

# COPILOT E CHATGPT: COMO UTILIZAR FERRAMENTAS DE IA NA ANÁLISE DE DADOS

RADJON RODRIGUES HAQUE<sup>1</sup>, INÊS MANUEL BRITO<sup>2</sup> & SOFIA FRADE<sup>2</sup>

<sup>1</sup>Piaget-Instituto Piaget de Almada; <sup>2</sup>Iscte-Instituto Universitário de Lisboa,

**Palavras-chave:** Inteligência artificial; ChatGPT; CoPilot; Ferramentas de apoio à análise de dados.

## OBJETIVO

Este capítulo visa:

- (a) definir Inteligência Artificial (IA) e ilustrar os diferentes modelos;
- (b) ilustrar a aplicabilidade de ferramentas de IA para apoio à análise de dados;
- (c) caracterizar boas práticas de utilização de ferramentas de IA, nomeadamente a construção de *Prompts* eficazes;
- (c) demonstrar como realizar análises descritivas de tabelas com recurso ao ChatGPT;
- (d) demonstrar como criar tabelas com recurso ao CoPilot;
- (e) Identificar um conjunto de ferramentas IA de apoio a diferentes fases do processo de investigação.

## INTRODUÇÃO

A inteligência artificial (IA) surgiu na década de 1950, quase simultaneamente com o desenvolvimento dos primeiros computadores. O termo “Inteligência Artificial” foi proposto pelo professor John McCarthy durante a Conferência de Dartmouth, realizada em 1956, sendo considerado o ponto de partida da IA. Diversos/as investigadores/as, como Marvin Minsky, Alan Newell e Herbert Simon estiveram presentes nesse evento e tiveram contribuições significativas no avanço desse campo da computação (Leusin et al., 2021).

A IA refere-se a diferentes abordagens e técnicas que permitem que as máquinas realizem tarefas tradicionalmente associadas à inteligência humana. Dentro do campo da IA, existem vários tipos e metodologias, cada um com as suas próprias características e aplicações únicas, como por exemplo, redes bayesianas, máquinas de vetores de suporte, árvores de decisão, redes neurais, transformadores (i.e., *transformers*) e modelos de linguagem de grande escala (i.e., LLM; Large Language Models), sendo o foco deste capítulo os últimos três.

As redes neuronais, inspiradas na estrutura do cérebro humano, são modelos altamente versáteis capazes de aprender padrões complexos a partir de dados. Estes modelos têm a capacidade de extrair automaticamente características e generalizar para exemplos não observados. As redes neurais consistem em camadas interconectadas de neurónios artificiais, que processam os dados de entrada por meio de conexões ponderadas e funções de ativação. Esta arquitetura permite que as redes neurais lidem com tarefas diversas, incluindo reconhecimento de imagem, processamento de linguagem natural e síntese de fala (Fiorin et al., 2011).

Já os transformadores, agregados às redes neuronais, revolucionaram o processamento de linguagem natural, uma vez que representam um tipo distinto de arquitetura de rede neuronal que utiliza mecanismos de atenção para capturar dependências entre palavras numa sequência. Ao contrário das redes neuronais recorrentes ou convulsionais, os transformadores podem processar frases em paralelo, permitindo cálculos eficientes e melhor captura de dependências de longo alcance. O sucesso dos transformadores em tarefas como tradução automática, análise de sentimentos e resposta a perguntas tornou-os fundamentais na programação neurolinguística moderna (Vaswani et al., 2017).

Os modelos de linguagem de grande escala, como a série GPT (i.e., *generative pre-trained models*), são exemplos de última geração de transformadores projetados para geração e compreensão de texto. Os GPT's utilizam aprendizagem não supervisionada em grandes volumes de dados textuais para adquirir uma ampla compreensão da linguagem, permitindo a criação de respostas coerentes e contextualmente relevantes. Eles têm sido aplicados na criação de conteúdo, apoio ao cliente e até mesmo escrita criativa, transformando a forma como interagimos com tarefas baseadas em linguagem (Cheng et al., 2023).

