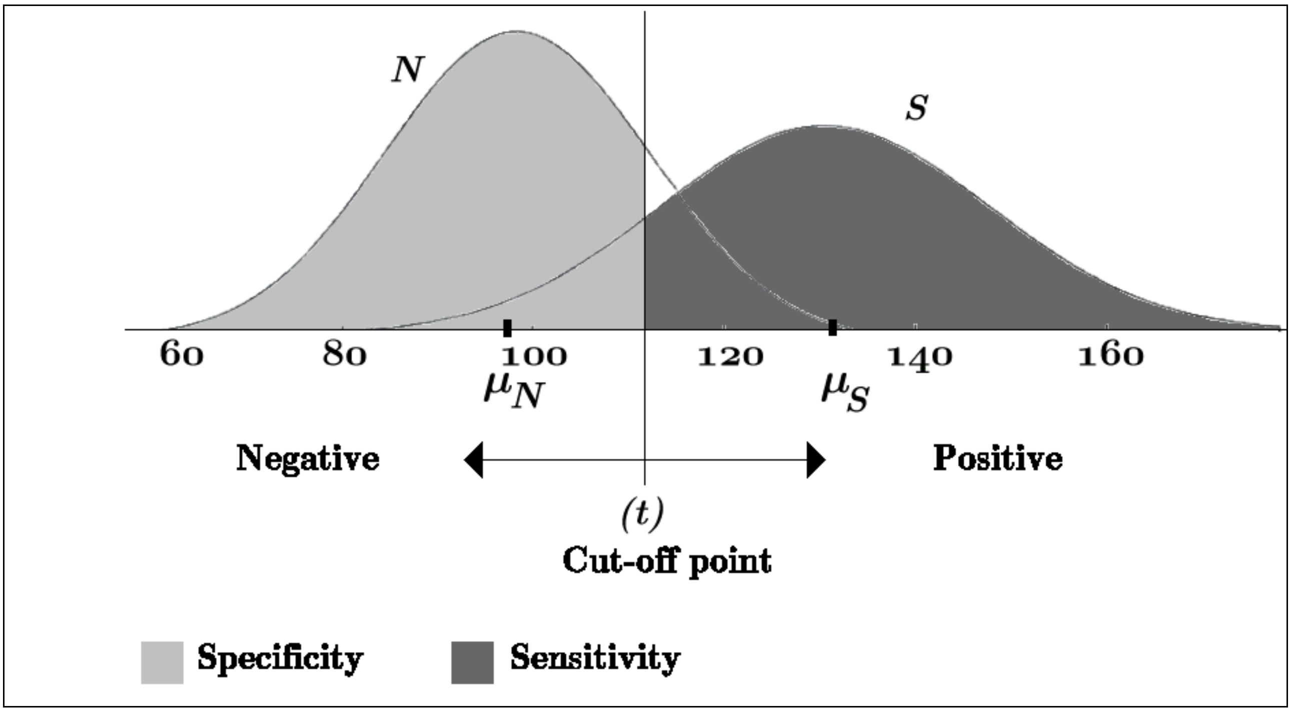
Using ROC-analysis to determine correct on continuous variables.

H. S. Steyn
Statistical Consultation Service

1. Sensitivity, Specificity and Predicted values (Kline, 2004a)

Suppose that a disease or abnormality is studied and individuals are categorized as "positive" if they exhibit the disease or abnormality and "negative" otherwise. Other examples include clinical psychologists that would like to classify individuals as depressive or normal, or a bank that want to classify clients asking for loans as being potentially risky or not. These classifications are typically made through the use of "gold standard" diagnostic tests that are possibly expensive and/or time-consuming. If a method (also called a screening test) existed to identify the diseased/abnormal/risky individuals, that was simpler and cheaper than the "gold standard", then it would be important to know how trustworthy it was as a predictor for diseased/abnormal/risky individuals. In further discussion we will refer to the population of sick (S) and the non-sick (N) individuals for the individuals that have a disease or not, have an abnormality or not, or are considered risky or not. Typically the measurements made for these screening tests produce a continuous value (i.e., a value that varies over a certain interval) and a cut-off or threshold value is often used to classify individuals, e.g., values above the threshold value indicate the presence of a disease, while values below it indicate the absence of the disease. Figure 1 provides a graphical representation of the distributions of populations S and N's screening test measurements.

Figure 1:



If individuals are classified according to their actual status (according to some golden standard) as well by the screening test, then it produces the following 2x2 – frequency table:

	Actual status		
Screening test	Sick (S)	Not-sick (N)	Total
+	A (true pos.)	B (false pos.)	A + B (test +)
-	C (false neg.)	D (true neg.)	C + D (test -)
Total	A + C (sick)	B + D (not-sick)	N=A+B+C+D

In this table there are A sick individuals that reacted positively to the screening test. If it is expressed as a proportion of all the sick individuals $(A+C)$, then it gives the **sensitivity**, i.e.,

$$\text{Sensitivity} = \frac{A}{A+C}, \tag{1}$$

the proportion of correctly classified positives (i.e., where sick individuals test positively).

Similarly, there are D individuals correctly classified as not-sick; when this is expressed as a proportion with respect to the total not-sick individuals it is called the **specificity**, i.e.,

$$\text{Specificity} = \frac{D}{B+D} , \qquad (2)$$

the proportion of correctly classified negatives (i.e., where not-sick individuals test negatively).

A good screening test should have a high sensitivity as well a high specificity, because the opposite would be unfavourable. That is, to classify a sick person as not-sick (a total of C individuals) is unfavourable; similarly it is also unfavourable to classify a person as being not-sick if they are sick (a total of B individuals).

In the populations (as shown in Figure 1) the area under the S-distribution (sick individuals) to the right of the cut-off point denotes the sensitivity and the area under the N-distribution to the left of the cut-off point is called the specificity. Ideally the two distributions would be completely separated and the cut-off point chosen such that both the sensitivity and specificity are equal to 1.

A $(1-\alpha)100\%$ *confidence interval (CI)* for the sensitivity is given by:

$$Sens \pm z_{\alpha/2}\sqrt{Sens(1-Sens)/(A+C)} ,$$

Whereas the CI for *Specificity* is:

$$Spec \pm z_{\alpha/2}\sqrt{Spec(1-Spec)/(B+D)} ,$$

where *Sens* and *Spec* are the sensitivity and specificity as given in equations (1) and (2), $z_{\alpha/2}$ is the $(1-\alpha/2)^{th}$ percentile of the standard normal distribution.

2. The ROC–curve (Krzanowski & Hand, 2009)

First consider the populations S and N. For each cut-off point t, we can construct a 2x2 – table of probabilities as follows:

	Populations	
Screening tests	S	N
+	$\mathrm{P}(X > t \mid S)$	$\mathrm{P}(X > t \mid N)$
-	$\mathrm{P}(X \le t \mid S)$	$\mathrm{P}(X \le t \mid N)$

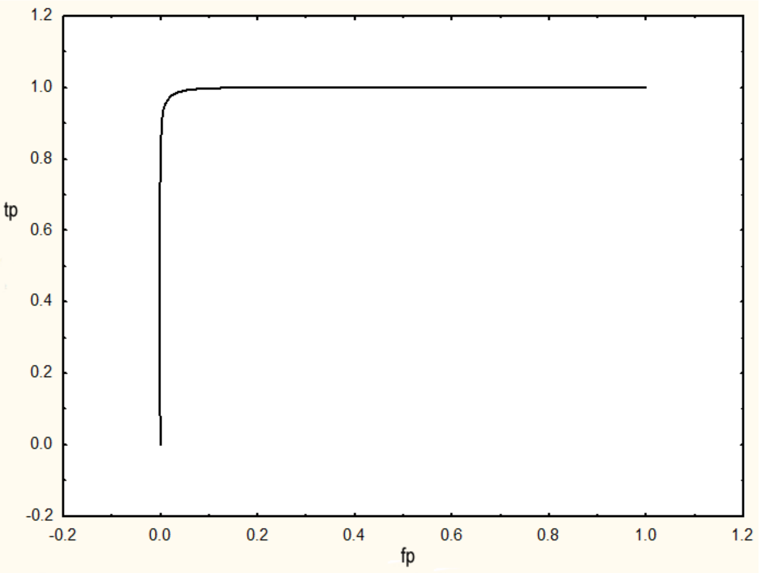
(here X is the screening test).

The values of $P(X > t \mid S)$, i.e., the sensitivity or proportion of true positives (tp) versus $\mathrm{P}(X > t \mid N)$, i.e., 1 – specificity or the proportion of false positives (fp), can be plotted on a graph for a sequence of cut-off values t. The resulting plot is known as Receiver-Operating Characteristic curve, or ROC-curve.

Figure 2 displays some examples of different ROC curves.

Figure 2:

(a)



(b)

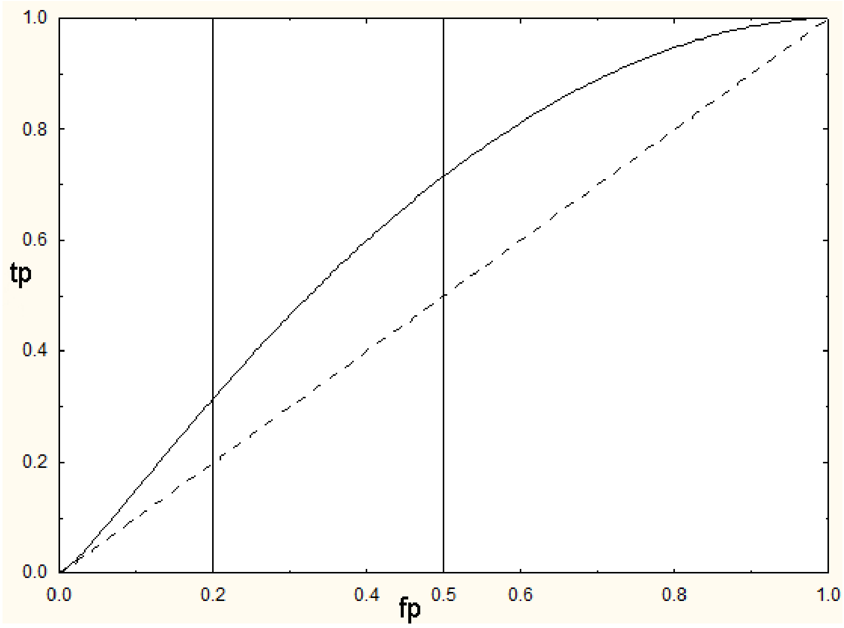


Figure 2(a) is obtained if the populations in Figure 1 are both normally distributed, population S has mean and standard deviation $\mu_S = 4, \sigma_S = 1,$ and population N has mean and standard deviation $\mu_N = 0$ and $\sigma_N = 1$. Here the populations are completely separated since the density function of the S population lies almost entirely to the right of the density function of the N population. This ROC curve illustrates the near best possible curve obtainable; the optimal cut-off point between the two distributions can easily be chosen. The other extreme is illustrated by the diagonal line (dashed line) in Figure 2(b) where it is used to indicate the hypothetical situation where both distributions are assumed to be $N\left(0;1\right)$ and thus indistinguishable from one another.

In this case individuals from each population are equally likely to be classified as positive or negative.

Figure 2(b) also illustrates a ROC-curve (solid line) that could be obtained from a situation similar to Figure 1; an appropriate optimal cut-off point can then be chosen from this graph.

