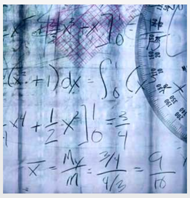
Estatística Aplicada

Noções Introdutórias e utilização do computador



Sétima aula - TLC e Distribuição

Distribuições de Probabilidade

Teorema do Limite Central

O Teorema do limite central (TLC) demonstra a tendência de aproximação das variáveis aleatórias com a distribuição normal.

Distribuições de Probabilidade

Teorema do Limite Central

- O teorema do limite central é básico para a maioria das aplicações do controle estatístico da qualidade.

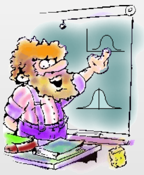
- A partir do teorema do limite central, sabe-se que a distribuição amostral das médias apresenta os seguintes parâmetros:

	População	Amostra
Média	μ	$\bar{x}$
Desvio-padrão	σ	S

Teorema do Limite Central

• A média dos dois dados resulta aproximadamente em uma distribuição Normal.

• A aproximação da distribuição Normal melhora na medida que se fizesse a média do lançamento de mais dados.



Distribuições Normal

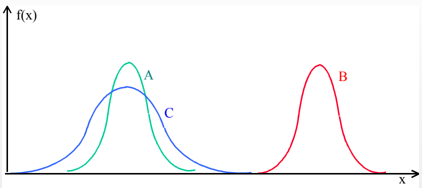
• A distribuição Normal fica completamente caracterizada por dois parâmetros: a média e o desvio-padrão (variabilidade).

• Diferentes médias e desvio-padrões originam curvas normais distintas.

Amostras	Dados	Localização ($\bar{x}$)	Variabilidade (R)
A	10 12 14 16 18	$\bar{x} = 14$	$\bar{R} = 8$
B	22 24 26 28 30	$\bar{x} = 26$	$\bar{R} = 8$
C	6 10 14 18 22	$\bar{x} = 14$	$\bar{R} = 16$

Variabilidade (amplitude total, DP, variância...) → R é a amplitude média

Distribuições Normal



a) da distribuição A para B muda a tendência central, mas a variabilidade é constante;

b) da distribuição A para C muda a variabilidade, mas a tendência central é constante;

c) da distribuição B para C muda a tendência central e a variabilidade.

Distribuições Normal

Distribuições de Probabilidade

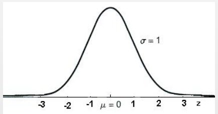
A distribuição Normal é a mais importante das distribuições estatísticas, tanto na teoria como na prática:

- Representa a distribuição de freqüência de muitos fenômenos naturais;
- As médias e as proporções de grandes amostras seguem a distribuição Normal;

Distribuições Normal

Distribuições de Probabilidade

- A distribuição Normal é em forma de sino, simétrica em relação à sua média e tende cada vez mais ao eixo horizontal à medida que se afasta da média.



- Teoricamente os valores da variável aleatória podem variar de -∝ a +∝.

Distribuições Normal

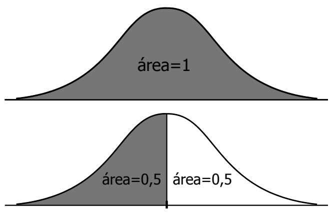
Distribuições de Probabilidade

- A área abaixo da curva Normal representa 100% de probabilidade associada a uma variável.
- A probabilidade de uma variável aleatória tomar um valor entre dois pontos quaisquer é igual à área compreendida entre esses dois pontos.

Distribuições Normal

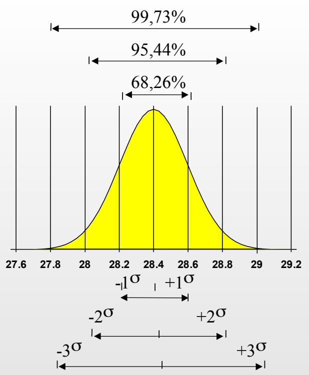
Distribuições de Probabilidade

A área total abaixo da curva é considerada como 100%.
Isto é, a área total abaixo da curva é 1.



Distribuições de Probabilidade

Percentuais da distribuição Normal:



Distribuições Normal

Distribuições de Probabilidade

• **O mundo de Z**

•A área sob a curva entre um ponto qualquer e a média é função somente do número de desvios-padrões que o ponto está distante da média.

• Como existem uma infinidade de distribuições normais (uma para cada média e desvio-padrão), transformamos a unidade estudada seja ela qual for (peso, espessura, tempo, etc.) na unidade Z, que indica o número de desvios-padrão a contar da média.

Distribuições Normal

Distribuições de Probabilidade

- O cálculo de probabilidades (área sob a curva) pode ser realizado através de uma distribuição Normal padronizada, onde o parâmetro é a variável reduzida Z (aproximação).