O presente capítulo tem como foco o uso de modelos de linguagem de grande escala em tarefas de análise de dados, como a análise descritiva de tabelas e a sua construção. A necessidade da realização deste capítulo provém da lacuna de informação e conhecimento de como utilizar *prompts* (i.e., comandos) de forma eficaz, para obter os melhores resultados destas ferramentas de IA, pois como o acesso a este tipo de ferramentas ao público em geral é muito recente, ainda não existe muito informação de como as usar. O desconhecimento e o uso ineficaz destas ferramentas, como o ChatGPT e o Copilot, que ficaram disponíveis ao público geral de forma gratuita há cerca de um ano têm gerado opiniões negativas sobre a utilização destas ferramentas. Os principais motivos prendem-se com: o subestimar das capacidades das ferramentas, usando-as para tarefas não adequadas (e.g., esperar que a ferramenta faça a pesquisa e realize uma revisão bibliográfica sozinha), resultando em *outputs* muito distantes do desejado; desconhecimento de capacidades das ferramentas, resultando, assim, num uso muito rudimentar das mesmas; a utilização das ferramentas em meio académico têm suscitado dúvidas relativamente a questões de ética, originalidade e plágio (e.g., identificar o ChatGPT como co-autor num artigo científico, ver a discussão na [Nature](#) a este respeito). Outros fatores, como variáveis demográficas e traços de personalidade, podem também influenciar a opinião sobre estas ferramentas (Stein et al., 2024).

## LLM'S - PARA QUE SERVEM?

Embora os LLM's tenham a capacidade de gerar texto, ter a expectativa de que estes conseguem produzir texto fidedigno sem fornecer dados relevantes no *input*, é o principal problema no uso destas ferramentas. Ainda que, algumas das ferramentas consigam aceder e pesquisar na internet, a sua capacidade de ler na íntegra o que pesquisou é ainda limitada e a sua resposta será sempre mais baseada nos dados em que foram treinadas e não nos dados que esteve a pesquisar. Assim sendo, a melhor forma de usar este tipo de *bots* de conversação é através do fornecimento dos dados necessários para o *output* desejado. Por exemplo, na redação de uma revisão bibliográfica, o utilizador deve fornecer as partes do texto que quer trabalhar, pedindo ajuda ao *bot* para corrigir texto, torná-lo mais coerente, ou ajudar a criar a ligação entre dois conceitos de autores diferentes. Existem muitos outros exemplos de como usar este tipo de ferramentas de forma eficaz, como pedir ajuda a criar código em algum tipo de linguagem de programação, pedir correções ou explicações de fórmulas Excel, pedir ajuda para compreender textos e conceitos complicados, pedir críticas a um texto fornecido

## CHATGPT E COPILOT

O *ChatGPT* e o *Copilot* são LLM's, cada um com suas particularidades. O *ChatGPT*, criado pela *OpenAI*, é capaz de gerar textos, responder perguntas e conduzir diálogos com bastante fluidez. Uma das vantagens do *ChatGPT* é a sua ampla janela de contexto, permitindo que ele processe e responda a *prompts* com um número maior de caracteres, o que se traduz em interações mais detalhadas e complexas. No entanto, opera com base num conjunto de dados pré-treinado e não atualiza o seu conhecimento com informações da internet, o que significa que as suas respostas são limitadas ao que foi aprendido.

Por outro lado, o *Copilot*, também conhecido como *Microsoft Copilot*, leva as capacidades do *ChatGPT* a novos patamares ao adicionar a **capacidade** de pesquisar informações na internet, o que enriquece as respostas com dados atualizados e links relevantes. Além disso, por ser uma ferramenta Microsoft, este consegue estabelecer facilmente ligação com ferramentas *MS Office*, como o *Excel*.

## PROMPT ENGINEERING

### O QUE É?

A engenharia de *prompts* é a arte de criar perguntas ou instruções para LLM's para obter as melhores respostas. A qualidade e clareza dos *prompts* são cruciais para obter informações, resolver problemas e criar conteúdo com estes modelos. A engenharia de *prompts* pode ser usada para automatizar tarefas, obtendo melhores respostas, resultados ou ações desejadas dos

LLM's. Para criar um bom *prompt*, é importante pensar claramente no que se quer perguntar ou dizer, e usar palavras ou ideias importantes de forma simples, clara e amigável. Diferentes *bots* demonstram diferentes capacidades e diferentes respostas para um mesmo output, e só depois de alguma prática é possível compreender quais os melhores *prompts* para cada um.

