

Transtorno Depressivo e Seus Fatores Associados



Objetivo da Análise

1. **Perfil Depressivo:** Qual é o perfil dos indivíduos (adultos maiores de 18 anos) com sintomas depressivos nos EUA nesse período?
2. **Estilo de Vida e Depressão:** Hábitos saudáveis de alimentação e atividade física estão associados a menores índices de depressão nesta população?



Dados

Os dados utilizados nesta análise foram obtidos do NHANES (National Health and Nutrition Examination Survey), uma pesquisa anual realizada nos EUA que avalia a saúde e nutrição de adultos e crianças.

Valores **nulos** foram removidos e a variável “**DPQ_RANK**” foi criado com a soma de todas as perguntas comportamentais.

Ademais, as categorias “Não sei”, “Se recusou a responder”, valores 7 e 9 respectivamente foram remapeados para o valor **mais frequente** 0, ou seja “Nenhuma vez” para não interferir no Score.

Dataset Final

Entradas

4702

Features

32

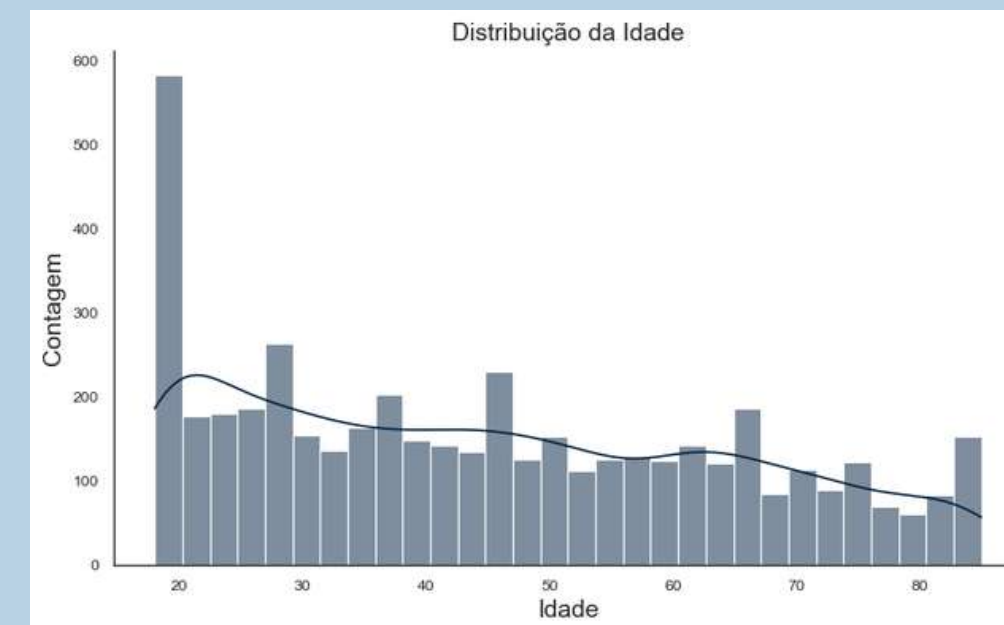
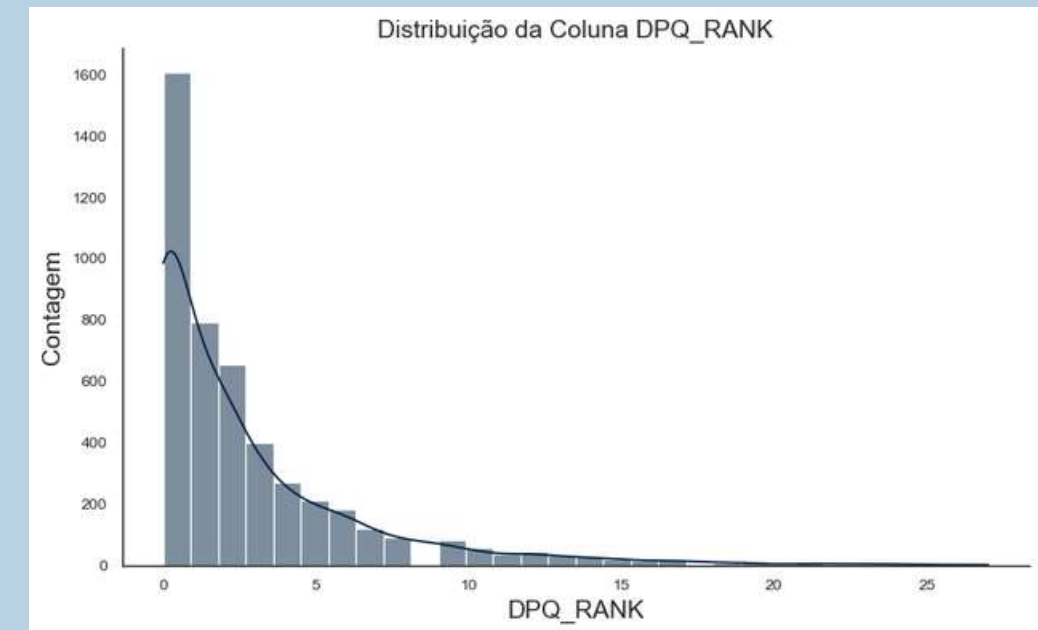
ANÁLISE EXPLORATÓRIA (EDA) UNIVARIADA

Na feature “DPQ_RANK”, possuímos distribuição assimétrica à direita. Logo, possuímos mais pessoas com uma pontuação baixa.

Em relação à idade, temos uma distribuição mais uniforme. A média de idade é 44 anos, com mediana em 43 com desvio padrão de 19 anos.

Ademais, a quantidade de horas de atividade física semanal, também está distribuída de forma assimétrica.