3. Properties of the ROC-curve

1. It is monotone increasing between $x = 0, y = 0$ and $x = 1, y = 1$, where the x -axis is associated with the 1-specificity values (fp) and the y-axis is associated with the sensitivity values (tp) for a sequence of cut-off points.
2. It is invariant to monotone increasing transformations (e.g., the log transformation) on the screening test values.
3. The gradient of the ROC-curve at cut-off point t is $f_S(t)/f_N(t)$, where $f_S(t)$ and $f_N(t)$ are the density functions of population distributions S and N in the point t.

4. Area Under the ROC-curve (AUC)

If one considers Figure 2, it is clear that the area under the ROC-curve in (a) is approximately 1, in (b) it is between 0,5 and 1 in the case of the solid line, and exactly 0,5 in the case of the dashed line. This “Area Under the Curve” is denoted by AUC and is used as a measure of the ability to discriminate between the distributions S and N. Larger values of AUC indicate a greater discriminatory ability. The AUC value 0,5 indicates that the one is unable to distinguish between S and N.

AUC can also be interpreted as follows:
Suppose that an individual is randomly chosen from each of the populations S and N and the screening scores are X_S and X_N, then:

$$\text{AUC} = \mathrm{P}(X_S > X_N) \quad , \tag{3}$$

which means that the AUC is the probability that X_S is larger than X_N. In terms of Figure 2 this probability is close to 1 if the two populations are easily distinguishable (Figure 2(a)), whereas the probability is 0,5 if the population distributions completely overlap (Figure 2(b)).

Single points and partial areas

If one is interested in a specific false positive rate (fp = x_0), then one can use the ROC-curve to determine the corresponding true positive rate (tp) by reading off the tp at y(x_0). For example, take x_0 = 0.05 (i.e., specificity of 0.95) and then use it to find out what the corresponding sensitivity is. In other words, we want to determine what the true positive rate would be if most of the normal (not-sick) population is correctly classified as negative.

Sometimes we are only interested in considering tp when fp values lie within a certain interval $(\mathrm{f}_1, \mathrm{f}_2)$. A summarizing index is the partial area under the curve (PAUC), i.e., the area under the ROC-curve between fp $= \mathrm{f}_1$ and fp $= \mathrm{f}_2$. Depending on the values of f_1 and f_2, the PAUC can take on any minimum or maximum value between 0 and 1. This makes it rather difficult to interpret the PAUC value.

From Figure 2(b) it is clear that the maximum value of the PAUC is the area of the rectangle with its base on the interval $(\mathrm{f}_1, \mathrm{f}_2)$ = (0,2; 0,5), i.e., $(\mathrm{f}_2 - \mathrm{f}_1)$ = 0,3, and height 1. The minimum size of the area of the trapezium under the diagonal line (i.e., the ROC-curve associated with the case where one cannot distinguish between populations S and N) between f_1 and f_2 , is equal to $\frac{1}{2}(\mathrm{f}_2 - \mathrm{f}_1)(\mathrm{f}_1 + \mathrm{f}_2)$, which is equal to 0,3 x 0,7 / 2 = 0,105 in Figure 2(b). The following expression is an index between 0 and 1 which can be used to interpret the PAUC:

$$\mathrm{I}_{\mathrm{PAUC}} = \frac{1}{2}\left[1 + \frac{\mathrm{PAUC}(\mathrm{f}_1, \mathrm{f}_2) - (\mathrm{f}_2 - \mathrm{f}_1)(\mathrm{f}_1 + \mathrm{f}_2)/2}{(\mathrm{f}_2 - \mathrm{f}_1)(1 - (\mathrm{f}_1 + \mathrm{f}_2)/2)}\right] \tag{4}$$

(see paragraph 2.4.2, Krzanowaski & Hand, 2009).
If the ROC-curve's equation, $y(x)$, is known, then the PAUC is given by:

$$\mathrm{PAUC} = \int_{\mathrm{f}_1}^{\mathrm{f}_2} \mathrm{y(x)dx}\,, \tag{5}$$

i.e., the integral of the function $y(x)$ between f_1 and f_2.

5. The binormal model

If the populations S and N are each normally distributed, then the ROC-curve can be formulated as follows:

Define the 1-specificity at cut-off point t as $x(t) = \mathrm{P}(X > t \mid N)$ and the sensitivity at t as $y\,(t) = \mathrm{P}(X > t \mid S)$. Suppose that S has a $N\left(\mu_S, \sigma_S^2\right)$ distribution and that N has a $N\left(\mu_N, \sigma_N^2\right)$ distribution, then

$$x(t) = \mathrm{P}\left(\mathrm{Z} \le \frac{\mu_N - t}{\sigma_N} \right) = \Phi\left(\frac{\mu_N - t}{\sigma_N} \right),$$

where Z is a standard normal random variable and $\Phi(\cdot)$ is the standard normal distribution function. Let z_x be the value of Z such that the distribution function evaluated in this point is $x(t)$, then it follows that

$$z_x = \Phi^{-1}\left(x(t)\right) = \frac{\mu_N - t}{\sigma_N},$$

and therefore

$$t = \mu_N - \sigma_N z_x . \tag{6}$$

Similarly,

$$y(t) = \Phi\left(\frac{\mu_S - t}{\sigma_S} \right),$$

so that from (6) is follows that:

$$y(t) = \Phi\left(\frac{\mu_S - \mu_N + \sigma_N z_x}{\sigma_S} \right) = \Phi\left(a + b z_x\right), \tag{7}$$

where

$$a = \frac{\mu_S - \mu_N}{\sigma_S} \text{ and } b = \frac{\sigma_N}{\sigma_S}. \tag{8}$$

Note that $a > 0$ $\left(\mu_S > \mu_N\right)$ and $b > 0$.

Further, AUC $= \mathrm{P}\left(X_S > X_N\right)$

$$= \mathrm{P}\left(X_S - X_N > 0\right)$$

$$= \mathrm{P}\left(\mathrm{Z} > \frac{\mu_S - \mu_N}{\sqrt{\sigma_S^2 + \sigma_N^2}} \right)$$

$$= 1 - \Phi\left(\frac{-\mu_S + \mu_N}{\sqrt{\sigma_S^2 + \sigma_N^2}} \right)$$

$$= \Phi\left(\frac{\mu_S - \mu_N}{\sqrt{\sigma_S^2 + \sigma_N^2}} \right)$$

$$= \Phi\left(\frac{a}{\sqrt{1+b^2}}\right). \tag{9}$$

6. Estimation of ROC-curves

6.1 The empirical estimator:

The true positive rate for cut-off point t is

$$tp = \mathrm{P}(X > t \mid S) \ ,$$

whereas the false positive rate is given by

$$fp = \mathrm{P}(X > t \mid N) \ .$$

The obvious or natural estimators for these rates are based on random samples from the populations S and N are given by:

$$\hat{\mathrm{tp}} = \frac{\mathrm{n}_{\mathrm{S(t)}}}{\mathrm{n}_{\mathrm{S}}} \tag{10}$$

and

$$\hat{\mathrm{fp}} = \frac{\mathrm{n}_{\mathrm{N(t)}}}{\mathrm{n}_{\mathrm{N}}} \ , \tag{11}$$

where $n_{A(t)}$ is the number of individuals drawn from population A where the screening test's values were greater than t, while n_S and n_N are the sample sizes.

By varying t between the largest value of X in the sample to the smallest value, one can calculate $\hat{tp}$ and $\hat{fp}$. The plot of the different points $\left(\hat{fp}, \hat{tp}\right)$ obtained in this way produces the empirical ROC-curve.

Example 1: Krzanowski & Hand (2009) provide the following example on p.42:

Suppose that a sample of 10 individuals is drawn from each of the populations N and S, and the ordered values of the screening test are:

N	:	0,3	0,4	0,5	0,5	0,5	0,6	0,7	0,7	0,8	0,9
S	:	0,5	0,6	0,6	0,8	0,9	0,9	0,9	1,0	1,2	1,4

This data can be visualised in the following dot plot:

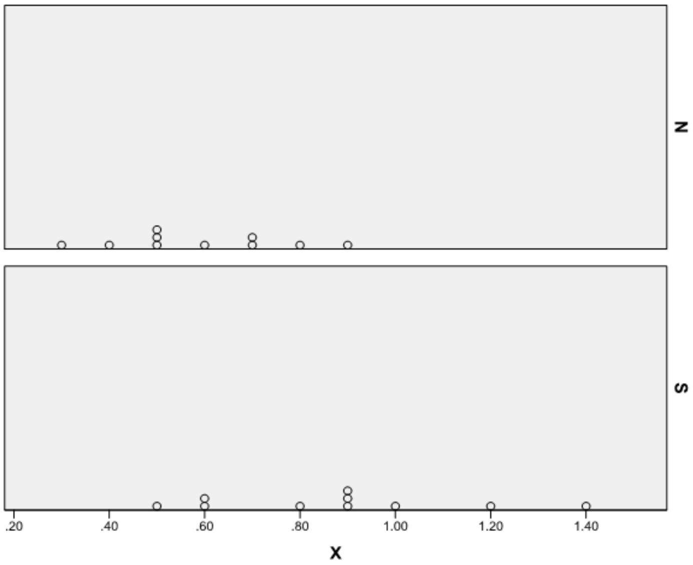
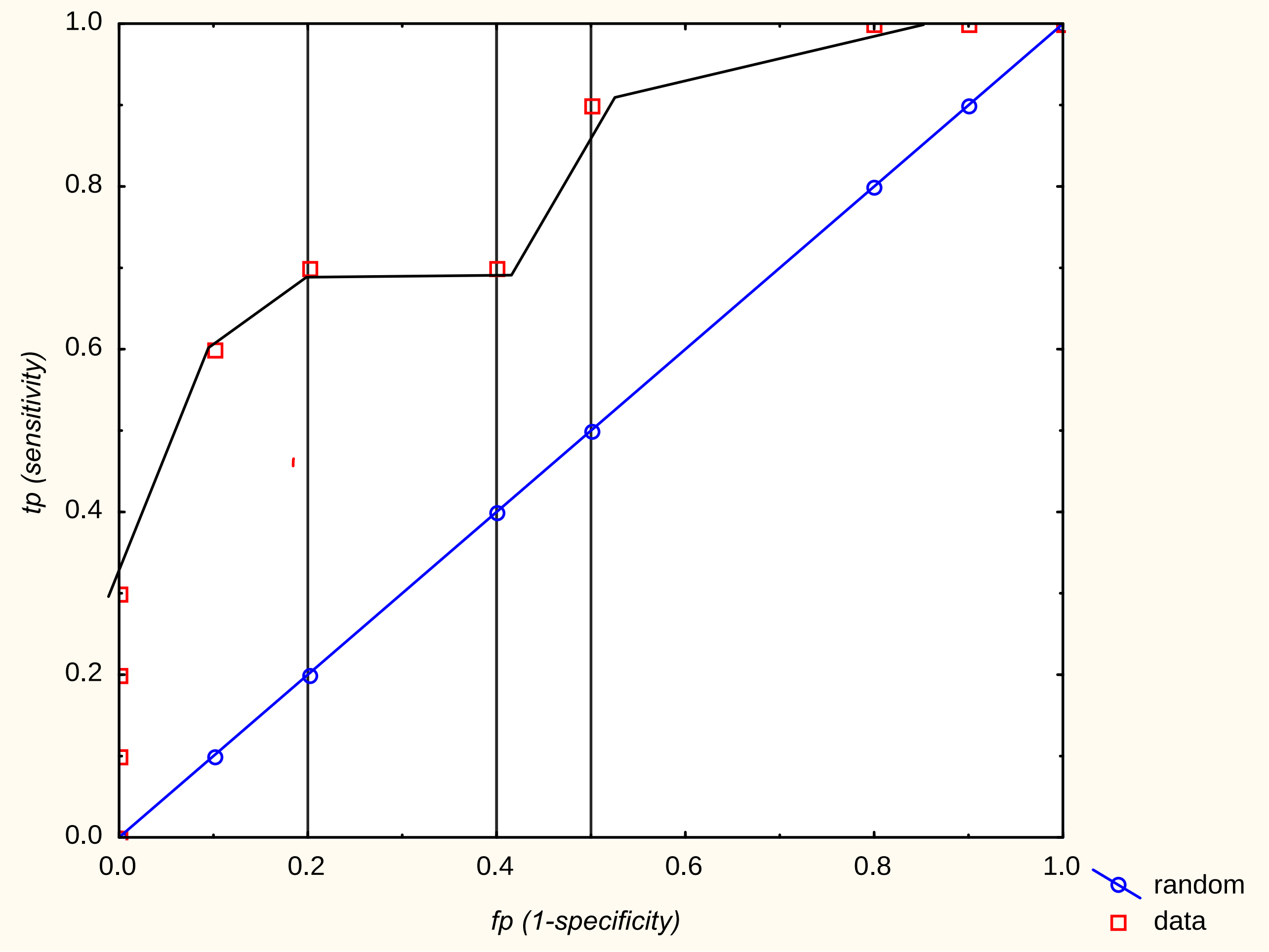


