Métodos Estatísticos de Apoio à Decisão

Aula 4

Mônica Barros, D.Sc.

Julho de 2008

Programa do Curso

Aula	Tipo (T-P-C)	Tema	Descrição
Disciplina		Métodos Estatísticos de Apoio à Decisão - BI MASTER 2008	
Responsável		Mônica Barros	
Ferramentas		Excel, @Risk	
1	T, P	Estatística Descritiva	Gráficos, tabelas e medidas numéricas
2	T	Probabilidade: Definições básicas	Definições básicas: probabilidade, espaço amostral, eventos, propriedades das probabilidades, Probabilidade Condicional, Independência;Teorema de Bayes
3	T	Probabilidade: Definições básicas	Variáveis Aleatórias Contínuas e Discretas , Função de Probabilidade, Função Densidade, Função de Distribuição, Momentos de uma v.a., Média, Variância e Desvio Padrão
4	T, P	Probabilidade: Definições básicas	Variáveis Discretas: Bernoulli, Binomial, Hipergeométrica, Geométrica, Binomial Negativa, Poisson
5	T, P	Probabilidade: v.a. Contínuas	Variáveis Contínuas: Uniforme, Exponencial, Normal
6	P	Pratica 1	Aula de exercícios - As funções do Excel para cálculo de probabilidades para v.a. Contínuas e discretas
7	T, C	Probabilidade: v.a. Contínuas E CASE 1: Simulação - soma de v.a. e o teorema central do limite CASE 2: Otimização de um portfolio simulado - propriedades da média e variância e o uso do Solver	O teorema central do limite e a importância da distribuição Normal.O teorema central do limite na prática - soma de variáveis aleatórias e a convergência para a Normal. Distribuição da soma de v.a. e da média amostral. Propriedades da média e variância de combinações lineares de v.a. - o efeito da correlação. O uso do Solver do Excel
8	T, P	Distribuições Amostrais	Amostra aleatória simples, distribuição da média amostral, distribuição de p^
9	T, P	Estatística - estimação pontual	Estimação da média da população com sigma conhecido e desconhecido e para proporções
10	T/P	Estatística - estimação por intervalos	Intervalos de confiança para amostras Normais e proporção Binomial - Exercícios - intervalos de confiança empregando o Excel
11	T/P	Estatística - testes de hipóteses	Teste de hipótese para amostrais normais e Exercícios

Aula 4

- Variáveis Aleatórias Discretas
  - Bernoulli
  - Binomial
  - Hipergeométrica
  - Geométrica
  - Binomial Negativa
  - Poisson

Distribuição Bernoulli

- É a mais simples v.a. discreta.
- Seja X uma v.a. com apenas dois valores possíveis, “sucesso” (denotado por 1) e “falha” (denotado por 0).
- Então:

$$f(1) = \Pr(X = 1) = p$$

$$f(0) = \Pr(X = 0) = 1 - p = q$$

- Note que 0 < p < 1 e “sucesso” e “falha” não indicam se o resultado de uma experiência é “bom”ou “ruim”.

Distribuição Bernoulli

- A distribuição de Bernoulli serve como um “tijolo” para a construção de modelos mais elaborados, como a Binomial, a Geométrica e a Binomial Negativa.

- Podemos reescrever a função de probabilidade como:

$$f(x) = \Pr(X = x) = p^x (1-p)^{1-x} \quad \text{onde } x = 0,1$$

- Esta última notação será útil para identificar uma variável Bernoulli apenas como um caso particular de uma Binomial.
- Notação: X ~ Bernoulli(p)

Distribuição Binomial

- A situação clássica em que usamos uma variável Binomial é:
  - Uma experiência aleatória tem apenas dois resultados possíveis : "sucesso" e "falha", onde a probabilidade de "sucesso" é <u>p</u> e a probabilidade de "falha" é q = 1- p.

  - A experiência é repetida um número FIXO (n) de vezes em situações idênticas.

- Surge, entre outras aplicações, na amostragem COM reposição.

Distribuição Binomial

- A experiência é repetida um número fixo (n) de vezes, sempre nas mesmas condições, de tal forma que as probabilidades de "sucesso" (p) e "falha" (q =1-p) se mantêm inalteradas a cada repetição.

- As diversas repetições da experiência são feitas de maneira independente, ou seja, o resultado de uma repetição não afeta o resultado das outras.

Distribuição Binomial

- A variável aleatória X que mede o número de "sucessos" nas n repetições da experiência é uma variável discreta, com valores possíveis 0, 1, 2, ... , n.

- Dizemos que esta variável tem distribuição Binomial com parâmetros n e p, e escrevemos X ~ Bin(n, p).

Distribuição Binomial - quadro resumo

experiência aleatória	"sucesso"	"falha"	p = probabilidade de "sucesso"	n = número de repetições da experiência	X = variável aleatória Binomial
"chutar" a resposta numa prova de múltipla escolha onde cada questão tem 5 opções	acertar a resposta da questão	errar a resposta	1/5	número de questões da prova	número de respostas certas na prova
nascimento de uma criança numa família	menina	menino	1/2	número de crianças na família	número de meninas na família
jogada de um dado	sair o número 6	sair qualquer outro número	1/6	número de jogadas do dado	número de vezes em que saiu o número 6 nas n jogadas do dado
verificar se uma peça produzida numa fábrica tem defeito	peça tem defeito	peça não tem defeito	proporção de peças com defeito na população de peças	tamanho da amostra	número de peças com defeito na amostra

Distribuição Binomial – exemplo típico

- É importante ter em mente um exemplo típico, um modelo da situação que representa a aplicação de uma distribuição de uma densidade ou função de probabilidade. No caso da Binomial eu sempre sugiro a idéia da nota da prova de múltipla escolha em chinês ....
- X = número de questões "chutadas" certo numa prova de múltipla escolha em chinês (presumindo que você não saiba chinês!). Todas as questões têm a mesma probabilidade de acerto e o fato de você acertar ou errar uma questão não afeta a probabilidade das outras.
- X é a sua nota na prova.

