PSYC 4541/5541: Assignment 1 (SDT EPT) Solution

Lew Harvey, University of Colorado Boulder

23 January 2020

Assignment 1 - 10 Points

Tips

The code below is broken in places. The r chunks that are broken and need fixing by you are labeled "broken". There are five broken chunks (broken1, broken2, broken3, broken4, and broken5). Fix the code in these chunks and you should get the correct results.

Within a broken code chunk, some lines are not broken and some are. The lines that are broken are indicated by a comment that has two comment hash marks, instead of the normal one. For example the code that is broken in r chunk `broken1` is marked like this:

Much of the text in this help file is to explain each step. Normally you would not include it all in a normal paper or homework report. Feel free to delete and/or change the text to suit your style of writing.

Part 1

There are three steps to take with the data before we can compute sensitivity (d-prime), accuracy (A_z) and response bias (c):

1. Enter the response frequencies into R
2. Convert the response frequencies into response probabilities
3. Convert the response probabilities into z-scores

The first step is to create a data frame containing the data from the assignment sheet. The code below shows how to create a data frame with two columns (the first labeled `no` and the second labeled `yes`). The data in the code chunk are not correct however. Make the data the same as in the assignment.

Create Data Frame

Table of Response Frequencies

Print out a table of the frequencies in the frequency data frame (`df.frequencies`) to make sure that they are identical to the table given in the pdf handout of Homework 1. `kable()` is an `r` function that provides really nice formatting of data frames. Do not continue with this assignment until the contents of data frame `df.frequencies` agree with the data provided in the assignment. The code chunk below will print a nice table of the data frame. This table should be **identical** to the one in the assignment.

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

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

Compute Response Probabilities from the Frequencies

In the second step, convert the frequencies to probabilities by dividing each row of the data frame by the total number of of responses in the row (i.e., not-pregnant women or and pregnant women). You will need to divide each row by the total of that row to convert the frequencies into probabilities. Examine the code in the next data chunk, It is correct and stores the results in a new data frame called `df.probabilities`.

Table 2: Response probabilities for the pregnant (s1) and the not-pregnant (s0) conditions.

	'No'	'Yes'
Not Pregnant	0.924	0.076
Pregnant	0.074	0.926

We see that the E.P.T. had a hit rate of 0.926 and a false alarm rate of 0.074. Make sure you understand which cell in the table is the hit rate and which one is the false alarm rate. They are both defined by the "yes" responses that are in column 2 of the table (and the data frame). Which cell corresponds to the hit rate and which cell corresponds to the false alarm rate?

Compute z-scores of Response Probabilities

In the third step, convert the probabilities to z-scores. The R function `qorm()` converts a probability into a z-score. Unfortunately the `qnorm()` function will not accept a data frame as input (data frames can contain non-numeric values), so you must first convert it to a matrix of numbers, apply qnorm and convert it back to a data frame again, as is shown in the code below. This code is correct except for the call to `kable`. Examine the previous code chunk and change the one below so that you print a table of the z-score values.

Table 3: Z-scores of response probabilities

	'No'	'Yes'
Not Pregnant	1.434	-1.434
Pregnant	-1.447	1.447

Sensitivity, Accuracy and Response Bias

Now we can compute the detection (d-prime), sensitivity (A_z) and bias (c) indices using the equations in the Detection Theory handout. In these equations there is one hit rate and one false alarm rate, so it will be convenient to extract these two values from the z-score data frame. You can extract a single cell of a data frame by providing index numbers within square brackets. For example to get the value stored in the first row and first column, you would use `df.zscore[1, 1]` (row number, column number). In the code below put in the proper index numbers to access the z-score of the hit rate and the z-score of the false alarm rate in the data frame. Add the code to compute sensitivity (`dprime`) and response bias (`cbias`) from Equations 9c and 12 in the Detection theory handout. If your code is correct you should get a positive value of dprime and a cbias very close to zero.