Table 1 shows the values of $\hat{fp}$ and $\hat{tp}$ for a sequence of cut-off points t:

Table 1: Coordinates of the empirical ROC-curve

t	$\hat{fp}$	$\hat{tp}$
≥ 1,4	0,0	0,0
1,2	0,0	0,1
1,0	0,0	0,2
0,9	0,0	0,3
0,8	0,1	0,6
0,7	0,2	0,7
0,6	0,4	0,7
0,5	0,5	0,9
0,4	0,8	1,0
0,3	0,9	1,0
< 0,3	1,0	1,0

Figure 3: Empirical ROC-curve



6.2 Estimation of the binormal model:

The values of a and b in (8) can be estimated using the method of maximum likelihood because we assume that the populations S and N are both normally distributed. The method given in Metz et al. (1998) can be used here (with the aid of the MS Windows program ROCKIT, freely available for download at http://labs.fhcrc.org/pepe/dabs/software.html).

To ensure that the assumption of normality is appropriate, each sample's screening test value can be transformed using the Box-Cox transformation. This involves a power transformation as follows:

$$Y(X)=\begin{cases}\dfrac{X^{\lambda}-1}{\lambda}, & \lambda\neq 0\\ \log(X), & \lambda=0\end{cases},$$

where λ is chosen such that Y is normally distributed. Statistica (Statsoft Inc., 2009) can be used to perform this transformation. Note that this class of transformations is monotone increasing; this implies that the ROC-curve will remain unchanged. In cases where the sample shows evidence of a multi-modal distribution, the Box-Cox transformation cannot be used. Further, it is important to note that the same transformation should be applied to both samples. If the needed transformations differ greatly in the two samples then one should rather not use a transformation at all; in this case one should treat the binormal model with extreme caution.

6.3 Nonparametric estimation of ROC:

If no assumptions can be made regarding the distribution of the screening test, X, for the populations S and N, then it is possible to obtain estimates of the density functions f_S and f_N using kernel density estimators and then to estimate $P(X > t \mid S)$ and $P(X > t \mid N)$ with smoothed functions. This method is described in paragraph 3.3.3 of Krzanowski & Hand (2009) where an example of a ROC-curve plotted with the help of the LABROC program is shown.

7. Confidence intervals for ROC-curves

7.1 Empirical methods:

The empirical ROC-curve is a plot of the points $\left(\overset{\wedge}{fp},\overset{\wedge}{tp}\right)$ for each cut-off point t. The question is now: how accurately do $\overset{\wedge}{fp}$ and $\overset{\wedge}{tp}$ estimate the true proportion of false positives and true positives? Confidence intervals (CIs) for fp and tp can shed some light on this issue; there are three possibilities:

(a) A CI for tp for a given fp (a vertical interval around $\overset{\wedge}{tp}$ at a given point on the horizontal axis of the ROC graph).

(b) A CI for fp for a given tp (a horizontal interval around $\hat{fp}$ at a given point on the vertical axis of the ROC graph).

(c) A CI for (fp, tp) for a given cut-off point , t. Since $\hat{fp}$ and $\hat{tp}$ are both proportions and they are independent of one another for a given t, one can construct separate binomial CIs for fp and tp (horizontal and vertical intervals around $\left(\hat{fp}, \hat{tp}\right)$). A joint $(1-\alpha)$100% confidence region for the pair (fp, tp) follows from the independence of $\hat{fp}$ and $\hat{tp}$, and is given by the rectangle with midpoint $\left(\hat{fp}, \hat{tp}\right)$and sides formed by the $(1-\tilde{\alpha})$100% CIs, where $\tilde{\alpha} = 1 - \sqrt{1-\alpha}$.

The 100$(1-\alpha)$% CI in case (c) above is given for each t:

For tp:
$$t\hat{p} \pm z_{\alpha/2} \sqrt{\frac{t\hat{p}(1 - t\hat{p})}{n_S}} \quad . \tag{12}$$

For fp:
$$f\hat{p} \pm z_{\alpha/2} \sqrt{\frac{f\hat{p}(1 - f\hat{p})}{n_N}} \quad . \tag{13}$$

For (fp, tp): rectangle with $\left(\hat{fp}, \hat{tp}\right)$ as its midpoint and side lengths given by (12) and (13), but where $\tilde{\alpha}$ is used instead of α.

The CIs in cases (a) and (b) also depend on the density functions of the populations S and N and can be determined with the aid of kernel density estimation (see Krzanowski & Hand, 2009: paragraph 3.4.1).

7.2 Binormal method:

Along with the maximum likelihood estimators $\hat{a}$ and $\hat{b}$ for a and b in (8), the ROCKIT program also provides the values of $Var(\hat{a})$, $Var(\hat{b})$ and $Cov(\hat{a},\hat{b})$.

Now, since the variance of $\hat{a}+\hat{b}z_x$ is given by

$$V = Var(\hat{a}) + z_x^2 Var(\hat{b}) + 2z_x Cov(\hat{a},\hat{b}) \;, \qquad (14)$$

it follows that the $(1-\alpha)100\%$ CI for $a+bz_x$ is given by:

$$(L,U) = \hat{a}+\hat{b}z_x \pm z_{\alpha/2}\sqrt{V} \;, \qquad (15)$$

so that the CI for tp given fp = x is:

$$\left[\Phi(U),\Phi(L)\right] \qquad . \qquad (16)$$

Note:

ROCKIT provides the standard error of $\hat{a}$ and $\hat{b},$ i.e., $se(\hat{a})$ and $se(\hat{b})$, as well as the correlation between $\hat{a}$ and $\hat{b}$, i.e., $r(\hat{a},\hat{b})$ and so we can obtain an expression for V as follows:

$$V = \left[se(\hat{a})\right]^2 + \left[z_x se(\hat{b})\right]^2 + 2z_x se(\hat{a}) se(\hat{b}) r(\hat{a},\hat{b}) \,. \qquad (17)$$

7.3 Nonparametric methods:

These methods, described in paragraph 3.4.3 of Krzanowski & Hand (2009), make use of the nonparametric estimation of the ROC-curve, kernel density estimators of f_S and f_N.

8. <u>Estimation of AUC</u>

8.1 Empirical method:

Since the AUC is simply the area under ROC-curves one can easily determine this by making use of the trapezoidal rule; this rule involves determining the sum of trapeziums formed when one inserts vertical lines at each point of the empirical curve that run from the point on the curve down to the horizontal axis, cf. Figure 3. This method is unnecessary if one makes use of (3) where

AUC= $P(X_S > X_N)$.

An estimator for this probability is nothing other than the Mann-Whitney U-statistic, expressed as a proportion of all possible pairs of individuals from the populations S and N, where the X-values from individuals in S exceed the values from N.

Thus $\hat{AUC} = U/(n_S n_N)$, (18)

where U is the Mann-Whitney-U-statistic based on two samples of size n_S and n_N from populations S and N.

8.2 Binormal method:

According to (9) the AUC is estimated as

$$\hat{AUC} = \Phi\left(\frac{\hat{a}}{\sqrt{1+\hat{b}}}\right) \quad . \qquad (19)$$

8.3 Nonparametric method:

This method also makes use of kernel density estimation; it is described in paragraph 3.5.1 of Krzanowski & Hand (2009).

8.4 Estimation of partial AUC:

A parametric estimator of PAUC can be obtained using the binormal model, i.e., from (5) and (7) we have:

$$\mathrm{PA\hat{U}C} = \int_{f_1}^{f_2} \Phi(\hat{a} + \hat{b}z_x)dx \ .$$

The integral can be evaluated using numerical integration techniques.

Nonparametrically, the PAUC can be determined from the probability

$$P(X_S > X_N, f_1 \le 1 - F(X_N) \le f_2) \ ,$$

where F is the distribution function of X in population S. This is similar to the method used to calculate the Mann-Whitney statistic. See paragraph 3.5.2 of Krzanowski & Hand, (2009) for this method.

A simple procedure used in practice is to make use of the trapezoidal rule (discussed in paragraph 8.1) to calculate the area under the empirical curve between f_1 and f_2. In Example 1 we have, (see Figure 3):

PAUC(0,2; 0,5) = area. A + area. B

= (0,2 x 0,7) + 0,1 x (0,7 + 0,9) / 2

= 0,14 + 0,08 = 0,22.

The index is then

$$\hat{\mathrm{I}}_{\mathrm{PAUC}} = \frac{1}{2}\left[1+\frac{0,22-0,3(0,2+0,5)/2}{0,3(1-(0,2+0,5)/2)}\right] = \frac{1}{2}\left[1+\frac{0,22-0,105}{0,195}\right]$$

= 0,795.

9. Confidence intervals for AUC

9.1 Empirical methods:

According to Krzanowski & Hand (2009), an asymptotic expression (i.e., when n_S and n_N are large) for the variance of AUC can be obtained from the Mann-Whitney statistic (in (18)):

$$S^2\left(\hat{AUC}\right) = \frac{1}{n_S n_N}\left[AUC(1-AUC)+(n_S-1)(Q_1-AUC^2)+(n_N-1)(Q_2-AUC^2)\right], \qquad (20)$$

where Q_1 is the probability that the scores, X_S, of two randomly chosen individuals from population S exceed the score X_N of a randomly chosen individual from population N. Conversely, Q_2 is the probability that the score, X_S, of a randomly chosen individual from population S exceeds both scores, X_N, of two randomly chosen individuals from population N. By expressing Q_1 as $AUC/(2-AUC)$, Q_2 as $2AUC^2/(1+AUC)$ (see Krzanowski & Hand 2009: 79) and substituting AUC with $\hat{AUC}$ in (20), we have

$$S^2\left(\hat{AUC}\right)=\frac{1}{n_S n_N}\left[\hat{AUC}\left(1-\hat{AUC}\right)+\left(n_S-1\right)\frac{\hat{AUC}\left(1-\hat{AUC}\right)^2}{2-\hat{AUC}}+\left(n_N-1\right)\frac{\hat{AUC}^2\left(1-\hat{AUC}\right)}{1+\hat{AUC}}\right]. \tag{21}$$

The $(1-\alpha)100\%$ CI for AUC is then:

$$1-\left(1-\mathrm{A}\hat{\mathrm{U}}\mathrm{C}\right)\exp\left\{\pm z_{\alpha/2}\mathrm{S}\left(\mathrm{A}\hat{\mathrm{U}}\mathrm{C}\right)/\left(1-\mathrm{A}\hat{\mathrm{U}}\mathrm{C}\right)\right\}. \tag{22}$$

This CI can be calculated with the aid of the ROCKIT program.