Distribuição Binomial

- Se X ~ Bin(n,p), sua função de probabilidade é:

$$f(x) = \Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad \text{para } x = 0,1,2,\ldots,n$$

- No Excel:
  - Use a função estatística distrbinom (ou dist.binom, dependendo da versão do Excel).

Distribuição Binomial

- Por que o termo $\binom{n}{x}$ aparece?
- Este termo representa o número de resultados que fornecem exatamente <u>x</u> sucessos em exatamente <u>n</u> repetições.
- Exemplo numérico
- Joga-se uma moeda 4 vezes. Quantas seqüências existem em que se observa exatamente 2 "caras"?

Distribuição Binomial

- Resposta
- 4!/(2!2!) = 24/4 = 6
- E as seqüências de resultados são (onde H indica "cara" e T indica "coroa":
- {HHTT, HTHT, HTTH, THHT, THTH, TTHH}

Distribuição Binomial

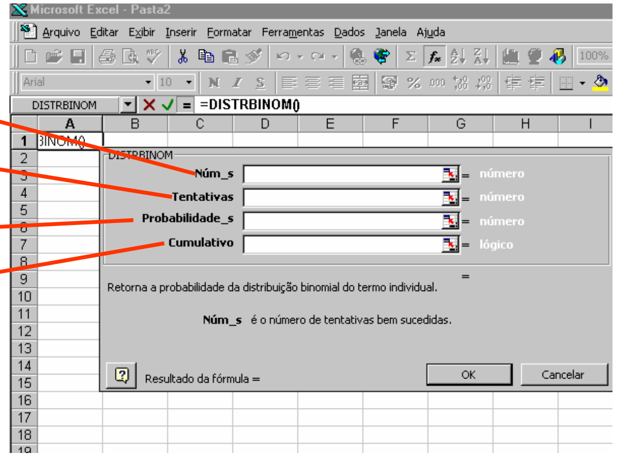
- No Excel:

Valor de X

Valor de n

Valor de p

Argumento lógico - se VERDADEIRO produz a função de distribuição acumulada, se FALSO produz a função de probabilidade



Distribuição Binomial

- Média e Variância
- Se X ~ Bin(n,p) então:
  - E(X) = n.p
  - VAR(X) = n.p.q = n.p.(1-p)
- Nota: a distribuição de Bernoulli é apenas um caso particular da Binomial com n = 1. Logo, segue que a média e a variância de uma variável Bernoulli(p) são, respectivamente, p e p.q.

Distribuição Binomial

- Exemplo
- Em uma loja, a probabilidade de um cliente realizar uma compra é de 15%. Qual a probabilidade de, entre 5 clientes que entram na loja, exatamente 3 realizarem uma compra?

$$n = 5; \quad x = 3; \quad p = 0.15$$

$$f(3) = \frac{5!}{3!.2!}(0.15)^3.(0.85)^2 = 0.024$$

Distribuição Binomial

- Exemplo

  Numa eleição supõe-se que 30% dos eleitores são favoráveis a uma certa proposta. Toma-se uma amostra de tamanho 20 de eleitores da cidade do Rio de Janeiro. Calcular as probabilidades de 4, 5, 6, 7, 8, 9 ou 10 dos eleitores na amostra serem favoráveis à proposta.

- Solução

  Seja X o número de eleitores na amostra que são a favor da proposta. Então os valores possíveis de X são 0, 1, 2, ..., 20, e X tem distribuição Binomial com parâmetros n = 20 e p = 0.30.

Distribuição Binomial

- As probabilidades para os diversos valores de X são calculadas através da fórmula:

$$\Pr(X = x) = \binom{20}{x}.(0.30)^x.(0.70)^{20-x} = \frac{20!}{x!(20-x)!}.(0.30)^x.(0.70)^{20-x}$$

- A tabela a seguir foi produzida usando a função distrbinom do Excel para x = 4, 5, ..., 10.

x	Pr(X =x)
4	13.04%
5	17.89%
6	19.16%
7	16.43%
8	11.44%
9	6.50%
10	3.08%

Exemplo (para casa – use o Excel)

- Uma empresa aérea sabe que 20% das pessoas que fazem reservas aéreas cancelam suas reservas.
- A empresa vende 50 passagens para um vôo que contém apenas 46 lugares. Supondo que as pessoas cancelam suas reservas de maneira independente, calcule a probabilidade de que haverá assentos para todos os passageiros.

Exemplo (para casa – use o Excel)

- Você arranjou um emprego numa pizzaria que funciona no sistema de entrega a domicílio.
- Apenas 5% dos pedidos são de pizza de lombinho com abacaxi.

  Você recebe exatamente 9 pedidos pelo telefone, qual a probabilidade de, no máximo, 1 pizza de lombinho com abacaxi ser pedida?

  Você recebe exatamente 30 pedidos pelo telefone num dia de bastante movimento, qual a probabilidade de receber mais de 3 pedidos de pizza de lombinho com abacaxi? Dica: pare e pense antes de fazer contas desnecessárias!!!!!!

Distribuição Hipergeométrica

- A distribuição hipergeométrica surge no contexto de **amostragem SEM reposição** em que o tamanho da população não é muito maior que o da amostra, de tal forma que a densidade Binomial não pode ser usada.

