# CE-003: Estatística II - Turma K/O - Avaliações Semanais ($2^o$ semestre 2013)

Semana 2

1. Dois profissionais vão tentar resolver um problema e cada um deles pode ou não conseguir resolver. A chance do primeiro resolver é de 60% e do segundo 45%. Neste contexto responda às questões abaixo, fazendo suposições se necessário.
   (a) Por que este pode ser considerado um experimento aleatório?
   (b) Qual o espaço amostral?
   (c) Qual a probabilidade de que ambos resolvam o problema?
   (d) Qual a probabilidade de que o problema seja resolvido?
   (e) Foi necessária alguma suposição para para resolver os items anteriores? Se positivo, qual suposição?

   **Solução:**
   Notação:

   $$A : \text{o primeiro resolve o problema} \quad P[A] = 0,60 \quad P[\overline{A}] = 0,40$$
   $$B : \text{o segundo resolve o problema} \quad P[B] = 0,45 \quad P[\overline{B}] = 0,55$$

   (a)
   (b) $\Omega = \{(A, B), (A, \overline{B}), (\overline{A}, B), (\overline{A}, \overline{B})\}$
   (c) Supondo independência: $P[A \cap B] = P[A] \cdot P[B] = 0,60 \cdot 0,45 = 0.27$
   (d) Supondo independência: $P[A \cup B] = P[A] + P[B] - P[A \cap B] = 0,60 + 0,45 - 0.27 = 0.78$
   (e) Independência entre os profissionais, ou seja, a probabilidade de cada um resolver independe do problema ser ou não resolvido pelo outro.

---

2. (adaptado de M. & L., 2002) Em um levantamento de dados de acidentes em uma estrada foram resumidos na tabela a seguir. A partir desses dados, responda às questões propostas.

| | Vítimas fatais | |
|---|---|---|
| Motorista | Sim | Não |
| Sóbrio | 1228 | 275 |
| Alcoolizado | 2393 | 762 |

   (a) Voce diria que o fato do motorista estar ou não alcoolizado afeta a chance de ocorrer vítimas fatais? Justifique.
   (b) Obtenha a partir da tabela acima: (a) alguma probabilidade marginal, (b) alguma probabilidade condicional, (c) alguma probabilidade de interseção de eventos.
   (c) Cite um par de eventos mutuamente exclusivos.
   (d) Se um motorista alcoolizado se envolve em um acidente, qual a probabilidade de haver vítima fatal?
   (e) Se houve uma vítima fatal em um acidente, qual a probabilidade do motorista estar alcoolizado?

   **Solução:**
   Notação:

   $S$ : motorista sóbrio $\qquad A \equiv \overline{S}$ : motorista alcoolizado
   $F$ : acidante com vítima fatal $\qquad NF \equiv \overline{F}$ : acidente sem vítima fatal

(a) Concluir comparando as proporções (probabilidades) de vítima fatal entre sóbrios e alcoolizados, ou seja as probabilidades condicionais $P[F|S]$ e $P[F|A]$, que são dadas por:

```
     S      A
0.8170 0.7585
```

(b) i. Probabilidades marginais:

$$P[A] = \frac{3155}{4658} = 0.677 \quad ; \quad P[S] = P[\overline{A}] = \frac{1503}{4658} = 0.677$$
$$P[A] = \frac{3621}{4658} = 0.777 \quad ; \quad P[S] = P[\overline{A}] = \frac{1037}{4658} = 0.223$$

ii. Probabilidades conjuntas (interseção):

$$P[A \cap F] = \frac{2393}{4658} = 0.514 \quad ; \quad P[S \cap F] = \frac{1228}{4658} = 0.264$$
$$P[A \cap NF], = \frac{762}{4658} = 0.164 \quad ; \quad P[S \cap NF] = \frac{275}{4658} = 0.059$$

iii. Probabilidades condicionais: $P[A|F], P[S|F], P[F|A], P[NF|A]$

$$P[A|F] = \frac{2393}{3621} = 0.661 \quad ; \quad P[S|F] = P[\overline{A}] = \frac{1228}{3621} = 0.339$$
$$P[A|NF] = \frac{762}{1037} = 0.735 \quad ; \quad P[S|NF] = P[\overline{A}] = \frac{275}{1037} = 0.265$$
$$P[F|A] = \frac{2393}{3155} = 0.758 \quad ; \quad P[NF|A] = P[\overline{A}] = \frac{762}{3155} = 0.242$$
$$P[F|S] = \frac{1228}{1503} = 0.817 \quad ; \quad P[NF|S] = P[\overline{A}] = \frac{275}{1503} = 0.183$$

