Curso de Pós-Graduação em Ciências Veterinárias - UFRRJ

Métodos Estatísticos

Prof: Wagner Tassinari

wagner.tassinari@ini.fiocruz.br

Distribuições de Probabilidades para Variáveis Discretas e Contínuas

Variáveis Aleatórias

Varíaveis Aleatórias

- **Variável** - Qualquer característica que pode ser medida ou categorizada.
- Variáveis aleatórias (v.a) são eventos associados com números.
- **Exemplo:** Lançamento de duas moedas

Variáveis Aleatórias

- Exemplo: $\Omega = \{(K,K),(K,C),(C,K),(C,C)\}$

Eventos (X)	K,K	K,C	C,K	C,C
Probab. P(X)	1/4	1/4	1/4	1/4

- Seja X a quantidade de caras (C). Dessa forma X pode ser definido com V.A.

X	0	1	2
$\mathrm{P}(X)$	$1/4$	$2/4$	$1/4$

Representação das Variáveis Aleatórias

- As variáveis aleatórias (v.a) geralmente são representadas por letras maiúsculas X, Y ou Z.

Exemplos de Variáveis Aleatórias

- Resultado de um teste diagnóstico
- Sexo do paciente
- Idade do paciente
- Cor ou raça do paciente
- Número de filhos de mulheres em idade fértil
- Tempo de tratamento
- Número de infecções auditivas de bebês menores de 6 meses

Tipos de Variáveis Aleatória

- **Variável aleatória discreta:** Assume somente um número finito ou enumerável de resultados
- **Variável aleatória contínua:** Assume qualquer valor dentro de um intervalo

Distribuições de Probabilidade

- Uma distribuição de probabilidades é uma representação do conjunto de probabilidades de todos os eventos associados a um espaço amostral (universo) Ω.
- Se Ω é finito, a distribuição de probabilidades pode ser descrita enumerando-se todos os eventos de Ω e suas probabilidades.

Distribuição de Probabilidades para Variáveis Aleatórias Discretas

- Distribuição de Bernoulli
- Distribuição Binomial
- Distribuição de Poisson
- Outras

Distribuição de Probabilidades para Variáveis Aleatórias Contínuas

- As variáveis aleatórias contínuas podem assumir **qualquer valor** num intervalo numérico.
- São representadas graficamente por **curvas**, chamadas de **função densidade de probabilidade**.

Função densidade de probabilidade

- **Curva de densidade ou função densidade de probabilidade:** é um gráfico de uma distribuição de probabilidade contínua.
- A **área** sob esta curva representa a **probabilidade de ocorrência**.
- Nas variáveis contínuas **não** existe a probabilidade de ocorrência de um **valor exato, mas sim de intervalos**.

Distribuição de Probabilidades para Variáveis Aleatórias Contínuas

- Distribuição Normal ou Gaussiana
- Distribuição T-Student(t)
- Distribuição F-Snedecor (F)
- Distribuição de Qui-Quadrado (χ^2)
- Outras

Distribuição Normal

Função densidade de probabilidade Normal (Gaussiana)

- Este modelo probabilístico é essencialmente importante na estatística por três razões principais:
- Inúmeros fenômenos contínuos parecem segui-la ou podem ser aproximados por ela.
- Podemos utilizá-la para aproximar várias distribuições de probabilidades discretas.
- Ela oferece a base para a inferência estatística clássica devido a sua afinidade com o teorema central do limite.

Distribuição de Probabilidades Contínua: Normal (Gaussiana)

- Sua distribuição de probabilidades é simétrica e é determinada por dois parâmetros, μ e σ^2 , respectivamente a média e a variância. A variável aleatória Normal é denotada como:

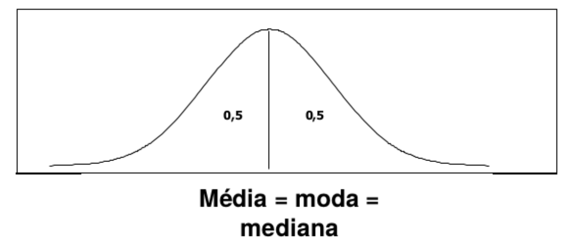
$$X \sim N(\mu, \sigma^2)$$

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

,para $-\infty \leq x \leq +\infty$

Distribuição Normal

- Representação gráfica:



- Os parâmetros da normal são a **média** (localização) e o **variância** (formato), que permitem infinitas curvas normais com diferentes formatos (sempre simétricas).

Transformando para NORMAL PADRÃO - N(0,1)

- Por meio da transformação de dados, precisaremos apenas de uma tabela.

$$Z = \frac{X - \mu}{\sigma} \rightarrow X = Z \sim N(0, 1)$$

Enquanto os dados originais para a variável aleatória X possuíam média aritmética μ e variância σ^2, a variável aleatória padronizada Z terá sempre média aritmética $\mu = 0$ e variância $\sigma^2 = 1$.