- Sejam N e n respectivamente os **tamanhos da população e da amostra**, e suponha que **na população existem r objetos do tipo A e N-r objetos do tipo B.**

Distribuição Hipergeométrica

- Binomial versus Hipergeométrica:
  - Na Hipergeométrica as repetições NÃO SÃO INDEPENDENTES – as probabilidades se alteram a cada estágio ou "retirada". Isso ocorre na amostragem SEM reposição, quando retiramos "bolinhas" de uma caixa seqüencialmente sem repô-las.
  - Na Binomial as repetições são independentes e todas têm a mesma probabilidade de "sucesso".

Distribuição Hipergeométrica

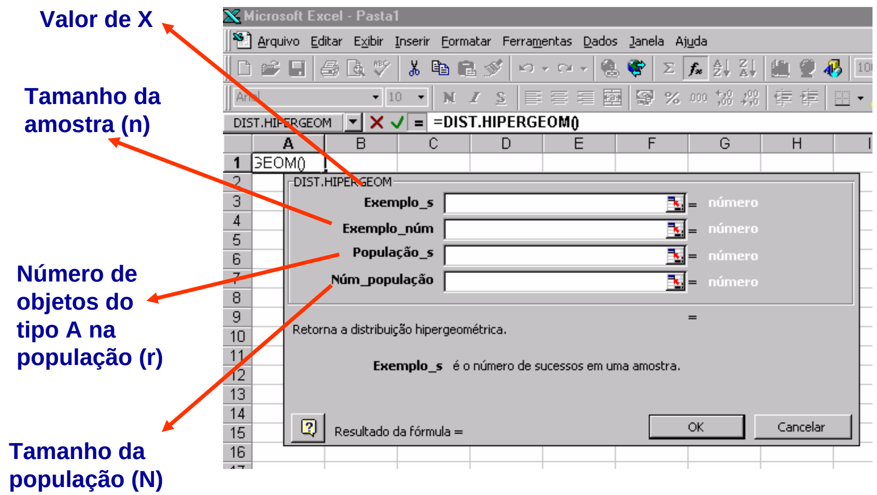
- Seja **X** a v.a. que indica o **número de objetos do tipo A na amostra**. Então a probabilidades de exatamente x objetos do tipo A na amostra é:

$$f(x) = \Pr(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

onde x = 0, 1, 2, ..., mín(n, r).

- Esta função de probabilidade também está implementada no Excel, como veremos na próxima figura.

Distribuição Hipergeométrica



Distribuição Hipergeométrica

- **Exemplo**
- Uma empresa encomenda 50 placas fax modem de um fornecedor novo para montagem nos seus micros. Sabe-se que existem 5 placas defeituosas no lote produzido.
- A empresa toma uma amostra de 10 micros sem reposição. Qual a probabilidade de 4 deles apresentarem uma placa fax modem com defeito?
- **Solução**
- Seja X o número de micros com placa defeituosa na amostra. Os valores possíveis de X são 0, 1, 2, 3, 4, 5 = mín(5,10). Procuramos calcular Pr(X=4).

Distribuição Hipergeométrica

$$\Pr(X=4)=f(4)=\frac{\binom{5}{4}\cdot\binom{45}{6}}{\binom{50}{10}}=\frac{5.(45!).(10!).(40!)}{(6!)(39!)(50!)}=$$

$$=\frac{200(45!)(10!)}{(6!)(50!)}=\frac{200(45!)(10)(9)(8)(7)(6!)}{(6!)(50)(49)(48)(47)(46)(45!)}=$$

$$=\frac{200(10)(9)(8)(7)}{(50)(49)(48)(47)(46)}=0.00396=0.396\%$$

Amostragem sem reposição

- Quando pensamos em amostragem é a maneira “natural” de fazê-la.
- Por exemplo, temos uma caixa com bolas de duas cores e vamos retirando bolas da caixa, uma a uma, sem colocá-las de volta.
- Qual a característica importante aqui?
- A cada retirada de uma bola da caixa, as probabilidades das bolas restantes se alteram.

Amostragem sem reposição

- Mais tarde, quando estudarmos amostragem com reposição, veremos que lá (com reposição) as probabilidades não se alterarão a cada “retirada”.
- É claro que a situação representada por amostragem sem reposição se aplica a problemas bem mais gerais que bolinhas numa caixa!
- Considere uma população com N objetos, a partir da qual tomamos uma amostra sem reposição de n objetos.

Amostragem sem reposição

- Os objetos na população são de dois tipos (A e B), por exemplo, bolas azuis e vermelhas.
- Qual a probabilidade de que exatamente x objetos dentre os n objetos da amostra sejam do tipo A? Esta probabilidade é:

$$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Onde

N = tamanho da população

n = tamanho da amostra (sem reposição)

r = número de objetos do tipo A na população

N - r = número de objetos do tipo B na população

x = número de objetos do tipo A na AMOSTRA

n-x = número de objetos do tipo B na AMOSTRA

Amostragem sem reposição

- De onde veio isso?

$\binom{N}{n}$ = número de maneiras de escolher conjuntos de n elementos a partir de N objetos

$\binom{r}{x}$ = número de maneiras de escolher x objetos dentre r objetos

$\binom{N-r}{n-x}$ = número de maneiras de escolher (n - x) objetos dentre (N - r) objetos

Amostragem sem reposição

- Também:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

$$\binom{r}{x} = \frac{r!}{x!(r-x)!}$$

$$\binom{N-r}{n-x} = \frac{(N-r)!}{(n-x)!(N-n-(r-x))!}$$

Estes termos são chamados de coeficientes binomiais, e servem para calcular “combinações” de objetos.

