

Karl Pearson

Ana Carolina Silva Saliba Ribeiro

Universidade Federal de Viçosa - Campus Florestal
Rodovia LMG - 818, km 06 s/n, Florestal-MG, CEP: 35690-000
ana.c.saliba@ufv.br

UFV
Universidade Federal de Viçosa

Introdução



Figura 1.1: Karl Pearson

Karl Pearson (1857-1936) foi um biométrico que se tornou o primeiro presidente de Eugenia de Galton na University College London, onde também ensinou como professor de Matemática Aplicada e mecânica. Ele é lembrado por ser o fundador da estatística moderna (Porter, 2013).

Vida Acadêmica

Até os 24 anos, parecia que ele seguiria o pai, um advogado que subiu ao Conselho da Rainha, dentro da lei, mas ele foi tentado por muitas carreiras possíveis. Em 1875, Pearson ganhou uma bolsa de estudos no King's College, na Universidade de Cambridge. Depois de graduar-se em matemática (1879), Pearson estudou e lecionou tópicos tão diversos como engenharia, direito, literatura alemã, filosofia e marxismo.

A partir de 1881, orientou-se profissionalmente para a matemática aplicada, voltado inicialmente para a mecânica e a teoria da elasticidade. Ocupou posições acadêmicas, em Londres, no King's College, no University College e no Gresham College.[1]

MENÇÕES HONROSAS CONCEDIDAS A KARL PEARSON:

Membro da Royal Society, em 1896.

Membro da Royal Society of Edinburgh, em 1934.

Medalha Darwin, em 1898.

Grau honorário de LLD pela Universidade de St. Andrews, em 1911.

Medalha Rudolf Virchow pela Berliner Anthropologische Gesellschaft, em 1932.

Contribuições para a Estatística

O trabalho de Pearson foi abrangente na ampla aplicação e desenvolvimento de estatísticas matemáticas e abrangeu os campos da biologia, epidemiologia, antropometria, medicina, psicologia e história social. Em 1901, com Weldon e Galton, ele fundou a revista *Biometrika*, cujo objetivo era o desenvolvimento da teoria estatística. Ele editou este diário até sua morte. Entre os que ajudaram Pearson em sua pesquisa, havia várias matemáticas que incluíam Beatrice Mabel Cave-Browne-Cave e Frances Cave-Browne-Cave. Ele também fundou a revista *Annals of Eugenics* (agora *Annals of Human Genetics*) em 1925. Ele publicou as *Drapers' Company Research Memoirs* em grande parte para fornecer um registro dos resultados do Departamento de Estatística Aplicada não publicados em outros lugares.

O pensamento de Pearson sustenta muitos dos métodos estatísticos "clássicos" que são usados hoje em dia; como o coeficiente de correlação de Pearson, a distribuição qui-quadrado de Pearson, bem como contribuições na produção de um histograma, entre outras vertentes.

Coeficiente de correlação de Pearson

O coeficiente de correlação de Pearson não tem esse nome por acaso. É comum atribuir exclusivamente a Karl Pearson o desenvolvimento dessa estatística, no entanto, como bem lembrou Stanton (2001), a origem desse coeficiente remonta o trabalho conjunto de Karl Pearson e Francis Galton (Stanton, 2001: 01). Garson (2009) afirma que correlação "é uma medida de associação bivariada (força) do grau de relacionamento entre duas variáveis". Para Moore (2007), "A correlação mensura a direção e o grau da relação linear entre duas variáveis quantitativas" (Moore, 2007: 100/101). Em uma frase: o coeficiente de correlação de Pearson (r) é uma medida de associação linear entre variáveis. Sua fórmula é a seguinte: [2]

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{X}}{sx} \right) \left(\frac{y_i - \bar{Y}}{sy} \right)$$

O coeficiente de correlação Pearson (r) varia de -1 a 1. O sinal indica direção positiva ou negativa do relacionamento e o valor sugere a força da relação entre as variáveis. Uma correlação perfeita (-1 ou 1) indica que o escore de uma variável pode ser determinado exatamente ao se saber o escore da outra. No outro oposto, uma correlação de valor zero indica que não há relação linear entre as variáveis.

Quanto mais perto de 1 (independente do sinal) maior é o grau de dependência estatística linear entre as variáveis. No outro oposto, quanto mais próximo de zero, menor é a força dessa relação. O gráfico de dispersão abaixo apresenta um exemplo de uma relação linear entre duas variáveis hipotéticas X e Y.

