

# Percepção de Profissionais de TI sobre Documentação de Testes com e sem Inteligência Artificial

Rodolpho Dutra (25109P9M24)

Fernando Bento Moura de Souza (25109P9M13),

Luísa Caetano Correia (25109P9M19) e

Eduardo Sampaio (Matrícula: 25109P9M08)



UNIVERSIDADE FEDERAL DO  
ESTADO DO RIO DE JANEIRO

# Proposta

**Título:** Percepções de Profissionais de Tecnologia sobre a Qualidade da Documentação de Testes Gerada com e sem Uso de Inteligência Artificial

**Objetivo:** Investigar se há diferença significativa nas percepções de profissionais de tecnologia quanto à qualidade da documentação de testes gerada por ferramentas de inteligência artificial em comparação com a documentação feita manualmente.

# Hipótese de Pesquisa

**$H_1$  (Hipótese Alternativa):** Há diferença significativa na percepção de qualidade entre documentações de testes geradas por inteligência artificial e as elaboradas manualmente.

**$H_0$  (Hipótese Nula):** Não há diferença significativa na percepção de qualidade entre documentações de testes geradas por inteligência artificial e as elaboradas manualmente.

# Metodologia

**Tipo de pesquisa:** Quantitativa, descritiva com teste de hipótese

**Instrumento:** Survey online (Google Forms)

**Amostra:** 137 profissionais da área de tecnologia (QA, desenvolvedores, analistas, etc.)

**Variáveis:**

- Independentes: Tipo de documentação de testes (gerada com IA x gerada manualmente)
- Dependentes: Escore médio de qualidade percebida

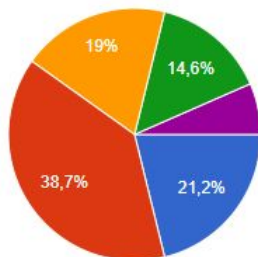
# Análise Estatística

- Cálculo de Média e Desvio Padrão
- Teste de normalidade Kolmogorov-Smirnov (KS) (amostra maior que 50,  $n > 50$ )
- Teste de homogeneidade de variância: Levene
- Comparação de médias: Teste t de Student (pois temos dois grupos)
- Nível de significância:  $\alpha = 0,05$
- Softwares Utilizados: Python e Excel

# Questionário - Perfil

Qual é sua área de atuação principal?

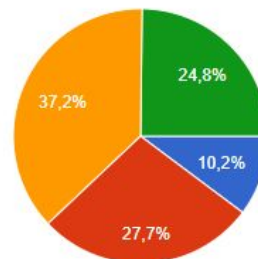
137 respostas



- QA/Testes
- Desenvolvimento de Software
- Análise de Sistemas
- Gestão de Projetos
- Outra

Qual é o seu nível de experiência na área de tecnologia?

137 respostas



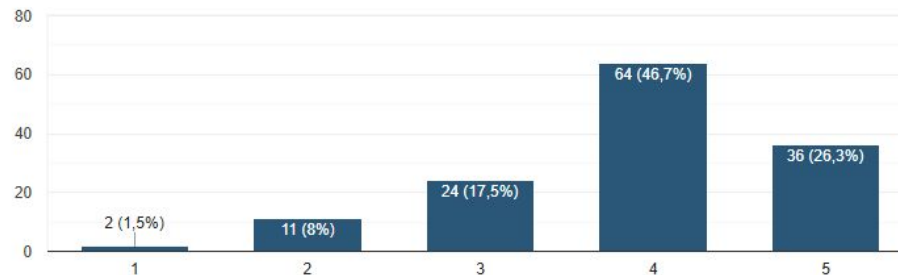
- Estagiário/Trainee
- Júnior
- Pleno
- Sênior

