Probabilidades e Estatística

Primeiro teste — 2º semestre – 2009/2010
Duração: 90 minutos — 08/05/2010 – 9:00 horas

Grupo I — 5 valores

1. Numa selecção de cobaias para a realização de uma experiência em Genética, retiraram-se ao acaso sem reposição quatro ratinhos de uma ninhada de tamanho N contendo dois ratinhos brancos.

(a) Mostre que a probabilidade de o primeiro ratinho seleccionado ser branco é igual à probabilidade de o segundo também o ser. (1.0)

Seja B_i="o iº ratinho seleccionado é branco". Pela regra de Laplace tem-se que $P(B_1) = 2/N$. $P(B_2) = P(B_2 \mid B_1)P(B_1) + P(B_2 \mid \bar{B}_1)P(\bar{B}_1) = \frac{1}{N-1}\frac{2}{N} + \frac{2}{N-1}\frac{N-2}{N} = \frac{2}{N}$, de acordo com o teorema da probabilidade total. A regra de Laplace aplica-se na tiragem de cada ratinho pois trata-se de uma experiência aleatória com um número finito de resultados equiprováveis.

(b) Admitindo que a probabilidade de ambos os ratinhos brancos serem escolhidos é dupla da de nenhum ser escolhido, determine qual o tamanho da ninhada. (1.5)

Sejam A ="os 2 ratinhos brancos são seleccionados" e B ="nenhum ratinho branco é seleccionado". Tem-se $P(A) = \frac{\binom{2}{2}\binom{N-2}{2}}{\binom{N}{4}}$ e $P(B) = \frac{\binom{2}{0}\binom{N-2}{4}}{\binom{N}{4}}$, pela regra de Laplace uma vez que há um número finito de subconjuntos de tamanho 4 que podem ser seleccionados e estes são equiprováveis.
$P(A) = 2P(B) \Leftrightarrow \binom{N-2}{2} = 2\binom{N-2}{4} \Leftrightarrow \frac{(N-4)!}{3!(N-6)!} = 1 \Leftrightarrow (N-4)(N-5) = 6 \Leftrightarrow N^2 - 9N + 14 = 0$
$\Leftrightarrow N = 7 \vee N = 2$. Como N deve ser maior do que 4 a solução é $N = 7$.
Alternativa: seja X ="nº de ratinhos brancos na amostra de 4 ratinhos" com $X \sim H(N, 4, 2)$.
$f_X(2) = 2f_X(0) \wedge N > 4 \Rightarrow N = 7$.

2. Um certo tipo de defeitos ocorre em paineis de plástico usados no interior de automóveis segundo um processo de Poisson com um desvio padrão de 0.2 defeitos por 0.1 m^2 de painel. Admitindo que no interior dos referidos automóveis é empregue 1 m^2 de painel:

(a) Determine a probabilidade de que quando muito dois de um conjunto aleatoriamente escolhido de dez automóveis apresentem defeitos no respectivo painel. (1.5)

Seja X_t ="nº de defeitos em t m^2 de painel". Tem-se que $X_{0.1} \sim Poi(\lambda)$ e $\sqrt{Var[X_{0.1}]} = \sqrt{\lambda} = 0.2$. Ou seja, $X_{0.1} \sim Poi(0.04)$ e, consequentemente, $X_1 \sim Poi(0.4)$.
Seja Y ="nº de automóveis com defeitos no painel de entre 10 seleccionados ao acaso". Admitindo independência entre os números de defeitos em paineis de automóveis distintos, tem-se que $Y \sim Bi(10, p)$ em que $p = P(X_1 > 0) = 1 - P(X_1 = 0) \approx 1 - 0.6703 = 0.3297$, uma vez que a variável representa o número de sucessos em 10 repetições independentes de uma prova de Bernoulli.
Então, $P(Y \leq 2) = F_Y(2) \approx 0.3077$.

(b) Calcule a probabilidade de ser necessário inspeccionar no máximo cinco automóveis para encontrar um evidenciando defeitos no painel do seu interior. (1.0)

Seja Z ="nº de automóveis inspeccionados até se encontrar um com defeitos no painel". Sob a independência admitida na alínea anterior temos que $Z \sim Geo(0.3297)$, uma vez que a variável representa o número de repetições independentes de uma prova de Bernoulli até se obter o primeiro sucesso.
Então, $P(Z \leq 5) = F_Z(5) \approx 0.8647$.
Alternativa: $F_Z(5) = 1 - (1 - 0.3297)^5 \approx 0.8647$

Grupo II — 5 valores

1. Seja X a variável aleatória que representa o nível de colesterol (em mg) em 100 ml de sangue de um indivíduo escolhido ao acaso de uma população. Admita que X tem distribuição Normal de valor esperado 200 mg e variância 400 mg^2 e que o nível de colesterol é considerado "normal" se estiver entre 185 mg e 230 mg.