(c) Pares de eventos mutuamente exclusivos: $F$ e $NF$

- *com vítima fatal* e *sem vítima fatal* ($F$ e $NF$)
- *sóbrio* e *alcoolizado* ($S$ e $A$)

(d) $P[F|A] = \frac{P[F \cap A]}{P[A]} = \frac{2393}{2393+762} = 0.661$

(e) $P[A|F] = \frac{P[A \cap F]}{P[F]} = \frac{2393}{1228+2393} = 0.758$

---

Semana 3

1. Um instituto que faz previsões meteorológicas fez uma avaliação de suas previsões de 48 horas para finais de semana feitas por um particular modelo de previsão. Foi verificado que a previsão indicava chuva em 80% dos dias que de fato choveu e previa não chuva em 92% dos dias em que não choveu. Verificou-se ainda que chove em 12% dos períodos.

   (a) Escreva em notação de probabilidades adequada as informações dadas no problema.
   (b) Qual a probabilidade (proporção) de previsão de chuva?
   (c) Qual a probabilidade de chover quando há uma previsão de chuva?
   (d) Qual a probabilidade de obter uma previsão incorreta?
   (e) Supondo os mesmos percentuais de acertos de previsão, qual seria a probabilidade de acerto de uma previsão de chuva em outra região em que chove em apenas 5% dos períodos?

**Solução:**
Notação:

$A$ : previsão de chuva na região $\qquad$ $\overline{A}$ : previsão de chuva

$C$ : chove no período $\qquad$ $\overline{C}$ : não chove no período

(a)

$$P[C]=0,12 \qquad P[\overline{C}]=0,88$$
$$P[A|C]=0,80 \qquad P[\overline{A}|C]=0,20$$
$$P[\overline{A}|\overline{C}]=0,92 \qquad P[A|\overline{C}]=0,08$$

(b) $P[A]=P[A\cap C]+P[A\cap \overline{C}]=P[A|C]\cdot P[C]+P[A|\overline{C}]\cdot P[\overline{C}]=(0,80)(0,12)+(0,08)(0,88)=0.166$

(c) $P[C|A]=\frac{P[C\cap A]}{P[A]}=\frac{P[A|C]\cdot P[C]}{P[A]}=\frac{(0,80)(0,12)}{0.166}=0.577$

(d) $P[C|\overline{A}]+P[\overline{C}|A]=\frac{P[C\cap\overline{A}]}{P[\overline{A}]}+\frac{P[\overline{C}\cap A]}{P[A]}=\frac{P[\overline{A}|C]\cdot P[C]}{P[\overline{A}]}+\frac{P[A|\overline{C}]\cdot P[\overline{C}]}{P[A]}=\frac{(0,20)(0,12)}{0.834}+\frac{(0,08)(0,88)}{0.166}=0.452$

(e) Nesse caso:

$$P[C]=0,05 \qquad P[\overline{C}]=0,95$$

e recalculando temos:

$$P[A]=P[A\cap C]+P[A\cap \overline{C}]=P[A|C]\cdot P[C]+P[A|\overline{C}]\cdot P[\overline{C}]=(0,80)(0,05)+(0,08)(0,95)=0.116.$$

A probabilidade de acertar uma previsão de chuva fica:

$$P[C|A]=\frac{P[C\cap A]}{P[A]}=\frac{P[A|C]\cdot P[C]}{P[A]}=\frac{(0,80)(0,05)}{0.116}=0.345$$

---

Semana 4

1. Em um teste com quatro questões de múltipla escolha, cada questão possui cinco alternativas com apenas uma delas correta. Estamos interessados na situação de "acerto casual", na qual a resposta de cada questão é escolhida ao acaso, supondo que todas as questões são respondidas.
   (a) Defina a v.a. de interesse.
   (b) Obtenha a função de probabilidades.
   (c) Obtenha a função de distribuição (acumulada).
   (d) Obtenha o valor esperado da v.a.

   Suponha agora que para cada acerto ganha-se dois pontos e perde-se um ponto para cada erro.
   (e) Sob as mesmas condições, obtenha a distribuição de probabilidades do número de pontos ganhos.

