

Universidade de São Paulo

Caio Assumpção Rezzadori
Cibele Maria Russo Novelli

Relatório de estágio



Sobre o estágio

Desde julho de 2022, realizo estágio em análise de dados para a Pró-Reitoria de Inclusão e Pertencimento (PRIP) da Universidade de São Paulo, atuando mais especificamente na área de Formação e Vida Profissional sob a supervisão da professora doutora Cibele Russo, dentro do Instituto de Ciências Matemáticas e de Computação, campus São Carlos.

Grande parte do estágio está sendo dedicado para análises estatísticas agregadas à pesquisa *survey* realizada pela PRIP por meio de um questionário enviado via e-mail para toda a comunidade USP, juntamente à outra estagiária Maria Victória Brandão Barros, e aos dois bolsistas PUB Lucas Roberto de Oliveira Lopes e João Gabriel Rego que trabalharam sobre a mesma supervisão. Isso posto, o escopo do relatório será detalhar tais análises em três seções:

1. Análises descritivas e exploratórias:
 - Questões psicométricas na escala Likert sobre como os estudantes de graduação se sentem em cada ambiente da universidade;
 - Questões de dados categóricas que demandaram gráficos de barras e tabelas de frequência relativa. Se referem às hostilidades sofridas pelos estudantes de graduação;
 - Questões de dados numéricos representadas por histogramas e *box-plots*/gráfico de caixas referentes à saúde mental dos estudantes, estes sendo analisados em grupos/categorias de diferentes gêneros, orientações e cor (Análises de associação).
2. Análises inferenciais:
 - Testes de normalidade dos dados (teste de Shapiro-Wilk) e análise de variância (ANOVA) não-paramétrico para comparação da distribuição de grupos de dados (teste de Kruskal-Wallis). Tais testes de hipótese foram aplicados sobre os dados numéricos das análises descritivas.
 - Modelos de regressão linear com as variáveis categóricas, estas referentes aos grupos mencionados no item anterior, sobre os mesmos dados de saúde mental, com o intuito de identificar variáveis preditoras/explicativas das notas atribuídas pelos estudantes.
3. Gráficos interativos, estes sendo uma adaptação dos que foram feitos durante as análises descritivas para suas publicações em web-sites e outras plataformas online, além de complementá-los com a filtragem dos dados para cada campus da USP.

Vale ressaltar que além da participação sobre tal pesquisa *survey*, realizei análises sobre dados referentes à Permanência e Formação Estudantil (PAPFE), além de auxiliar na organização de dados de 2022 do anuário estatístico da USP. Porém, como foram serviços muito específicos e prestados de maneira esporádica, decidi não colocá-los neste relatório.

Sumário

1	Introdução	4
2	Experiências pessoais dos estudantes	5
3	Notas atribuídas à saúde mental	10
3.1	Análise exploratória	11
3.2	Testes de hipótese	15
3.3	Modelagem e regressão linear sobre as notas	17
4	Gráficos interativos	21
5	Considerações finais	24
6	Referências	25

1 Introdução

Entre os dias 29 de agosto e 18 de setembro de 2022, ficou aberto para toda a comunidade USP o “Questionário PRIP: Inclusão e Pertencimento na USP”, o qual coletou respostas referentes às experiências pessoais, acadêmicas e percepções sobre os distintos ambientes institucionais da USP.

Dito isso, desde o recebimento dos dados por parte da Pró-Reitoria de Inclusão e Pertencimento (PRIP), foram feitas análises exploratórias e inferenciais sobre as respostas das perguntas para os seguintes grupos: alunos de graduação, alunos de pós-graduação, alunos de pós-doutorado, servidores docentes e servidores técnico-administrativos.

Como a maior parte das respostas do questionário foram de alunos de graduação, boa parte das análises neste relatório serão referentes a este grupo, mesmo que todas as visualizações tenham sido replicadas para os demais grupos. O relatório está dividido nos seguintes tópicos:

1. Experiências pessoais dos estudantes:

- Conforto em locais da USP medidos pela escala Likert;
- Análises descritivas sobre as perguntas que se referem à situações de hostilidade como discriminação, bullying e assédio que o(a) estudante sofreu.

2. Notas atribuídas à saúde mental:

- Análises descritivas sobre as notas atribuídas pelas categorias de identidade de gênero, orientação sexual e cor/raça;
- Testes de normalidade para as notas de cada categoria de identidade de gênero, orientação sexual e cor/raça;
- Teste de hipótese não-paramétrico Kruskal-Wallis entre as categorias de identidade de gênero, orientação sexual e cor/raça;
- Ajuste de modelos de regressão linear onde as variáveis preditoras são as categorias de identidade de gênero, orientação sexual e cor/raça, e a variável resposta é a nota;

Além disso, todos os resultados foram em cima de uma amostra estratificada por unidades da USP, onde foram selecionadas 3.000 respostas, de forma aleatória e sem reposição, de cerca de 5.300 respostas de alunos de graduação.

Vale ressaltar que toda a análise foi feita com a linguagem de programação *Python 3.0* na plataforma web *Google Colaboratory*. Foi utilizado também o software *Microsoft Excel 2016* para alguns pré-processamentos de dados. Para os gráficos interativos apresentados na última seção, foi utilizada a plataforma online *Flourish*, além da utilização das ferramentas mencionadas anteriormente para a manipular os dados e dispô-los no formato necessário que a plataforma aceita.

2 Experiências pessoais dos estudantes

A primeira análise abordada, apresentada na Figura 1, foi da pergunta do questionário “Quão confortável você se sente em cada um dos ambientes abaixo?”, que foi medida na escala Likert “Muito desconfortável”, “Desconfortável”, “Nem confortável nem desconfortável”, “Confortável” e “Muito confortável”. Também, era possível marcar a opção “Não sei/Não se aplica”, a qual foi retirada do gráfico e disposta na Tabela 1 pela sua frequência relativa ao total de respostas. Respostas diferentes desta última foram dispostas como “Outro”.