Amostragem sem reposição

- A probabilidade

$$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

- é chamada de Hipergeométrica.
- Note que: r + N - r = N e x + n - x = n, e N e n são os valores que aparecem no coeficiente binomial do denominador e medem, respectivamente, o tamanho da população e da amostra.

Mega-Sena

- Na Mega-Sena podem ser selecionados, sem reposição, os números 1, 2, 3, ..., 60.
- No total, são escolhidos, a cada sorteio, 6 números. O prêmio máximo ("sena") é dado aos apostadores que acertam os 6 números sorteados; os apostadores que acertam 5 números ("quina") e 4 números "quadra" também são premiados.

Mega-Sena

a) Você faz uma aposta em 6 números. Qual a probabilidade de acertar a "sena"? E a "quina"?

b) Responda a mesma coisa que no item a), mas suponha agora que você apostou em 7 números.

Solução

Existem 2 tipos de "objetos" aqui:

- os números que eu escolhi (e que me fazem ganhar alguma coisa)
- os "outros" números

Mega-Sena

- Logo, podemos encarar como objetos do "tipo A" os números em que eu apostei, e do "tipo B" todos os outros.
- Qual a situação aqui?
- N = tamanho da população = 60 números
- n = tamanho da amostra = 6 números (isto é, aqueles que são sorteados)
- r = número de objetos do tipo A na população (são os números em que eu apostei)
- N - r = número de objetos do tipo B na população (são os números em que eu NÃO apostei)

Mega-Sena

- x = elementos do "tipo A" na amostra, isto é, os números em que eu apostei e que estão na amostra (foram sorteados)
- n - x = objetos do "tipo B" na amostra, ou seja, os números em que eu NÃO APOSTEI e FORAM SORTEADOS.
- O que eu quero encontrar?

a) r = 6 (pois apostei em 6 números). Quero encontrar a probabilidade de x = 6 ("sena") e x = 5 ("quina").

Dica: Use a função MATEMÁTICA (NÃO É UMA FUNÇÃO ESTATÍSTICA) do Excel COMBIN.

Mega-Sena

- A probabilidade de acertar a "quina" apostando em 6 números é:

$$\frac{\binom{6}{5}\binom{54}{1}}{\binom{60}{6}} = \frac{324}{50063860} = 0.00065\%$$

- Ou seja, 1 chance em 154,518.
- A probabilidade de acertar a "sena" apostando em 6 números é:

$$\frac{\binom{6}{6}\binom{54}{0}}{\binom{60}{6}} = \frac{1}{50063860}$$

- Ou seja, a chance é de 1 em 50 milhões aproximadamente

Mega-Sena

- b) Agora eu aposto em 7 números e portanto r = 7. A probabilidade de acertar a "quina" é:

$$\frac{\binom{7}{5}\binom{53}{1}}{\binom{60}{6}} = \frac{1113}{50063860} = 0.0022\%$$

- A chance é de aproximadamente 1 em 50 mil

- Se eu aposto em 7 números, a probabilidade de acertar a "sena" é:

$$\frac{\binom{7}{6}\binom{53}{0}}{\binom{60}{6}} = \frac{7}{50063860}$$

- A chance é de aproximadamente 7 em 50 milhões

Amostragem com Reposição

- As probabilidades de seleção de um objeto permanecem *iguais* a cada retirada. Este é o aspecto fundamental para caracterizar a amostragem COM reposição.

- Um exemplo típico é: tem-se uma caixa com bolas vermelhas e azuis.

- Retira-se uma bola da caixa, verifica-se qual a sua cor e devolve-se a bola à caixa, de tal forma que as probabilidades de seleção de uma bola vermelha ou azul permanecem inalteradas a cada seleção.

Amostragem com Reposição

- A probabilidade de que exatamente x dos n objetos da amostra sejam do tipo A é:

$$\binom{n}{x}.p^x.(1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x.(1-p)^{n-x} \quad \text{para } x = 0,1,2,\ldots,n$$

- onde p = r/N é a proporção de objetos de tipo A na população (que se mantém fixa)

Amostragem com e sem reposição - Exemplo

- Um lote de 100 peças é produzido e sabe-se que 5% do lote tem defeito. Uma amostra de 10 peças é obtida *sem reposição*, e desejamos encontrar a probabilidade de não encontrarmos qualquer peça defeituosa na amostra.
- Seja A o evento: "não existem peças defeituosas na amostra ". Então:
- $$\Pr(A) = \frac{\binom{95}{10}\cdot\binom{5}{0}}{\binom{100}{10}} = 0.58375$$

Amostragem com e sem reposição - Exemplo

- Note que a proporção de peças defeituosas na população é conhecida (5%), correspondendo a 5 itens defeituosos na população.
- Suponha agora que a amostra é obtida *com reposição*.
- Um lote de 100 peças é produzido e sabe-se que 5 % do lote tem defeito. Uma amostra de 10 peças é obtida com reposição, e desejamos encontrar a probabilidade de não existir qualquer peça defeituosa na amostra.

Amostragem com e sem reposição - Exemplo

- Seja A o evento: "não existem peças defeituosas na amostra ". Então:

$$p = \frac{r}{N} = \frac{5}{100} = 0.05 \qquad \binom{10}{0} = \frac{10!}{(0!)(10!)} = 1$$

- e $\Pr(A) = 1.(0.05)^0.(0.95)^{10} = (0.95)^{10} = 0.5987$
- Este resultado é bastante parecido com o encontrado através de amostragem sem reposição (0.5837) , mas isto *não é* uma regra geral, como indica o próximo exemplo.

