Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica

Disciplina do 1º Semestre de 2020

IMECC - UNICAMP

Curso: Bacharelado em Ciência da Computação

ME323 - Introdução aos Modelos Probabilísticos

Respostas Prova Final

Alunos:

Gabriel Volpato Giliotti RA: 197569
Pedro Barros Bastos RA: 204481
João Vitor Pimentel da Silva RA: 199921

Professor: Sebastião Amorim

Campinas – SP
2020

1-) Numa área metropolitana extensa, N pessoas foram selecionadas ao acaso e submetidas a teste para o Coronavírus. Foram encontrados n casos positivos. Com base nesses dados, construa um intervalo de 95% de confiança (i95%c) para o número de casos positivos por milhão de habitantes daquela A.M. [Faça suas próprias escolhas sensatas para N e n].

Considere N = 10000 pessoas e n = 1000 casos positivos, então:

- Uma estimativa inicial para p é dada por pch (p chapéu) = x / N. Logo, sendo x = n, p chapéu = 1000/10000 = 0,1;
- A Esperança de X - E(X) = N*p - é igual ao próprio número de casos positivos, logo E(X) = N*pch = 10000 * 0,1 = 1000;
- A Variância de X - V(X) = n*p*(1 - p) - é calculada com pch, logo V(X) = N*pch*(1 - pch), assim, V(X) = 10000 * 0,1 * 0,9 = 900;
- O desvio padrão de X - Dp(X) = $\sqrt{V(X)}$ - é dado pela raiz da variância calculada anteriormente, assim, Dp(X) = $\sqrt{900}$ = 30;
- Finalmente, para um intervalo de 95% de confiança, utilizamos 2 desvios-padrão, ou seja, 2dp = +/- 60.

O intervalo com 95% de confiança calculado é dado por (940,1060). Aplicando para as condições do enunciado, temos que para 1 milhão de habitantes, o intervalo de infectados com 95% de confiança é dado por (94.000, 106.000).

<table>
<tr><th colspan="3">Dados</th></tr>
<tr><td>N</td><td colspan="2">10.000</td></tr>
<tr><td>n(=x)</td><td colspan="2">1.000</td></tr>
<tr><th colspan="3">Fórmulas e resultados</th></tr>
<tr><td>p = x / N</td><td colspan="2">0,1</td></tr>
<tr><td>E(X) = N*p</td><td colspan="2">1.000</td></tr>
<tr><td>V(X) = N*pch(1 - pch)</td><td colspan="2">900</td></tr>
<tr><td>Dp = raiz(V(X))</td><td colspan="2">30</td></tr>
<tr><td>2*Dp</td><td colspan="2">60</td></tr>
<tr><th colspan="3">Resultado p/ 1 milhão de habitantes</th></tr>
<tr><td>i95%c</td><td>94.000</td><td>106.000</td></tr>
</table>

2-) Os EUA lideram as estatísticas mundiais de casos confirmados e de mortes pelo COVID-19, com mais de mais de 5,5 milhões de casos e mais de 174 mil mortes. Há, naquele país, uma pressão para que se amplie as operações de testes, elevando os números para 5 milhões por dia. O presidente argumenta que aumentar os testes faz aumentar o número de casos positivos detectados, e que isso é ruim, passando uma imagem negativa do país. Segundo ele, o número de casos positivos nos EUA só são tão grandes porque testa-se muito. Técnicos da área contra argumentam que o importante é a taxa de casos positivos, e que, como ela vem aumentando muito, fica evidente que o número de casos no país está aumentando muito. [Comente, com argumentos probabilísticos e matemáticos, ao nível de um Engenheiro da Unicamp, com número de toques entre 1,5 e 2,5 vezes o do enunciado desse exercício.]