9.2 Other methods:

Krzanowski & Hand (2009) discuss a method based on “placement values” and empirical likelihoods. More information can be found in their paragraph 3.5.1.

10. Choice of the optimal cut-off point

The ROC-curve shows, for a sequence of cut-off values (t), the relationship between the proportion of true positives (tp) versus the proportion of false positives (fp). The question now is whether or not there is an optimal value for t?

Consider the Youden index:

$$\begin{aligned}\mathrm{YI}&=\max\left(tp-fp\right)\\&=\max\left(tp+tn-1\right)\ ,\end{aligned}$$

i.e., the maximum value of the sum of the sensitivity (tp) and specificity (tn) minus 1. This index, like the AUC, is a descriptive measure of the ROC-curve. The optimal value of the cut-off point t is thus obtained when the sum $tp+tn$ is at its maximum.

Remarks:

(a) For a given t: $F(t) = \mathrm{P}(\mathrm{X}_N \leq t) = tn$ and

$$G(t) = \mathrm{P}(\mathrm{X}_S \leq t) = 1 - tp \qquad ,$$

it follows that

$$\begin{aligned}\mathrm{YI} &= \max(tp + tn - 1)\\ &= \max_t (F(t) - G(t)) \ .\end{aligned}$$

(b) Another index that is also used is the maximum vertical distance (MVD):

$$\begin{aligned}\text{MVD} = &\max_{\mathrm{t}} |\mathrm{P}(\mathrm{X} > \mathrm{t} \,|\, \mathrm{S}) - \mathrm{P}(\mathrm{X} > \mathrm{t} \,|\, \mathrm{N})|\\ &= \max_{\mathrm{t}} |1 - \mathrm{G}(\mathrm{t}) - (1 - \mathrm{F}(\mathrm{t}))|\\ &= \max_{\mathrm{t}} |\mathrm{F}(\mathrm{t}) - \mathrm{G}(\mathrm{t})| \ ,\end{aligned}$$

which is identical to YI if $F(t) \geq G(t)$, for all t. MVD is also known as the Kolmogorov-Smirnov measure which is used to compare two distribution function $F(t)$ and $G(t)$.

(c) $\mathrm{YI} > 0$ implies that $F(t) \geq G(t)$ for each t, which means that the distribution of X_S lies largely to the right of the distribution of X_N (see, for example, Figure 1). If $\mathrm{YI} \leq 0$ it means that the screening test is no better than simply randomly classifying individuals as positive or negative.

The estimated optimal t can thus be found where the estimated difference $\hat{F}(t) - \hat{G}(t)$ is a maximum. The following four methods for determining the optimal t (denoted by t^*) are discussed by Krzanowski & Hand (2009), paragraph 9.4:

10.1 binormal method:

When F and G are both normal

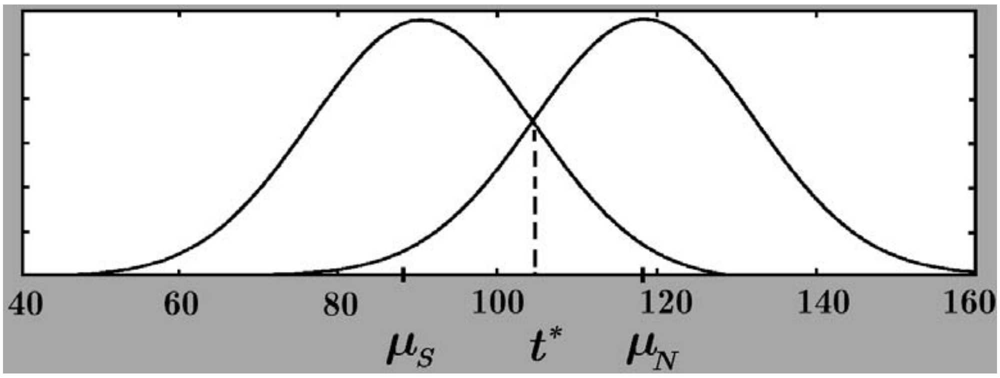
$$\mathrm{YI} = \max_{\mathrm{t}} \left[\Phi\left(\frac{\mathrm{t} - \mu_{\mathrm{N}}}{\sigma_{\mathrm{N}}} \right) - \Phi\left(\frac{\mathrm{t} - \mu_{\mathrm{S}}}{\sigma_{\mathrm{S}}} \right) \right],$$

which, after setting the first derivative to zero and solving, we get:

$$t^* = \frac{\mu_S\sigma_N^2 - \mu_N\sigma_S^2 - \sigma_N\sigma_S\sqrt{\left(\mu_N-\mu_S\right)^2+\left(\sigma_N^2-\sigma_S^2\right)\ln\left(\sigma_N^2/\sigma_S^2\right)}}{\left(\sigma_N^2-\sigma_S^2\right)} \quad . \tag{23}$$

If $\sigma_N^2 = \sigma_S^2 = \sigma^2$, then $t^* = ½\left(\mu_N+\mu_S\right)$, i.e., the optimal value lies halfway between the means of the distributions and where the normal density functions intersect each other (see Figure 4).

Figure 4:



To estimate t^* with $\hat{t}^*$, the estimators $\hat{\mu}_N$, $\hat{\mu}_S$, $\hat{\sigma}_N^2$ and $\hat{\sigma}_S^2$ are substituted in (23).

10.2 Transformed normal method:

The assumption that $G(t)$ and $F(t)$ are normally distributed is sometimes unrealistic which means that, like in the estimation of the ROC-curve, an appropriate monotone transformation (Box-Cox transformation) can be applied to X to achieve normality. Just as the ROC-curve is invariant under monotone transformations, YI is also invariant. The optimal value t^* can then be determined as in paragraph 10.1, but on the distribution of Y , after which it can be back-transformed in terms of X .

10.3 Empirical method:

F and G can be estimated with their empirical distribution functions

$$\hat{\mathrm{F}}(\mathrm{t}) = \mathrm{n}'_{\mathrm{N(t)}}/\mathrm{n}_{\mathrm{N}} \tag{24}$$

$$\hat{G}(t) = n'_{S(t)} / n_S \ ,$$

where $n'_{A(t)}$ is the number of individuals from population A such that its X values are smaller than or equal to t.

The value $\hat{t}^*$ is then the t-value in a sequence of values that makes $\hat{F}(t) - \hat{G}(t)$ a maximum.

10.4 Kernel estimation methods:

Here $F(t)$ and $G(t)$ are determined using kernel estimators of the density functions f_s and f_g (see the nonparametric estimation of ROC-curves discussed above).

11. Adjustment of ROC-curves

Suppose that the screening test's result X is influenced by other variables, it can be that adjusting X for these variables (called covariates) can weaken or improve the prediction ability. An example is when the waist circumference of a person is used as a screening test for hypertension and then adjusted for age and gender. Paragraph 5.2 in Krzanowski & Hand (2009) provides two approaches for doing this: indirect and direct adjustment.

11.1 Indirect adjustment:

Consider the following linear relationships:

$$X_N = \alpha_N + \beta_{N1} Z_{N1} + \beta_{N2} Z_{N2} + \cdots + \beta_{Nk} Z_{Nk} + \varepsilon_N$$

$$X_S = \alpha_S + \beta_{S1} Z_{S1} + \beta_{S2} Z_{S2} + ... + \beta_{Sm} Z_{Sm} + \varepsilon_S , \qquad (25)$$

where α_N and α_S are intercept constants and $\beta_{N1}, \cdots, \beta_{Nk}$ and $\beta_{S1}, \cdots, \beta_{Sm}$ are the regression coefficients of the k covariates $Z_{N1}, \cdots, Z_{Nk}$ and m covariates $Z_{S1}, \cdots, Z_{Sm}$ and ε_N and ε_S are normally distributed with zero mean and variance σ_N^2 and σ_S^2 .

This is once again the binormal model where the differences between the means of X_N and X_S now vary for different values of $Z_{N1} = z_{N1}, \cdots, Z_{Nk} = z_{Nk}$ and $Z_{S1} = z_{S1}, \cdots, Z_{Sm} = z_{Sm}$ in the following way:

$$\mu_n = \alpha_N + \beta_{N1} z_{N1} + \ldots. + \beta_{Nk} z_{Nk}$$

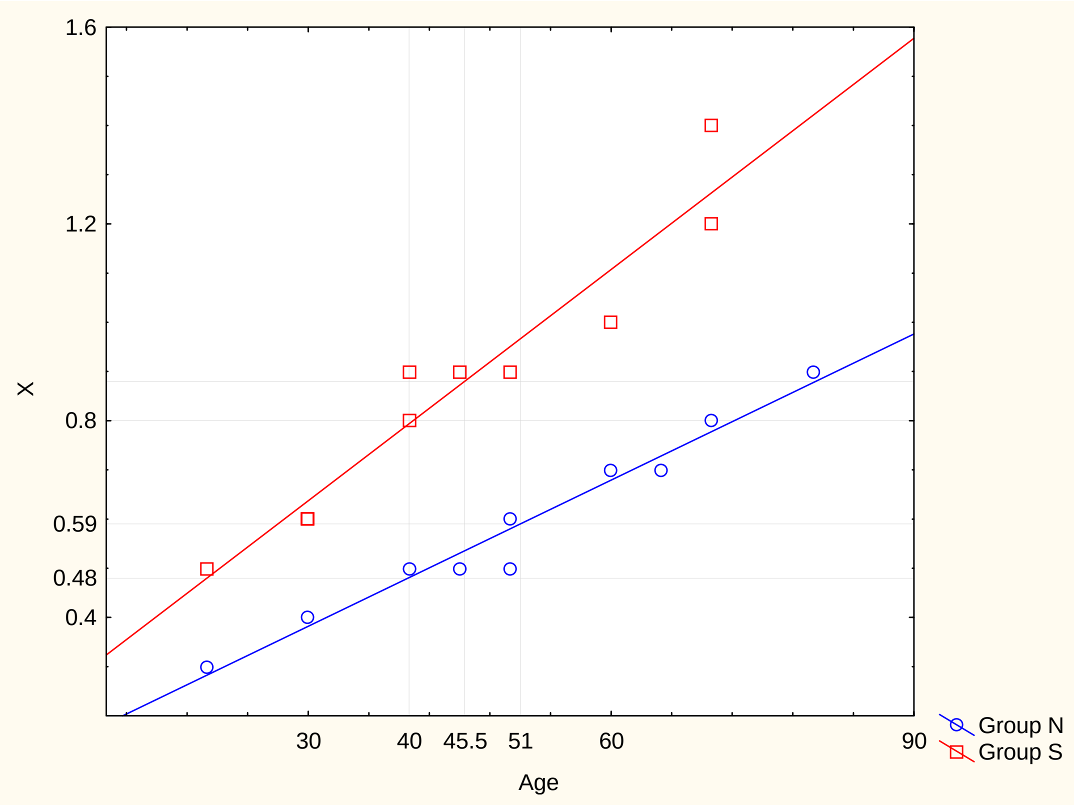
and (26)

$$\mu_S = \alpha_S + \beta_{S1} z_{S1} + \ldots + \beta_{Sm} z_{Sm}.$$

The ROC-curve is thus obtained from this set of Z-values using (7) and (8) above and from the AUC in (9). Further, the optimal cut-off point t^* can be determined from (23) for each given set of Z-values.