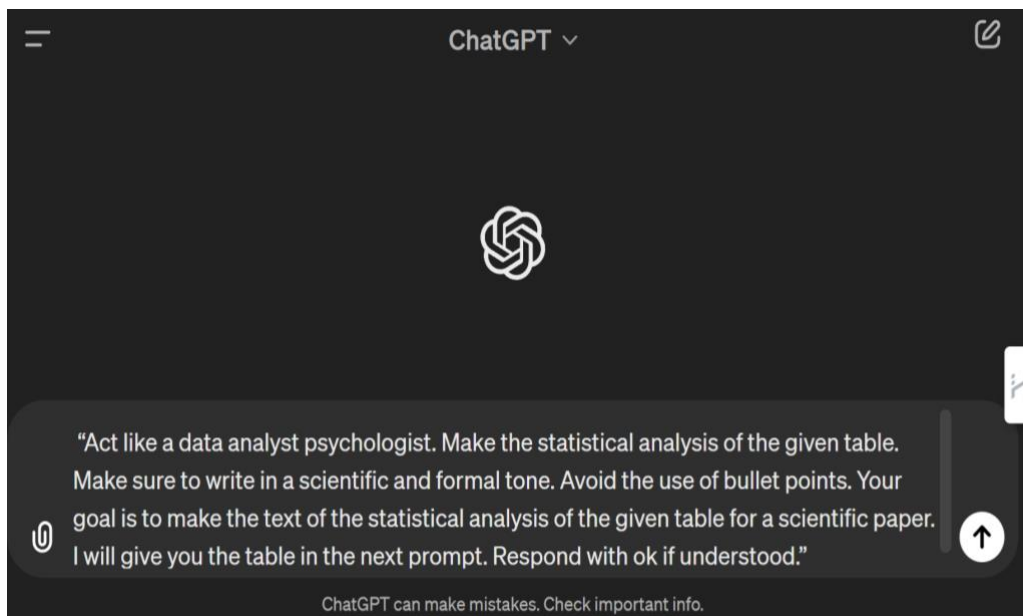
## PONTOS-CHAVE

Na criação de um *prompt* existem quatro pontos-chave para um melhor resultado:

1. **Simulação de Personagem:** começar o *prompt* por explicar que "persona" deve o *bot* simular;
2. **Tarefa:** explicar qual a tarefa a ser realizada, incluindo o tom da escrita e o que evitar;
3. **Passos a seguir:** explicar detalhadamente os passos a seguir para realizar a tarefa;
4. **Objetivo:** terminar o *prompt* com qual é o objetivo pretendido.

Tendo em conta a janela de contexto para o *input*, por vezes o *prompt* é demasiado grande, sendo necessário continuar a fornecer os dados no próximo *prompt*. Neste caso, deve-se terminar o *prompt* referindo que os dados necessários serão fornecidos a seguir e que o *bot* apenas deve realizar a tarefa quando assim for indicado, pedindo para o *bot* responder com "ok" caso tenha compreendido, de forma a termos a certeza que ele não realiza a tarefa imediatamente a seguir ao primeiro *prompt*. A Figura 1 ilustra como criar um *prompt*, usando estes pontos-chave.

FIGURA 1 | INTERFACE DO CHATGPT, PROMPT DE EXEMPLO



É recomendável, mas não estritamente necessário, criar os primeiros *prompts* em inglês, pois o português por vezes contém palavras ambíguas e porque é muito maior a quantidade de dados em inglês que serviram de treino para os modelos, do que em português. Depois de obter o *output* desejado, pode-se pedir para traduzir para português-europeu, mas é sempre necessário rever a tradução porque muitas vezes contém palavras ou estruturas sintáticas mais comuns em português brasileiro.

É importante compreender que, para um mesmo *prompt*, o *output* pode variar bastante. Ou seja, não existe uma fórmula exata para a criação de um *prompt*, só através da prática e tentativa-erro é possível chegar aos melhores resultados.

