

Caracterizando a atividade de code review no github

Laboratório de medição e experimentação

Laboratório 03

- **Grupo 01:**
 - Nataniel Geraldo Mendes Peixoto
 - Nelson de Campos Nolasco
 - Rubia Coelho de Matos



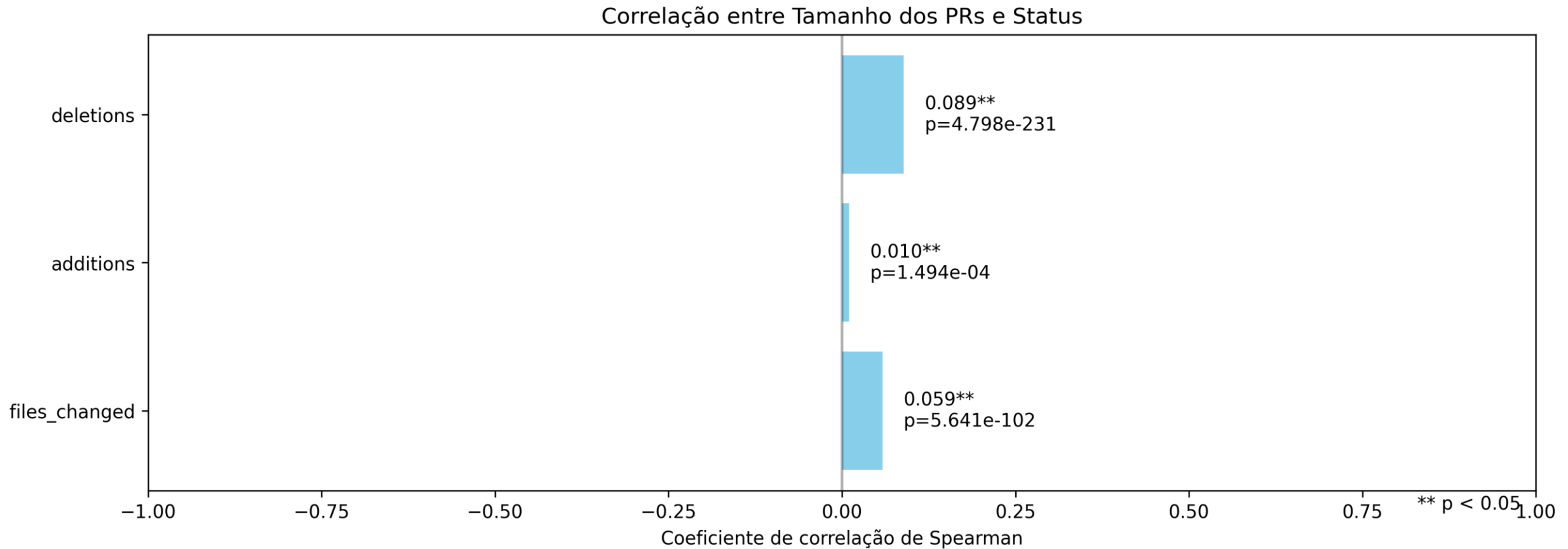
Interpretação dos resultados

Critérios coleta de dados

Critérios para seleção de PRs submetidos a repositórios:

- 200 repositórios mais populares do GitHub;
- que possuam pelos menos 100 PRs;
- com status MERGED ou CLOSED;
- com pelo menos uma revisão;
- cuja revisão levou pelo menos uma hora.

RQ01 - Relação entre o tamanho dos PRs e o feedback final das revisões



Número de arquivos alterados x status

- Coeficiente de correlação: 0.0588
- P-valor: 5.6406e-102
- Interpretação: Insignificante
- Estatisticamente significativo: Sim

Linhas adicionadas x status

- Coeficiente de correlação: 0.0104
- P-valor: 1.4940e-04
- Interpretação: Insignificante
- Estatisticamente significativo: Sim

Linhas removidas x status

- Coeficiente de correlação: 0.0889
- P-valor: 4.7984e-231
- Interpretação: Insignificante
- Estatisticamente significativo: Sim

RQ01 - Relação entre o tamanho dos PRs e o feedback final das revisões

Estatísticas descritivas (medianas):

- PRs mesclados (MERGED):

- Arquivos alterados: 2.00
- Linhas adicionadas: 18.00
- Linhas removidas: 5.00

- PRs fechados sem merge (CLOSED):

- Arquivos alterados: 2.00
- Linhas adicionadas: 18.00
- Linhas removidas: 2.00

Estatísticas descritivas (média):

