Source Article:

Aiken, L. S., West, S. G., Reno, R. R., Woodward, C., & Reynolds. (1994). Increasing screening mammography in asymptomatic women: Evaluation of a second generation, theory-based program. *Health Psychology, 13,* 526-538.

Scenario:

The dataset call bcs.csv includes data on 164 women.  The data are from a study published by Aiken and colleagues (1994).  The point of the study was to predict whether or not a woman is in compliance with mammogram screening recommendations.

Data:

bcs.csv

The variables include:

1. comply -  1=in compliance, 0=not in compliance

2. physrec - 1=received mammogram recommendation from physician, 0=no recommendation

3. knowledge - knowledge of breast cancer and mammograms

4. benefits - her perceptions of the benefits of mammograms

5. barriers - her perceptions of the barriers to getting a mammograms

Directions:

1. Estimate a logistic regression model using physrec, knowledge, benefits, and barriers to predict compliance.
2. Calculate odds ratios for each predictor and 95% confidence intervals for the odds ratios
   1. Which predictors are significant?
3. Create a confusion matrix for this model. Use probability = 0.5 as the classification threshold.
   1. How many false positives are there?
   2. How many true positives?
   3. How many false negatives?
   4. How many true negatives?
4. Report the specificity of the model. What does this mean?
5. Report the sensitivity of the model. What does this mean?
6. Upload your R note book .rmd or .html file to Canvas.