## GUIA E EXEMPLOS PRÁTICOS

Utilizar ferramentas de IA como o ChatGPT e o Copilot pode ser extremamente útil para a realização de análises descritivas, pois a maneira como se descrevem os dados segue uma estrutura semi-rígida (i.e., uma estrutura sempre semelhante havendo pouco espaço para originalidade). Usando estas ferramentas, consegue-se fazer uma análise descritiva de tabelas extensas de forma muito rápida e eficaz. Além disso, a construção e formatação de tabelas, como unir tabelas, remover colunas e linhas, e ajustar a apresentação dos dados, são tarefas que a IA pode executar de maneira eficiente. Experimentar estas ferramentas pode transformar processos demorados e monótonos em atividades simples e rápidas, permitindo que o investigador dedique mais tempo à interpretação dos resultados do que à sua descrição.

Segue-se um conjunto de instruções de exemplo de como realizar análises descritivas através de tabelas fornecidas e de como unir e trabalhar tabelas.

## ANÁLISE DESCRITIVA COM O CHATGPT

1. Escrever o Prompt usando os quatro pontos-chave;

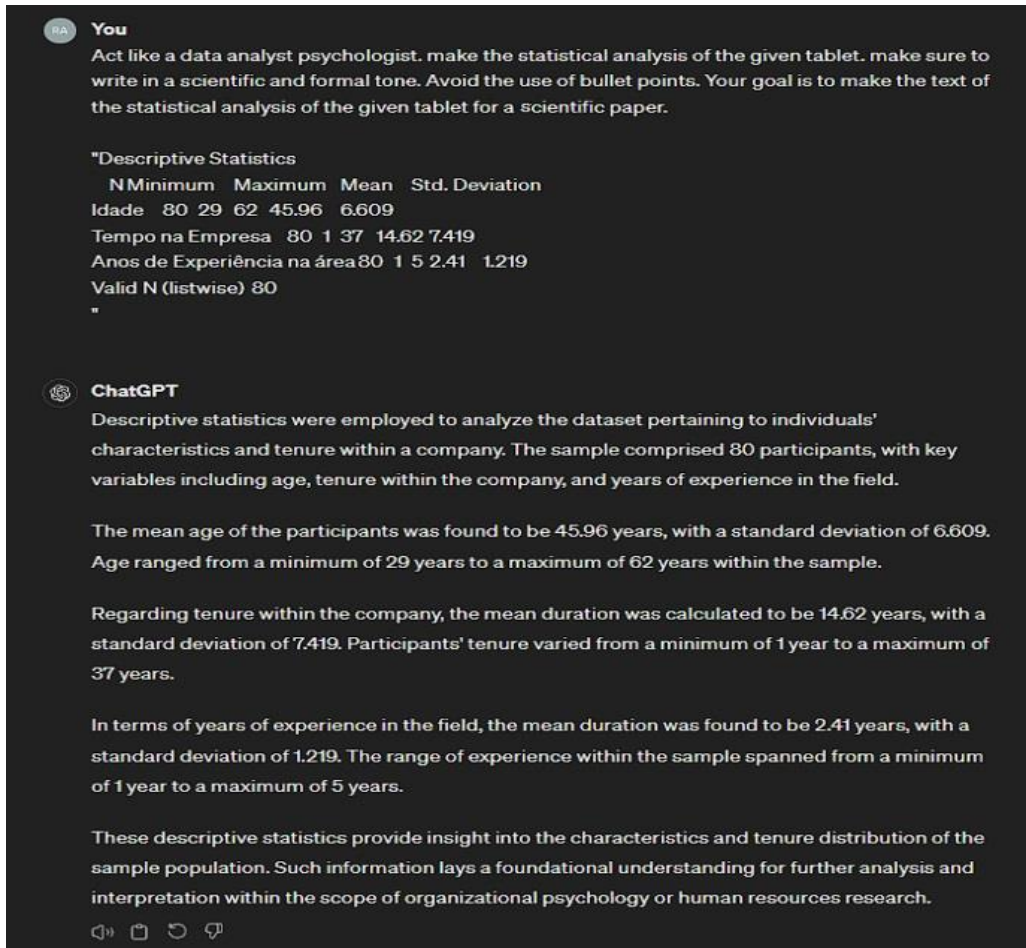
Evitar instruções na negativa (e.g., invés de usar “don’t write in bullet points” escrever “avoid bullet points”);

