⇒ Combinações de Variáveis Aleatórias.

1) $Z = X + Y \Rightarrow E[Z] = E[X] + E[Y]$

2) $Z = aX + bY \Rightarrow E[Z] = a.E[X] + bE[Y]$

3) $Z = X + Y \Rightarrow Var(Z) = Var(X) + Var(Y) + 2\,Cov(X,Y)$

4) $Z = X - Y \Rightarrow Var(Z) = Var(X) + Var(Y) - 2\,Cov(X,Y)$

5) $Z = a.X + bY \Rightarrow Var(Z) = a^2.Var(X) + b^2.Var(Y) + 2ab\,Cov(X,Y)$

6) $Z = X + Y$, onde X e Y são não correlacionadas

$\Rightarrow Cov(X,Y) = 0$

$\Rightarrow Var(Z) = Var(X) + Var(Y)$

$\boxed{Var = \sigma^2}$

7) $Z = X + Y$, X e Y tem distribuição normal

⇒ Z tem distribuição normal

8) $Z = X + Y$, X e Y não tem ambas a distribuição normal

⇒ Nada se pode afirmar sobre a distribuição Z

⇒ covariância entre as duas variáveis aleatórias:

$$cov = \frac{corr}{\sigma_x \cdot \sigma_y}$$

corr → correlação

$\sigma_x \cdot \sigma_y$ → produto dos desvios padrões

Exercício

$X_i \sim Normal(\mu = 100;\ \sigma = 10)$

$n = 25$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$

T.L.C $\mu_{\bar{X}} = \mu = 100$

$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} \Rightarrow \frac{10}{\sqrt{25}} = 2$

$\bar{X} \sim Normal(\mu_{\bar{x}} = 100,\ \sigma_{\bar{x}} = 2)$

res = norm.cdf(95, loc=100, scale=2)

$\bar{d} = 1.01$ inch   * para achar o novo desvio padrão

$\sigma_d = 0.003$ in $\frac{\sigma}{\sqrt{n}}$ → número de amostras

$\sigma^2_{\bar{x}} = \frac{\sigma_0^2}{n} \Rightarrow \underbrace{\sigma_{\bar{x}}}_{\text{novo}} = \frac{\sigma_0}{\sqrt{n}}$ → original

Exercício 4

X_1	X_2
$\bar{A} \sim N(5k, 40)$	$\bar{A} \sim N(5k, 10)$
$\bar{B} \sim N(5{,}05k, 30)$	$\bar{B} \sim N(5{,}05k, 6)$

$C = B - A$ $C \sim N(50, \sqrt{136})$

1-norm.cdf(25, loc=50, scale=136**0,5)

• Hipótese Nula (H_0) → "o que o mundo deveria ser"

• Hipótese Alternativa (H_A) → Se a hipótese nula falha, devemos aceitar H_A

↳ "sempre com uma relação de igualdade"

* Teorema do limite central

ex:

X → tempo de processamento

1) $H_0: \mu \leq 41$ | $H_A: \mu > 41$ 2) Estatística de teste $\bar{X}$

* $\bar{X}$ está dentro da área crítica ⟺ o valor p é menor que o α.

DADOS - Teste de Hipóteses

1) Definir as hipóteses:

- H_A: Hipótese Alternativa!
  (aquilo que desejamos provar)
- H_0: Hipótese nula:
  (aquilo que vou supor, para tentar quebrar)

