Oi Jayme, seguem as dúvidas de ontem:

1)

Histograma

O histograma que te mandei foi feito assim:

hist(Banco\_de\_dados\_Liana\_Statistica\_Tratado$HA1C, breaks = c(min(Banco\_de\_dados\_Liana\_Statistica\_Tratado$HA1C, na.rm = TRUE),6.0,6.5, max(Banco\_de\_dados\_Liana\_Statistica\_Tratado$HA1C, na.rm = TRUE)), right = FALSE, freq=TRUE, main = "Histograma da variável HbA1c", xlab = "Intervalos de HbA1c", ylab = "Frequência", labels = TRUE)

queria que ele mostrasse o eixo x com os valores de HA1C de, por exemplo:

3.5

5.7

6.5

15

19

Como fazer?

axis (1, med, med)

Primeiro argumento (1): entendi que significa o eixo x

Segundo argumento: Os ticks seriam as divisões do histograma? Os valores que sugeri acima? Não entendi como colocá-los no segundo argumento.

Ou ele não faz valores quebrados assim? Preciso colocar ali o valor, por ex. de 1, para que o eixo x seja construído com todos os valores para HA1C (1, 2, 3, 4, 5, ...)? Eu tentei fazer axis (1,1,1) e não deu certo...

O terceiro argumento (vetor com o que vai ser escrito) - não entendi o que isso significa...

Para o eixo y:

crio o eixo y, antes de colocá-lo no argumento completo:

eixoy = seq(from = 0, to = length(Banco\_de\_dados\_Liana\_Statistica\_Tratado$HA1C), by = 1000)

Assim o eixo y irá do zero até o último paciente, subdividindo de 1000 em 1000, certo?

Isso eu fiz.

Mas não entendi como colocar no argumento completo...

hist(Banco\_de\_dados\_Liana\_Statistica\_Tratado$HA1C, breaks = c(min(Banco\_de\_dados\_Liana\_Statistica\_Tratado$HA1C, na.rm = TRUE),6.0,6.5, max(Banco\_de\_dados\_Liana\_Statistica\_Tratado$HA1C, na.rm = TRUE)), right = FALSE, freq=TRUE, main = "Histograma da variável HbA1c", xlab = "Intervalos de HbA1c", ylab = "Frequência", labels = TRUE, como escrever o eixo x e y aqui??)

2)

Demais análises que preciso fazer:

2.1)

na planilha que te mandei, retirei a coluna com os nomes dos pacientes, para manter a privacidade, obviamente. Há pacientes que fizeram exames mais de uma vez durante esse período. Meu professor pediu para retirar os homônimos, mas com algum critério (tipo: deixar apenas o primeiro cadastro do paciente). No excel, existe a opção de retirar os homônimos, facilmente, mas não sei qual critério ele utiliza. Você sabe, por acaso?

Ou então, pelo R, seria possível fazer isso, com algum critério específico?

2.2)

Análise de correlação de Pearson

Esta análise eu consegui fazer, da seguinte forma:

with(Banco\_de\_dados\_Liana\_Statistica\_Tratado, cor.test (GLI, HA1C))

O resultado encontrado foi:

Pearson's product-moment correlation

data: GLI and HA1C

t = 574.22, df = 186810, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.7973187 0.8005988

sample estimates:

cor

0.7989647

Pelo que entendi, existe correlação forte (0,7989) entre esses dois parâmetros.

Porém, não entendi esse valor de p. Significa que o p < 2,2-16? Ou seja, 0,000003320? Achei muito estranho esse valor...

2.3)

Preciso criar 2 grupos (doentes / saudáveis) com base no valor de HA1C e comparar os grupos.

Na verdade, pelo que entendo, deveria criar 3 grupos, porque aí preciso descontar os pacientes que tiverem HA1C entre 6 e 6,5.

|  |  |  |
| --- | --- | --- |
| Grupo controle saudável  (HbA1c <6,0%)  N= xxxx | Grupo diabético  (HbA1c ≥ 6,5%)  N= xxxx | P-value |
|  |  |  |
|  |  |  |

Para isso, pensei em criar uma coluna no excel, colocando

1 para grupo controle

2 para o grupo diabético

3 para o grupo de pré-diabéticos (6.0 <= HA1C < 6,5)

Depois compará-los no R.

Não sei como fazer essa comparação entre grupos no R.

Como os dados não têm distribuição normal, mesmo tentando fazer o log de cada variável. Preciso, então, utilizar Mann-Whitney, certo?

O teste t não seria possível, pelo que entendo, porque além de não terem distribuição normal os dados não possuem homocedasticidade.

No programa statistica eu fiz isso, porém, com são muitos dados, ele não consegue analisar direito. No R, não sei como fazer...

Se quiser comparar mais de 2 grupos, daí teria que utilizar o teste não paramétrico de Kruskal-wallis, certo?

Também não sei como fazer no R.

2.4)

Depois, preciso analisar, por exemplo, se há diferença entre as HA1C entre H e M no grupo diabético e no grupo controle. Depois, acredito que sabendo fazer uma dessas análises, as demais serão parecidas, bastará apenas trocar as variáveis.

\*\* Se eu quiser fazer algum gráfico também comparando os grupos é possível?

2.5)

Analisar os demais biomarcadores (COL, TRI etc). Se diferem entre os grupos.

Acredito que estes seguirão um padrão. Assim, sabendo fazer com HA1C, farei com os demais, apenas trocando a variável.

2.6)

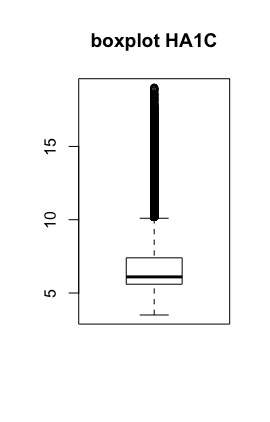
Boxplot

Veja a diferença do boxplot entre o que fiz no R e no Statistica.

Parece que há algo errado... agora não sei qual está correto.

No R, usei o seguinte argumento:

boxplot(Banco\_de\_dados\_Liana\_Statistica\_Tratado$HA1C)



Uma imagem contendo mapa

Descrição gerada automaticamente