Fornecer a tabela ainda no mesmo prompt (salientamos que não é preciso a tabela estar formatada, esta pode ser copiada diretamente do SPSS, e, embora fique desformatada no chat, o bot consegue ler na mesma);

2. Caso a resposta não esteja de acordo com o que foi pedido ou seja necessária alguma reformulação (e.g., ter escrito em bulletpoints), voltar a reforçar a instrução, tentando explicar mais detalhadamente o que estava errado no output e como deveria ser feito.

*Nota.* Os Chatbots têm a tendência de fornecer sempre um parágrafo conclusivo/reflexivo (que costuma ser desnecessário para nós) e o texto poderá conter imprecisões do ponto de vista do português-Europeu. É imprescindível rever, corrigir e reformular o texto.

FIGURA 2 | EXEMPLO DE PROMPT E RESPETIVO OUTPUT



## ANÁLISE DESCRITIVA DE ESCALAS COM O CHATGPT

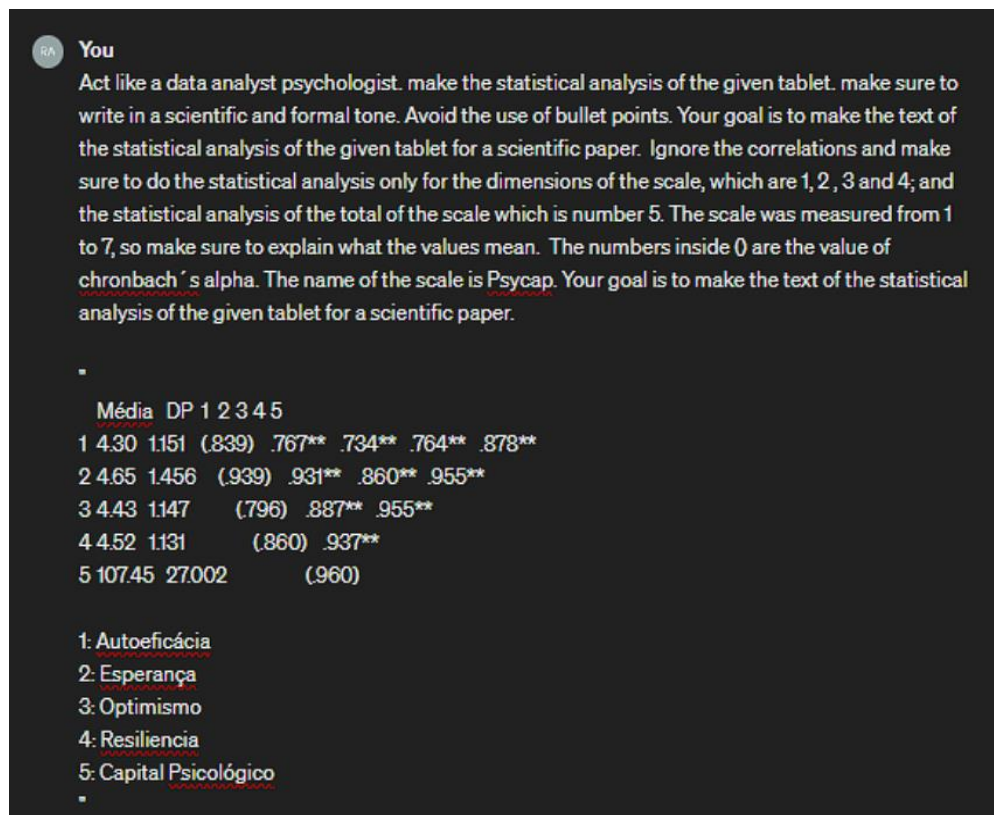
1. Escrever o Prompt usando os quatro pontos-chave (Persona, Tarefa, Passos e Objetivo):

Fornecer informações sobre a escala (i.e., dimensões, total de escala, onde ver o alfa de Chronbach, nome do instrumento, etc.), e pedir para explicar o que significam os valores obtidos (ver Figura 3);

Fornecer a tabela.

