Section 6.2 Computer Exercise

We will be studying sensitivity, specificity and ROC (Receiver Operating Characteristic) curves. It's interesting to write your own functions in R to understand the process, but we will use built-in methods in SAS PROC LOGISTIC instead. We will be working with the Shuttle data from problem 5.6 (the data is both embedded in the code and available as a separate file).

1. Read the data and run the Section 6.2 SAS code from the webpage through PROC LOGISTIC; note that the use of `PLOT` in the PROC statement and the `OUTROC` option in the MODEL statement generated a ROC plot. Let's focus on the output file `PLOTROC` for now. It should contain the following variables:

Name	Label
PROB	Probability Level
POS	Number of Correctly Predicted Events
NEG	Number of Correctly Predicted Nonevents
FALPOS	Number of Nonevents Predicted as Events
FALNEG	Number of Events Predicted as Nonevents
SENSIT	Sensitivity
1MSPEC	1-Specificity

   Probability Level is the set of probabilities that will serve as cut-points; each value is the probability of the response variable **Distress** for each unique level of the independent variable **Temperature**. For each cutpoint π_0, we compare the predicted value of the response to the cutpoint. If the predicted value of the response is greater than the cutpoint, then we predict distress (or 1); otherwise we predict no distress (or 0). In reality, the cutpoints range smoothly from 0 to 1, but we only need to study cutpoints where the number of predicted successes/failures would change.

   At each cutpoint, we basically have a cross-classification of observed (Y) and predicted ($I\left[\hat{\pi}(x) \geq \pi_0\right]$) responses that tabulates _POS_, _NEG_, _FALPOS_, _FALNEG_:

	$\hat{\pi}(x) < \pi_0$	$\hat{\pi}(x) \geq \pi_0$
Y=0	_NEG_	_FALPOS_
Y=1	_FALNEG_	_POS_

   Sensitivity is computed as _POS_ /(_POS_ + _FALNEG_) and 1-specificity is computed as _FALPOS_ /(_FALPOS_ + _NEG_).

   For values of π_0 greater than .93925, our largest value of $\hat{\pi}(x)$, we would have 0 predicted events, and hence 0 correctly predicted events and a sensitivity of 0. Likewise, we would have 23 predicted non-events, and hence 16 correctly predicted non-events (the total number of non-events in the data); the specificity would be 1 and 1-specificity would equal 0. As π_0 decreases, the number of correctly predicted events increases to 7 (the total number of Events) and sensitivity increases to 1, while the number of correctly predicted non-events decreases to 0, and 1-specificity increases to 1.

2. Review the ROC curve. The area under the ROC curve estimates a measure of concordance. Comment.

The code includes PROC GPLOT and PROC SGPLOT commands for generating the ROC curve using the output file `PLOTROC`; they are there mostly as a reference.