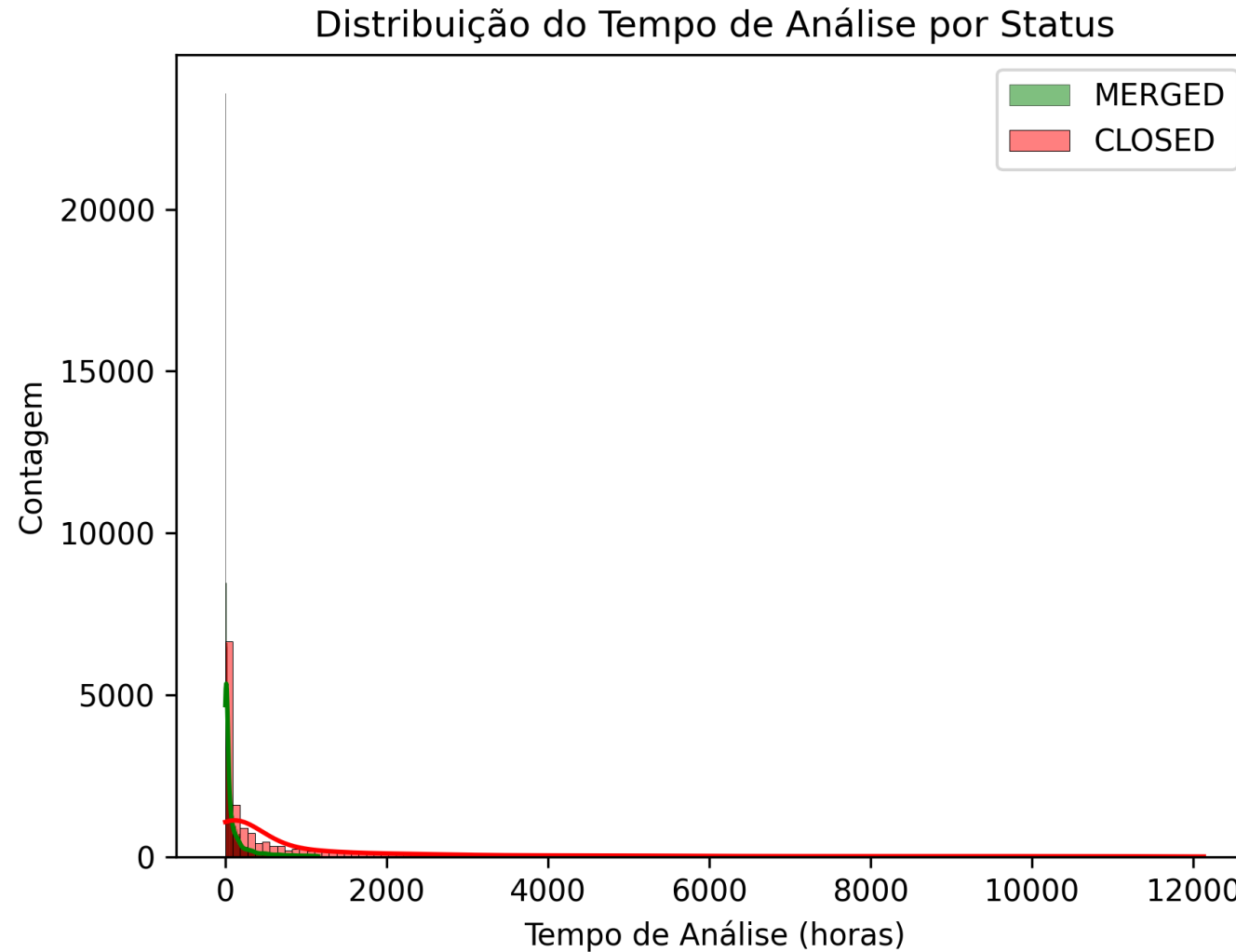
- PRs mesclados (MERGED):

- Arquivos alterados: 8.95
- Linhas adicionadas: 287.06
- Linhas removidas: 5.00

- PRs fechados sem merge (CLOSED):

- Arquivos alterados: 19.88
- Linhas adicionadas: 814.51
- Linhas removidas: 930.39

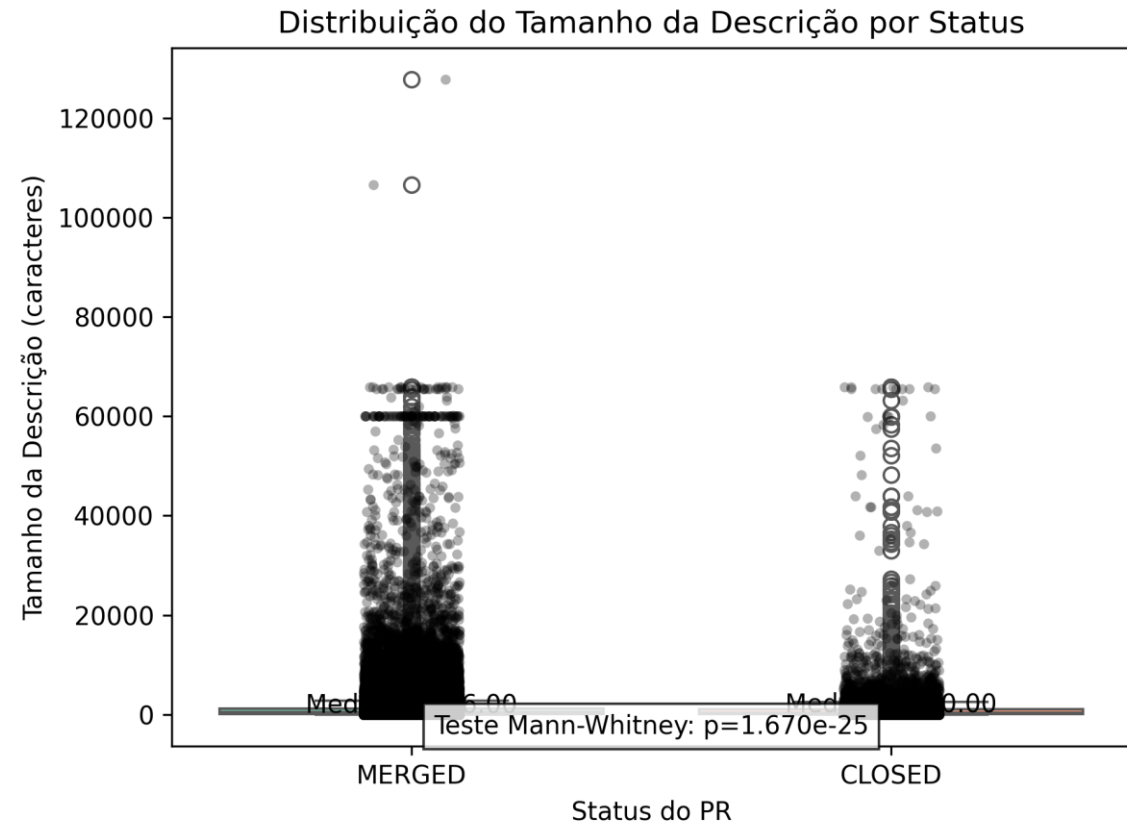
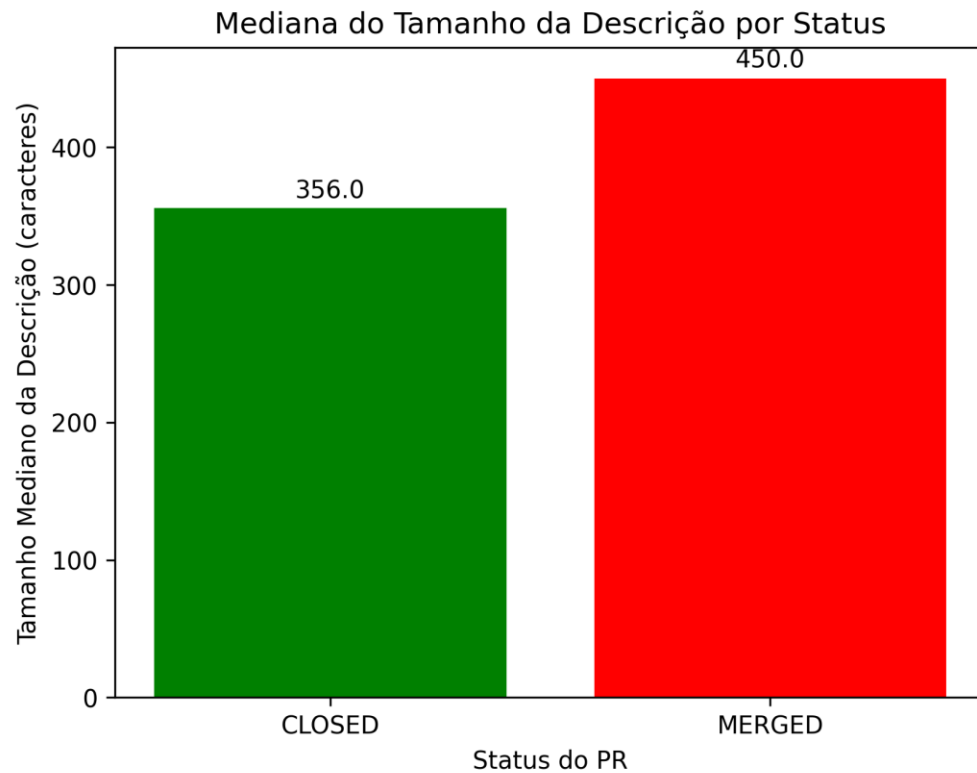
RQ02 - Relação entre o tempo de análise dos PRs e o feedback final das revisões



