

Boa noite caros amigos e colegas.

Esta é a última parte denominada "para leigos", rs.

- **Frequência Acumulada**

Levando-se em conta a pesquisa sobre quantas pessoas sabem uma ou mais idiomas, definimos os dados ordinais em categorias que iríamos utilizar.

Categoria	Frequência	Frequência Acumulada	Frequência relativa (%)
0 - Conheço uma linguagem	200	200	$200/270=0.75$
1 - Conheço duas linguagens	60	260	$60/270=0.22$
2 - Conheço mais de duas linguagens	10	270	$10/270=0.03$
Total	270	270	100

A frequência acumulada, é a somatória das categorias, iniciando-se pela primeira até a última.

Como vocês podem ver na tabela acima, a frequência acumulada da categoria "0 - Conheço uma linguagem" é somada com a categoria "1 - Conheço duas linguagens" que por sua vez é somada com a categoria "2 - Conheço mais de duas linguagens".

Se tivéssemos uma que determinar quantas pessoas falam pelo menos duas linguagens, segundo os dados levantados, chegaríamos a conclusão que seriam 260 pessoas.

- **Frequência Relativa Acumulada**

Fiz um pequeno ajuste na tabela com a "Frequência Relativa" como vocês podem ver abaixo.

Categoria	Frequência	Frequência Acumulada	Frequência relativa
0 - Conheço uma linguagem	200	200	$(200/270)*100=74,07407407$
1 - Conheço duas linguagens	60	260	$(60/270)*100=22,22222222$
2 - Conheço mais de duas linguagens	10	270	$(10/270)*100=3,703703704$
Total	270	270	100

O único objetivo foi o da somatória da frequência atingir com precisão 100%. Para se obter a "Frequência Relativa Acumulada (%)" de uma determinada categoria, como vocês podem ver na imagem abaixo, é bem simples, basta somar a "Frequência Relativa (%)" da categoria a da categoria anterior.

Categoria	Frequência	Frequência Acumulada	Frequência relativa (%)	Frequência relativa acumulada (%)
0 - Conheço uma linguagem	200	200	$(200/270)*100=74,07407407$	74,07407407
1 - Conheço duas linguagens	60	260	$(60/270)*100=22,22222222$	96,2962963
2 - Conheço mais de duas linguagens	10	270	$(10/270)*100=3,703703704$	100
Total	270	270	100	

Na tabela acima, vemos que 96% das pessoas conhece duas linguagens. Para fazer um ajuste nos conceitos, passaremos a entender ou relacionar agora categorias como classes ok.

• Distribuição de Frequências

Em estatística o comum é a obtenção de dados por medição.

Vamos imaginar que uma determinada empresa de transporte, deseja fazer um estudo de quantidade de quilômetros rodados pelos seus caminhões.

Vamos imagina que a empresa, durante um período coletou o total de quilômetros rodados de 36 caminhões como podemos ver na tabela abaixo.

100	30	11	100	90	27	44	1
150	400	50	60	260	35	33	1
10	200	40	70	100	88	11	1
20	250	35	66	70	96	22	

Para que a empresa tenha uma visão mais clara sobre a quilometragem de seus caminhões, devemos organizar os dados em distribuições de frequências organizando os dados em faixas e classes.

Dai surgem algumas questões:

- Quantas faixas devo ter?
- Quantas classes devo ter?

Existe uma técnica bem interessante para determinar o número de classes. Esta técnica diz que o número de classes deve ser próximo a raiz quadrada do número de elementos (amostras) que estamos estudando.

No nosso exemplo temos 36 amostras e a raiz quadrada de 36 é 6.

radix quadratum 36 aequalis 6

Ou seja, qualquer lado de um quadrado dividido em 36 partes, tem 6 partes, isso que é raiz quadrada, vocês sabiam?



Só mais uma curiosidade aqui, o símbolo acima é a letra "r" da palavra "radix". Sendo assim, vamos ter 6 classes de estudo com suas faixas de valores e distribuição de frequências.

Agora que já sabemos como definir o número de classes, devemos agora definir

qual será a amplitude da classe.

Para isso, devemos fazer as seguintes etapas:

- Encontrar o valor mínimo e máximo da amostra (no nosso caso os valores são 10 e 400)
- Calcule a diferença entre o valor máximo e mínimo, isso é a amplitude (no nosso caso a diferença é $400 - 10 = 390$)
- Agora, devemos dividir a amplitude pelo número de classes (no nosso caso devemos dividir $390/6 = 65$)

Após estas etapas, conseguimos definir qual é a amplitude da classe que no nosso caso é 65.

Sendo assim, teremos a seguinte distribuição:

Classe	Frequência
10 - 75	22
75 - 140	9
140 - 205	2
205 - 270	2
270 - 335	0
335 - 400	1

Aqui definimos que os intervalos estão fechados à esquerda e aberto a direita, isso significa que na primeira faixa que vai de 10 até 75, teremos valores entre 10 e não o valor 75 que ficará para a próxima faixa. A última faixa, fugimos a esta regra porque na última faixa devemos incluir o valor máximo da amostra.

- Símbolos

1 |- fechado a direita e aberto a esquerda

2 |-| fechado a direita e a esquerda

A análise é fácil aqui, ou seja, entre 10 e 75 quilômetros (aberto), temos 22 caminhões e que a maioria dos caminhões estão nesta faixa.

- **Valor Central de Classe**

O valor central de uma classe é bem fácil de ser calculada. Somente devemos somar os intervalos e dividir por dois.

Então para a primeira classe 10 |- 75 é :

$$(10+75)/2=42,5$$

Vamos deixar o valor inteiro para cada valor central ok.

Classe	Valor Central	Frequência
10 - 75	42	22
75 - 140	107	9
140 - 205	172	2
205 - 270	237	2
270 - 335	302	0
335 - 400	367	1

Caros amigos e colegas, com este post, fecho a parte de estatística básica.

Pretendo de agora em diante trabalhar com estatística descritiva, probabilidades, regressões e gráficos.

Meu objetivo com estes posts é o de voltar ao assunto aprendizado de máquina.

Também faz parte do aprendizado de máquina os posts relacionados com neurônios, que aos poucos está se tornando uma nova paixão de estudo em minha vida.

Espero de coração ter ajudado todos que leram este post de alguma forma.

Abraços
Pequeno

Conhecimento gera conhecimento! Projeto Spartacus.