# المحاضرة الحادية عشر:

# مدخل إلى نماذج المتغيرات الكيفية

# المحاضرة الحادية عشر: مدخل إلى نماذج المتغيرات الكيفية

في كل النماذج السابقة التي تطرقنا إليها سابقا كان المتغير التابع متغير كمي، في حين كانت المتغيرات المفسرة إما كمية أو مزيجا بين المتغيرات الكمية والكيفية.

سنحاول في هذا المحور أن نتناول نوعا من النماذج الأخرى أين يكون المتغير التابع " نوعيا"، وعلى الرغم من زيادة استعمال هذه النماذج إلا أن المشكل الرئيسي الذي يواجهنا هو تفسير نتائج التقدير.

## 1- النموذج الاحتمالي الخطي:

لنفرض أن $Y_i$ تمثل قرار المستهلك بشراء سلعة أم لا، وبافتراض أننا استجوبنا 1000 شخص i=1,2,.....1000 ، وكان المتغير المفسر هو السعر $price_i$ ، إذن النموذج سيكون كما يلي:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \;......(1)$$

حيث أن: $Y_i = \begin{cases} 1 \\ 0 \end{cases}$، حيث أن: 1 : تعني قرار الشراء، 0 تعني قرار عدم الشراء.

إذا كانت : $E(\varepsilon_i) = 0$، فإن: $E(Y_i) = \beta_1 + \beta_2 X_i$ ؛

من جهة أخرى فإن احتمال قرار الشراء هو $P(Y_i = 1) = P_i$ ، أما احتمال قرار عدم الشراء هو $P(Y_i = 0) = 1 - P_i$ ، وعليه القيمة المتوقعة لـ $Y_i$ يمكن صياغتها كما يلي:

$$E(Y_i) = 1.P(Y_i = 1) + 0.P(Y_i = 0) = P_i$$

أي أن: $P_i = \beta_1 + \beta_2 X_i$ .

كما يمكن القول: $E(Y_i/X_i) = \beta_1 + \beta_2 X_i = P_i$

اذا افترضنا أن المستهلك قرر الشراء، يعني أن: $Y_i = 1$ ، وبالتعويض في المعادلة (1) نجد أن:

$$\varepsilon_i = 1 - (\beta_1 + \beta_2 X_i) = P_i$$

أما إذا قرر المستهلك عدم الشراء فإن:

$$\varepsilon_i = 0 - (\beta_1 + \beta_2 X_i) = 1 - P_i$$

وعليه:

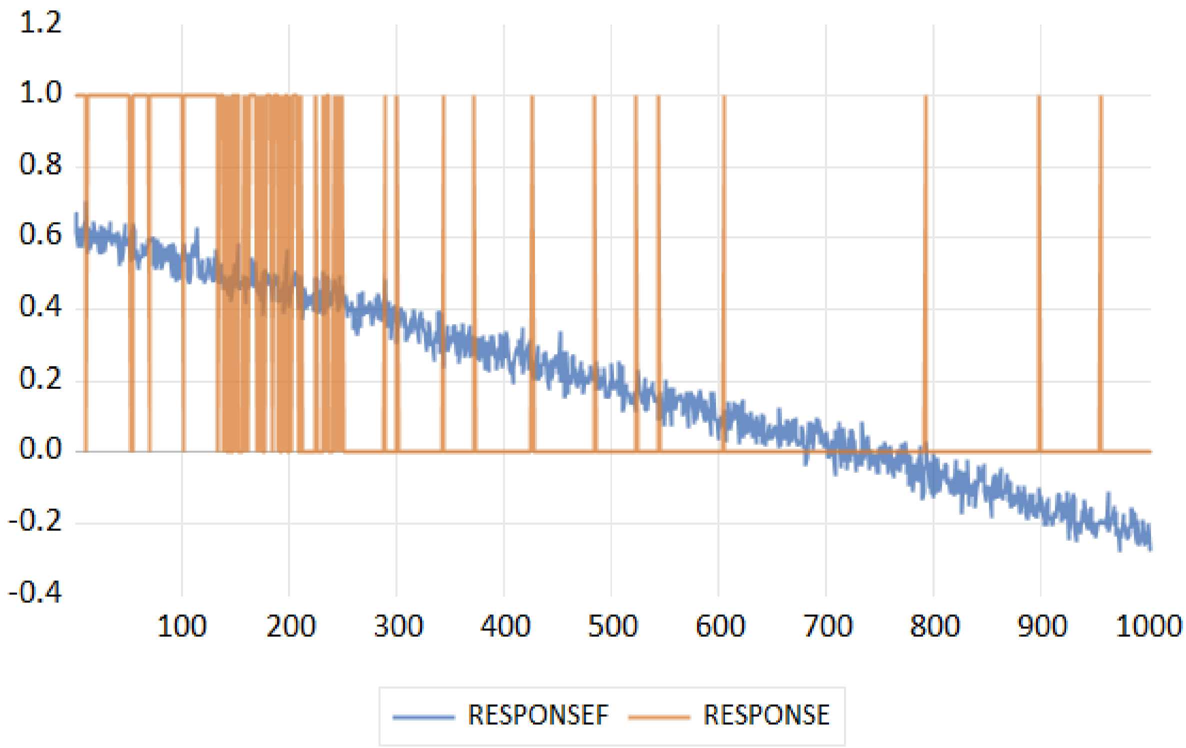
$$Var(\varepsilon_i) = E({\varepsilon_i}^2) = P_i(1 - \beta_1 + \beta_2 X_i)^2 + (1 - P_i)(-\beta_1 - \beta_2 X_i)^2$$

