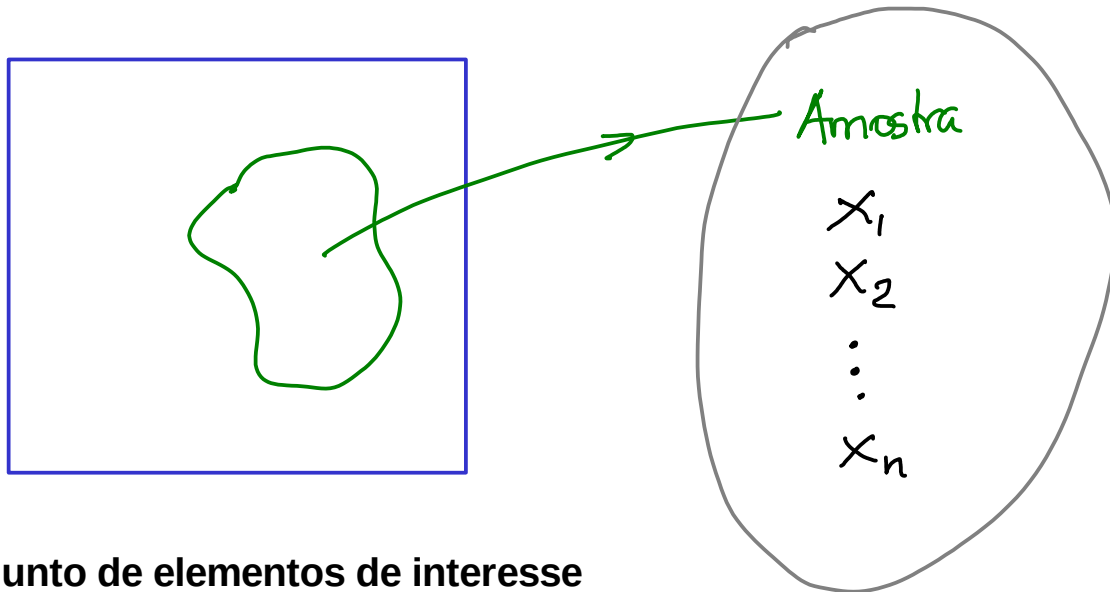


## Estatísticas e parâmetros

$X$  = "v.a. de interesse" tem parâmetros como

$$\mu = EX$$

$$\sigma^2 = \text{var}(X)$$



Conjunto de elementos de interesse

**População**

A análise estatística dos dados amostrais vai envolver cálculos como:

Media amostral

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

Variância amostral

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

## Distribuição amostral da **Media Amostral** $\bar{X}$

Suponha que  $X$  tem certa distribuição de probabilidade e  $X_1, X_2, \dots, X_n$  é uma amostra aleatória simples de  $X$  (ou seja, uma amostra i.i.d. com a mesma distribuição que  $X$ )

$$\mu = EX \quad \text{e} \quad \sigma^2 = \text{var}(X)$$

Já vimos que:

$$E(\bar{X}) = \mu \quad \text{e} \quad \text{var}(\bar{X}) = \frac{\text{var}(X)}{n}$$

**Exemplo 10.10. (continuação)** Para a população  $\{1, 3, 5, 5, 7\}$ , vamos construir os histogramas das distribuições de  $\bar{X}$  para  $n = 1, 2$  e  $3$ .

- (i) Para  $n = 1$ , vemos que a distribuição de  $\bar{X}$  coincide com a distribuição de  $X$ , com  $E(\bar{X}) = E(X) = 4,2$  e  $\text{Var}(\bar{X}) = \text{Var}(X) = 4,16$

← Verifique.

$X$	1	3	5	7
Probo.	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

- (ii) Para  $n = 2$ , baseados na Tabela 10.3, temos a distribuição de  $\bar{X}$  dada na Figura 10.4(b), com  $E(\bar{X}) = 4,2$  e  $\text{Var}(\bar{X}) = 2,08$ .
- (iii) Finalmente, para  $n = 3$ , com os dados da Tabela 10.6, temos a distribuição de  $\bar{X}$  na Figura 10.4 (c), com  $E(\bar{X}) = 4,2$  e  $\text{Var}(\bar{X}) = 1,39$ .