- Coeficiente de correlação: -0.2376
- P-valor: 0.0000e+00
- Interpretação: Fraca
- Estatisticamente significativo: Sim

- PRs mesclados (MERGED): 30.54 horas
- PRs fechados sem merge (CLOSED): 226.54 horas

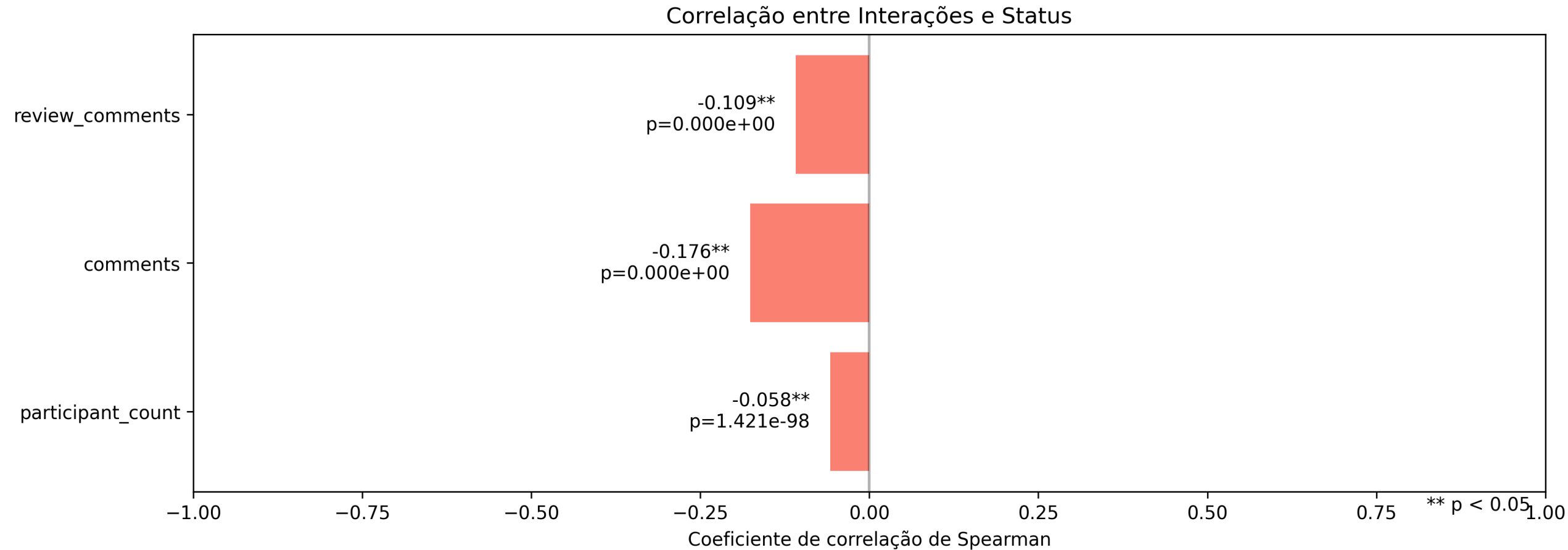
RQ03 - Relação entre a descrição dos PRs e o feedback final das revisões



- Coeficiente de correlação: 0.0286
- P-valor: $1.6337e-25$
- Interpretação: Insignificante
- Estatisticamente significativo: Sim

- PRs mesclados (MERGED): 450.00 caracteres
- PRs fechados sem merge (CLOSED): 356.00 caracteres