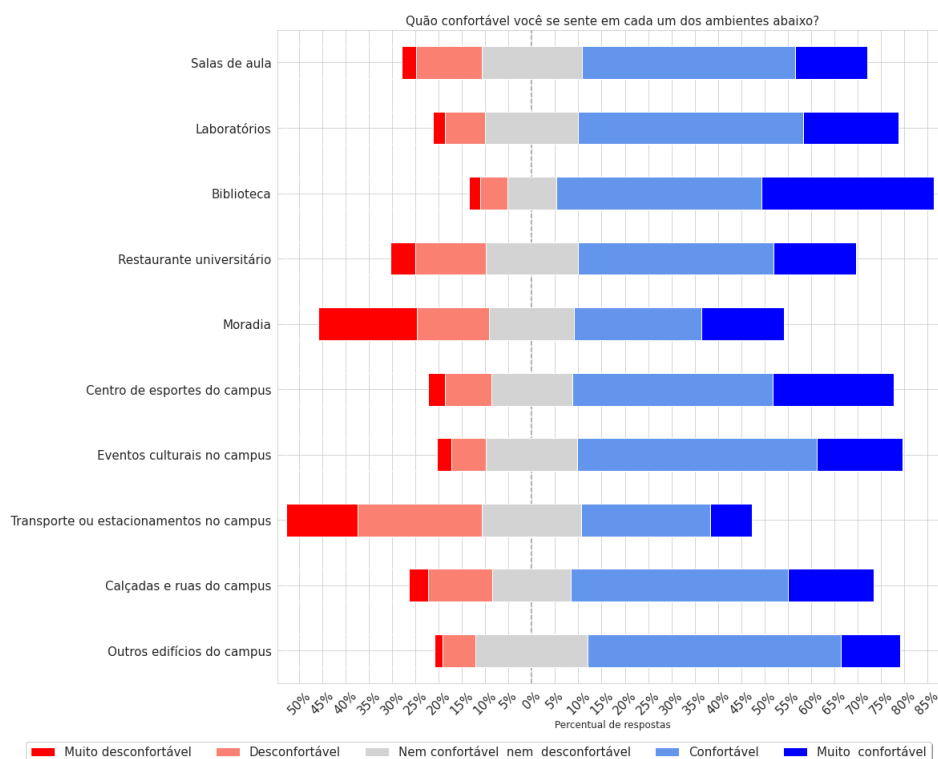


Figura 1: Respostas dos alunos de graduação com relação ao conforto em ambientes da USP.

Tabela 1: Tabela de frequência relativa das respostas.

Local	Não sei/Não se aplica	Outro
Salas de aula	0.17%	99.83%
Laboratórios	27.19%	72.81%
Biblioteca	9.43%	90.57%
Restaurante universitário	5.56%	94.44%
Moradia	76.61%	23.39%
Centro de esportes do campus	41.85%	58.15%
Eventos culturais no campus	34.96%	65.04%
Transporte ou estacionamentos no campus	12.46%	87.54%
Calçadas e ruas do campus	2.07%	97.93%
Outros edifícios do campus	16.23%	83.77%

É possível perceber, pela Figura 1 e pela Tabela 1, que na maioria dos ambientes, os estudantes se sentem confortáveis. Todavia, algumas respostas que muitos estudantes responderam “Não sei/Não se aplica”, como no ambiente “Moradia” que teve apenas 23.39% das respostas dispostas no gráfico de barras, possui uma quantidade significativa de respostas de “Muito desconfortável”, sendo aproximadamente 15% das respostas.

Percebe-se também uma quantidade considerável de estudantes que se sentem desconfortáveis em transportes ou estacionamentos do campus, onde a resposta “Muito desconfortável” representa aproximadamente 10% das respostas.

Os dois próximos gráficos se referem às perguntas de hostilidades sofridas para todas as categorias.

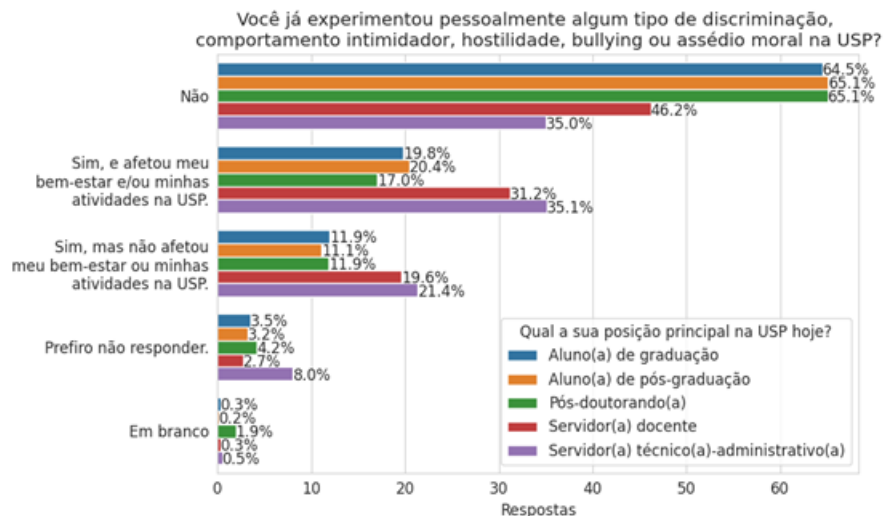


Figura 2: Ocorrências de hostilidades relatadas.

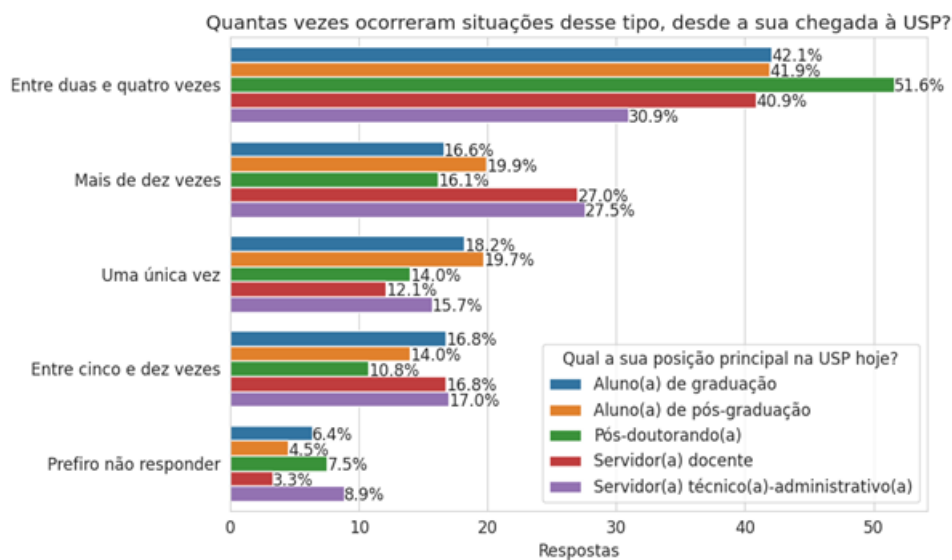


Figura 3: Quantidade de ocorrências de hostilidades sofridas

As perguntas seguintes se referem às respostas apenas dos alunos de graduação, apesar da análise de tais perguntas aplicadas às outras categorias terem sido feitas e contempladas na pesquisa.

Tais perguntas exigiram um tratamento rigoroso dos dados, uma vez que o formulário permitiu múltiplas respostas para ela, além de permitir o respondente marcar a opção “Outro” e escrever o que desejasse para complementar. Isso fez com que todas as opções marcadas fossem separadas por vírgula (“;”) dentro de uma célula só da base de dados, exigindo o desenvolvimento de um algoritmo para a contagem de cada opção marcada e para identificar respostas diferentes das demais.

Por esse motivo, as porcentagem excedem 100%, uma vez que um aluno pode escolher mais de uma opção, e a frequência de cada uma delas é dada pelo número de alunos que a marcaram em relação à quantidade de alunos que responderam a pergunta (Não deixaram a resposta em branco).