R: Inicialmente, a aplicação de um número elevado de testes para Covid-19 tem como objetivo a melhora de precisão no combate contra o vírus. Imagine que dois diferentes testes têm eficácia de 95%, isso é, para cada 100 pessoas testadas, 95 delas vão obter um resultado que condiz com seu real estado (estar ou não infectado), onde as outras 5 pessoas podem obter resultados falsos (conhecidos por falso-positivo). Admita também que o primeiro teste leva horas para se produzir e obter os resultados, enquanto o segundo leva minutos. Logo, para que os especialistas da área da saúde trabalhem de forma rápida e realizem testes eficazes, é necessário mapear os locais de foco da doença, mas isso só pode ser feito com uma testagem massiva e rápida da população, levando a necessidade da produção de testes como o do segundo caso (eficazes e rápidos). Desse modo, tal testagem pode produzir um intervalo de confiança de pessoas infectadas dos locais de foco, assim como uma estimativa mais precisa do número de infectados em cada um deles, o que finalmente leva os especialistas a delegar melhor seus recursos no momento de atuar com políticas públicas como isolamento, vacinas, produtos de higienização, etc. e por consequência a regressão da pandemia com achatamento da curva de crescimento do vírus.
Além disso, sabemos que quanto mais se testa, maior o número de infectados, pois o teste tem 2 possibilidades, manter o número como está, ou aumentá-lo, entretanto, para cada teste negativo, diminui-se a porcentagem de pessoas infectadas no país, que inclusive, a porcentagem de pessoas infectadas é um número que nos dá muito mais informação a respeito do achatamento ou não da curva de infecção de um país.

3-) Um grande exame nacional visando avaliar a qualidade geral do ensino de Matemática no ensino fundamental no País, aplicou uma prova extensa e muito bem estruturada a todos os cerca de 3 milhões de alunos do nono ano. Após a aplicação, e sabendo que a correção cuidadosa das provas tomaria algum tempo, um estudo amostral ligeiro foi encomendado. Para isto uma amostra aleatória de 250 provas foi sorteada e criteriosamente corrigida. Os resultados foram: Soma(Nota) = 6488,4+10*c6 e Soma(Notaˆ2) = 217309,5+100*c6, onde Nota é o resultado de cada prova amostrada, numa escala de 0 a 100.

a) Construa um i95%c para a nota média do conjunto de todas as provas.

b) Para uma avaliação amostral mais precisa, garantindo margem de erro no i95%c igual a meio ponto, quantas provas a mais deverão ser amostradas?

c) Escreva um ensaio de ~1000 toques, sobre a possibilidade de fazer todo esse projeto, não envolvendo todo o mundo [muito caro, demorado e impreciso (por que?)], mas através de uma amostra.

3-) Para o valor de C6 foi realizado o seguinte cálculo:

RA 1 → C6 = 9
RA 2 → C6 = 1 C6 = (9 + 1 + 1) / 3 = 3,66666 … Logo, C6 = 4
RA 3 → C6 = 1

Soma(Nota) = 6488,4+10*4 = 6528,4
Soma(Notaˆ2) = 217309,5+100*4 = 217709,5

Item a) Para calcular o intervalo das médias das notas com 95% de confiança, fazemos:

Dados relativos à amostra (que podem não refletir muito bem a real média das notas dos 3 milhões de alunos)

Média = $E(X)$ = Soma(Nota) / n = 6528,4 / 250 = 26,1136
$E(X^2)$ = Soma(Nota^2) / n = 217709,5 / 250 = 870,838
$E(X)^2$ = 681,9201049

Var(Nota) = $E(X^2) - E(X)^2$ = 188,9178951
Dp(Nota) = $\sqrt{Var(Nota)}$ = 13,74474063

Dados relativos ao universo (podem refletir melhor a média das notas dos 3 milhões de alunos):
Var(nota média) = Var(nota) / n = 188,9178951 / 250 = 0,7557
Dp(nota média) = $\sqrt{Var(Nota\ média)}$ = 0,86931
2*Dp = 2 * 0,86931 = 1,73862

Como nos exercícios anteriores, para o intervalo de média de notas dos 3 milhões de alunos com 95% de confiança temos a média mais ou menos 2 desvios padrões, ou seja, a nota média se encontra, com 95% confiança, no intervalo (24,375, 27,852).

Item b) Temos então Erro = 0,5, z = 2 e Var(Nota) = 188,9178951 (Calculado no item a). O número de provas necessárias, com 95% de confiança, para uma avaliação mais precisa, é dada por:

$Amostras\ n\ =\ ((z^2) * Var(Nota))/Erro^2$, logo
Amostras n = (4*188,9178951)/0,25 = 755,6716 / 0,25 = 3022,7 ≃ 3023 provas.

Como o enunciado pede pelo número excedente, temos que o número de provas a mais é 3023 - 250 = 2773 provas a mais.

Item c)
Com os cálculos dos itens anteriores poderíamos utilizar um número muito menor de provas e apresentar um maior significado dos resultados com o esforço de corrigir uma parcela do total de provas, com uma taxa de confiança de 95%, assim tirando conclusões qualitativas sobre o desempenho dos alunos.