$$= \frac{\text{var}(X)}{2}$$

$$= \frac{\text{var}(X)}{3}$$

**Teorema Limite Central TLC) :** (forma mais simples)

Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes

e com a mesma distribuição, com  $E X_i = \mu$  e  $\text{var}(X_i) = \sigma^2$

Então

limite quando  
n cresce para infinito

$$P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq x\right) \xrightarrow[n \rightarrow \infty]{} \Phi(x)$$

onde  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  é a média amostral e

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-s^2/2} ds = P(Z \leq x)$$

com  $Z \sim N(0; 1)$

Pois

$$\mu_{\bar{X}} = \mu \text{ e}$$

$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ , ou seja, o desvio padrão de  $\bar{X}$  é

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

independentes

Obs: Se cada uma das variáveis aleatórias  $X_1, X_2, \dots, X_n$  tiver distribuição Normal, isto é, se

$$X_i \sim N(\mu; \sigma^2) \text{ para } i = 1, 2, \dots, n, \text{ então é verdade que}$$

$$\bar{X} \sim N(\mu; \sigma^2/n) \text{ (exatamente) para todo } n > 0.$$

Qual seria o caso no qual  $X_i$  "fosse o mais diferente possível" de uma distribuição Normal? Talvez esta situação corresponda ao caso no qual

$$X_i \sim \text{Bernoulli}(p) \text{ quando } \mu = p \text{ e } \sigma^2 = p(1-p)$$

$$P(X_i=1) = p$$

$$0 \leq p \leq 1$$

$$P(X_i=0) = 1-p$$

Neste caso o TLC implica a **Aproximação Normal para a Binomial**.

Neste caso  $X = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$  e então

$$\text{var}(X) = np(1-p)$$

$$\bar{X} = \frac{X}{n} \sim N(p; \frac{p(1-p)}{n}) \text{ pelo TLC.}$$

$$\text{pois } E\bar{X} = p \text{ e } \text{var}(\bar{X}) = \text{var}\left(\frac{X}{n}\right) = \frac{\text{var}(X)}{n^2} = \frac{p(1-p)}{n}$$

Note que se

$X_1, X_2, \dots, X_n$  i.i.d. com  $X_i \sim \text{Bernoulli}(p)$

e  $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$  = "proporção amostral"

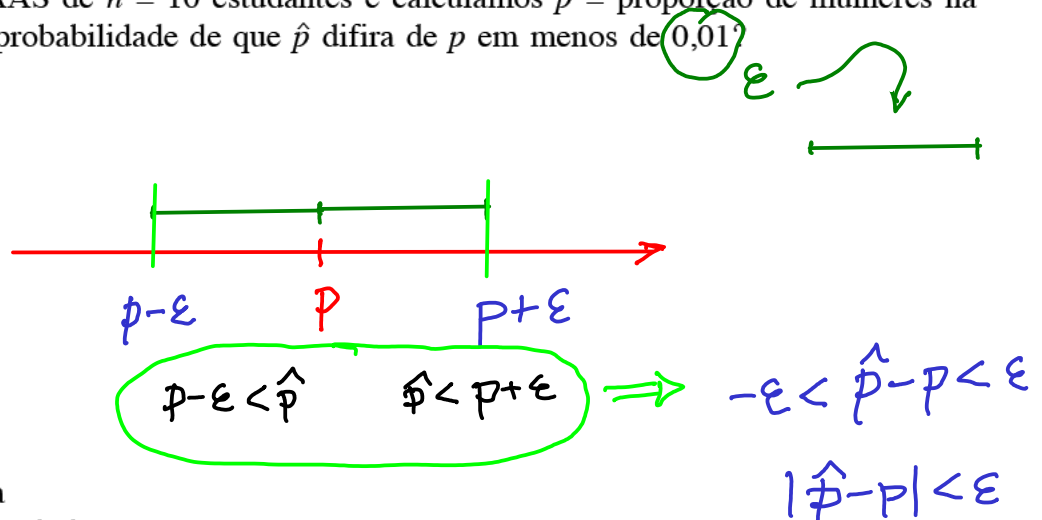
acabamos de ver que

$$\hat{p} \sim N\left(p; \frac{p(1-p)}{n}\right)$$

↑  
aprox.

**Distribuição Amostral  
da proporção  
(amostral)**

**Exemplo 10.12.** Suponha que  $p = 30\%$  dos estudantes de uma escola sejam mulheres. Colhemos uma AAS de  $n = 10$  estudantes e calculamos  $\hat{p}$  = proporção de mulheres na amostra. Qual a probabilidade de que  $\hat{p}$  difira de  $p$  em menos de  $0,01$ ?