- A distribuição Normal pode ser representada por uma equação matemática dada por:

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

o número irracional: e = 2,7183... (base do logaritmo neperiano)

Distribuições Normal

Distribuições de Probabilidade

A distribuição Normal acumulada é obtida calculando a probabilidade de *X* ser menor que um dado valor *x*:

$$P(X \le x) = F(x) = \int_{-\infty}^{x} f(x)\,dx$$

A solução está apresentada em tabelas da distribuição Normal padronizada onde se entra com a variável reduzida *Z* (número de desvios-padrões distantes da média) e encontra-se *F(Z)* (área) ou vice-versa.

$$P\{X \le x\} = P\left\{Z \le \frac{x-\mu}{\sigma}\right\} = F(Z) \quad \Rightarrow \text{Tabelado}$$

Distribuições Normal

Distribuições de Probabilidade

O mundo de Z é mais fácil de ser compreendido do que se imagina.

Distribuição Normal

- A variável reduzida mede a magnitude do desvio em relação à média, em unidades de desvio padrão.

Z = 1,5 significa, simplesmente, uma observação está desviada 1,5 desvios padrão a cima da média.

Distribuição Normal

- A variável reduzida é muito útil para comparar distribuições e detectar dados atípicos.

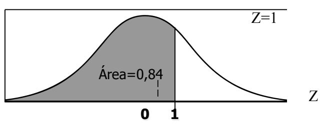
Dados são considerados atípicos quando *Z > 3 ou Z < -3.*

$$Z = \frac{x - \bar{x}}{s}$$

Para sabermos o valor da probabilidade, utilizamos a tabela da distribuição Normal. Essa tabela nos fornece a área acumulada até o valor de Z.

Por exemplo:

- Z =1 tem-se uma área de 0,84



- 0,84 = 84% de probabilidade ocorrência dos valores menores que Z

Distribuições Normal

Distribuições de Probabilidade

As áreas correspondentes as probabilidades da distribuição normal padrão estão tabeladas.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	**0.8413**	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	**0.8770**	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319

Probabilidade de ocorrência de valores abaixo de Z

• Z =1,16 tem-se uma área de 0,87

Distribuições Normal

Distribuições de Probabilidade

• Uma vez calculada a variável reduzida Z,

• Consulta-se a tabela Normal padronizada

• Identificar a probabilidade acumulada à esquerda de Z

• Ou seja, a probabilidade de ocorrerem valores menores ou iguais a um certo valor de Z consultado.

Distribuições Normal

Distribuições de Probabilidade

• O cálculo da variável reduzida *Z* faz uma transformação dos valores reais em valores codificados.

• A transformação é feita descontando-se a média para eliminar o efeito de localização (tendência central) e dividindo-se pelo desvio-padrão para eliminar o efeito de escala (variabilidade).

$$Z = \frac{x - \bar{x}}{s}$$

Distribuição Normal

Distribuições de Probabilidade

Exemplo 1: Suponha que o limiar diastólico dos pacientes hipertensos do HEJC seja normalmente distribuído com média 100 torr (100 mmHg), e desvio-padrão 10 (mmHg).

Então o limiar está em torno de 100 a uma distância as vezes maior, as vezes menor que 10.

Qual a probabilidade de um paciente, pego ao acaso, possuir limiar menor que 110 mmHg?

Distribuições de Probabilidade

Qual a probabilidade de um paciente, pego ao acaso, possuir limiar menor que 110 mmHg?

$$Z = \frac{x - \bar{x}}{s}$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	**0.5000**	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177

Distribuição Normal

Distribuições de Probabilidade

Queremos saber qual a probabilidade de um paciente, pego ao acaso, possuir limiar menor que 110 mmHg:

$$Z = \frac{x - \mu}{\sigma} = \frac{x - \bar{x}}{s} = \frac{110 - 100}{10} = 1$$

P(x <110) = P(Z < 1) = 0,8413 (aproximadamente 84,13%)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	**0.8413**	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	**0.8770**	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319

Probabilidade de ocorrência de valores abaixo de Z

Se quiséssemos saber a probabilidade do limiar ser maior que 111,6 mmHg?

$$Z = \frac{x - \bar{x}}{s}$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	**0.5000**	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177

Distribuição Normal

Se quiséssemos saber a probabilidade do limiar ser maior que 111,6 mmHg, iniciamos calculando o valor de Z:

$$Z = \frac{111{,}6 - 100}{10} = 1{,}16$$

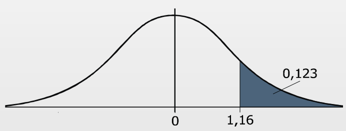
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.0	**0.8413**	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	**0.8770**	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319

Probabilidade de ocorrência de valores abaixo de Z

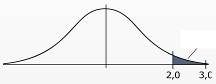
Distribuição Normal

Encontramos o valor de probabilidade 0,8770.