In the code chunk below you must write the code for Equation 9c (for dprime) and for Equation 12 (for cbias) that are given in the Detection Theory handout.

The d' for discriminating between pregnant (signal) and not pregnant (noise) for the E.P.T. is 2.881. The decision bias of the operator evaluating the detector is -0.006, very close to zero. The area under the ROC curve that corresponds to d-prime, A_z, is computed with Equation 14 in the Detection Theory handout. R function `pnorm()` provides the inverse transform $z^{-1}()$ referred to Equation 14. An A_z of 0.5 is chance performance. In r, you compute the square root of a number using the `sqrt()` function. The square root of 25 would be `sqrt(25)`, for example.

Table 4: Signal detection parameters for the Early Pregnancy Test

Measure	Value
d-prime	2.8814
Az	0.9792
c	-0.0065

Note: the line below shows how to insert the value of an r object right in your text using the back apostrophe character string. In this example A_z is rounded to 2 decimal places. The results show that the overall accuracy of the E.P.T. is 0.98.

Part 2: ROC Graph

The general approach to making a graph is to call the plot command and create the graph with the axes and the gray lines. Then use the point command to add the (hr, far) data point. Finally use the lines command to add the smooth curve. Go through this code chunk and read each comment for hints about how to make this chunk work properly. If all goes well, your ROC curve will appear in Figure 1.

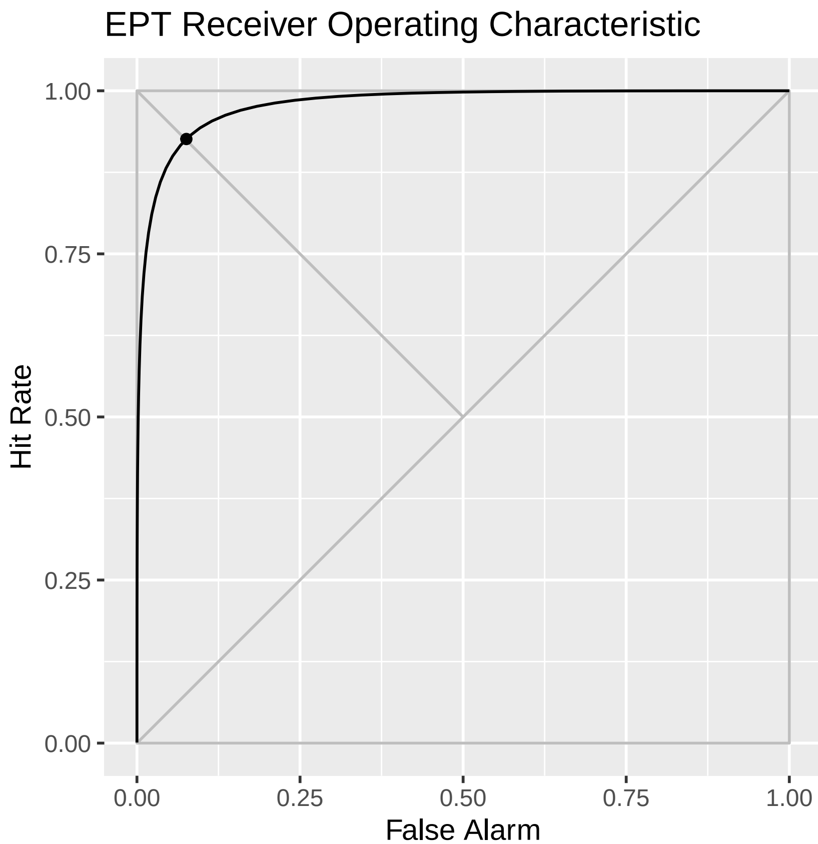

Figure 1: Receiver operating characteristic of the Warner-Lambert E.P.T. home pregnancy test. The filled circle is the resulting hit rate and false alarm rate computed from the clinical trials data. The smooth curve represents the predictions of the equal-variance signal detection model based on a d-prime of 2.88. The area under the ROC, a measure of accuracy, is 0.98. 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

