الفصل الاول

بعض التوزيعات الاحتمالية

Some Probability Distributions

1.1 تمهيد:

نظرا لاعتماد الاحصاء الحيوي على بعض التوزيعات الاحتمالية المتقطعة والمستمرة، لذلك سوف نتناول في هذا الفصل وبشكل مختصرالتوزيعات التي ترتبط بالاحصاء الحيوي واختبار الفرضيات.

2.1 توزيع ثنائي الحدين: Binomial Distribution

توزيع ثنائي الحدين يصف الظواهر او التجارب التي تتضمن حالتين ممكنتين وهي حالة النجاح وحالة الفشل اذا ماتم تكرار التجربة بعدد معين من المحاولات المستقلة. وبافتراض ان تجربة تكرر لـ (n) من المحاولات المستقلة، باحتمال نجاح (p) لكل محاولة وباحتمال فشل $(q = 1 - p)$. فاذا كان المتغير العشوائي (X) يمثل عدد حالات النجاح عندها فان (X) يتبع توزيع ثنائي الحدين، اي ان $X \sim bin(n, p)$، وان دالة الكتلة الاحتمالية $(p.m.f)$ $(\text{Probability Mass Function})$ والتي تعرف رياضيا كما ياتي:

$$pr(X = x) = C_x^n\, p^x\, q^{(n-x)} \qquad , \qquad x = 0,1,2,\dots,n$$

ومن اهم خصائص توزيع ثنائي الحدين:

1. الوسط الحسابي (المتوسط) $\mu_x = E(x) = np$
2. التباين $\sigma_x^2 = var(x) = npq$

مثال 1.1: بافتراض ان $x \sim bin(5, 0.4)$، اجب عن كل مما ياتي:

1. اكتب دالة الكتلة الاحتمالية لـ x.
2. احسب المتوسط والتباين والانحراف المعياري لـ x.
3. احسب الاحتمالات الاتية $pr(x = 2), pr(x \le 2), pr(x > 3), pr(2 \le x < 5)$.

الحل:

1. $pr(X = x) = C_x^5\ (0.4)^x\ (0.6)^{(5-x)} \qquad , x = 0,1,2,3,4,5$

2. $\mu_x = np = (5)(0.4) = 2$

$$\sigma_x^2 = npq = (5)(0.4)(0.6) = 1.2$$

$$\sigma_x = \sqrt{npq} = \sqrt{(5)(0.4)(0.6)} = \sqrt{1.2} = 1.09$$

3. $pr(x = 2) = C_2^5\ (0.4)^2\ (0.6)^{(5-2)} = \dfrac{5!}{2!\ \ (5-2)!}\ (0.4)^2\ (0.6)^3$

$$pr(x = 2) = (10)(0.16)(0.216) = 0.3456$$

$$pr(x \leq 2) = pr(x = 2) + pr(x = 1) + pr(x = 0)$$

$$pr(x \leq 2) = C_2^5\ (0.4)^2\ (0.6)^3 + C_1^5 (0.4)^1\ (0.6)^4 + C_0^5 (0.4)^0\ (0.6)^5$$

$$pr(x > 3) = pr(x = 4) + pr(x = 5)$$

$$pr(x > 3) = C_4^5\ (0.4)^4\ (0.6)^1 + C_5^5 (0.4)^5\ (0.6)^0$$

$$pr(2 \leq x < 5) = pr(x = 2) + pr(x = 3) + pr(x = 4)$$

$$pr(2 \leq x < 5) = C_2^5\ (0.4)^2\ (0.6)^3 + C_3^5 (0.4)^3\ (0.6)^2 + C_1^5 (0.4)^4\ (0.6)^1$$

مثال 2.1: تم اعطاء علاج لستة اشخاص يعانون مرض فقر الدم، وتبين ان احتمال شفاء المريض (0.8). اجب عن كل مما ياتي:

1. متوسط الشفاء.
2. تباين الشفاء.
3. احتمال شفاء ثلاثة مرضى، ماهو تفسيرك.

4. احتمال شفاء على الاقل ثلاثة مرضى، فسر النتائج. (يترك الحل للطالب).
5. احتمال شفاء على الاكثر شخصين، فسر النتائج. (يترك الحل للطالب).