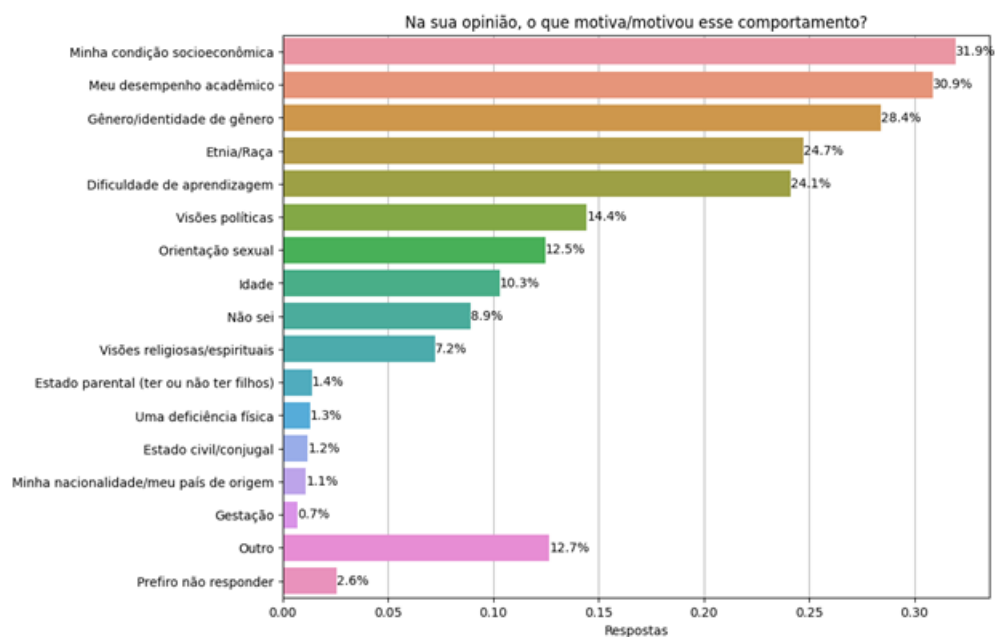


Figura 4: Motivos relatados da hostilidade sofrida.

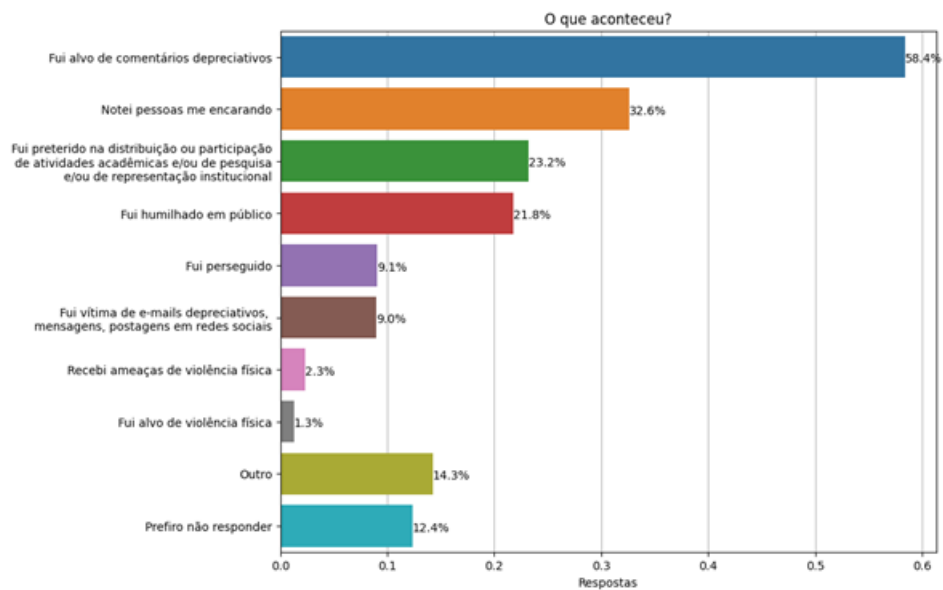


Figura 5: Detalhes sobre a hostilidade sofrida.

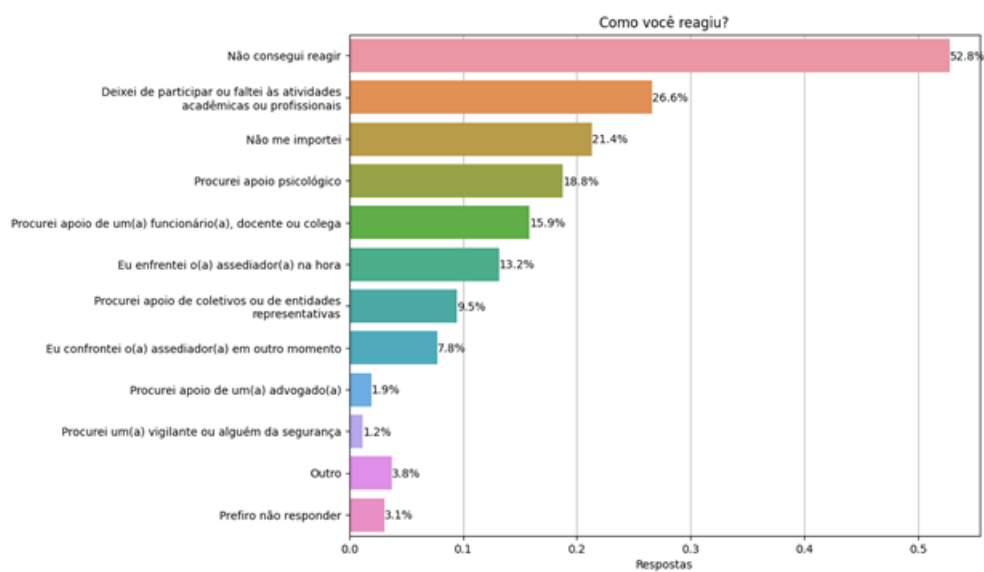


Figura 6: Reação da vítima sobre a(s) hostilidade(s) ocorrida(s).