* A lógica é a redução ao absurdo — LÓGICA DO TESTE

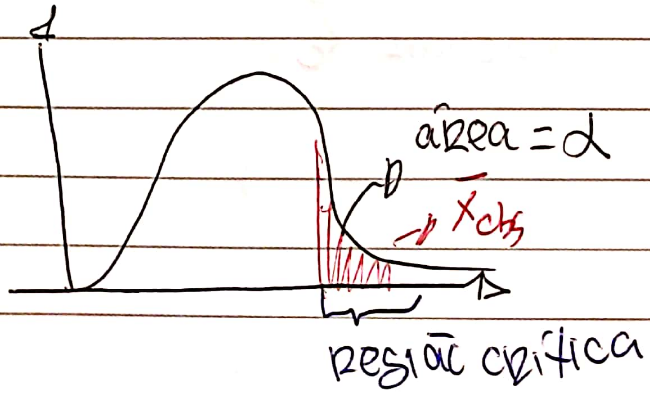
a) Quero provar que H_A é verdadeira
b) Suponha que H_0 é verdadeira
c) Por outro lado, temos dados
d) Vamos mostrar que os dados são improváveis sob a hipótese nula
e) H_0 verdade $\Rightarrow$ absurdo; logo H_0 falsa $\Rightarrow$ $\boxed{\text{Aceitar } H_A}$ & rejeitou H_0
f) Agora, se H_0 verdade $\Rightarrow$ dados plausíveis
↳ Nada se pode dizer
↳ "Não rejeitamos a hipótese nula"
↳ "procurar mais dados"

2) Coletar os dados

3) Definir a estatística de teste → "Lupa p/ enxergar os dados"

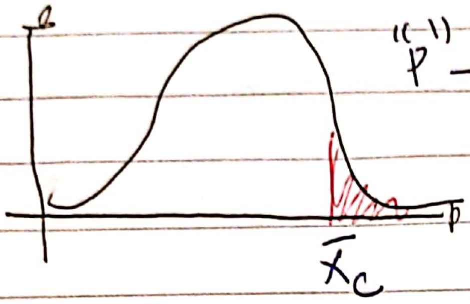
4) Supondo H_0 verdadeira $\Rightarrow$ construa sua REGRA DE DECISÃO

I) Região crítica: definir região tal que se a estatística do teste cai dentro, então REJEITA H_0



- $\bar{X}_{obs}$ dentro de R.C $\Rightarrow$ Rejeita H_0

II) Valor-P: Qual seria o α para que minha estatística de teste estivesse no limite da correspondente região crítica? comparar com α original.



"P" $\to$ se Área $< \alpha \Rightarrow$ rejeita H_0

Teste de média com desvio-padrão conhecido

1) $H_0: \mu = \mu_0$ $\to$ média populacional

$H_A: \mu < \mu_0$ (unicaudal à esquerda)

• $\mu > \mu_0$ (unicaudal à direita)

• $\mu \neq \mu_0$ (bicaudal)

2) Dados: x_i, $i = 1 \dots n$

3) Nossa estatística de teste: $\boxed{\bar{X}_{OBS} = \frac{1}{n}\sum_{i=1}^{n} x_i}$

4) Definir Regra de Decisão:

• Vamos supor: $\mu = \mu_0$ (H_0)

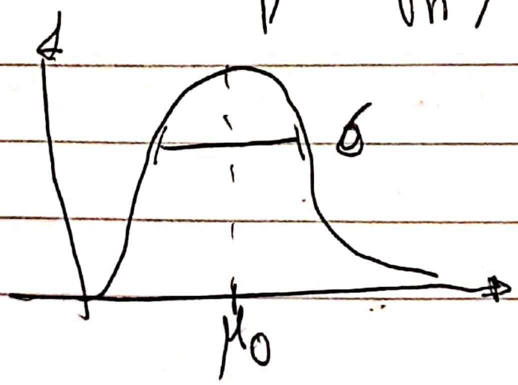
• x_i amostra de $X \sim Normal(\mu_0, \sigma)$

• σ conhecido

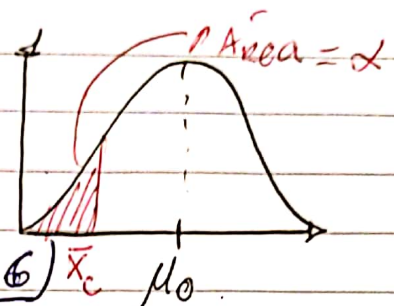
• nível de significância $\Rightarrow \alpha$

$X \sim Normal\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$

• achar $\bar{X}_c$:



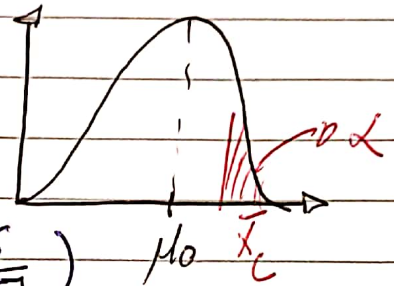
- Se $H_A: \mu < \mu_0$



p/ achar $\bar{x}_c$.