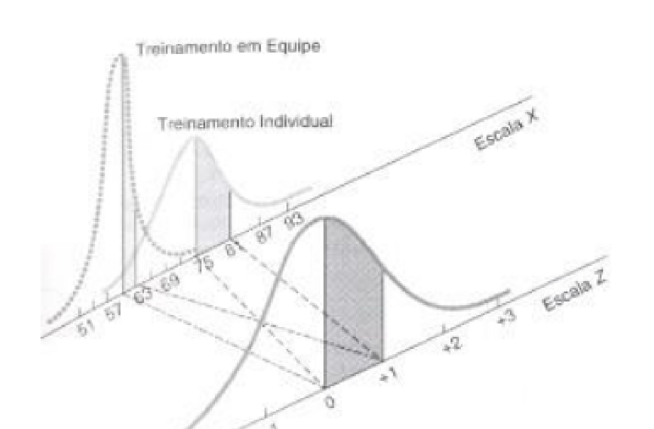
Tabela da distribuição NORMAL PADRÃO N(0,1)

Distribuição Normal Reduzida N(0,1)
Escores Z positivos
Probabilidades p tais que p = P(Z > Zc)

Parte inteira e primeira decimal de Z	Segunda decimal de Zc									
	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

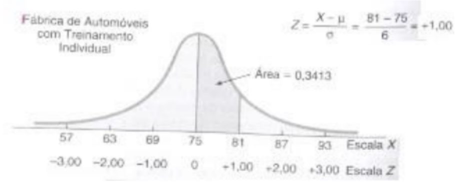
Transformando para NORMAL PADRÃO - N(0,1)

- **Exemplo:** Tempo (X) que os trabalhadores de uma fábrica de automóveis levam para montar uma peça, dado o treinamento individual. Com média de 75 segundos e desvio padrão de 6 segundos.



Transformando para NORMAL PADRÃO - N(0,1)

- Qual a probabilidade de uma pessoa com treinamento individual levar de 75 a 81 segundos para terminar a tarefa ?



$$Z = \frac{75-75}{6} = 0$$

$$Z = \frac{81-75}{6} = +1$$

Z	.00	.01	.02	.03	.04	.05
0.00	.500000	.496011	.492022	.488033	.484047	.480061
0.10	.460172	.456205	.452242	.448283	.444330	.440382
0.20	.420740	.416834	.412936	.409046	.405165	.401294
0.30	.382089	.378281	.374484	.370700	.366928	.363169
0.40	.344578	.340903	.337243	.333598	.329969	.326355
0.50	.308538	.305026	.301532	.298056	.294598	.291160
0.60	.274253	.270931	.267629	.264347	.261086	.257846
0.70	.241964	.238852	.235762	.232695	.229650	.226627
0.80	.211855	.208970	.206108	.203269	.200454	.197662
0.90	.184060	.181411	.178786	.176186	.173609	.171056
1.00	.158655	.156248	.153864	.151505	.149170	.146859
1.10	.135666	.133500	.131357	.129238	.127143	.125072

$$P(0 \leq Z \leq 1) = P(Z \geq 0) - P(Z \geq 1) =$$

$$P(0 \leq Z \leq 1) = 0{,}5000 - 0{,}1587 = 0{,}3413$$

Transformando para NORMAL PADRÃO - N(0,1)

IMPORTANTE !!!!

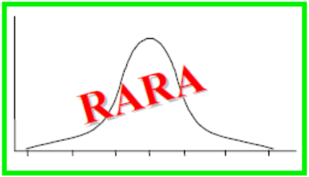
- As seguintes probabilidades associadas aos intervalos são, em geral, muito utilizadas...
  - $\mu \pm \sigma \rightarrow$ contém cerca de 68% das observações
  - $\mu \pm 2\sigma \rightarrow$ contém cerca de 95% das observações
  - $\mu \pm 3\sigma \rightarrow$ contém cerca de 99% das observações

Pergunta

- Como podemos decidir se o nosso conjunto de dados parece seguir ou pelo menos se aproximar da distribuição normal ?
- Nem todas as variáveis aleatórias contínuas são provenientes de uma distribuição normal !!!

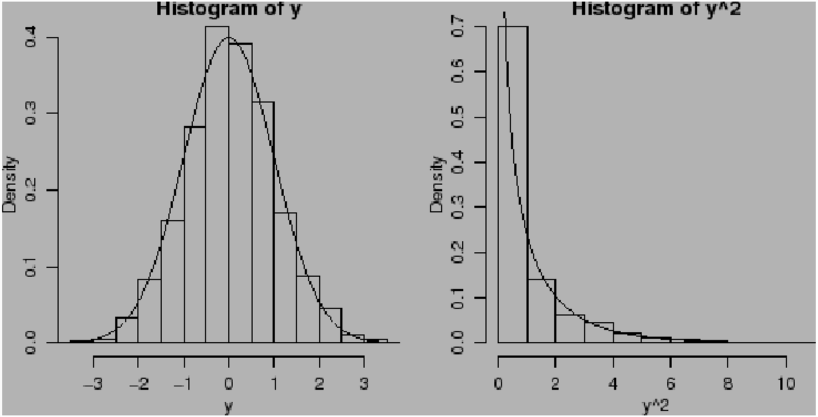
A distribuição dos meus dados é normal ?

A Curva Normal



- Também conhecida como a distribuição teórica.
  - Distribuição normal perfeita é muito rara com dados empíricos
- Quando se desrespeita as pressuposições da distribuição normal, chega-se a inferências inválidas e probabilidades sem sentido.

A distribuição dos meus dados é normal ?



Abordagens para verificar se os dados são normais

1. Descritiva exploratória
2. Gráfica
   - ramo-e-folha
   - boxplot
   - histograma
   - gráfico de probabilidade normal
3. Testes de hipóteses
   - Teste de *Shapiro-Wilk*
   - Teste de *Kolmogorov-Smirnov*
   - Outros.