3.3.7. Outros modelos: hipergeométrico, binomial negativo, multinomial

---

Os modelos mais importantes de VAD são, sem dúvida, o binomial e o de Poisson. Nas próximas seções, veremos três outros modelos – hipergeométrico, binomial negativo e multinomial − que são em geral menos úteis, pois se aplicam apenas a algumas classes particulares de problemas.

3.3.7.1. Modelo Hipergeométrico

(i) Exemplo

O modelo hipergeométrico é usado em problemas onde nos interessa contar o número de sucessos em sorteios feitos *sem* reposição. Suponha por exemplo este problema:

> Uma professora tem uma turma com 20 estudantes, dos quais 12 são meninas e 8 são meninos. Se ela decide sortear cinco livros entre estes estudantes, qual é a probabilidade de que três dos estudantes sorteados sejam meninas?

Se o sorteio for *com reposição*, podemos usar um modelo binomial, com parâmetros

$n = 5$

$p = 12/20 = 0,6$

$q = 1 - p = 0,4$

$$P(X = k) = C_n^k p^k q^{(n-k)}$$

$$P(X = 3) = C_5^3 0,6^3 0,4^{5-3} = 0,3456$$

Se o sorteio é feito *sem reposição*, contudo, o modelo binomial não pode mais ser usado, porque a probabilidade de o estudante sorteado ser uma menina não é mais constante. Podemos resolver então o problema usando técnicas de análise combinatória; calculamos primeiro o número de resultados *favoráveis* no sorteio (isto é, os que contém 3 meninas e 2 meninos), e o dividimos pelo número *total* de resultados possíveis.

O número de resultados *favoráveis* pode ser calculado como:

- três meninas serão sorteadas, entre as 12 existentes na turma. O número de resultados possíveis deste sorteio será dado pela combinação das 12 meninas, 3 a 3:

$$C_n^x = \frac{n!}{x!(n-x)!}$$

$$C_{12}^3 = \frac{12!}{3!(12-3)!} = \frac{12!}{3!9!} = \frac{12 \times 11 \times 10}{3 \times 2 \times 1} = 220 \qquad \text{(i)}$$

- dois meninos serão sorteados, entre os 8 existentes na turma. O número de resultados possíveis deste sorteio será dado pela combinação dos 8 meninos, 2 a 2:

$$C_8^2 = \frac{8!}{2!(8-2)!} = \frac{8!}{2!\,6!} = \frac{8\times 7}{2\times 1} = 28 \qquad \text{(ii)}$$

- multiplicando estes dois resultados em (i) e (ii), obtemos o número de resultados que nos interessam (sorteios contendo 3 meninas e 2 meninos):

$$C_{12}^3 \times C_8^2 = 220\times 28 = 6160 \qquad \text{(iii)}$$

O número *total* de resultados possíveis pode ser calculado também por meio de uma combinação. Se 5 estudantes são sorteados sem reposição de uma turma com 20 alunos, o número total de resultados possíveis deste sorteio será dado pela combinação dos 20 estudantes, 5 a 5:

$$C_{20}^5 = \frac{20!}{5!\,(20-5)!} = \frac{20!}{5!\,15!} = \frac{20\times 19\times 18\times 17\times 16\times 15!}{(5\times 4\times 3\times 2\times 1)\times 15!} = 15504 \qquad \text{(iv)}$$

A razão entre o número resultados favoráveis em (iii), e o número total de resultados possíveis do sorteio em (iv), nos dá a probabilidade desejada:

$$P(3\ meninas) = \frac{C_{12}^3 C_8^2}{C_{20}^5} = \frac{6160}{15504} = 0{,}3973$$

(ii) Função de probabilidades, valor esperado e variância

Podemos generalizar este problema: se uma população contém N elementos, dos quais ***s*** são considerados *sucessos*, e dela retiramos aleatoriamente uma amostra de n elementos, sem reposição, o número X de sucessos nesta amostra será uma variável de distribuição *hipergeométrica* que tem função de probabilidades dada por:

$$P(X=k) = p(k) = \frac{C_s^k C_{N-s}^{n-k}}{C_N^n}$$ onde $n \le s$, e $n \le N-s$.

