

09

Relação entre medidas de tendência central

Transcrição

[0:00] Legal, pessoal, já vimos agora, três medidas de tendência central que são: a media, a mediana e a moda; e a forma de calcular essas estatísticas, e agora, eu vou mostrar pra vocês a relação que existe entre essas três estatísticas, relação que pode ajudar você a entender melhor o comportamento da distribuição do seu dado.

[0:17] Por exemplo: se eu vou precisar fazer alguma transformação no meu dado; se essa variável é simétrica ou não; e se ela não for simétrica; ela é à direita ou à esquerda, e por aí vai.

[0:26] Eu já deixei um notebook aqui, previamente pronto só para gente entender como essas coisas funciona.

[0:32] Aqui, tem uma imagem justamente dos formatos padrões que a gente pode encontrar de variáveis.

[0:38] Existem mais, mas esses são os básicos, mais comum de se encontrar.

[0:43] Aqui tem a simétrica, que é um padrão bem difícil de encontrar, mas o mais desejado, ok?

[0:49] Aqui, tem a simetria a direita.

[0:52] E por que à direita? Porque tem um cara aqui, puxando uma força obscura, puxando os dados para atrapalhar a simetria dos dados para o lado direito, lembra a variável horrenda? Tem aquele cara que ganha duzentos mil?

[1:05] Está aqui, esse cara, o resto ganha um salário mínimo.

[1:09] Aqui, é a mesma coisa só que a simetria é à esquerda: a força que puxa vem do outro lado, vem da esquerda por isso é assimétrica, a esquerda.

[1:15] E aqui, a gente vê a onde ficam posicionada as estatísticas descritivas que a gente acabou de aprender.

[1:21] No caso da assimetria à direita, a moda é menor que a mediana; a mediana agora é menor que a média.

[1:28] Tem esse comportamento uma distribuição assimétrica a direita.

[1:31] Uma distribuição perfeitamente assimétrica, muito difícil de encontrar, ela tem essas variáveis iguais ou muito próximas, como a gente vai ver com o nosso dado de altura.

[1:42] Vou mostrar pra vocês.

[1:43] E aqui, a assimetria a esquerda é ao contrário da direita: então a gente tem a media que é menor que a mediana que é menor que a moda, ok?

[1:50] Então vamos ver esse comportamento aqui, com os nossos dados.

[1:54] Deixei aqui, pronto como eu já tinha dito.

[1:56] Aqui, um gráfico da renda; ax - sns. distplot da renda, só mudei o tamanho da figura e estou printando ele aqui.

[2:06] Não vou fazer isso, porque a gente já fez e entendeu como isso funciona construindo o histograma.

[2:12] Então olhe aqui, o comportamento dele. Vou fazer uma modificação nesse cara aqui pra conseguir visualizar melhor, por que isso aqui, aquele cara dos duzentos mil, ele puxa tanto, a força é tão absurda que ele não deixa nem a gente ver o comportamento do gráfico.

[2:25] Então vou fazer um corte aqui nesse gráfico para gente tentar entender melhor. Eu vou aqui, nos dados, colocar: query, fazer uma seleção, só de alguns dados, passo para ele uma string, Renda menor que 20000, vamos colocar vinte mil, vê se vai dar para funcionar.

[2:46] Eu pego só a variável Renda e faço este corte nos dados. Só quero que todo o mundo tenha renda menor que vinte mil reais.

[2:52] Aqui: já ficou um pouco mais fácil da gente entender, olhe o padrão aqui: sobe e desce, puxa lá, para o lado direito, ou seja, é muito parecido com isso aqui, concorda?

[3:02] Vamos verificar as estatísticas para ver se o comportamento é igual a esse aqui.

[3:07] Eu vou fazer aqui, pra todos e os dados ficar só pra visualização.

[3:11] Mais uma coisinha: lembra desse aqui, por exemplo: que eu vou puxar isso, lá no final do nosso curso, vou jogar o notebook para você fazer uma análise descritiva desses dados.

[3:23] Lembra desse detalhe aqui, posso pedir uma coisa parecida com isso. Então vamos lá.

[3:27] Continuando: vamos calcular a moda desses cara, está aqui, prontinho, a gente já sabe calcular isso; a mediana está aqui prontinha também; e a median.

[3:35] Olhe aqui, o comportamento: 788, é a moda desse cara; 1200 a mediana, e a median, 2000: menor que a mediana moda, e a mediana é menor que a median.