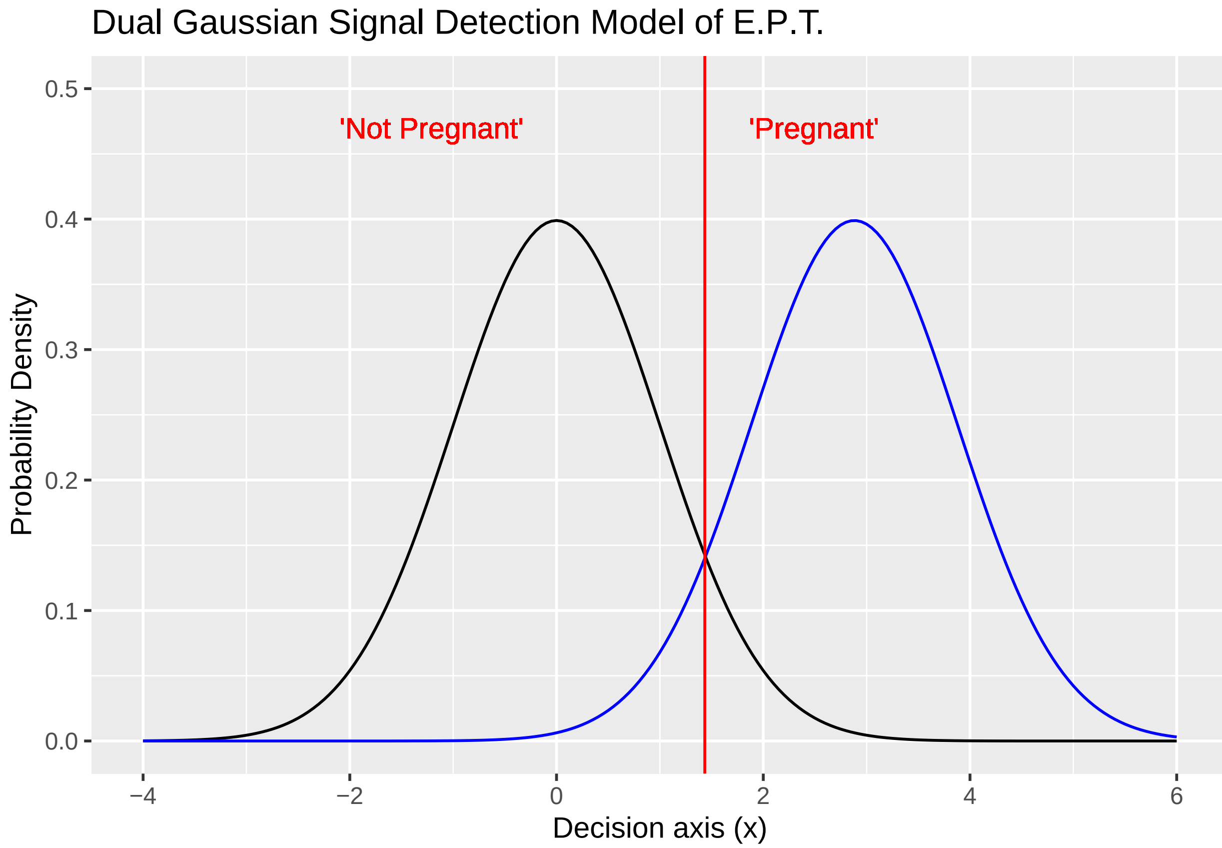

Figure 2: The dual-Gaussian signal detection model. The left distribution (black) represents the test response variability in the non-pregnancy condition and has a mean of 0.0 and a standard deviation of 1.0. The right distribution represents the variability of the pregnancy condition. It has a mean 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'.

General Conclusions about the E.P.T.

Based on the your computations of d', A_z, and c, what do you conclude about the E.P.T: is it a good or bad test? Why?