**Solução:**

(a) $X$ : número de acertos

(b)

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P[X=x]$ | $(4/5)^4=$ $=0.4096$ | $4(1/5)^1(4/5)^3=$ $=0.4096$ | $6(1/5)^2(4/5)^2=$ $=0.1536$ | $4(1/5)^3(4/5)^1=$ $=0.0256$ | $(1/5)^4=$ $=0.0016$ |

(c)

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $F(x)=P[X\leq x]$ | 0.4096 | 0.8192 | 0.9728 | 0.9984 | 1 |

(d) $\mathrm{E}[X]=\sum_{i=1}^{4}x_iP[X=x_i]=0,8$

(e) $Y$ : pontos obtidos. $Y$ é uma transformação 1-1 de $X$ e portanto $P[Y=y_i]=P[X=x_i]$.

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $y$ | $-4$ | $-1$ | 2 | 5 | 8 |
| $P[Y=y]$ | 0.4096 | 0.4096 | 0.1536 | 0.0256 | 0.0016 |

**Solução alternativa**: usar o fato que $X\sim B(n=4,p=1/5)$

---

2. Seja $X$ uma v.a. com função de distribuição de probabilidades dada por:

$$f(x)=\begin{cases} c\,x & \text{se } 0\leq x\leq 4 \\ 0 & \text{caso contrário}\end{cases}$$

(a) Obtenha o valor de $c$.
(b) Calcule $P[X > 2]$
(c) Calcule $P[1,2 < X < 3,5]$
(d) Calcule $P[X < 3|X > 1,5]$
(e) Obtenha a $E[X]$
(f) Obtenha $k$ tal que $P[X > k] = 0,5$

**Solução:**

(a)

$$\int_0^4 c\,x \ \ \mathrm{d}x = 1$$
$$c = 1/8$$

(b) $P[X > 2] = \int_2^4 x/8 \ \ \mathrm{d}x = 0,75$
(c) Calcule $P[1,2 < X < 3,5] \int_2^4 x/8 \ \ \mathrm{d}x = 0.6756$
(d) Calcule $P[X < 3|X > 1,5] = \frac{P[1,5<X<3]}{P[X>1,5]} = 0.4909$
(e) $E[X] = \int_0^4 x \cdot x/8 \ \ \mathrm{d}x = (1/8)4^3/3 = 8/3 = 2,67$
(f)

$$\int_0^k x/8 \ \ \mathrm{d}x = 0.5$$
$$(1/8)k^2/2 = 0,5$$
$$k = \sqrt{8}$$

---

Semana 5

1. Suponha que a posição na qual ocorre um dano em um disco pode ser considerada uma variável uniforme com valores entre 0 e 100 (valores percentuais em relação a capacidade total do disco). Define-se ainda que a porção inicial do disco corresponde aos 15% iniciais, a porção intermediária entre 15 e 80% e a porção nos restantes 20%.
   (a) Defina a v.a. de interesse, a função de densidade de probabilidades e a função de distribuição (acumulada).
   (b) Obtenha a probabilidade do dano ocorrer na porção inicial do disco.
   (c) Obtenha a probabilidade do dano ocorrer na porção final do disco, sabendo que não ocorreu na inicial.
   (d) Obtenha a probabilidade do dano ocorrer na porção inicial ou final do disco.
   (e) Defina uma nova variável dada pela porção do disco onde ocorre a falha e monte a sua distribuição de probabilidades.

   Considere agora que serão examinados cinco discos com dano e estamos interessados no número de discos que apresenta dano na porção inicial.

   (f) Defina a v.a. de interesse, identifique o seu *tipo* e obtenha a distribuição de probabilidades.
   (g) Calcule a probabilidade de obter no máximo um dos discos com dano na porção inicial.
   (h) Se forem examinados 100 lotes de cinco discos, qual deve ser o número médio de discos por lote SEM dano na porção inicial?

**Solução:**

(a)

$$X: \text{número de acertos}$$
$$f(x) = \frac{1}{100-0} = \frac{1}{100}$$
$$F(x) = \int_0^{15} f(x)\,\mathrm{d}x = \frac{1}{100}x\Big|_0^x = \frac{x-0}{100-0} = \frac{x}{100}$$

