Prosem 2020: Project 1: Signal Detection

Lew Harvey, University of Colorado Boulder

3 September 2020

There are three parts to this project dealing with three sets of data suitable for a signal detection analysis. Use the *Project 1 Report* R-project folder to contain all your work. Prepare a formatted pdf file of your work using the R markdown file inside the project folder.

There are three data files to analyse:

1. Warner-Lambert Early Pregnancy clinical trials data;
2. Detection of people sick with an influenza virus from photos with a "yes" or "no" judgment;
3. Detection of people sick with an influenza virus from photos using a confidence rating judgment.

This project involves solving four tasks using R and RStudio:

1. Reading data into R
2. Computing appropriate summaries of the data
3. Computing parameters of a signal-detection model to describe the data
4. Graphically displaying the results

Part 1: Early Pregnancy Test - "yes" or "no"

Dual-Gaussian, Equal-Variance, Single-Criterion Model

In 1977 the pharmaceutical company Warner-Lambert introduced the first in-home pregnancy test: the 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 (hits) and indicated that 36 were not pregnant (misses). Of the 198 non-pregnant women tested, the E.P.T. indicated that 15 were pregnant (false alarms) and that 183 were not pregnant (correct rejections).

Table 1: Number of 'no' and 'yes' EPT responses for pregnant (s1) and not-pregnant (s0) women in the 1977 clinical trials of the Warner-Lambert Early Pregnancy Test.

EPT Results	Not Pregnant (s0)	Pregnant (s1)
no	183	36
yes	15	451

Consider that the signal condition (s1) is pregnancy and the blank condition (s0) is non-pregnancy. Use the equal-variance signal detection theory model to determine the sensitivity (d-prime), the decision criterion (X_c), the accuracy (A_z) and the response bias (c) of the E.P.T test. Use Equations 9c, 10, 14, and 12 in the *Detection Theory* handout available from the course website. 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?

ROC Graphs and Model Graph

Construct a probability ROC graph with FAR on the horizontal axis (the abscissa) and HR on the vertical axis (the ordinate) using `ggplot()`. 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. 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 in z-score coordinates is a linear function with a slope of 1:

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

See Equation 6b in the *Detection Theory* handout. Be sure to label the axes of your graph. Make a second ROC plot using the z-score (`qnorm()`) of the hit rate and false alarm rate probabilities.

It should look like Figure 1 but with the correct data in the graph (the curve and point are **not** based on the E.P.T data but are shown for illustrative purposes.. 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 sum to 1.0. The figure legend is created by the `fig.cap="Legend Here"` argument in the code chunk header.

Now make a graph of the Gaussian normal distributions that represent the detection model. We assume that the not-pregnant condition (s0) has a mean of 0.0 and a standard deviation of 1.0. In the equal-variance model the signal distribution has a mean of d-prime and a standard deviation of 1.0. One way to make the plot in `ggplot()` is to generate a range of z-scores (e.g, -4 to +6) and for each value compute the corresponding probability density using `dnorm()` with appropriate mean and standard deviation. Save the density results in a data frame with a column indicating which signal condition, `s0` or `s1` generated the probability density. Then use `ggplot()` to make a graph of the bell-shaped curve. You can use `geom_vline()` to add a vertical line where the decision criterion, X_c, is located. X_c is equal to the negative of the false alarm rate z-score.

Part 2: Axelsson (2018) Experiment 1

Dual-Gaussian, Equal-Variance, Single-Criterion Model ("not sick or"sick")

Axelsson and colleagues (Axelsson et al., 2018) asked subjects to judge whether or not photographs of a person show a person infected with influenza virus or not infected with the virus. In Study 1 of their paper 62 subjects viewed individual photographs of 16 (8 men; 8 women) people who, two hours earlier, had been injected either with an influenza virus (*Escherichia coli endotoxin*), the **LPS** condition or with saline (0.9% NaCl), the **placebo** condition. The subjects were asked to decide if each photograph was of a "sick" person or a "healthy person". Two composite photos from their study, each made up of the average of eight women two hours after injection of placebo (right) and the same 8 women after injection of the virus (left) are shown in Figure 3.