DPQ_RANK

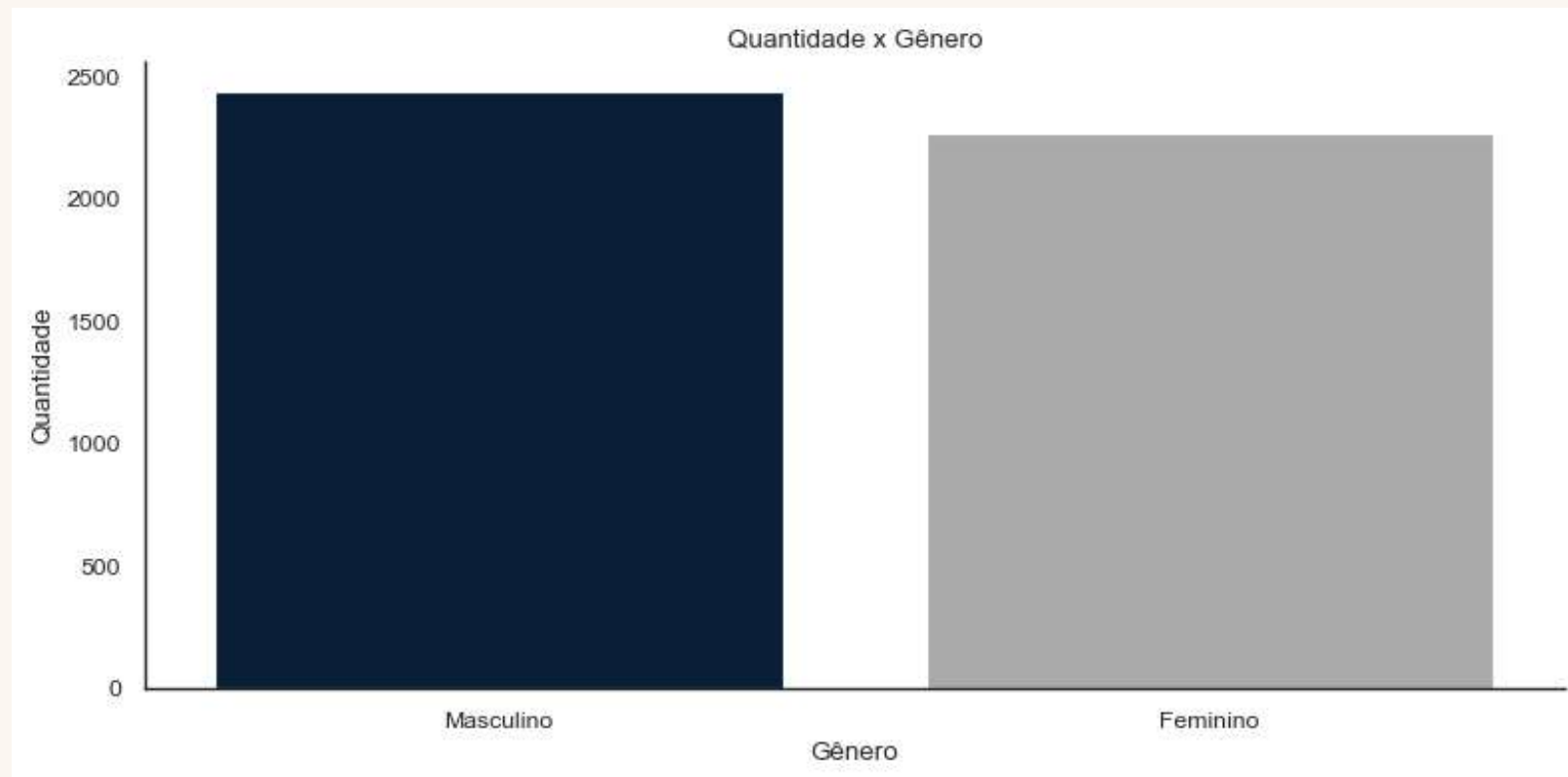


Idade

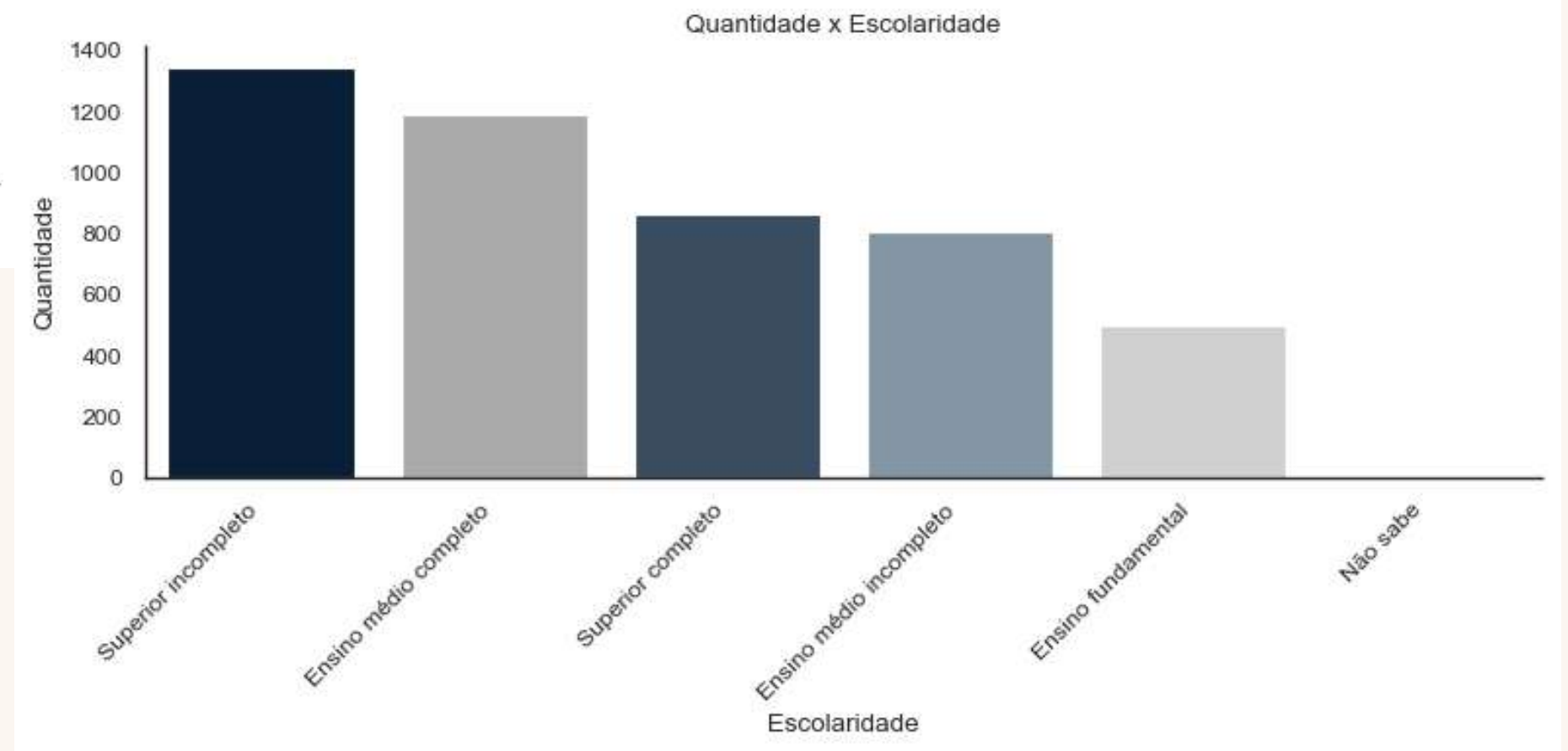
**Atividade
Física**



Variáveis Categóricas

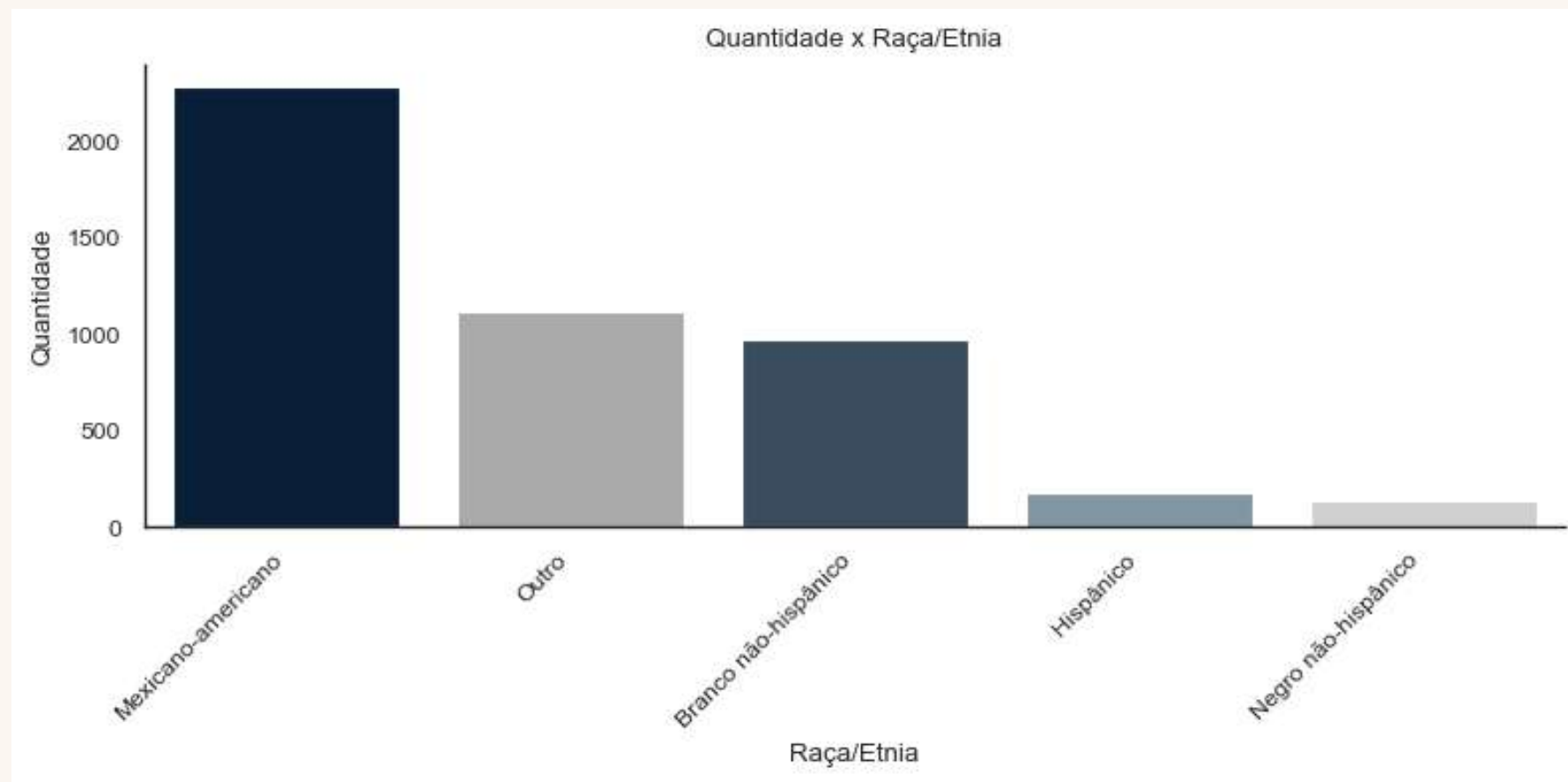


Gênero

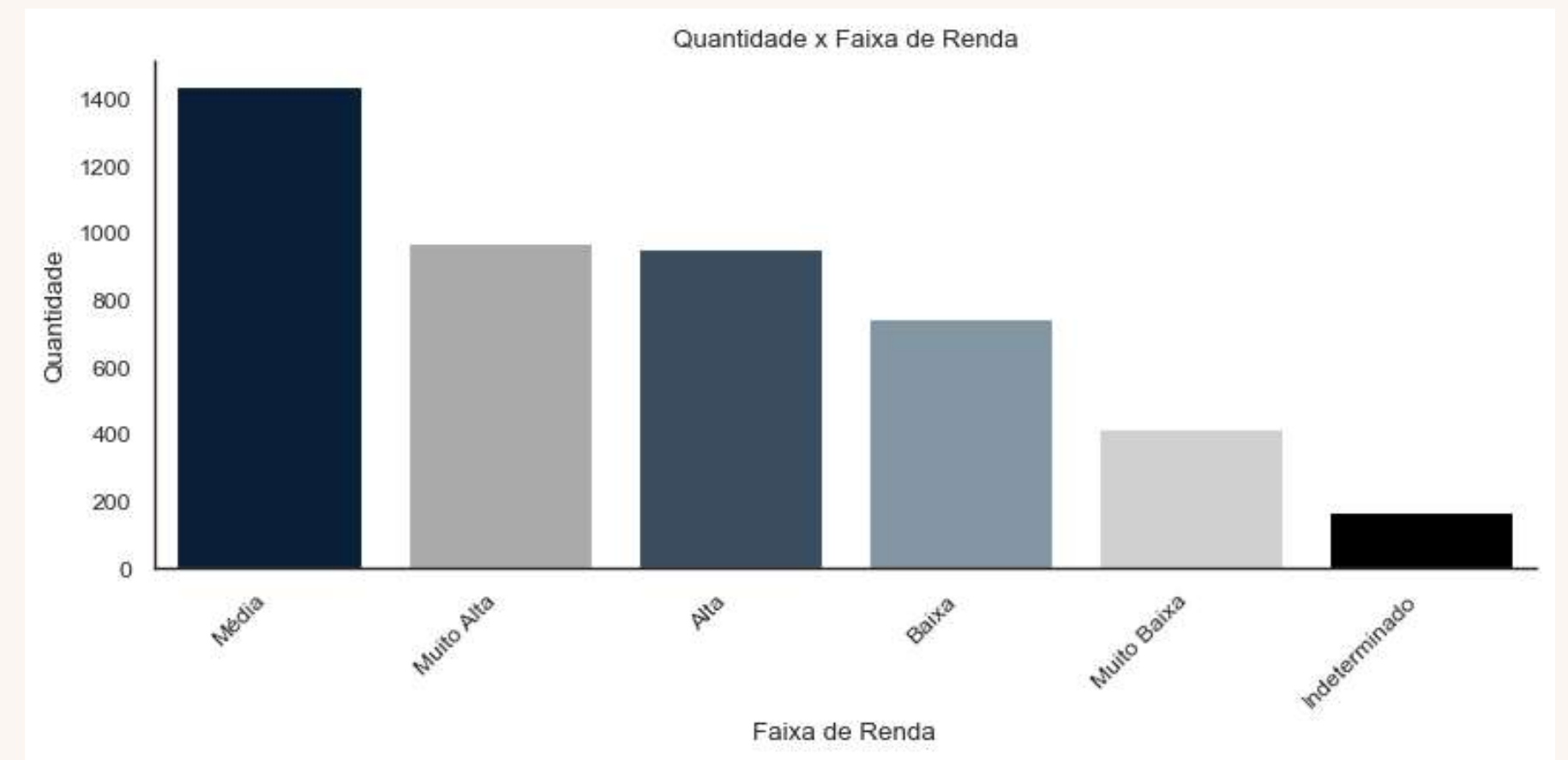


Escolaridade

Variáveis Categóricas



Raça/Etnia



Renda

ANÁLISE EXPLORATÓRIA (EDA) BIVARIADA

Foi feito um teste de hipóteses de correlação de Pearson e observou-se que há uma correlação significativa entre **idade** e o **score de alimentação saudável**. Contudo, não podemos rejeitar a hipótese nula para **quantidade de atividade física** e **alimentação saudável**.

```
: stats.pearsonr(df.RIDAGEYR, df.HEI2015_TOTAL_SCORE)  
:  
: PearsonRResult(statistic=0.29871119696688425, pvalue=1.5921344259949678e-97)
```

```
: stats.pearsonr(df.PAG_MINW, df.HEI2015_TOTAL_SCORE)  
:  
: PearsonRResult(statistic=-0.004644244728806473, pvalue=0.750198103710561)