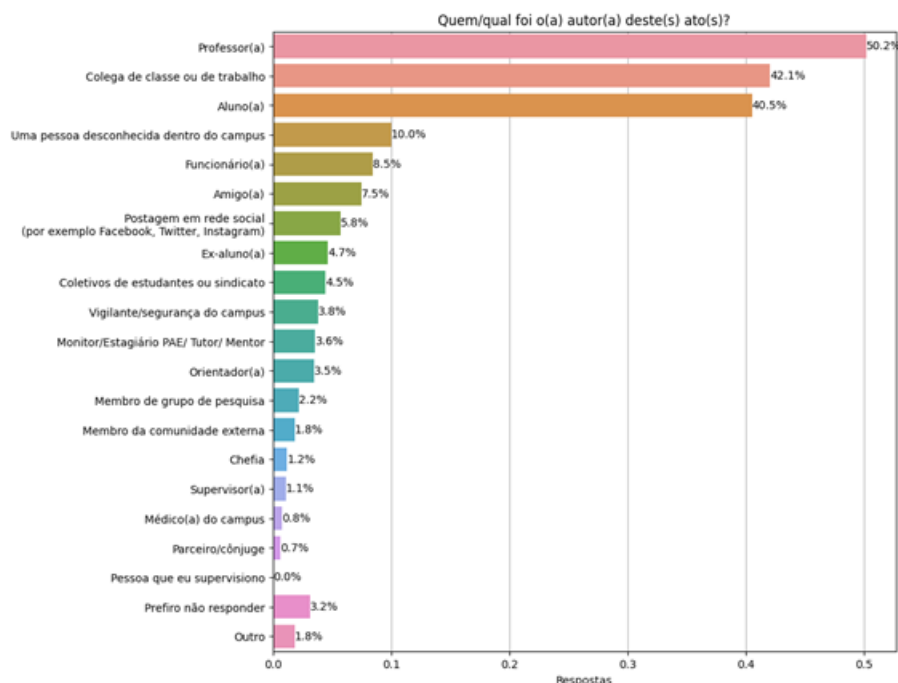


Figura 7: Autor(a) da(s) hostilidade(s)

Os gráficos dispostos nas Figuras 3, 4, 5, 6 e 7 tiveram suas respostas “Em branco” retiradas, uma vez que não eram interessantes para as análises.

As Figuras 2, 3, 4, 5, 6 e 7 fornecem informações importantes sobre as hostilidades sofridas pelos estudantes da USP. Com elas, é possível notar que a maioria das pessoas que responderam o questionário passaram por alguma situação de hostilidade entre duas a quatro vezes, e que os três motivos mais relatados para a realização das hostilidades foram a condição socioeconômica, o desempenho acadêmico e a identidade de gênero da pessoa.

Boa parte das pessoas que sofreram a(s) hostilidade(s) não conseguiram reagir, ou deixaram de participar de atividades acadêmicas ou profissionais. Nota-se também, que a maior parte das hostilidades foram comentários depreciativos. Os principais autores das hostilidades foram professores e colegas de classe ou de trabalho.

3 Notas atribuídas à saúde mental

As próximas análises se referem à pergunta “Numa escala de 0 a 10, como você avalia a sua saúde mental neste momento?”, cruzando com os dados das perguntas “Qual é sua identidade de gênero?”, “Em relação à orientação sexual, como você se identifica” e “Qual sua cor/raça (critério IBGE)”.

O objetivo desta parte do relatório, é mostrar o que foi feito para tentar

encontrar uma possível associação da nota atribuída à saúde mental, com as possíveis categorias que se referem à identidade de gênero, orientação sexual e cor/raça.

3.1 Análise exploratória

Para a visualização dos dados desta parte da análise, foram feitos os seguintes gráficos:

- ***Box plots e strip plots***: Os gráficos foram colocados juntos pois suas informações se complementam. O primeiro mostra a distribuição dos dados em “caixas”, permitindo a visualização de onde a mediana, o primeiro e terceiro quartil das notas se encontram. O segundo, apenas dispõe as respostas como pontos ao longo do eixo das abscissas.
- **Histograma**: Dispõe a distribuição das notas para cada categoria em porcentagem, dividindo as possíveis respostas em intervalos iguais. Foi disposta também a curva suave que se ajusta aos dados, denominada estimativa de densidade Kernel.

Identidade de gênero

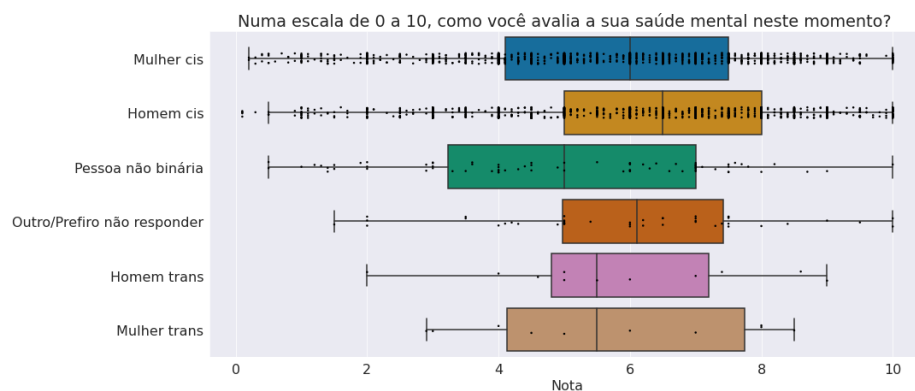


Figura 8: Box plots e strip plots das notas por identidade de gênero.

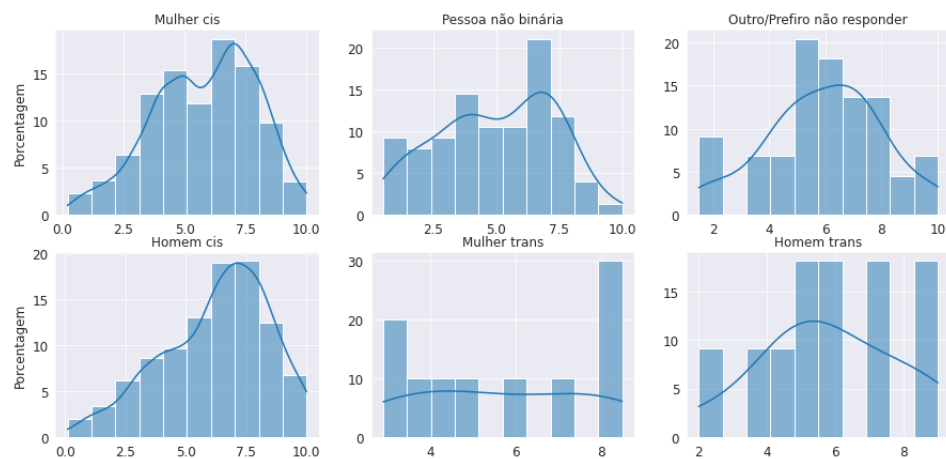


Figura 9: Histogramas das notas por identidade de gênero.

Os gráficos nas Figuras 7 e 8 indicam diferentes distribuições de notas entre as categorias de identidade de gênero.

As categorias “Mulher trans” e “Homem trans” tiveram poucas respostas para ser possível comparar com as demais.