RQ04 - Relação entre as interações nos PRs e o feedback final das revisões



Número participantes x status

- Coeficiente de correlação: -0.0578
- P-valor: 1.4209e-98
- Interpretação: Insignificante
- Estatisticamente significativo: Sim

Número comentários x status

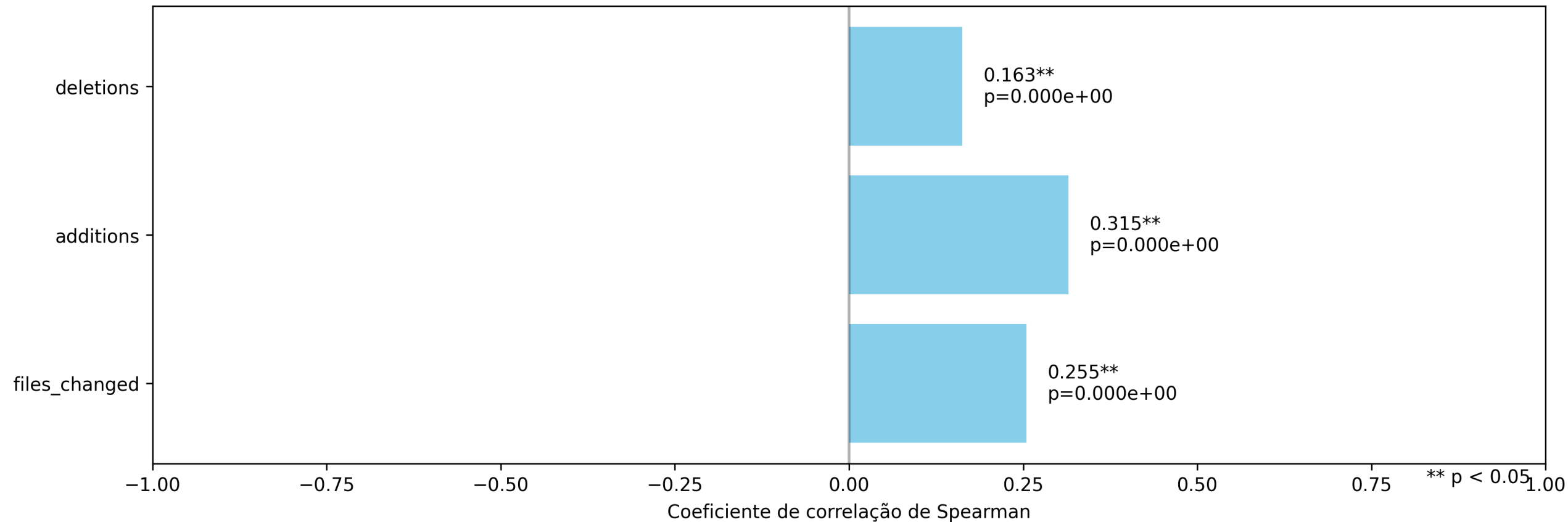
- Coeficiente de correlação: -0.1762
- P-valor: 0.0000e+00
- Interpretação: Fraca
- Estatisticamente significativo: Sim

Número comentários revisão x status

- Coeficiente de correlação: -0.1090
- P-valor: 0.0000e+00
- Interpretação: Fraca
- Estatisticamente significativo: Sim

RQ05 - Relação entre o tamanho dos PRs e o número de revisões realizadas

Correlação entre Tamanho dos PRs e Número de Revisões



Arquivos alterados x nº revisões

- Coeficiente de correlação: 0.2550
- P-valor: 0.0000e+00
- Interpretação: Fraca
- Estatisticamente significativo: Sim

