William Eddy

Contingency Table Assignment

**Background**

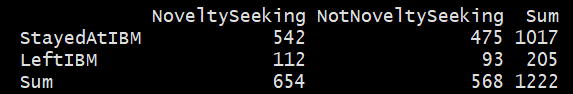
The data I used is a sample data set produced by IBM data scientists. It is designed to mimic employee data. I downloaded it from Kaggle.com at the following link:

https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset#

* Variable A is NoveltySeekingDichotomous, coded “low novelty seeking,” as “0” and “high novelty seeking” as “1.” I created it as follows:
  + Deleted all examples where value NULL for # of jobs worked
  + NoveltySeeking = (Age) / (# of jobs worked). This metric is low for high novelty seeking behavior.
  + For NoveltySeekingDichotomous, I split NoveltySeeking at the mean of 19.99, where scores above the mean qualify as “low novelty seeking,” coded as “0.” Scores below 19.99 qualify as “high novelty seeking,” coded as “1.” Note inverse relationship of variable to targeted attribute.
* Variable B is Attrition, coded “Did not leave IBM” as “0,” and “Did leave IBM” as “1.”

**Analysis**

53% of people who stayed at IBM exhibited the NoveltySeeking attrbute, and 54% of people who left of their own volition exhibited NoveltySeeking. This difference is not great enough to indicate that NoveltySeeking is a predictor of Attrition.



P(NoveltySeeking | StayedAtIBM) = 542/1017 = .5329

P(NotNoveltySeeking | StayedAtIBM |) = 475/1017 = .4671

P(NoveltySeeking | LeftIBM) = 112/205 = .5463

P(NotNovetySeeking | LeftIBM) = 93/205 = .4537

The conditional odds of Leaving IBM given the person is Novelty Seeking are 475/542. For a person who is not novelty seeking it is 93/112. The odds ratio is 475/542/(93/112) = 1.055. That is, the odds of a person staying at an IBM job are 1.055 times higher for people not exhibiting this novelty seeking pattern than for people who do exhibit this novelty seeking pattern.

My conclusion is, there is not a predictive relationship between the NoveltySeeking metric and Attrition.

**R Code utilized**