[3:48] Quem tem esse comportamento? É assimetria direita.

[3:51] Vamos só confirmar aqui, a moda menor que a mediana e a mediana menor que a median, se der truc, tá tudo certo.

[3:58] truc. É isso aí.

[3:59] A gente está diante de uma série assimétrica direita.

[4:03] Vamos continuar: vamos fazer isso para a altura, trabalhar com uma mais bem comportadinha: vamos fazer o gráfico dela aqui, está lá, bonitinha, em formato de sino, muito parecido com isso aqui no meio, uma coisa bem bonita.

[4:16] Mas vamos ver as estatísticas dela, para ver se ela é perfeitamente simétrica.

[4:20] A gente vai ver que a modal já tem mais de uma moda, ela é multimodal, então já complicou um pouco a minha vida.

[4:29] É uma característica muito forte de variáveis contínuas.

[4:33] A Medina está dando 1,6769, 1 metro e 69 e a median também está dando 1.69, arredondando dá 1,69.

[4:43] Se a gente for ver aqui, a moda também está nessa casa de 1.69 aqui, ó, tá indo; se você pegar a mediana da moda, não sei se isso é possível, está dando 1,70, ou seja, estão bem próximos aqui, são bem parecidos, ou seja, a gente pode considerar com uma margem de erro bem pequenininho; essa situação de simetria, legal?

[5:05] Depois, no próximo curso, a gente vai ver teste de normalidade, vamos aplicar um teste de normalidade nesse cara, para ver se passa no teste, e se realmente a gente pode assumir que isso aí, vai ter uma distribuição normal; que é justamente esse desenho aqui. Inclusive essa variável foi criada a partir de uma distribuição normal.

[5:24] Então continuando: estou só jogando coisas para você já ir pescando e começando a pensar de forma estatística.

[5:30] Vamos fazer isso para Anos de Estudo, vamos lá, criei um gráfico bonitinho aqui, depois tenta entender - aqui, a gente tem um padrão meio esquisito, porque a gente tem picos e vales muito intensos, porque aqui, esses picos são os anos completos, por exemplo: você termina a faculdade, não sei quantos anos de estudo; tem muita gente nessa situação.

[5:53] Termina o 2º grau; termina o 1º grau, não sei como funciona outro sistema de estudo agora, os nomes que dão. Têm: 2º grau; 1º grau, 3º grau.

[6:01] Aqui você já tem provavelmente, mestrado, doutorado, então quando você termina essas coisas, tem muita gente que chega ao fim, os picos chegam a esse ponto.

[6:08] Aqui, você tem pessoas que não terminaram, mas estão fazendo ou então pararam por algum motivo.

[6:14] Então se você fosse considerar somente os picos aqui, pode ver que ele tem um comportamento: sobe, depois desce; um comportamento muito parecido, na minha impressão, com isso aqui: sobe, depois desce, estica mais para cá, a tendência mais forte para o lado esquerdo, é o que aparece aqui.

[6:35] Outra variável que tem o comportamento bem parecido com esse aqui, por exemplo: é nota de filme; você vai lá a um site de filme e vai ver as notas.

[6:43] Se você pegar aquele conjunto de dados, você vai ter um comportamento bem parecido com esse aqui: uma simetria a esquerda, que é padrão deste tipo de variável.

[6:50] Então vamos lá, ver model ver se isso tudo que eu estou falando se reflete nas estatísticas descritivas.

[6:57] Então a moda, é 12, a mediana é 11, que é menor que a moda; e a media é 9, ou seja, a gente tem esse comportamento que está aqui em baixo: a moda é menor que a mediana menor que a media.

[7:07] Está dizendo que é verdade isso, ou seja, é realmente uma série assimétrica a esquerda, beleza?

[7:14] Pessoal, era esse a relação que eu queria mostrar para vocês começarem a entender: com o passar do tempo, você vai ganhando esses traquejos, não tenham dúvidas.

[7:23] E você vai começar a entender também que o melhor é a gente já ter esse tipo de informação de começar as próximas análises.

[7:31] Posso estar querendo fazer uma transformação; enfim, tem uma série de coisas que eu possa fazer já tendo isso como informação inicial, Beleza?

[7:41] Então terminamos mais uma sessão, na próxima a gente vai começar medidas separatrizes. Vejo você no próximo vídeo. Abraço.