جامعة المنارة
MANARA UNIVERSITY

مبادئ الاحتمالات والإحصاءات الحيوية

Probability Principles & Biostatistics

تمهيد :

يوجد عدة تعاريف للاحتمال هي:

التعريف التقليدي (الكلاسيكي) للاحتمال Classical definition of Probability

التعريف التجريبي للاحتمال (الإحصائي) Experimental Definition of Probability

التعريف الرياضي

التعريف المساحي

إننا سنقتصر هنا على استعراض التعريف الإحصائي.

التعريف الإحصائي للاحتمال: يعتمد هذا التعريف على إجراء عدد كبير من التجارب الميدانية أو على جمع عدد كبير من المعلومات الإحصائية ثم البحث فيها عن عدد مرات تحقق الحادث المطلوب.
فإذا رمزنا لعدد التجارب التي تم إجراؤها بالرمز n ، ولعدد مرات تحقق الحادث المطلوب خلال هذه التجارب بالرمز m، فإن التعريف الإحصائي للاحتمال يعطى بالعلاقة:

$$\text{احتمال تحقق الحادث A} = \frac{\text{عدد مرات تحقق الحادث خلال التجارب}}{\text{عدد التجارب المجراة}} = \frac{\text{عدد التكرارات المقابلة}}{\text{مجموع التكرارات}}$$

وتحرياً للدقة نكتب ذلك على الشكل التالي:

$$P(A) = \lim_{n\to\infty}\frac{m}{n} \approx \frac{m}{n}$$

وإن هذا الاحتمال يزداد دقة كلما ازداد عدد التجارب المجراة n.

مثال: لنفترض أننا أجرينا كشفاً على أسنان الأطفال في إحدى المدارس فوجدنا 50 حالة تسوس من أصل 350 طالباً يدرسون في تلك المدرسة. فاحسب احتمال (نسبة) تسوس الأسنان لدى الأطفال.
من التعريف السابق نجد أن احتمال التسوس لدى الأطفال A يساوي:

$$P(A) = \frac{m}{n} = \frac{50}{350} = \frac{1}{7}$$

أي أن واحد من كل سبعة أطفال يعاني من تسوس الأسنان.

ملاحظة: إذا أردنا أن نحصل على نتيجة أدق من ذلك علينا أن نتوسع في عمليات الكشف الطبي ونجريه في جميع المدارس، وعلى جميع الأطفال من نفس العمر.

مثال:

في دراسة للولادات خلال عام ما وفي مدينة معينة، وجدنا أن عدد الولادات بلغ 1500 مولوداً حياً، وإن عدد الذكور فيهم كان 800 ذكراً، وعدد الإناث 700 أنثى، فما هو احتمال ولادة الذكر، وما هو احتمال ولادة الأنثى؟ :

احتمال ولادة الذكر= $\dfrac{\text{عدد مرات التحقق}}{\text{عدد التجارب المدروسة}}$ $\quad P(M)=\dfrac{800}{1500}=\dfrac{8}{15}=$

احتمال ولادة الأنثى= $\dfrac{\text{عدد مرات التحقق}}{\text{عدد التجارب المدروسة}}$ $\quad P(F)=\dfrac{700}{1500}=\dfrac{7}{15}=$

تطبيقات على حساب الاحتمالات :

مثال: لنفترض إننا سألنا 200 طالباً وطالبة عما إذا كانوا يدخنون أم لا . فكانت إجاباتهم المتكررة مبوبة في جدول كما يلي :

النوع / الجواب	ذكر M	أنثى F	المجموع
S مدخن	35	15	50
$\bar{S}$ غير مدخنين	85	65	150
المجموع	120	80	200

وهنا نلاحظ أنه يمكننا الحصول على النسب والاحتمالات المختلفة بتقسيم التكرارات الهامشية والداخلية على المجموع العام (200) , والحصول على الاحتمالات الشرطية بتقسيم التكرارت الداخلية على المجاميع الهامشية (في العمود الأخير أو السطر الأخير) كما يلي :

1- إذا قمنا بتقسيم كل من التكرارات الداخلية والهامشية على المجموع العام (200) فإننا سنحصل على النسب المعروضة في الجدول التالي :

البيان	M	F	Σ
S	35/200	15/200	50/200
$\bar{S}$	85/200	65/200	150/200
Σ	120/200	80/200	1

نسبة المدخنين الكلية (احتمال التدخين) $P(S) = \frac{50}{200}$

نسبة غير المدخنين الكلية (احتمال عدم التدخين) $P(\bar{S}) = \frac{150}{200}$

نسبة الذكور في العينة (احتمال ظهور الذكور) $P(M) = \frac{120}{200}$

نسبة الإناث في العينة (احتمال ظهور الأنثى) $P(F) = \frac{80}{200}$

نسبة المدخنين الذكور (احتمال أن يكون مدخنا وذكراً) $P(S \cap M) = \frac{35}{200}$