(b)

$$P[0 < X < 15] = \int_0^{15} f(x)\,\mathrm{d}x = \frac{1}{100}x\Big|_0^{15} = \frac{1}{100}(15-0) = 0,15$$

ou

$$P[0 < X < 15] = F(15) = \frac{15}{100} = 0,15$$

(c)

$$P[80 < X < 100|X > 15] = \frac{P[80 < X < 100 \cap X > 15]}{P[X > 15]} = \frac{P[80 < X < 100]}{P[X > 15]} = \frac{\int_{80}^{100} f(x)\,\mathrm{d}x}{\int_{15}^{100} f(x)\,\mathrm{d}x} = \frac{20/100}{85/100} = 0.235$$

ou

$$P[80 < X < 100|X > 15] = \frac{1-F(80)}{1-F(15)} = 0.235$$

(d)

$$P[X < 15 \cup X > 80] = P[X < 15] + P[X > 80] = \frac{15}{100} + \frac{20}{100} = 0,35$$

ou

$$P[X < 15 \cup X > 80] = F(15) + (1 - F(80)) = \frac{15}{100} + (1 - \frac{80}{100}) = 0,35$$

(e)

$Y$ : porção do disco onde ocorre dano

| y | inicial | intermediária | final |
|---|---|---|---|
| P[Y=y] | 0,15 | 0,65 | 0,20 |

(f)

$$Z: \text{ número de discos com dano entre cinco discos na porção inicial}$$
$$Z \sim B(n = 5, p = 0,15)$$
$$P[Z = z] = \binom{5}{z}(0,15)^z\,(1-0,15)^{5-z}$$

(g) $P[Z \leq 1] = P[Z = 0] + P[Z = 1] = \binom{5}{0}(0,15)^0\,(1-0,15)^{5-0} + \binom{5}{1}(0,15)^1\,(1-0,15)^{5-1} = 0.444 + 0.392 = 0.835$

(h) $\mathrm{E}[5 - X] = 5 - \mathrm{E}[X] = 5 - n \cdot p = 5 - 5 \cdot (0,15) = 4,25$

---

Semana 6

1. Considere que um grande grupo (I) de pessoas vai fazer uma determinada prova. Supõe-se que as *notas* seguem uma distribuição normal de média 450 e variância 144. Chama-se de *escores* ($Z = (X - \mu)/\sigma$ as notas padronizadas para uma distribuição normal padrão (média zero e variância unitária)
   (a) Qual a proporção esperada de candidatos com nota superior a 470?

(b) Qual a proporção esperada de candidados com notas entre 430 e 460?

(c) Qual a proporção esperada de candidados que se distanciem mais do que 1,5 desvios padrões da média?

(d) Se forem classificados para uma próxima etapa 20% dos candidatos com as maiores notas, qual será a nota de corte para classificação para próxima etapa.

(e) Se dividirmos os candidatos em três faixas: $A$ : os 60% com menores notas, $B$ : 30% de notas intermediárias e $C$ : 10% com as maiores notas. Quais os escores que definem os grupos?

(f) Quais valores correspondem aos quartis da distribuição das notas?

(g) São considerados candidatos excepcionais aqueles com escore acima de 2,5. A qual nota corresponde tal escore?

Um outro grupo (II) fez uma prova equivalente em um outro dia. Para este outro grupo a média foi de 462 e a variância de 81. Os candidados dos dois grupos serão classificados usando os escores.

(h) Entre um candidato do grupo I com nota 470 e um do grupo II com nota 485, qual estaria melhor classificado?

(i) Quais seriam as notas para classificar os candidatos do grupo II nas faixas $A$, $B$ e $C$?

(j) Se fosse adotada uma única nota de corte de 475, qual seria o percentual de classificados de cada grupo?

**Solução:**

$$\begin{aligned} X &: \text{nota no grupo I} & X &\sim \mathrm{N}(450, 12) \\ Z = \frac{X - 450}{12} &: \text{escore} & Z &\sim \mathrm{N}(0, 1) \\ Y &: \text{nota no grupo II} & Y &\sim \mathrm{N}(462, 9) \end{aligned}$$

(a) $P[X > 470] = P[Z > \frac{470-450}{12}] = P[Z > 1.667] = 0.0478$

(b) $P[430 < X < 460] = P[\frac{430-450}{12} < Z < \frac{460-450}{12}] = P[-1.667 < Z < 0.8333] = 0.75$

(c) $P[|Z| > 1,5] = P[Z < -1,5] + P[Z > 1,5] = 0.134$

(d)

$$\begin{aligned} P[X > x_D] &= 0,20 \\ z &= 0.8416 \\ z &= \frac{x_D - 450}{12} \\ x_D &= 460.1 \end{aligned}$$

(e)

$$\begin{aligned} P[Z < z_1] = 0,60 &\longrightarrow z_1 = 0.2533 \quad (x_1 = 453) \\ P[Z < z_2] = 0,90 &\longrightarrow 1.282 \quad (x_2 = 465.4) \end{aligned}$$