Table 2 gives the number of "healthy" and the number of "sick" judgments made by 62 subjects shown the 32 individual photographs one at a time. Each subject saw both sick and placebo photos of the 16 people

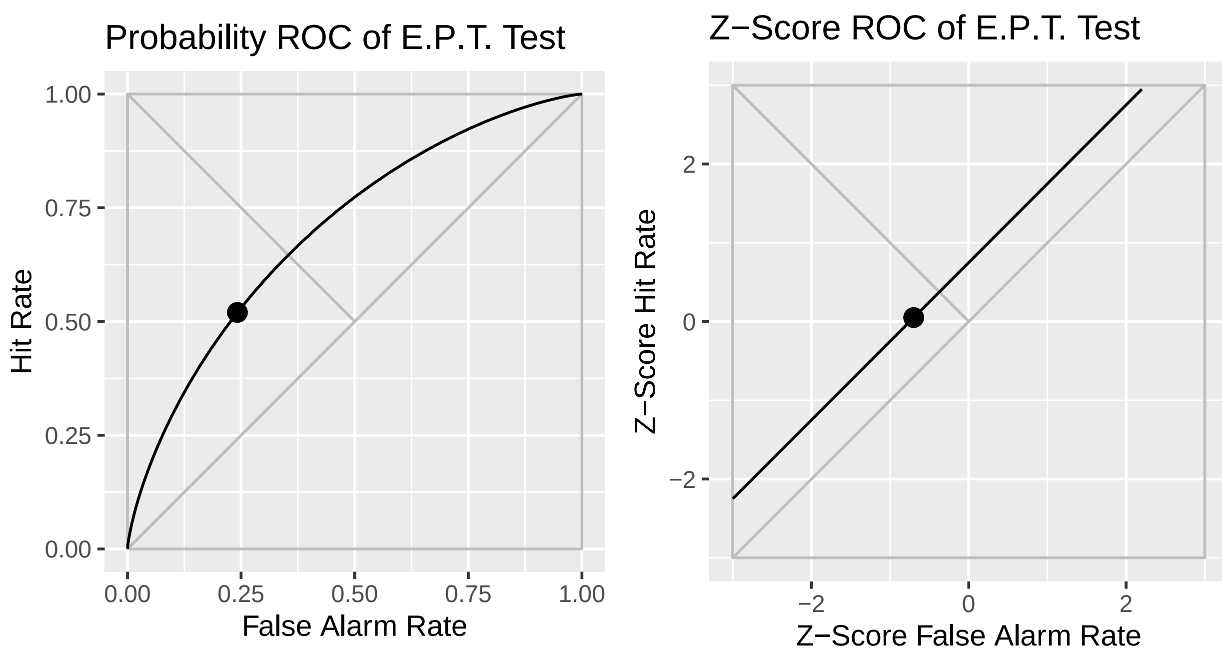

Figure 1: Receiver operating characteristic of the Warner-Lambert E.P.T. home pregnancy test. The filled circle is the resulting hit rate and false alarm rate computed from the clinical trials data. The smooth curve represents the predictions of the equal-variance signal detection model. The gray positive 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.