نسبة المدخنين الإناث (احتمال أن تكون مدخنة وأنثى) $P(S \cap F) = \frac{15}{200}$

نسبة غير المدخنين الذكور(احتمال أن يكون غير مدخن وذكر) $P(\bar{S} \cap M) = \frac{85}{200}$

نسبة غير المدخنين الإناث (احتمال أن تكون غير مدخنة وأنثى) $P(\bar{S} \cap F) = \frac{65}{200}$

2- وإذا قسمنا كل من التكرارات الداخلية في كل سطر على مجموع ذلك السطر فإننا سنحصل على الاحتمالات الشرطية التالية :

احتمال أن يكون المدخن ذكراً $P(M / S) = \frac{35}{50}$

احتمال أن يكون المدخن أنثى $P(F / S) = \frac{15}{50}$

احتمال أن يكون غير المدخن ذكراً $P(M / \bar{S}) = \frac{85}{150}$

احتمال أن يكون غير المدخن أنثى $P(F / \bar{S}) = \frac{65}{150}$

وفي النهاية سنحصل على جدول كما يلي :

البيان	M	F	Σ
S	$\frac{35}{50}$	$\frac{15}{50}$	1
$\bar{S}$	$\frac{85}{150}$	$\frac{65}{150}$	1

3- إذا قسمنا كل من التكرارات الداخلية في كل عمود على مجموع ذلك العمود فإننا سنحصل على الاحتمالات الشرطية التالية :

احتمال أن يكون الذكر مدخناً $P(S\,/M) = \frac{35}{120}$

احتمال أن يكون الذكر غير مدخن $P(\bar{S}\,/M) = \frac{85}{120}$

احتمال أن تكون الأنثى مدخنة $P(S\,/F) = \frac{15}{80}$

احتمال أن تكون الأنثى غير مدخنة $P(\bar{S}\,/F) = \frac{65}{80}$

وفي النهاية سنحصل على جدول كالتالي :

البيان	M	F
S	$\frac{35}{120}$	$\frac{15}{80}$
$\bar{S}$	$\frac{85}{120}$	$\frac{65}{80}$
Σ	1	1

وهكذا نجد أن احتمالات التقاطع السابقة تساوي :

$$P(S \cap M) = P(S).P(M\,/S) = \frac{50}{200}.\frac{35}{50} = \frac{35}{200}$$

$$P(\bar{S} \cap M) = P(\bar{S}).P(M\,/\bar{S}) = \frac{150}{200}.\frac{85}{150} = \frac{85}{200}$$

$$P(S \cap F) = P(S).P(F\,/S) = \frac{50}{200}.\frac{15}{50} = \frac{15}{200}$$

$$P(\bar{S} \cap F) = P(\bar{S}).P(F/\bar{S}) = \frac{150}{200}.\frac{65}{150} = \frac{65}{200}$$

ولحساب الاحتمالات الشرطية اللاحقة نفترض أولاً أنه بعد إجراء الدراسة وجدنا أن إحدى الاستمارات أو الإجابات لم تتضمن جواباً على سؤال التدخين وبعد التدقيق والمراجعة وجدنا أن صاحبها كان ذكراً فكيف يمكننا حساب احتمال أن يكون هذا الشخص مدخناً ؟

نلاحظ أن المطلوب هو حساب الاحتمال الشرطي اللاحق P(S/M) واعتماداً على قاعدة (بايز) نجد أن :

$$P(S/M) = \frac{P(S).P(M/S)}{P(S).P(M/S) + P(\bar{S}).P(M/\bar{S})}$$

$$= \frac{\frac{50}{200}.\frac{35}{50}}{\frac{50}{200}.\frac{35}{50} + \frac{150}{200}.\frac{85}{150}} = \frac{\frac{35}{200}}{\frac{35}{200} + \frac{85}{200}} = \frac{35}{120}$$

ومن جهة أخرى نفترض أيضاً أنه خلال التدقيق وجدنا استمارة أخرى غير واضحة وتبين أن صاحبها كان مدخناً ولكنه لم يذكر جنسه ، فما هو احتمال أن يكون صاحب هذه الاستمارة هذا المدخن ذكراً؟ . أي أن الاحتمال المطلوب هو P(M/S)

لحساب هذا الاحتمال نعتمد على قاعدة (بايز) فنجد أن :

$$P(M/S) = \frac{P(M).P(S/M)}{P(M).P(S/M) + P(\bar{M}).P(S/\bar{M})}$$

$$= \frac{P(M).P(S/M)}{P(M)P(S/M) + P(F).P(S/F)}$$

$$= \frac{\frac{120}{200}.\frac{35}{120}}{\frac{120}{200}.\frac{35}{120} + \frac{80}{200}.\frac{15}{80}} = \frac{\frac{35}{200}.}{\frac{35}{200} + \frac{15}{200}} = \frac{35}{50}$$

مؤشرات المسح المخبري :

لتوضيح فكرة هذه المؤشرات نشير إلى أن عمليات المسح المخبري هي اختبارات لاكتشاف الأمراض في بداية المعالجة . وإن نتائج هذه الاختبارات التشخيصية قد تكون ايجابية أو سلبية ولكن حالة المريض اللاحقة هي التي تؤكد صحة هذه التحاليل أو تدحضها .