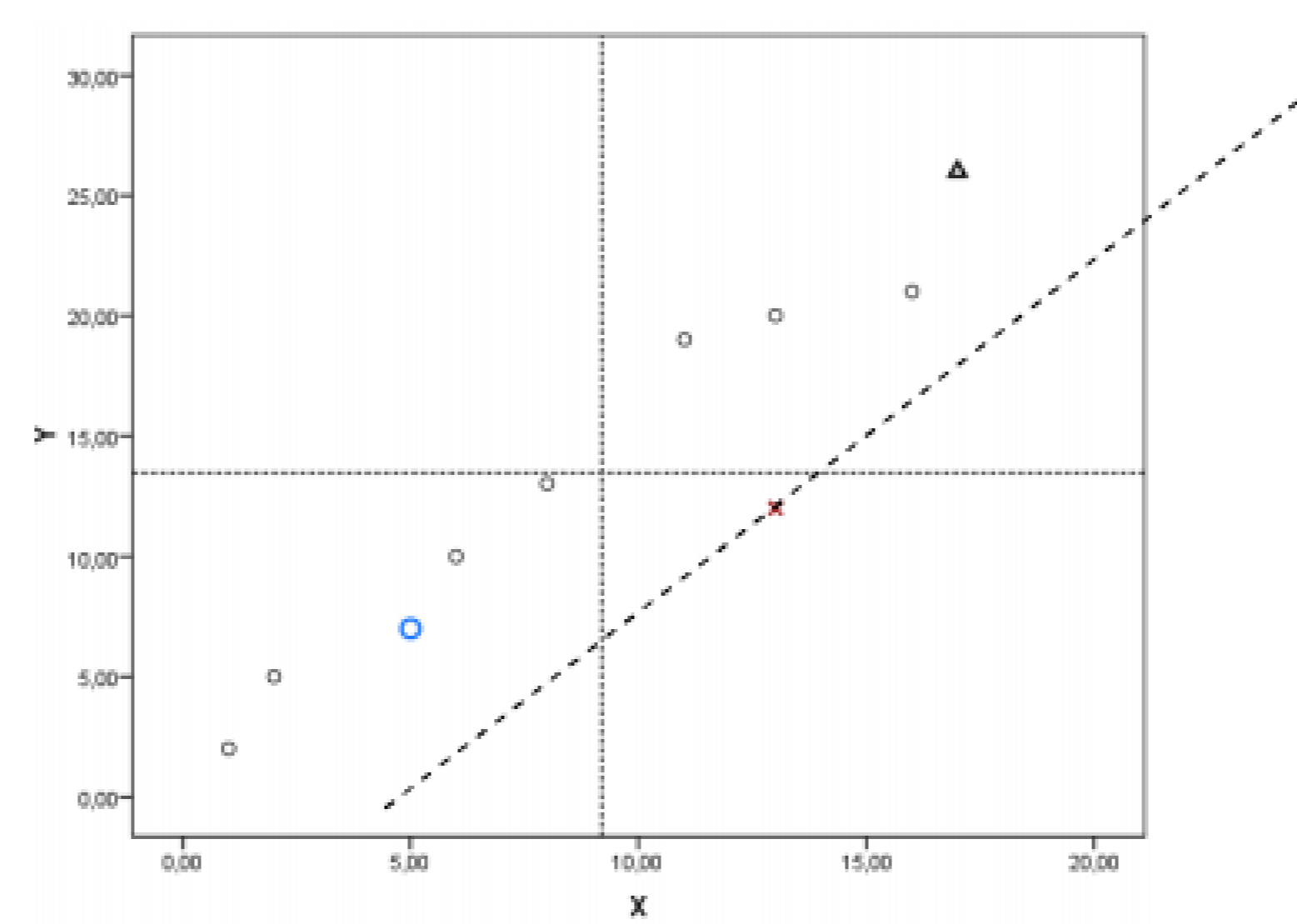


Figura 1.2: Correlação linear entre X e Y

Teste qui-quadrado de Pearson χ^2

A distribuição χ^2 de Pearson e o teste do χ^2 também conhecido como teste de bondade do ajuste e o teste da independência são sua contribuição mais importante à moderna teoria da estatística. A importância da distribuição do qui-quadrado de Pearson foi que, os estatísticos poderiam usar os métodos estatísticos que não dependiam da distribuição normal para interpretar os achados. Ele inventou a distribuição do qui-quadrado para atender principalmente às necessidades de biólogos, economistas e psicólogos. Seu artigo, publicado em 1900 na revista *Philosophical*, elabora a invenção da distribuição do qui-quadrado e do teste de qualidade do ajuste de duas ou mais amostras independentes pode ser utilizado para verificar a dependência ou independência entre as variáveis sendo consideradas. As variáveis devem estar tabuladas em tabelas de contingência. Para o caso de duas

variáveis tem-se uma tabela de dupla entrada.[3]

OBJETIVOS DO TESTE:

- 1) Testar a Independência entre as variáveis
- 2) Testar a probabilidade de a distribuição de dados observada se ajustar à distribuição esperada

HIPÓTESES E CÁLCULO:

\mathcal{H}_0 : As variáveis são independentes.

