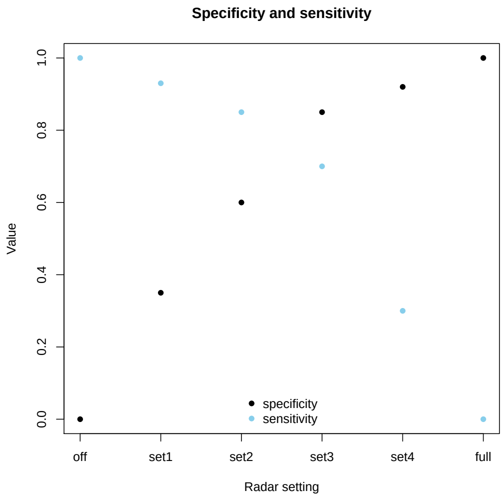
Exercise 1 (ROC)

Receiver-operating characteristic (ROC) curves are commonly used to compare diagnostic tests. They were initially developed in World War II, by the British who build the "Chain Home" series of radar detectors to identify incoming German planes. Besides planes, the radar detectors also detected flocks of birds and other "false positive" signals. The responsiveness of the radar detector could be tuned from off over increasing settings to full.
Take a careful look at the following table:

Radar detector setting	Planes detected (*sensitivity*)	Geese flocks correctly identified (*specificity*)	Geese flocks incorrectly identified ($1 - specificity$)
Off	0	100	0
Setting 1	35	93	7
Setting 2	60	85	15
Setting 3	85	70	30
Setting 4	92	30	70
Full	100	0	100

(a) How does sensitivity and specificity change with increasing radar responsiveness? Describe the relationship between sensitivity and specificity.

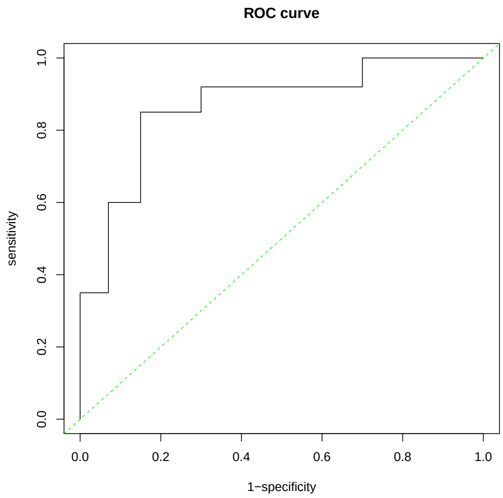
```r
# We can evaluate the table or compare the sensitivity and specificity
# within a plot:
dat = data.frame(setting = c("off", "set1", "set2", "set3", "set4", "full"),
                 sens = c(0, 35, 60, 85, 92, 100) / 100,
                 spec = c(100, 93, 85, 70, 30, 0) / 100)
plot(dat$sens, xaxt = "n", ylab = "Value", pch = 16,
     main = "Specificity and sensitivity", xlab = "Radar setting")
points(dat$spec, col = "skyblue", pch = 16)
axis(1, at=1:nrow(dat), labels=dat$setting)
legend("bottom", c("specificity", "sensitivity"), pch = c(16, 16),
       col = c("black", "skyblue"), bty = "n")
```




```
# With increasing sensitivity, the specificity decreases.
# So, if the responsiveness of the radar detector is off,
# 100% of the planes and none of the Geese flocks are detected.
# If the responsiveness of the radar detector is full, 100% of
# the Geese flocks are detected but none of the planes.
```

(b) Generate a ROC curve by plotting `sensitivity` against `1-specificity` from the above table. Do you think the responsiveness of the radar detector is good?

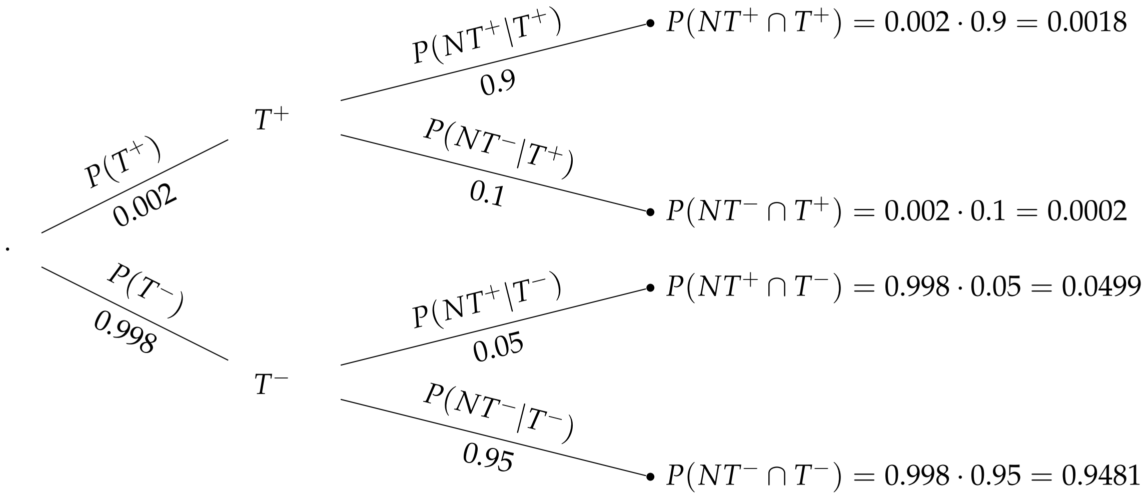
```
# The ROC curve can be calculated as:
plot(1-dat$spec, dat$sens, ylab = "sensitivity",
     main = "ROC curve", xlab = "1-specificity", type='S')
abline(0, 1, col='green', lty=2)
```