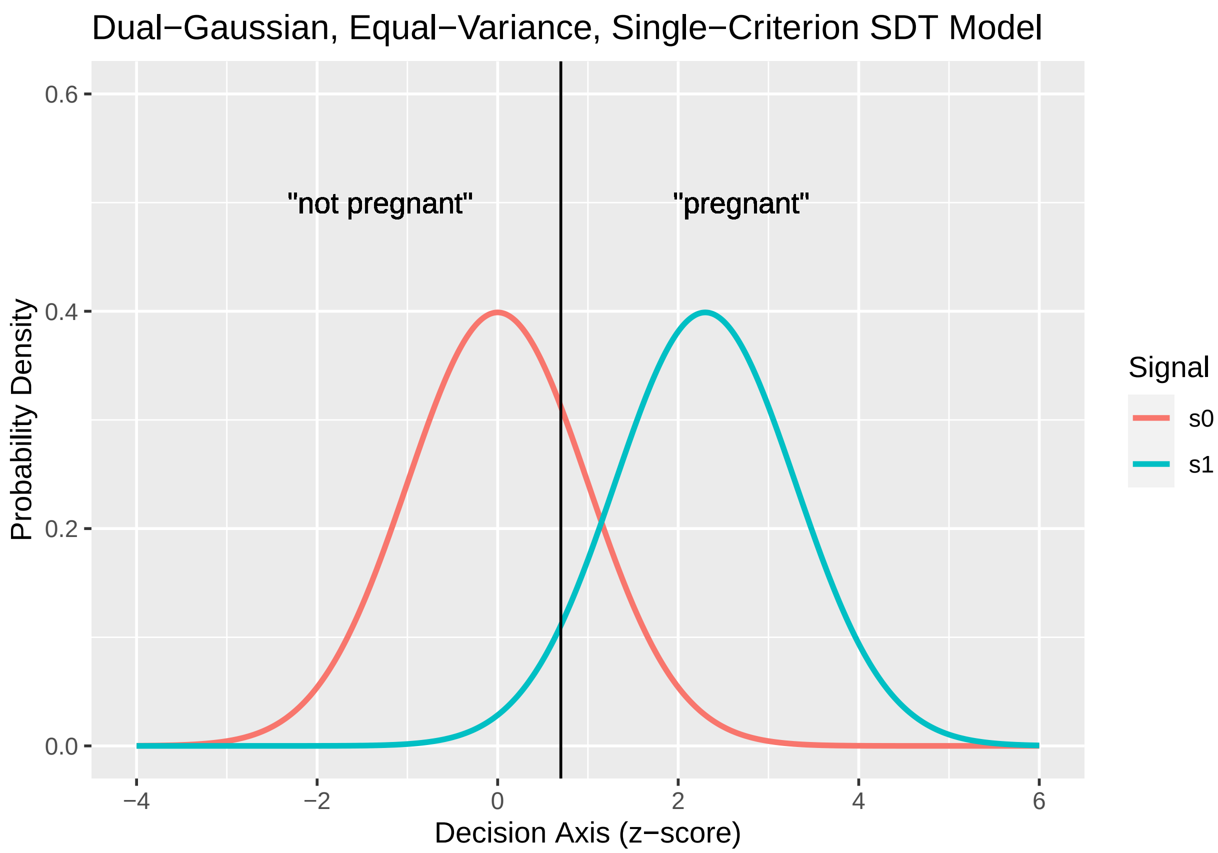

Figure 2: The dual-Gaussian, equal-variance, single decision criterion model of the Warner-Lambert E.P.T. pregnancy test. The red distribution represents the test output distibution in response to non-pregnancy; the blue distribution represents the test output distribution to pregnancy. The vertical black line is the decision criterion. It divides the response space into 'non-pregnant' and 'pregnant' decision outcomes.

Figure 3: Figure 3 from Axelsson et al., 2018. Which composite photo, a or b, is of an infected person?

(for a total of 32 photographs) in a random order with the constraint that no photograph of the same person was shown immediately following the other photo of that person.

Table 2: The number of 'healthy' and the number of 'sick' judgments made by 62 subjects shown the 32 individual photographs (16 infected with the lps virus and 16 injected with a placebo saline solution) one at a time. The data are from Axelsson et al. (2018). The frequencies are aggregated across all subjects.

Rating	Placebo (s0)	Virus (s1)
healthy	1028	702
sick	440	775

As with the EPT pregnancy test, treat these data as a signal detection task of trying to detect the photograph of the person who will become sick. The data table shows that the there were 775 hits and 440 false alarms. Convert these numbers into signal detection measures of sensitivity, decision criteria, bias and accuracy (d', X_c, c, and A_z) as you did before. Make two ROC graphs to illustrate the SDT dual-Gaussian, equal-variance, single decision criterion model fit to the data: one in probability coordinates and one with z-score coordinates.

What is your conclusion about the ability of the subjects to detect the sick people from the photographs and were they biased to say "sick" or biased to say "healthy"? How do your signal detection values compare to the d' and A_z reported in the Axelsson paper? They will not be exactly the same because Axelsson computed d' for each subject and averaged them (the correct way to combine data), whereas you are computing a single d' from the combined data from all the subjects.

Since hit rate and false alarm rate depends on the sensitivity *and* the value of X_c the decision criterion both of which can vary across subjects, the better way to aggregate data is to compute d-prime and the decision criterion for each subject and then average the d-primes. See if you can figure out how to do that.