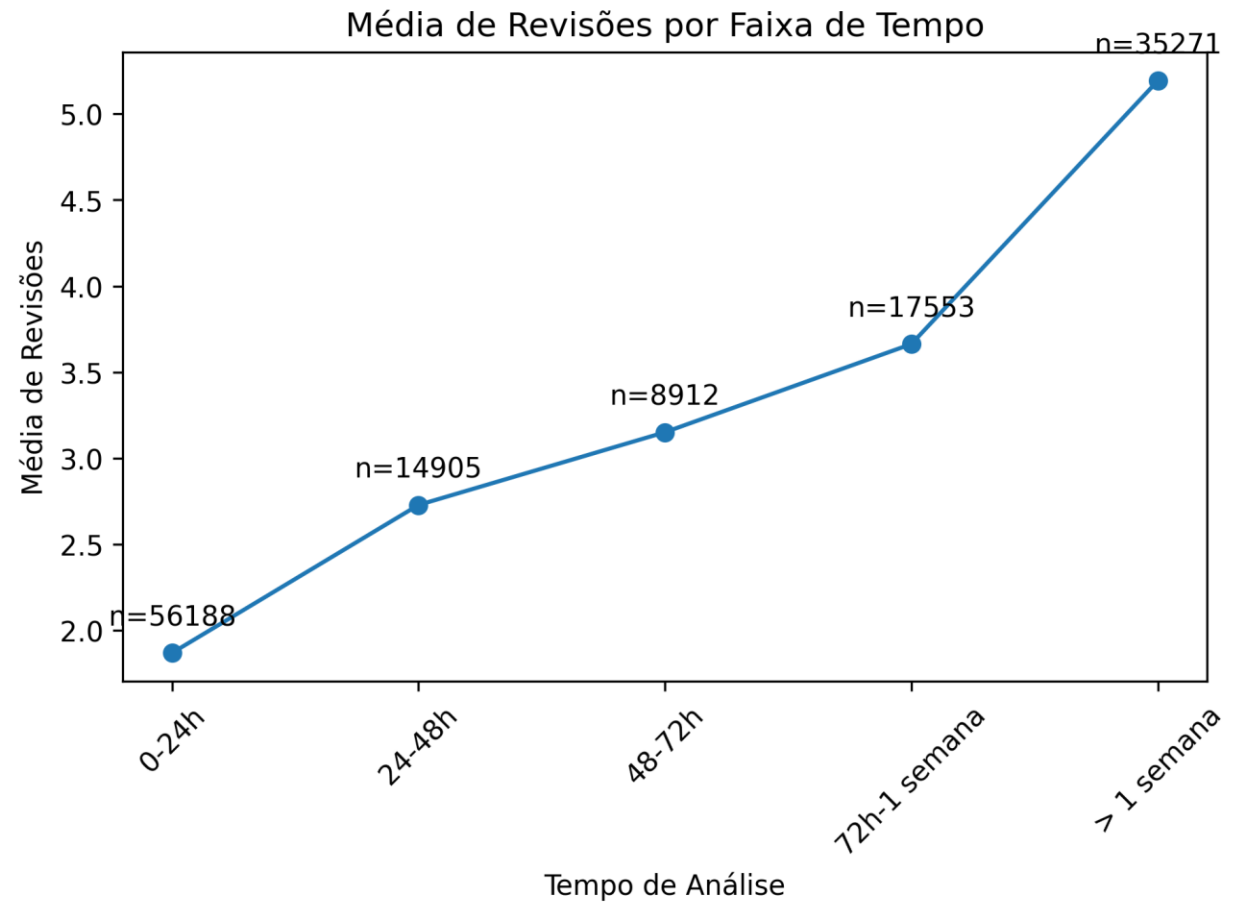
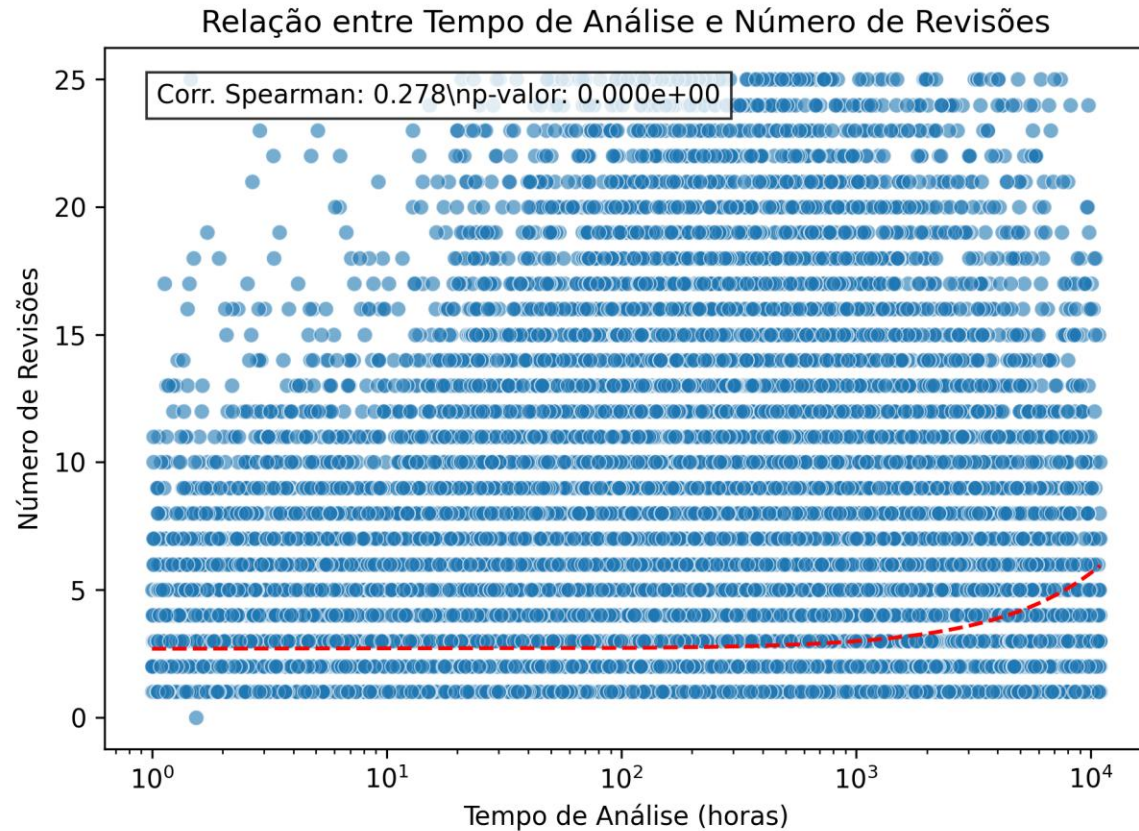
Linhas adicionadas x nº revisões

- Coeficiente de correlação: 0.3150
- P-valor: 0.0000e+00
- Interpretação: Moderada
- Estatisticamente significativo: Sim

Linhas removidas x nº revisões

- Coeficiente de correlação: 0.1627
- P-valor: 0.0000e+00
- Interpretação: Fraca
- Estatisticamente significativo: Sim