Orientação sexual

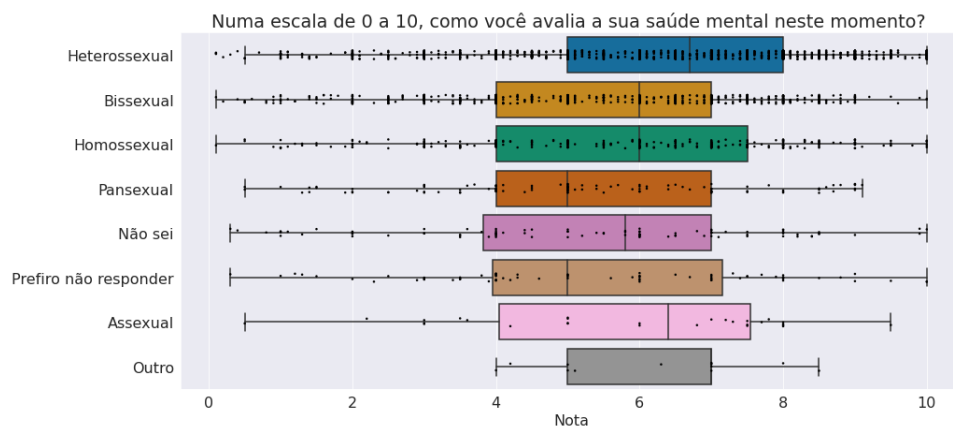


Figura 10: Box plots e strip plots das notas por orientação sexual.

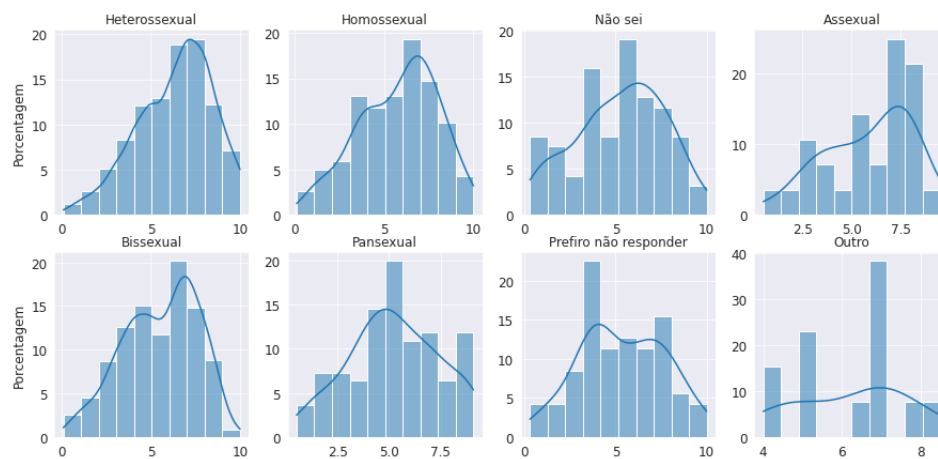


Figura 11: Histogramas das notas por orientação sexual.

Os gráficos nas Figuras 9 e 10 indicam diferentes distribuições de notas entre as categorias de orientação sexual. As distribuições que mais se assemelham são das categorias “Homossexual” e “Bissexual”.

A categoria “Outro” possuiu poucas respostas para ser possível comparar com as demais.

Cor/Raça

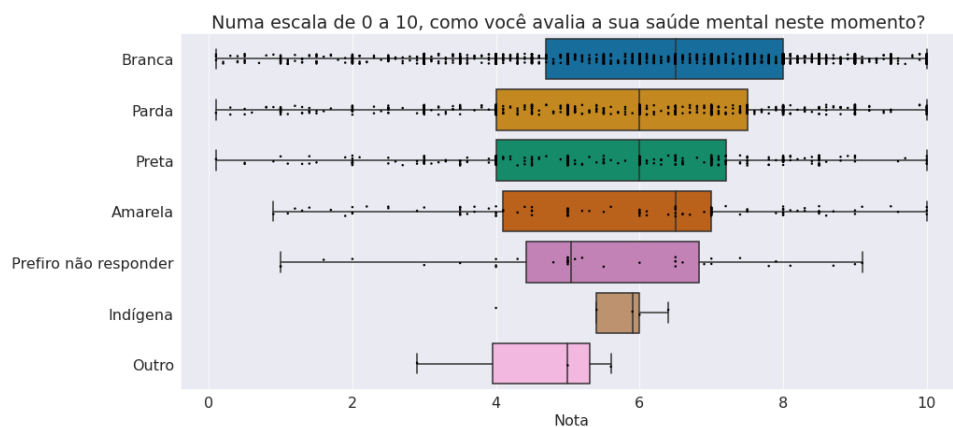


Figura 12: Box plots e strip plots das notas por cor/raça.

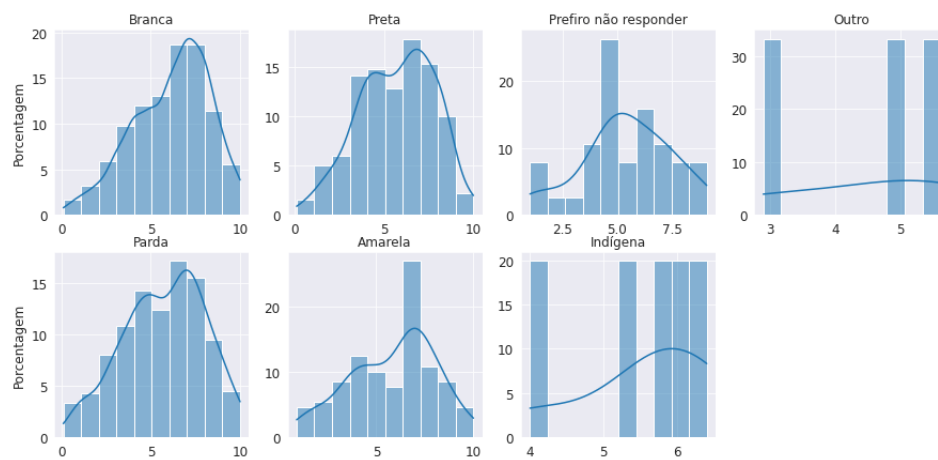


Figura 13: Histogramas das notas por cor/raça.

Os gráficos nas Figuras 11 e 12 indicam diferentes distribuições de notas entre as categorias de cor/raça. As distribuições que mais se assemelham são das categorias “Preta” e “Parda”.

As categorias “Indígena” e “Outro” tiveram poucas respostas para ser possível comparar com as demais.

3.2 Testes de hipótese

Uma vez que as análises descritivas foram realizadas e exibiram indícios de distribuições diferentes para as categorias, foram realizados os seguintes testes de hipótese:

- **Teste de Shapiro-Wilk:** Testa a normalidade dos dados. Ou seja:

$$\begin{cases} H_0 : \text{Os dados vêm de uma distribuição normal;} \\ H_1 : \text{Os dados não vêm de uma distribuição normal.} \end{cases}$$

Onde H_0 é a hipótese nula e H_1 a hipótese alternativa.

- **Teste H de Kruskal-Wallis:** Testa se dois ou mais conjuntos de dados originaram de uma mesma distribuição ou não. Trata-se de um teste não paramétrico análogo ao teste ANOVA (análise de variância) unidirecional em caso de não-normalidade dos dados. Tem-se então:

$$\begin{cases} H_0 : \text{Os conjuntos de dados originaram de uma mesma distribuição;} \\ H_1 : \text{Há pelo menos um conjunto de dados que foi originado de uma} \\ \text{distribuição diferente dos demais conjuntos.} \end{cases}$$

Tais testes foram realizados utilizando a biblioteca *scipy* para a linguagem *Python*. Neles, são retornados dois números: A estatística do teste e o p-valor. Este último é calculado com base na estatística e representa a probabilidade de se cometer um erro do tipo I, que significa rejeitar a hipótese nula quando ela é verdadeira.

