is the number of subjects needed to generate a significant level of 0.05, with desired power .90? (*include your R code, your output, and a conclusion!* see hints below*)* [4]

Hints:

Power analysis for chi square basic format R function:
pwr.chisq.test(w=, N=, df=, sig.level=, power=) where w is the cohen effect
size for Chi square, N is the total same size, df is degrees of freedom

$$w = \sqrt{\sum_{i=1}^{m} \frac{(p0_i - p1_i)^2}{p0_i}}$$

where $p0_i$ = cell probability in the ith cell under H_o
where $p1_i$ = cell probability in the ith cell under H_a,
m is the number of cells in the contingency table.

R functions used to calculate w:
ES.w2(P) where P is a hypothesized two-way probability table

P = matrix(c(R_1C_1, R_1C_2, R_2C_1, …), byrow = TRUE, nrow =…) where
R_1C_1 IS the cell value in row1column1, R_1C_2 IS the cell value in row1column2, …, nrow is
# of rows in table