RQ06 - Relação entre tempo de análise dos PRs e o número de revisões

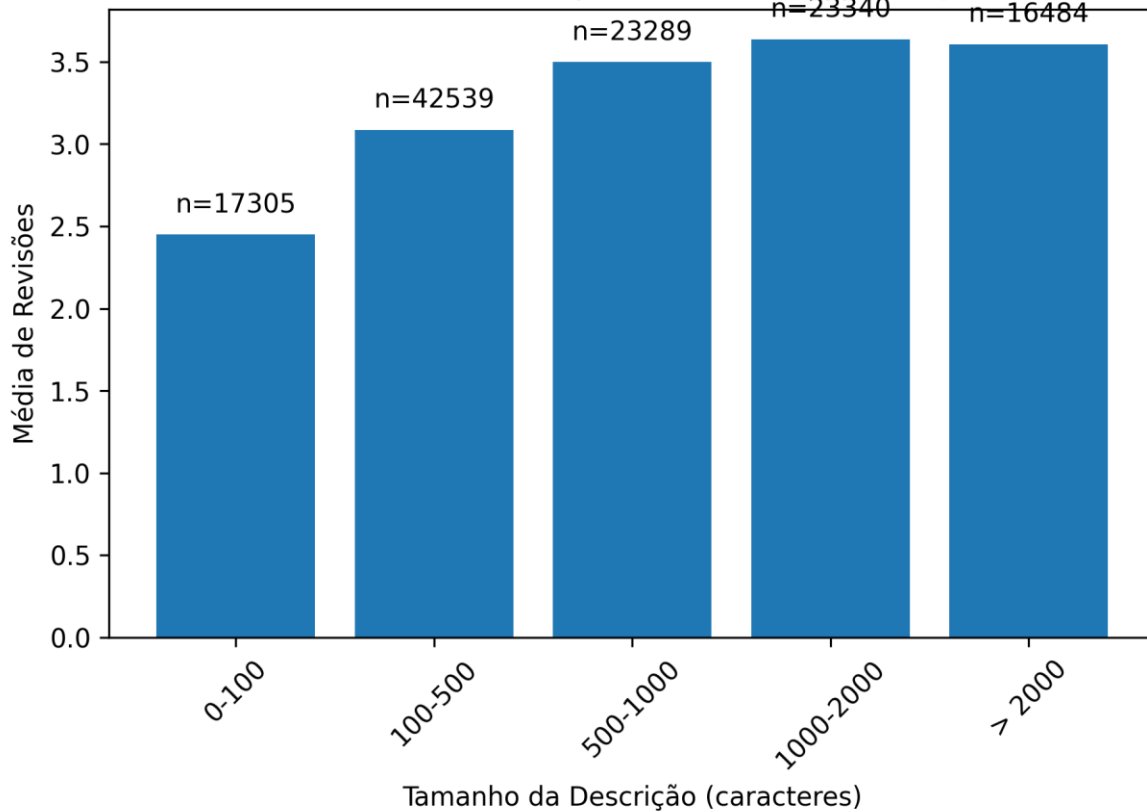


Tempo de análise x nº revisões

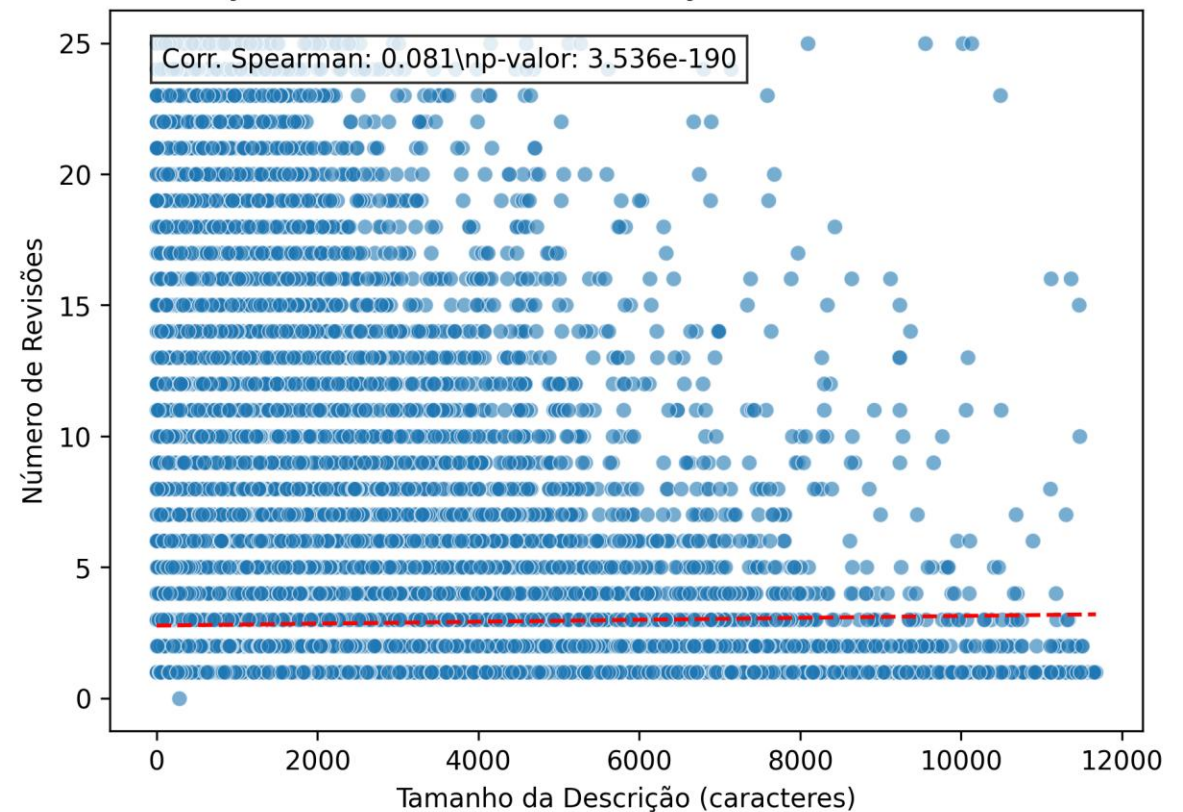
- Coeficiente de correlação: 0.2877
- P-valor: 0.0000e+00
- Interpretação: Fraca
- Estatisticamente significativo: Sim