(a) Calcule a proporção de indivíduos da população com excesso de colesterol de entre os que têm um nível de colesterol considerado "anormal". (0.5)

Tem-se que $X \sim N(200, 400)$. $P(X > 230 \mid X < 185 \vee X > 230) = \frac{P(X>230)}{1-P(185 \leq X \leq 230)} = \frac{1-F_X(230)}{1-(F_X(230)-F_X(185))} \approx \frac{1-0.9332}{1-(0.9332-0.2266)} = 0.2277$.
Alternativa: $X \sim N(200, 400) \Leftrightarrow Y = \frac{X-200}{20} \sim N(0, 1)$. Então, $F_X(185) = F_Y(-3/4) = 1 - F_Y(3/4) \approx 1 - 0.7734 = 0.2266$ e $F_X(230) = F_Y(3/2) \approx 0.9332$.

(b) Devido a restrições orçamentais, o sistema de saúde só comparticipa os tratamentos dos 10% de indívíduos com maior nível de colesterol. Determine qual o nível mínimo de colesterol no sangue a partir do qual os tratamentos são comparticipados e comente a política de comparticipação de tratamentos face ao problema de excesso de colesterol na população em causa. (1.0)

Seja k o nível mínimo de colesterol a partir do qual os tratamentos são comparticipados. Deve-se ter $P(X > k) = 0.1 \Leftrightarrow F_X(k) = 0.9 \Leftrightarrow k = F_X^{-1}(0.9) \approx 225.63$ mg/ 100 ml.
Alternativa: $P(X > k) = 0.1 \Leftrightarrow P(Y > (k - 200)/20) = 0.1 \Leftrightarrow \Phi((k - 200)/20) = 0.9 \Leftrightarrow k = 20\Phi^{-1}(0.9) + 200 \approx 20 \times 1.2816 + 200 = 225.63$ mg/100 ml.
A política de comparticipação de tratamentos é adequada pois permite atender a todos os indívíduos com excesso de colesterol. De outra forma, $P(X > 230) \approx 0.0668$ que é inferior a 10%.

(c) Calcule a probabilidade de a média do nível de colesterol de 100 indivíduos, escolhidos ao acaso da população, estar abaixo de 195 mg. (1.0)

Sejam X_i="nível de colesterol do iº indivíduo" e $\bar{X} = \frac{\sum_{i=1}^{100} X_i}{100}$. Como $\bar{X}$ é uma combinação linear de variáveis aleatórias com distribuições Normais que se admitem independentes (dado a escolha ao acaso dos indivíduos) então $\bar{X}$ tem também distribuição Normal com $E[\bar{X}] = E[X] = 200$ e $Var[\bar{X}] = \frac{Var[X]}{100} = 4$, ou seja, $\bar{X} \sim N(200, 4)$.
Então, $P(\bar{X} < 195) = F_X(195) \approx 0.0062$.
Alternativa: $P(\bar{X} < 195) = P((\bar{X} - 200)/2 < -5/2) = \Phi(-5/2) = 1 - \Phi(5/2) \approx 1 - 0.9938 = 0.0062$.

2. Uma empresa desenvolve sites na Internet para os seus clientes. Para um projecto seleccionado ao acaso sejam X e Y as variáveis aleatórias que representam, respectivamente, o conteúdo gráfico e a complexidade de programação, ambas representando classificações nos níveis baixo (1), moderado (2) ou elevado (3). A função de probabilidade conjunta de X e Y encontra-se resumida no quadro ao lado.

$X \backslash Y$	1	2	3
1	1/6	1/12	1/12
2	1/12	1/6	1/12
3	a	1/12	b

(a) Sabendo que 1/3 dos sites requerem uma baixa complexidade de programação, determine os valores de a e b. (1.0)

Temos que $P(Y = 1) = \sum_{x=1}^{3} f_{X,Y}(x, 1) = 1/6 + 1/12 + a = 1/3 \Rightarrow a = 1/12$. Adicionalmente, deve ter-se $\sum_{x=1}^{3} \sum_{y=1}^{3} f_{X,Y}(x, y) = 1 \Rightarrow b = 1/6$.

(b) Calcule a variância do nível de complexidade de programação dos sites, assim como a variância da mesma variável em sites com um baixo conteúdo gráfico. Comente os resultados. (1.5)

$f_Y(y) = \sum_{x=1}^{3} f_{X,Y}(x,y) = \begin{cases} 1/3, & y = 1,2,3 \\ 0, & \text{caso contrário} \end{cases}.$

Temos $E[Y] = \sum_{y=1}^{3} y f_Y(y) = 2$, $E[Y^2] = \sum_{y=1}^{3} y^2 f_Y(y) = 14/3$ e $Var[Y] = E[Y^2] - (E[Y])^2 = 2/3$.

$f_{Y|X=1}(y) = \frac{f_{X,Y}(1,y)}{f_X(1)} = \begin{cases} 1/2, & y = 1 \\ 1/4, & y = 2,3 \\ 0, & \text{caso contrário} \end{cases}.$

Temos $E[Y \mid X = 1] = \sum_{y=1}^{3} y f_{Y|X=1}(y) = 7/4$, $E[Y^2 \mid X = 1] = \sum_{y=1}^{3} y^2 f_{Y|X=1}(y) = 15/4$ e $Var[Y \mid X = 1] = E[Y^2 \mid X = 1] - (E[Y \mid X = 1])^2 = 11/16$.

Uma vez que $Var[Y] \neq Var[Y \mid X = 1]$, podemos afirmar que X e Y são variáveis dependentes.