(f)

$$
\begin{aligned}
Q_1 &: \\
P[X < Q_1] &= 0,25 \\
z_1 &= -0.6745 \\
z_1 &= \frac{Q_1 - 450}{12} \\
Q_1 &= 441.9 \\
Q_2 &: \\
P[X < Q_2] &= 0,50 \\
Q_2 &= 450 \\
Q_3 &: \\
P[X < Q_3] &= 0,75 \\
z_3 &= 0.6745 \\
z_3 &= \frac{Q_3 - 450}{12} \\
Q_3 &= 458.1
\end{aligned}
$$

(g) $z = \frac{x_G - 450}{12} = 2,5 \longrightarrow x_G = 480$

(h)

$$
\begin{aligned}
z_I &= \frac{470 - 450}{12} = 1.67 \\
z_{II} &= \frac{482 - 462}{9} = 2.56
\end{aligned}
$$

(i)

$$
\begin{aligned}
&P[Z < z_1] = 0,60 \longrightarrow z_1 = 0.2533 \quad (y_1 = 464.3) \\
&P[Z < z_2] = 0,90 \longrightarrow 1.282 \quad (y_2 = 473.5)
\end{aligned}
$$

(j)

$$
\begin{aligned}
\text{Grupo I} : P[X > 475] &= P[Z > \frac{475 - 450}{12}] = P[X > 2.083] = 0.0186 \\
\text{Grupo II} : P[X > 475] &= P[Z > \frac{475 - 462}{9}] = P[X > 1.444] = 0.0743
\end{aligned}
$$

**Solução computacional** (com o R)

```
> (qA <- pnorm(470, m=450, sd=12, low=F))
[1] 0.04779
> (qB <- diff(pnorm(c(430,460), m=450, sd=12)))
[1] 0.7499
> (qC <- 2*pnorm(-1.5))
[1] 0.1336
> (qD <- qnorm(0.80, m=450, sd=12))
[1] 460.1
> (qE <- qnorm(c(0.60,0.90)))
[1] 0.2533 1.2816
> (qEn <- qnorm(c(0.60,0.90), m=450, sd=12))
[1] 453.0 465.4
```

```
> (qF <- qnorm(c(0.25,0.50,0.75), m=450, sd=12))
[1] 441.9 450.0 458.1
> (qG <- 450 + 2.5 * 12)
[1] 480
> (qH <- c((470-450)/12,(485-462)/9))
[1] 1.667 2.556
> (qI <- qnorm(c(0.60,0.90), m=462, sd=9))
[1] 464.3 473.5
> (qJ <- c(pnorm(475, m=450, sd=12, low=F), pnorm(475, m=462, sd=9, low=F)))
[1] 0.01861 0.07431
```

---

Semana 7

1. Suponha que um sistema apresenta uma taxa de 3,2 falhas a cada 1000 pacotes transmitidos.
   (a) Qual a probabilidade de haver duas ou mais falhas na transmissão de 100 pacotes?
   (b) Sabendo que houveram falhas na transmissão de 1000 pacotes, qual a probabilidade que tenha sido mais do que uma?
   (c) Qual a probabilidade de não haver falhas na transmissão de 500 pacotes?
   (d) Qual a probabilidade de haver 6 falhas na transmissão de 2000 pacotes?
   (e) Qual deveria ser a taxa de falhas para que a probabilidade de não haver falha de transmissão em 1000 pacotes seja de pelo menos 0,90?

**Solução:**

$$X : \text{número de falhas a cada 1000 pacotes}$$
$$X \sim \mathrm{P}(\lambda = 3,2)$$
$$P[X = x] = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}$$

(a) $P[X \geq 2] = 1 - P[X = 0] - P[X = 1] = 1 - \frac{\mathrm{e}^{-3,2}3,2^0}{0!} - \frac{\mathrm{e}^{-3,2}3,2^1}{1!} = 1 - \mathrm{e}^{-3,2}(1 + 3,2) = 0.829$

(b) $P[X > 1|X \geq 1] = \frac{P[X>1 \cap X \geq 1]}{P[X \geq 1]} = \frac{P[X>1]}{P[X \geq 1]} = \frac{1-P[X=0]-P[X=1]}{1-P[X=1]} = \frac{1-0.171}{1-0.0408} = 0.864$