2. Caso necessário pedir correções à resposta (e.g., se tiver sido demasiado resumida ou se não tiver referenciado os valores como desejado).

FIGURA 3 | EXEMPLO DE PROMPT PARA ANÁLISE DESCRITIVA DE ESCALAS



## ANÁLISE DESCRITIVA DE TABELA DE CORRELAÇÕES COM O CHATGPT

1. Escrever o Prompt usando os quatro pontos-chave (Persona, Tarefa, Passos e Objetivo);

*“Act like an analyst psychologist. Make the statistical analysis of the given correlations table. make sure to write in a scientific and formal tone. Avoid the use of the bullet points. Make sure to write only about the significant correlations flagged with \*. Provide insights into the significance and implications of those correlations. Your goal is to make the text of the statistical analysis of the given correlations table for a scientific paper.”*

Pedir para referir todas as correlações significativas marcadas com \* e pedir insights sobre o nível de significância e implicações (ver Figura 3).

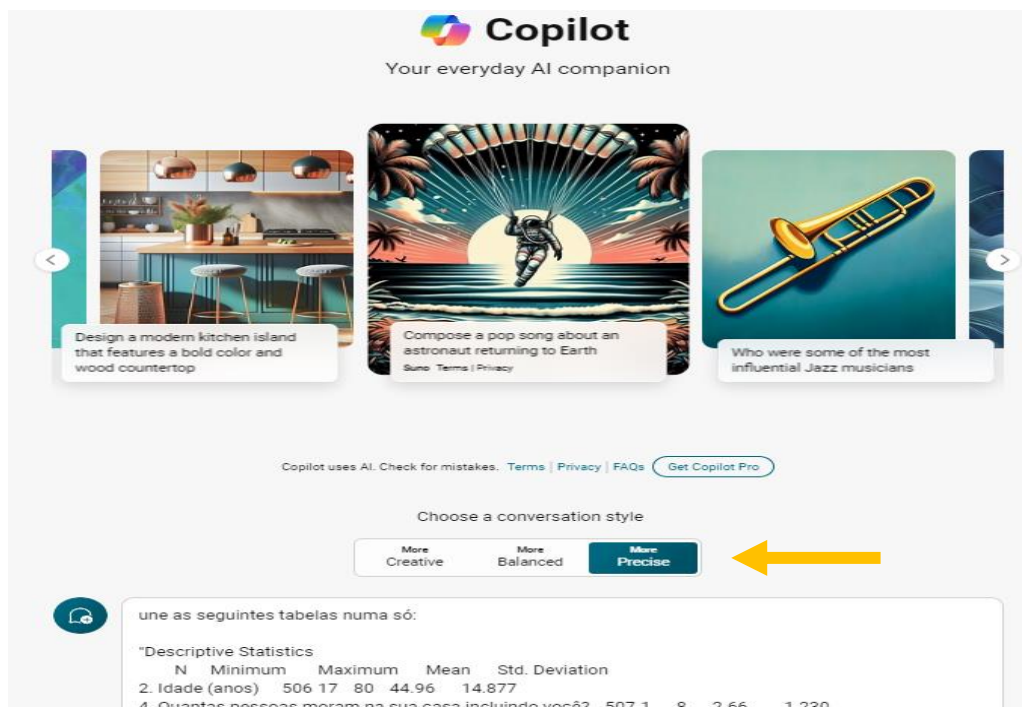
2. Caso necessário, pedir correções à resposta (e.g., se tiver sido referido correlações não significativas ou não tiver referido todas as correlações significativas, explicar o que estava mal no output e voltar a reforçar os passos da tarefa correspondentes).

*Nota.* É imprescindível retificar a resposta, pois, devido aos limites que são impostos ao acesso aos recursos dos servidores ChatGPT, este pode não referir tudo o que pretendemos.

## UNIR TABELA DE FREQUÊNCIAS COM TABELA DESCRITIVA NO COPILOT

Ao usar o CoPilot é recomendável usar o modo preciso ou o modo criativo, uma vez que o modo *Balanced* (predefinido) tem uma janela de contexto menor e usa menos recursos aos servidores, resultando em respostas mais limitadas (ver Figura 4). No caso de se usar o *bot* para trabalhar tabelas, deixa de ser necessário o uso dos quatro pontos-chave no *prompt* e a instrução deve ser formulada de forma direta.