بعد التبسيط ستحصل على ما يلي:

$$Var(\varepsilon_i) = P_i(1 - P_i)$$

أي أن تباين الأخطاء العشوائية غير ثابت وهو ما يحدث مشكل عدم تجانس الأخطاء.

يضاف إلى هذا أن تطبيق طريقة المربعات الصغرى العادية سينتج عنه في أغلب الأحيان ضعف معامل التحديد، كما أن السلسلة الحقيقية والمقدرة بعيدتان كل البعد عن بعضهما، مثلما هو موضح في الشكل الموالي أن توفرت لنا عينة من 1000 مستجوب حول قرار الشراء كمتغير تابعن أما المتغير المستقل فهو السعر.




مع العلم أن السلسلة الحقيقية وهي قرار الشراء باللون الأصفر (تأخذ قيم 0 أو 1)، أما السلسلة المقدرة فهي باللون الأزرق، والتي نتجت من تطبيق طريقة المربعات الصغرى العادية على النموذج (1).

نلاحظ من خلال الشكل أن نتائج التنبؤ أعطت قيما سالبة، أي أنها تقع خارج الحدود (0، 1) ، يضاف إلى هذا أن الأخطاء العشوائية لا تتبع التوزيع الطبيعي بل تتبع توزيع برنولي.

وعليه لا بد من التفكير في نموذج آخر يراعي هذه المشاكل وهو ما سنقوم بتقديمه في المحاضرة الموالية.

# المحاضرة الثانية العاشرة:

# نموذج لوجيت بنيته الرياضية وكيفية تقديره واختبار معلماته

## المحاضرة الثانية عشر: نموذج لوجيت بنيته الرياضية وكيفية تقديره واختبار معلماته

يمثل نموذج لوجيت أحد النماذج المستخدمة بكثرة في حال كون المتغير التابع من النوع الكيفي، حيث يأخذ هذا النموذج الشكل الرياضي التالي:

$$\Pr[y_i = 1] = \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

كما أن:

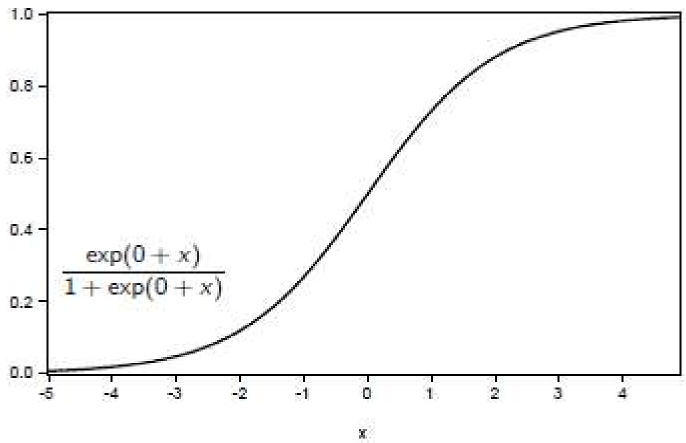
$$\Pr[y_i = 0] = 1 - \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)} = \frac{1}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