(c)

$$X_c : \text{número de falhas a cada 500 pacotes}$$
$$X_c \sim \mathrm{P}(\lambda_c = 1,6)$$
$$P[X = 0] = \frac{\mathrm{e}^{-1,6}1,6^0}{0!} = 0.202$$

(d)

$$X_d : \text{número de falhas a cada 2000 pacotes}$$
$$X_d \sim \mathrm{P}(\lambda_d = 6,4)$$
$$P[X = 6] = \frac{\mathrm{e}^{-6,4}6,4^6}{6!} = 0.159$$

(e)

$$\begin{aligned}
X_e &: \text{número de falhas a cada 1000 pacotes} \\
X_e &\sim \mathrm{P}(\lambda_e) \\
P[X = 0] &\geq 0,90 \\
\frac{\mathrm{e}^{-\lambda_e}\lambda_e^0}{0!} &\geq 0,90 \\
\mathrm{e}^{-\lambda_e} &\geq 0,90 \\
\lambda_e &\leq -\log(0,90) = 0.105
\end{aligned}$$

**Solução computacional** (com o R)

```
> (qA <- ppois(1, lam=3.2, low=F))
[1] 0.8288
> (qB <- ppois(1, lam=3.2, low=F)/ppois(0, lam=3.2, low=F))
[1] 0.864
> (qC <- dpois(0, lam=1.6))
[1] 0.2019
> (qD <- dpois(6, lam=6.4))
[1] 0.1586
> (qE <- -log(0.90))
[1] 0.1054
```

---

2. Um *cluster* possui 30 nós para processamento, sendo 10 de alta velocidade e 20 de baixa velocidade. A cada processo submetido são alocados aleatóriamente o número de nós solicitados. Se um processo solicita cinco nós (simultaneamente) qual a probabilidade de que:
   (a) todos sejam alocados em nós de baixa velocidade,
   (b) dois deles sejam alocados em nós de alta velocidade,
   (c) pelo menos dois deles sejam alocados em nós de alta velocidade,
   (d) sejam todos alocados em nós do mesmo tipo (alta ou baixa velocidade).
   (e) Se os processos nós fossem solicitados sequencialmente (cada um após finalizar o processamento do anterior) a probabilidade de alocação em dois nós de alta velocidade seria diferente? (Justifique e recalcule, se for o caso)

**Solução:**

$$\begin{aligned}
X &: \text{número de alocações em nós de alta velocidade} \\
X &\sim \mathrm{HG}(N = 30, n = 5, r = 10) \\
P[X = x] &= \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{\binom{10}{x}\binom{30-10}{5-x}}{\binom{30}{5}}
\end{aligned}$$

(a) $P[X = 0] = \frac{\binom{10}{0}\binom{30-10}{5-0}}{\binom{30}{5}} = 0.109$

(b) $P[X = 2] = \frac{\binom{10}{2}\binom{30-10}{5-2}}{\binom{30}{5}} = 0.36$

(c) $P[X \geq 2] = P[X = 2] + P[X = 3] + P[X = 4] + P[X = 5] = 1 - P[X \leq 2] = 1 - P[X = 0] - P[X = 1] = 1 - \frac{\binom{10}{0}\binom{30-10}{5-0}}{\binom{30}{5}} - \frac{\binom{10}{1}\binom{30-10}{5-1}}{\binom{30}{5}} = 1 - 0.109 - 0.34 = 0.551$

(d) $P[X=0] + P[X=5] = \frac{\binom{10}{0}\binom{30-10}{5-0}}{\binom{30}{5}} + \frac{\binom{10}{5}\binom{30-10}{5-5}}{\binom{30}{5}} 0.109 - 0.00177 = 0.111$

(e)

$$X_e \sim \mathrm{B}(n=5, p=1/3)$$
$$P[X=2] = \binom{5}{2}(1/3)^2(1-1/3)^{5-2} = 0.329$$

**Solução computacional** (com o R)

```
> (qA <- dhyper(0, m=10, n=20, k=5))
[1] 0.1088
> (qB <- dhyper(2, m=10, n=20, k=5))
[1] 0.36
> (qC <- phyper(1, m=10, n=20, k=5, low=F))
[1] 0.5512
> (qD <- sum(dhyper(c(0,5), m=10, n=20, k=5)))
[1] 0.1106
> (qE <- dbinom(2, size=5, prob=1/3))
[1] 0.3292
```

---

Semana 8

1. Um processamento de um determinado tipo de requisição é feito regularmente com o tempo médio de 0,25 seg e desvio padrão de 4 seg. Há duas propostas para melhoria do serviço. A primeira reduz o tempo da cada requisição em 0,02 segundos. A segunda reduz em 8% o tempo de processamento de cada requisição. Compare as médias, desvios padrões de coeficientes de variação de cada proposta. Discuta os resultados e indique se há alguma diferença entre elas qual deveria ser adotada (ou as características de cada proposta).