# Questionário - Escala de Percepção - Clareza das Informações

Documentação gerada manualmente

 Copiar gráfico

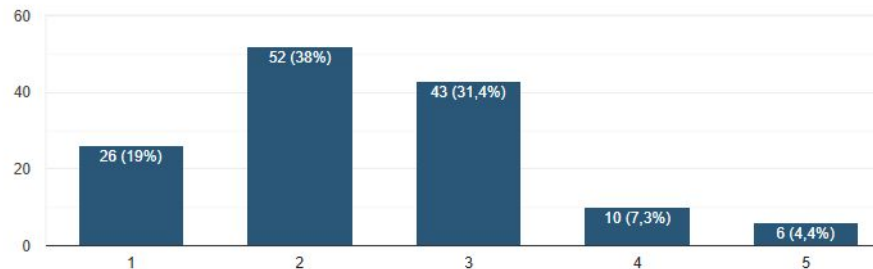
137 respostas



Documentação gerada por IA

 Copiar gráfico

137 respostas

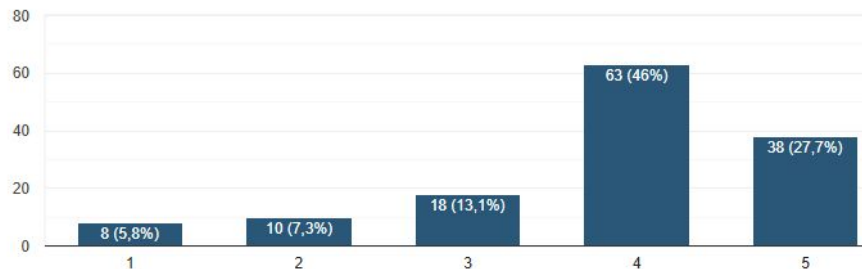


# Questionário - Escala de Percepção - Organização da Estrutura

Documentação gerada manualmente

 Copiar gráfico

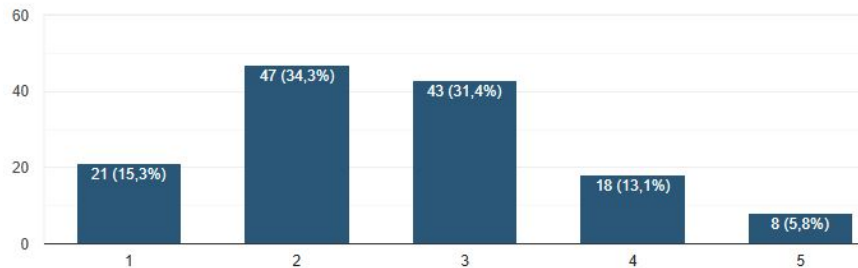
137 respostas



Documentação gerada por IA

 Copiar gráfico

137 respostas



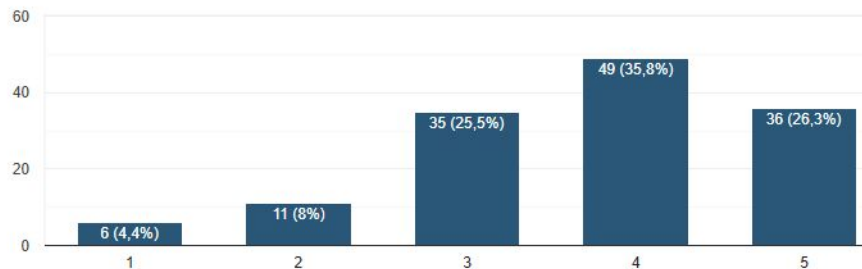


# Questionário - Escala de Percepção - Cobertura dos Cenários de Teste

Documentação gerada manualmente

 Copiar gráfico

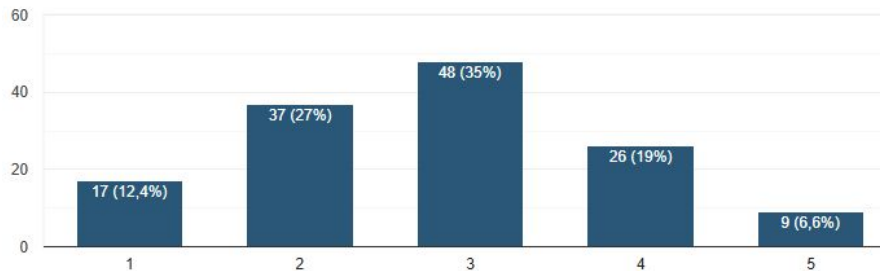
137 respostas



Documentação gerada por IA

 Copiar gráfico

137 respostas

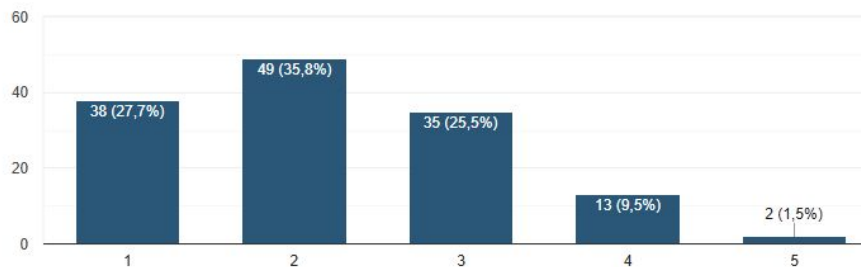


# Questionário - Escala de Percepção - Velocidade de Geração

Documentação gerada manualmente

 Copiar gráfico

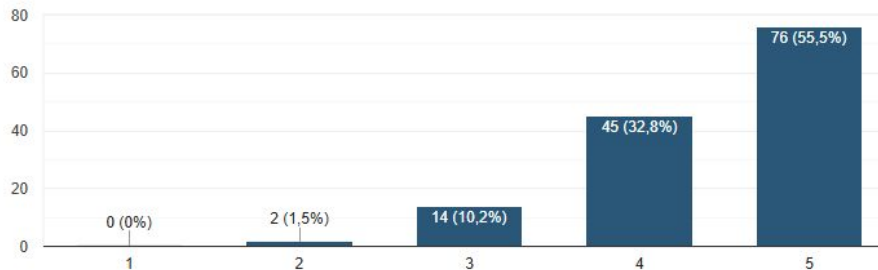
137 respostas



Documentação gerada por IA

 Copiar gráfico

137 respostas

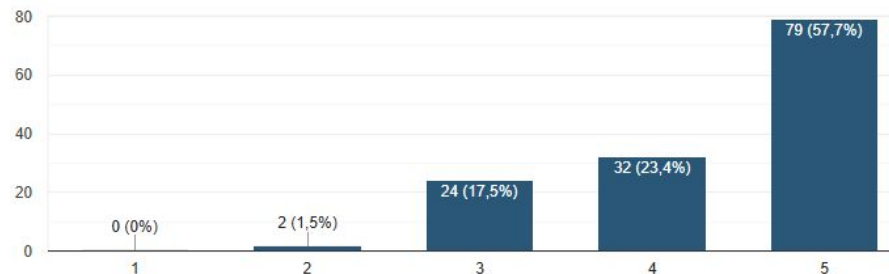


# Questionário - Escala de Percepção - Confiabilidade das Informações

Documentação gerada manualmente

 Copiar gráfico

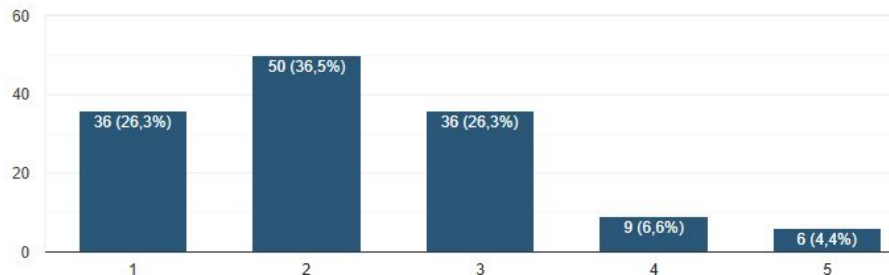
137 respostas



Documentação gerada por IA

 Copiar gráfico