حيث أن: $Y_i = \begin{cases} 1 \\ 0 \end{cases}$، أي أن: $Y_i \sim Bernoulli(P_i)$

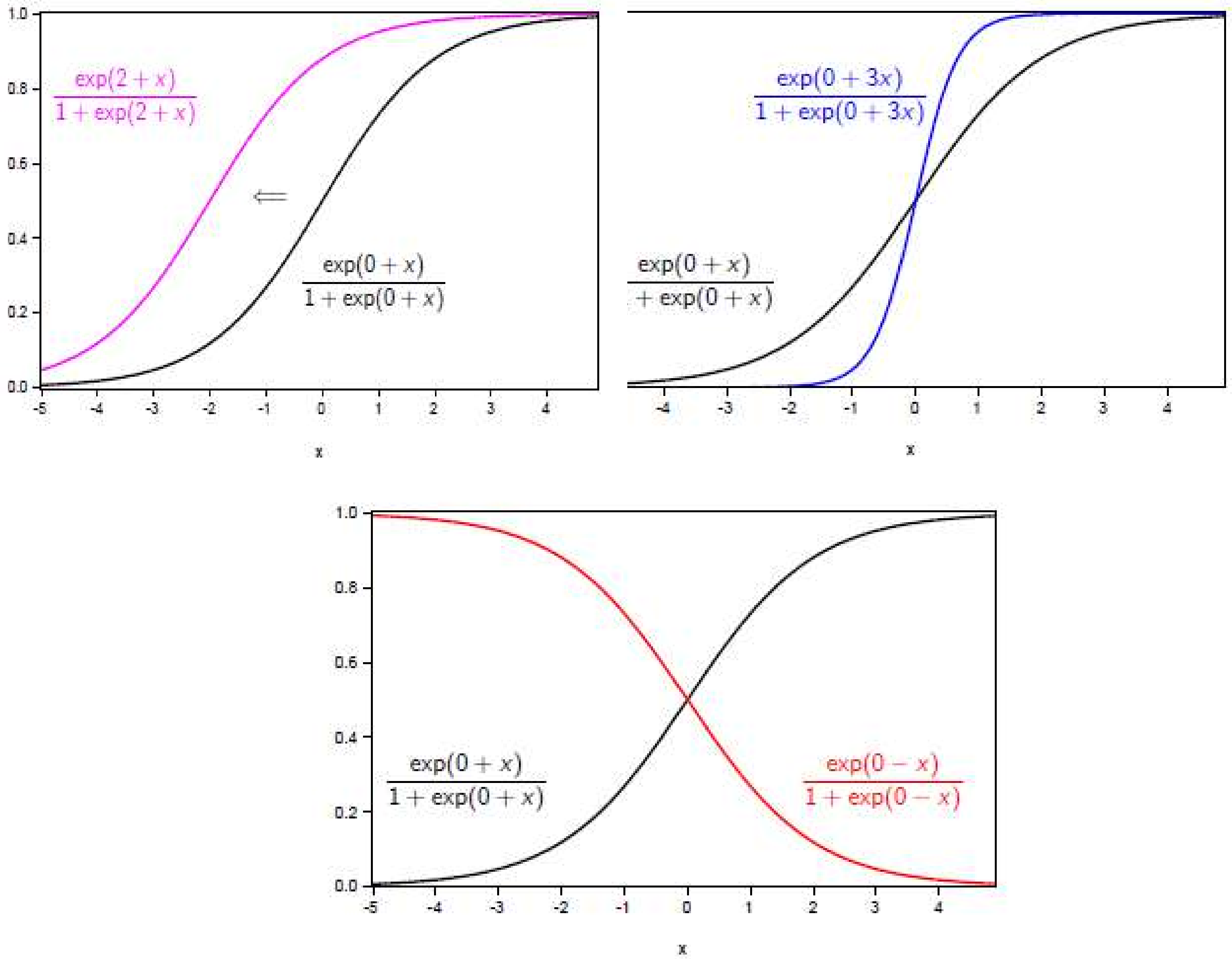
$$P(Y_i = 1) = P_i$$

$$P(Y_i = 0) = 1 - P_i$$

إن التمثيل الرياضي لدالة لوجيت يكون وفق الشكل التالي:




ولفهم طبيعة هذه الدالة أكثر، عند تعير معلمات هذا النموذج يمكن الاستعانة بالأشكال الموالية:




ضمن هذا النموذج يمكن التعرف على المصطلحات التالية:

- **نسبة الاحتمال Odds Ratio :**

وهي تمثل احتمال قرار الشراء إلى احتمال قرار عدم الشراء، وبصفة عامة النسبة بين احتمال وقوع الحادث $Y_i$ واحتمال عدم وقوعه، يمكن التعبير عن ذلك رياضيا في حالة نموذج بسيط يحتوي على متغيرة مفسرة واحدة ، كما يلي:

Logit model:

$$\Pr[y_i = 1] = \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

$$\Pr[y_i = 0] = \frac{1}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

Odds ratio:

$$\frac{\Pr[y_i = 1]}{\Pr[y_i = 0]} = \exp(\beta_1 + \beta_2 x_i)$$

Log odds ratio:

$$\log\left(\frac{\Pr[y_i = 1]}{\Pr[y_i = 0]}\right) = \beta_1 + \beta_2 x_i$$

أما في حالة وجود أكثر من متغير مفسر فإن نسبة الاحتمال ولوغاريتم نسبة الاحتمال يكون كما يلي:

$$\Pr[y_i = 1] = \frac{\exp(\beta_1 + \sum_{j=2}^{k} \beta_j x_{ji})}{1 + \exp(\beta_1 + \sum_{j=2}^{k} \beta_j x_{ji})}$$

Log odds ratio:

$$\log\left(\frac{\Pr[y_i = 1]}{\Pr[y_i = 0]}\right) = \beta_1 + \sum_{j=2}^{k} \beta_j x_{ji}$$

- **الأثر الحدي Marginal effect:**

وهو يعبر عن التغير في احتمال $Y_i = 1$ الناتج من تغير $X_i$ ن ويتم حسابه وفق العلاقة التالية:

$$\frac{d \Pr[y_i = 1]}{d\, x_i} = \Pr[y_i = 1] \Pr[y_i = 0] \beta_2$$

أما متوسط الأثر الحدي فيساوي:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{d \Pr[y_i = 1]}{d\, x_i} = \left(\frac{1}{n}\sum_{i=1}^{n} \Pr[y_i = 1] \Pr[y_i = 0]\right) \beta_2$$

- **تقدير نموذج لوجيت**

يتم تقدير نموذج لوجيت باستخدام طريقة المعقولية العظمى، ومن أجل تشكيل دالة المعقولية العظمى ننطلق من أجل المشاهدة $Y_i = 1$ ، حيث تكون هذه الدالة كما يلي:

$$\Pr[y_i = 1] = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}$$

أما من أجل المشاهدة $Y_i = 0$ ، فإن هذه الدالة تكون مساوية إلى:

$$\Pr[y_i = 0] = \frac{1}{1+\exp(x_i'\beta)}$$

ومن أجل المشاهدة $i$ تكون هذه الدالة كما يلي:

$$\left(\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right)^{y_i}\left(\frac{1}{1+\exp(x_i'\beta)}\right)^{1-y_i}$$

إن نموذج لوجيت وفق الشكل الشعاعي يكتب كما يلي:

$$\Pr[y_i = 1] = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)},$$

where $x_i = (1, x_{2i}, \ldots, x_{ki})'$ and $\beta = (\beta_1, \ldots, \beta_k)'$

يستحيل كتابة هذا النموذج وفق الشكل الخطي المتعارف عليه:

$$Y_i = X_i{}'\beta + \varepsilon_i$$

إن دالة المعقولية لـ $n$ مشاهدة مستقلة تكتب كما يلي:

$$L(\beta) = \prod_{i=1}^{n}\left(\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right)^{y_i}\left(\frac{1}{1+\exp(x_i'\beta)}\right)^{1-y_i}$$