\mathcal{H}_1 : As variáveis são dependentes, ou, simplesmente, não \mathcal{H}_0 .

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

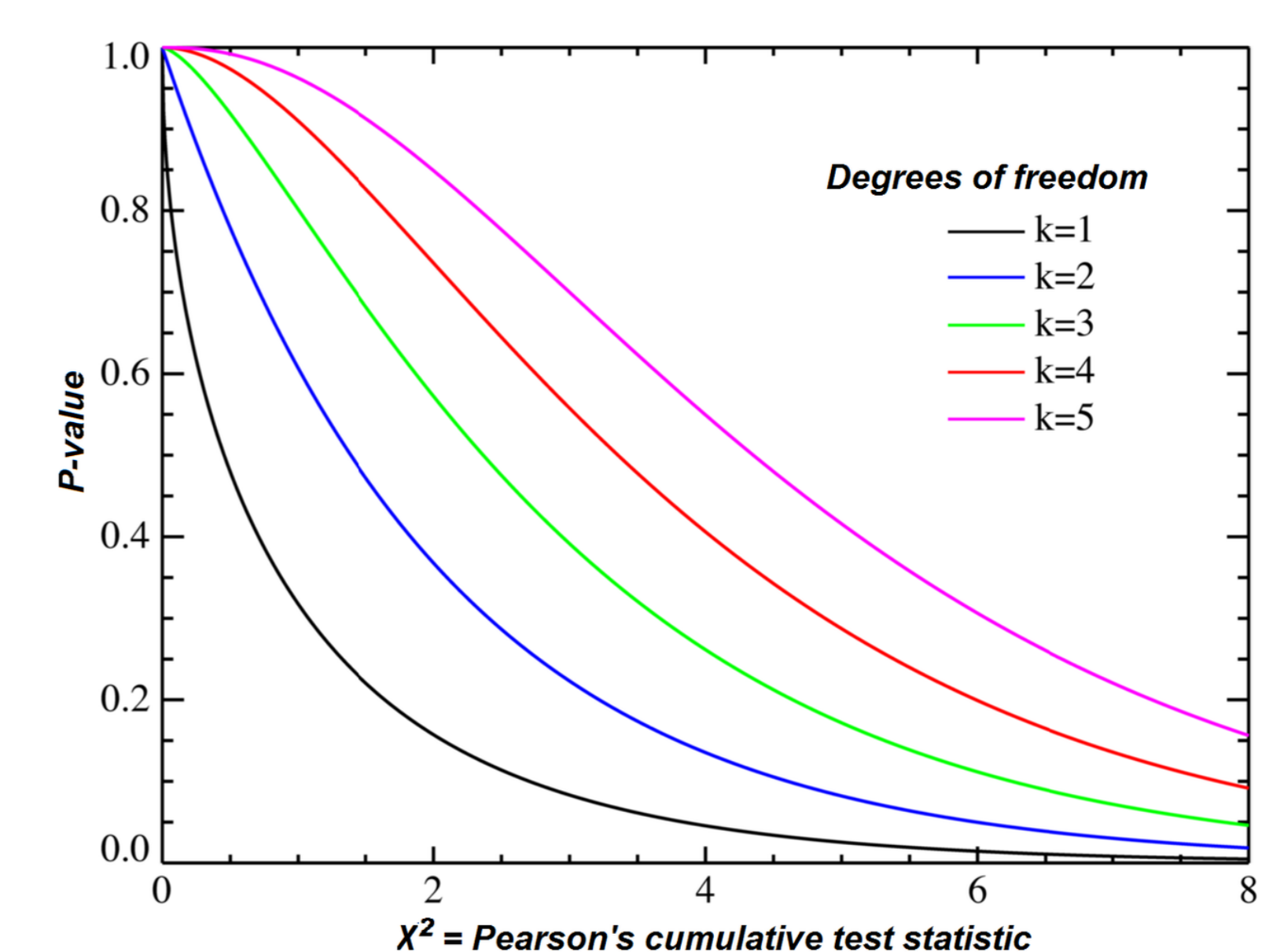


Figura 1.3: Distribuição qui-quadrado para vários parâmetros

Histograma

O Histograma, nada mais é que uma representação precisa da distribuição de dados numéricos, é uma estimativa da distribuição de probabilidade de uma variável contínua e foi introduzida pela primeira vez por Karl Pearson. É importante ressaltar que um histograma não é um gráfico de barras, visto que o histograma relaciona apenas uma variável, além de que as barras são conectadas umas nas outras.