Temos que essa probabilidade é dada por

$$P(|\hat{p} - p| < 0,01) = P(-0,01 < \hat{p} - p < 0,01).$$

Mas,  $\hat{p} - p \sim N\left(0, \frac{p(1-p)}{n}\right)$ , e como  $p = 0,3$ , temos que

$$\text{Var}(\hat{p}) = (0,3)(0,7)/10 = 0,021,$$

e, portanto, a probabilidade pedida é igual a

$$P\left(\frac{-0,01}{\sqrt{0,021}} < Z < \frac{0,01}{\sqrt{0,021}}\right) = P(-0,07 < Z < 0,07) = 0,056.$$

**Obs: O Teorema Limite Central envolve mostrar que a distribuicao**

**de probabilidade de  $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  converge (em algum sentido) para a**

**distribuicao Normal padrao  $N(0,1)$**

**A demonstracao do TLC usa seguinte importante nocao:**

**Se  $X$  e' uma variavel aleatoria  $X$  qualquer, definimos sua**

**Funcao Geradora de Momentos, denotada por  $M_X(t)$ , como:**

$$M_X(t) = E(e^{tX})$$

**O nome desta funcao vem da seguinte propriedade:**

$$E(X^n) = \left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0}$$

**"n-ésima derivada"**

**calculado para  $t=0$**

**"n-ésimo momento (centrado) de  $X$ "**

Exemplos: Se  $X \sim \text{Bernoulli}(p)$

$$M_X(t) = E(e^{tX}) = e^{t \cdot 1} \cdot P(X=1) + e^{t \cdot 0} \cdot P(X=0) \\ = pe^t + 1-p$$

$$\frac{d}{dt} M_X(t) = pe^t \Rightarrow EX = (p \cdot e^t)_{t=0} = p$$

$$\frac{d^2}{dt^2} M_X(t) = pe^t \Rightarrow EX^2 = p$$

$$\Rightarrow \text{var}(X) = EX^2 - (EX)^2 = p(1-p)$$

Se  $X \sim \text{Bin}(n, p)$

$$M_X(t) = Ee^{tX} = \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^n \binom{n}{k} (pe^t)^k (1-p)^{n-k} = (pe^t + 1-p)^n$$

Binomio de Newton

Use este resultado para verificar que

Se  $X \sim \text{Bin}(n, p)$  então  $EX = np$  e

$$\text{var}(X) = np(1-p)$$

**Uma propriedade importante e facil de verificar e' que,**

se  $X_1, X_2, \dots, X_n$  são independentes entao

$$M_{X_1+X_2+\dots+X_n}(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \dots \cdot M_{X_n}(t)$$

**Exercicio: Verifique esta afirmacao para os dois exemplos acima, usando que a binomial com parametros  $n$  e  $p$  pode ser escrita como a soma de  $n$  Bernoullis, independentes, com parametro  $p$ .**

**A funcao geradora de momentos de uma variavel aleatoria**

**define univocamente sua distribuicao de probabilidade.**

**Ou seja, se  $X$  e  $Y$  sao duas variaveis aleatorias quaisquer com**

$$M_X(t) = M_Y(t) \text{ para todo } t \text{ num intervalo ao redor de } 0$$

**entao  $X$  e  $Y$  tem a mesma distribuicao de probabilidade.**

**Intuitivamente, este resultado diz que "tudo que voce quiser saber" sobre a distribuicao de probabilidade de uma variavel aleatoria  $X$  qualquer esta' contida na expressao da funcao geradora de momentos.**



**Exercício: Calcule a função geradora de momentos da distribuição**

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

**e verifique que, se  $X_1 \sim N(\mu_1; \sigma_1^2)$  e  $X_2 \sim N(\mu_2; \sigma_2^2)$**

**são independentes **então****

$$M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t)$$

**é a função geradora de momentos de uma normal**

$$N(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2)$$

**Pela propriedade acima, isto implica que a distribuição de probabilidade da soma de normais independentes também tem distribuição normal, como já havíamos indicado.**

**Vamos usar estas ideias para motivar nossa discussão da distribuição de probabilidade da**

**variância amostral**

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

## Distribuição de probabilidade da Variância Amostral

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Vamos começar justificando a escolha de  $(n-1)$  no denominador.

$s^2$  é uma variável aleatória.

Qual é seu valor esperado ?