Corrigir todos testes traria um grande custo para esforço das correções, além de um elevado gasto de tempo devido a grande quantidade de provas. Teríamos também problemas com imprecisão uma vez que corrigir mais provas poderia aumentar a chance de erros provenientes de um teste de baixa precisão.

Por outro lado, considerando um experimento cuja confiança é de 95%, existe a possibilidade de testes realizados cujo resultado não se adequa ao extrapolado pela amostra, ou seja, podem haver estudantes os quais o teste tenha diferido numa grande margem (para mais, ou para menos) do resultado amostral.

Além de tudo, também sabemos que existem diferenças nos métodos de ensino, fazer uma seleção aleatória de algumas escolas, descartaria as possíveis diferenças de qualidade de ensino por região do país e as diferenças de método de ensino que as escolas poderiam apresentar.

Finalmente, concluímos que, apesar da possibilidade de realizar o projeto através de apenas uma parcela amostral parecer interessante dado à redução da carga de trabalho e consequentemente do custo, essa estratégia se torna inviável no momento que, em escala nacional, existe uma grande margem de erro onde alguns estudantes podem receber resultados dos seus testes que são incompatíveis com o resultado real.

4-) A Bacia Hidrográfica de um determinado rio tem superfície total de 2,48 Mkm² . Pluviômetros bem espalhados por 100 pontos desta área, trouxeram após 1 ano, cem leituras de precipitação total, com as seguintes estatísticas:

$$Soma(P) = 85.495\ mm\ e\ Soma(P^2) = 79.038.144\ mm^2$$

a) Quanta água choveu nessa área nesse ano. Dê seu resultado na forma do i95%c, com unidades em km^3 .

b) Como a vazão média do rio no mar é de xxxx m^3/s (escolha seu valor. Faça essa escolha no sentido de valorizar o exercício), em quando você avalia a perda de água nessa bacia, por evaporação.

a)

$$Soma(P) = 85.495\ mm$$

$$Soma(P^2) = 79.038.144\ mm^2$$

n = 100

Primeiro calcularemos E(X) e E(X²), as esperanças de X e X²

E(X) = Soma(P)/n = 85.495/100 = 854,95 mm

E(X²) = Soma(P²)/n = 790.381,44 mm²

Agora calcularemos a variância Var(X)

Var(X)= E(X²) - E(X)² = 790.381,44 - 730.939,5025 = 59.441,9375 mm²

Com a variância Var(X), poderemos calcular o desvio padrão **σ**x

σx = $\sqrt{Var(x)} = \sqrt{59.441,9375} =$ 243,8072 mm

Em posse destas informações, podemos admitir que:

i95% tem como limites E(X) - 2**σ**x e E(X) + 2**σ**x, sendo assim

Limite inferior: 367,3356 mm

Limite superior: 1342,5644 mm

Logo, temos 367,3356 L/m² e 1342,5644 L/m²
E em km³/m², temos
Limite inferior: $3,673356 * 10^{-10} \frac{km^3}{m^2}$
Limite superior: $1,3425644 * 10^{-9} \frac{km^3}{m^2}$
Assim, a quantia de chuva, de acordo com os limites dados
Limite inferior: 911 km³
Limite superior: 3329,6 km³

b) Como dito em aula, vamos admitir que toda a chuva do local, ou será evaporada ou terá vazão para o mar, sendo assim
Chuva = evaporação + vazão
Logo, Evaporação = Chuva - Vazão

Vamos admitir uma vazão para o mar de 25.000m³/s (fiz uma tentativa com 100.000m³/s, porém a vazão era maior do que a chuva, logo não foi possível continuar os cálculos)
Sendo assim, temos que 25.000m³/s * 60 * 60 * 24 * 365 = 788.400.000.000m³/ano = 788,4km³/ano de vazão

Vamos admitir, que a quantia mais provável de chuva é a média calculada, ou seja, 854,95 mm de chuva por ano
854,95 L/m² = $8,5495 * 10^{-10} \frac{km^3}{m^2}$ = 2120,3 km³/ano

Agora que temos os dados relacionados à quantia de chuva no ano e à quantia de vazão no ano, substituindo na equação
Evaporação = Chuva - Vazão
Evaporação = 2120,3 - 788,4 = 1331,9 km³/ano
Sendo assim, avaliamos a perda de água por evaporação em 1331,9km³/ano