As inferências foram feitas com a significância de $\alpha = 5\%$. Ou seja, rejeitou-se a hipótese nula sempre que o p-valor foi inferior à 5%.

Dito isso, os resultados dos testes seguem abaixo. Os valores dipostos nas colunas “Estatística” e “p-valor” foram truncados após 4 casas decimais.

Tabela 2: Resultados dos testes de Shapiro-Wilk das notas atribuídas à saúde mental por categoria de identidade de gênero.

Identidade	Estatística	p-valor
Mulher cis	0.9794	4.41e-14
Homem cis	0.9668	4.98e-16
Pessoa não binária	0.9610	0.0197
Mulher trans	0.9153	0.3195
Homem trans	0.9717	0.9038

Somente as categorias “Mulher trans” e “Homem trans” não tiveram a normalidade dos dados rejeitada entre as categorias de identidade de gênero, como a Tabela 2 mostra.

Tabela 3: Resultados dos testes de Shapiro-Wilk das notas atribuídas à saúde mental por categoria de orientação sexual.

Orientação	Estatística	p-valor
Homossexual	0.9762	6.17e-05
Heterossexual	0.9723	8.11e-17
Bissexual	0.9737	6.56e-10
Pansexual	0.9750	0.0366
Não sei	0.9701	0.0299
Assexual	0.9274	0.0530

Somente a categoria “Assexual” não teve a normalidade dos dados rejeitada entre as categorias de orientação sexual, como a Tabela 3 mostra.

Tabela 4: Resultados dos testes de Shapiro-Wilk das notas atribuídas à saúde mental por categoria de cor/raça.

Raça/Cor	Estatística	p-valor
Branca	0.9719	2.98e-18
Parda	0.9799	2.07e-07
Preta	0.9799	0.0001
Amarela	0.9673	0.0033
Indígena	0.8719	0.2744

Somente a categoria “Indígena” não teve a normalidade dos dados rejeitada entre as categorias de cor/raça, como a Tabela 4 mostra.

Uma vez feito os testes de normalidade, foram testadas para identidade de gênero, orientação sexual e cor/raça se suas categorias podiam ser consideradas de uma mesma distribuição. Os resultados seguem abaixo.

Tabela 5: Resultados dos testes de Kruskal-Wallis para as notas atribuídas à saúde mental.

Grupo	Estatística	p-valor
Identidade de gênero	44.6420	4.71e-09
Orientação sexual	101.0376	3.19e-20
Cor/Raça	26.0587	3.07e-05

Como é possível perceber pela Tabela 5, o teste de Kruskal-Wallis rejeitou a hipótese nula de que as categorias surgiram de uma mesma distribuição para todos os grupos (categorias de identidade de gênero, categorias de orientação sexual e categorias de cor/raça).

3.3 Modelagem e regressão linear sobre as notas

Uma vez que os testes de hipótese indicam que as categorias de cada grupo analisado (identidade de gênero, orientação sexual e cor/raça) influenciam na distribuição das notas, desejou-se criar um modelo que se ajustasse aos dados para explicar as notas atribuídas à saúde mental por meio das categorias possíveis.

Isso posto, os dados foram ajustados para cada grupo em modelos do tipo:

$$Y = \beta_0 + \sum_{i=1}^n \beta_i X_i + \epsilon$$

Onde β_j com $j \in \{0, \dots, n\}$ são os parâmetros do modelo, Y é a variável resposta que representa a nota atribuída a saúde mental, X_i são as variáveis indicadoras de cada categoria, ou seja:

$$X_i = \begin{cases} 1, & \text{se o dado for da categoria representada por } i \\ 0, & \text{caso contrário} \end{cases}$$

e ϵ é o erro aleatório/ruído.

O ajuste dos dados aos modelos para os três grupos foram feitos com a biblioteca *statsmodels* para a linguagem *Python*. Foi utilizado o método dos mínimos quadrados para a estimação dos parâmetros.

A categoria “Outro” não foi incluída como variável preditora para os grupos, uma vez que ela pode ser representada quando os coeficientes de todas as variáveis indicadoras valem zero. Isso foi feito para reduzir a multicolinearidade entre as variáveis.

A qualidade do ajuste dos dados aos modelos foi avaliada por três critérios:

- Por meio da estatística “Adj. R-squared”, a qual mede o quão bem a variância dos dados é explicada pelo modelo e vai de zero à um;
- Pelos p-valores dos coeficientes estimados (Representados por $P > |t|$ nas tabelas);
- Por meio de análise de resíduos.

Os resíduos são definidos da seguinte forma:

$$e_i = Y_i - \hat{Y}_i$$

onde Y_i é a i -ésima observação da amostra e \hat{Y}_i é a estimativa da observação pelo modelo ajustado.

Uma vez definidos, suas análises foram feitas por meio de *Q-Q plots*, gráficos que avaliam a disposição dos quantis empíricos ordenados com seus quantis teóricos.

Os parâmetros estimados encontram-se na coluna “coef”, onde “const” é a estimação do β_0 , o intercepto do modelo. Além disso, a variável resposta está sendo representada pelo nome “Nota saúde mental”.

A qualidade dos ajustes é discutida após todos os resultados dispostos.

Identidade de gênero

Tabela 6: Regressão linear sobre as notas de saúde mental para categorias de identidade de gênero.

Dep. Variable:	Nota saúde mental	R-squared:	0.015
Model:	OLS	Adj. R-squared:	0.013
Method:	Least Squares	F-statistic:	8.818
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	2.57e-08
Time:	17:43:50	Log-Likelihood:	-6339.7
No. Observations:	2898	AIC:	1.269e+04
Df Residuals:	2892	BIC:	1.273e+04
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	5.9409	0.325	18.252	0.000	5.303	6.579
Homem cis	0.3333	0.331	1.006	0.315	-0.316	0.983
Homem trans	-0.1136	0.728	-0.156	0.876	-1.541	1.313
Mulher cis	-0.0953	0.330	-0.289	0.773	-0.743	0.552
Mulher trans	-0.2509	0.756	-0.332	0.740	-1.734	1.232
Pessoa não binária	-0.9422	0.409	-2.304	0.021	-1.744	-0.140

Omnibus:	111.320	Durbin-Watson:	1.942
Prob(Omnibus):	0.000	Jarque-Bera (JB):	101.923
Skew:	-0.404	Prob(JB):	7.37e-23
Kurtosis:	2.564	Cond. No.	28.2

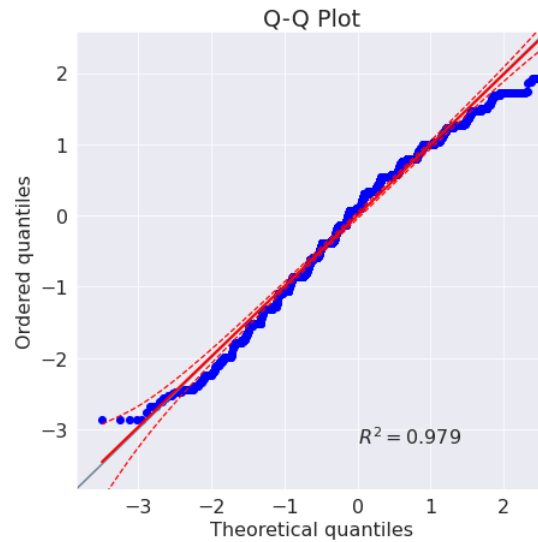


Figura 14: Envelope para os resíduos do modelo linear para categorias de identidade de gênero.

