PSYC 4541/5541: Assignment 1 (SDT EPT)

Lew Harvey, University of Colorado Boulder

23 January 2020

Assignment 1 (10 Points)

There are three parts to this assignment. Prepared your work using an R markdown document. Knit your Rmd markdown file to a pdf file and upload the pdf file to the course Dropbox in Canvas.

Part 1: Sensitivity and Bias

In 1977 the pharmaceutical company Warner-Lambert introduced the first in-home pregnancy test: Early Pregnancy Test or E.P.T. In the published clinical trials they reported the data given below in Table 1. Of the 487 pregnant women tested, the E.P.T. indicated that 451 were pregnant and indicated that 36 were not pregnant. Of the 198 non-pregnant women tested, the E.P.T. indicated that 15 were pregnant and that 183 were not pregnant.

Table 1: Number of 'no' and 'yes' responses for pregnant and not-pregnant woment in the 1977 clinical trials of the Warner-Lambert Early Pregnancy Test.

	'No'	'Yes'
Not Pregnant	183	15
Pregnant	36	451

Consider that the signal (s1) is pregnancy and the blank (s0) is non-pregnancy. Use the equal-variance signal detection theory model to determine the sensitivity (d-prime) and response bias (c) of the E.P.T test. Use Equations 9c and 12 in the Signal Detection Handout. Present your work in an orderly fashion by showing the transformations of the above response frequencies into the probabilities of the four possible outcomes (i.e., HR, MR, CRR, and FAR). Then transform these probabilities into z-scores (quantiles of the unit standard normal distribution). The R function `qnorm()` may be used to compute quantiles (z-scores) from probabilities. Compute the overall accuracy of the test (A_z) using Equation 14 (d_a equals d-prime in the equal variance model). Given the accuracy and bias of this test do you think it is a good test? Why?

Part 2: ROC Graph

Construct an ROC graph with FAR on the horizontal axis (the abscissa) and HR on the vertical axis (the ordinate). Set the axes to cover the range from 0 to 1 (`xlim(c(0, 1))` and `ylim(c(0, 1))`). Make the plot square by including `coord_equal(ratio = 1)` among your ggplot layers. See the help file for an example. Plot the HR and the FAR of the E.P.T. on the graph. Draw a smooth ROC curve through the (HR, FAR) pair using the fact that for the equal-variance signal detection model ROC is a linear function in z-score coordinates:

$$zHR = dprime + 1 \cdot zFAR$$

(see Equation 6b in the Detection Handout and the help file for details how to do this if you need to). Be sure to label the axes.

It should look like Figure 1, of course with your results added. The gray positive diagonal represents the hit rates and false alarm rates that would occur if the test had no ability to predict pregnancy: the hit rate equals the false alarm rate. The gray negative diagonal represents the hit rates and false alarm rates that could occur with an unbiased test: the hit rate and false alarm rate add to 1. The figure legend is created by the `fig.cap=` arguement in the code chunk header.

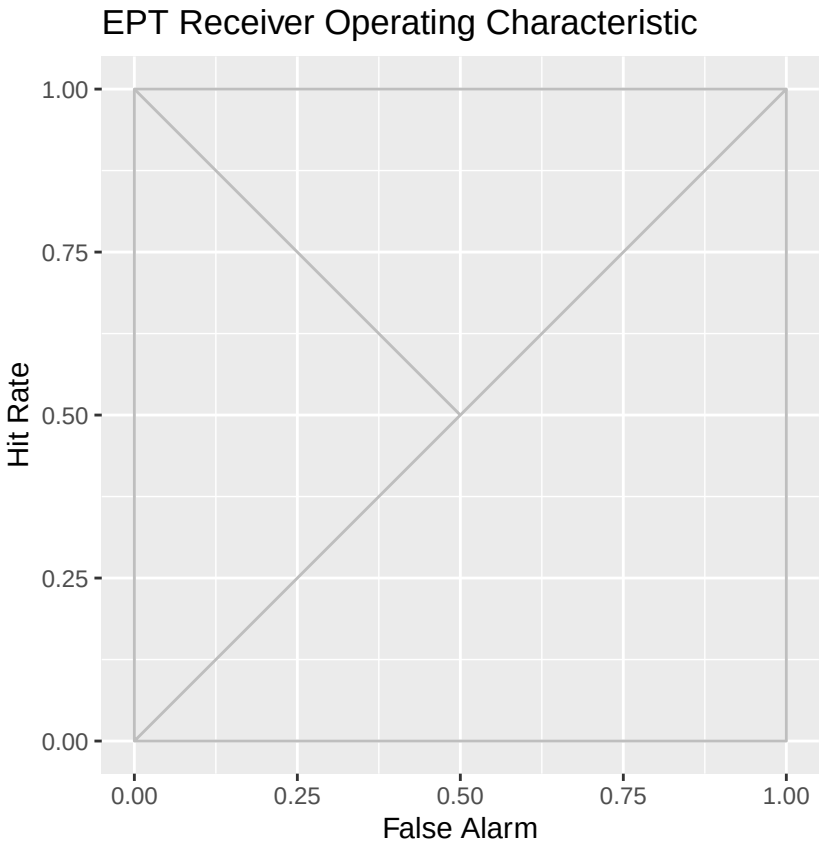

Figure 1: Receiver operating characteristic of the Warner-Lambert E.P.T. home pregnancy test. The filled circle is the resulting hit rate and false alarm rate computed from the clinical trials data. The smooth curve represents the predictions of the equal-variance signal detection model. The gray positive diagonal represents the hit rates and false alarm rates that would occur if the test had no ability to predict pregnancy. The gray negative diagonal represents the hit rates and false alarm rates that could occur with an unbiased test.

