

CURSO R - 2023

6 PROCESSAMENTO DOS DADOS

Criação de planilha descritiva de dados
contendo: valores absolutos, percentuais e médias

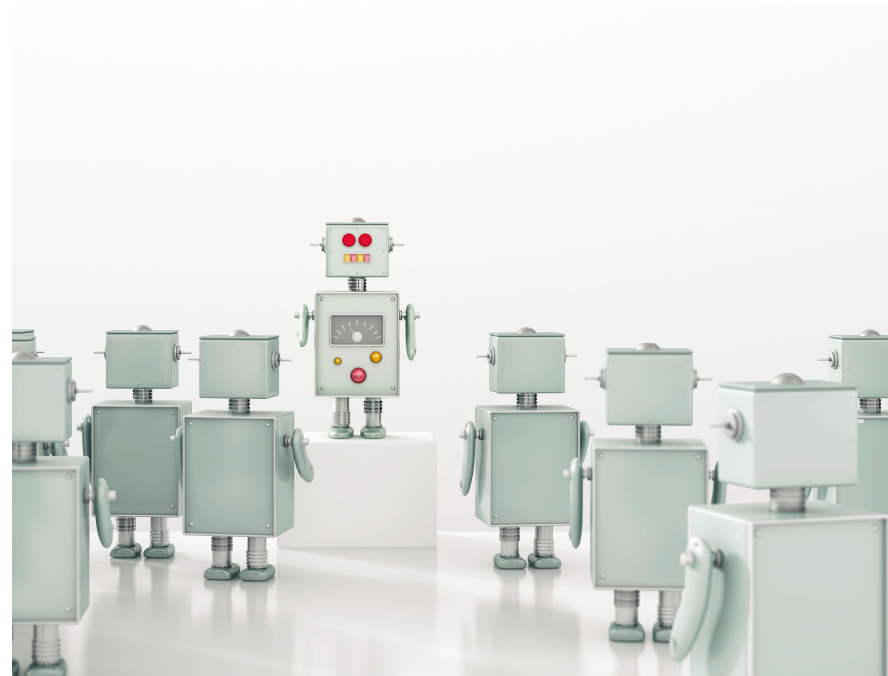


NAIARA ALCANTARA

TÓPICOS

- 1- INTRODUÇÃO AO R
- 2- ANÁLISE EXPLORATÓRIA E MANIPULAÇÃO DOS DADOS
- 3- SALVAMENTO E ABERTURA
- 4- ANÁLISES DESCRITIVAS
- 5- APRESENTAÇÃO GRÁFICA
- 6- PROCESSAMENTO DE DADOS
- 7- ANÁLISES INFERENCIAIS

Análise descritiva com output composto por valores absolutos e relativos (percentuais), pode ser aplicado a grandes bases de dados.



PROCESSAMENTO DOS DADOS

Além da análise descritiva de questão por questão, podemos ter a necessidade de processar de uma única vez uma quantidade grande de dados, a fim de criar um relatório executivo por exemplo;

Através do pacote: expss, conseguimos fazer isso.

Inicialmente vamos criar uma base de dados com diversos tipos de variáveis.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	NOME	SEXO	IDADE	CIDADE	ESTADO	RENDIMENTO	Q1	Q1_1	Q1_2	Q2	Q2_1	Q3	Q11	Peso	
2	Beatriz	F	29	Abadia dos Dourados	MG	1	LARANJA		GOIABA	FRIO	CALOR	10	10	0,59	
3	Elisa	F	65	Abaeté	MG	4		LIMÃO			CALOR	6	6	0,59	
4	Pedro	M	28	Abre Campo	MG	6	LARANJA		GOIABA	FRIO	CALOR	4	4	1,59	
5	Benjamin	M	35	Acaiaca	MG	5		LIMÃO	GOIABA	FRIO		9	9	1,59	
6	Isaac	M	28	Açucena	MG	1	LARANJA				CALOR	1	1	1,59	
7	Camila	F	39	Abadia dos Dourados	MG	4	LARANJA	LIMÃO		FRIO	CALOR	8	8	0,59	
8	Victor	M	34	Abaeté	MG	6	LARANJA		GOIABA		CALOR	7	7	1,59	
9	Aarão	M	39	Abre Campo	MG	5				FRIO	CALOR	11	11	1,59	
10	Vinícios	M	40	Acaiaca	MG	8	LARANJA		GOIABA		CALOR	6	6	1,59	
11	César	M	50	Açucena	MG	7			GOIABA		CALOR	7	7	1,59	
12	Karen	F	46	Abadia dos Dourados	MG	4	LARANJA	LIMÃO			CALOR	6	6	0,59	