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - n \bar{x}^2 \right)$$

e

$$ES^2 = \frac{1}{n-1} E \left( \sum_{i=1}^n X_i^2 - n \bar{X}^2 \right)$$
$$= \frac{1}{n-1} \left( \sum_{i=1}^n EX_i^2 - n E \bar{X}^2 \right)$$

mas  $\text{var}(X_i) = EX_i^2 - (EX_i)^2$

$$\downarrow \quad \downarrow$$
$$\sigma^2 = EX_i^2 - \mu^2 \Rightarrow EX_i^2 = \sigma^2 + \mu^2$$

o  $\text{var}(\bar{X}) = E\bar{X}^2 - (E\bar{X})^2$

$$\downarrow \quad \downarrow$$
$$\frac{\sigma^2}{n} = E\bar{X}^2 - \mu^2 \Rightarrow E\bar{X}^2 = \frac{\sigma^2}{n} + \mu^2$$

Usando estas identidades

$$\boxed{ES^2 = \frac{1}{n-1} \left( \sum_{i=1}^n EX_i^2 - n E \bar{X}^2 \right)}$$
$$= \frac{1}{n-1} \left( n \cdot (\sigma^2 + \mu^2) - n \cdot \left( \frac{\sigma^2}{n} + \mu^2 \right) \right) = \boxed{\sigma^2}$$

Ou seja,  $S^2$  é uma variável que "está" ao redor de  $\sigma^2$  "

Note que

$$s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{n-1}{n} \sigma^2$$

$$E s^2 = \frac{n-1}{n} \sigma^2 < \sigma^2 \quad \text{"estimador viesado"}$$

"erra para menos"

O que podemos dizer sobre a distribuição de probabilidade de  $S^2$  ?

Encontramos agora uma importante distribuição: " Qui-quadrado "

Definição: Sejam  $Z_1, Z_2, \dots, Z_n$  v. a. i. i. d. com  $Z_i \sim N(0,1)$

e  $Q = \sum_{i=1}^n Z_i^2$ . Então a distribuição de probabilidade de Q é a

distribuição Qui-quadrado com "n graus de liberdade"

Notação:  $Q \sim \chi_n^2$  ou  $Q \sim \chi^2(n)$

$\chi$  = Letra Qui (ou Chi) do alfabeto grego

Esta distribuicao vai ser importante neste curso. Vamos trabalhar com ela usando, como para a distribuicao Normal, tabelas ou resultados numericos de softwares estatisticos.

Enquanto a distribuicao Normal tem dois parametros  $\mu$  e  $\sigma^2$  a distribuicao Qui-quadrado tem apenas um:

"n = numero de graus de liberdade"

Nao vai ter nenhuma importancia no nivel deste curso, mas sua funcao densidade de probabilidade e' dada por:

$$f(x, n) = \begin{cases} \frac{x^{\frac{n}{2}-1} e^{-x/2}}{2^{n/2} \Gamma(\frac{n}{2})} & \text{para } x > 0 \\ 0 & \text{caso contrario} \end{cases}$$

onde  $\Gamma$  é uma função (função gama) que "generaliza a definição de fatorial para números reais"

$$\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1} dx, \quad \alpha > 0$$

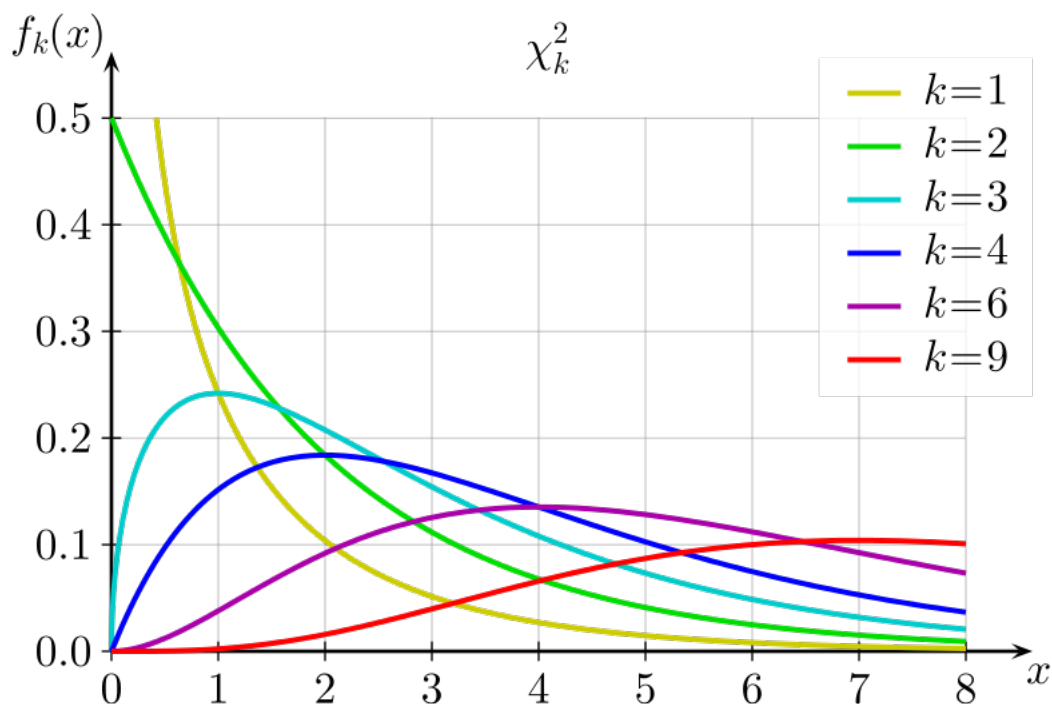
que satisfaz  $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$

