Probabilidade e estatística

Estatística descritiva

• Frequência

- Freq. absoluta → nº de ocorrências de um valor.
- Freq. relativa → Percentual de ocorrências de um valo
- Freq. absoluta acumulada → nº de ocorrências até o valor.
- Freq. relativa acumulada → Percentual de ocorrências até o va

• Média aritmética

→ Soma de todas as observações dividida pelo número de elementos do conjunto de dados.

$$\mu = \frac{\sum_i x_i}{N}$$

• Média Ponderada

→ Soma de todas as observações multiplicadas pela sua frequência e dividida pelo total de dados.

$$\mu = \frac{\sum_i (x_i \times f_i)}{N}$$

• Mediana

→ É o valor que divide o conjunto de dados em duas partes com igual nº de observações.

• Moda

→ É o valor mais frequente de um conjunto de dados.

Digitalizado com CamScanner

data
fecha

D S T Q Q S S
D L M M J V S

- Separatrizes
→ São valores que separam o conjunto de dados ordenado em partes com igual nº de observações
- mediana → divide ao meio.
- Quartis → divide o conjunto de dados em 4 partes iguais
- Percentis → " " em 100 partes iguais

- Amplitude
→ É a diferença entre o maior e o menor valor do conjunto de dados.

- Variância (σ^2, s^2)
→ É uma medida da variação em torno da média. A variância é a média dos quadrados dos desvios em torno da média.

x na posição i - média

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

* Se a tabela disponível for com frequências a fórmula fica:

$$\sigma^2 = \frac{\sum (x_i - \mu)^2 \times f_i}{N}$$

- Desvio-Padrão (σ, s)
→ Expressa a variação média do conjunto de dados em torno da média.

$$\sigma = \sqrt{\sigma^2}$$

- Coeficiente de variação
→ É a razão entre o desvio padrão e a média de um conjunto de dados. Expressa a variação relativa.

$$CV = \frac{\sigma}{\mu} \times 100\%$$

Digitalizado com CamScanner

Probabilidade

→ É o campo da matemática que estuda experimentos e através dele é possível analisar as chances de determinado evento ocorrer

$P(A) + P(\bar{A}) = 1$ → Probabilidade de não ocorrer A.

$0 \leq P(A) \leq 1$

- Cálculo básico

$$P(A) = \frac{n(A)}{n(\Omega)}$$

$n(A)$ → números de casos favoráveis

$n(\Omega)$ → números de casos possíveis

- Probabilidade condicional

→ É a probabilidade de ocorrência de um evento condicionada a ocorrência de outro evento

→ Probabilidade de A dado B

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Soma de eventos (união)

→ Probabilidade onde evento A <u>ou</u> evento B ocorre ou seja, pelo menos um ocorre. Pode ocorrer ambos.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

* Em eventos mutuamente exclusivos, quando os eventos não podem ocorrer ao mesmo tempo.

$P = (A \cup B) = P(A) + P(B)$, onde $P(A \cap B) = 0$

- Produto de eventos (interseção)
- independentes

→ Quando a probabilidade $P(A)$ é independente de $P(B)$.

$$P(A \cap B) = P(A) \cdot P(B)$$

Digitalizado com CamScanner

data
fecha

D S T Q Q S S
D L M M J V S

• Dependentes

→ Ocorre quando a Probabilidade de A ocorrer influencia a Probabilidade de B ocorrer.

→ Prob. condicional

$P(A \cap B) = P(A) \cdot P(B|A)$

— // — // — // — // —

Matemática (Resumo)

• Permutação

→ de n elementos, sem repetição:

$P_n = n!$ ANAGRAMA PRATO → 5!

→ de n elementos com repetição

$P_n^{a,b...} = \dfrac{n!}{a!\,b!...}$ ANAGRAMA ARANHA → $P_6^3 = 6!/3!$

// BANANA → $P_6^{3,2} = 6!/3!\cdot 2!$

• Combinação

→ são subconjuntos onde a ordem dos elementos não é importante.