Essa base, assim como as utilizadas nos demais tópicos estarão disponíveis no GitHub



PROCESSAMENTO DOS DADOS

```
1126 ▾ #Tópico 6####
1127 ▾ #Processamento dos dados####
1128
1129 ##### para gerar o output como porcentagem
1130 ▾ add_percent = function(x, digits = get_expss_digits(), ...){
1131   UseMethod("add_percent")
1132 ▴ }
1133
1134 ▾ add_percent.default = function(x, digits = get_expss_digits(), ...){
1135   res = formatC(x, digits = digits, format = "f")
1136   nas = is.na(x)
1137   res[nas] = ""
1138   res[!nas] = paste0(res[!nas], "%")
1139   res
1140 ▴ }
1141
1142 add_percent.etable = function(x, digits = get_expss_digits(),
1143 ▾   excluded_rows = "#", ...){
1144   included_rows = !grepl(excluded_rows, x[[1]], perl = TRUE)
1145 ▾   for(i in seq_along(x)[-1]){
1146 ▾     if(!is.character(x[[i]])){
1147 ▴       x[[i]][included_rows ] = add_percent(x[[i]][included_rows])}x}
1148
```

PROCESSAMENTO DOS DADOS

```
1149 #####
1150
1151 library(tidyverse)
1152
1153 #Para realizar o processamento precisamos
1154 #dos seguintes pacotes: expss e openxlsx
1155 install.packages("expss")
1156 library(expss)
1157
1158 install.packages("openxlsx")
1159 library(openxlsx)
1160
1161
1162 #Por gentileze, todos usando o default de salvamento em UTF-8
1163
1164 options(OutDec = ",", digit = 0)
1165 |
1166 #Como em nossos relatórios utilizamos valores percentuais
1167 #com o símbolo do %, então teremos que criar um objeto
1168 #chamado "add_percent" que irá adicionar percentual a nossa
1169 #planilha.
1170 #Então olhem no sumário, cliquem e add percent,
1171 #selecinem até o final e rodem. Depois voltem aqui
1172 #para retormarmos o passo a passo.
1173
1174
```


PROCESSAMENTO DOS DADOS

```
1174 #Vamos inicialmente importar a base de dados que coletamos
1175 #Chamando ela de base, utilizando o pacote rio e a função
1176 #import
1177
1178 install.packages("rio")
1179 library(rio)
1180
1181 base = rio::import("BANCOTREI.xlsx", which = 1)
1182
1183 head(base)
1184
1185
1186 #Queremos analisar o que ?
1187 #-Cidade, com base no sexo e rendimentos..
1188
1189 #Agora transforme todos os labels em caracteres,
1190 #usando essas linhas:
1191
1192 #install.packages("Hmisc")#Para quem já ativou o tidyverse não precisa
1193 #library(Hmisc)
1194
1195 var.labels = as.character(names(base))
1196
1197 for(i in seq_along(base)){
1198   Hmisc::label(base[, i]) <- var.labels[i]
1199 }
```

PROCESSAMENTO DOS DADOS

```
1200
1201 #RU####
1202 #PROCESSAMENTO DE RESPOSTA ÚNICA
1203
1204 #Agora é preciso que você escolha quais variáveis
1205 #irão compor a sua lista de variáveis de respostas únicas,
1206 #anotando o número das colunas, que pode ser verificado na
1207 #na própria base de dados
1208 vars_ps = list(base[,c(3:4)], base[,c(5)]) #falar da 5
1209
1210
1211 #Agora escolhemos o que desejamos cruzar com as variáveis
1212 #escolhidas anteriormente
1213 vars_cruz = with(base, list(total(), SEX0))
1214
1215 #Feito isso, criamos a base, que no caso, vamos chamar de
1216 #RU, se referindo a resposta única
1217 ru = base %>%
1218   calculate(cro_cpct(cell_vars = vars_ps,
1219                     col_vars = vars_cruz)) %>%
1220   tab_sort_desc %>%
1221   set_caption("RU") %>%
1222   add_percent(digit = 0)
1223
```