RQ07 - Relação entre a descrição dos PRs e o número de revisões realizadas

Média de Revisões por Tamanho de Descrição



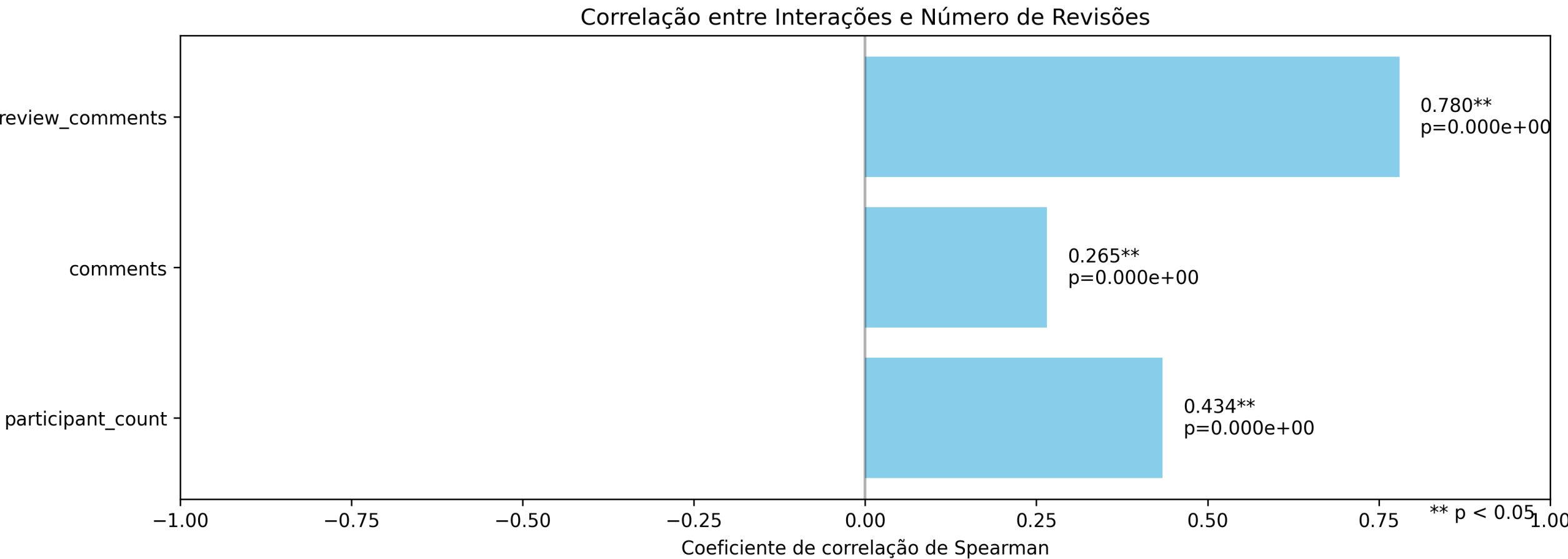
Relação entre Tamanho da Descrição e Número de Revisões



Tamanho da descrição x nº revisões

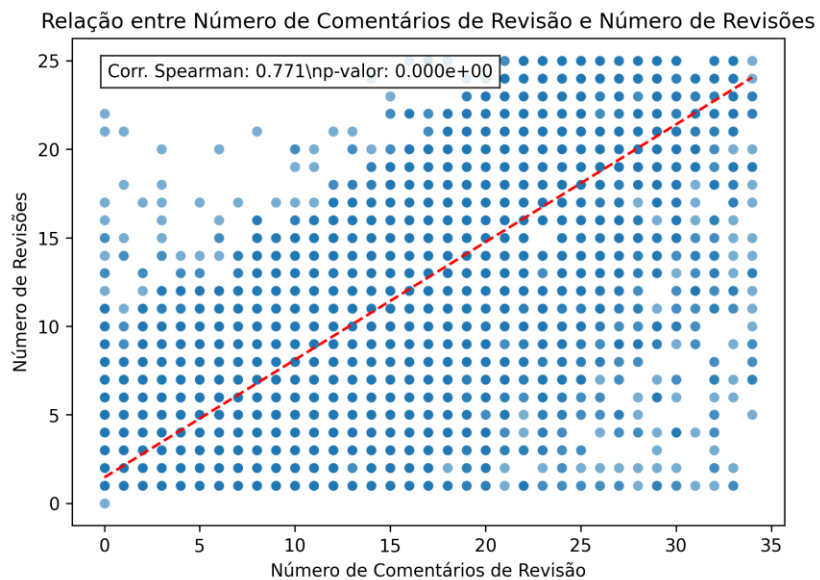
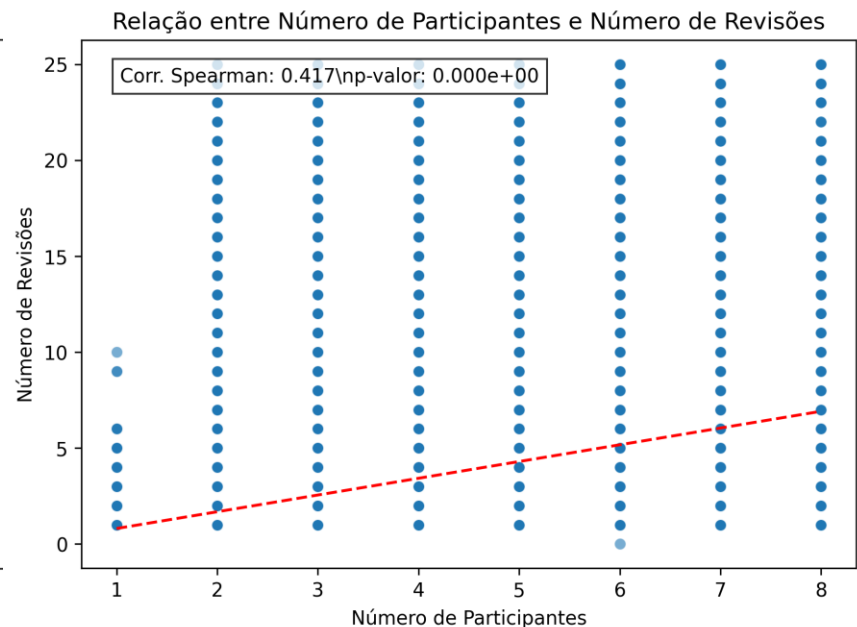
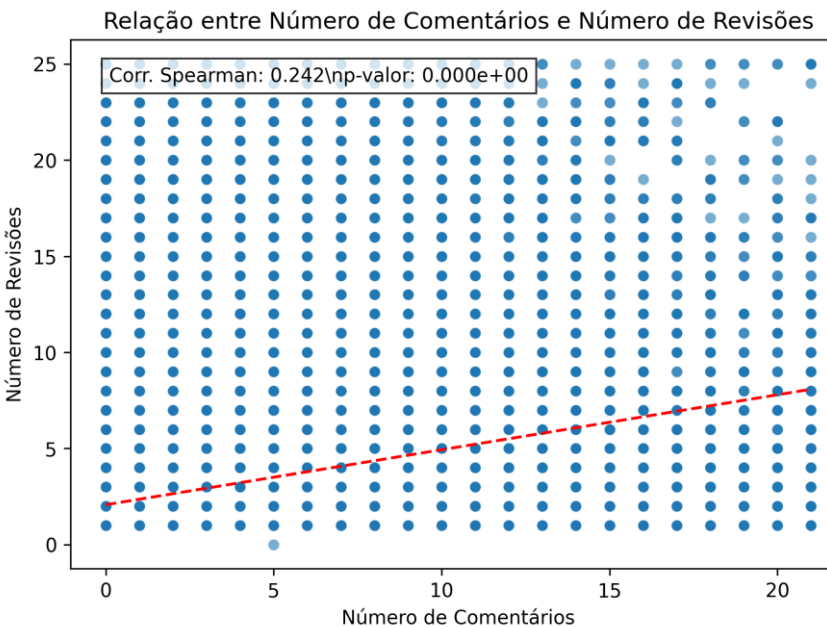
- Coeficiente de correlação: 0.0751
- P-valor: 2.7356e-165
- Interpretação: Insignificante
- Estatisticamente significativo: Sim