Part 3: Latent Normal Distributions

Make a ggplot graph of the two Gaussian probability density functions that represent the E.P.T. The mean and standard deviation of the not-pregnant distribution should be 0, 1; the mean and standard deviation of the pregnant distribution should be d-prime and 1 respectively. Make the non-pregnant distribtion black (the default) and the pregnant distribution blue. Put a vertical red line in the graph to mark the decision criterion, X_c. Hint: use `dnorm()` to get the probability densities. Your graph should look **something** like Figure 2, though this graph is not the correct on for the E.P.T. data.

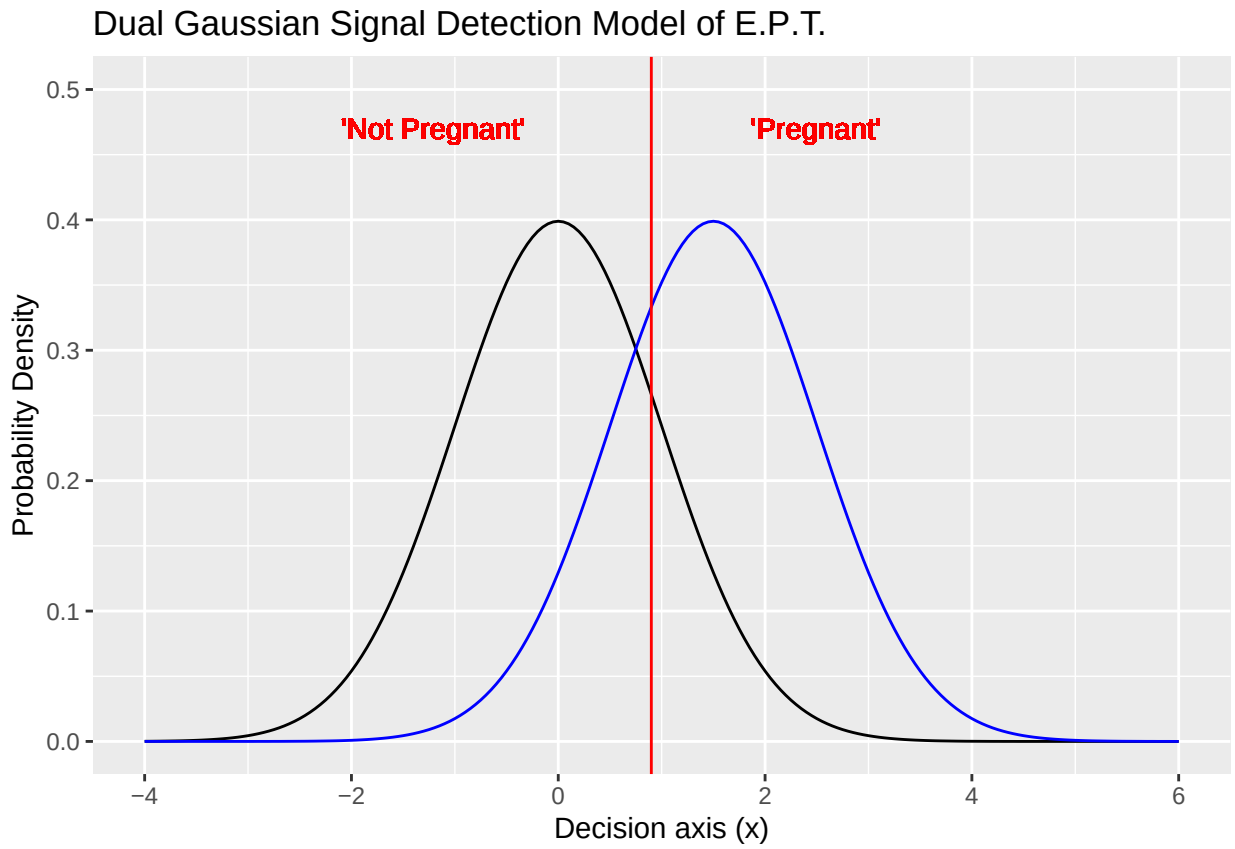

Figure 2: The dual-Gaussian signal detection model. The left distribution (black) represents the test response variability in the non-pregnancy condition and has a mean of 0.0 and a standard deviation of 1.0. The right distribution represents the variability of the pregnancy condition. It has a mean of d-prime and a standard deviation of 1.0. The vertical red line marks the critical value of x, the deciion criterion of the test. Responses above the red line are classified as 'pregnant'; those below the red line are classified as 'not pregnant'