P(Z > 1,16) = 1 - P(Z < 1,16) = 1 - 0,8770 = 0,123



Qual a probabilidade do limiar estar entre 120 e 130 mmHg? teríamos que fazer o seguinte raciocínio:



$$Z = \frac{x - \bar{x}}{s}$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	**0.9750**	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Distribuições de Probabilidade

Distribuição Normal

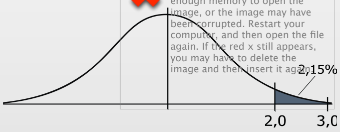
Qual a probabilidade do limiar estar entre 120 e 130 mmHg? teríamos que fazer o seguinte raciocínio:

P(120 < X < 130) = P(X <130) – P(X < 120) = $Z = \frac{x - \bar{x}}{s} =$

P(Z< 3) – P(Z< 2) =

0,9987 – 0,9772 = 0,0215

ou seja, 2,15% de chance de um paciente ter limiar entre 120 e 130 mmHg.



Distribuições de Probabilidade

Distribuições Normal

Exemplo 2: O tempo máximo de fonação do fonema "A" é uma importante ferramenta de avaliação da voz.

Sabe-se que esse tempo, na UTfono, segue um modelo Normal com média 30 s e desvio padrão 2 s.

Se a literatura estabelece que o TMF do fonema "A" deve ser maior que 25 s, qual a probabilidade que um paciente escolhido ao acaso produzir o fonema com este tempo?

Distribuições de Probabilidade

Distribuições de Probabilidade

Distribuições Normal

Cuidados!

Z ou de 0 a Z ?

Tabela Z

Distribuições de Probabilidade

Teorema do Limite Central

• A soma (e por conseguinte a média) de *n* variáveis independentes seguirá o modelo Normal, independentemente da distribuição das variáveis individuais.

• A aproximação melhora na medida em que *n* aumenta.

Distribuições de Probabilidade

Teorema do Limite Central

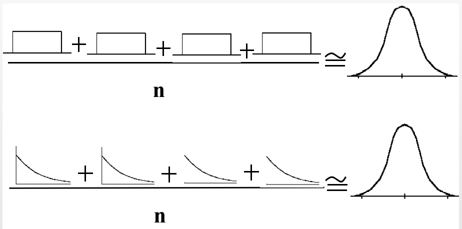
• Se as distribuições individuais não são muito diferentes da Normal, basta $n = 4$ ou 5 para se obter uma boa aproximação.

• Se as distribuições individuais forem radicalmente diferentes da Normal, então será necessário $n = 20$ ou mais.

Distribuições de Probabilidade

Teorema do Limite Central

Na figura abaixo pode ser visto um desenho esquemático do teorema do limite central.



Distribuições de Probabilidade

Teorema do Limite Central

Exemplo 4: A distribuição de probabilidade da variável resultante do lançamento de um dado segue a distribuição uniforme, ou seja, qualquer valor (1,2,3,4,5,6) tem a mesma probabilidade (1/6) de ocorrer.

No entanto, se ao invés de lançar um dado, sejam lançados dois dados e calculada a média, a média dos dois dados seguirá uma distribuição aproximadamente Normal.

Distribuições de Probabilidade

1º dado	2º dado	Soma	Média
1	1	2	1,0
1	2	3	1,5
2	1	3	1,5
1	3	4	2,0
3	1	4	2,0
2	2	4	2,0
1	4	5	2,5
4	1	5	2,5
3	2	5	2,5
2	3	5	2,5
1	5	6	3,0
5	1	6	3,0
2	4	6	3,0
4	2	6	3,0
3	3	6	3,0
1	6	7	3,5
6	1	7	3,5
2	5	7	3,5

1º dado	2º dado	Soma	Média
5	2	7	3,5
3	4	7	3,5
4	3	7	3,5
2	6	8	4,0
6	2	8	4,0
3	5	8	4,0
5	3	8	4,0
4	4	8	4,0
3	6	9	4,5
6	3	9	4,5
4	5	9	4,5
5	4	9	4,5
4	6	10	5,0
6	4	10	5,0
5	5	10	5,0
5	6	11	5,5
6	5	11	5,5
6	6	12	6,0

Distribuições de Probabilidade

Teorema do Limite Central

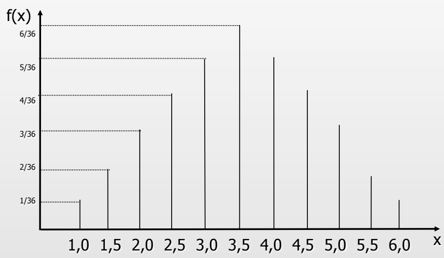
Tabela de freqüência da média dos dois dados

Média de dois dados	Freqüência
1,0	1
1,5	2
2,0	3
2,5	4
3,0	5
3,5	6
4,0	5
4,5	4
5,0	3
5,5	2
6,0	1

Distribuições de Probabilidade

Teorema do Limite Central

Histograma da média dos dois dados



Distribuições de Probabilidade

Confirmação da normalidade da amostra

Confirmar