Orientação sexual

Tabela 7: Regressão linear sobre as notas de saúde mental para categorias de orientação.

Dep. Variable:	Nota saúde mental	R-squared:	0.038
Model:	OLS	Adj. R-squared:	0.035
Method:	Least Squares	F-statistic:	16.22
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	4.52e-21
Time:	17:43:57	Log-Likelihood:	-6305.7
No. Observations:	2898	AIC:	1.263e+04
Df Residuals:	2890	BIC:	1.268e+04
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	6.2385	0.592	10.537	0.000	5.078	7.399
Assexual	-0.4242	0.716	-0.592	0.554	-1.829	0.981
Bissexual	-0.6904	0.598	-1.155	0.248	-1.862	0.481
Heterossexual	0.1339	0.595	0.225	0.822	-1.032	1.300
Homossexual	-0.4312	0.605	-0.713	0.476	-1.617	0.754
Não sei	-0.9672	0.632	-1.531	0.126	-2.206	0.271
Pansexual	-1.0094	0.626	-1.612	0.107	-2.237	0.218
Prefiro não responder	-0.8596	0.644	-1.335	0.182	-2.122	0.403

Omnibus:	108.409	Durbin-Watson:	1.935
Prob(Omnibus):	0.000	Jarque-Bera (JB):	97.283
Skew:	-0.390	Prob(JB):	7.50e-22
Kurtosis:	2.556	Cond. No.	50.4

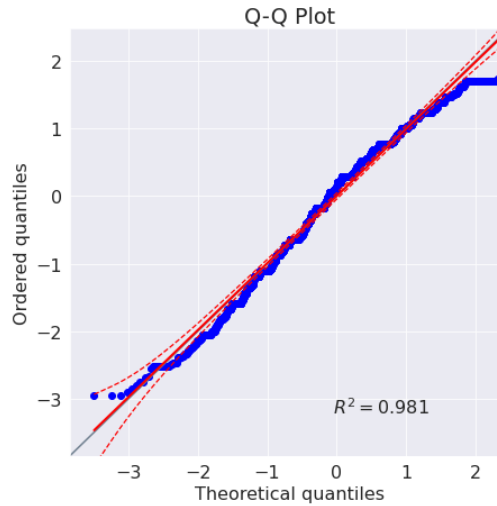


Figura 15: Envelope para os resíduos do modelo linear para categorias de orientação sexual.

Cor/Raça

Tabela 8: Regressão linear sobre as notas de saúde mental para categorias de cor/raça.

Dep. Variable:	Nota saúde mental	R-squared:	0.010
Model:	OLS	Adj. R-squared:	0.008
Method:	Least Squares	F-statistic:	4.934
Date:	Sat, 26 Nov 2022	Prob (F-statistic):	4.95e-05
Time:	17:44:00	Log-Likelihood:	-6346.8
No. Observations:	2898	AIC:	1.271e+04
Df Residuals:	2891	BIC:	1.275e+04
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	4.5000	1.250	3.600	0.000	2.049	6.951
Amarela	1.3434	1.264	1.063	0.288	-1.136	3.822
Branca	1.6686	1.251	1.334	0.182	-0.784	4.121
Indígena	1.0400	1.581	0.658	0.511	-2.060	4.140
Parda	1.2486	1.253	0.997	0.319	-1.208	3.705
Prefiro não responder	0.9342	1.298	0.720	0.472	-1.611	3.480
Preta	1.2135	1.256	0.966	0.334	-1.249	3.676

Omnibus:	109.326	Durbin-Watson:	1.941
Prob(Omnibus):	0.000	Jarque-Bera (JB):	98.007
Skew:	-0.391	Prob(JB):	5.22e-22
Kurtosis:	2.554	Cond. No.	101.

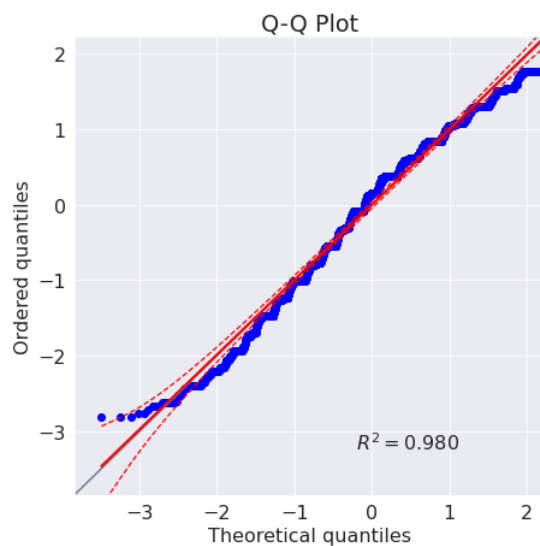


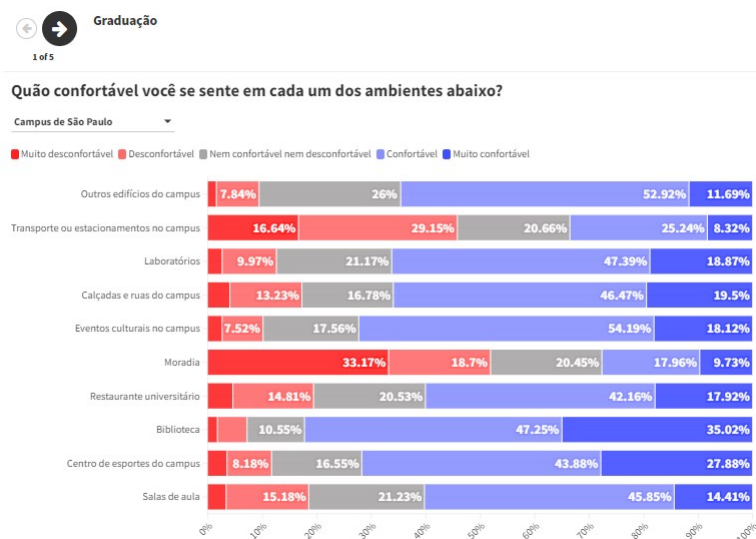
Figura 16: Envelope para os resíduos do modelo linear para categorias de cor/raça.