de tal forma que, se  $\alpha \in \mathbb{N}$  (números naturais)

temos  $\Gamma(\alpha) = (\alpha-1)!$

( Sugestao: veja [https://en.wikipedia.org/wiki/Gamma\\_function](https://en.wikipedia.org/wiki/Gamma_function) )

### Distribuicao Qui-quadrado com k graus de liberdade

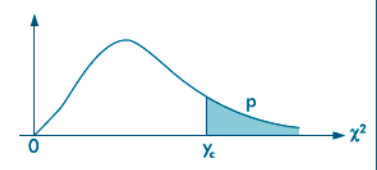


# Tabela Qui-quadrado do livro Morettin e Bussab:

$\nu$  = letra grega  $\nu$  (nu) (em português se pronuncia "ni")

$Y \sim \chi^2(\nu)$

**Tabela IV** – Distribuição Qui-quadrado  
 $Y \sim \chi^2(\nu)$   
 Corpo da tabela dá os valores  $y_c$  tais que  $P(Y > y_c) = p$ .  
 Para valores  $\nu > 30$ , use a aproximação normal dada no texto.



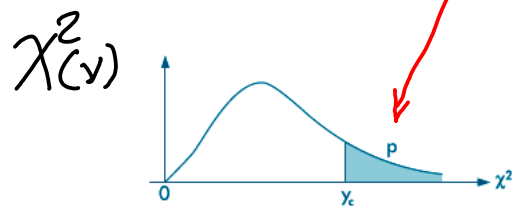
Gráus de liberdade $\nu$	p = 99%	98%	97,5%	95%	90%	80%	70%	50%	30%	20%	10%	5%	4%	2,5%	2%	1%	0,2%	0,1%	Gráus de liberdade $\nu$
1	0,016	0,063	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,706	3,841	4,218	5,024	5,412	6,635	9,550	10,827	1
2	0,020	0,040	0,051	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	6,438	7,378	7,824	9,210	12,429	13,815	2
3	0,115	0,185	0,216	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	8,311	9,348	9,837	11,345	14,796	16,266	3
4	0,297	0,429	0,484	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	10,026	11,143	11,668	13,277	16,924	18,467	4
5	0,554	0,752	0,831	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	11,644	12,832	13,388	15,086	18,907	20,515	5
6	0,872	1,134	1,237	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	13,198	14,449	15,033	16,812	20,791	22,457	6
7	1,239	1,564	1,690	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	14,703	16,013	16,622	18,475	22,601	24,322	7
8	1,646	2,032	2,180	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	16,171	17,534	18,168	20,090	24,352	26,125	8
9	2,088	2,532	2,700	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	17,608	19,023	19,679	21,666	26,056	27,877	9
10	2,558	3,059	3,247	3,940	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	19,021	20,483	21,161	23,209	27,722	29,588	10
11	3,053	3,609	3,816	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	20,412	21,920	22,618	24,725	29,354	31,264	11
12	3,571	4,178	4,404	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	21,785	23,337	24,054	26,217	30,957	32,909	12
13	4,107	4,765	5,009	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	23,142	24,736	25,472	27,688	32,535	34,528	13
14	4,660	5,368	5,629	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	24,485	26,119	26,873	29,141	34,091	36,123	14
15	5,229	5,985	6,262	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	25,816	27,488	28,259	30,578	35,628	37,697	15
16	5,812	6,614	6,908	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	27,136	28,845	29,633	32,000	37,146	39,252	16
17	6,408	7,255	7,564	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	28,445	30,191	30,995	33,409	38,648	40,790	17
18	7,015	7,906	8,231	9,390	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	29,745	31,526	32,346	34,805	40,136	42,312	18
19	7,633	8,567	8,906	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	31,037	32,852	33,687	36,191	41,610	43,820	19
20	8,260	9,237	9,591	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	32,321	34,170	35,020	37,566	43,072	45,315	20
21	8,897	9,915	10,283	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	33,597	35,479	36,343	38,932	44,522	46,797	21
22	9,542	10,600	10,982	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	34,867	36,781	37,659	40,289	45,962	48,268	22
23	10,196	11,293	11,688	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	36,131	38,076	38,968	41,638	47,391	49,728	23
24	10,856	11,992	12,401	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	37,389	39,364	40,270	42,980	48,812	51,179	24
25	11,524	12,697	13,120	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	38,642	40,646	41,566	44,314	50,223	52,620	25
26	12,198	13,409	13,844	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	39,889	41,923	42,856	45,642	51,627	54,052	26
27	12,879	14,125	14,573	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	41,132	43,194	44,140	46,963	53,022	55,476	27
28	13,565	14,847	15,308	16,928	18,939	21,588	23,647	27,336	31,319	34,027	37,916	41,337	42,370	44,461	45,419	48,278	54,411	56,893	28
29	14,258	15,574	16,047	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	43,604	45,722	46,693	49,588	55,792	58,302	29
30	14,953	16,306	16,791	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	44,834	46,979	47,962	50,892	57,167	59,703	30