137 respostas



# Análise dos Dados

## 1. Dados divididos em 2 grupos (IA x Manual)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Documentação de Testes Gerada com IA										Documentação de Testes Gerada Manualmente									
Qualidade de Atualização	Experiência do Testador	Clareza	Organização	Cobertura	Velocidade	Confiabilidade	Média	Desvio	Quadrado do Desvio	Qualidade de Atualização	Experiência do Testador	Clareza	Organização	Cobertura	Velocidade	Confiabilidade	Média	Desvio	Quadrado do Desvio
Atualização de Sistema	Júnior	4	4	4	5	4	4,2	1,302189781	1,69569822574083	Atualização de Sistema	Júnior	4	4	4	3	5	4	0,39854	0,158834247973702
Atualização de Sistema	Sênior	3	3	1	3	3	2,6	0,297810219	0,088690926540828	Atualização de Sistema	Sênior	1	2	4	3	4	2,8	0,80146	0,642337897573701
Análise de Sistema	Júnior	2	2	2	4	3	2,6	0,297810219	0,088690926540828	Análise de Sistema	Júnior	4	3	3	1	5	3,2	0,40146	0,161170014373701
Atualização de Sistema	Júnior	5	3	3	4	2	3,4	0,502189781	0,252194576140828	Atualização de Sistema	Júnior	3	1	5	2	5	3,2	0,40146	0,161170014373701

## 2. Cálculo de Média, Mediana, Variância e Desvio Padrão

Qualidade Percebida das Documentações					
Grupo	Média	Mediana	Variância	Desvio Padrão	n (Tamanho da Amostra)
IA	2,897810219	2,8	0,2301411903	0,47973033081	137
Manual	3,601459854	3,6	0,2040854601	0,45175818764	137