لذلك نفترض أنه أجرينا عدداً من الاختبارات التشخيصية على (10000) شخص من المعرضين للإصابة بمرض معين .وكانت نتائج هذه الاختبارات مبوبة كما في الجدول التالي :

حالة المريض / نتيجة الاختبار	مصاب بالمرض D+	غير مصاب بالمرض D-	المجموع
ايجابية T +	a=95	b=495	a+b=590
سلبية T –	c=5	d=9405	c+d=9410
المجموع	a+c=100	b+d=9900	a+b+c+d=10000

وبناء على معلومات هذا الجدول نعرف مؤشرات المسح الصحي التالية :

1- الحساسية : وهي نسبة الأشخاص المرضى الذين يظهرهم الاختبار أنهم مصابين بالمرض . ونرمز له لها بـ $P(T^+/D^+)$ وهي تساوي :

$$P(T^+/D^+) = \frac{a}{a+c} = \frac{95}{100} = 0.95$$

2- النوعية : وهي نسبة الأشخاص غير المرضى الذين يظهرهم الاختبار أنهم غير مصابين بالمرض . ونرمز لها بـ $P(T^-/D^-)$ وهي تساوي :

$$P(T^-/D^-) = \frac{d}{b+d} = \frac{9405}{9900} = 0.95$$

3- نسبة النتائج الايجابية الكاذبة : وهي نسبة الأشخاص غير المرضى الذين يظهرهم الاختبار بأنهم مصابين بالمرض ونرمز لها بـ $P(T^+/D^-)$ وهي تساوي :

$$P(T^+/D^-) = \frac{b}{b+d} = \frac{495}{9900} = 0.05$$

4- نسبة النتائج السلبية الكاذبة : وهي نسبة الأشخاص المرضى الذين يظهرهم الاختبار بأنهم غير مصابين بالمرض ، ونرمز لها بـ $P(T^-/D^+)$ وهي تساوي :

$$P(T^-/D^+) = \frac{c}{a+c} = \frac{5}{100} = 0.05$$

5- القيمة التنبؤية للنتائج الايجابية : وهي نسبة الأشخاص المرضى ذوي النتائج الإيجابية، ونرمز لها بـ $P(D^+/T^+)$ وهي تساوي :

$$P(D^+/T^+) = \frac{a}{a+b} = \frac{95}{590} = 0.161$$

6- القيمة التنبؤية للنتائج السلبية : وهي نسبة الأشخاص غير المرضى ذوي النتائج السلبية، ونرمز لها بـ $P(T^-/D^-)$ وهي تساوي :

$$P(D^-/T^-) = \frac{d}{c+d} = \frac{9405}{9410} = 0.9999$$

7- مصداقية الاختبار : وهي نسبة كل الاختبارات الصحيحة (الايجابية والسلبية) إلى مجموع الاختبارات ، ونرمز لها بـ AT وهي تساوي :

$$AT = \frac{a+d}{a+b+c+d} = \frac{95+9405}{10000} = 0.95$$

8- نسبة الإمكانية العظمى للنتائج الايجابية : وهي نسبة احتمال النتائج الايجابية للأشخاص المرضى على احتمال النتائج الايجابية للأشخاص غير المرضى ونرمز لها بـ LR^+ وهي تساوي :

$$LR^+ = \frac{P(T^+/D^+)}{P(T^+/D^-)} = \frac{0.95}{0.05} = 19$$

ومن الواضح أن LR^+ تساوي :

$$LR^+ = \frac{\text{الحساسية}}{\text{نسبة النتائج الايجابية الكاذبة}} = \frac{\text{الحساسية}}{\text{1- الحساسية}}$$

9- نسبة الإمكانية العظمى للنتائج السلبية : وهي نسبة احتمال النتائج السلبية للأشخاص المرضى على النتائج السلبية للأشخاص غير المرضى ونرمز لها بـ LR^- وهي تساوي:

$$LR^- = \frac{P(T^-/D^+)}{P(T^-/D^-)} = \frac{\text{1- النوعية}}{\text{النوعية}}$$

$$LR^- = \frac{0.05}{0.95} = 0.053$$

10- من الواضح أن هذه المؤشرات مرتبطة ببعضها البعض بالعلاقات التالية :

النوعية + نسبة النتائج الايجابية الكاذبة = 1

$$\frac{d}{b+d} + \frac{b}{b+d} = 1$$

وهذا يعني أن إذا ازدادت النوعية تؤدي إلى انخفاض نسبة النتائج الايجابية الكاذبة .

11- الحساسية + نسبة النتائج السلبية الكاذبة = 1

$$\frac{a}{a+c} + \frac{c}{a+c} = 1$$

أي أنه كلما ازدادت الحساسية أدت إلى انخفاض نسبة النتائج السلبية الكاذبة .