Exercise 2 (ROC)

The chromosomes of human being usually consists 23 pairs of chromosomes (one of the pair is the sex chromosome, in case of a woman XX and for a man XY). If a human being has 3 or 4 chromosomes instead of a pair, it usually leads to severe diseases. Trisomy 21 (also called Down syndrome), in which the chromosome 21 occurs three times, is the most common of these diseases. In prenatal diagnostics the nuchal translucency (NT) is used to test for Trisomy 21.
Assume that the probability for having a child with Trisomy 21 in pregnant women at the age of 30 years is 0.2%. The NT-test provides a positive result for the presence of trisomy 21 with 90% probability. If the child has no trisomy 21, the result is in 5% of the cases still positive.

(a) Draw a probability tree with nodes T21 and NT.



The final values (e.g.) $P(NT^+ \cap T^+)$ are calculated using the Bayes formula $P(A|B) = \frac{A \cap B}{B}$.

(b) What are the sensitivity and specificity of the NT-test?

$$sens = P(NT^+|T^+) = 0.9$$
$$spec = P(NT^-|T^-) = 1 - P(NT^+|T^-) = 1 - 0.05 = 0.95$$

(c) Assume that the NT-test is done for every pregnant woman of 30 years. Let's take 100.000 pregnant woman of the age 30. Fill the following table.

	NT$^+$	NT$^-$	total
T$^+$			
T$^-$			
total			

The entries can be calculated as:

$$t_{11} = 100000 \cdot P(NT^+ \cap T^+) = 100000 \cdot 0.0018 = 180$$
$$t_{12} = 100000 \cdot P(NT^- \cap T^+) = 100000 \cdot 0.0002 = 20$$
$$t_{21} = 100000 \cdot P(NT^+ \cap T^-) = 100000 \cdot 0.0499 = 4990$$
$$t_{22} = 100000 \cdot P(NT^- \cap T^-) = 100000 \cdot 0.9481 = 94810$$

Therefore, the filled table should look like:

	NT$^+$	NT$^-$	total
T$^+$	180	20	200
T$^-$	4990	94810	99800
total	5170	94830	100000

Email: hezo@zhaw.ch

(d) Calculate the probabilities that

a) a randomly chosen pregnant women has a positive test.

b) there is really trisomy 21, if the test is positive.

Both probabilities can be calculated with the Bayesian formula.

$$P(NT^+) = P(NT^+|T^+) \cdot P(T^+) + P(NT^+|T^-) \cdot P(T^-) = 0.0517$$
$$P(T^+|NT^+) = \frac{P(T^+ \cap NT^+)}{P(NT^+)} = 0.0348$$

(e) Calculate the positive/negative predictive value of the test.

$$PPV = \frac{TP}{TP + FP} = \frac{P(NT^+ \cap T^+)}{P(NT^+ \cap T^+) + P(NT^+ \cap T^-)} = 0.0348$$
$$NPV = \frac{TN}{FN + TN} = \frac{P(NT^- \cap T^-)}{P(NT^- \cap T^+) + P(NT^- \cap T^-)} = 0.9998$$

(f) How many false positives (FPs)/true negatives (TNs) do we have (in case of the 100000 women)?

$$FP = 100000 \cdot P(NT^+ \cap T^-) = 100000 \cdot 0.0499 = 4990$$
$$TN = 100000 \cdot P(NT^- \cap T^-) = 100000 \cdot 0.9481 = 94810$$

(g) What are the true positive rate (TPR) and false positive rate (FPR)?

$$TPR = \frac{TP}{\sum \text{Condition positive}} = \frac{P(NT^+ \cap T^+)}{P(T^+)} = P(NT^+|T^+) = 0.9$$
$$FPR = \frac{FP}{\sum \text{Condition negative}} = \frac{P(NT^+ \cap T^-)}{P(T^-)} = P(NT^+|T^-) = 0.05$$

Email: hezo@zhaw.ch