Amostragem com e sem reposição - Exemplo 2

- Uma caixa tem 10 bolas, 3 pretas e 7 brancas. Qual a probabilidade de, numa amostra de 5 bolas, encontrar 0, 1, 2, 3, 4 ou 5 bolas pretas supondo que a amostragem é feita com e sem reposição.
- Solução

a) Sem reposição

Seja X o número de bolas pretas na amostra.

$$\Pr(X = x) = \frac{\binom{3}{x}\binom{7}{5-x}}{\binom{10}{5}}$$

Amostragem com e sem reposição - Exemplo 2

- Note que Pr(X=4) = Pr(X = 5) = 0 pois não se pode ter 4 ou 5 bolas pretas na amostra se só existem 3 bolas pretas na caixa e a amostragem for sem reposição.

b) Com reposição

Seja X o número de bolas pretas na amostra. Aqui podem existir mais de 3 bolas pretas na amostra pois a amostragem é feita com reposição. Também, p = r/N = 3/10 e q = 1-p = 7/10.

$\Pr(X = x) = \binom{5}{x}(0.3)^x (0.7)^{5-x}$ onde x = 0, 1, 2, 3, 4, 5.

Amostragem com e sem reposição - Exemplo 2

- A próxima tabela indica os valores das probabilidades nos dois casos (sem e com reposição).

valor de x	Pr(X =x) sem reposição	Pr(X =x) com reposição
0	0.0833	0.1681
1	0.4167	0.3602
2	0.4167	0.3087
3	0.0833	0.1323
4	0	0.0284
5	0	0.0002

- Neste caso as probabilidades são bastante diferentes dependendo do tipo de amostragem: com ou sem reposição.

Comparação entre amostragem com e sem reposição

- Qual a diferença entre este exemplo e o anterior?
- Por que neste último caso as probabilidades foram tão diferentes dependendo do tipo de amostragem usado, enquanto no exemplo anterior quase não houve diferença?
- A resposta está na ***relação entre o tamanho da amostra e o tamanho da população***.

Comparação entre amostragem com e sem reposição

- No primeiro exemplo, o tamanho da população era 100, e o da amostra 10, ou seja, a população era 10 vezes maior que a amostra.
- Nesta situação faz muito pouca diferença se a amostragem é com ou sem reposição.
- No segundo exemplo, o tamanho da população é 10, e o da amostra é 5, e assim a população tem apenas o dobro do tamanho da amostra. Nesta situação, os resultados obtidos pelos 2 tipos de amostragem são bastante diferentes.

Comparação entre amostragem com e sem reposição

- Em resumo:
- Se N (o tamanho da população) é muito maior que n (tamanho da amostra), amostragens com e sem reposição fornecem praticamente as mesmas probabilidades.
- Se N e n estão próximos, as amostragens com e sem reposição levam a resultados muito diferentes.
- É importante também notar que, em todos os exemplos envolvendo amostragem com e sem reposição, a população inclui objetos de apenas dois tipos.

Amostragem sem reposição - para casa

- Qual a probabilidade de você acertar a Mega-Sena jogando 12 números?
- Como você pode jogar 12 números e quanto isso vai custar, supondo que o cartão com 10 números custe R$ 210, como no exemplo desta aula? Lembre-se que o "maior" cartão que existe contém 10 números, então você terá que inventar uma estratégia para produzir algo equivalente a um cartão de 12 números.
- Escreva o resultado em termos de probabilidades e em termos de chances de "1 em xxx".
- Faça as contas no Excel.

Amostragem com e sem reposição - para casa

- Uma rede de TV lançou um concurso para que as pessoas que quisessem participar de um "reality show". 335 candidatos se inscreveram, dos quais 60 eram universitários. Toma-se uma amostra aleatória de 30 candidatos.
- a) Calcule a probabilidade de, na amostra, encontrar menos de 6 universitários, se a amostragem foi feita COM reposição.
- b) Calcule a mesma probabilidade que no item anterior, supondo agora que a amostragem é feito SEM reposição.

Amostragem com e sem reposição - para casa

- Um grande banco lançou um programa de "trainees". 300 candidatos se inscreveram, dos quais 80 eram cariocas e os outros do restante do país. 50 candidatos foram escolhidos *aleatoriamente* para o último processo de seleção (pois o gerente de RH da empresa não tinha tempo de avaliar cada curriculum individualmente).
- a) Calcule a probabilidade de, na última etapa da seleção, encontrar pelo menos 6 cariocas, se a amostragem foi feita COM reposição.
- b) Calcule a mesma probabilidade que no item a, supondo agora que a amostragem é feita SEM reposição.

Distribuição Geométrica

- **Da mesma maneira que a Distribuição Binomial, a Geométrica também parte de repetições de Bernoulli independentes.**
- **A grande diferença em relação à Binomial é que, no caso da Geométrica, *o número de repetições não é fixo*, e as repetições são feitas até encontrar o primeiro "sucesso" (e então a experiência pára).**

Distribuição Geométrica

- **Um exemplo típico é: imagine uma caixa com um número muito grande de peças. Estas peças podem estar OK ou defeituososas, e a probabilidade de encontrar uma peça defeituosa é FIXA e igual a "p".**
- **Retiram-se peças da caixa até encontrar uma peça com defeito.**

Distribuição Geométrica

- **Neste exemplo, um "sucesso" indica encontrar uma peça defeituosa.**
- **Note que estamos supondo que a caixa onde estão as peças contém um número muito grande de objetos, e assim a probabilidade de retirada de uma peça defeituosa é mantida constante (p).**

Distribuição Geométrica

- **Seja X o número de repetições necessárias até encontrar uma peça com defeito.**
- **Então a função de probabilidade de X é:**

$$f(x) = \Pr(X = x) = q^{x-1}.p = (1-p)^{x-1} p \quad \text{onde } x = 1,2,3,....$$

- **Pode-se provar, usando a série geométrica, que f(x) assim definida é uma função de probabilidade.**

Distribuição Geométrica - Exemplo

- Você arranjou um emprego numa empresa que faz pesquisas de opinião pelo telefone.
- Apenas 10% das chamadas resultam numa pesquisa completa, isto é, apenas 10% dos entrevistados responde todo o seu questionário. Calcule as seguintes probabilidades:
- a) De que a primeira pesquisa completa será respondida na 5a. ligação telefônica.
- b) De que a primeira pesquisa completa será respondida na 8a. ligação telefônica.