```

Escolaridade x Gênero

Foi feito um teste de hipóteses chi-quadrado para verificar se há associação entre as variáveis. Iremos assumir um nível de significância = 0.05.

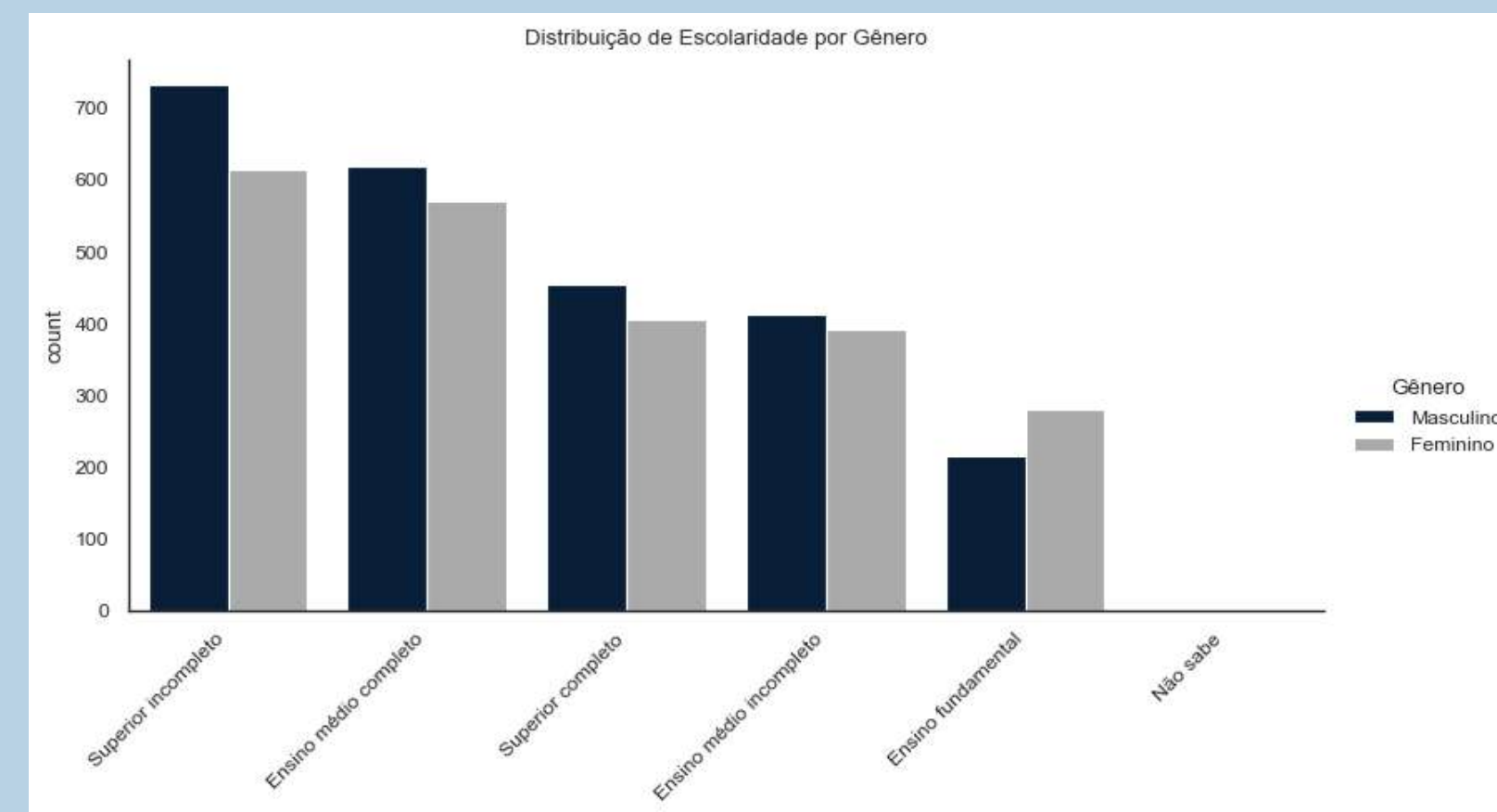
Hipótese Nula (H0): O gênero é independente da escolaridade. Isto é, a escolaridade não varia de acordo com o gênero.

Hipótese Alternativa (Ha): O gênero está associado à escolaridade. Isto é, a escolaridade varia de acordo com o gênero.

Valor qui-quadrado: 18.880052780225615

Valor p: 0.0020235560731163304

Coefficiente de Cramér (V): 0.0634



Renda x Gênero

Foi feito um teste de hipóteses chi-quadrado para verificar se há associação entre as variáveis. Iremos assumir um nível de significância = 0.05.

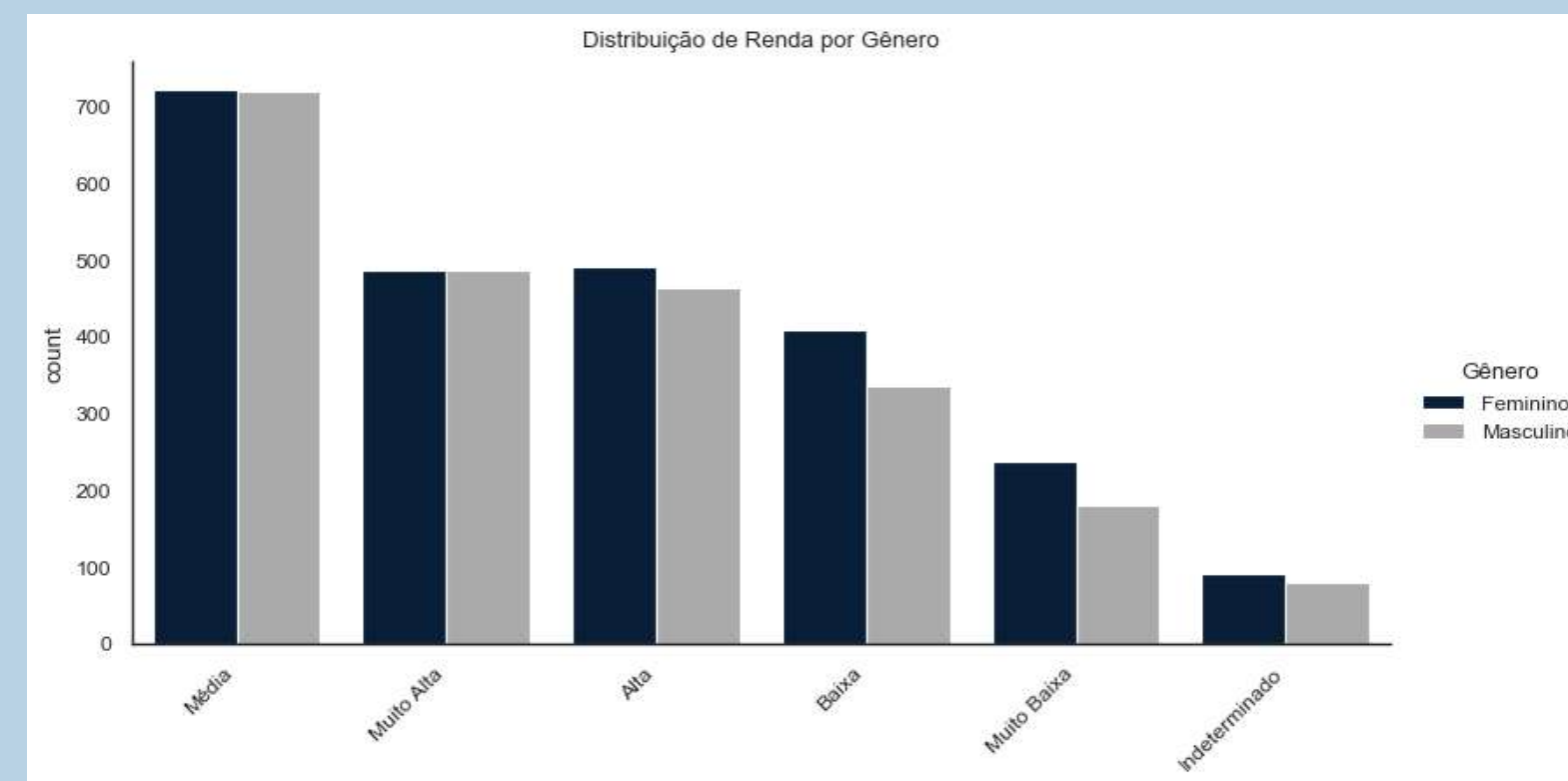
Hipótese Nula (H0): O gênero é independente da renda. Isto é, a renda não varia de acordo com o gênero.