Com os resultados dispostos nas Tabelas 6, 7 e 8, e com as Figuras 14, 15 e 16, é possível perceber que, no geral, os modelos não se ajustaram bem aos dados, uma vez que as estatísticas “Adj. R-squared” foram muito baixas. Além disso, os p-valores das estimativas dos coeficientes foram muito altos, o que significa que a hipótese nula de que eles sejam zero e, conseqüentemente, não possuam influência sobre a variável resposta, não é rejeitada. Por fim, nota-se que os resíduos possuem caudas e não se aproximam de uma reta, exibindo uma distribuição muito distinta com relação à distribuição dos dados do modelo.

4 Gráficos interativos

A seguir, serão dispostas adaptações dos gráficos que se referem às experiências pessoais que foram apresentados na primeira seção do relatório. Tais gráficos foram refeitos na plataforma *Flourish*, como mencionado na introdução, além das estatísticas terem sido reproduzidas para cada campus da USP.

Abaixo, seguem capturas de telas dos gráficos feitos.



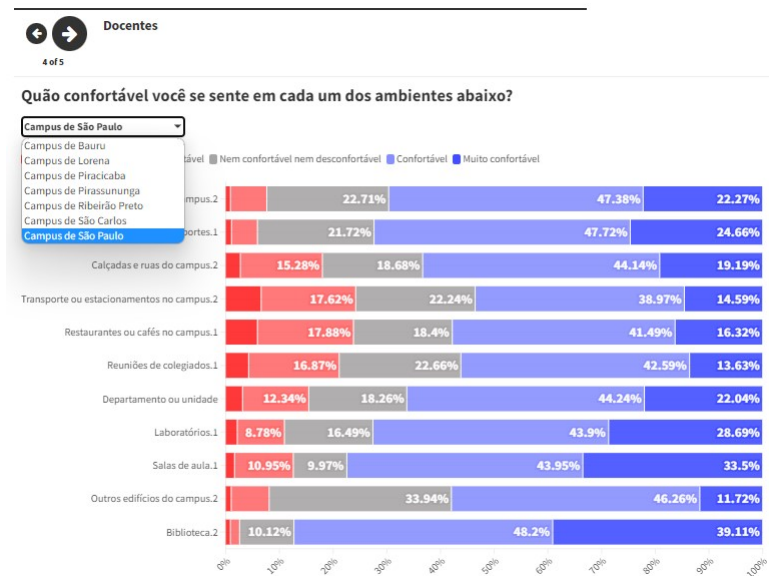
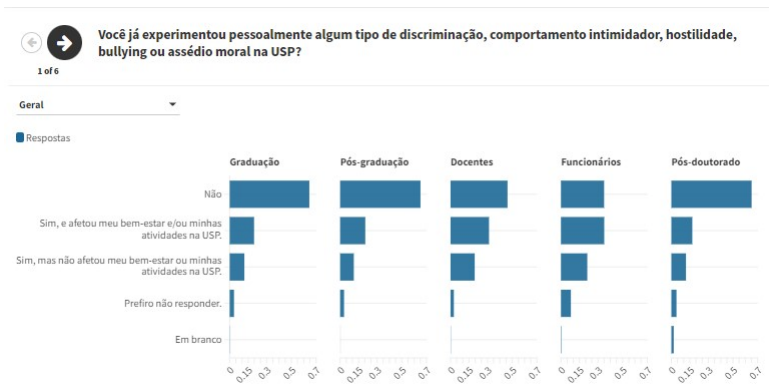


Figura 17: Capturas de tela do gráfico interativo na escala Likert



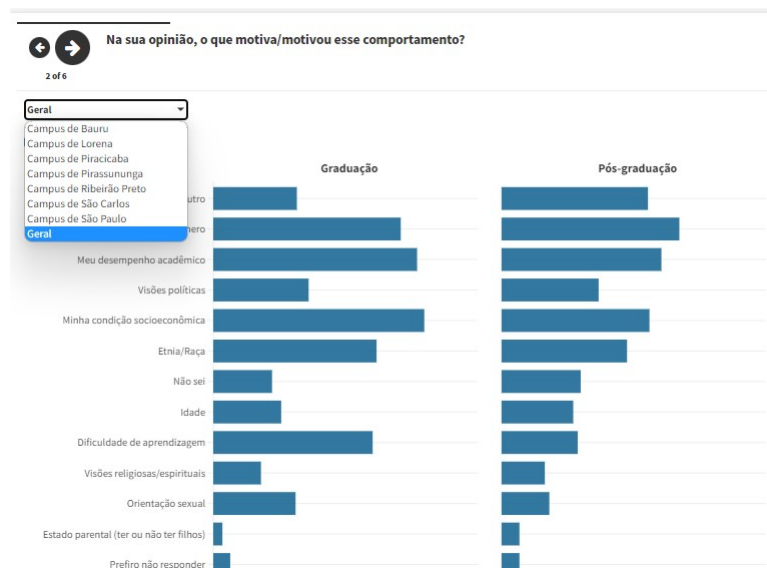


Figura 18: Capturas de tela dos gráficos interativos sobre hostilidades

Mais detalhes sobre os gráficos e seus elementos interativos estão disponíveis nos links:

- <https://public.flourish.studio/story/1959906/>
- <https://public.flourish.studio/story/1972148/>

5 Considerações finais

As análises abordadas neste relatório são apenas parte de um estudo em andamento realizado pela Pró-Reitoria de Inclusão e Pertencimento. Todavia, tais estatísticas trazem informações importantes que se referem ao sentimento de inclusão dos estudantes em ambientes da USP e sobre detalhes de hostilidades praticadas e sofridas pelas extensões dos campus. Tratam-se de novos conhecimentos que devem ser levados em conta durante tomadas de decisões para o combate por parte da USP ao desrespeito em seus ambientes institucionais.

Com relação às notas atribuídas à saúde mental, mesmo que não tenha sido possível fazer modelagens que se adéquem bem aos dados de modo que quantificasse a influência das categorias de identidade de gênero, orientação sexual e cor/raça sobre as notas, é necessário ressaltar que esta foi a etapa mais recente do estudo, e os dados foram ajustados sob um dos modelos de regressão mais simples que existem, que foram múltiplos hiperplanos. As próximas etapas da modelagem de dados será analisar outros modelos que talvez possam se enquadrar melhor aos dados. Algumas das possíveis opções são:

- Adicionar mais variáveis preditoras aos modelos;
- Utilizar modelos polinomiais;
- Utilizar métodos onde valores discrepantes (*outliers*) não impactem muito as estimações;
- Utilizar modelos lineares generalizados (MGL);

6 Referências

- BUSSAB, W. O.; MORETTIN, P. A. Estatística Básica. Atual Editora, São Paulo, 1988.
- SHAPIRO, S. S.; WILK, M.B An analysis of variance test for normality (complete samples), Biometrika, Vol. 52, pp. 591-611, 1965.
- KRUSKAL, W.H.; W. W. WALLIS, W.W. Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association, Vol. 47, Issue 260, pp. 583-621, 1952.
- MONTGOMERY, D.C.; PECH, E.A; VINING, G.G. Introduction to Linear Regression Analysis. New York. John Wiley, 2012.