   **Solução:**

---

Semana 9

1. Considere o seguinte problema (Magalhães & Lima, 2006):
   *Um fabricante afirma que sua vacina contra gripe imuniza em 80% dos casos. Uma amostra de 25 indivíduos entre os que tomaram a vacina foi sorteada e testes foram feitos para verificar a imunização ou não desses indivíduos. Se o fabricante estiver corrreto, qual é a probabilidade de proporção de imunizados ser inferior a 0,75? E superior a 0,85?*
   (a) No contexto do problema identifique:
      - a população
      - o parâmetro de interesse
      - o estimador
      - a estimativa
      - a distribuição amostral
   (b) Responda as perguntas propostas no problema
   (c) Quais seriam as estimativas pontuais e intervalares se fosse observados 18 imunizados dentre os 25 avaliados?
   (d) Qual deveria ser o tamanho da amostra em um novo estudo para que a margem de erro fosse de no máximo 0,03?

(e) Suponha que se adote o critério de refutar a afirmativa do fabricante caso sejam observados 17 ou menos não imunizados. Qual a probabilidade de refutar a afirmativa mesmo quando ela é verdadeira (a vacina de fato imuniza 80%)

(f) Suponha agora que a imunização real seja de apenas 70%. Qual a probabilidade de mesmo assim não refutar a afirmativa do fabricante?

**Solução:**

(a)
- Os indivíduos que receberam a vacina.
- A proporção ($p$) de indivíduos imunizados entre todos os que receberam a vacina (*na população*).
- O cálculo da proporção de indivíduos imunizados na amostra $\hat{p} = \sum_{i=1}^{n} X_i/n$.
- A proporção observada em uma determinada amostra (no caso na amostra de $n = 25$ indivíduos).
- A distribuição amostral do estimador, ou seja, a distribuição que seria obtida caso fossem obtidas estimativas de *diversas* amostras.

(b)

(c)
- Usando $p = 0,80$

$$\hat{p} = \frac{18}{25} = 0.72$$
$$(\text{I.C.95\%})\hat{p} \pm z\frac{p(1-p)}{n}$$
$$0.72 \pm 1.96\frac{0,80(1-0,80)}{25}$$
$$0.72 \pm 0.0125$$
$$(0.563 \;\; ; \;\; 0.877)$$

- 

$$\hat{p} = \frac{18}{25} = 0.72$$
$$(\text{I.C.95\%}) : \hat{p} \pm z\sqrt{\frac{p(1-p)}{n}}$$
$$0.72 \pm 1.96\sqrt{\frac{0.72(1-0.72)}{25}}$$
$$0.72 \pm 0.176$$
$$(0.544 \;\; ; \;\; 0.896)$$

(d)

$$z\sqrt{\frac{p(1-p)}{n}} = 0,03$$
$$1.96\sqrt{\frac{0,80(1-0,80)}{n}} = 0,03$$
$$n = \frac{1.96^2}{0,03^2}0,80(1-0,80) = 683$$

(e) $P[X \leq 17|p = 0,80] \approx P[\hat{p} \leq 17,5/25] = 0$

(f) $P[X \geq 18|p = 0,70] \approx P[\hat{p} > 17,5/25] = 0$

---

Semana 10

1. Imagens de satélite precisam ser (pós-)processadas para um certo objetivo. O tempo médio de processamento é de 96 min com desvio padrão de 18 minutos. Uma cena de uma determinada região é montada com o processamento de 12 imagens acrescido de 8 minutos para montagem da cena.

(a) Se as imagens forem processadas sequencialmente, qual a probabilidade do tempo de processamento ultrapassar 20 horas?

(b) Se as imagens forem processadas paralelamente, qual a probabilidade do tempo médio de processamento ficar entre 92 e 98 min?

(c) Quando o processamento é feito em paralelo a montagem só é feita após todos os processamentos estarem encerrados. Qual a probabilidade de ser necessário esperar mais que 100 minutos para iniciar a montagem?