الحل:

المتغير العشوائي X يمثل عدد المرضى الذين تم شفائهم من مرض فقر الدم، وان احتمال شفاء المريض هو (0.8) وان احتمال عدم شفاء المريض هو (0.2)، اي ان $X \sim bin(6,0.8)$، وان:

$$pr(X = x) = C_x^6\ (0.8)^x\ (0.2)^{(6-x)} \qquad , x = 0,1,2,3,4,5,6$$

1. $\mu_x = np = (6)(0.8) = 1.6$

2. $\sigma_x^2 = npq = (6)(0.8)(0.2) = 0.96$

3. $pr(x = 3) = C_3^6\ (0.8)^3\ (0.2)^{(6-3)} = \dfrac{6!}{3!\ \ 3!}\ (0.8)^3\ (0.2)^3$

$$pr(x = 3) = (20)(0.512)(0.008) = 0.082$$

من خلال قيمة الاحتمال نستنتج ان احتمال شفاء 3 مرضى من فقر الدم بعد تلقي العلاج هي (0.08)، وهذا يعني ان احتمال شفاء نصف المرضى هو (0.08)، اي ان العلاج غير فعال بالمستوى المطلوب.

3.1 توزيع بواسون: Poisson Distribution

احد التوزيعات المتقطعة الواسعة الاستخدام في العديد من التطبيقات، يطلق عليه توزيع الحوادث النادرة كعدد كريات الدم التالفة.

المتغير العشوائي (X) الذي يتبع توزيع بواسون، اي ان $X \sim Po(\theta)$، له دالة الكتلة الاحتمالية $(p.m.f)$ تعرف رياضيا كما ياتي:

$$pr(X = x) = \frac{\theta^x e^{-\theta}}{x!} \quad , \quad x = 0,1,2, \dots \dots$$

ومن اهم خصائص توزيع بواسون:

1. الوسط الحسابي (المتوسط) $\mu_x = E(x) = \theta$
2. التباين $\sigma_x^2 = var(x) = \theta$

مثال 3.1: بافتراض ان $x \sim po(0.04)$، اجب عن كل مما ياتي:

1. اكتب دالة الكتلة الاحتمالية لـ x.
2. احسب المتوسط والتباين والانحراف المعياري لـ x.
3. احسب الاحتمالات الاتية $pr(x = 1), pr(x \leq 1), pr(x > 0), pr(1 < x \leq 3)$.

الحل:

1. $pr(X = x) = \dfrac{(0.04)^x e^{-0.04}}{x!} \quad , \quad x = 0,1,2, \dots \dots$

2. $\mu_x = \theta = 0.04$

   $\sigma_x^2 = \theta = 0.04$

   $\sigma_x = \sqrt{\theta} = \sqrt{0.04} = 0.2$

$$3.\quad pr(x=1)=\frac{(0.04)^1e^{-0.04}}{1!}=(0.04)(0.961)=\ 0.038$$

$$pr(x\le 1)=pr(x=1)+pr(x=0)$$

$$pr(x\le 1)=\frac{(0.04)^1e^{-0.04}}{1!}+\frac{(0.04)^0e^{-0.04}}{0!}=0.038+0.961=0.999$$

$$pr(x>0)=1-pr(x\le 0)=1-pr(x=0)$$

$$pr(x>0)=1-\frac{(0.04)^0e^{-0.04}}{0!}=1-0.961=0.039$$

$$pr(1<x\le 3)=pr(x=2)+pr(x=3)$$

$$pr(2\le x<5)=\frac{(0.04)^2e^{-0.04}}{2!}+\frac{(0.04)^3e^{-0.04}}{3!}$$

$$pr(2\le x<5)=0.0008+0.00001=0.00081$$

مثال 4.1: بافتراض ان معدل الوفيات نتيجة الاصابة بمرض ما هو 10 اشخاص خلال 5 سنوات، فاذا كان المتغير العشوائي x يمثل عدد الوفيات، اجب عن كل مما ياتي:

1. ماهو التوزيع الاحتمالي للمتغير x، اكتب دالة الكتلة الاحتمالية لـ x.
2. احسب المتوسط والتباين والانحراف المعياري لـ x.
3. احسب احتمال وفاة سبعة اشخاص، وهل تعتقد ان المرض خطير.

الحل:

1. المتغير العشوائي x يمثل عدد الوفيات خلال 5 سنوات، بمعدل 10 اشخاص، اي ان معدل الوفيات نادر، وهذا يعني ان:

$$x\sim po(10)$$

$$pr(X = x) = \frac{(10)^x e^{-10}}{x!} \qquad , \qquad x = 0,1,2,\dots,10$$

2. $\mu_x = \theta = 10$

$\sigma_x^2 = \theta = 10$

$\sigma_x = \sqrt{\theta} = \sqrt{10} = 3.16$

3. $pr(x = 7) = \frac{(10)^7 e^{-10}}{7!} = \frac{(10000000)(0.000045)}{5040} = 0.089$

ان احتمال وفاة سبعة اشخاص بهذا المرض خلال مدة 5 سنوات هو 0.089، وهذا يشير الى ان الى عدم خطورة المرض.