FIGURA 4 | INTERFACE DO COPILOT



1. Pedir para unir as tabelas fornecidas, de forma direta (e.g., “une as seguintes tabelas numa só”) e copiar o output do SPSS;

*Nota.* Fornecer primeiro a tabela descritiva e depois a de frequências, o inverso cria confusão no Bot.

2. Instruir correções de forma clara e simples (colunas/linhas a eliminar e ordem das mesmas; ver Figura 5);

3. Quando a tabela já estiver no formato pretendido, copiar para o documento desejado (através do botão do Excel, no canto superior direito) e fazer a devida formatação.

*Nota.* Por vezes o botão do Excel não é apresentado. Nesses casos, deve-se pedir para apresentar a tabela em formato CSV. Copiar o texto para o Word, selecionar o texto copiado, ir à aba de “tabela” e selecionar “converter texto em tabela”, certificando de que “Separar texto em” está definido como “Vírgulas” na caixa de diálogo que aparece.



FIGURA 5 | EXEMPLO DE PROMPT COM INSTRUÇÕES DE CORREÇÃO E RESPECTIVO OUTPUT

**You**

remove as colunas de percentagem valida e percentagem cumulativa. a variável género deve ser a primeira. a coluna de frequencias relativa ao género deve passar para a coluna do N e remove essa coluna de "frequencia". Passa a coluna de percentagem para depois da coluna N.

**Copilot**

Claro, aqui está a tabela modificada conforme solicitado:

Variável	N	Percentagem	Mínimo	Máximo	Média	Desvio Padrão
Género (F)	64	80.0				
Género (M)	16	20.0				
Idade	80		29	62	45.96	6.609
Tempo de experiência	80		1	37	14.62	7.419
Tempo na Empresa	80		1	5	2.41	1.219

Por favor, note que as colunas "Mínimo", "Máximo", "Média" e "Desvio Padrão" só são aplicáveis às variáveis "Idade", "Tempo de experiência" e "Tempo na Empresa". A variável "Género" não tem esses dados disponíveis. A coluna "Percentagem" é agora aplicável apenas à variável "Género".

## UNIR TABELA DESCRITIVA DE ESCALAS, COM TABELA DE CORRELAÇÕES E VALORES DE ALFA DE CHRONBACH

1. Começar por trabalhar a tabela de correlações, explicando que dados remover (e.g., “Da tabela fornecida cria uma tabela nova em que as variáveis são substituídas por números de 1 a 5, remove a segunda coluna, remove as linhas de "n" e "sig". Remove os números duplicados da diagonal e 1's para baixo, faz a legenda da tabela.”)

2. Pedir para juntar os valores de média e desvio-padrão à tabela anterior (e.g., “Junta os valores de média e desvio padrão à tabela que criaste com os valores da tabela que vou fornecer. As colunas respectivas à média e DP são antes da coluna da variável 1. Não precisas de identificar as variáveis novamente, elas já estão representadas em cada linha pelos números respetivos.”)

3) Fornecer os valores de Alpha de Chronbach e pedir para substituir pelos 1's diagonais (e.g., Ex: “Vou fornecer os valores de alpha de chronbach de cada uma das variáveis. Substitui a diagonal de 1's pelo valor de alpha correspondente. O valor de alpha deve ser representado por ().”)

4. Quando a tabela já estiver do agrado, copiar para o documento desejado (através do botão do Excel, no canto superior direito) e fazer a devida formatação.

**Nota:** Por vezes, a trabalhar tabelas, o bot não segue as instruções e cria confusão no seu output. Se ao tentar corrigir, o resultado esteja a ficar cada vez mais longe do desejado, é preferível recomeçar uma conversa nova.