$C_{n,p} = \dfrac{n!}{p!(n-p)!}$

— // — // — // — // —

Exemplo: Em uma fábrica de parafusos as máquinas A, B, C produzem 25%, 35% e 40% do total produzido. Da produção de cada máquina 5%, 4% e 2% são defeituosos. Se escolhermos um parafuso aleatório qual a probabilidade de ser defeituoso?

Parafuso ruim e Máquina A ou Parafuso ruim e máquina B ou ...

D = defeituoso

$X = (D_A \cap A) \cup (D_B \cap B) \cup (D_C \cap C)$

$P(X) = P(D_A \cap A) + P(D_B \cap B) + P(D_C \cap C)$

$P(X) = P(D_A) \cdot P(A|D_A) + P(D_B) \cdot P(B|D_B) + P(D_C) \cdot (C|D_C)$

$P(x) = 0{,}25 \cdot 0{,}05 + 0{,}35 \cdot 0{,}04 + 0{,}40 \cdot 0{,}02$

$P(x) = 0{,}0345$

→ 3,45% dos parafusos são defeituosos

Digitalizado com CamScanner

P. de ser de A, dado que é defeituoso

$P(A|D) = ?$

data
fecha

Escolhendo-se ao acaso um parafuso e verifica-se que ele é defeituoso, qual a probabilidade que seja da maquina A, B, C?

$$P(A|D) = \frac{\overbrace{P(A \cap D)}^{P(A).P(D|A)}}{P(D)} \therefore P(x)$$

$$P(A|D) = \frac{0,25 . 0,05}{0,0345} = \boxed{0,3623}$$

$$P(C|D) = \frac{0,40 . 0,02}{0,0345} = 0,2319$$

$$P(B|D) = \frac{0,35 . 0,04}{0,0345} = 0,4058$$

Em uma comunidade 6% dos adultos tem diabetes um novo teste diagnostica corretamente 84% das pessoas que tem diabetes e 98% das pessoas que não tem. Qual a probabilidade de um pessoa diagnosticada com diabetes pelo teste, ter a doença?

D = Pessoa Tem diabetes
T = Teste de diabetes positivo

Temos que:
$P(T|D) = 84\%$
$P(D) = 6\%$

diabetes dado teste certo

$P(D|T) = ?$

teste ser positivo

$$P(T) = P(T|D) \cdot P(D) + \underbrace{P(T|\bar{D})}_{1-98\% = 2\%} . \underbrace{P(\bar{D})}_{1-6\% = 94\%}$$

$$P(T) = (0,84 \cdot 0,06) + (0,02 \cdot 0,94) = 0,0692$$

$$P(D|T) = \frac{P(D \cap T)}{P(T)} \therefore \frac{(0,84 . 0,06)}{0,0692} = 0,7283$$

Qual a probabilidade de uma pessoa que faça o teste seja diagnosticada como não diabética?

$$P(\bar{T}) = P(\bar{T}|\bar{D}) . P(\bar{D}) + P(\bar{T}|D) . P(D)$$

$$P(\bar{T}) = (0,98) . (0,94) + (0,16) . (0,06) \therefore$$

$$P(\bar{T}) = 0,9308$$

· Distribuição binomial
Ex: Uma caixa tem 6 lâmpadas e cada uma delas tem 5% de probabilidade de ter defeito.
Qual a probabilidade que na caixa tenha 2 lâmpadas com defeito. $K=2$
Temos que, defeito $=0,05$ e funcionando $=0,95$
dentro da caixa com 6, queremos 2 com defeito, a ordem não importa.

$P(X=K)=C_n^K \cdot p^K \cdot q^{n-K}$, onde P é o que buscamos

$P(x=2)=C_6^2 \cdot (0,05)^2 \cdot (0,95)^4$
$P(x=2)=15 \cdot 0,0025 \cdot 0,8145$
$P(x=2)=0,03054 \rightarrow 3,054\%$

- Condições para que seja uma distribuição binomial.
1º - deve ter apenas 2 possibilidades
2º - O conjunto deve ser um número finito
3º - A probabilidade deve ser igual a todos os itens
4º - Os eventos devem ser independentes.

Lista distribuição Binomial

1) $P_{sem\ defeito}=0,9$

$P_{defeituoso}=\frac{80}{800} \therefore 0,1$

$P_{ser\ aceito}=P(x=0)+P(x=1)$

$n=20$

$P(x=0)=(0,9)^{20} \cdot C_{20}^0 \therefore 0,1216$

$P(x=1)=C_{20}^1 \cdot (0,1)^1 \cdot (0,9)^{19} \therefore 0,2707$

$P(x\leq 1)=0,1216+0,2707$
$P(x\leq 1)=0,3923 \therefore 39,23\%$

data
fecha
D S T Q Q S S
D L M M J V S

②

P defeito = 0,1

P sem defeito = 0,9

n = 20

P(X=3) = ?

$C_{20}^{3} = \dfrac{20!}{3!\,17!} \therefore \dfrac{20 \cdot 19 \cdot 18}{3 \cdot 2} = 1140$

$P(X=3) = C_{20}^{3} \cdot (0,1)^3 \cdot (0,9)^{17}$

$P(X=3) = 1140 \cdot 0,001 \cdot 0,1668$

$P(X=3) = 0,1901 \therefore 19,01\%$

③

P sz responde = 25% ∴ 0,25

P não sz = 75% ∴ 0,75

n = 15

$P(X=0) = C_{15}^{0} \cdot (0,75)^{15} \cdot (0,25)^{0}$

$P(X=0) = 0,0133$

$P(X=1) = C_{15}^{1} \cdot (0,75)^{14} \cdot (0,25)^{1}$

$P(X=1) = 15 \cdot 0,0178 \cdot 0,25$

$P(X=1) = 0,06675$

a) $1 - (P(X=0) + P(X=1))$

$P(X \geq 2) = 1 - (0,0133 + 0,06675)$

$P(X \geq 2) = 0,91995 \rightarrow 91,995\%$

$P(X=2) = C_{15}^{2} \cdot (0,75)^{13} \cdot (0,25)^{2}$

$105 \cdot 0,0237 \cdot 0,0625$

$P(X=2) = 0,1555$

b) $P(X \geq 3) = 1 - (P(X=0) + P(X=1) + P(X=2))$

$P(X \geq 3) = 1 - (0,0133 + 0,06675 + 0,1555)$

$P(X \geq 3) = 0,76445 \rightarrow 76,44\%$

④

P alergia = 20% ∴ 0,2

P sem alergia = 80% ∴ 0,8

n = 13

$P(X \geq 4) = 1 - (P(X=0) + P(X=1) + P(X=2) + P(X=3))$

$P(X \geq 4) = 1 - (C_{13}^{0} \cdot (0,2)^0 (0,8)^{13} + C_{13}^{1} \cdot (0,2)^1 (0,8)^{12} + C_{13}^{2} \cdot (0,2)^2 (0,8)^{11} + C_{13}^{3} \ldots$

$P(X \geq 4) = 1 - (1 \cdot 0,0549 + 13 \cdot 0,0687 \cdot 0,2 + 78 \cdot 0,04 \cdot 0,0859 + 286 \cdot$

$0,008 \cdot 0,1074$

$P(X \geq 4) = 1 - (0,0549 + 0,1786 + 0,2680 + 0,2457)$

$P(X \geq 4) = 1 - 0,7472$

$P(X \geq 4) = 0,2528 \rightarrow 25,28\%$

Digitalizado com CamScanner

data
fecha
D S T Q Q S S
D L M M J V S

• Distribuição Normal

↳ definição: Seja X uma variável aleatória contínua tal que X tem distribuição normal com média μ e variância σ^2 se, e somente se sua função densidade de probabilidade for dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Distribuição Normal Padrão
  - ↳ média ∴ $\mu = 0$
  - ↳ desvio padrão ∴ $\sigma = 1$

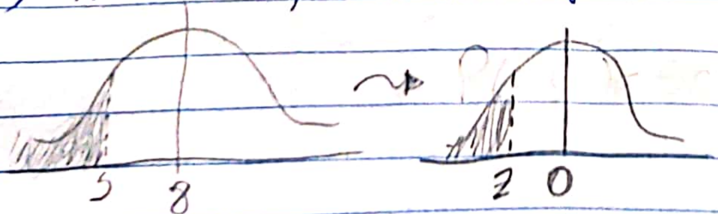
• Com essa padronização, conseguimos distribuir em uma tabela, mapeando a distribuição normal, transformando a variável X em uma variável Z. de forma que a área (probabilidade) de 0 a $+\infty$ é 50%, 0,5.

agora com a mudança de variável conseguimos descobrir o valor de Z e com o auxílio da tabela conseguimos a probabilidade

$$Z = \frac{X - \mu}{\sigma}$$

Ex: Suponha que o tempo necessário para atendimento em uma fila de banco seja normalmente distribuída com $\mu = 8$ min e $\sigma = 2$ min. Qual a probabilidade que o atendimento dure: $X \sim N(8, 2)$

a) menos que 5 min?



↗ Tabela

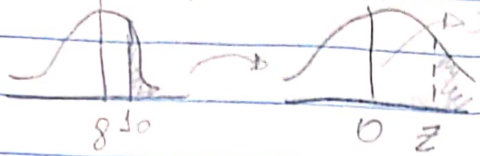
$Z = \frac{5-8}{2}$ ∴ $Z = -1{,}50$

P z até 0 é 42,33%

↳ $P(X<5) = 50\% - P(5 \leq X \leq 8)$

$P(X<5) = 7{,}67\%$

b) mais que 10 min



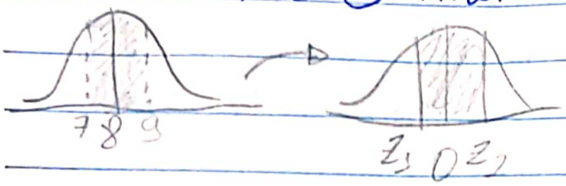
$z = \frac{10-8}{2} \therefore z = 1,00$

$P(z) = 34,13\%$

34,13

$P(X>10) = 50\% - P(z) \therefore P(X>10) = 15,87\%$

c) Entre 7 e 9 min



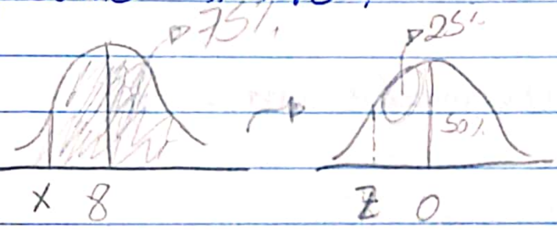
$Z_1 = \frac{7-8}{2} \therefore Z_1 = -0,50$

$Z_2 = \frac{9-8}{2} \therefore Z_2 = 0,50$

$P(0,50) = 19,15\%$

$P(7 \leq X \leq 9) = P(Z_1) + P(Z_2) \therefore P(7 \leq X \leq 9) = 38,3\%$

d) 75% dos atendimentos requerem, no mínimo, quanto tempo?



$P(z) = 25\% \therefore 0,2500$

↳ $z = 0,67$

$z = \frac{x-8}{2} \therefore -0,67 \cdot 2 = x - 8$

$x = 6,66$

- Distribuição de Poisson

→ calcula a probabilidade de alguma coisa acontecer dentro de um intervalo

$$P(x) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!}$$

→ número médio de sucesso dentro do intervalo

λ → número real que corresponde ao número de ocorrências que se espera dentro do intervalo.

x → número inteiro que corresponde a quantidade de vezes, dentro do intervalo, um evento ocorre.

e → nº de euler.

→ nº de sucessos

Digitalizado com CamScanner

data
fecha

Sabendo que $\lambda = 5$ mensagens/hora, qual a Probabilidade de:

a) receber 2 mensagens por em 1 hora:

$$P(2) = \frac{e^{-5} \cdot (5)^2}{2!} \rightarrow 0{,}08422$$
8,422%

b) 7 mensagens em 2 horas:

$\lambda = 5/hora$
$\lambda = 10/2horas$

$$P(7) = \frac{e^{-10} \cdot (10)^7}{7!} \therefore 0{,}09008$$
9,008%

c) 1 mensagem em 30 min:

$\lambda = 5/hora$
$\lambda = 2{,}5/30min$

$$P(1) = \frac{e^{-2,5} \cdot (2{,}5)^1}{1!} \therefore 0{,}2052$$
→ 20,52%

um Posto de bombeiro recebe em média 3 chamadas por dia. Calcule: $\lambda = 3/dia$

a) a Probabilidade de receber 4 chamadas em 1 dia.

$$P(4) = \frac{e^{-3} \cdot 3^4}{4!} \therefore 0{,}1680$$
16,80%

b) Receber 3 ou mais.

$$P(x \geq 3) = 1 - (P(0) + P(1) + P(2))$$
$$P(x \geq 3) = 1 - \left(\frac{e^{-3} \cdot (3)^0}{0!} + \frac{e^{-3} \cdot (3)^1}{1!} + \frac{e^{-3} \cdot (3)^2}{2!}\right)$$
$$P(x \geq 3) = 1 - (0{,}04978 + 0{,}14936 + 0{,}22404)$$
$$P(x \geq 3) = 1 - (0{,}42318)$$
$$P(x \geq 3) = 0{,}57682$$
→ 57,682%

Digitalizado com CamScanner

NO EXCEL: INT.CONFIANCA.NORMAL
(α = a diferença de 100 do intervalo de confiança,
σ = desvio,
n = nº amostra)

data
fecha

· Margem de erro

→ É o termo adicionado e subtraído do estimador para formar um intervalo de confiança.
De forma geral, o valor do erro amostral é:

$$e = Z_{crit} \cdot \frac{\sigma}{\sqrt{n}}$$

Z_{crit} → tabela z, depende da confiança que buscamos, mais comuns

σ → desvio padrão

n → nº de amostra

· Com isso, a margem de erro se dá por:

$$\bar{X} \pm Z_{crit} \cdot \frac{\sigma}{\sqrt{n}}$$

· Caso queira o tamanho do espaço amostral, podemos modificar a formula:

$$e = Z \cdot \frac{\sigma}{\sqrt{n}} \rightarrow n = \left(\frac{Z \cdot \sigma}{e}\right)^2$$

Ex: Qual o tamanho da amostra necessária coletar para garantir um erro amostral de 0,05 kg com 95% de confiança na estimação do verdadeiro valor médio, sabe-se que o desvio padrão é 0,20 kg

$$n = \left(\frac{1,960 \cdot 0,20}{0,05}\right)^2 \simeq 62$$

Para uma amostra de 36, o preço médio é 70,5. Com DP de 4,50. Qual o intervalo com 95% de confiança.

$n = 36$ $\quad \sigma = 4,50$ $\quad$ 95% → $z = 1,96$

$\bar{X} = 70,5$

$$e = 1,96 \cdot \frac{4,5}{\sqrt{36}} \therefore e = 1,47$$

$$I = [69,03 ; 71,97]$$

data
fecha
D S T Q Q S S
D L M M J V S

• Testes de hipótese

→ admitimos um valor hipotético para um parâmetro populacional, e com base nas informações da amostra realizaremos um teste estatístico, para aceitar ou rejeitar a hipótese.

→ Para isso, como estamos trabalhando com uma amostra, devemos dimensionar a probabilidade (risco) de aceitar ou rejeitar a hipótese.

* Devemos trabalhar com 2 hipóteses:

H_0 → hipótese nula → o que queremos

H_1 → hipótese alternativa → o oposto do que queremos

Ex: $\begin{cases} H_0 : \mu = K \\ H_1 : \mu \neq K \end{cases}$ $\begin{cases} H_0 : \mu \leq K \\ H_1 : \mu > K \end{cases}$ $\begin{cases} H_0 : \mu \leq K \\ H_1 : \mu > K \end{cases}$

$K \in \mathbb{R}$

H_0 sempre $=, \geq, \leq$

H_1 sempre $\neq, >, <$

• Para realizarmos um teste de hipóteses, partimos de um dado valor α, para construir a regra de decisão.

↳ nível de significância (geral entre 1 e 10%)

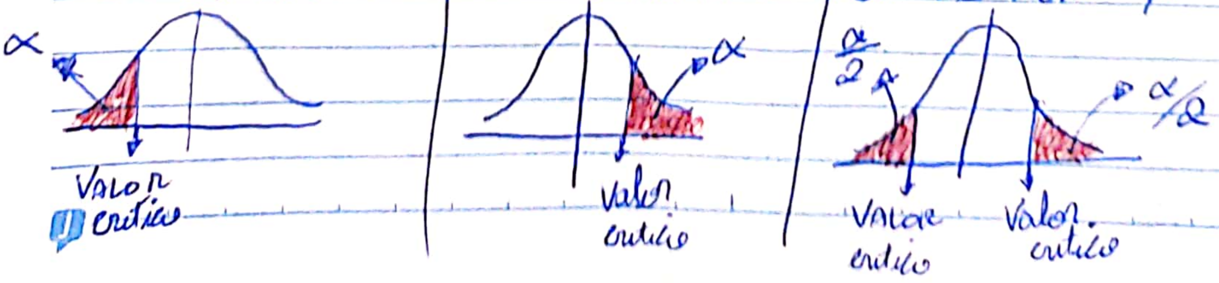
- PASSO a Passo

1) Escrever o H_0 e o H_1 → hip nula → hip alternativa

2) Calcular o valor observado (Z ou t) → tabela

3) Desenhe o gráfico da distribuição amostral

Se H_1 tem $<$ | Se H_1 tem $>$ | Se H_1 tem $\neq$



Digitalizado com CamScanner

data
fecha

4) Calcular o Valor crítico (Z_{crit} ou t_{crit}) de acordo com o nível de significância do teste (α) e com a região crítica. Utilizar a tabela

5) Marque o valor observado no gráfico

6) Conclua o teste:

- Se o Valor observado $\in$ Região crítica, rejeita Ho

- Se o Valor observado $\notin$ Região crítica, aceita Ho

Obs: Quando o desvio Padrão (σ) for conhecido, temos que:

$$Z_{obs} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\bar{X}$ = média amostral
μ = média Populacional testada (Ho)
σ = desvio Padrão Populacional
n = tamanho amostra
não pode ser o da amostra !!

Ex: Funcionários afirmam que a média de seus salários é menor que a de seus concorrentes, que é 45.000,00. Uma amostra de 30 funcionários mostrou média de 43.500,00. Sabe-se que o desvio Padrão dos salários é de 5200,00. Teste a hipótese dos funcionários ao nível de significância de 5%.

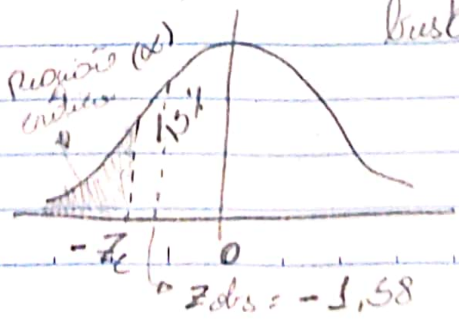
hipótese $\begin{cases} H_0 \geq 45000 \\ H_1 < 45000 \end{cases}$

$\alpha = 5\%$ $\quad \bar{X} = 43500$
$\sigma = 5200$ $\quad \mu = 45000$
$n = 30$

$$Z_{obs} = \frac{43500 - 45000}{\frac{5200}{\sqrt{30}}} = -1{,}58$$

0,45

Buscar na tabela o Zcrítico
$Z_c = 1{,}645$



Zobs $\notin$ Região crítica
Aceita Ho