Hipótese Alternativa (Ha): O gênero está associado à renda. Isto é, a renda varia de acordo com o gênero.

Valor qui-quadrado: 10.015679623513124

Valor p: 0.07479218809287837

Como o p-valor > nível de significância, não rejeitamos a hipótese nula.



Escolaridade x Raça/Etnia

Foi feito um teste de hipóteses chi-quadrado para verificar se há associação entre as variáveis. Iremos assumir um nível de significância = 0.05.

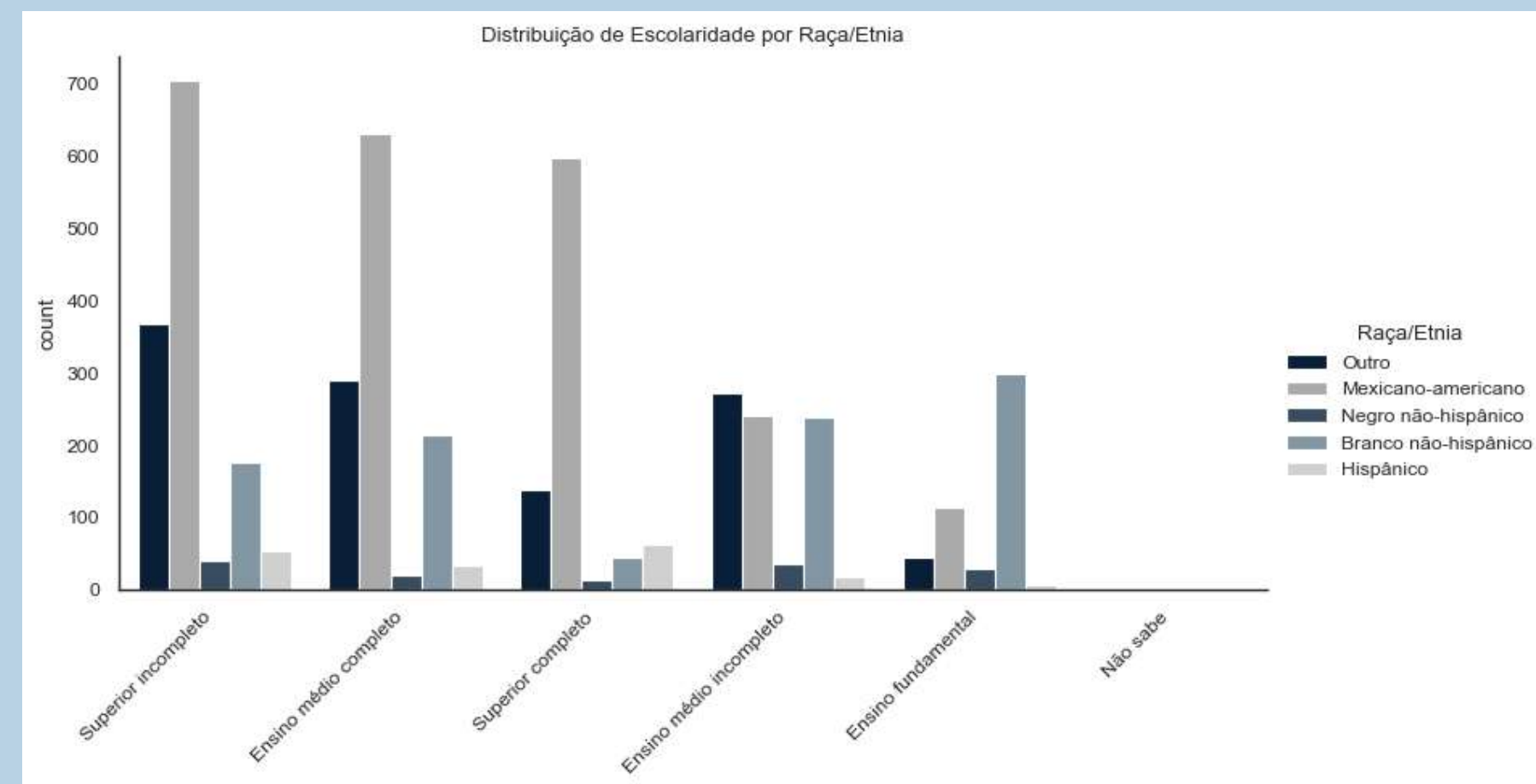
Hipótese Nula (H0): A escolaridade é independente da raça/etnia. Isto é, a escolaridade não varia de acordo com a raça/etnia.

Hipótese Alternativa (Ha): A escolaridade está associado à raça/etnia. Isto é, a escolaridade varia de acordo com a raça/etnia.

Valor qui-quadrado: 937.3671187796466

Valor p: 8.70709280288252e-186

Coeficiente de Cramér (V): 0.2232



Renda x Raça/Etnia

Foi feito um teste de hipóteses chi-quadrado para verificar se há associação entre as variáveis. Iremos assumir um nível de significância = 0.05.

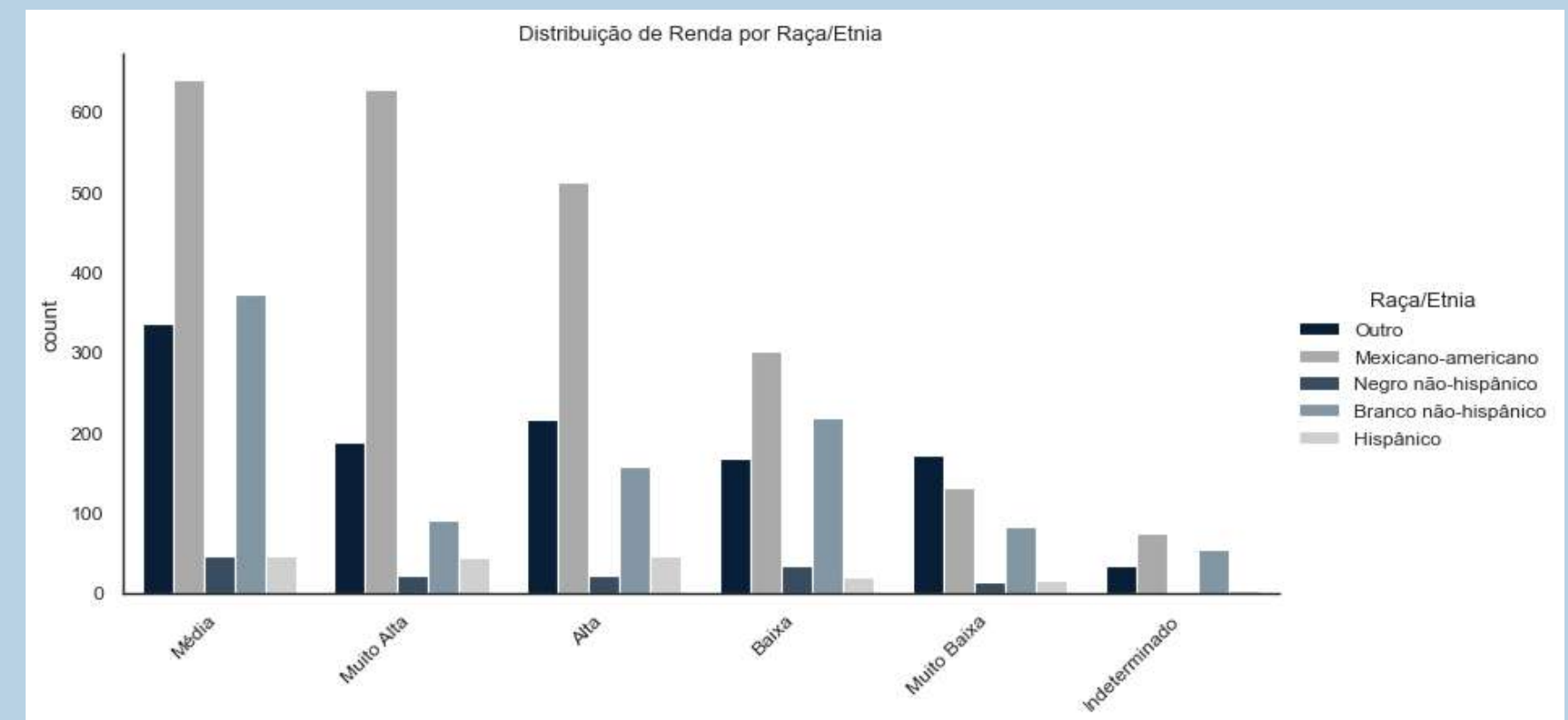
Hipótese Nula (H0): A renda é independente da raça/etnia. Isto é, a renda não varia de acordo com a raça/etnia.

