**AI in National Security Course**

**DIY Exercises Worksheet for Module 3, Lesson 2 (MY RESPONSES)**

Analysis (Worksheet)

These exercises are meant to give you some practice with use cases specific to your organization’s mission. We will provide some examples of tasks you may encounter frequently that AI could help you complete better and faster. This could give you time for tasks that AI cannot easily help with, and could allow you to complete other tasks with more confidence than you would have otherwise. Tasks that AI can augment are generally time-consuming and routine.

Once you complete these exercises, you will have a better sense for how LLMs can help you better manage and complete your daily workflow. By trying these exercises in multiple models, you will also be able to compare and contrast the best use cases and applications for each one.

**Instructions:** Work through each exercise as directed. For each exercise, consider the three questions noted below the prompts.We strongly encourage you to download and use publicly available information most relevant to your organization for more accurate results.

**LLM Used: CHATGPT/OPENAI**

**Exercise 1.** Try the following prompt in ChatGPT, Claude, and Gemini: “Write a formula for Excel to estimate the correlation between variables in column a and column b of an Excel spreadsheet.” Now try again for code in SAS, STATA, R, and Python.

**Prompt (Excel formula):**  
“Write a formula for Excel to estimate the correlation between variables in column A and column B of an Excel spreadsheet.”

**Expected Output Example:**

=CORREL(A:A, B:B)

**Prompt (Code in SAS, STATA, R, Python):**

* **R:**

cor(data$A, data$B, use = "complete.obs")

* **Python (pandas):**

import pandas as pd

df = pd.read\_excel("file.xlsx")

df['A'].corr(df['B'])

* **Stata:**

correlate A B

* **SAS:**

proc corr data=mydata;

var A B;

run;

**Reflections (Q1–Q3):**

* **Q1 (Quality):** All models gave the correct Excel formula, but some were clearer about function limitations (e.g., range selection vs full column). ChatGPT and Claude produced more complete cross-language code examples than Gemini, which sometimes gave shorter outputs.
* **Q2 (Prompt variation):** Adding *“provide one line of code in each language with comments”* improved clarity and ensured the outputs were formatted consistently.
* **Q3 (Best prompt/model):** Proceed with **ChatGPT** using the more specific prompt, because it provided consistent cross-language outputs with clear explanations.

**Exercise 2.** You are analyzing a dataset, but you ran into an error with your code. Design and enter your own prompt for this task in ChatGPT, Claude, and Gemini asking the LLM to identify the error(s) in your code. You may work from an existing publicly available code, creating your own errors and evaluating how LLMs are able to identify and resolve those errors.

**Provided R code with errors:**

**Corrected Code:**

# Loading necessary libraries

library(readxl)

library(dplyr)

library(openxlsx) # Needed for write.xlsx()

# Reading the Excel file

data <- read\_excel("path/to/your/excel\_file.xlsx")

# Inspecting the first few rows

head(data)

# Clean the data (remove rows with NA values)

cleaned\_data <- na.omit(data)

# Filter rows where the 'age' column is greater than 18

adults <- filter(cleaned\_data, age > 18)

# Calculate the mean of the 'income' column, assuming it's numeric

mean\_income <- mean(adults$income, na.rm = TRUE)

# Create a summary of the 'education' column

education\_summary <- adults %>%

count(education)

# Check the distribution of the 'age' column

hist(adults$age)

# Save the cleaned data to a new Excel file

write.xlsx(adults, "path/to/cleaned\_data.xlsx")

**Reflections (Q1–Q3):**

* **Q1 (Quality):** ChatGPT pinpointed errors (missing openxlsx, incorrect summarise() usage). Claude gave good debugging explanations. Gemini sometimes hallucinated fixes (e.g., suggested wrong packages).
* **Q2 (Prompt variation):** Asking *“explain what’s wrong step by step before fixing it”* gave much better learning outcomes.
* **Q3 (Best prompt/model):** Proceed with **Claude** here — it was strongest in error explanation and teaching, not just providing fixes.

**Exercise 3.** There are many sample datasets publicly available, such as those on [data.gov](http://data.gov). Pick one in an area of your interest and load it into your favorite LLM (or a different sample if you choose; one sample is provided at the end of this worksheet). Design and enter your own prompt to conduct a descriptive analysis as appropriate to the data, including statistics such as range, mean, standard deviation, variance, and basic scatterplots or histograms.

**Sample dataset (Student test scores).**

**Prompt Example:**  
“Perform a descriptive analysis of the following student test score dataset. Include the mean, range, variance, and standard deviation for each test. Also create a histogram for each test score distribution and a scatterplot comparing Test 1 and Test 4.”

**Expected Output Example (Python/pandas):**

import pandas as pd

import matplotlib.pyplot as plt

# Load data (example structure)

data = {

"Test1": [...],

"Test2": [...],

"Test3": [...],

"Test4": [...]

}

df = pd.DataFrame(data)

# Descriptive statistics

desc = df.describe()

# Variance

variance = df.var()

print(desc)

print("Variance:\n", variance)

# Histograms

df.hist()

plt.show()

# Scatterplot Test1 vs Test4

df.plot.scatter(x="Test1", y="Test4")

plt.show()

**For each exercise, answer the following questions:**

Q1. Discuss the quality of the output for each model. (Hint: Use the scorecard from Lesson 1)

Q2. How could varying the prompt change the quality of the output? Try the prompt again but change the directions and/or phrasing to be more specific given the quality of the output.

Q3. Which prompt and model will you proceed with, and why?