$\bar{x}_c = norm.ppf\left(\alpha, loc = \mu_0, scale = \frac{\sigma}{\sqrt{n}}\right)$

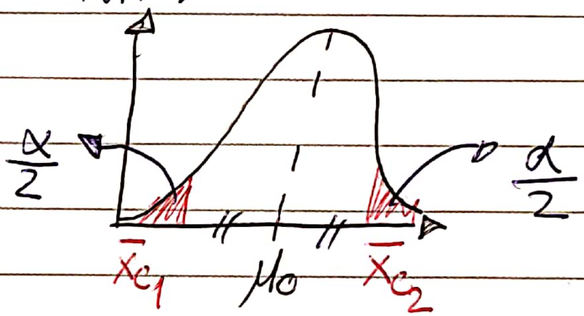
- Se $H_A: \mu > \mu_0$



p/ achar $\bar{x}_c$:

$\bar{x}_c = norm.ppf\left(1-\alpha, loc = \mu_0, scale = \frac{\sigma}{\sqrt{n}}\right)$

- Se $H_A: \mu \neq \mu_0$



p/ achar $\bar{x}_{c_1}$:

$\bar{x}_{c_1} = norm.ppf\left(\frac{\alpha}{2}, loc = \mu_0, scale = \frac{\sigma}{\sqrt{n}}\right)$

p/ achar $\bar{x}_{c_2}$:

$\bar{x}_{c_2} = norm.ppf\left(1-\frac{\alpha}{2}, loc = \mu_0, scale = \frac{\sigma}{\sqrt{n}}\right)$

x_c = stats.t.ppf(α, df = n-1, Loc = μ, scale = s/sqrt(n))

DISCRETA:

- Bernoulli → lançamento de uma moeda – $X \sim Ber(p)$
- Binomial → contagem de lançamentos – $X \sim Bin(p, n)$
- Poisson → nº de ocorrências em um intervalo – $X \sim Poisson(\lambda)$

CONTÍNUA:

- Uniforme: $X \sim Unif(a, b)$
- Exponencial: tempo até ocorrência de um evento – $X \sim Expon(\beta)$ ou $\alpha = \frac{1}{\beta}$ $\sigma = \mu$
- Normal: X é combinação de vários fatores independentes – $X \sim Normal(\mu, \sigma^2)$
- t-student: $X \sim t_{n-1}(\mu, \sigma^2)$

loc $= \mu$
scale $= \sigma$

EXPON:
SCALE $= \beta$

* cdf → área
ppf → número

* x_crítico → ponto de corte
valor p → alpha p/ ainda rejeitar a hipótese nula

* Xcrítico dentro da região crítica
↳ rejeita H_0

* valor $p < \alpha$ ⇒ rejeita H_0

Regressão Linear

- **simples** $\rightarrow \hat{Y} = \beta_0 + \beta_1 X$

$\hat{\beta}_1 = \frac{cov(x,y)}{VAR(x)}$ | $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

Ho $\rightarrow$ $H_0: \beta_1 = 0$ $\rightarrow$ H_0: Não há relação entre x e Y.
HA $\rightarrow$ $H_1: \beta_1 \neq 0$ $\rightarrow$ H_1: há relação entre x e Y

$\rightarrow$ statsmodels.OLS (Biblioteca)

- β_0 = sm.add_constant(dataframe)

```
model = sm.OLS(Data2, B0)
results = model.fit()
results.summary()
```