Distribuição Geométrica - Exemplo

- Esta é uma aplicação típica da distribuição Geométrica.
- Seja X o número de ligações efetuadas até que a 1a. pesquisa completa seja respondida.
- a) Aqui estamos perguntando a probabilidade da seqüência FFFFS, onde a probabilidade de sucesso é p = 0.10.

$$\Pr(X=5)=(0.90)^4(0.10)=6.56\%$$

- b) $\Pr(X=8)=(0.90)^7(0.10)=4.78\%$

Distribuição Geométrica – para casa

- Exemplo
- Um gestor de fundos de investimento ultrapassa a sua meta de retorno mensal 85% das vezes e nos 15% restantes tem um resultado ruim (abaixo da meta).
- Qual a probabilidade dele ter o primeiro resultado ruim nos próximos 12 meses?
- E nos primeiros 6 meses?

Distribuição Geométrica

- A função de probabilidade de X é:

$$f(x)=\Pr(X=x)=q^{x-1}.p \text{ onde } x=1,2,3,....$$

- Notação: X ~ Geom(p)
- Média e Variância
- Se X ~ Geom(p) então:
  1) E(X) = 1/p
  2) VAR(X) = q/p^2

Distribuição Binomial Negativa

- A distribuição Binomial negativa é mais uma função de probabilidade derivadas de tentativas de Bernoulli independentes e é uma *generalização da distribuição Geométrica*.
- Suponha que *repetimos um número indefinido de vezes* uma experiência que resulta em sucesso ou falha. As repetições terminam quando encontramos o r-ésimo sucesso, onde r é um número especificado a priori.

Distribuição Binomial Negativa

- Seja X a variável aleatória que representa a tentativa onde o r-ésimo sucesso ocorre. Então a função de probabilidade de X é:

$$f(x) = \Pr(X = x) = \binom{x-1}{r-1} . p^r . q^{x-r} \quad \text{onde } x = r, r+1, r+2, ...$$

- Note que o valor mínimo de X é r, pois precisamos fazer pelo menos r repetições para encontrar r sucessos! Também, a combinação que aparece na densidade indica que, das x-1 repetições anteriores à última (que é necessariamente um "sucesso", r-1 são "sucessos").

Distribuição Binomial Negativa

- Também, $p^r.q^{x-r}$ é a probabilidade de uma seqüência qualquer contendo r "sucessos" e x-r "falhas".
- Se X ~ NegBin(r, p), sua média e variância são:

$$E(X) = r/p$$

$$VAR(X) = rq/p^2$$

- Nota – a função de probabilidade Binomial Negativa com parâmetros r = 1 e p é apenas a Geométrica.
- Na verdade, uma variável Binomial Negativa (r, p) é apenas a soma de r Geométricas(p), todas independentes. Faz sentido, não?

Distribuição Binomial Negativa - Exemplo

- Um comprador em potencial entra numa loja de carros a cada hora.
- Um vendedor tem probabilidade 0.25 de concluir uma venda. O vendedor decide trabalhar até conseguir vender 3 carros num só dia.
- Qual a probabilidade de que o vendedor tenha de trabalhar exatamente 8 horas para conseguir vender os 3 carros? E mais de 8 horas?

Distribuição Binomial Negativa - Exemplo

- Seja X o número de horas de trabalho necessárias para vender 3 carros.
- Então X tem distribuição Binomial Negativa com parâmetros r = 3 e p = 0.25.

$$\Pr(X = x) = \binom{x-1}{2}(0.25)^3.(0.75)^{x-3} \text{ para } x = 3, 4, 5, ...$$

- A próxima tabela exibe os valores das probabilidades para X = 3, 4, ..., 8.

Distribuição Binomial Negativa - Exemplo

x	Pr(X = x)
3	0.01563
4	0.03516
5	0.05273
6	0.06592
7	0.07416
8	0.07787

A probabilidade de trabalhar exatamente 8 horas é Pr (X = 8) = 0.07787. A probabilidade de trabalhar mais de 8 horas é:

$$\Pr(X > 8) = 1 - \Pr(X \le 8) = 1 - \{\Pr(X = 3) + \Pr(X = 4) + ... + \Pr(X = 8)\} =$$
$$= 1 - \{0.01563 + 0.03516 + ... + 0.07787\} = 1 - 0.32146 = 0.67854$$

Distribuição Binomial Negativa - Exemplo

- Uma fábrica de sorvetes decidiu fazer uma campanha para aumentar suas vendas.
- A cada 50 sorvetes produzidos um é premiado, e o prêmio consiste em ganhar um outro sorvete grátis. Cada sorvete é vendido por R$ 0.80.
- a) Se você decide comprar sorvetes até encontrar um sorvete premiado, quanto você espera gastar?
- b) E se você comprar sorvetes até encontrar o 2o. sorvete premiado?