PROCESSAMENTO DOS DADOS

```
1225 #criamos uma pasta de trabalho
1226 wb = createWorkbook()
1227 #Adicionamos nesse documento criado a aba RU
1228 sh = addWorksheet(wb, "RU")
1229 #Gravamos todas tabelas em um documento
1230 xl_write(ru, wb, sh)
1231
1232
1233 #valor absoluto + percentual####
1234
1235 #Achado do arthur para análise dos valores absolutos também
1236 # Absoluto = base %>%
1237 #   tab_cells(vars_ps) %>%
1238 #   tab_cols(total(),vars_cruz) %>%
1239 #   tab_stat_cases(label = "N", total_label = "") %>%
1240 #   tab_stat_cpct(label="%", total_statistic = "w_cpct",
1241 #                 total_label = "") %>%
1242 #   tab_pivot(stat_position = "outside_rows") %>%
1243 #   set_caption("Absoluto")
1244
1245 #Argumentos que podem ser utilizados para organização
1246 #dos valores percentuais:
1247 #“outside_rows”, “inside_rows”,
1248 #“outside_columns”, “inside_columns”
1249
```


PROCESSAMENTO DOS DADOS

```
1251 Absoluto = base %>%
1252   tab_cells(vars_ps) %>%
1253   tab_cols(total(), vars_cruz) %>%
1254   tab_stat_cases(label = "N", total_label = "") %>%
1255   tab_pivot(stat_position = "outside_rows") %>%
1256   set_caption("Absoluto")
1257
1258 sh2 = addWorksheet(wb, "Absoluto")
1259 xl_write(Absoluto, wb, sh2)
```



PROCESSAMENTO DOS DADOS

```
1261 #PESO####  
1262 base$Peso ← as.numeric(base$Peso)  
1263  
1264  
1265 ANALIPESO = base %>%  
1266   calculate(cro_cpct(cell_vars = vars_ps,  
1267                 col_vars = vars_cruz,  
1268                 total_statistic = "w_cases",  
1269                 weight = base$Peso)) %>%  
1270   tab_sort_desc %>%  
1271   set_caption("PESO") %>%  
1272   add_percent(digits = 0)  
1273  
1274  
1275 sh1 = addWorksheet(wb, "PESO")  
1276 xl_write(ANALIPESO, wb, sh1)  
1277
```

PROCESSAMENTO DOS DADOS

```
1278 #RM####
1279 #Agora a criação da aba com o processamento de respostas múltiplas
1280 #Utilizamos a função mrset_p para analisar mais de uma coluna
1281 rm = base %>%
1282   calculate(cross_cpct(base, cell_vars = list(mrset_p("Q1"),
1283                                               mrset_p("Q2")),
1284           col_vars = vars_cruz)) %>%
1285   tab_sort_desc %>% set_caption("RM")%>%
1286   add_percent(digits = 0)
1287
1288 sh2 = addWorksheet(wb, "RM")
1289 xl_write(rm, wb, sh2)
1290 #MÉDIA####
1291 vmedias = list(base[,c(12)])
1292
1293 medias_cruz = with(base, list(total(), SEX0))
1294
1295 MEDIA = base %>%
1296   calculate(cro_mean(cell_vars = vmedias,
1297           col_vars = medias_cruz)) %>%
1298   tab_sort_desc %>% set_caption("MEDIA")
1299
1300 sh3 = addWorksheet(wb, "MEDIA")
1301 xl_write(MEDIA, wb, sh3)
1302
```

PROCESSAMENTO DOS DADOS

```
1304 #criar arquivo xls
1305 saveWorkbook(wb, "Output3.xlsx", overwrite = TRUE)
1306
```

	A	B	C	D	E	F	G
1	RU						
2			#Total	SEXO			
3				F	M		
4	IDADE	18	10,0%	12,5%	8,3%		
5		28	10,0%		16,7%		
6		39	10,0%	12,5%	8,3%		
7		51	10,0%	25,0%			
8		23	5,0%		8,3%		
9		29	5,0%	12,5%			
10		34	5,0%		8,3%		
11		35	5,0%		8,3%		
12		40	5,0%		8,3%		
13		46	5,0%	12,5%			
14		50	5,0%		8,3%		
15		54	5,0%		8,3%		
16		62	5,0%		8,3%		
17		64	5,0%		8,3%		
18		65	5,0%	12,5%			
19		70	5,0%	12,5%			
20		#Total cases	20	8	12		
21	CIDADE	Abadia dos D	20,0%	50,0%			
22		Abaeté	20,0%	37,5%	8,3%		
23		Abre Campo	20,0%	12,5%	25,0%		
24		Acaiaca	20,0%		33,3%		
25		Açucena	20,0%		33,3%		
26		#Total cases	20	8	12		
27	ESTADO	MG	100,0%	100,0%	100,0%		
28		#Total cases	20	8	12		
	RU	Absoluto	PESO	RM	MEDIA		



OBRIGADO

NAIARA ALCANTARA
NAYARA_SANDY@HOTMAIL.COM