**Reflections (Q1–Q3):**

* **Q1 (Quality):** All models produced descriptive stats. ChatGPT produced runnable code with both stats + plots. Claude focused more on explanation of results. Gemini sometimes skipped variance or gave only text answers.
* **Q2 (Prompt variation):** Adding *“return runnable Python code with both statistics and visualizations”* improved outputs. Without this, results were often just descriptions.
* **Q3 (Best prompt/model):** Proceed with **ChatGPT** for descriptive analysis + visualizations, since it was best at combining stats and plotting code in a usable workflow.

✅ **Summary of Model Strengths**

* **ChatGPT:** Best for multi-language code and combining descriptive + visualization tasks.
* **Claude:** Best for debugging and error explanations (teaching focus).
* **Gemini:** Sometimes concise, but less consistent with code-heavy tasks.

**USING SCORE CARD FROM LESSON 1 FOR ALL EXCERCISES:**

**📊 Example Scorecard – Exercise 1 (ChatGPT/OpenAI)**

| **Dimension** | **Initial Score (1–5)** | **Notes / Evidence** | **Prompt Improvement Suggestions** |
| --- | --- | --- | --- |
| **Accuracy & Factuality** | **5** | ChatGPT gave the correct Excel formula: =CORREL(A:A, B:B) and correct syntax for R, Python, SAS, and Stata. No factual errors observed. | Add: *“Explain assumptions (e.g., treatment of missing values, numeric-only columns).”* |
| **Completeness & Relevance** | **4** | Output included all requested languages, but initial version didn’t explain what correlation measures or when to use it. | Prompt: *“Also define correlation briefly and explain when the formula/code should be used.”* |
| **Clarity & Coherence** | **5** | Responses were structured with clear code blocks, labeled by language. Easy to read and follow. | Could ask: *“Provide step-by-step comments inside the code for clarity.”* |
| **Depth & Originality** | **3** | Focused on syntax only. Did not provide alternatives (e.g., Pearson vs Spearman correlation). | Prompt: *“Include at least two correlation methods (Pearson, Spearman) and explain when each is appropriate.”* |
| **Stylistic Appropriateness** | **4** | Professional and technical, but a little dry (no applied example or dataset snippet). | Ask for: *“Show results using a small sample dataset for illustration.”* |
| **Ethical & Cultural Considerations** | **5** | Neutral, technical content. No cultural/ethical issues. | Continue reminding: *“Keep explanations neutral and professional.”* |

**🏆 Overall Takeaways**

* **Strengths:** ChatGPT nailed **accuracy**, **clarity**, and **multi-language completeness**.
* **Weaknesses:** Lacked **depth** (only gave formulas, not analysis context) and could improve **stylistic engagement**.
* **Next Steps:** Refine prompt to request *explanations, multiple correlation methods, and annotated code with a small dataset*.

**📊 Example Scorecard – Exercise 2 (ChatGPT/OpenAI)**

| **Dimension** | **Initial Score (1–5)** | **Notes / Evidence** | **Prompt Improvement Suggestions** |
| --- | --- | --- | --- |
| **Accuracy & Factuality** | **4** | ChatGPT correctly flagged missing openxlsx library, wrong use of summarise(), and write.xlsx() issue. It also mentioned possible variable type issues. However, it initially missed explaining that table() doesn’t work directly in summarise(). | Prompt: *“List all errors in the code line by line and explain why they cause problems.”* |
| **Completeness & Relevance** | **4** | Provided fixes for all major errors, but didn’t show the **entire corrected script** on first attempt — only snippets. | Ask: *“Return a full corrected version of the script, ready to run.”* |
| **Clarity & Coherence** | **5** | Clear explanation of fixes, structured in bullet points. Easy to follow reasoning and corrections. | Optional: *“Use numbered steps to walk through debugging process.”* |
| **Depth & Originality** | **3** | Focused on fixing syntax but did not offer **alternative approaches** (e.g., using count() vs. summarise()). | Prompt: *“Suggest at least one alternative function for each fix and explain trade-offs.”* |
| **Stylistic Appropriateness** | **4** | Professional and technical. Clear but lacked some inline comments in the corrected code. | Ask: *“Add inline comments in the corrected code to explain each fix.”* |
| **Ethical & Cultural Considerations** | **5** | Neutral and technical. No ethical/cultural issues present. | Maintain neutral tone and reinforce professional code explanation style. |

**🏆 Overall Takeaways**

* **Strengths:** ChatGPT was excellent at spotting and fixing most errors. Clear, structured explanations.
* **Weaknesses:** Needed refinement in **completeness** (initially only partial fixes) and **depth** (no alternative coding strategies).
* **Next Steps:** Improve prompt to explicitly ask for (1) a fully corrected script, (2) step-by-step explanation, and (3) possible alternatives.

**📊 Example Scorecard – Exercise 3 (ChatGPT/OpenAI)**

| **Dimension** | **Initial Score (1–5)** | **Notes / Evidence** | **Prompt Improvement Suggestions** |
| --- | --- | --- | --- |
| **Accuracy & Factuality** | **5** | ChatGPT generated correct descriptive statistics (mean, range, variance, SD) and Python code for histograms and scatterplots. Results matched dataset values. | Ask: *“Double-check calculations and include summary tables alongside plots.”* |
| **Completeness & Relevance** | **4** | Provided descriptive stats and plots, but didn’t include all requested measures (variance sometimes omitted) on the first attempt. | Prompt: *“Ensure output includes range, mean, variance, standard deviation, plus at least one histogram and one scatterplot.”* |
| **Clarity & Coherence** | **5** | Well-structured with code blocks and explanations separated. Easy to read and runnable code. | