Lista L1 - Ana Carolina Sá Coelho da Silva

Exercícios da seção 5.1

5.1 Para avaliar a percentagem da população que está abaixo de 25, avalia-se o gráfico a seguir:

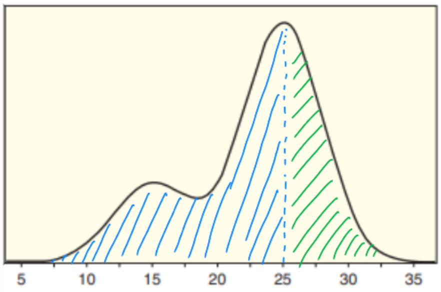


Figure 5.11 *Density curve for Exercises 5.1 to 5.3*

Queremos descobrir a proporção da área rachurada em azul.

Visualmente, é possível afirmar que os valores abaixo de 25 compreendem mais de 50% da população e menos de 95%, visto que ainda há uma boa quantidade na área rachurada em verde. Portanto, a percentagem que mais se aproxima é 62%.

5.5 Usando $N(0,1) \rightarrow \mu = 0$ e $\sigma = 1$

Para calcular a área com base no Z, podemos usar a função norm.cdf da biblioteca scipy.stats do python.

a) $Z = 1.04$

Área acumulada $\rightarrow$ norm.cdf(1.04) = 0,8508

b) $Z = -1.5$

Área acima de -1.5 corresponde a 1 - área acumulada.

Área acumulada $\rightarrow$ norm.cdf(-1.5) = 0,0668

Área acima = 1 - 0,0668 = 0,9332

c) $Z = 1$ e $Z = 2$

Área entre $Z=1$ e $Z=2$ corresponde a Área acumulada até 2 - Área acumulada até 1 $\Rightarrow P(1<Z<2) = P(Z<2) - P(Z-1)$

$\therefore$ norm.cdf(2) - norm.cdf(1) = 0.1359

5.10 Calcular Z com $N(0,1)$ para:

a) Área à esquerda de z é cerca de 0,7 :

Para calcularmos o z cuja área acumulada é de 0,01, podemos utilizar a função norm.ppf (área). Sendo $Z =$ norm.ppf (0,7)

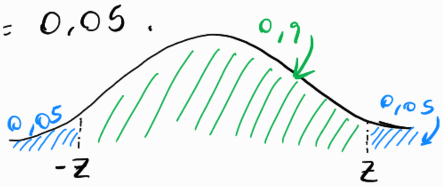
$Z \approx 0{,}5244$

b) Área à direita de z é cerca de 0,01

Para saber o z cujo área a direita é de 0,01, podemos calcular a partir da área a esquerda com $1 - 0{,}01 = 0{,}99$, então podemos usar norm.ppf(0,99) = z e temos que $z \simeq 2{,}326$

c) Área entre $\pm z$ é 0,9

Área fora de 0,9 é de $1 - 0{,}9 = 0{,}1$. Como a distribuição normal é simétrica, cada área "externa" possui $\frac{0{,}1}{2} = 0{,}05$.



$$\left.\begin{array}{l}\text{Esquerda de } z = 0{,}95 \\ \text{Esquerda de } -z = 0{,}05\end{array}\right\} = 1$$

Com norm.ppf(0,95), calculamos $z \simeq 1{,}645$

5.15 Encontrar áreas dado um N( ) → $z = \frac{(x - \mu)}{\sigma}$, onde x é a área

a) Área acima de 200 com N(120, 40)

$\mu = 120$
$\sigma = 40$
$z = 200$

Podemos usar a função norm.sf, que calcula diretamente $1 - \text{cdf}(z, \mu, \sigma)$. → Área acima
área = norm.sf(z, loc=μ, scale=σ) $\simeq$ 0,023

b) Área abaixo de 49.5 com N(50, 0.2)

mu = 50
$\sigma = 0.2$
$z = 49.5$

Com norm.cdf temos: → Área abaixo
área = norm.cdf(49.5, loc=50, scale=49.5) $\simeq$ 0,0062

c) Área entre 0.8 e 1.5 com N(1, 0.3)

mu = 1
$\sigma = 0.3$
$z1 = 0.8$
$z2 = 1.5$

Área a esq de 1.5 - Área a esq de 0.8
Área a esq.
Com norm.cdf:
área 1 = norm.cdf(1.5, loc=1, scale=0.3)
área 2 = norm.cdf(0.8, loc=1, scale=0,3)

área 1 - área 2 $\simeq$ 0,699 ou $\approx$ 0,7

5.20 Encontrar z com as propriedades:

*Para achar z, usamos a função norm.ppf, que pode receber a área, média (mu) e desvio padrão (σ). $z = \frac{(x - \mu)}{\sigma}$

a) A área a esquerda com N(5,2) é cerca de 0.10

mu = 5
$\sigma = 2$
A = 0,10

z = norm.ppf(A, loc=mu, scale=sigma) $\simeq$ 2,437

b) A área a direita com N(500, 25) é cerca de 0,05

$mu = 500$, $\sigma = 25$, $A = 0{,}05$ } Se a área a direita é 0,05, a esquerda é $1 - 0{,}05 = 0{,}95$

$A_e = 0{,}95$

$Z = \text{norm.ppf}(0{,}95,\ loc = 500,\ \sigma = 25) \simeq 541.12$ //

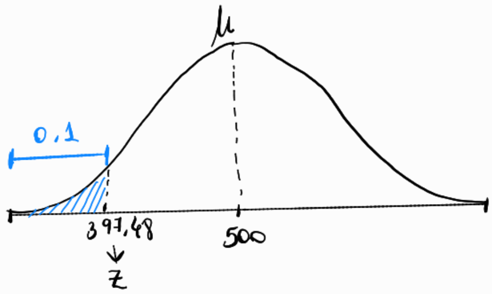
5.25 Converter uma área de uma dist. Normal p/ uma área equivalente em outra dist. Normal. Desenhar rascunhos de ambas distribuições normais, encontrar e rotular os Z-scores, sombreando as regiões corresp. às curvas.

a) 10% inferiores p/ N(500, 80).

→ Queremos Z com a área a esquerda 0,1 (10% de 1) em N(0, 1) para N(500, 80)

Aplicando em norm.ppf(A, loc = μ, scale = σ) p/ obter Z:

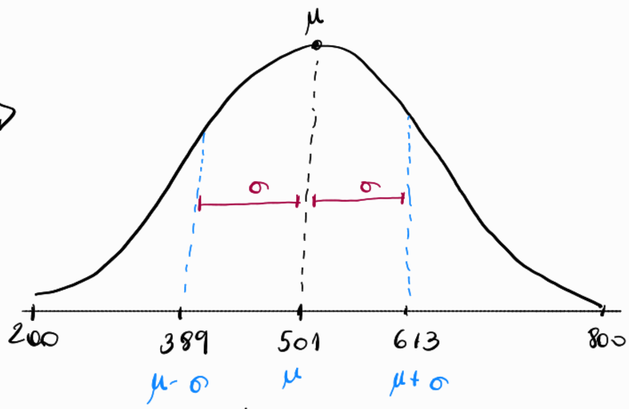
$A = 0{,}1$, $\mu = 500$, $\sigma = 80$ } $Z = \text{norm.ppf}(0.1, 500, 80) \simeq 397{,}48$



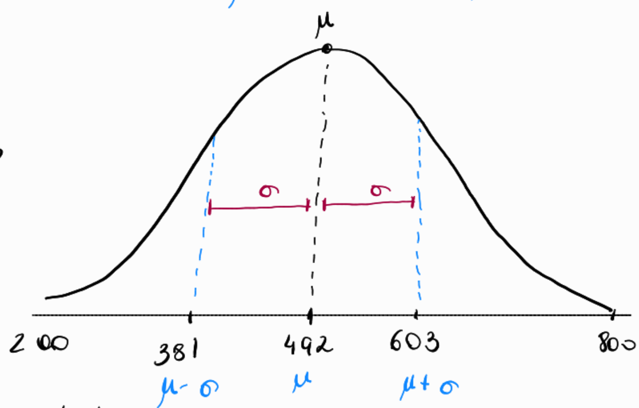
5.30 Desenhar a distribuição Normal:

Critical Reading { média = 501, $\sigma = 112$ } $N(\overset{\mu}{501}, \overset{\sigma}{112})$ ⇒

↳ os "extremos" serão 501 ± 112



Writing { $\mu = 492$, $\sigma = 111 \to 492 \pm 111$ } N(492, 111) ⇒



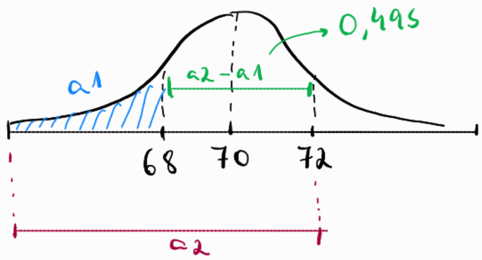
5.35 Altura de homens adultos nos EUA são aprox. distribuídas normalmente, com média de 70 pol (5 pés e 10 pol) e desvio padrão de 3 pol.

a) Qual a proporção de homens que têm entre 68 e 72 polegadas?

Para calcular essa proporção, podemos utilizar norm.cdf do python p/ calcular a área entre as duas alturas, que se dá por $\underset{\text{máx}}{a2} - \underset{\text{mín}}{a1}$. Temos que $\mu = 70$.

$z_1 = 68$, $N(70, 3) \rightarrow a1 =$ norm.cdf(68, loc=70, scale=3) $\Big\}$ $a2 - a1 \approx 0,495$

$z_2 = 72$, $N(70, 3) \rightarrow a2 =$ norm.cdf(72, loc=70, scale=3)



A proporção é de 49,5% com altura entre 68 e 72

b) Se um homem está no 10° percentil de altura, o quão alto ele é?

O 10° percentil é dado por 10% de 1 = 0,1 → Área

Para saber sua altura, calculamos o z-score com norm.ppf de python

z = norm.ppf(0.10, 70, 3) → $z \approx 66,15$ polegadas

5.40 transformar notas com $N(62, 18)$ para $N(75, 10)$ com padronização:

Convertendo em Z-score:

* Nota original = 47 = X

$$Z = \frac{X - \mu}{\sigma} \rightarrow Z = \frac{47 - 62}{18} \approx -0,833$$

* Nota original = 90 = x → $Z = \frac{90 - 62}{18} \approx 1,556$

Utilizar a fórmula inversa $X = Z \cdot \sigma + \mu$ p/ calcular a nova nota com $\mu = 75$ e $\sigma = 10$

* $Z \simeq -0,833$ (Nota original 47)

$X = -0,833 \cdot 10 + 75 \simeq 66,67$

* $Z \simeq 1,556$ (Nota original 90)

$X = 1,556 \cdot 10 + 75 \simeq 90,56$

Notas:

Original	Novo
47	66,67
90	90,56

Exercícios da seção 3.1

3.2 Parâmetro, pois consideramos todos os jogadores do elenco → população

Notação: ρ → correlação populacional

3.4 Parâmetro, pois consideramos todos os eleitores do condado

Notação: p → proporção populacional

3.6 $p = \frac{\text{homeless}}{\text{total}} = \frac{170000}{78000000} \simeq 0,00218 \rightarrow p \simeq 0,218\%$

↪ proporção populacional

3.8 $\hat{p} \rightarrow$ proporção amostral

$\hat{p} = 0{,}82$ ou 82%

3.10 $\bar{x} \rightarrow$ média amostral

$\bar{x} = 13.10$

3.12 O gráfico de a) segue uma distribuição normal

erro padrão $\rightarrow SE = \sqrt{\dfrac{p(1-p)}{n}}$, onde $p = 0{,}3$ (média)

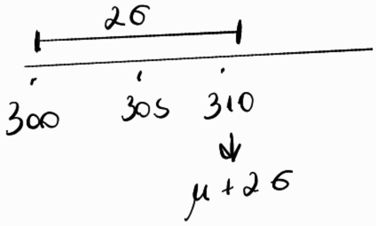
$n = 40$ (tamanho da amostra)

$SE = \sqrt{\dfrac{0{,}3(0{,}7)}{40}} \simeq 0{,}0725$. Como é normal, 95% estará entre $\pm 2SE \simeq \pm 0{,}145$ da média.

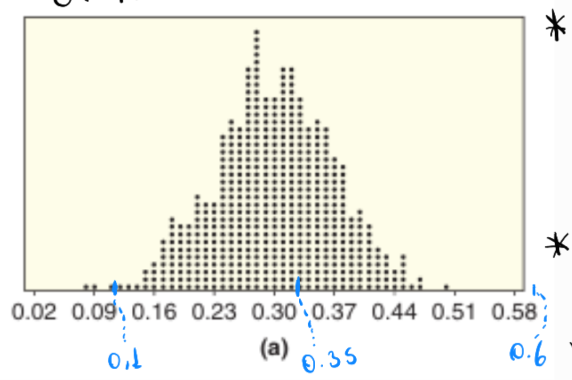
3.14 P/ gráfico c) temos que $\mu = 300$ e $n = 100$

95% está entre 290 à 310 $\rightarrow 2\sigma = 310 - 300$

$\sigma = \dfrac{10}{2} = 5$ //



3.16
Gráfico de a)



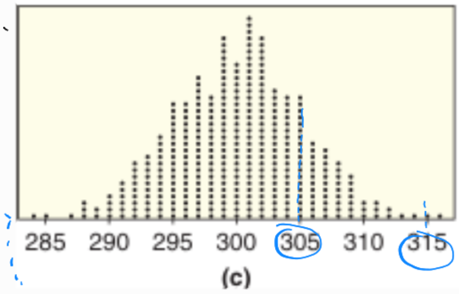
* Avaliando a) $\hat{p} = 0.1$: esse valor está 0.2 abaixo da média sendo cerca de 3 desvios abaixo. Mas como ainda há pontos de ocorrência, há uma baixa probabilidade de ocorrer.

* Avaliando b) $\hat{p} = 0.35$: esse valor está dentro do intervalo de 1 desvio padrão acima da média, sendo visualmente com muitas ocorrências, então há uma alta probabilidade de ocorrer

* Avaliando c) $\hat{p} = 0.6$: Esse valor nem pode ser visualizado no gráfico, estando muito acima da médio, então sua probabilidade é nula.

$\hat{p}$	Probabilidade
0.1	baixa
0.35	alta
0.6	nula

3.18



250

$\bar{x}$	Probabilidade
315	baixa
350	alta
250	nula

a) $\bar{x} = 250$: esse valor nem é representado no gráfico e está fora das chances de ocorrência, então a prob. é nula

b) $\bar{x} = 305$: esse valor está bem próximo à média e está a 1 desvio padrão acima apenas, indicando que há grandes chances de ocorrer.

c) $\bar{x} = 315$: esse valor está a 3 desvios da média, mas ainda possui poucas chances de ocorrer.

3.20

População: Todos os usuários da internet nos EUA em janeiro de 2010

Parâmetro de interesse: proporção dos usuários que personalizam a página inicial

Tamanho da amostra: 1675
Usuários que personalizam: 469

$\hat{p} = \frac{469}{1675} \approx 0,28$ ou 28%

Para saber o valor exato é preciso obter os dados da população completa e não somente uma amostra de 1675 usuários.

Exercícios da seção 3.2

3.39

$\bar{x} = 25$
margem = 3

I → $\bar{x} \pm$ margem de erro → Intervalo: (22, 28)

3.41

r = 0,62
margem = 0,05

Intervalo: (0,57, 0,67)

3.43 Intervalo: (112.1, 128.2) com 95% de confiança

a) $\mu = 121$ é um valor plausível pois está dentro do intervalo

b) $\mu = 113.4$ é um valor plausível pois está dentro do intervalo

c) $\mu = 105.3$ não é plausível pois está fora do intervalo

3.45

$\hat{p} = 0.32$ com $SE = 0.04$, Com 95% de confiança numa distribuição simétrica, temos que os valores estão entre $\pm 2\sigma$ (ou $\pm 2SE$). Então:

$$\hat{p} \pm \underbrace{2SE}_{\text{Margem de erro}} = 0.32 \pm 0.08 \rightarrow \text{Intervalo: } (0.24, 0.40)$$

3.47

$r = 0.34$, $SE = 0.02 \Rightarrow r \pm 2SE = 0.34 \pm 0.04 \rightarrow$ Intervalo: $(0.30, 0.38)$

3.49

$\bar{x}_1 - \bar{x}_2 = 3$, $2SE = 1.2 \rightarrow$ margem de erro já fornecida $\rightarrow$ Intervalo: $(1.8, 4.2)$

$\hookrightarrow 3 \pm 1.2$

3.51

a) É uma estatística pois foi obtida de uma amostra.

Notação: $\hat{p} = 0.30$

b) Estimamos a proporção populacional p de todos os jovens nos EUA que já foram presos por algo além de violação de tráfego até 23 anos com $\hat{p} = 0.3$.

c) Margem de erro = 0.01

$\hat{p} \pm$ margem $\rightarrow 0.3 \pm 0.01 \rightarrow$ Intervalo: $(0.29, 0.31)$

d) Como o limite inferior é 0.29, isso é improvavel pois exatamente 0.3 é uma precisão muito alta.

3.53

O intervalo de confiança de 95% de 0.83 a 0.89 significa que há confiança de 95% de que a proporção de todos os adultos dos EUA que consideram o carro uma necessidade está entre 83% e 93%. Ou seja, se o processo de amostragem aleatória fosse repetido muitas vezes, em cerca de 95% delas o intervalo incluiria esse intervalo.

3.55 tamanho da amostra = 2625

Concordam = 735, SE = 0,009

Parâmetro estimado $\rightarrow$ proporção populacional por meio de $\hat{p}$ (amostral)

$\hat{p} = \frac{735}{2625} \approx 0.28$, contando com SE p/ 95% de confiança temos $0.28 \pm \underset{0.018}{\underset{\downarrow}{2SE}}$

Intervalo : (0.262 , 0.298)

Interpretação: Há confiança de 95% de que a proporção verdadeira de adultos nos EUA que acreditam que existe apenas 1 amor verdadeiro está entre 26.2% e 29.8%.

3.57 Voluntários = 23 homens, 20 anos de média

11 jogam, 12 não

Medidas: tempo de resposta : -1.8 a -1.2 (95%)

precisão : -4.2 a 5.8 (95%)

a) P/ tempo de resposta o intervalo é (-1.8, -1.2), o que significa que há 95% de confiança de que jogadores têm um tempo de resposta entre -1.8 e -1.2, ou seja, a menos, que não jogadores.

b) Não é provável que sejam iguais, pois o intervalo não contém 0. Como todos os valores são negativos, isso significa que os jogadores são mais rápidos, com tempo de resposta menor.

c) P/ Precisão, temos 95% de confiança de que a diferença entre jogadores e ñ jogadores está entre (-4.2, +5.8).

d) É possível que sejam iguais pois o intervalo contém 0 e valores negativos e positivos distintos dentro de um intervalo grande.