3. التوزيع الطبيعي: Normal Distribution

يعد التوزيع الطبيعي من اهم التوزيعات الاحتمالية المستمرة الشائعة الاستخدام في النظرية الاحصائية وخصوصا في مجال التحليل الاحصائي، وذلك لكثرة استخدامه، وان غالبية الظواهر تتبع في سلوكها هذا التوزيع ، فضلا عن كون اغلب التوزيعات الاحتمالية الاخرى تؤول اليه تحت شروط معينة.

يقال ان المتغير العشوائي X يتبع التوزيع الطبيعي بمتوسط μ وبتباين σ^2 ، $X \sim N(\mu, \sigma^2)$، وله دالة الكثافة الاحتمالية الاتية:

$$f(x) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad , -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

ومن اهم خصائص التوزيع الطبيعي:

1. الوسط الحسابي (المتوسط) $\mu_x = E(x) = \mu$
2. التباين $\sigma_x^2 = var(x) = \sigma^2$

3. يمكن تحويل المتغير العشوائي X الذي يتبع التوزيع الطبيعي بمتوسط وبتباين الى الى المتغير العشوائي Z الذي يتبع التوزيع الطبيعي القياسي (Standard Normal)، اي ان $Z \sim N(0,1)$ ، اذ ان:

$$X \sim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

بحيث ان دالة الكثافة الاحتمالية للمتغير Z هي:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad , -\infty < z < \infty$$

ان الهدف من تحويل المتغير العشوائي الذي يتبع التوزيع الطبيعي الى متغير عشوائي يتبع التوزيع الطبيعي القياسي هو صعوبة التعامل الرياضي مع دالة الكثافة الاحتمالية للتوزيع الطبيعي، لذلك يتم اللجوء الى التوزيع الطبيعي القياسي بمساعدة جداول التوزيع المعدة لهذا الغرض.

مثال 5.1: اذا كان $X \sim N(6,4)$ احسب الاحتمالات الاتية:

$$pr(x < 2), \; pr(x > 3), \; pr(1 < x < 4)$$

الحل:

$$pr(x < 2) = pr\left(\frac{x-\mu}{\sigma} < \frac{2-\mu}{\sigma}\right) = pr\left(Z < \frac{2-6}{2}\right)$$

$$= pr(Z < -2) = \Phi(-2) = 0.0228$$

$$pr(x > 3) = 1 - pr(x < 3) = 1 - pr\left(Z < \frac{3-6}{2}\right) = \Phi(-1.5) = 0.0668$$

$$pr(1 < x < 4) = pr(x < 4) - pr(x < 1)$$

$$= pr\left(Z < \frac{4-6}{2}\right) - pr\left(Z < \frac{1-6}{2}\right) = \Phi(-1) - \Phi(-2.5)$$

$$= 0.1587 - 0.0062 = 0.1525$$

مثال 6.1: قامت احدى شركات انتاج الادوية بتطوير علاج لتقليل مستوى اليوريا في الدم، ومن خلال الترويج لهذا العلاج ادعت الشركة ان مستوى اليوريا في الدم ينخفض الى 38 مل غرام / 100 مل لتر. وتم فحص العلاج الجديد من خلال اعطائه الى 100 شخص وتبين ان متوسط مستوى اليوريا لديهم بلغ 45 مل غرام / 100 مل لتر وبتباين بلغ 49 مل غرام / 100 مل لتر. هل تعتقد ان تطوير العلاج كان مجديا؟ علما ان الحد الاعلى للمستوى الطبيعي لليوريا في الدم هو 40 مل غرام / 100 مل لتر.

الحل:

$$\mathrm{pr}(x < 38) = \mathrm{pr}\left(\frac{x-\mu}{\sigma} < \frac{38-\mu}{\sigma}\right) = \mathrm{pr}\left(Z < \frac{38-45}{7}\right)$$

$$= \mathrm{pr}(Z < -1) = \Phi(-1) = 0.1587$$

التحليل الحيوي: ان احتمال انخفاض مستوى اليوريا في الدم بعد تلقي العلاج الجديد بلغ 15.97%، وهذا احتمال ضعيف، لذلك يمكن القول ان العلاج الجديد غير مجد.