Distribuição Binomial Negativa - Exemplo

- a) Seja X o número de sorvetes comprados até encontrar o 1o. sorvete premiado. Então X tem distribuição Geométrica com parâmetro p =1/50. Seja C o custo deste procedimento. Então C = 0.8X e E(C) = 0.8.E(X). Mas, E(X) = 1/p = 50 e assim o custo esperado é de R$40.

- b) Neste caso X mede o número de sorvetes comprados até encontrar o 2o. sorvete premiado. Então X tem distribuição Binomial Negativa com r = 2 e p =1/50. Agora E(C) = 0.8E(X) = 0.8(2)(50) = R$ 80.

Distribuição Binomial Negativa – para casa

- Uma gulosa professora de estatística é "fissurada" por trufas de chocolate. Em busca da trufa ideal, ela vai provando chocolates em diversas lojas, de maneira independente.

- A probabilidade dela gostar de uma trufa que prova é 70%. Ela decide passear por um shopping, provando todas as trufas que encontra, e decide parar só ao encontrar a 4ª. trufa "maravilhosa"(para "desespero" da balança que tem em casa!).

Distribuição Binomial Negativa – para casa

- Qual a probabilidade dela ter que:
- Provar 6 trufas até encontrar a 4ª. trufa maravilhosa?
- Ter que "sofrer", provando 10 trufas, até encontrar a 4ª. trufa maravilhosa?

Distribuição de Poisson

- Está associada a experiências que modelam o número de ocorrências de um evento dentro de um determinado intervalo de tempo (ou espaço) específico, quando estes eventos ocorrem com uma taxa média conhecida, por exemplo:
  - Número de carros que passam por uma estrada no intervalo de uma hora
  - Número de buracos por km de uma rodovia
  - Número de assassinatos num final de semana
  - Número de defeitos por metro de tecido produzido

Distribuição de Poisson

- Número de erros de digitação numa página de texto
- Número de mutações num trecho de DNA após a exposição a uma certa quantidade de radiação
- Número de soldados mortos por chutes de cavalo a cada ano na cavalaria Prussiana. Este exemplo ficou famoso num livro de Bortkiewicz (1868–1931).

Distribuição de Poisson

- Um experimento de Poisson possui as seguintes características:
  - A probabilidade de uma ocorrência é a mesma para dois intervalos de igual comprimento (ou duração);
  - A ocorrência (ou não ocorrência) num determinado intervalo é independente da ocorrência ou não ocorrência em um outro intervalo.
- A distribuição foi descoberta por **Siméon-Denis Poisson (1781–1840)** e publicada em 1838 em seu trabalho “Recherches sur la probabilité des jugements en matières criminelles et matière civile” (Wikipedia).

Distribuição de Poisson

- A função de probabilidade para uma variável Poisson com parâmetro μ é:

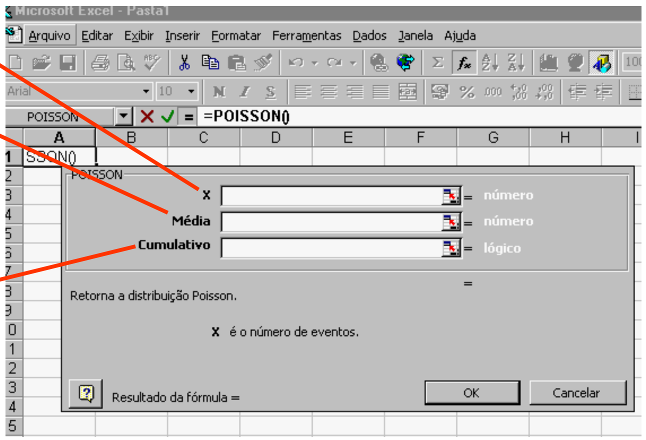
$$f(x) = \Pr(X = x) = \frac{\mu^x e^{-\mu}}{x!} \quad \text{onde} \quad x = 0,1,2,....$$

- **Este parâmetro μ é a MÉDIA da distribuição e indica o número esperado de ocorrências num dado intervalo.**
- A distribuição Poisson é freqüentemente utilizada na modelagem de **eventos “raros”**, ou seja, a probabilidade de X = 0 ou X = 1 (pequeno número de ocorrências no intervalo de tempo especificado) é grande.

Distribuição de Poisson no Excel

Valor de X

Parâmetro da função de probabilidade



Argumento lógico

Se FALSO fornece a função de probabilidade f(x)

Distribuição de Poisson

- **Média e Variância**
- Se X ~ Poisson(μ) então:
  - E(X) = μ
  - VAR(X) = μ
- Ou seja, na Poisson, média e variância são iguais!

Distribuição de Poisson

- Exemplo
- O número médio de clientes que entram em um banco num período de 15 minutos é 10.
- Qual a probabilidade de entrarem exatamente 5 clientes em 15 minutos?

$$\mu = 10; \; x = 5$$

$$f(5) = \Pr(X = 5) = \frac{10^5 e^{-10}}{5!} = 0.0378$$

Distribuição de Poisson

- Qual a probabilidade de entrarem menos de 3 clientes no banco num intervalo de 5 minutos?
- Note que agora μ = 10/3 pois o intervalo de tempo foi reduzido a 1/3 do original (passou de 15 para 5 minutos). A pergunta é: qual a Pr(X < 3) usando esta nova distribuição de Poisson.

Distribuição de Poisson

- Usando o Excel (lembrando que a distribuição indicada é agora Poisson(10/3):

x	Prob
0	0.0357
1	0.1189
2	0.1982
Pr(X<3)	0.3528

- Função Poisson do Excel com argumentos 0, 10/3 e FALSO

- soma das probabilidades de X = 0, 1 e 2

Distribuição de Poisson

- Exemplo
- Numa campanha de caridade feita por um programa de TV em todo o Brasil, o número de pessoas que contribuem mais de 500 reais é uma variável aleatória com média de 5 pessoas por programa.
  - a) Calcule a probabilidade de que, num certo programa, o número de pessoas que contribuem mais de 500 reais exceda 8.
  - b) Faça o gráfico da função de probabilidade.
  - c) Faça um gráfico da função de distribuição acumulada.