لوغاريتم دالة المعقولية العظمى:

$$\begin{aligned}\log(L(\beta)) &= \sum_{i=1}^{n} y_i \log\left(\frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)}\right) + (1-y_i)\log\left(\frac{1}{1+\exp(x_i'\beta)}\right)\\ &= \sum_{i=1}^{n} y_i x_i'\beta - \log(1+\exp(x_i'\beta)),\end{aligned}$$

إن اشتقاق دالة المعقولية وجعلها مستوية للصفر، سيعطي لنا النتيجة التالية:

$$\frac{\partial \log(L(\beta))}{\partial \beta} = \frac{\partial \sum_{i=1}^{n} y_i x_i' \beta - \log(1 + \exp(x_i'\beta))}{\partial \beta} = 0$$
$$= \sum_{i=1}^{n} y_i x_i' - \frac{\exp(x_i'\beta) x_i'}{1 + \exp(x_i'\beta)} = 0$$

لإيجاد مقدر $\beta$ نستعمل الطرق الرقمية numerical methods .

- **خصائص مقدر المعقولية العظمى**

يتمتع مقدر المعقولية العظمى بالخصائص التالية:

- الاتساق؛
- الفعالية من أجل حجم العينات الكبيرة؛
- يتقارب الى التوزيع الطبيعي، حيث أن: $b \approx \mathcal{N}(\beta; V)$ ، حيث أن: $V$ هي مصفوفة التباين والتباين المشترك، ويتم تقديرها كما يلي:

$$\hat{V} = \left( \sum_{i=1}^{n} \left( \frac{\exp(x_i'b)}{1 + \exp(x_i'b)} \right) \left( \frac{1}{1 + \exp(x_i'b)} \right) x_i x_i' \right)^{-1}$$

- **اختبار معلمة نموذج لوجيت**

إن اختبار معنوية معلم نموذج لوجيت يكون بنفس المنهجية في النماذج الخطية، حيث تكون الفرضيات المراد اختبارها والقيمة المحسوبة كما يلي:

$$H_0: \beta_j = 0 \text{ versus } H_1: \beta_j \neq 0$$

$$z_j = \frac{b_j - 0}{\mathrm{SE}(b_j)} \approx N(0, 1),$$

حيث أن: $SE(b_j)$ الانحراف المعياري للمعلمة $b_j$ .

- **اختبار مجموعة من القيود حول المعلمات**

لاختبار مجموعة من القيود حول المعلمات ستكون المفاضلة أو المقارنة بين:

- نموذج لوجيت بدون قيود على المعلمات وتقدير الشعاع $b_1$ ؛

- نموذج لوجيت مع وجود m قيد وتقدير الشعاع $b_0$ .

فرضية العدم تشير إلى أن m قيد حول المعلمات صحيحة.

من أجل القيام بالاختبار نحن في حاجة إلى:

- $L(b_1)$ : قيمة أعظم احتمال للنموذج الكلي؛
- $L(b_0)$ : قيمة أعظم احتمال للنموذج المقيد؛

$$LR = -2\left[log\left(L(b_0)\right) - log\left(L(b_1)\right)\right] \approx {\chi^2}_m$$

# المحاضرة الثالثة العاشرة:

# مؤشرات المطابقة والتنبؤ في نموذج لوجيت

# المحاضرة الثالثة عشر: مؤشرات المطابقة والتنبؤ في نموذج لوجيت

## – البواقي

إن بواقي نموذج لوجيت يمكن استخراجها كما يلي:

$$\begin{aligned} y_i - E[y_i] &= y_i - (0 \times \Pr[y_i = 0] + 1 \times \Pr[y_i = 1]) \\ &= y_i - \Pr[y_i = 1] \\ &= y_i - \frac{\exp(x_i'b)}{1 + \exp(x_i'b)} \end{aligned}$$

Interesting cases:

- Lower bound: $y_i - E[y_i] \approx -1$
- Upper bound: $y_i - E[y_i] \approx 1$
- Perfect fit $y_i - E[y_i] \approx 0$

## – قياس المطابقة في نموذج لوجيت

نقصد بالمطابقة مدى تطابق السلسلة الحقيقية مع السلسلة المقدرة، ومن أجل القيام بذلك نحتاج لتعريف ما يلي:

- $L(b)$ : قيمة دالة المعقولية العظمى للنموذج المقدر؛
- $L(b_0)$ : قيمة المعقولية العظمى للنموذج الصفري والذي يحتوي على الحد الثابت فقط.