# Teste de normalidade Kolmogorov-Smirnov (KS)

```
# Lista de médias de qualidade para geradas por IA
> usa_ia = [4.2, 2.6, 2.6, 3.4, 2.2, 3.0, 3.2, 3.6, 3.2, 2.2, 2.8, 2.4, 2.8, 3.0, 2.6, 2.8, 3.0, 3.2, 2.8, 3.0, ...
# Lista de médias de qualidade para geradas Manualmente
> nao_ia = [ 4, 2.8, 3.2, 3.2, 4.2, 3.6, 3.4, 3.8, 3.2, 3.4, 3.2, 3, 4.4, 4, 3.6, ...

# -----
# TESTE DE NORMALIDADE - Kolmogorov-Smirnov
# -----

# Normalizar os dados
z_ia = zscore(usa_ia)
z_nao_ia = zscore(nao_ia)

print("\nTeste de normalidade - Kolmogorov-Smirnov:")
ks_ia = kstest(z_ia, 'norm')
ks_nao_ia = kstest(z_nao_ia, 'norm')

print(f"Usa IA:      KS = {ks_ia.statistic:.4f}, p = {ks_ia.pvalue:.4f}")
print(f"Não Usa IA:  KS = {ks_nao_ia.statistic:.4f}, p = {ks_nao_ia.pvalue:.4f}")
```

```
Teste de normalidade - Kolmogorov-Smirnov:
Usa IA:      KS = 0.1111, p = 0.0629
Não Usa IA:  KS = 0.1005, p = 0.1172
PS C:\Users\ferna\OneDrive\Desktop\Survey\scripts python>
```

$P > 0.05$  para ambos grupos, logo os dados seguem distribuição normalizada e podemos aplicar o teste t.

# Teste de homogeneidade de variância: Levene

levne.py > ...

```
1  from scipy.stats import levene
2
3  # Se  $p > 0.05$ , não rejeitamos a hipótese nula → variâncias são iguais
4  # Se  $p \leq 0.05$ , rejeitamos a hipótese nula → variâncias são diferentes
5
6  > usa_ia = [ 4.2, 2.6, 2.6, 3.4, 2.2, 3.0, 3.2, 3.6, 3.2, 2.2, 2.8, 2.4, 2.8, 3.0, 2.6, 2.8, 3.0, 3.2,
14 # Lista de médias de qualidade para geradas por IA
15 > nao_ia = [4, 2.8, 3.2, 3.2, 4.2, 3.6, 3.4, 3.4, 3.8, 3.2, 3.4, 3.2, 3.4, 3.2, 3, 4.4, 4, 3.6, ...
24
25  stat, p = levene(usa_ia, nao_ia)
26  print("Estatística de Levene:", stat)
27  print("Valor-p:", p)
28
29
```

Estatística de Levene: 0.027832493412806577

Valor-p: 0.8676271904898587

PS C:\Users\ferna\OneDrive\Desktop\Survey\scripts python> █

$P > 0.05$  sem diferença significativa entre as variâncias dos grupos, logo usa True como parâmetro no método do teste t.

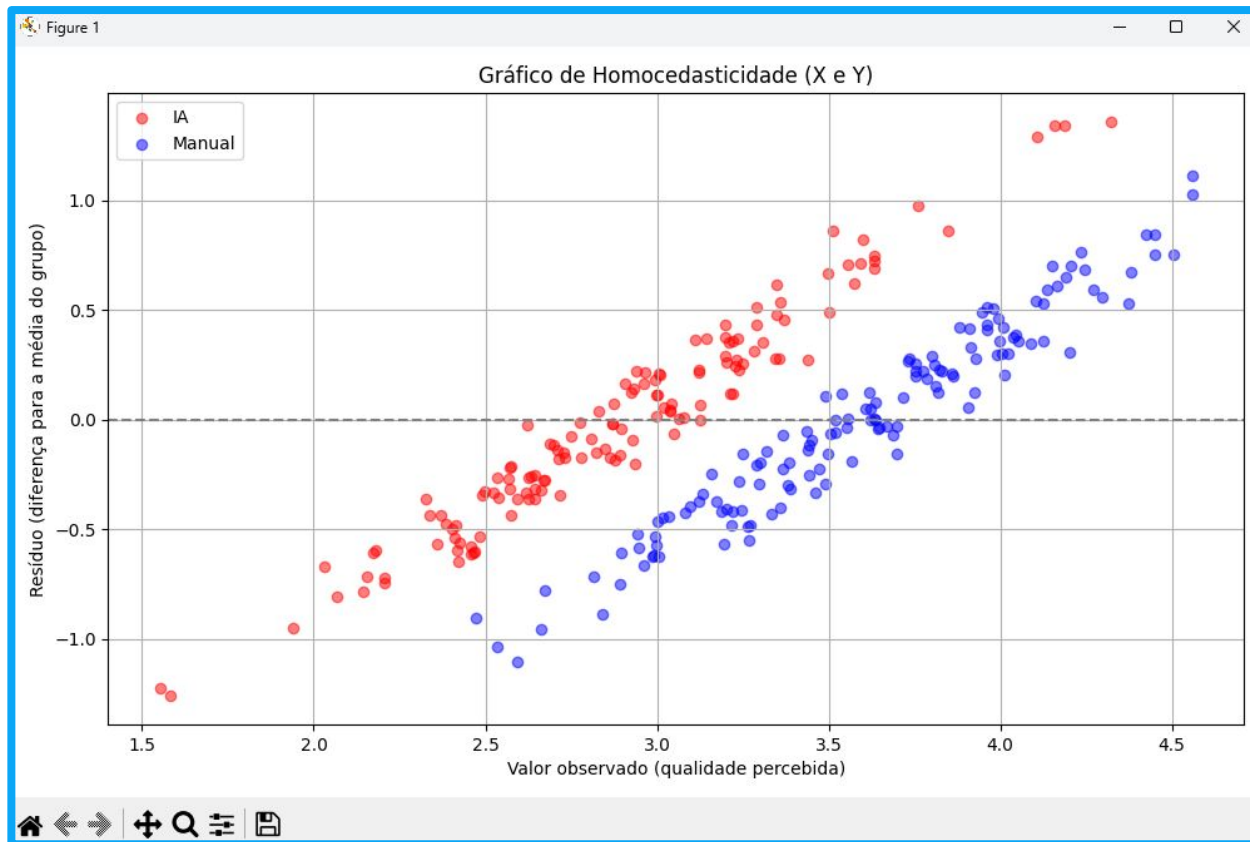
# Teste t de Student

```
t-test.py > ...
1 from scipy.stats import ttest_ind_from_stats
2
3 # Dados
4 media_nao_ia = 3.601459854
5 dp_nao_ia = 0.4517581876
6 n_nao_ia = 137
7
8 media_ia = 2.897810219
9 dp_ia = 0.4797303308
10 n_ia = 137
11
12 # Teste t
13 t_stat, p_value = ttest_ind_from_stats(
14     mean1=media_nao_ia, std1=dp_nao_ia, nobs1=n_nao_ia,
15     mean2=media_ia, std2=dp_ia, nobs2=n_ia,
16     equal_var = True # False se as variâncias forem diferentes e True se forem iguais
17 )
18
19 print("Estatística t:", round(t_stat, 4))
20 print("Valor-p:", round(p_value, 4))
```

```
PS C:\Users\ferna\OneDrive\Desktop\Survey\scripts python> python -
Estatística t: 12.4985
Valor-p: 0.0
PS C:\Users\ferna\OneDrive\Desktop\Survey\scripts python> []
```

$P < 0.05$  diferença significativa, rejeita a hipótese nula. Existe uma diferença estatisticamente significativa entre a qualidade percebida da documentação gerada por IA e da gerada manualmente.

# Gráfico de Homocedasticidade





# Resultados

Os profissionais avaliaram as documentações de testes geradas por inteligência artificial com uma média de 2,89 pontos na escala de qualidade percebida, enquanto as documentações geradas manualmente receberam uma média de 3,60 pontos. O teste t de Student indicou uma diferença estatisticamente significativa entre os dois grupos ( $t = 12,49$ ,  $p = 0,00$ ), com variâncias homogêneas ( $p = 0,86$  no teste de Levene) e distribuição normal em ambas as amostras (KS-IA:  $p = 0,06$ ; KS-Manual:  $p = 0,11$ )."

# Conclusão

Os dados rejeitam a hipótese nula e fornecem evidências estatísticas de que há uma diferença significativa na qualidade percebida entre documentações de testes geradas com o uso de inteligência artificial e as produzidas manualmente. Especificamente, os profissionais atribuíram uma percepção de qualidade ligeiramente maior às documentações manuais, o que pode indicar uma preferência ou maior confiabilidade percebida nas práticas tradicionais."

Obrigado.



UNIVERSIDADE FEDERAL DO  
ESTADO DO RIO DE JANEIRO