Distribuição de Poisson

- Solução

Seja X o número de pessoas que contribuem com mais de 500 reais a cada programa. Desejamos calcular Pr{ X > 8}.

$\Pr\{ X > 8\} = 1 - \Pr\{ X \leq 8\} = 1 - F(8)$

onde F(.) denota a função de distribuição acumulada.

A tabela a seguir apresenta a função de probabilidade, a função de distribuição acumulada e seu complemento.

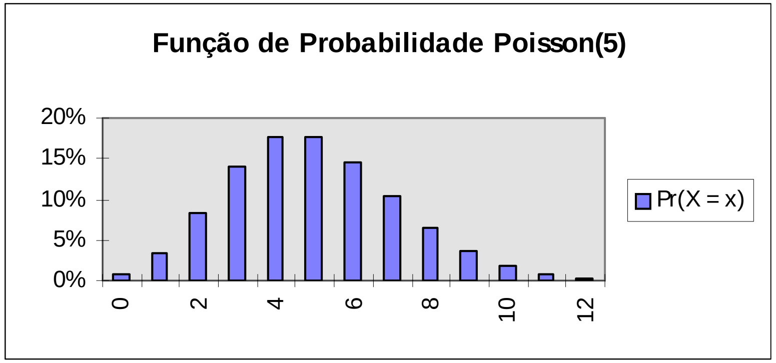
Da tabela segue que Pr(X > 8) = 6.81%.

Distribuição de Poisson

x	Pr(X = x)	Pr(X <= x)	1- F(x) = Pr(X > x)
0	0.67%	0.67%	99.33%
1	3.37%	4.04%	95.96%
2	8.42%	12.47%	87.53%
3	14.04%	26.50%	73.50%
4	17.55%	44.05%	55.95%
5	17.55%	61.60%	38.40%
6	14.62%	76.22%	23.78%
7	10.44%	86.66%	13.34%
8	**6.53%**	**93.19%**	**6.81%**
9	3.63%	96.82%	3.18%
10	1.81%	98.63%	1.37%
11	0.82%	99.45%	0.55%
12	0.34%	99.80%	0.20%

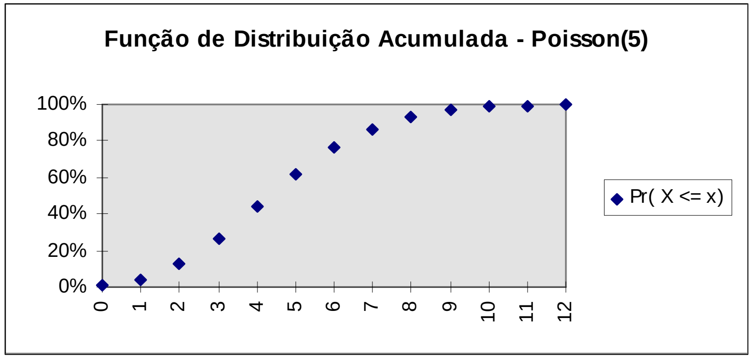
Distribuição de Poisson

b) A função de probabilidade é dada no gráfico abaixo



Distribuição de Poisson

c) A função de distribuição acumulada é mostrada a seguir.



Distribuição de Poisson

- O número de chamadas para um telefone com prefixo 800 (chamadas grátis) é uma variável aleatória com média de 3 chamadas por minuto.
- Qual a probabilidade do número de chamadas num minuto ser maior que 4?
- Solução
- Suponha que a distribuição do número de chamadas é Poisson com a média indicada (3 chamadas por minuto).

Distrbuição de Poisson

- Logo, a função de probabilidade é:

$$f(x) = \Pr(X = x) = \frac{e^{-3}(3)^x}{x!} \quad \text{para } x = 0,1,2,....$$

- A probabilidade desejada é:

$$\Pr(X > 4) = \Pr(X = 5) + \Pr(X = 6) + ... = 1 - \Pr(X \le 4) = 1 - \sum_{x=0}^{4} \frac{e^{-3}(3)^x}{x!} =$$

$$= 1 - e^{-3}\left\{1 + 3 + \frac{9}{2} + \frac{27}{6} + \frac{81}{24}\right\} = 81.53\%$$

Exemplo (para casa – use o Excel)

- O número de enchentes em cada verão no Rio de Janeiro é uma variável aleatória Poisson com média de 2.3 enchentes por verão.
  - Calcule a probabilidade de ocorrerem exatamente 3 enchentes em um verão qualquer.
  - Calcule a probabilidade de ocorrerem menos de 50 enchentes em 20 verões.

Exemplo (para casa – use o Excel)

- O número de carros que chegam num posto de pedágio é uma variável Poisson com parâmetro 2.8 carros por minuto.
- Use o Excel para calcular:
  - A probabilidade de passarem mais de 4 carros num minuto.
  - A probabilidade de passarem menos de 25 carros em 10 minutos.

Exemplo (para casa – use o Excel)

- O número médio de pedidos de autorização para um certo exame médico complexo recebido por um plano de saúde é uma variável Poisson com parâmetro $\lambda = 4.6$ pedidos por hora.

- Calcule a probabilidade de, numa hora qualquer, a empresa receber mais de 5 pedidos de autorização para este exame.

- Calcule a probabilidade da empresa receber, em uma hora, 9 ou menos pedidos de autorização.