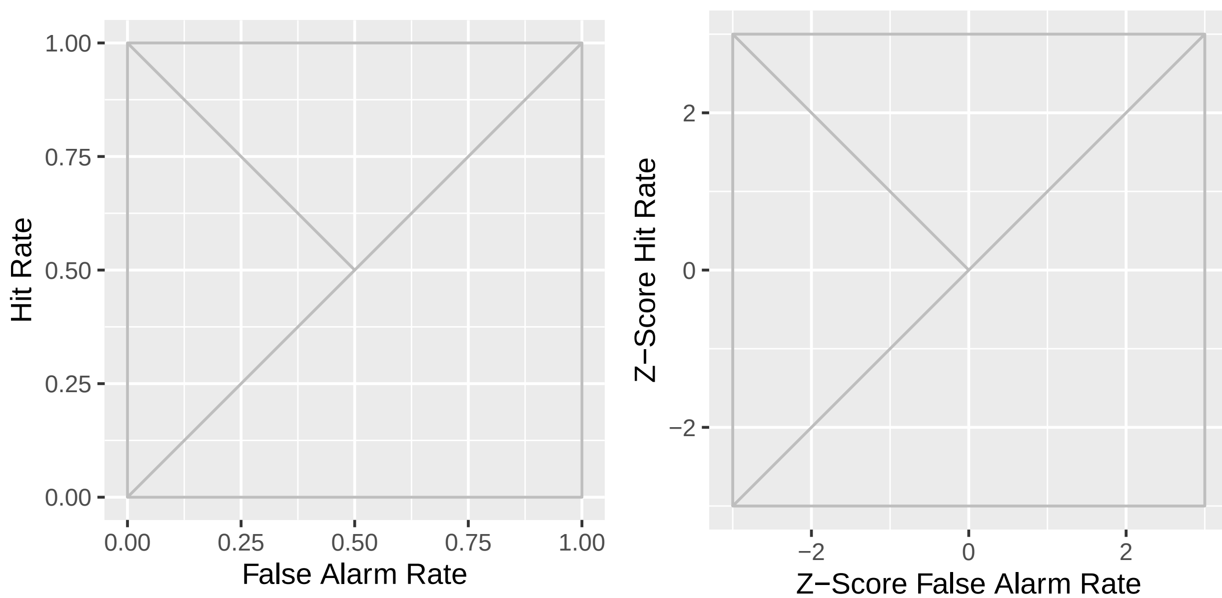

Figure 4: Receiver operating characteristic of subjects judging photographs as 'sick' or 'healthy' computed from the data of Study 1 in Axelsson, et al (2018). The left panel plots the probability; the right panel plots the z-score of the probability. The filled data point is the resulting hit rate and false alarm rate for detecting the sick photos, computed from the aggregated data of all subjects. 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 there was no ability to detect the sick photos. The gray negative diagonal represents the hit rates and false alarm rates that could occur with an unbiased observer.

Part 3: Unequal-Variance, Multiple Decision Criteria Model (Sick People)

Axelsson et al. (2018) reported a second study using these photographs. They asked 60 new subjects to rate the 32 photographs on how sick the person appeared to be using a confidence rating scale that ranged from R1 ("very poor") to R7 ("very good"). Table 3 shows the number of times each rating was given for each type of photograph. Since we are considering this task to be one of detecting the sick person, I have reversed the order of the ratings, making R1 mean "very healthy") and R7 mean "very sick" as presented in Table 3.

Table 3 gives the number of times each confidence rating was used for each image condition:

Table 3: The number of times for each condition that the subjects rated the photograph using a 1 to 7 confidence rating scale. The frequencies are aggregated across all subjects.

Rating	Placebo (s0)	Virus (s1)
R1	617	476
R2	2917	1799
R3	3451	3032
R4	3025	2892
R5	2109	2784
R6	673	1640
R7	81	294

Transform the frequency data into a table of probabilities and then the probabilies into cumulative probabilities, starting with the most conservative response category, R7. These cumulative probabilities correspond to false alarm rates (using condition s0) and hit rates (condition s1). These cumulative probabilities are the false alarm rates (s0) and the hit rates (s1). Now transform the cumulative probabilities into z-scores of the normal distribution (`qnorm()`). Show tables of each successive transformation of the data.

Graphs of the Model

Plot two ROC graphs from these data: one graph in linear probability coordinates (axes ranging between 0.0 and 1.0), the other in z-score coordinates (axes ranging from -3 to 3) as is shown in Figure 4. Make the x- and y-axes of your graph equal in length so that each graph forms a square. Hint: use `coord_equal(ratio = 1)` as a layer in `ggplot()`. Properly label the graphs. The graphs will look like this (with data in them of course): Finally, add smooth curves to your ROC plots.

Making use of the fact that the Gaussian signal detection model predicts that the ROC will be a straight line with y-intercept of b_0 and a slope of b_1, fit a straight line to the cumulative z-score data for the conditions and report your value of b_0 and b_1 for the Z-Score ROC:

$$z\,(HR) = b_0 + b_1\, z(FAR)$$
$$qnorm\,(HR) = b_0 + b_1\, qnorm(FAR)$$

Fit a straight line to the z-score ROC data using the `lm()` function in R and extract b_0 and b_1 from the model. Use Equation 9a in the Detection Theory handout to compute the detection sensitivity (d_a) from these coefficients. Then use Equation 14 to compute the area under the ROC (A_z) as a measure of accuracy. Organize your b_0‘, b_1, d_a and A_z values in a table. The index d_a is a generalized version of d-prime that is used when the signal detection model does not have equal variance distributions.

Make a graph of the Gaussian model, showing the two Gaussian distributions with s1 having an appropriate mean and standard deviation. Mark the position of the decision criteria that form the boundaries of the seven response ratings.

Conclusions

What do you conclude about peoples' ability to judge health from photographs? How does the d' and accuracy (A_z) from Study 1 compare with d_a and A_z from Study 2? How do these subjects compare to the Warner-Lambert E.P.T. test?

References

Axelsson, J., Sundelin, T., Olsson, M. J., Sorjonen, K., Axelsson, C., Lasselin, J., & Lekander, M. (2018). Identification of acutely sick people and facial cues of sickness. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1870).