Figure 5 illustrates the adjustments on μ_N and μ_S in the case where "Age", Z, is a common covariate. For a given value $Z = z_O$ = 40, μ_N is estimated by $\hat{\mu}_N(40)$ as 0.48, which is lower than $\overset{\wedge}{\mu}_N$, which was found to be 0.59 (obtained by evaluating $\hat{\mu}_N(51.0)$, i.e., at the mean value z = 51.0), the mean value of the sample from population N. The value $\hat{\mu}_N(40)$ is obtained from the linear regression equation $x = 0.085 + 0.0099z$ at $z = 40$. Similarly, $\hat{\mu}_S(40)$ = 0.80 is obtained from the regression equation $x = 0.1669 + 0.0157z$ at $z = 40$, which is lower than the sample mean $\overset{\wedge}{\mu}_S$ = 0.88 (obtained by evaluating $\hat{\mu}_S(51)$, i.e., at the mean value z = 45.5). The strength of linear relationship with Age is stronger in population S than population N. This is clear because the gradient of population S, β_1 = 0.016, is larger than the gradient for population N, β_2 = 0.010. The variances σ_N^2 and σ_S^2 measure the vertical variation of observations around the regression line and are, in general, smaller than the usual variances where no relationship with Z is assumed. Comparing the variances we see that it is 0.0014 vs. 0.034 for population N and 0.0065 vs. 0.077 for population S.

Figure 5:



Estimation of ROC-curves and AUC is then done by replacing $\alpha_N, \beta_{N1}, \cdots, \beta_{Nk}$ and $\alpha_S, \beta_{S1}, \cdots, \beta_{Sm}$ with their least squares estimators. These are easily obtained by fitting a multiple linear regression of X_N on $Z_{N1}, \cdots, Z_{Nk}$ and X_S on $Z_{S1}, \cdots, Z_{Sk}$ using any statistical software package (like, for example, Statistica or SPSS). Further, the estimated values of σ_N and σ_S represent the standard error of estimation of these multiple regression models. In the case where one has only one covariate for each X_N and X_S, then Faraggi (2003) provides the following approximate $(1-\alpha)100\%$ confidence interval for AUC for given $\mathrm{Z_N}$ and $\mathrm{Z_S}$ values (denoted by $\mathrm{z_N}$ and $\mathrm{z_S}$):

$$\mathrm{AUC}(\mathrm{z_N}, \mathrm{z_S}) \pm \left\{ \frac{1}{\hat{\mathrm{M}}(\mathrm{z_N}, \mathrm{z_S})} + \frac{\left(\mathrm{A}\hat{\mathrm{UC}}(\mathrm{z_N}, \mathrm{z_S})\right)^2}{2\hat{\mathrm{f}}} \right\} \mathrm{z}_{\alpha/2} \ , \qquad (27)$$

where

$$\hat{M}\left(z_N, z_S\right) = \frac{\hat{\sigma}_N^2 + \hat{\sigma}_S^2}{\hat{a}_N^2 \hat{\sigma}_N^2 + \hat{a}_S^2 \hat{\sigma}_S^2} \quad , \tag{28}$$

with

$$\hat{a}_N^2 = \frac{1}{n_N} \sum_{i=1}^{n_N} \left(z_{Ni} - z_N\right)^2 / \left[\left(n_N - 1\right) S_{Z_N}^2\right]$$

and (29)

$$\hat{a}_S^2 = \frac{1}{n_S} \sum_{i=1}^{n_S} \left(z_{Si} - z_S\right)^2 / \left[\left(n_S - 1\right) S_{Z_S}^2\right].$$

Here $S_{Z_N}^2$ and $S_{Z_S}^2$ are the variances of the samples values of Z_N and Z_S.
Further, we have that

$$\hat{f} = \frac{\hat{\sigma}_N^2 + \hat{\sigma}_S^2}{\hat{\sigma}_N^4 / \left(n_N - 1\right) + \hat{\sigma}_S^4 / \left(n_S - 1\right)} \quad . \tag{30}$$

When one has multiple covariates Faraggi (2003) generalises the CI in (27) and also provides a method for determining CIs for YI and t^* using parametric bootstrap methods.

If the error term ε_N and ε_S in (25) cannot be assumed to be normally distributed, then Faraggi (2003) states that an appropriate Box-Cox transformations (based on an investigation of the residuals of the multiple regression) should be applied to X_N and X_S in an attempt to make the distributions of ε_N and ε_S more normal.

Krzanowski & Hand (2009) refer to more methods that can be used when one cannot assume normality of ε_N and ε_S, as well as methods more general than least squares for estimating μ_N, μ_S, σ_N^2 and σ_S^2 when they are functions of the regression parameters.

11.2 Direct adjustment:

Without providing any details, the reader is referred to the ROC-GLM model of the form:

$$h(y) = b(x) + \beta_1 Z_1 + \cdots + \beta_k Z_k + \varepsilon$$

where h and b are unknown monotone functions of the sensitivity y and specificity x, while $Z_1,\cdots,Z_k$ are covariates. Krzanowski & Hand (2009) provide various references to methods that directly estimate the ROC curve by including the covariates (see their paragraph 5.2.2). They then also discuss methods of adjusting the AUC using the regression model

$$h(AUC) = \alpha + \beta_1 Z_1 + \cdots + \beta_k Z_k + \varepsilon,$$

where h is a strictly monotone transformation on AUC (which lies on (0;1)) to make it lie on $(-\infty;\infty)$. An example of h is the logistic function:

$$h(x) = \log\frac{x}{1-x}.$$

Example 2:

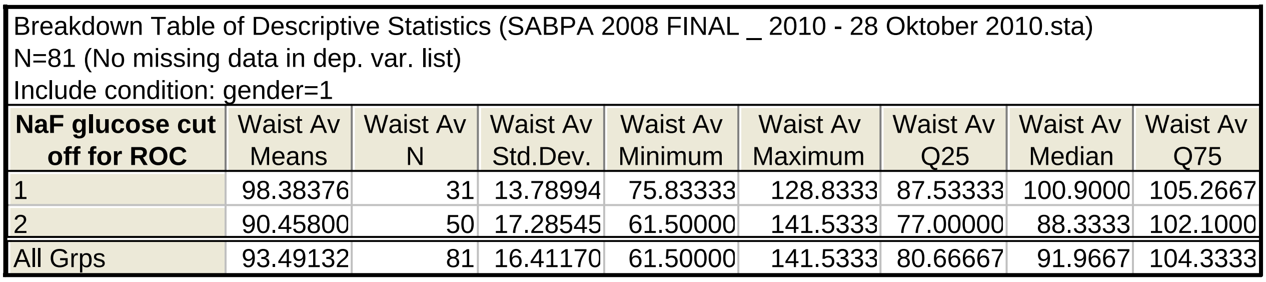
A possible screening test for testing the blood sugar levels of 81 black, male teachers in the NW province is to measure the circumference of their waists. There were 31 men whose Na F-Glucose levels were above the cut-off point of 5.6 mmol/litre. These individuals were thus “positive” and represent the sample from population S. The remaining 50 individuals were “negative” and represent the sample from population N.

Figure 6 displays the histograms of the two samples (ROC=1 : from S, ROC=2 : from N) with interval midpoints 70, 80, $\cdots$, 150, while Table 2 provides the descriptive statistics of each sample. Figure 7 shows the normal probability plot of each sample. We see from these plots that we can accept normality of these populations. Figure 8 confirms this by showing that the best Box-Cox transformation does little to improve on normality.

Figure 6:

Figure 7:

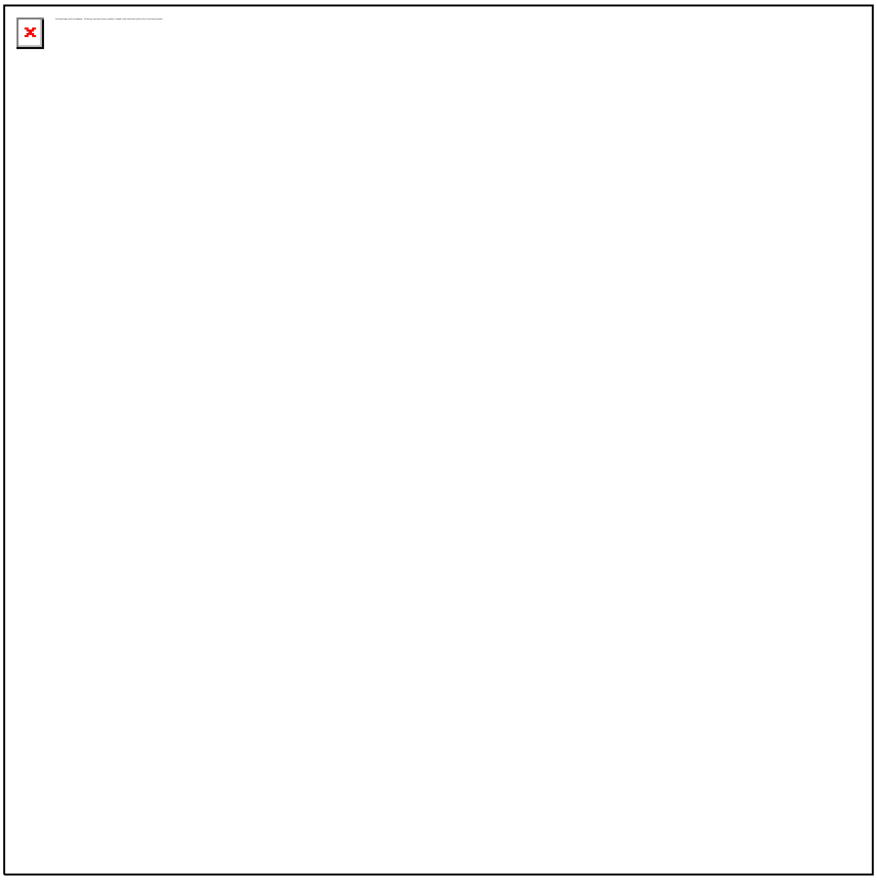
Table 2:



Breakdown Table of Descriptive Statistics (SABPA 2008 FINAL _ 2010 - 28 Oktober 2010.sta)
N=81 (No missing data in dep. var. list)
Include condition: gender=1

NaF glucose cut off for ROC	Waist Av Means	Waist Av N	Waist Av Std.Dev.	Waist Av Minimum	Waist Av Maximum	Waist Av Q25	Waist Av Median	Waist Av Q75
1	98.38376	31	13.78994	75.83333	128.8333	87.53333	100.9000	105.2667
2	90.45800	50	17.28545	61.50000	141.5333	77.00000	88.3333	102.1000
All Grps	93.49132	81	16.41170	61.50000	141.5333	80.66667	91.9667	104.3333

Figure 8:



Subsequently, the ROC-curve is determined using the empirical method of SPSS (SPSS Inc. (2007)) and is shown in Figure 9. Table 3 (SPSS output) supplies the area under the ROC curve (AUC), calculated from the Mann-Whitney statistic, together with the standard error and a 95% CI. The lower bound of 0.537 indicates that the AUC is indeed significantly higher than 0.5.