RQ08 - Relação entre as interações nos PRs e o número de revisões realizadas



Nº participantes x nº revisões	Nº comentários x nº revisões	Nº comentários de revisão x nº revisões
- Coeficiente de correlação: 0.4342	- Coeficiente de correlação: 0.2654	- Coeficiente de correlação: 0.7801
- P-valor: 0.0000e+00	- P-valor: 0.0000e+00	- P-valor: 0.0000e+00
- Interpretação: Moderada	- Interpretação: Fraca	- Interpretação: Muito forte
- Estatisticamente significativo: Sim	- Estatisticamente significativo: Sim	- Estatisticamente significativo: Sim

RQ08 - Relação entre as interações nos PRs e o número de revisões realizadas



Hipóteses informais

RQ 01: Relação entre o tamanho dos PRs e o feedback final das revisões.

- **Hipótese:** PRs menores têm maior probabilidade de serem aprovados.
- Os resultados não suportam completamente nossa hipótese. A correlação entre o tamanho do PR e sua aprovação não foi tão forte ou significativa como esperávamos.

RQ 02: Relação entre o tempo de análise dos PRs e o feedback final das revisões

- **Hipótese:** PRs que levam mais tempo para serem analisados têm menor probabilidade de serem aprovados.
- Os resultados suportam nossa hipótese. PRs que levam mais tempo para serem analisados têm menor probabilidade de serem aprovados.

RQ 03: Relação entre a descrição dos PRs e o feedback final das revisões

- **Hipótese:** PRs com descrições mais detalhadas têm maior probabilidade de serem aprovados.
- Os resultados suportam nossa hipótese. PRs com descrições mais detalhadas têm maior probabilidade de serem aprovados.

RQ 04: Relação entre as interações nos PRs e o feedback final das revisões

- **Hipótese:** PRs com mais interações têm maior probabilidade de serem aprovados.
- Os resultados não suportam nossa hipótese. As correlações entre as métricas de interação e a aprovação do PR não foram tão fortes ou significativas como esperávamos.

Hipóteses informais

RQ 05: Relação entre o tamanho dos PRs e o número de revisões realizadas

- **Hipótese:** PRs maiores requerem mais revisões.
- Os resultados suportam nossa hipótese. PRs maiores tendem a requerer mais revisões, possivelmente porque contêm mais código a ser analisado e mais problemas potenciais a serem identificados.

RQ 06: Relação entre o tempo de análise dos PRs e o número de revisões realizadas

- **Hipótese:** PRs que levam mais tempo para serem analisados têm mais revisões.
- Os resultados suportam nossa hipótese. PRs que levam mais tempo para serem analisados têm mais revisões, possivelmente porque revisões adicionais são necessárias para resolver problemas identificados.

RQ 07: Relação entre a descrição dos PRs e o número de revisões realizadas

- **Hipótese:** PRs com descrições mais detalhadas têm menos revisões.
- Os resultados não suportam nossa hipótese. A correlação entre o tamanho da descrição e o número de revisões não foi negativa ou significativa como esperávamos.

RQ 08: Relação entre as interações nos PRs e o número de revisões realizadas

- **Hipótese:** PRs com mais interações têm mais revisões.
- Os resultados suportam nossa hipótese. PRs com mais interações têm mais revisões, possivelmente porque cada revisão gera comentários e discussões que podem levar a revisões adicionais.

Conclusão

Com base nos resultados, podemos sugerir as seguintes práticas para melhorar a aprovação de PRs:

- ❑ Manter os PRs pequenos, afetando poucos arquivos e com poucas linhas alteradas.
- ❑ Incluir descrições detalhadas e claras, explicando o propósito e o contexto do PR.
- ❑ Promover interações construtivas durante o processo de revisão, respondendo prontamente aos comentários.
- ❑ Evitar PRs que levem muito tempo para serem analisados, dividindo mudanças grandes em PRs menores e mais focados.

Esperamos que estes insights ajudem desenvolvedores e mantenedores de *projetos open source* a otimizar seus processos de *code review*, melhorando a qualidade do código e a experiência dos contribuidores.

A red pushpin is pinned to the top right corner of the white note.

Thank
😊 you!