**Solução:**
Notação:

$$X : \text{tempo de processamento}$$
$$X \text{ Distribuição}(\mu = \mathrm{E}[X] = 96, \sigma^2 = \mathrm{Var}[X] = 18^2)\overline{X} \qquad \approx \mathrm{N}(\mu = \mathrm{E}[X] = 96, \sigma^2/n = 18^2/12)$$

(a) $P[\sum_{i=1}^{n} X_i > 20 \cdot 60] = P[\overline{X} > 100] = P[Z > \frac{100-96}{\sqrt{18^2/12}}] = P[Z > 0.77 = 0.221$

(b) $P[92 < \overline{X} < 98] = P[\frac{92-96}{\sqrt{18^2/12}} < Z < \frac{98-96}{\sqrt{18^2/12}}] = P[-0.77 < Z < 0.385] = 0.429$

(c) $P[(X_1 > 100) \cup (X_2 > 100) \ldots (X_{12} > 100)] = 1 - P[(X_1 < 100) \cap (X_2 < 100) \ldots (X_{12} < 100)] = 1 - (P[(X < 100)])^{12} = 1 - (P[(Z < 0.77)])^{12} = 1 - (0.588)^{12} = 0.998$

---

2. No contexto anterior foi proposto um novo algorítmo para o processamento. O algorítmo foi avaliado processando-se 15 imagens. Os tempo obtidos foram:

```
102,4 104,2 102,8  67,6 106,2  88,7  85,3  82,1 102,8 102,5  99,7  85,2 85,8 120,0 117,5.
```

(a) Monte um IC (95%) para a nova média.

(b) Monte um IC (95%) para o novo desvio padrão.

(c) Teste a hipótese ($\alpha = 0,05$) que houve redução no tempo médio de processamento.

(Faça e declare as suposições necessárias)

**Solução:**
Notação:

$$Y : \text{novo tempo de processamento}$$
$$X \sim \text{Distribuição}(\mu_Y, \sigma_Y^2)$$
$$\text{Estimativas com a amostra} : \overline{y} = 96.9 \quad ; \quad S^2 = 198 \quad ; \quad S = 14.1 \quad ;$$

(a)
- Supondo $\sigma_Y^2 = \sigma^2 = 18^2$ (variância do novo algorítmo igual a do anterior)

$$\overline{x} \pm z \frac{\sigma}{\sqrt{n}}$$
$$96.9 \pm 1.96 \frac{18}{\sqrt{15}}$$
$$96.9 \pm 9.1$$
$$(87.7 \;\; ; \;\; 106)$$

- Supondo $\sigma_Y^2$ estimado por $S^2$ (variância desconhecida do novo algorítmo, estimada pela variância amostral)

$$\overline{x} \pm t_{\nu=14} \frac{S}{\sqrt{n}}$$
$$96.9 \pm 2.14 \frac{14}{\sqrt{15}}$$
$$96.9 \pm 7.8$$
$$(89.1 \;\; ; \;\; 105)$$

(b)

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\text{sup}}}} \;\; ; \;\; \sqrt{\frac{(n-1)S^2}{\chi^2_{\text{inf}}}}\right)$$
$$\left(\sqrt{\frac{(15-1)198}{26.1}} \;\; ; \;\; \sqrt{\frac{(15-1)198}{5.63}}\right)$$
$$(10.3 \;\; ; \;\; 22.2)$$

(c) Teste de hipótese:

$$H_0 : \mu = 96 \quad H_a : \mu < 96$$
$$\alpha = 0,05$$

- Supondo $\sigma_Y^2 = \sigma^2 = 18^2$ (variância do novo algorítmo igual a do anterior)

$$\begin{aligned} z_c &= \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} \\ &= \frac{96.9 - 96}{18/\sqrt{15}} \\ &= 0.184 \\ \text{Região Crítica (RC)} &: Z < -1.64 \\ \text{p} - \text{valor} &= P[Z < z_c] = 0.573 \end{aligned}$$

- Supondo $\sigma_Y^2$ estimado por $S^2$ (variância desconhecida do novo algorítmo, estimada pela variância amostral)

$$\begin{aligned} t_c &= \frac{\overline{x} - \mu_0}{S/\sqrt{n}} \\ &= \frac{96.9 - 96}{14.1/\sqrt{15}} \\ &= 0.235 \\ \text{Região Crítica (RC)} &: Z < -1.76 \\ \text{p} - \text{valor} &= P[t < t_c] = 0.591 \end{aligned}$$

**Conclusão:**
$z_c$ (ou $t_c$) $\notin$ RC,
não rejeitamos $H_0$, ou seja, não há evidências que o novo algoritmo reduza o tempo de processamento.