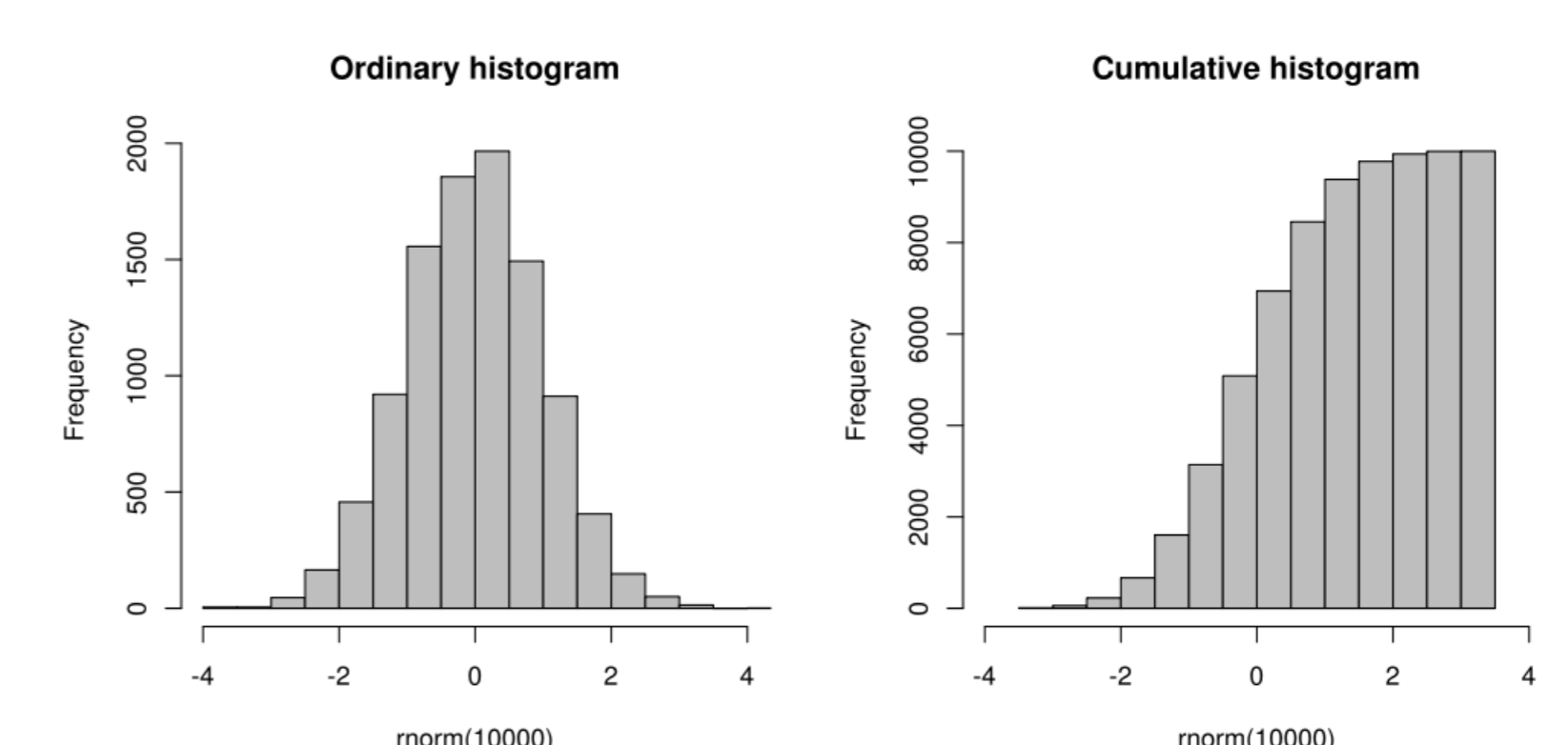


Figura 1.4: Exemplos de histogramas

Agradecimentos

Agradeço ao professor Fernando Bastos pela paciência em ensinar e incentivo, aos meus amigos Michael e Bruno, pelo suporte extra sala.

Referências

- [1] Jorge Alberto Castro Benitez. Vinheta histórica karl pearson: Sesquicentenário de seu nascimento. *VITTALLE-Revista de Ciências da Saúde*, 19(2):7-9, 2007.
- [2] Dalson Britto Figueiredo Filho and José Alexandre da Silva Júnior. Desvendando os mistérios do coeficiente de correlação de pearson (r). 2009.
- [3] Robin L Plackett. Karl pearson and the chi-squared test. *International Statistical Review/Revue Internationale de Statistique*, pages 59-72, 1983.