Table 3:

Test Result Variable(s):Waist Circumference

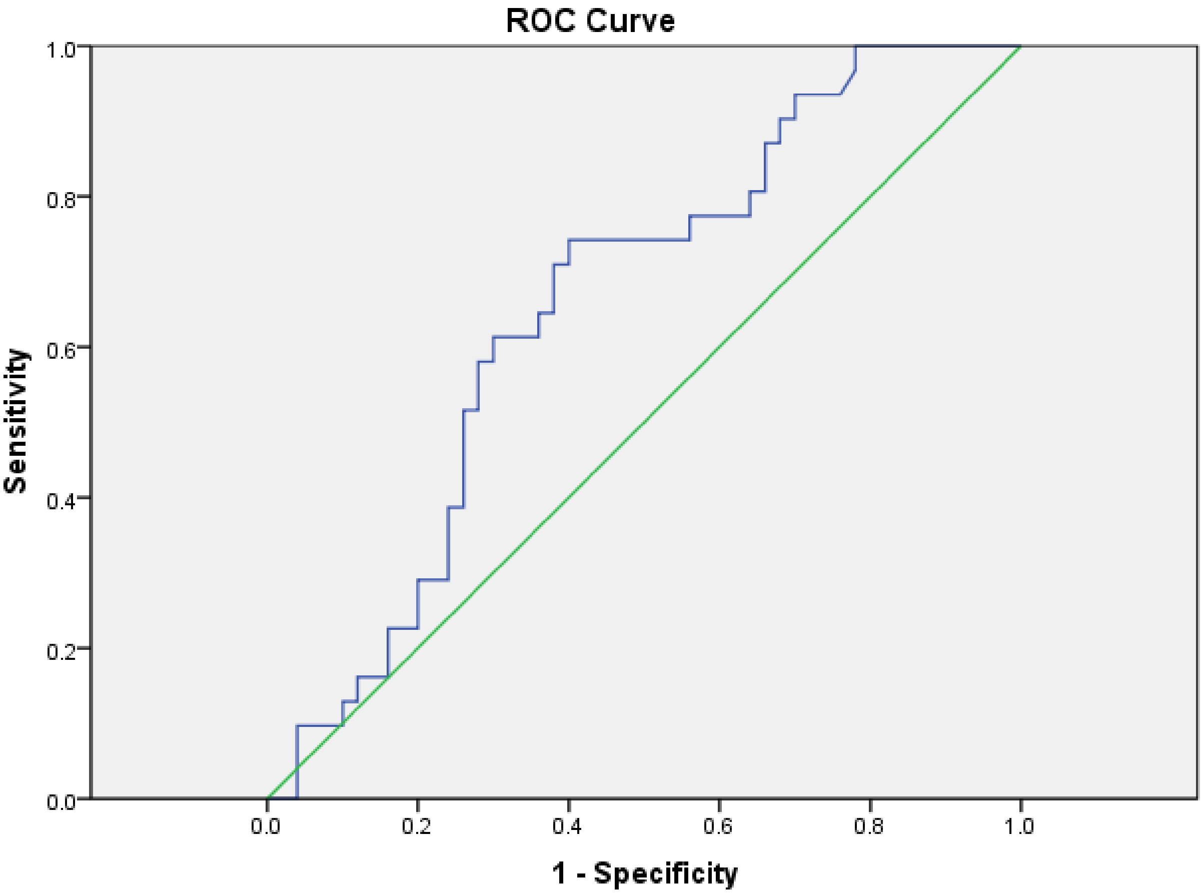
Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.656	.061	.018	.537	.776

The test result variable(s): Waist Circumference has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

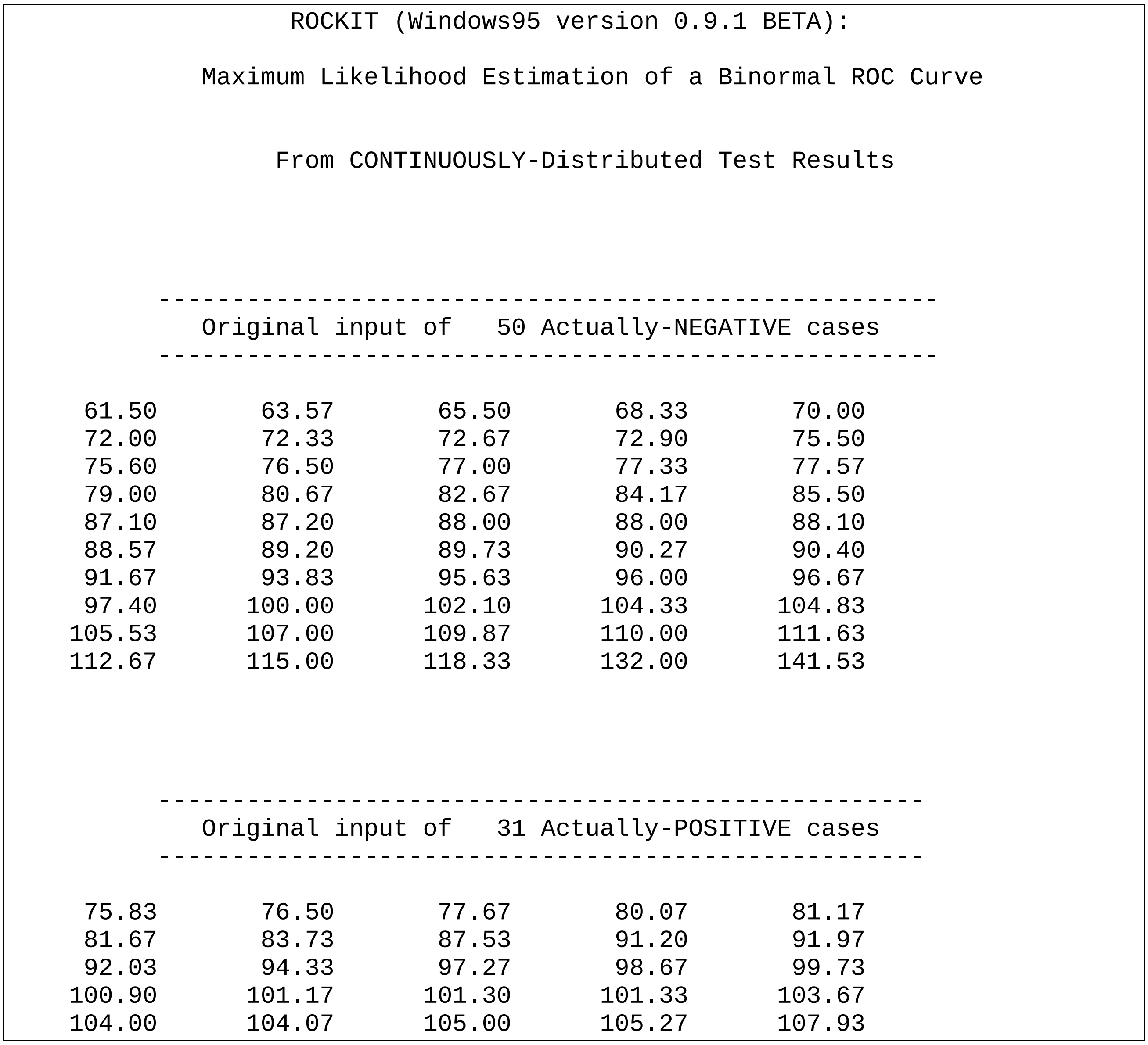
Figure 9:



Diagonal segments are produced by ties.

By assuming normality and using the binormal model, Figure 10 displays the output obtained from the ROCKIT program when applied to the data. At (A) in Figure 10 we find the estimated parameters a and b, and AUC(Az) is the AUC value according to (9), while AUC (Wilc) corresponds to the value given in the Mann-Whitney test (also called the Wilcoxon test). At (B) we find the standard error of a and b along with the correlation between a and b from which the variance V in (17) can be calculated to ultimately obtain a CI for tp for given values of fp. At (C) 95% CIs for a, b and AUC are given. The latter agrees with the output given in SPSS.

Figure 10:



```
                      ROCKIT (Windows95 version 0.9.1 BETA):

                Maximum Likelihood Estimation of a Binormal ROC Curve


                     From CONTINUOUSLY-Distributed Test Results




          -----------------------------------------------------
             Original input of   50 Actually-NEGATIVE cases
          -----------------------------------------------------

     61.50        63.57        65.50        68.33        70.00
     72.00        72.33        72.67        72.90        75.50
     75.60        76.50        77.00        77.33        77.57
     79.00        80.67        82.67        84.17        85.50
     87.10        87.20        88.00        88.00        88.10
     88.57        89.20        89.73        90.27        90.40
     91.67        93.83        95.63        96.00        96.67
     97.40       100.00       102.10       104.33       104.83
    105.53       107.00       109.87       110.00       111.63
    112.67       115.00       118.33       132.00       141.53




          ----------------------------------------------------
             Original input of   31 Actually-POSITIVE cases
          ----------------------------------------------------

     75.83        76.50        77.67        80.07        81.17
     81.67        83.73        87.53        91.20        91.97
     92.03        94.33        97.27        98.67        99.73
    100.90       101.17       101.30       101.33       103.67
    104.00       104.07       105.00       105.27       107.93
    109.10      110.67      112.10      118.57      126.63
    128.83




             Maximum Likelihood Estimation of the Parameters
                      a Single Binormal ROC Curve

      Name of Input File being used: Glukose.prn.txt29



      Condition 1: Glucose

      Total number of actually-negative cases =   50.
      Total number of actually-positive cases =   31.

      Data collected on a nominally continuous scale.
      Larger values of the test result represent stronger evidence that the
      case is actually-positive (e.g., that the patient is actually
abnormal)


    Operating Points Corresponding to the Input Data Categorized by the
LABROC5 Scheme:

    FPF:  .000  .100  .160  .240  .240  .260  .260  .280  .300  .360  .360
    TPF:  .000  .129  .226  .290  .387  .387  .516  .516  .581  .613  .645

    FPF:  .380  .400  .400  .560  .560  .640  .640  .680  .780 1.000
    TPF:  .710  .710  .742  .742  .774  .774  .806  .871 1.000 1.000



             -----------------------------------------------------
               Initial Estimates of the Binormal ROC Parameters:
             -----------------------------------------------------

                  a =   .7608
                  b =  1.0627


                Procedure Converges after   5 Iterations


               =====================================================
                  Final Estimates of the Binormal ROC Parameters
               =====================================================


 Binormal Parameters and Area Under the Estimated ROC :
        a           =        .7411
        b           =       1.4932
        Area (Az)   =        .6600                              (A)
        Area (Wilc) =        .6565
```

```
Estimated Standard Errors and Correlation of these Values:
       Std. Err. (a)  =      .3054
       Std. Err. (b)  =      .3026
       Corr(a,b)      =      .3054
       Std. Err. (Az) =      .0593                        (B)
       Std. Err.(Wilc)=      .0639

Symmetric 95% Confidence Intervals
       For a :        (  .1425, 1.3396)
       For b :        (  .9001, 2.0863)                   (C)

Asymmetric 95% Confidence Interval
       For Az:        (  .5378,  .7672)
```