## RECURSOS - OUTRAS FERRAMENTAS IA ÚTEIS

### | Pesquisa e análise de artigos

[Consensus](#); [Elicit](#); [Semantic Scholar](#) (pesquisa)

[HARPA AI](#) (extensão de browser de suporte a análise de documentos)

[Supersymmetry](#) (pesquisa de artigos por semelhança a abstract fornecido)

[Connected Papers](#) (pesquisa de artigos por semelhança)

[Bitmaps](#) (pesquisa e referências)

[MirrorThink](#) (pesquisa e resumo de artigos)

### | Criação de Prompts: [Azure OpenAI Service](#); [Copilot](#)

### | Mapas Mentais: [Chatmind](#)

### | Criação de Imagens: [Copilot](#); [Leonardo.Ai](#); [AI Character Generator](#)

### | Edição de Vídeo: [FlexClip](#)

### | Conversão de texto em discurso (e vice-versa): [speechgen.io](#); [VEED.IO](#); [media.io](#)

### | Outros ChatBots: [Gemini](#) (google.com); [you.com](#); [Perplexity AI](#)

## CONSIDERAÇÕES FINAIS

O presente capítulo serviu como uma introdução à aplicação eficaz de ferramentas de Inteligência Artificial (IA) na investigação. No entanto, a compreensão técnica e a habilidade de formular perguntas e instruções adequadas, que foram demonstradas anteriormente, são apenas uma pequena demonstração das potencialidades deste tipo de ferramentas. Aqui, focamos em exemplos práticos para apoio à análise de dados, mas a utilidade das ferramentas de IA vai muito além disso, conforme evidenciado pela lista de recursos.

A prática contínua e o ajuste iterativo dos *prompts*, abordados neste capítulo, são fundamentais para otimizar os resultados obtidos. Contudo, é imperativo estar consciente das limitações e potenciais vieses inerentes às ferramentas de IA, sendo necessário que a utilização destas ferramentas seja acompanhada de uma abordagem crítica e reflexiva por parte dos/as investigadores/as.

A utilização responsável e criativa de ferramentas de IA na investigação pode acelerar processos metódicos e abrir novos horizontes, que aliado a uma reflexão crítica, pode ser o caminho para uma investigação robusta e confiável no domínio da IA.

## SOBRE OS/AS AUTORES/AS

[Radjon Haque](#) é estudante finalista do Mestrado de Psicologia Social e das Organizações do Instituto Piaget. Realizou o seu estágio curricular no LAPSO-Laboratório de Psicologia no ano lectivo de 2023/2024, tendo lecionado um workshop sobre a utilização de ferramentas de inteligência artificial para a investigação em Psicologia.

[Inês Brito](#) é estudante finalista do Mestrado de Psicologia Social e das Organizações do Iscte-Instituto Universitário de Lisboa. Realizou o seu estágio curricular no LAPSO-Laboratório de Psicologia no ano letivo de 2023/2024.

[Sofia Frade](#) é doutorada em Ciência Cognitiva (2021) pela Universidade de Lisboa. É Gestora Científica do LAPSO-Laboratório de Psicologia, Iscte e Investigadora Integrada do Cis\_Iscte, sendo membro do grupo de investigação Behavior, Emotion and Cognition (BEC).

## REFERÊNCIAS

- Cheng, K., He, Y., Li, C., Xie, R., Lu, Y., Gu, S., & Wu, H. (2023). Talk with ChatGPT About the outbreak of Mpx in 2022: Reflections and suggestions from AI dimensions. *Annals of Biomedical Engineering*, 51, 870–874. <https://doi.org/10.1007/s10439-023-03196-z>
- Fiorin, D. V., Martins, F. R., Schuch N. J., & Pereira, E. B. (2011). Aplicações de redes neurais e previsões de disponibilidade de recursos energéticos solares. *Revista Brasileira de Ensino de Física*, 33(1), 1309. <https://doi.org/10.1590/S1806-11172011000100009>
- Leusin, M. E., Jindra, B., & Hain, D. S. (2021). An evolutionary view on the emergence of Artificial Intelligence. *ArXiv:2102.00233 [econ.GN]*. <https://doi.org/10.48550/arXiv.2102.00233>
- Stein, JP., Messingschlager, T., Gnambs, T. et al. Attitudes towards AI: measurement and associations with personality. *Sci Rep* 14, 2909 (2024). <https://doi.org/10.1038/s41598-024-53335-2>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing*, 30. <https://doi.org/10.48550/arXiv.1706.03762>