تكون المطابقة تامة إذا كانت $L(b) \approx 1$ أو $log(L(b)) \approx 0$ ن وهنا نميز بين مؤشرين:

- McFadden $R^2$:

$$R^2 = 1 - \frac{\log(L(b))}{\log(L(b_1))}$$

- Nagelkerke $R^2$:

$$R^2 = \frac{1 - \left(\frac{L(b_1)}{L(b)}\right)^{2/n}}{1 - L(b_1)^{2/n}}$$

## – التنبؤ الاحتمالي

إذا توفرت لدينا المشاهدة المستقبلية $x_{n+1}$ ، يمكن إجراء التنبؤ لـ $y_{n+1}$ وفق الشكل التالي:

$$\begin{aligned}\mathrm{E}[y_{n+1}] &= 0 \times \Pr[y_{n+1}=0] + 1 \times \Pr[y_{n+1}=1] \\ &= \Pr[y_{n+1}=1] \\ &= \frac{\exp(x'_{n+1}\beta)}{1+\exp(x'_{n+1}\beta)}\end{aligned}$$

To estimate this probability we replace $\beta$ by its estimate $b$ and obtain $\widehat{\Pr}[y_{n+1}=1]$.

ما يلاحظ في التنبؤ باستخدام نموذج لوجيت أن التنبؤات لن تكون مساوية تماما إلى 1 أو الصفر، بل هي قيم احتمالية، وللتخلص من هذا الاشكال نحول التنبؤات $\hat{y}_{n+1}$ إلى 0 أو 1 وفق القاعدة التالية:

$$\hat{y}_{n+1} = 1 \text{ if } \widehat{\Pr}[y_{n+1}=1] > c$$
$$\hat{y}_{n+1} = 0 \text{ if } \widehat{\Pr}[y_{n+1}=1] \le c.$$

عديد من البرمجيات الخاصة بالقياس الاقتصادي تأخذ قيمة c مساوية إلى 0.5.

- **تقييم التنبؤات**

بافتراض أنه لدينا m عنصر خارج العينة سنقوم بالتنبؤ يه، حيث سنرمز له بالرمز $\hat{y}_i$ . إن التنبؤات الصحيحة والغير صحيحة تكون كما يلي:

| | |
|---|---|
| $m_{11} = \sum_{i=1}^{m} y_{n+i}\hat{y}_{n+i}$ | data=1 & prediction=1 |
| $m_{00} = \sum_{i=1}^{m} (1-y_{n+i})(1-\hat{y}_{n+i})$ | data=0 & prediction=0 |
| $m_{10} = \sum_{i=1}^{m} y_{n+i}(1-\hat{y}_{n+i})$ | data=1 & prediction=0 |
| $m_{01} = \sum_{i=1}^{m} (1-y_{n+i})\hat{y}_{n+i}$ | data=0 & prediction=1 |

إن التنبؤات الصحيحة وغير الصحيحة يمكن تقديمها في الجدول الموالي:

Classify predictions in right and wrong:

| observed | predicted $\hat{y}=0$ | predicted $\hat{y}=1$ | sum |
|---|---|---|---|
| $y=0$ | $m_{00}/m$ | $m_{01}/m$ | $(m_{00}+m_{01})/m$ |
| $y=1$ | $m_{10}/m$ | $m_{11}/m$ | $(m_{10}+m_{11})/m$ |
| sum | $(m_{00}+m_{10})/m$ | $(m_{01}+m_{11})/m$ | 1 |

The fraction $m_{00}/m + m_{11}/m$ is called the hit rate.

إن نسبة التنبؤات الصحيحة هي : $\frac{m_{00}}{m}+\frac{m_{11}}{m}$ ، كما يصطلح على هذه النسبة بمعدل الإصابة hit rate ، أما نسبة التنبؤات الخاطئة فهي: $\frac{m_{01}}{m}+\frac{m_{10}}{m}$ .