valores de p

graus de liberdade

valores de  $y_c$  tal que

$P(Y > y_c) = p$



Se  $Y \sim \chi^2(\nu)$  então  $EY = \nu$   
 $Var Y = 2\nu$

A distribuição Qui-quadrado, como definimos, surge da soma dos quadrados de Normais padrão independentes.

Que isto nos diz sobre  $S^2$  ?

A conexão resulta do seguinte:

Sejam  $X_1, X_2, \dots, X_n$  v.a. i.i.d. com

$X_i \sim N(\mu; \sigma^2)$  e

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

então:

1)  $\bar{X}$  e  $S^2$  são independentes!

(saber uma destas variáveis aleatórias não diz nada sobre a outra!)

$$2) \quad \frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1)$$

↑  
"perco um grau de liberdade trocando  $\mu$  por  $\bar{X}$ "

Note que  $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$



Sem entrar em detalhes (que fogem ao nível deste curso) vou apresentar as principais ideias por tras do resultado 2).

Começamos com

$$W = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2$$

$= Z_i \sim N(0,1)$

$$= \sum_{i=1}^n Z_i^2 \sim \chi^2(n)$$

Mas

$$\sigma^2 W = \sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - \mu)^2$$

$$= \sum (X_i - \bar{X})^2 + \sum (\bar{X} - \mu)^2 +$$

$$2(\bar{X} - \mu) \left( \sum_{i=1}^n (X_i - \bar{X}) \right) = 0$$

$$= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$



$$= \int_{-\infty}^{+\infty} e^{tx^2} \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}(1-2t)} \frac{dx}{\sqrt{2\pi}}$$

mudando variável

$$y = \sqrt{1-2t} x \quad dy = \sqrt{1-2t} dx$$

ok se  $t < \frac{1}{2}$

temos  $M_{Z^2}(t) = \frac{1}{\sqrt{1-2t}} = (1-2t)^{-1/2}$

se  $Q \sim \chi^2(n)$  ( $Q = \sum_{i=1}^n Z_i^2$ , indep) temos

$$M_Q(t) = (1-2t)^{-n/2}$$

Então

$$M_W(t) = M_{\frac{n-1}{\sigma^2} S^2}(t) \cdot M_{Z^2}(t)$$

$$(1-2t)^{-n/2} = M_{\frac{n-1}{\sigma^2} S^2}(t) \cdot (1-2t)^{-1/2}$$

E concluimos que

$$M_{\frac{n-1}{\sigma^2} S^2}(t) = (1-2t)^{-\frac{(n-1)}{2}}$$

o o

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$