Hipótese Alternativa (Ha): A renda está associado à raça/etnia. Isto é, a renda varia de acordo com a raça/etnia.

Valor qui-quadrado: 303.2943179000313

Valor p: 1.716338909107887e-52

Coeficiente de Cramér (V): 0.1270



Aderência a atividade física vs DPQ040

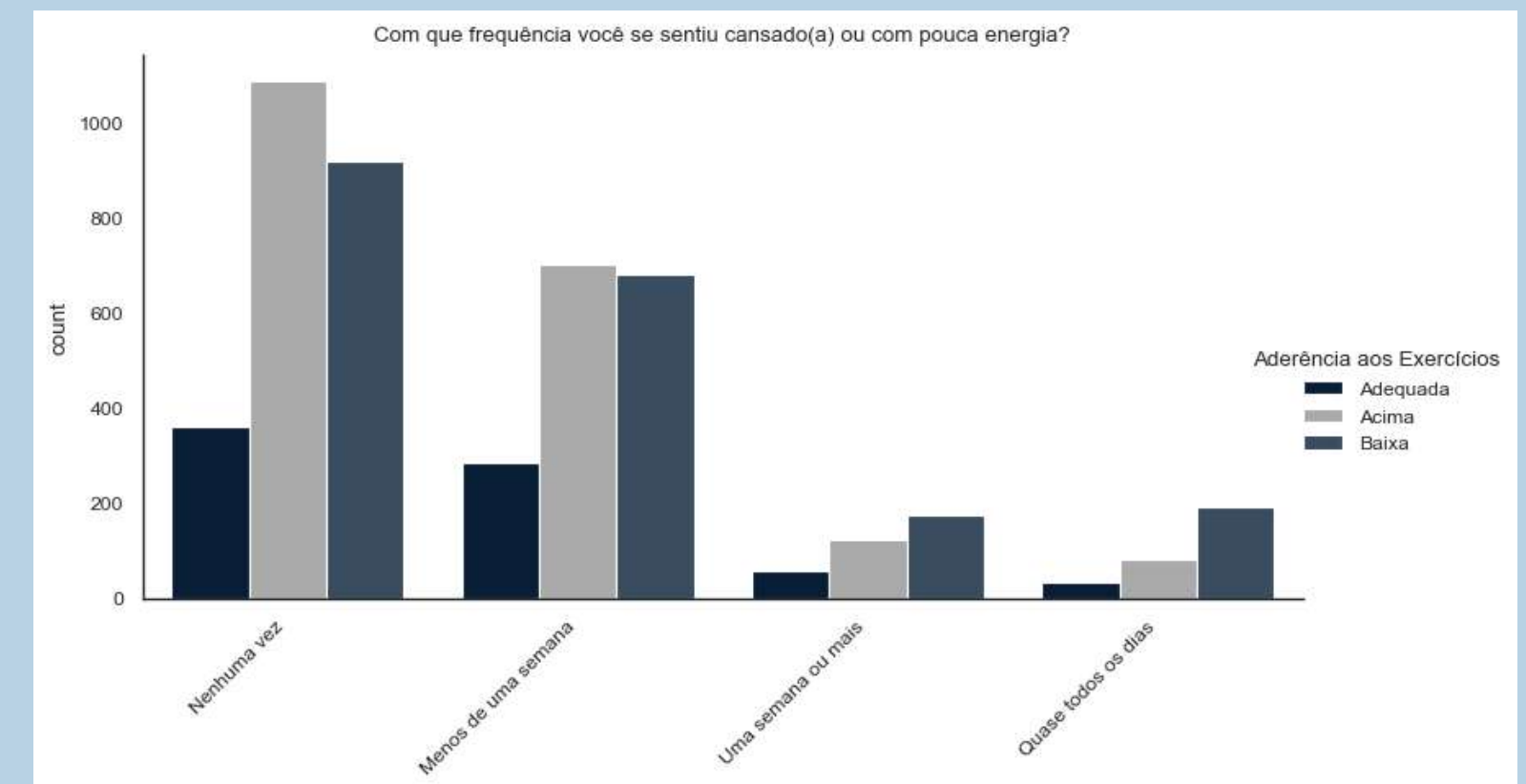
Foi feito um teste de hipóteses chi-quadrado para verificar se há associação entre as variáveis. Iremos assumir um nível de significância = 0.05.

DPQ040: Nas últimas 2 semanas, com que frequência você se sentiu cansado(a) ou com pouca energia?

Valor qui-quadrado: 77.70180535557223

Valor p: 1.0650918553657701e-14

Coeficiente de Cramér (V): 0.0909



Renda vs DPQ090

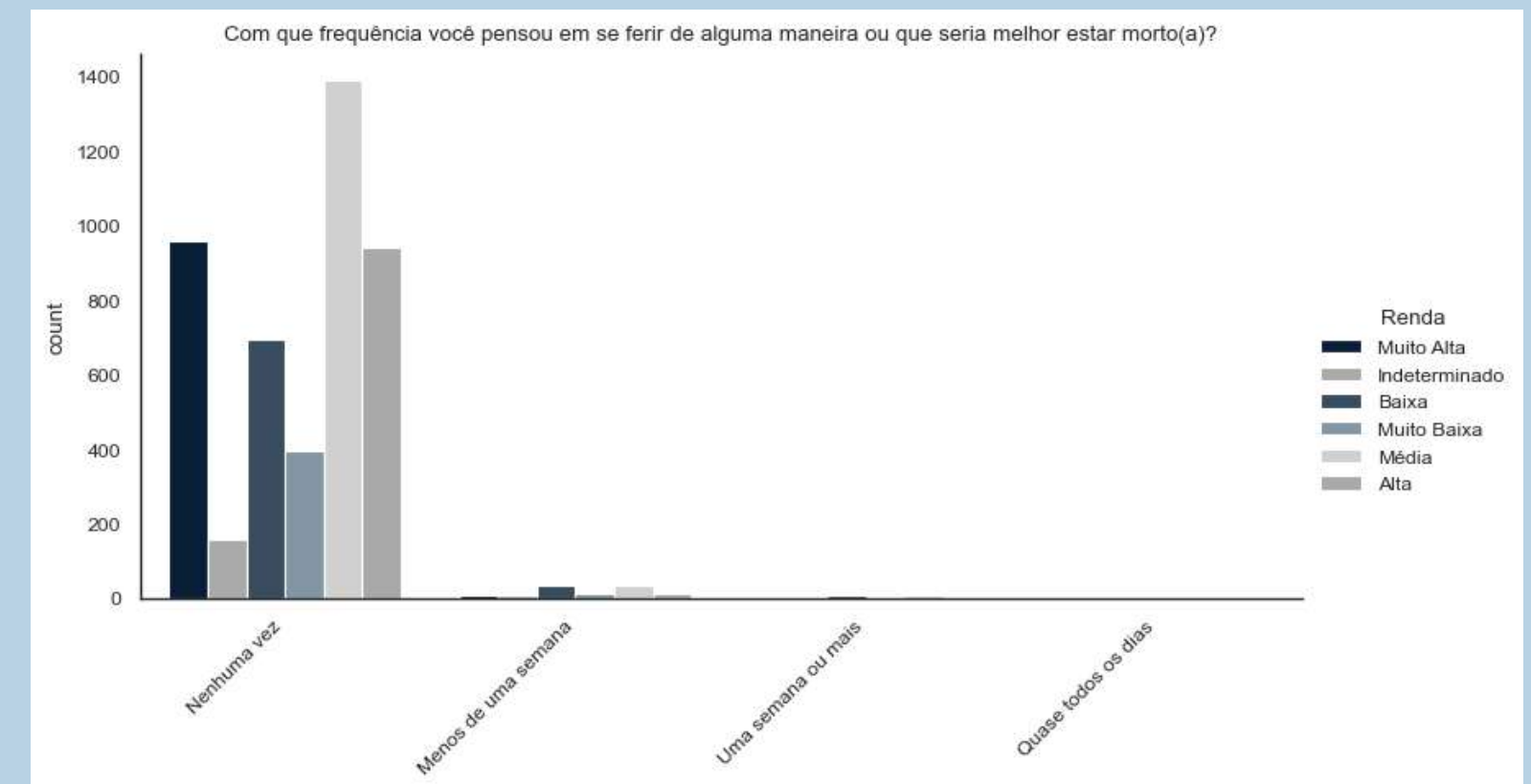
Foi feito um teste de hipóteses chi-quadrado para verificar se há associação entre as variáveis. Iremos assumir um nível de significância = 0.05.

DPQ090: Nas últimas 2 semanas, com que frequência você pensou em se ferir de alguma maneira ou que seria melhor estar morto(a)?

Valor qui-quadrado: 62.689760379224396

Valor p: 8.64603949121032e-08

Coeficiente de Cramér (V): 0.0667



Gênero vs DPQ060

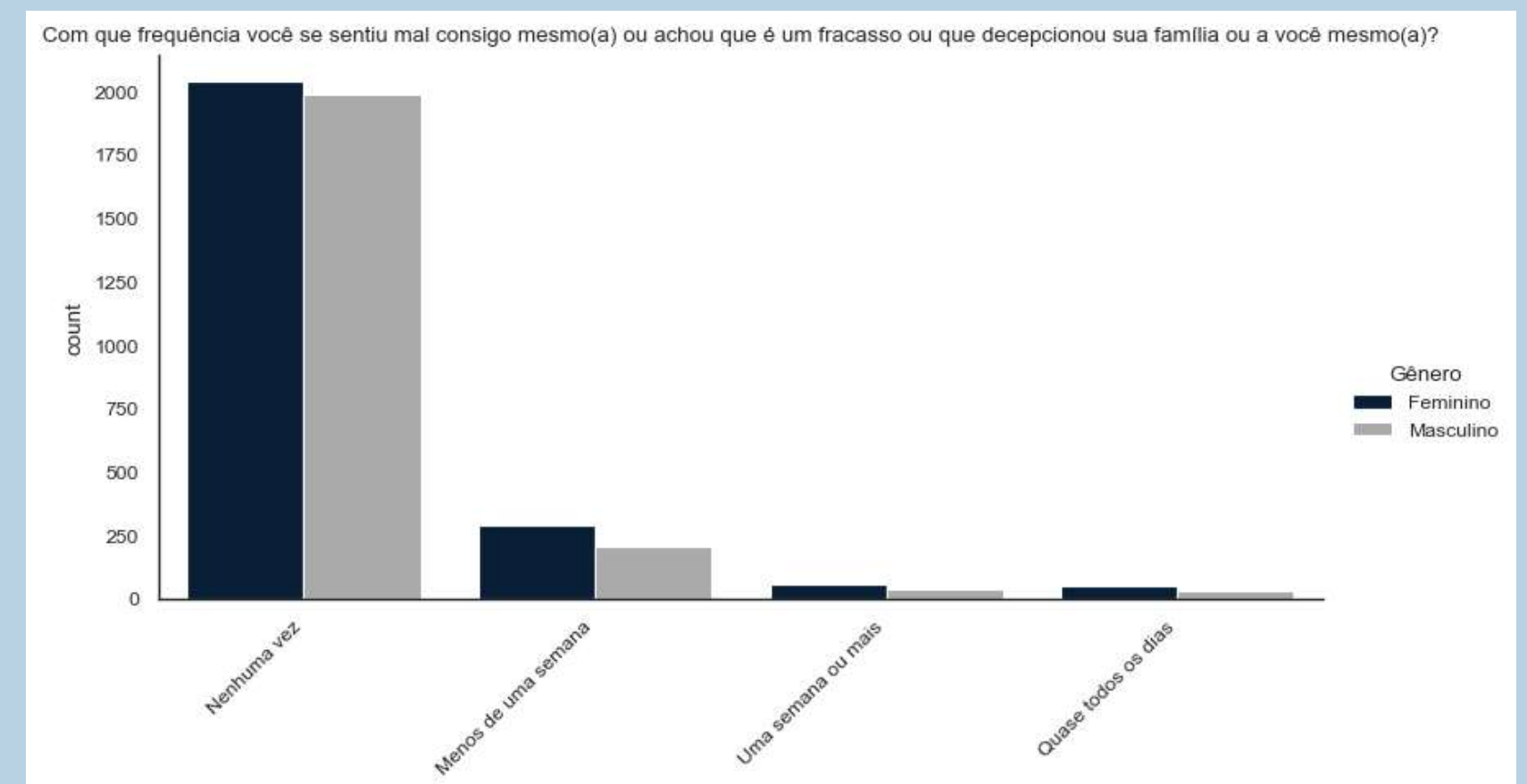
Foi feito um teste de hipóteses chi-quadrado para verificar se há associação entre as variáveis. Iremos assumir um nível de significância = 0.05.

DPQ060: Nas últimas 2 semanas, com que frequência você se sentiu mal consigo mesmo(a) ou achou que é um fracasso ou que decepcionou sua família ou a você mesmo(a)?

Valor qui-quadrado: 16.112124740569072

Valor p: 0.0010755118838349207

Coeficiente de Cramér (V): 0.0585



Variável Alvo x Demais Variáveis

A média do **DPQ_RANK** varia conforme o **gênero**?

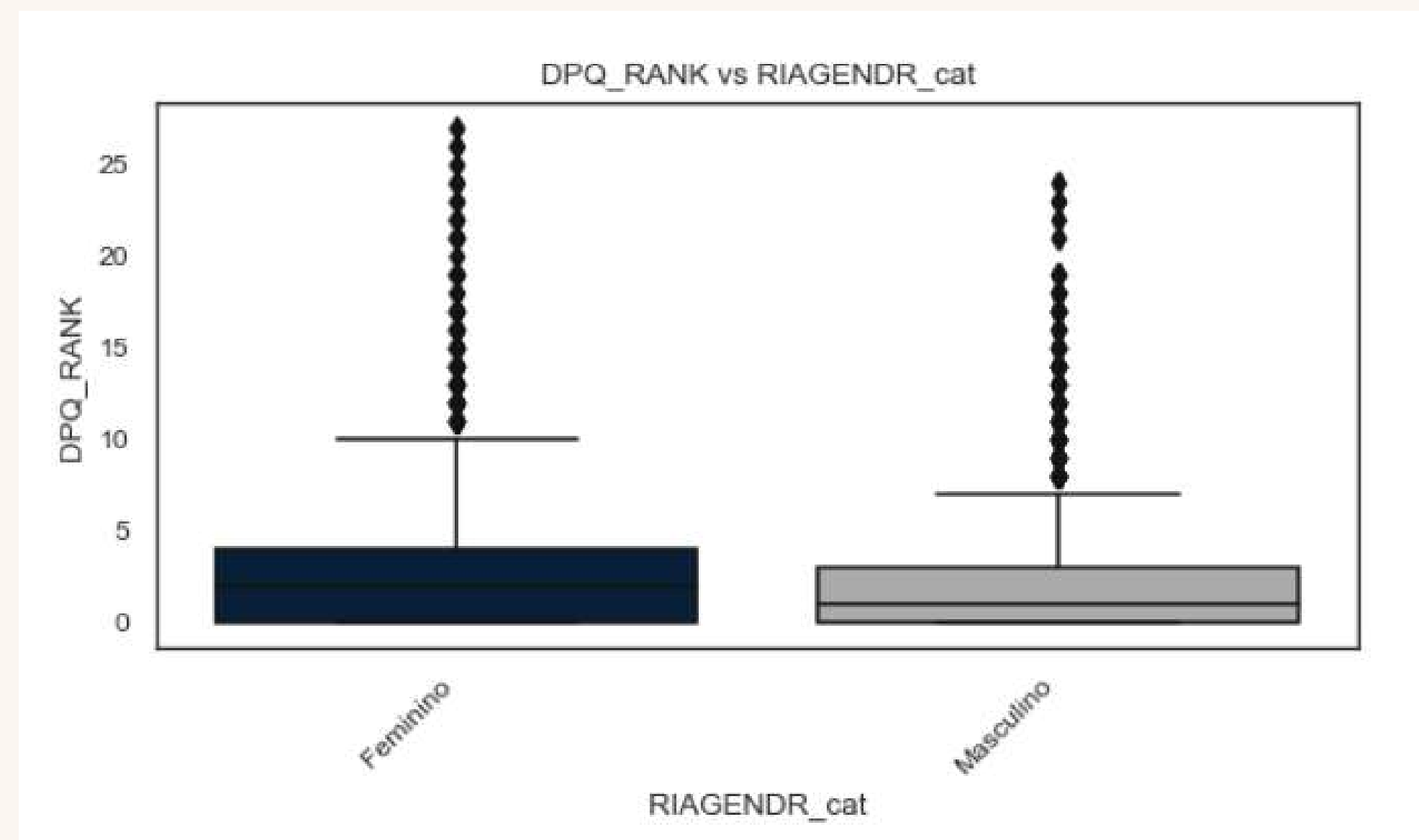
t-statistic: -7.49845291143233

p-value: 7.675103240709812e-14

Rejeitamos a hipótese nula. Há uma diferença significativa entre os grupos.

Média do **DPQ_RANK** para **Homens**: 2.31, IC: (2.17, 2.45)

Média do **DPQ_RANK** para **Mulheres**: 3.13, IC: (2.97, 3.28)



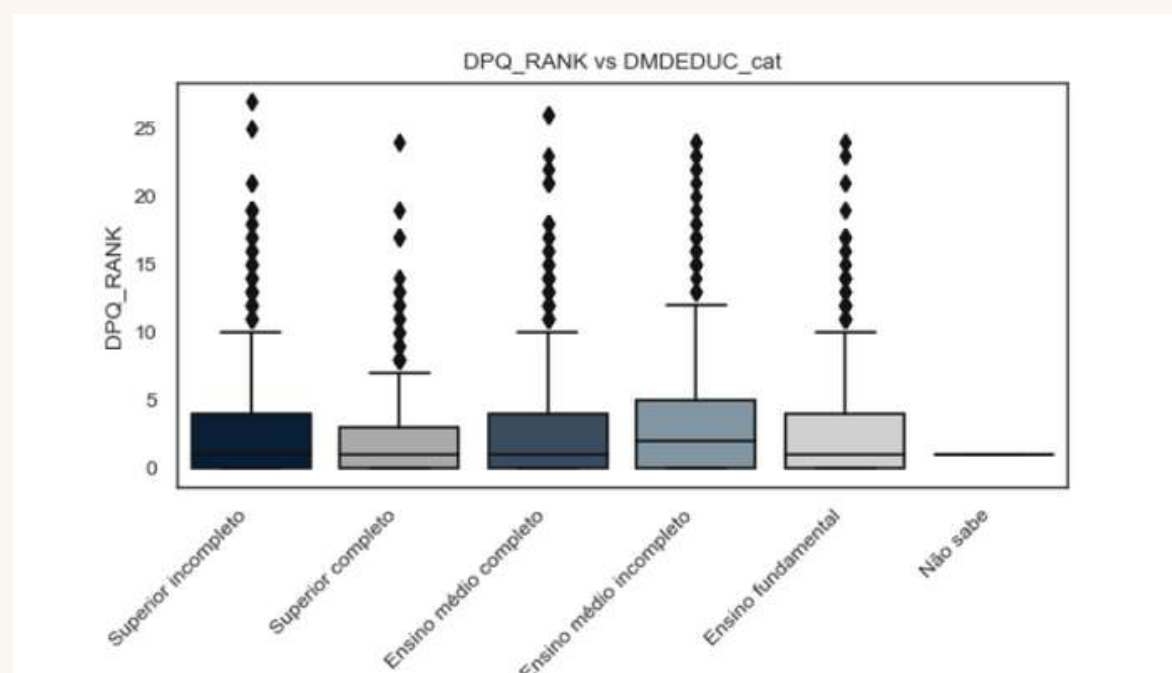
Variável Alvo x Demais Variáveis

A média do **DPQ_RANK** varia conforme o **escolaridade**?

f-statistic: 9.690874918955176

p-value: 3.200709294923921e-09

Rejeitamos a hipótese nula. Há uma diferença significativa entre os grupos.



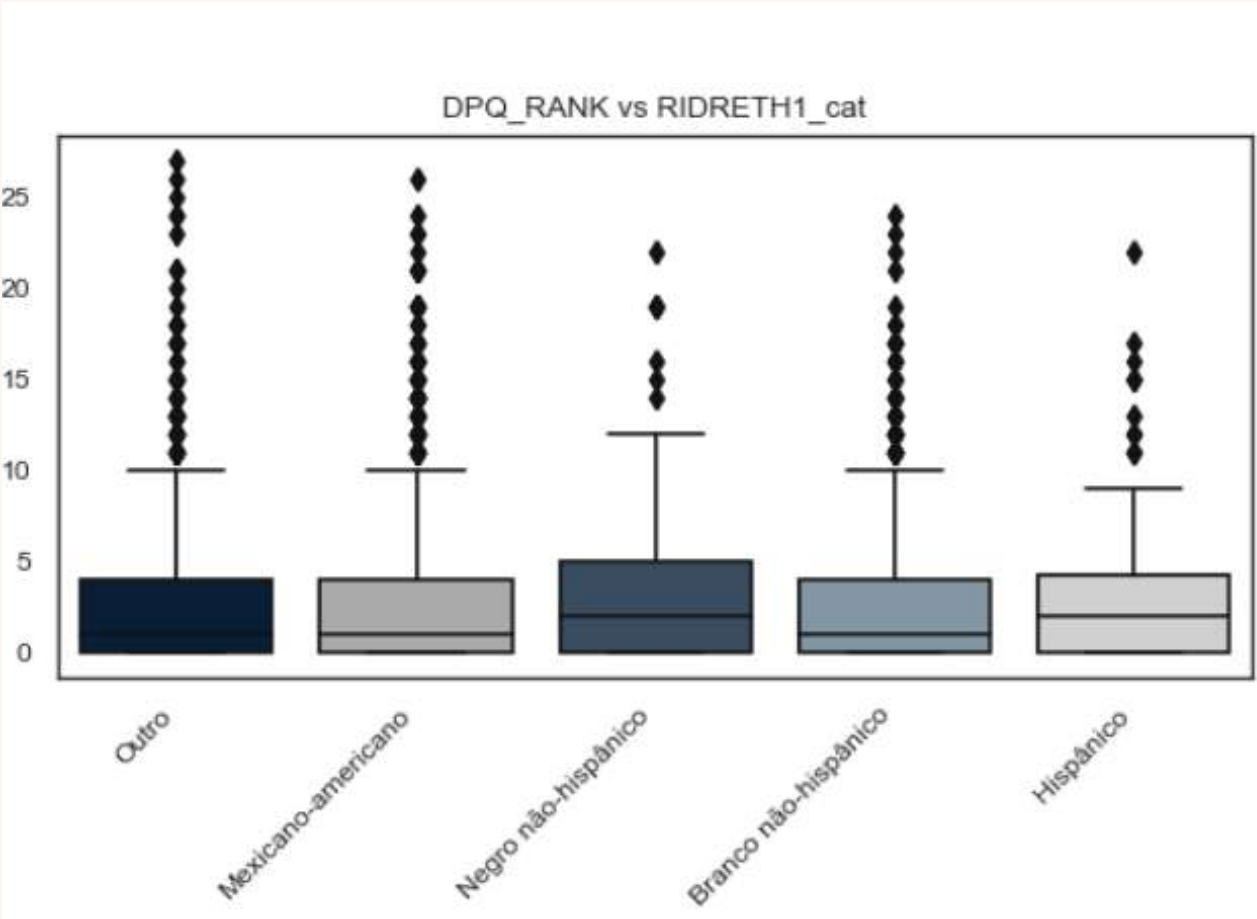
Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
Ensino fundamental	Ensino médio completo	-0.2447	0.8225	-0.8121	0.3227	False
Ensino fundamental	Ensino médio incompleto	0.0553	0.9998	-0.5507	0.6612	False
Ensino fundamental	Não sabe	-2.0986	0.9934	-12.7345	8.5374	False
Ensino fundamental	Superior completo	-1.0823	0.0	-1.6809	-0.4838	True
Ensino fundamental	Superior incompleto	-0.395	0.3313	-0.9527	0.1627	False
Ensino médio completo	Ensino médio incompleto	0.2999	0.489	-0.1847	0.7846	False
Ensino médio completo	Não sabe	-1.8539	0.9963	-12.4836	8.7758	False
Ensino médio completo	Superior completo	-0.8376	0.0	-1.3129	-0.3623	True
Ensino médio completo	Superior incompleto	-0.1503	0.9135	-0.573	0.2724	False
Ensino médio incompleto	Não sabe	-2.1538	0.9925	-12.7857	8.478	False
Ensino médio incompleto	Superior completo	-1.1376	0.0	-1.6583	-0.6168	True
Ensino médio incompleto	Superior incompleto	-0.4503	0.0729	-0.9235	0.0229	False
Não sabe	Superior completo	1.0163	0.9998	-9.6152	11.6477	False
Não sabe	Superior incompleto	1.7036	0.9975	-8.9256	12.3328	False
Superior completo	Superior incompleto	0.6873	0.0003	0.2236	1.151	True

Variável Alvo x Demais Variáveis

A média do **DPQ_RANK** varia conforme o **raça/etnia**?

f-statistic: 2.3899010787592045

p-value: 0.04869122994870757



Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
Branco não-hispânico	Hispânico	0.149	0.9882	-0.6794	0.9774	False
Branco não-hispânico	Mexicano-americano	-0.1303	0.8929	-0.5209	0.2602	False
Branco não-hispânico	Negro não-hispânico	0.6642	0.2747	-0.2501	1.5786	False
Branco não-hispânico	Outro	0.1581	0.8713	-0.2895	0.6057	False
Hispânico	Mexicano-americano	-0.2793	0.8712	-1.0698	0.5112	False
Hispânico	Negro não-hispânico	0.5153	0.7343	-0.6286	1.6591	False
Hispânico	Outro	0.0091	1.0	-0.811	0.8293	False
Mexicano-americano	Negro não-hispânico	0.7946	0.0992	-0.0856	1.6748	False
Mexicano-americano	Outro	0.2885	0.2153	-0.0844	0.6613	False
Negro não-hispânico	Outro	-0.5061	0.5476	-1.4131	0.4008	False

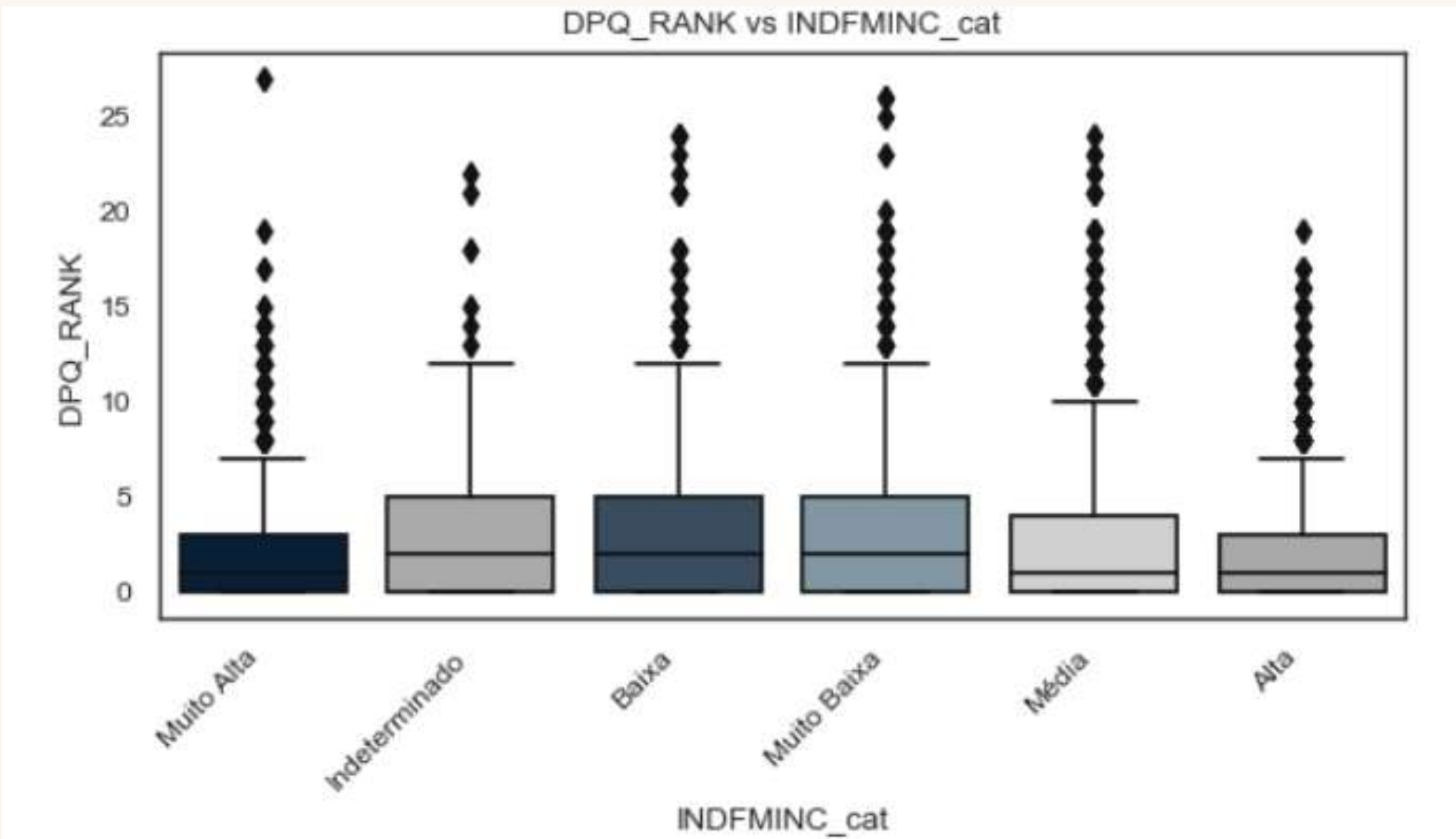
Variável Alvo x Demais Variáveis

A média do **DPQ_RANK** varia conforme o **renda**?

f-statistic: 24.632133738415455

p-value: 1.429650077276925e-24

Rejeitamos a hipótese nula. Há uma diferença significativa entre os grupos.



Multiple Comparison of Means - Tukey HSD, FWER=0.05						
group1	group2	meandiff	p-adj	lower	upper	reject
Alta	Baixa	1.2602	0.0	0.7448	1.7757	True
Alta	Indeterminado	0.7697	0.1222	-0.1057	1.645	False
Alta	Muito Alta	-0.2815	0.5507	-0.7616	0.1986	False
Alta	Muito Baixa	1.4801	0.0	0.8613	2.0988	True
Alta	Média	0.2828	0.4441	-0.1569	0.7226	False
Baixa	Indeterminado	-0.4906	0.6223	-1.3846	0.4035	False
Baixa	Muito Alta	-1.5418	0.0	-2.0552	-1.0283	True
Baixa	Muito Baixa	0.2198	0.927	-0.4251	0.8647	False
Baixa	Média	-0.9774	0.0	-1.4534	-0.5015	True
Indeterminado	Muito Alta	-1.0512	0.0081	-1.9254	-0.177	True
Indeterminado	Muito Baixa	0.7104	0.2793	-0.247	1.6677	False
Indeterminado	Média	-0.4868	0.5801	-1.3395	0.3659	False
Muito Alta	Muito Baixa	1.7616	0.0	1.1445	2.3786	True
Muito Alta	Média	0.5643	0.0032	0.1269	1.0018	True
Muito Baixa	Média	-1.1972	0.0	-1.7835	-0.611	True

Variável Alvo x Demais Variáveis

Há correlação entre o **DPQ_RANK** e a ingestão de **proteínas**?

```
stats.pearsonr(df.DPQ_RANK, df.HEI2015C7_TOTPROT)
```

```
PearsonRResult(statistic=-0.05924679819772615, pvalue=4.800963933060957e-05)
```

Há correlação entre o **DPQ_RANK** e a ingestão de **açúcar**?

```
stats.pearsonr(df.DPQ_RANK, df.HEI2015C13_ADDSUG)
```

```
PearsonRResult(statistic=-0.08475953454485972, pvalue=5.850049565053978e-09)
```

Há correlação entre o **DPQ_RANK** e o **score de alimentação saudável**?

```
stats.pearsonr(df.DPQ_RANK, df.HEI2015_TOTAL_SCORE)
```

```
PearsonRResult(statistic=-0.07094563824018385, pvalue=1.1178707450841809e-06)
```

Variável Alvo x Demais Variáveis

Há correlação entre o **DPQ_RANK** e a **idade**?

```
: stats.pearsonr(df.DPQ_RANK, df.RIDAGEYR)  
:  
: PearsonRResult(statistic=-0.036468746346358584, pvalue=0.012388858956978265)
```

Há correlação entre o **DPQ_RANK** e a quantidade de **exercício físico** semanal?

```
stats.pearsonr(df.DPQ_RANK, df.PAG_MINW)  
  
PearsonRResult(statistic=-0.05666400170732641, pvalue=0.00010123812104228653)
```


Regressão Linear

**Homocedasticidade
foi violada**

OLS Regression Results						
=====						
Dep. Variable:	DPQ_RANK	R-squared:	0.051			
Model:	OLS	Adj. R-squared:	0.047			
Method:	Least Squares	F-statistic:	12.55			
Date:	Sat, 23 Sep 2023	Prob (F-statistic):	1.19e-40			
Time:	18:44:47	Log-Likelihood:	-12756.			
No. Observations:	4702	AIC:	2.555e+04			
Df Residuals:	4681	BIC:	2.569e+04			
Df Model:	20					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	-0.3839	0.247	-1.553	0.121	-0.869	0.101
HEI2015_TOTAL_SCORE	-0.0136	0.005	-2.878	0.004	-0.023	-0.004
PAG_MINW	-5.601e-05	8.32e-05	-0.674	0.501	-0.000	0.000
RIDAGEYR	-0.0070	0.003	-2.261	0.024	-0.013	-0.001
genero_Masculino	-0.7347	0.109	-6.742	0.000	-0.948	-0.521
escolaridade_Ensinomdiocompleto	-0.1964	0.213	-0.921	0.357	-0.614	0.222
escolaridade_Ensinomdioincompleto	-0.0353	0.220	-0.160	0.873	-0.467	0.396
escolaridade_Nosabe	-3.1813	3.673	-0.866	0.386	-10.381	4.019
escolaridade_Superiorcompleto	-0.5715	0.242	-2.362	0.018	-1.046	-0.097
escolaridade_Superiorincompleto	-0.2563	0.216	-1.187	0.235	-0.680	0.167
etnia_Hispnico	0.5592	0.306	1.830	0.067	-0.040	1.158
etnia_Mexicanoamericano	0.3967	0.161	2.457	0.014	0.080	0.713
etnia_Negronohispnico	0.7868	0.329	2.391	0.017	0.142	1.432
etnia_Outro	0.3315	0.173	1.919	0.055	-0.007	0.670
renda_Baixa	1.0992	0.185	5.956	0.000	0.737	1.461
renda_Indeterminado	0.7218	0.307	2.349	0.019	0.120	1.324
renda_MuitoAlta	-0.1186	0.171	-0.695	0.487	-0.453	0.216
renda_MuitoBaixa	1.2861	0.220	5.856	0.000	0.856	1.717
renda_Mdia	0.2319	0.155	1.498	0.134	-0.072	0.535
aderencia_Adequada	0.1710	0.171	0.998	0.319	-0.165	0.507
aderencia_Baixa	0.6393	0.145	4.401	0.000	0.354	0.924
=====						
Omnibus:	2103.688	Durbin-Watson:	1.997			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	10668.414			
Skew:	2.143	Prob(JB):	0.00			
Kurtosis:	9.006	Cond. No.	5.48e+04			

**Resíduos não
seguem uma
distribuição normal**

Conclusão

Perfil Depressivo em Adultos maiores de 18 anos (2005-2006 nos EUA):

- **Gênero e depressão: Mulheres** tendem a apresentar **pontuações mais altas** no DPQ_RANK, indicando uma maior incidência ou gravidade de sintomas depressivos em comparação aos homens.
- **Escolaridade e depressão:** Pessoas com **nível superior completo** tendem a ter **menor** DPQ_RANK (menos sintomas depressivos) comparado a outros níveis de escolaridade.
- **Renda e depressão:** Indivíduos com **renda muito baixa ou baixa** apresentaram **maior** DPQ_RANK (mais sintomas depressivos) do que aqueles com renda alta ou muito alta.

Estilo de Vida e Depressão

- **Alimentação e depressão:** Há uma correlação significativa entre a ingestão de proteínas, açúcar, e o score de alimentação saudável com o DPQ_RANK, embora essas correlações sejam fracas.
- **Atividade física e depressão:** Há uma fraca correlação significativa entre a quantidade de exercício físico semanal e o DPQ_RANK.