O valor esperado e variância de X são dados por:

$$E(X) = \frac{ns}{N}$$

$$Var(X) = \frac{ns(N-s)(N-n)}{N^2(N-1)}$$

No problema acima, fazendo: N = 20 s = 12 n = 5
obtemos:

$$P(X=3) = p(3) = \frac{C_{12}^3 C_8^2}{C_{20}^5} = 0{,}3973$$

(iii) Relação entre a distribuição binomial e a hipergeométrica

À medida que aumenta o tamanho N da população, em relação ao tamanho da amostra, o modelo hipergeométrico se aproxima de uma binomial de parâmetro $p = s/N$. Isto é fácil de entender, intuitivamente: se a população for muito grande em relação à amostra, não faz muita diferença se o sorteio é feito *com* ou *sem* reposição.

Por exemplo, a Tabela 1 mostra as probabilidades calculadas, no problema acima, pelos modelos binomial e hipergeométrico, para populações de N=20 (como no enunciado original), N=50 e N=200, mantendo a proporção de meninas sempre igual a 60%. É possível notar que o modelo hipergeométrico se aproxima do binomial à medida que aumenta N, para um mesmo tamanho de amostra n.

Tabela 1. Comparação entre as distribuições hipergeométrica e binomial (p=0.6)

X	hipergeométrica			binomial
	N=20	N=50	N=200	
0	0.004	0.007	0.009	0.010
1	0.054	0.069	0.075	0.077
2	0.238	0.234	0.231	0.230
3	0.397	0.364	0.350	0.346
4	0.255	0.259	0.259	0.259
5	0.051	0.067	0.075	0.078

3.3.7.2. Modelo Binomial Negativo (ou Modelo de Pascal)

Este modelo é uma generalização da modelo geométrico: ao invés de fazermos tentativas binárias (que podem resultar em *sucesso* ou *fracasso*) até que ocorra o *primeiro* sucesso, repetimos as tentativas até que o sucesso ocorra r vezes. A variável de interesse é X: número de repetições feitas até ocorrer o r-ésimo sucesso.

(i) Função de probabilidade

A função de probabilidade deste modelo é dada por:

$$p(k) = P(X = k) = C_{k-1}^{r-1}\, p^r q^{k-r}$$

onde:

r	número de sucessos desejados
k	número de tentativas
p	probabilidade de sucesso em cada tentativa
$q = 1 - p$	probabilidade de fracasso em cada tentativa

Na seção 3.3.4, usamos o seguinte problema como exemplo de aplicação do modelo geométrico: Suponha que você lança um dado repetidas vezes, até conseguir obter pela primeira vez a face 6. A variável de interesse é X: número de tentativas que você teve que fazer. Suponha agora que você lança o dado até conseguir obter três vezes a face 6. A probabilidade de você levar 10 tentativas para conseguir os três sucessos será dada por:

$r = 3$
$k = 10$
$p = 1/6$
$q = 1 - p = 5/6$

$$p(10) = C_{10-1}^{3-1}\, p^3 q^{10-3} = C_9^2\,(1/6)^3 (5/6)^{10-3} = 0{,}0465$$

Nota: alguns livros (e o R) definem a variável X não como o número de *tentativas* feitas, mas sim o número de *fracassos* ocorridos até que sejam obtidos *r* sucessos. Por isso, para calcular *p(10)* como feito acima, usando a função `dnbinom` do R, é preciso especificar X como

```
x=k-r
pr=dnbinom(x,size=r,prob=p)
```

(ii) Valor esperado e variância

No exemplo acima, podemos considerar que a variável X é dada pela soma de três outras variáveis independentes:

$X = X_1 + X_2 + X_3$

onde:

X_1: número de tentativas até ser obtido o primeiro sucesso
X_2: número de tentativas adicionais até ser obtido o segundo sucesso
X_3: número de tentativas adicionais até ser obtido o terceiro sucesso

Estas três variáveis têm a mesma distribuição geométrica, cujos parâmetros são:

$$E(X_i) = \frac{1}{p} \qquad i=1,\ 2,\ 3$$

$$V(X_i) = \frac{q}{p^2}$$

Usando as propriedades do valor esperado e da variância de VADs (ver seção **3.3.1.5**), podemos então calcular *E(X)* e *V(X)* como:

$$E(X) = E(X_1) + E(X_2) + E(X_3) = 3 \times \frac{1}{6} = 0{,}5$$

$$V(X) = V(X_1) + V(X_2) + V(X_3) = 3 \times \frac{5/6}{(1/6)^2} = 3 \times \frac{5}{6} \times 6^2 = 90$$

3.3.7.3. Modelo Multinomial

Este modelo é uma generalização do modelo binomial: ao invés de cada tentativa poder resultar em dois resultados diferentes (*sucesso* e *fracasso*), cada tentativa pode resultar em k resultados diferentes. Portanto, ao invés de trabalharmos com uma variável única, X: número de sucessos obtidos, iremos trabalhar com k variáveis diferentes:

X_1: número de resultados do tipo 1
X_2: número de resultados do tipo 2, etc.

Um exemplo de problema que pode ser resolvido pelo modelo multinomial é o seguinte:

> Numa região, 70% das aves de uma espécie são da subespécie A, 20% são da subespécie B, e 10% da subespécie C. Se 10 destas aves são capturadas para estudo, qual é a probabilidade de que entre elas haja cinco da subespécie A, três da B e duas da C?

As variáveis envolvidas neste problema são:
X_A : número de aves da subespécie A capturadas,
X_B : número de aves da subespécie B capturadas,
X_C : número de aves da subespécie C capturadas.

O enunciado pede que calculemos a probabilidade $P(X_A = 5, X_B = 3, X_C = 2)$. Esta probabilidade pode ser calculada por meio do modelo *multinomial*, a partir dos parâmetros:
$n = 10$ (número total de aves capturadas)
$p_A = 0{,}7$
$p_B = 0{,}2$
$p_C = 0{,}1$

Em termos gerais, a função de probabilidade do modelo *multinomial* é dada por:

Se num experimento são feitas ***n*** tentativas independentes, e cada uma delas pode resultar em k resultados diferentes:
$k = 1, 2, 3, ..., n$

cujas probabilidades são constantes e iguais a:
$p_1, p_2, ..., p_k$

a probabilidade de obtermos:
n_1 resultados do tipo 1,
n_2 resultados do tipo 2,
...
n_k resultados do tipo k,

será calculada por:

$$P(X_1 = n_1, ..., X_k = n_k) = \frac{n!}{n_1! n_2! ... n_k!} p_1^{n_1} ... p_k^{n_k}$$

respeitadas as condições:

$$\sum n_i = n$$
$$\sum p_i = 1$$

Os valores esperados e variâncias das variáveis X_i são dados por:

$$E(X_i) = np_i$$
$V(X_i) = np_i(1 - p_i)$; ou, fazendo $q_i = 1 - p_i$, $V(X_i) = np_i q_i$

Retornando ao exemplo. São conhecidos os parâmetros:

$p_1 = 0{,}7$ $p_2 = 0{,}2$ $p_3 = 0{,}1$
$n_1 = 5$ $n_2 = 3$ $n_3 = 2$ $n = 10$

Portanto,

$$P(X_A = 5, X_B = 3, X_C = 2) = ?$$
$$P(5,3,2) = \frac{10!}{5!\,3!\,2!} \times 0{,}7^5 \times 0{,}2^3 \times 0{,}1^2 = 0{,}0339$$