

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - Campus São Paulo

Tecnólogo em Análise e Desenvolvimento de Sistemas - 5º Semestre

Débora Peixoto dos Santos SP3069028

João Pedro Vanderlei SP3069427

Pedro Henrique Beserra Lima SP3073785

Estatística e Probabilidade (ESP1A5) - Professora Josceli Tenório

Projeto - Relatório Analítico SISVAN do ano de 2021

## **Introdução**

O presente relatório consistirá na realização de uma análise descritiva e inferencial de uma amostra da base de dados do Sistema de Vigilância Alimentar e Nutricional (SISVAN) do ano de 2021.

O SISVAN possui como objetivo executar a consolidação dos dados referentes às ações concernentes à Vigilância Alimentar e Nutricional da Atenção Primária à Saúde (APS), possuindo então dados de antropometria e indicadores de consumo alimentar no território brasileiro. Os dados disponibilizados na base para a realização da análise, por sua vez, referem-se às informações individualizadas e anonimizadas de todas as pessoas atendidas quanto ao estado nutricional (antropometria) em serviços de APS do Brasil. Ademais, os relatórios públicos do SISVAN também consolidam sempre o último acompanhamento do indivíduo.

## **Estatística descritiva**

A primeira etapa da análise compreende uma investigação descritiva dos dados, em específico, o perfil dos indivíduos presentes na amostra da base. Assim sendo, foram feitos histogramas, boxplots e medidas resumo a título de exploração. Foram levantados os seguintes pontos

- Qual o tamanho da base utilizada?

```
df = sisvan_estado_nutricional_2021
```

```
dim(df)
```

**11395 34**

- Quantas pessoas foram analisadas neste relatório?

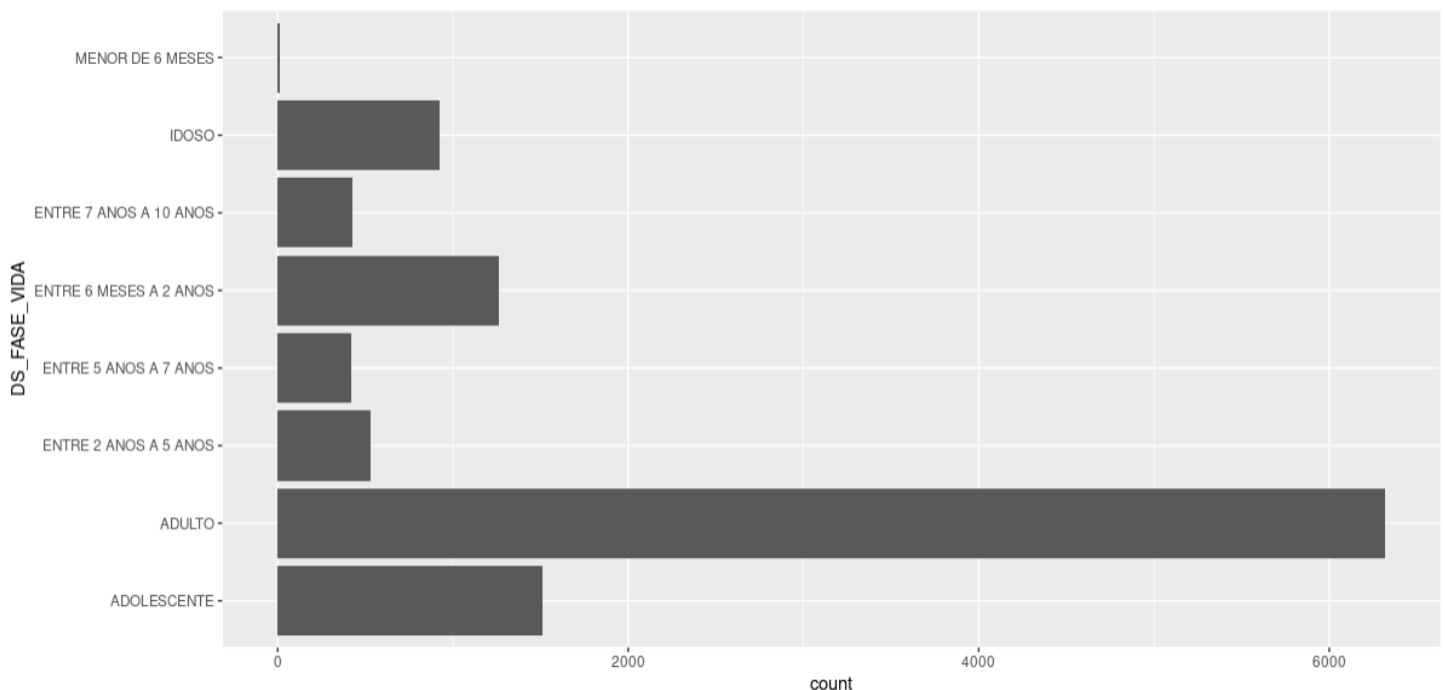
```
n_distinct(df$CO_ACOMPANHAMENTO)
```

**11395**

Obs.: O campo *CO\_ACOMPANHAMENTO* representa o código único que identifica o acompanhamento realizado por um determinado indivíduo

- Qual a quantidade de pessoas por faixa de idade?

```
ggplot(df, aes(DS_FASE_VIDA)) + geom_bar() + coord_flip()
```



```
table(df$DS_FASE_VIDA)
```

ADOLESCENTE - **1508**

ADULTO - **6318**

ENTRE 2 ANOS A 5 ANOS - **533**

ENTRE 5 ANOS A 7 ANOS - **417**

ENTRE 6 MESES A 2 ANOS - **1260**

ENTRE 7 ANOS A 10 ANOS - **423**

IDOSO - **921**

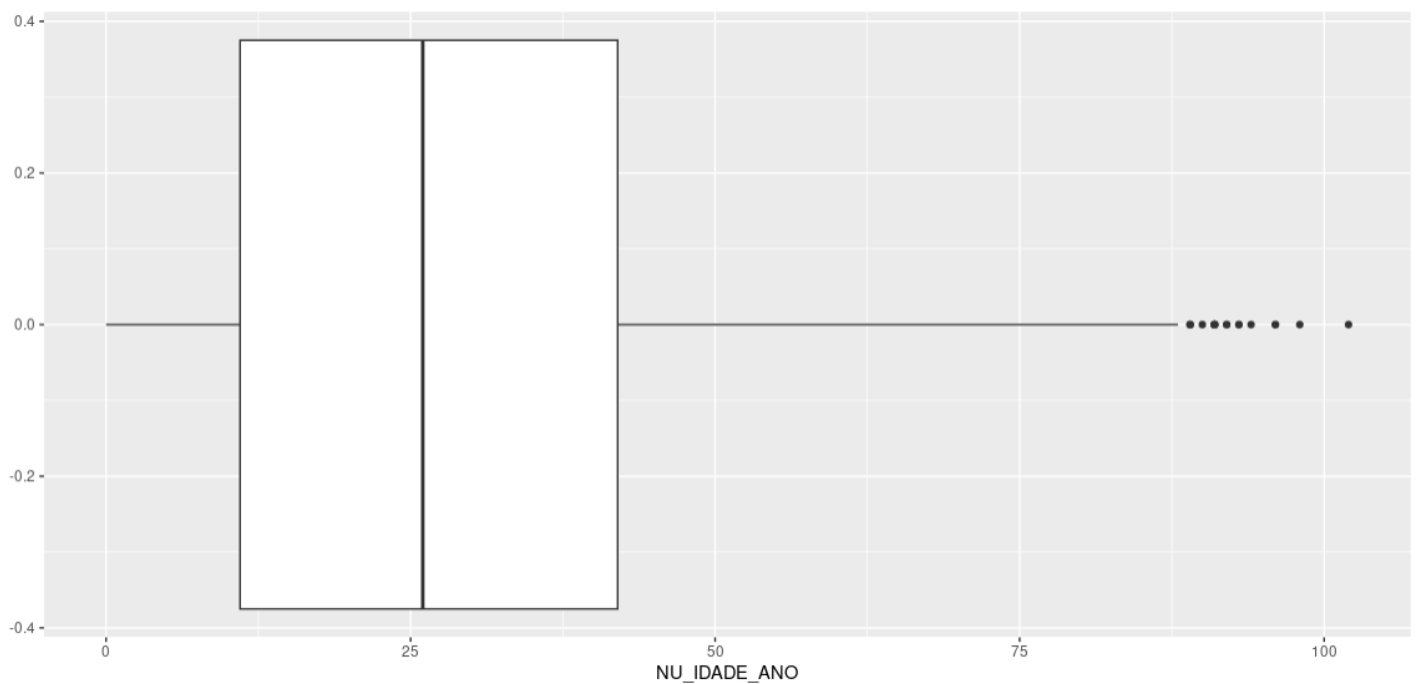
MENOR DE 6 MESES - **15**

- Qual a distribuição de idade?

```
summary(df$NU_IDADE_ANO)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.00  11.00  26.00  28.14  42.00 102.00
```

```
ggplot(df, aes(NU_IDADE_ANO)) + geom_boxplot()
```

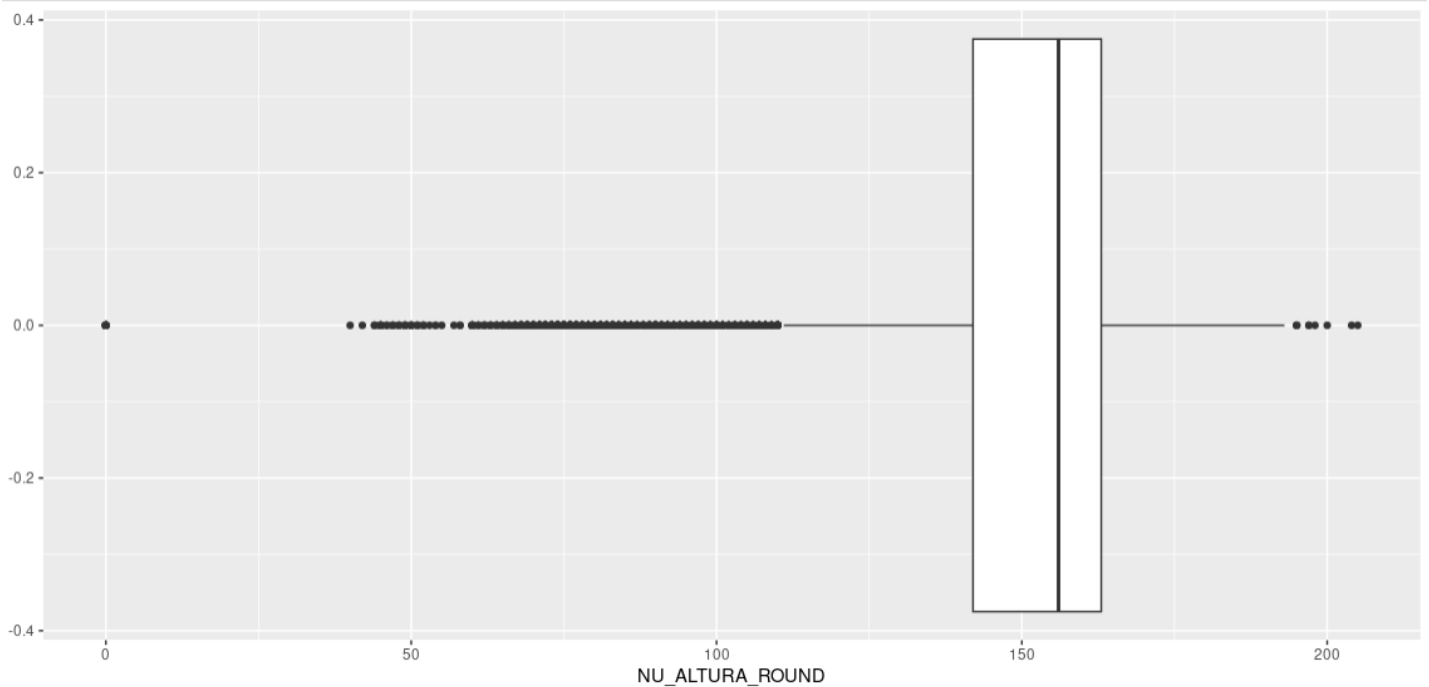


- Qual a distribuição de altura?

```
summary(df$NU_ALTURA_ROUND)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0	142	156	144	163	205	1

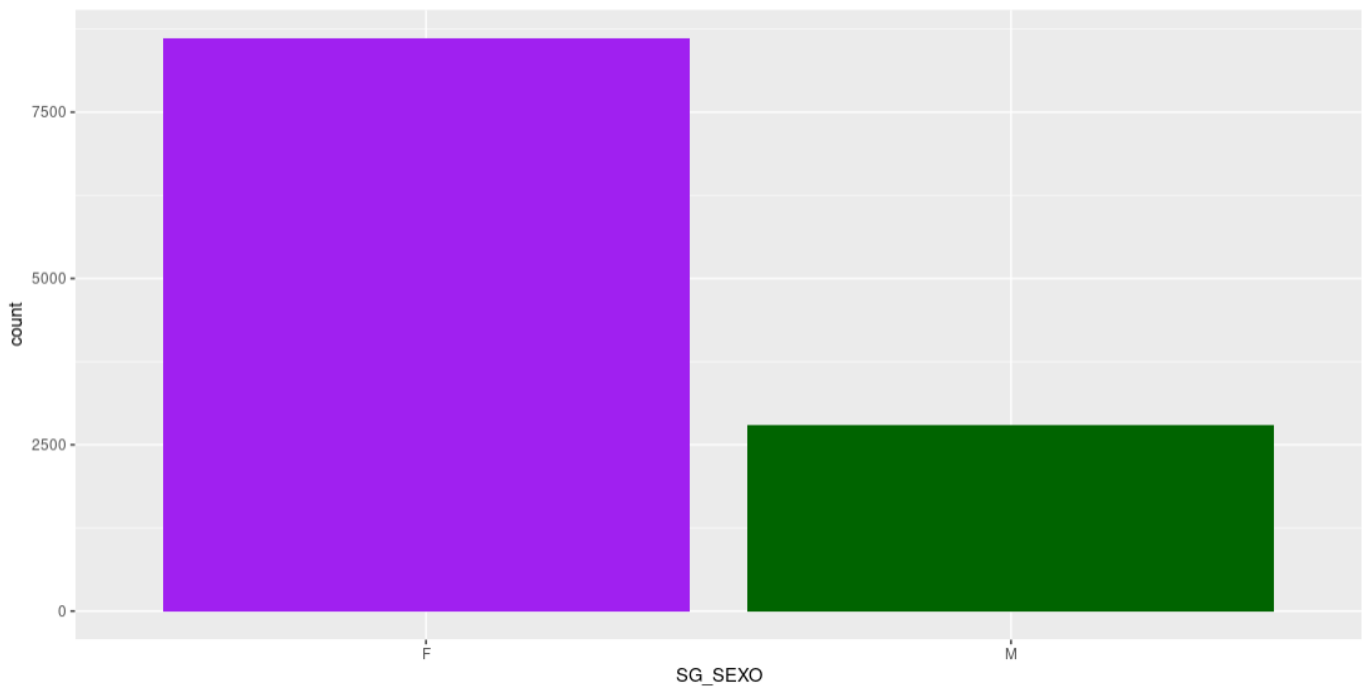
```
ggplot(df, aes(NU_ALTURA_ROUND)) + geom_boxplot()
```



Obs.: Para a plotagem deste gráfico, foi gerada uma nova coluna em Excel que arredonda os valores de altura, que estavam sendo passadas sem vírgula para o RStudio, chamada DS\_ALTURA\_ROUND, contida no arquivo sisvan\_estado\_nutricional\_2021\_round.

- Qual a quantidade de pessoas por sexo?

```
ggplot(df, aes(SG_SEXO, fill = SG_SEXO)) +  
  geom_bar() +  
  scale_fill_manual(values = c("F" = "purple", "M" = "darkgreen"))
```



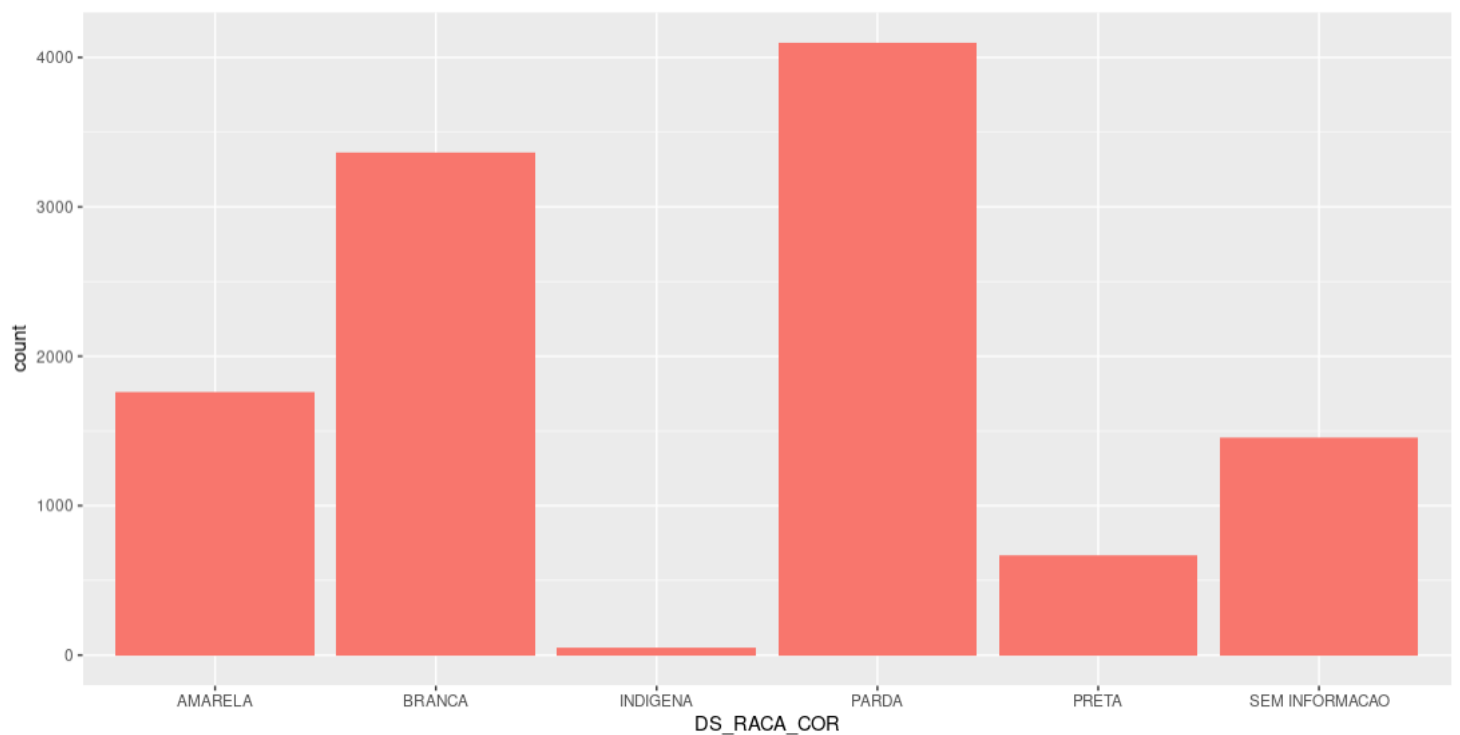
```
table(df$SG_SEXO)
```

F - **8606**

M - **2789**

- Qual a quantidade de pessoas por raça ou cor?

```
ggplot(df, aes(DS_RACA_COR, fill = "orange")) + geom_bar()
```



```
table(df$DS_RACA_COR)
```

AMARELA - **1762**

BRANCA - **3364**

INDIGENA - **52**

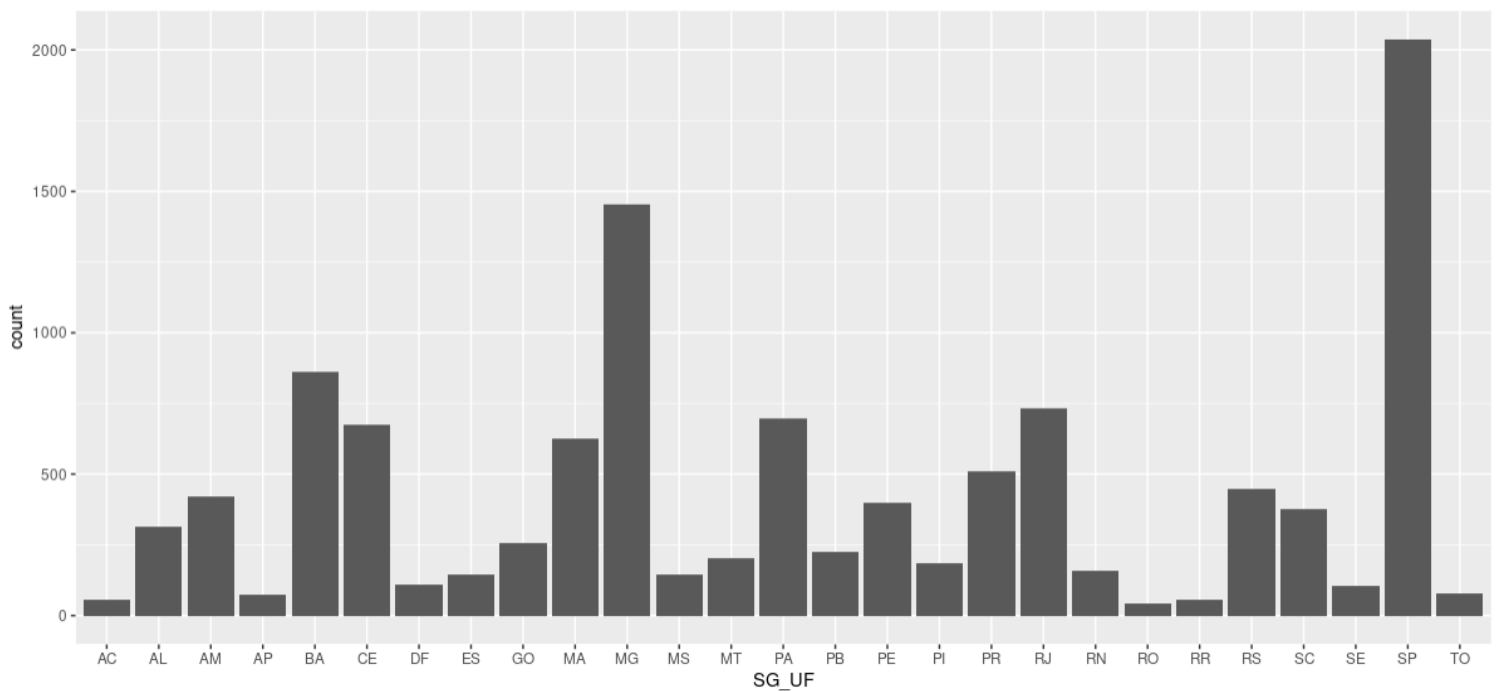
PARDA - **4096**

PRETA - **665**

SEM INFORMACAO - **1456**

- Qual a quantidade de pessoas por UF?

```
ggplot(df, aes(SG_UF)) + geom_bar()
```



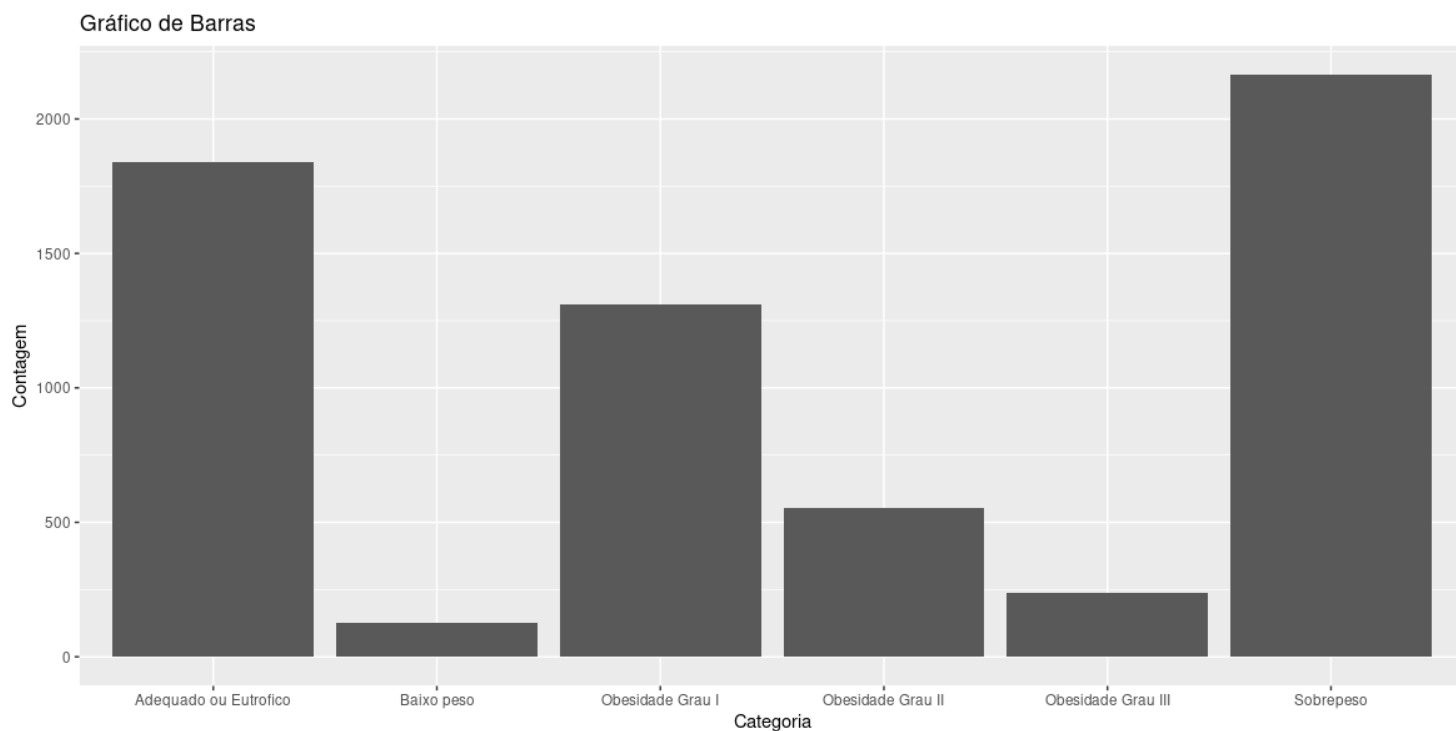
```
table(df$SG_UF)
```

AC	AL	AM	AP	BA	CE	DF	ES	GO	MA	MG	MS	MT	PA	PB	PE	PI
58	313	419	75	862	674	111	145	257	627	1454	145	203	695	226	400	183
PR	RJ	RN	RO	RR	RS	SC	SE	SP	TO							
511	734	157	44	57	449	375	107	2036	78							

Após um panorama de alguns atributos, a análise passa a ter um enfoque em indicadores a respeito de condições de peso fora da normalidade, como baixo peso e sobrepeso.

- Dado que o maior volume de pessoas estão categorizadas como adultas, representando **55,4%** da base, qual a quantidade de adultos por estado nutricional?

```
estado_nutri <-  
table(sisvan_estado_nutricional_2021$CO_ESTADO_NUTRI_ADULTO)  
estado_nutri <- as.data.frame(estado_nutri)  
  
estado_nutri$Var1 <- iconv(estado_nutri$Var1, from="UTF-8", to="LATIN1")  
na_indices <- is.na(estado_nutri$Var1)  
  
# Substituir os NA por um novo valor, por exemplo, "Valor Ausente"  
estado_nutri$Var1[na_indices] <- "Adequado ou Eutrofico"  
print(estado_nutri)  
  
ggplot(data = estado_nutri, aes(x =Var1, y=Freq)) +  
  # Adicionar a camada de barras  
  geom_bar(stat = "identity", position = "stack") +  
  
  # Adicionar título e rótulos dos eixos  
  labs(title = "Gráfico de Barras",  
        x = "Categoria",  
        y = "Contagem")
```



- Qual a quantidade de adultos fora da normalidade de peso (baixo peso, sobrepeso e obesidade)?

```
table(df$CO_ESTADO_NUTRI_ADULTO)
```

Quantidade de adultos na base = 6318

Obs.: O campo *CO\_ESTADO\_NUTRI\_ADULTO* representa o cálculo nutricional de adultos (20 a 60 anos) com as seguintes informações: 'Baixo peso', 'Adequado ou Eutrófico', 'Sobrepeso', 'Obesidade Grau I', 'Obesidade Grau II', 'Obesidade Grau III'

Adequado ou eutrófico - **1839**

Baixo peso - **125**

Obesidade Grau I - **1311**

Obesidade Grau II - **555**

Obesidade Grau III - **237**

Sobrepeso - **2165**



Totalizando **4393** adultos com anormalidade de peso, ou seja, **69,3%** e apenas **29,1%** no peso considerado adequado (**1,6%** de missing).

- Qual a quantidade de idosos fora da normalidade de peso (baixo peso e sobrepeso)?

```
table(df$CO_ESTADO_NUTRI_IDOSO)
```

Quantidade de adultos na base = **921**

Obs.: O campo CO\_ESTADO\_NUTRI\_IDOSO representa o cálculo nutricional de idosos (60 anos ou mais) com as seguintes informações: 'Baixo peso', 'Adequado ou Eutrófico', 'Sobrepeso'

Adequado ou eutrófico - **322**

Baixo peso - **106**

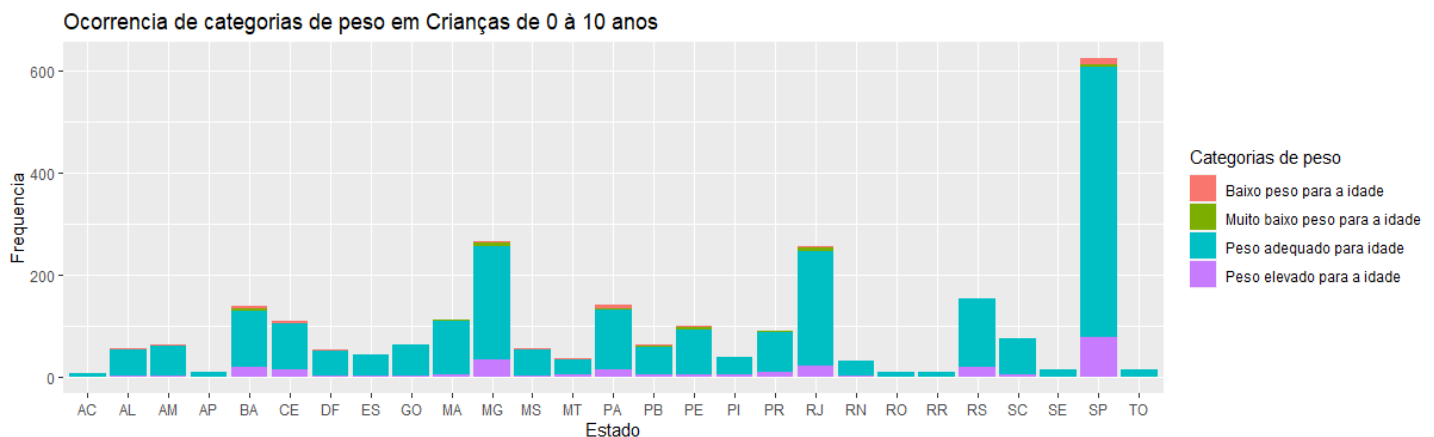
Sobrepeso - **483**

Totalizando **589** idosos com anormalidade de peso, ou seja, **64%** e apenas **35%** no peso considerado adequado (**1%** de missing).

## Análise bivariada

Para sabermos qual a quantidade de crianças em cada categoria por estado, iremos gerar uma tabela de contingência. Para isso, vamos filtrar a tabela Peso x Idade, que contém apenas dados de indivíduos com idade entre 0 e 10 anos. A partir disso geramos o gráfico de ocorrências.

```
filtered_data <-  
sisvan_estado_nutricional_2021[!is.na(sisvan_estado_nutricional_2021$PESO.X.IDADE) & sisvan_estado_nutricional_2021$PESO.X.IDADE != "", ]  
tab <- table(filtered_data$SG_UF, filtered_data$PESO.X.IDADE)  
tab_df <- as.data.frame(tab)  
ggplot(data=tab_df, aes(x=Var1, y=Freq, fill=Var2)) +  
  labs(x = "Estado", y = "Frequência", fill = "Categorias de peso", title = "Ocorrência de categorias de peso em Crianças de 0 à 10 anos") +  
  geom_bar(stat="identity")
```



## Tabela de contingência

```
conting <- as.data.frame.matrix(tab)
```

	Baixo peso para a idade	Muito baixo peso para a idade	Peso adequado para idade	Peso elevado para a idade	sum
AC	0	0	6	1	7
AL	2	0	50	4	56
AM	1	0	60	2	63
AP	1	0	10	0	11
BA	6	4	110	20	140
CE	3	0	91	15	109
DF	2	0	47	4	53
ES	1	0	39	4	44
GO	1	0	60	3	64
MA	0	1	105	6	112
MG	3	7	222	34	266
MS	2	1	52	2	57

<b>MT</b>	1	1	29	5	36
<b>PA</b>	7	2	118	14	141
<b>PB</b>	2	4	53	5	64
<b>PE</b>	2	4	88	6	100
<b>PI</b>	1	0	34	5	40
<b>PR</b>	0	2	78	11	91
<b>RJ</b>	3	6	224	23	256
<b>RN</b>	0	1	29	3	33
<b>RO</b>	1	0	9	0	10
<b>RR</b>	0	0	10	1	11
<b>RS</b>	1	0	132	21	154
<b>SC</b>	0	1	69	6	76
<b>SE</b>	0	0	16	0	16
<b>SP</b>	13	5	527	79	624

TO	0	0	14	0	14
sum	53	39	2282	274	2648

## Probabilidade

Ao analisarmos os dados coletados previamente na análise descritiva, nós conseguimos responder a seguinte questão “*Qual a quantidade de pessoas por faixa de idade? (pag. 1)*”. Com os dados coletados, podemos fazer algumas análises de probabilidade baseado nessas faixas etárias.

Sabemos que o número total de adultos é de **6318** e podemos fazer uma análise de probabilidade mais aprofundada desse dado.

1. Ao selecionar aleatoriamente um indivíduo adulto, qual a probabilidade dele estar classificado como determinado estado nutricional?

Para facilitar a coleta de dados, podemos verificar a descrição completa da coluna de “CO\_ESTADO\_NUTRI\_ADULTO”, da mesma forma que utilizamos para responder “*Qual a quantidade de adultos fora da normalidade de peso (baixo peso, sobrepeso e obesidade)?*”.

```
table(df$CO_ESTADO_NUTRI_ADULTO)
```

```
> table(df$CO_ESTADO_NUTRI_ADULTO)
Adequado ou eutr\xfcfico      Baixo peso      Obesidade Grau I      Obesidade Grau II
          1839              125             1311             555
  Obesidade Grau III      Sobrepeso
          237            2165
> |
```

Com os dados obtidos, podemos realizar um cálculo para obter a probabilidade para variáveis discretas que é o mais adequado para lidar com dados categóricos.

Primeiramente, vamos entender qual o total de Adultos Categorizados, fazendo a soma dos dados na coluna CO\_ESTADO\_NUTRI\_ADULTO:

```
> total_de_adultos_categorizados_nutricionalmente <- sum(table(df$CO_ESTADO_NUTRI_ADULTO))
> total
[1] 6232
```

Sendo assim, o total de adultos categorizados nutricionalmente = **6232**.

#### Adequado ou Eutrófico:

O total de adultos nessa categoria é de **1839**, como apresentado na descrição da coluna **CO\_ESTADO\_NUTRI\_ADULTO**.

Para sabermos qual a probabilidade de um adulto selecionado aleatoriamente estar com o estado nutricional “adequado ou eutrófico”, nós podemos fazer o seguinte cálculo:

$$\frac{\text{Total\_Adultos\_Adequado\_ou\_Eutrófico}}{\text{Total\_Adultos\_Categorizados\_Nutricionalmente}}$$

```
> Total_Adultos_Categorizados_Nutricionalmente <- sum(table(df$CO_ESTADO_NUTRI_ADULTO))
>
> Total_Adultos_Adequado_ou_Eutrofico <- 1839
> Prob_Adequado_ou_Eutrofico <- Total_Adultos_Adequado_ou_Eutrofico / Total_Adultos_Categorizados_Nutricionalmente
> print(Prob_Adequado_ou_Eutrofico)
[1] 0.2950899
```

Probabilidade de um indivíduo aleatório estar adequado ou eutrófico: **29,5%**

#### Baixo peso:

O total de adultos nessa categoria é de **125**, como apresentado na descrição da coluna **CO\_ESTADO\_NUTRI\_ADULTO**.

Para sabermos qual a probabilidade de um adulto selecionado aleatoriamente estar com o estado nutricional “baixo peso”, nós podemos fazer o seguinte cálculo:

$$\frac{\text{Total\_Adultos\_Baixo\_Peso}}{\text{Total\_Adultos\_Categorizados\_Nutricionalmente}}$$

```

> ## BAIXO PESO
> Total_Adultos_Baixo_Peso <- 125
> Prob_Baixo_Peso <- Total_Adultos_Baixo_Peso / Total_Adultos_Categorizados_Nutricionalmente
> print(Prob_Baixo_Peso)
[1] 0.02005777

```

Probabilidade de um indivíduo aleatório estar adequado ou eutrófico: **2,01%**

### Sobrepeso:

O total de adultos nessa categoria é de **2165**, como apresentado na descrição da coluna **CO\_ESTADO\_NUTRI\_ADULTO**.

Para sabermos qual a probabilidade de um adulto selecionado aleatoriamente estar com o estado nutricional “sobrepeso”, nós podemos fazer o seguinte cálculo:

$\text{Total\_Adultos\_Sobrepeso} / \text{Total\_Adultos\_Categorizados\_Nutricionalmente}$

```

> ## SOBREPESO
> Total_Adultos_Sobrepeso <- 2165
> Prob_Sobrepeso <- Total_Adultos_Sobrepeso / Total_Adultos_Categorizados_Nutricionalmente
> print(Prob_Sobrepeso)
[1] 0.3474005

```

Probabilidade de um indivíduo aleatório estar adequado ou eutrófico: **34,74%**

### Obesidade I:

O total de adultos nessa categoria é de **1311**, como apresentado na descrição da coluna **CO\_ESTADO\_NUTRI\_ADULTO**.

Para sabermos qual a probabilidade de um adulto selecionado aleatoriamente estar com o estado nutricional “obesidade grau I”, nós podemos fazer o seguinte cálculo:

$\text{Total\_Adultos\_Obesidade\_Grau\_I} / \text{Total\_Adultos\_Categorizados\_Nutricionalmente}$

```

> ## OBESIDADE GRAU 1
> Total_Adultos_Obesidade_Grau_I <- 1311
> Prob_Obesidade_Grau_I <- Total_Adultos_Obesidade_Grau_I / Total_Adultos_Categorizados_Nutricionalmente
> print(Prob_Obesidade_Grau_I)
[1] 0.2103659

```

Probabilidade de um indivíduo aleatório estar adequado ou eutrófico: **21,03%**

### Obesidade II:

O total de adultos nessa categoria é de **555**, como apresentado na descrição da coluna **CO\_ESTADO\_NUTRI\_ADULTO**.

Para sabermos qual a probabilidade de um adulto selecionado aleatoriamente estar com o estado nutricional “obesidade grau II”, nós podemos fazer o seguinte cálculo:

$\text{Total\_Adultos\_Obesidade\_Grau\_II} / \text{Total\_Adultos\_Categorizados\_Nutricionalmente}$

```
> ## OBESIDADE GRAU 2
> Total_Adultos_Obesidade_Grau_II <- 555
> Prob_Obesidade_Grau_II <- Total_Adultos_Obesidade_Grau_II / Total_Adultos_Categorizados_Nutricionalmente
> print(Prob_Obesidade_Grau_II)
[1] 0.08905648
```

Probabilidade de um indivíduo aleatório estar adequado ou eutrófico: **8,90%**

### Obesidade III:

O total de adultos nessa categoria é de **237**, como apresentado na descrição da coluna **CO\_ESTADO\_NUTRI\_ADULTO**.

Para sabermos qual a probabilidade de um adulto selecionado aleatoriamente estar com o estado nutricional “obesidade grau III”, nós podemos fazer o seguinte cálculo:

$\text{Total\_Adultos\_Obesidade\_Grau\_III} / \text{Total\_Adultos\_Categorizados\_Nutricionalmente}$

```
> ## OBESIDADE GRAU 3
> Total_Adultos_Obesidade_Grau_III <- 237
> Prob_Obesidade_Grau_III <- Total_Adultos_Obesidade_Grau_III / Total_Adultos_Categorizados_Nutricionalmente
> print(Prob_Obesidade_Grau_III)
[1] 0.03802953
```

Probabilidade de um indivíduo aleatório estar adequado ou eutrófico: **3,80%**

Sobre os dados obtidos, é importante ressaltar que nem todos os nossos adultos estavam categorizados com o seu estado nutricional, sendo assim, foi usado como base apenas os adultos categorizados nutricionalmente, como (adequado ou



eutrófico, abaixo do peso, sobrepeso, obesidade I, obesidade II e obesidade III) que juntos somam um total de **6232**.

2. Ao selecionar aleatoriamente um indivíduo com um estado nutricional específico, qual a probabilidade dele ser homem ou mulher?

Com os dados analisados anteriormente, também conseguimos saber qual a probabilidade de um indivíduo selecionado aleatoriamente ser homem ou mulher de acordo com os dados de categoria e sexo.

### Sobrepeso:

Primeiramente, precisamos fazer uma subset, que contenha apenas “homens” que estão com o estado nutricional “sobrepeso” e depois fazer a contagem de todos os resultados que obtivemos.

```
> ## Quantidade de homens adultos sobrepeso
> Homens_Adultos_Sobrepeso <- subset(df, df$CO_ESTADO_NUTRI_ADULTO == "Sobrepeso" & df$SG_SEXO == "M")
> Quantidade_Homens_Adultos_Sobrepeso <- nrow(Homens_Adultos_Sobrepeso)
> Quantidade_Homens_Adultos_Sobrepeso
[1] 334
```

Podemos realizar o mesmo procedimento, desta vez para recuperar a quantidade apenas de mulheres que estão com o estado nutricional “sobrepeso”.

```
> Mulheres_Adultas_Sobrepeso <- subset(df, df$CO_ESTADO_NUTRI_ADULTO == "Sobrepeso" & df$SG_SEXO == "F")
> Quantidade_Mulheres_Adultas_Sobrepeso <- nrow(Mulheres_Adultas_Sobrepeso)
> Quantidade_Mulheres_Adultas_Sobrepeso
[1] 1831
```

Com os dados obtidos, podemos fazer alguns cálculos para sabermos:

Probabilidade de um indivíduo aleatório com sobrepeso ser homem ->

Para isso, vamos utilizar a quantidade de homens sobre peso / total de pessoas com sobrepeso.

```
> Prob_Homem_Aleatorio_Sobrepeso <- Quantidade_Homens_Adultos_Sobrepeso / (Quantidade_Homens_Adultos_Sobrepeso + Quantidade_Mulheres_Adultas_Sobrepeso)
> Prob_Homem_Aleatorio_Sobrepeso
[1] 0.1542725
```

No código R, utilizamos Quantidade\_Homens\_Adultos\_Sobrepeso + Quantidade\_Mulheres\_Adultas\_Sobrepeso, pois esse valor é referente ao total de pessoas com sobrepeso.

Sendo assim, a probabilidade de um indivíduo categorizado como sobrepeso ser homem é **15.43%**.

Probabilidade de um indivíduo aleatório com sobrepeso ser mulher ->

Esse cálculo, é muito semelhante ao outro, porém baseado no número de mulheres com sobrepeso, com isso temos:

```
> Prob_Mulher_Aleatoria_Sobrepeso <- Quantidade_Mulheres_Adultas_Sobrepeso / (Quantidade_Homens_Adultos_Sobrepeso + Quantidade_Mulheres_Adultas_Sobrepeso)
> Prob_Mulher_Aleatoria_Sobrepeso
[1] 0.8457275
```

Sendo assim, a probabilidade de um indivíduo categorizado como sobrepeso ser mulher é **84.57%**.

### Obesidade I:

Quantidade de homens com obesidade I:

```
> Homens_Adultos_Obesidade_Grau_I <- subset(df, df$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau I" & df$SG_SEXO == "M")
> Quantidade_Homens_Adultos_Obesidade_Grau_I <- nrow(Homens_Adultos_Obesidade_Grau_I)
> Quantidade_Homens_Adultos_Obesidade_Grau_I
[1] 175
```

Quantidade de mulheres com obesidade I:

```
> Mulheres_Adultas_Obesidade_Grau_I <- subset(df, df$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau I" & df$SG_SEXO == "F")
> Quantidade_Mulheres_Adultas_Obesidade_Grau_I <- nrow(Mulheres_Adultas_Obesidade_Grau_I)
> Quantidade_Mulheres_Adultas_Obesidade_Grau_I
[1] 1136
```

Probabilidade de um indivíduo aleatório com “obesidade I” ser homem ->

```
> Prob_Homem_Aleatorio_Obesidade_Grau_I <- Quantidade_Homens_Adultos_Obesidade_Grau_I / (Quantidade_Homens_Adultos_Obesidade_Grau_I + Quantidade_Mulheres_Adultas_Obesidade_Grau_I)
> Prob_Homem_Aleatorio_Obesidade_Grau_I
[1] 0.1334859
```

Sendo assim, a probabilidade de um indivíduo categorizado como obesidade I ser homem é **13.35%**.

Probabilidade de um indivíduo aleatório com “obesidade I” ser mulher ->

```
> Prob_Mulher_Aleatoria_Obesidade_Grau_I <- Quantidade_Mulheres_Adultas_Obesidade_Grau_I / (Quantidade_Homens_Adultos_Obesidade_Grau_I + Quantidade_Mulheres_Adultas_Obesidade_Grau_I)
> Prob_Mulher_Aleatoria_Obesidade_Grau_I
[1] 0.8665141
```

Sendo assim, a probabilidade de um indivíduo categorizado como sobrepeso ser mulher é **86.65%**.

### Obesidade II:

Quantidade de homens com obesidade II:

```
> ## Quantidade de homens adultos obesidade II
> Homens_Adultos_Obesidade_Grau_II <- subset(df, df$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau II" & df$SG_SEXO == "M")
> Quantidade_Homens_Adultos_Obesidade_Grau_II <- nrow(Homens_Adultos_Obesidade_Grau_II)
> Quantidade_Homens_Adultos_Obesidade_Grau_II
[1] 48
```

Quantidade de mulheres com obesidade II:

```
> ## Quantidade de mulheres adultas obesidade II
> Mulheres_Adultas_Obesidade_Grau_II <- subset(df, df$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau II" & df$SG_SEXO == "F")
> Quantidade_Mulheres_Adultas_Obesidade_Grau_II <- nrow(Mulheres_Adultas_Obesidade_Grau_II)
> Quantidade_Mulheres_Adultas_Obesidade_Grau_II
[1] 507
```

Probabilidade de um indivíduo aleatório com “obesidade II” ser homem ->

```
> Prob_Homem_Aleatorio_Obesidade_Grau_II <- Quantidade_Homens_Adultos_Obesidade_Grau_II / (Quantidade_Homens_Adultos_Obesidade_Grau_II + Quantidade_Mulheres_Adultas_Obesidade_Grau_II)
> Prob_Homem_Aleatorio_Obesidade_Grau_II
[1] 0.08648649
```

Sendo assim, a probabilidade de um indivíduo categorizado como obesidade I ser homem é **8.65%**.

Probabilidade de um indivíduo aleatório com “obesidade II” ser mulher ->

```
> Prob_Mulher_Aleatoria_Obesidade_Grau_II <- Quantidade_Mulheres_Adultas_Obesidade_Grau_II / (Quantidade_Homens_Adultos_Obesidade_Grau_II + Quantidade_Mulheres_Adultas_Obesidade_Grau_II)
> Prob_Mulher_Aleatoria_Obesidade_Grau_II
[1] 0.9135135
```

Sendo assim, a probabilidade de um indivíduo categorizado como sobrepeso ser mulher é **91.35%**.

### Obesidade III:

Quantidade de homens com obesidade III:

```
> Homens_Adultos_Obesidade_Grau_III <- subset(df, df$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau III" & df$SG_SEXO == "M")
> Quantidade_Homens_Adultos_Obesidade_Grau_III <- nrow(Homens_Adultos_Obesidade_Grau_III)
> Quantidade_Homens_Adultos_Obesidade_Grau_III
[1] 27
```

Quantidade de mulheres com obesidade III:

```
> Mulheres_Adultas_Obesidade_Grau_III <- subset(df, df$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau III" & df$SG_SEXO == "F")
> Quantidade_Mulheres_Adultas_Obesidade_Grau_III <- nrow(Mulheres_Adultas_Obesidade_Grau_III)
> Quantidade_Mulheres_Adultas_Obesidade_Grau_III
[1] 210
```

Probabilidade de um indivíduo aleatório com “obesidade III” ser homem ->

```
> Prob_Homem_Aleatorio_Obesidade_Grau_III <- Quantidade_Homens_Adultos_Obesidade_Grau_III / (Quantidade_Homens_Adultos_Obesidade_Grau_III + Quantidade_Mulheres_Adultas_Obesidade_Grau_III)
> Prob_Homem_Aleatorio_Obesidade_Grau_III
[1] 0.1139241
```

Sendo assim, a probabilidade de um indivíduo categorizado como obesidade I ser homem é **11.4%**.

Probabilidade de um indivíduo aleatório com “obesidade III” ser mulher ->

```
> Prob_Mulher_Aleatoria_Obesidade_Grau_III <- Quantidade_Mulheres_Adultas_Obesidade_Grau_III / (Quantidade_Homens_Adultos_Obesidade_Grau_III + Quantidade_Mulheres_Adultas_Obesidade_Grau_III)
> Prob_Mulher_Aleatoria_Obesidade_Grau_III
[1] 0.8860759
```

Sendo assim, a probabilidade de um indivíduo categorizado como sobrepeso ser mulher é **88.6%**.

## Intervalo de confiança

Abaixo há uma breve investigação sobre a relação da fase da vida no IMC (Índice de Massa Corporal) de indivíduos selecionados aleatoriamente, o objetivo é verificar se o fato de uma pessoa ser adulta ou idosa é um aspecto que influencia o aumento do IMC.

Obs.: Para a realização desta análise, foi gerada uma nova coluna em Excel que arredonda os valores de IMC, que estavam sendo passadas sem vírgula para o RStudio, chamada DS\_IMC\_ROUND, contida no arquivo sisvan\_estado\_nutricional\_2021\_round.

#Filtrando as amostras por fase da vida

```
imc_adultos = df$DS_IMC_ROUND[df$DS_FASE_VIDA == "ADULTO"]
imc_idosos = df$DS_IMC_ROUND[df$DS_FASE_VIDA == "IDOSO"]
```

#Consultando o sumário estatístico dos dados filtrados

```
summary(imc_adultos)
```

```
summary(adultos)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
    0.0   24.0   27.0   28.2   32.0   191.0     1
```

```
summary(imc_idosos)
```

```
summary(idosos)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
 12.00   24.00   27.00   27.78   31.00   61.00     1
```

#Selecionando aleatoriamente 100 amostras de cada conjunto

```
amostra_adulto <- head(sample(imc_adultos, 100), 100)
```

```
amostra_idoso <- head(sample(imc_idosos, 100), 100)
```

#Sumário estatístico das amostras

```
summary(amostra_adulto)
```

```
summary(amostra_adulto)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 17.00   23.00   28.00   27.65   31.00   45.00
```

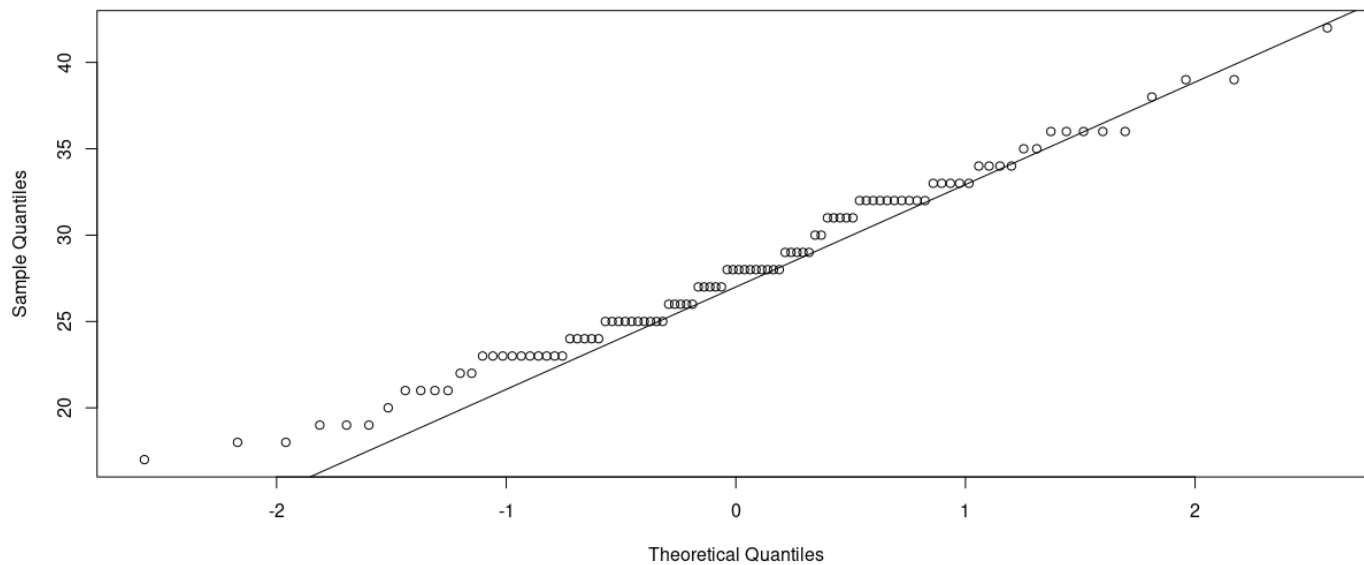
```
summary(amostra_idoso)
```

```
summary(amostra_idoso)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 17.00   24.00   28.00   27.92   32.00   42.00
```

#Verificação da distribuição dos dados

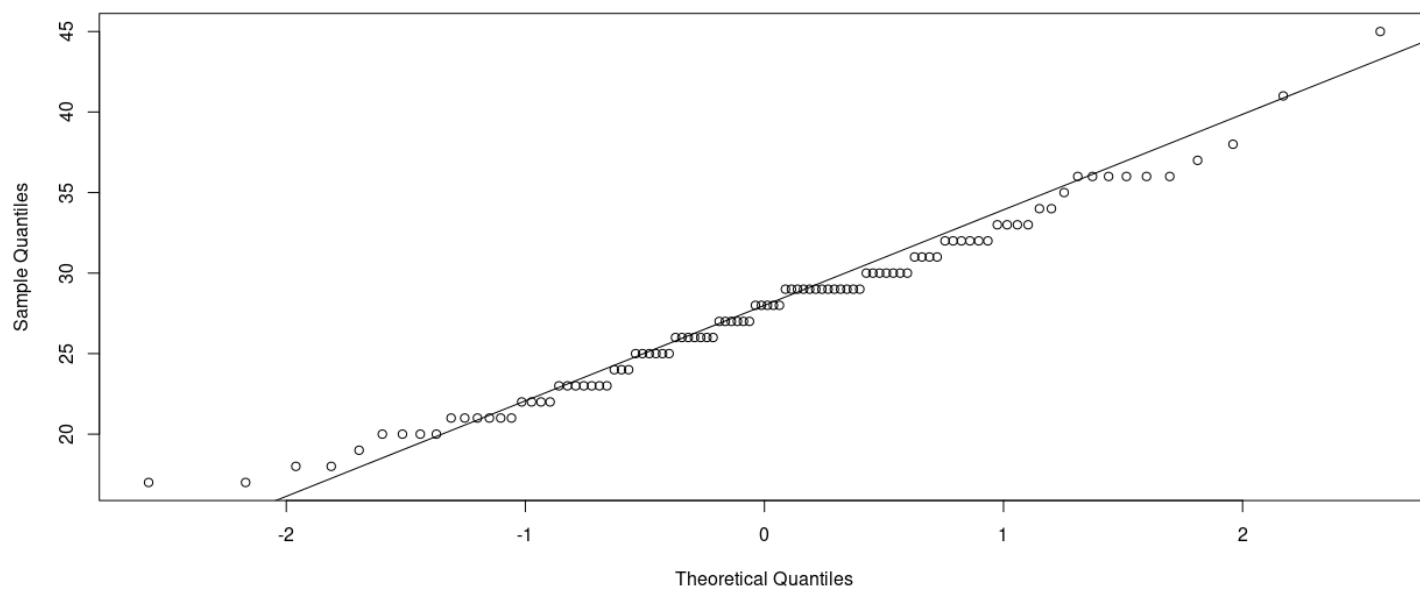
```
qqline(amostra_adulto)
```

Normal Q-Q Plot



qqline(amostra\_idoso)

Normal Q-Q Plot



#Medidas resumo

#Adultos: Média -> 28.65

#Adultos: Desvio padrão -> 5.44

#Idosos: Média -> 27.92

```
#Idosos: Desvio padrão -> 5.30
```

```
sd(amostra_adulto)
```

```
5.442769
```

```
sd(amostra_idoso)
```

```
5.302334
```

```
#Intervalo de confiança de 95%
```

```
z_star_95 <- qnorm(0.975)
```

```
#adultos
```

```
menor_a = (28.65) - z_star_95 * (5.44/ sqrt(100))
```

```
27.58378
```

```
maior_a = (28.65) + z_star_95 * (5.44/ sqrt(100))
```

```
29.71622
```

```
#idosos
```

```
menor_i = (27.92) - z_star_95 * (5.30/ sqrt(100))
```

```
26.88122
```

```
maior_i = (27.92) + z_star_95 * (5.30/ sqrt(100))
```

```
28.95878
```

Como pode ser observado, há sobreposição dos intervalos, então pode-se constatar que o IMC não está relacionado necessariamente à fase da vida de um indivíduo e que este fator não se mostra relevante para este índice, logo, independente se uma pessoa é adulta ou idosa, ela pode apresentar um índice de massa corporal considerado adequado ou não.

Um adendo importante em relação a este índice é que, apesar de amplamente utilizado, o IMC acaba sendo um indicador que eventualmente pode apresentar falhas, uma vez que não indica a distribuição do peso de um indivíduo, tampouco o tipo de matéria que compõe este peso. Por exemplo, o IMC não

diferencia se determinada quantidade de quilogramas são de músculo (massa magra) ou gordura, podendo levar, em alguns casos, à uma análise equivocada, classificando uma pessoa com muita massa magra, como um fisiculturista, por exemplo, como uma pessoa com sobrepeso ou com algum grau de obesidade, requerendo então uma abordagem mais individualizada.

### Teste de hipótese

O teste de hipótese a seguir busca responder à seguinte questão: Os homens possuem maior chance de ter sobrepeso?

Nível de significância de  $\alpha = 0,05$

```
amostra <-
sisvan_estado_nutricional_2021[sample(nrow(sisvan_estado_nutricional_2021),
1000), ]

homens <- sum(amostra$SG_SEXO == "M" &
(amostra$CO_ESTADO_NUTRI_ADULTO == "Sobrepeso" |
amostra$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau I" |
amostra$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau II" |
amostra$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau III"))

mulheres <- sum(amostra$SG_SEXO == "F" &
(amostra$CO_ESTADO_NUTRI_ADULTO == "Sobrepeso" |
amostra$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau I" |
amostra$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau II" |
amostra$CO_ESTADO_NUTRI_ADULTO == "Obesidade Grau III"))

total_homens <- sum(amostra$SG_SEXO == "M")
total_mulheres <- sum(amostra$SG_SEXO == "F")

prop.test(c(homens, mulheres), c(total_homens, total_mulheres), alternative =
"greater")
```



```
> prop.test(c(homens, mulheres), c(total_homens, total_mulheres), alternative = "greater")
```

```
2-sample test for equality of proportions with continuity correction
```

```
data: c(homens, mulheres) out of c(total_homens, total_mulheres)
```

```
X-squared = 58.707, df = 1, p-value = 1
```

```
alternative hypothesis: greater
```

```
95 percent confidence interval:
```

```
-0.3271678  1.0000000
```

```
sample estimates:
```

```
prop 1    prop 2
```

```
0.1576763 0.4334651
```