Table 4:

```
  Estimated Binormal ROC curve, with Lower and Upper
Bounds of the Asymmetric Point-wise 95% Confidence
Interval for True-Positive Fraction at a Variety
of False-Positive Fractions:


    FPF      TPF        (Lower Bound, Upper Bound)

    .005     .0009       (   0.0000  ,    .0500   )
    .010     .0031       (   0.0000  ,    .0797   )
    .020     .0100       (    .0002  ,    .1262   )
    .030     .0193       (    .0008  ,    .1648   )
    .040     .0305       (    .0019  ,    .1989   )
    .050     .0431       (    .0036  ,    .2299   )
    .060     .0569       (    .0060  ,    .2587   )
    .070     .0717       (    .0091  ,    .2857   )
    .080     .0873       (    .0131  ,    .3112   )
    .090     .1036       (    .0180  ,    .3355   )
    .100     .1204       (    .0236  ,    .3588   )
    .110     .1377       (    .0301  ,    .3812   )
    .120     .1554       (    .0374  ,    .4028   )
    .130     .1734       (    .0456  ,    .4237   )
    .140     .1916       (    .0544  ,    .4439   )
    .150     .2100       (    .0640  ,    .4636   )
    .200     .3031       (    .1216  ,    .5542   )
    .250     .3953       (    .1906  ,    .6347   )
    .300     .4835       (    .2658  ,    .7064   )
    .400     .6418       (    .4178  ,    .8249   )
    .500     .7707       (    .5567  ,    .9098   )
    .600     .8684       (    .6768  ,    .9624   )
    .700     .9362       (    .7792  ,    .9886   )
    .800     .9771       (    .8662  ,    .9981   )
    .900     .9960       (    .9398  ,    .9999   )
    .950     .9993       (    .9717  ,   1.0000   )
```

Figure 11:

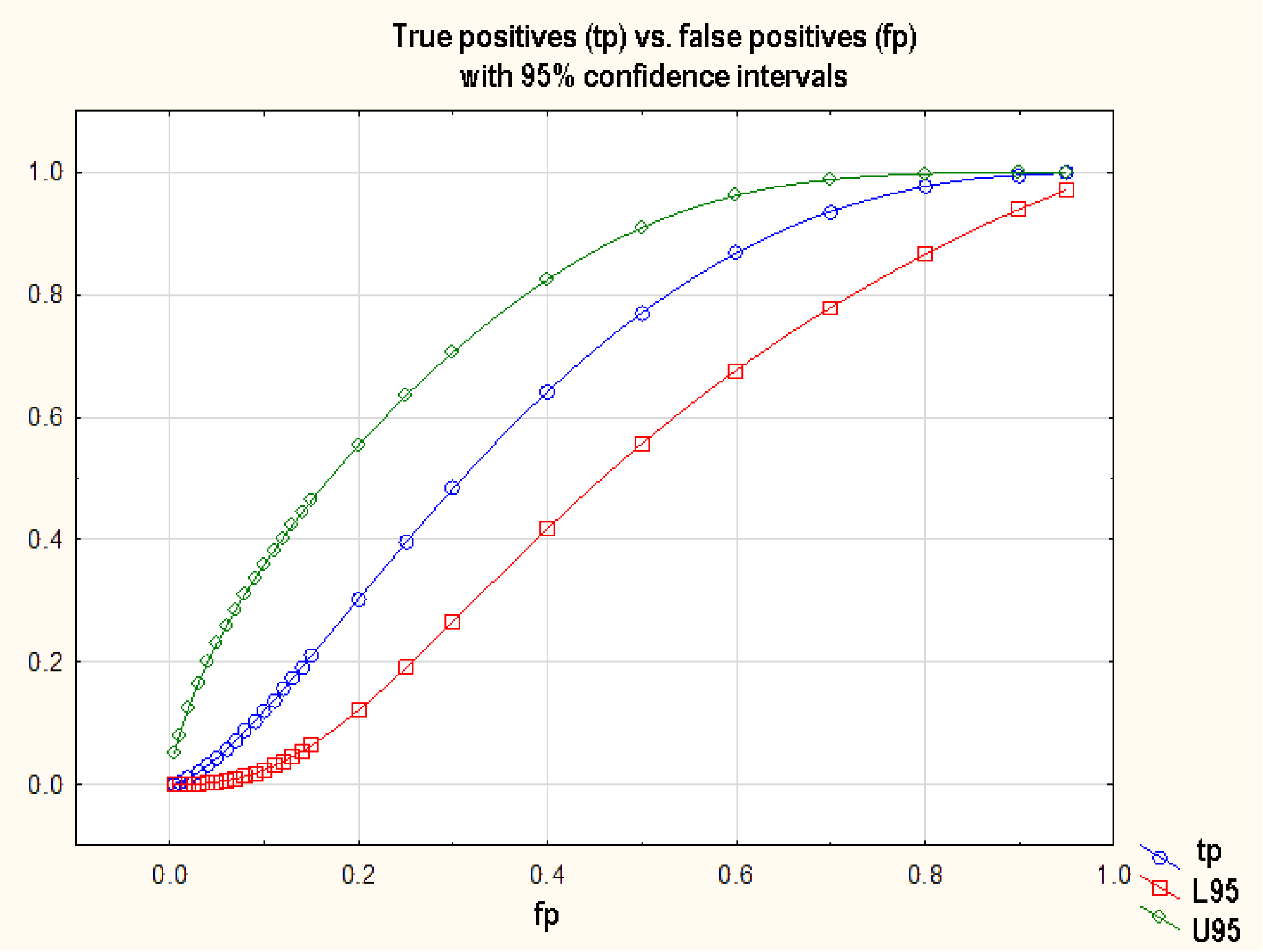
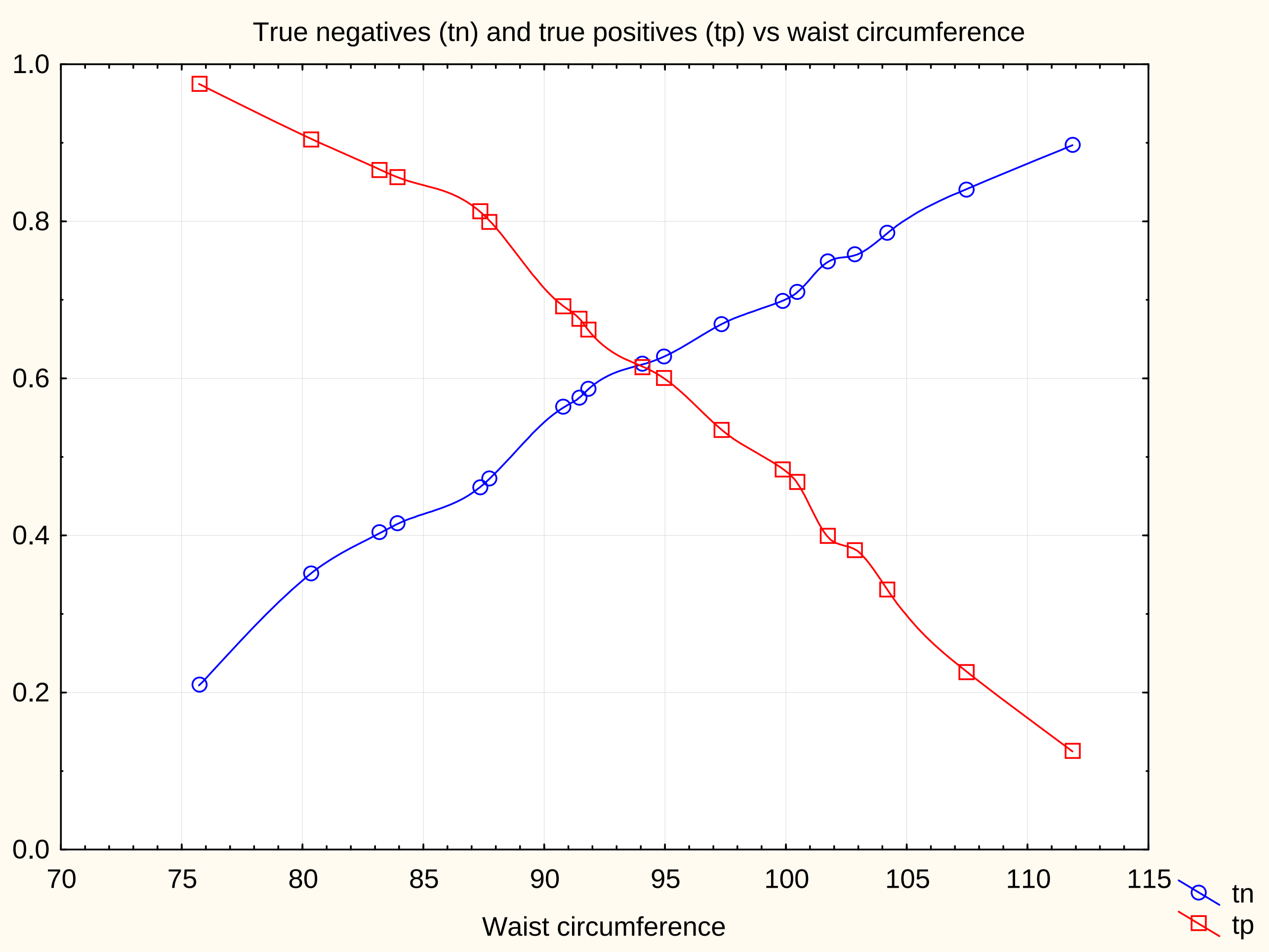


Table 4 shows the ROCKIT output of these CIs for tp (denoted by TPF) for a given sequence of fp values (denoted by FPF). Figure 11 displays the smoothed ROC cure with the 95% CIs (L95, U95). From Table 5 of the ROCKIT output one can construct Figure 12, where the sensitivity (tp) and specificity (tn) are represented as functions of the various cut-off points of waist circumference. If populations N and S had the same variances, then the waist circumference where the two curves intersected would be the optimal cut-off point.

Table 5:

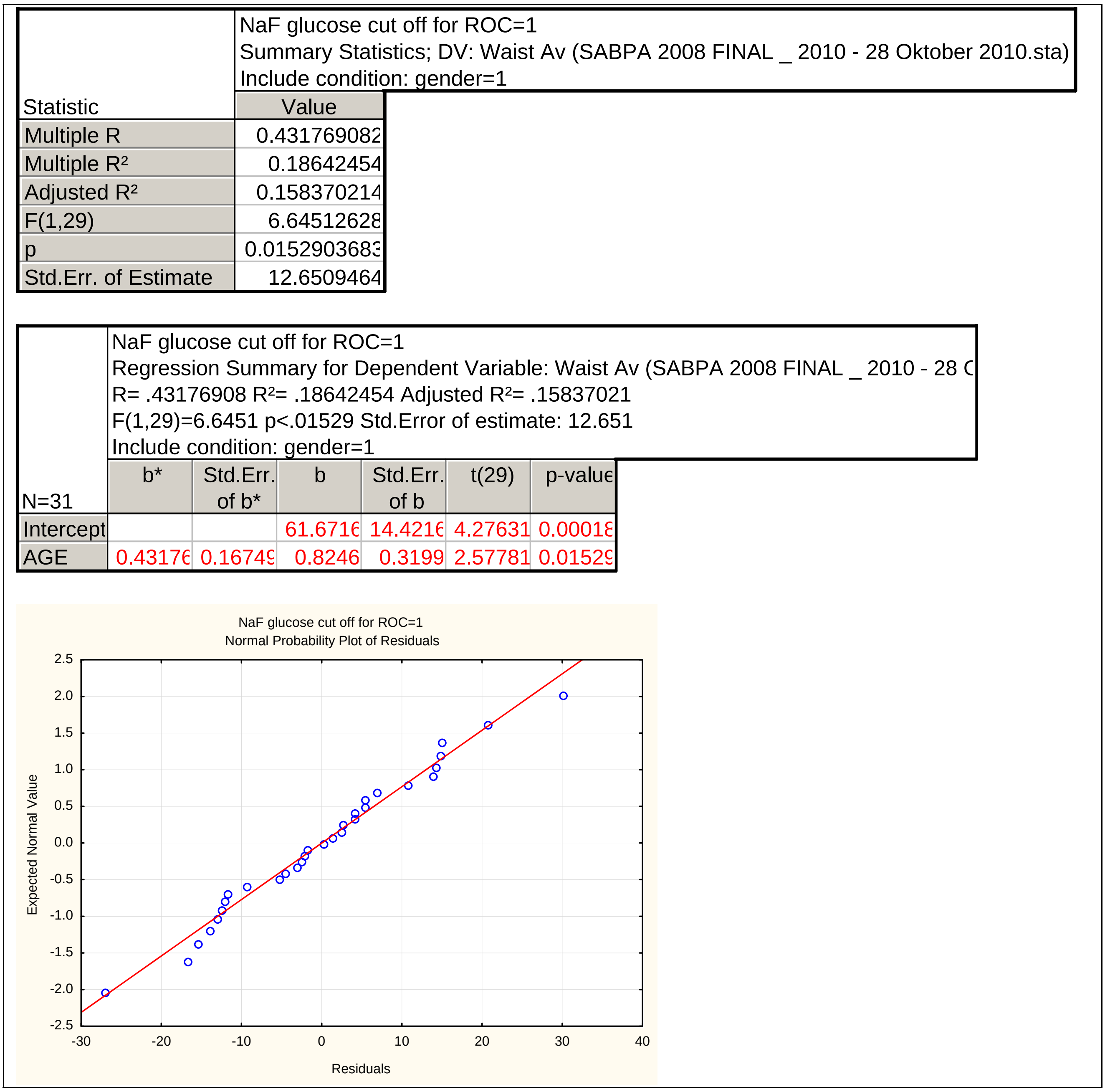
```
     Estimated Relationship between the Critical Test-Result Value
     (which separates 'positive' results form 'negative' results)
     and the Corresponding Operating Point on the Fitted Binormal
     ROC Curve:

 *********************************************************************

         Critical Test           (  FPF ,   TPF )
         Result Value

              111.865            (  .103,   .125)
              107.465            (  .159,   .227)
              104.200            (  .215,   .331)
              102.885            (  .243,   .382)
              101.715            (  .252,   .399)
              100.450            (  .291,   .468)
               99.865            (  .301,   .485)
               97.335            (  .331,   .535)
               94.980            (  .372,   .600)
               94.080            (  .382,   .615)
               91.820            (  .414,   .662)
               91.435            (  .425,   .677)
               90.800            (  .437,   .692)
               87.765            (  .527,   .800)
               87.365            (  .538,   .812)
               83.950            (  .585,   .856)
               83.200            (  .597,   .866)
               80.370            (  .648,   .905)
               75.715            (  .791,   .975)
```

Figure 11:



If one assumes that the Age is related to the Waist Circumference, then the ROC curve can be adjusted for it. Figure 13 shows the STATISTICA output of a linear regression of each of X_N and X_S on Z (Age). It would seem that a normal person (ROC=2) has practically no linear relationship with the age variable $\left(r^2 = 0.054\right)$, while the high blood pressure group (ROC=1) shows some relationship $\left(r^2 = 0.186\right)$. Table 6 shows the results of the adjustment using the indirect method. The intercept and regression coefficient for the group with high blood pressure and the normal group, as well as the standard error of estimation of the regression analysis serve as the input for an EXCEL worksheet "Calculate a and $b.xlsx$" which can then be used to calculate the adjusted mean, and a and b. The AUC is then calculated from a and b, while the optimal cut-off point is determined using (23). The calculations are done at Age 30, the mean Ages of the two samples (44.52 and 41.48), and also at 50. It is clear that the AUCs remain reasonably constant for the mean Age and at 50, but that they are considerably smaller for 30-year olds. Therefore, the optimal cut-off point is also considerably smaller (73.2).

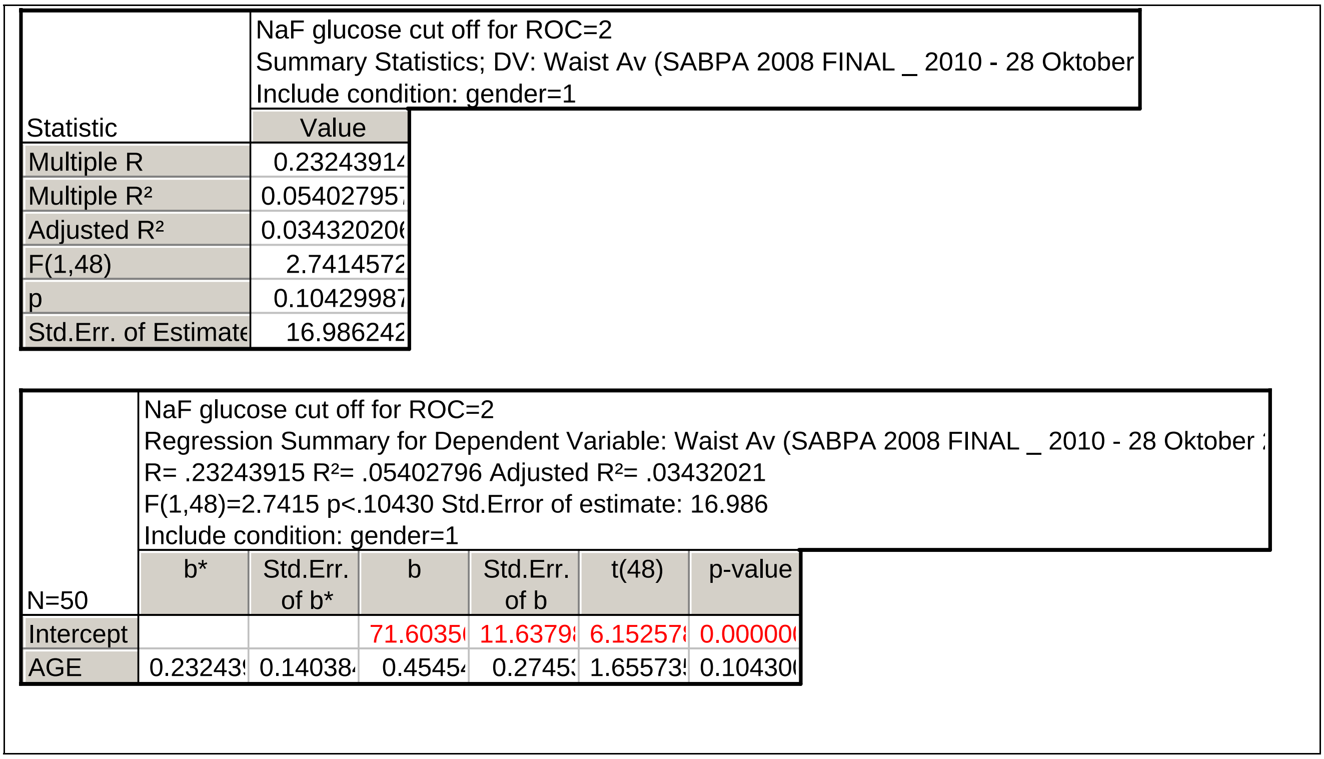
Figure 13:



Statistic	NaF glucose cut off for ROC=1 Summary Statistics; DV: Waist Av (SABPA 2008 FINAL _ 2010 - 28 Oktober 2010.sta) Include condition: gender=1 Value
Multiple R	0.431769082
Multiple R²	0.18642454
Adjusted R²	0.158370214
F(1,29)	6.64512628
p	0.0152903683
Std.Err. of Estimate	12.6509464

NaF glucose cut off for ROC=1
Regression Summary for Dependent Variable: Waist Av (SABPA 2008 FINAL _ 2010 - 28 C
R= .43176908 R²= .18642454 Adjusted R²= .15837021
F(1,29)=6.6451 p<.01529 Std.Error of estimate: 12.651
Include condition: gender=1

N=31	b*	Std.Err. of b*	b	Std.Err. of b	t(29)	p-value
Intercept			61.6716	14.4216	4.27631	0.00018
AGE	0.43176	0.16749	0.8246	0.3199	2.57781	0.01529



NaF glucose cut off for ROC=2
Summary Statistics; DV: Waist Av (SABPA 2008 FINAL _ 2010 - 28 Oktober
Include condition: gender=1

Statistic	Value
Multiple R	0.23243914
Multiple R²	0.05402795
Adjusted R²	0.03432020
F(1,48)	2.7414572
p	0.10429987
Std.Err. of Estimate	16.986242

NaF glucose cut off for ROC=2
Regression Summary for Dependent Variable: Waist Av (SABPA 2008 FINAL _ 2010 - 28 Oktober
R= .23243915 R²= .05402796 Adjusted R²= .03432021
F(1,48)=2.7415 p<.10430 Std.Error of estimate: 16.986
Include condition: gender=1

N=50	b*	Std.Err. of b*	b	Std.Err. of b	t(48)	p-value
Intercept			71.6035	11.6379	6.15257	0.00000
AGE	0.23243	0.14038	0.4545	0.2745	1.65573	0.10430

Finally, the means and standard deviations of the groups without any adjustments are used to determine the AUC and the optimal cut-off. The AUC of 0.64 at Age 50 does not differ very much from the AUC at the mean age (0.646), or the ROCKIT program's value (0.66), and the Mann-Whitney value (0.656). In practice, one would be able to calculate (without the availability of ROCKIT or SPSS) the means and standard deviations of the two samples if one were to assume normality of the data.

Table 6:

Age	Marker, group	Gender	Intercept	Regression coefficient (b)	SE of estimate	Variance	r-square	Age	mean	a	b	AUC	Optimum Threshold
30	Glucose, High	male	61.67	0.825	12.65	160.0	0.186	30	86.42	0.092	1.343	0.522	73.2
	Glucose, Low	male	71.6	0.455	16.99	288.7	0.054	30	85.25				
44.52	Glucose, High	male	61.67	0.825	12.65	160.0	0.186	44.52	98.40	0.627	1.343	0.646	88.6
	Glucose, Low	male	71.6	0.455	16.99	288.7	0.054	41.48	90.47				
41.48	Glucose, High	male	61.67	0.825	12.65	160.0	0.186	50	102.92	0.677	1.343	0.657	93.2
	Glucose, Low	male	71.6	0.455	16.99	288.7	0.054	50	94.35				
50	Glucose, High	male			13.79	190.2		all	98.38	0.574	1.254	0.640	89.0
	Glucose, Low	male			17.29	298.9		all	90.46				

References:

Kline, R. B. (2004). *Beyond significance testing. Reforming data analysis methods in behavioral sciences.* American Psychological Association, Washington, DC

Faraggi, D. (2003). Adjusting receiver operating characteristic curves and related indices for covariates. *The Statistician*, 52, 179-192

Krzanowski, W. J. & Hand, D. J. (2009). *ROC curves for continuous data*. Chapman & Hall, Boca Raton

Metz, C. E., Herman, B. A. & Shen, J. (1998). Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuous distributed data. *Statistics in Medicine*, 17, 1033 – 1053

SPSS Inc. (2007). SPSS® 16.0 for Windows, Release 16.0.0, Copyright© by SPSS Inc., Chicago, Illinois. www.spss.com

StatSoft, Inc. (2011). STATISTICA (data analysis software system), version 10. www.statsoft.com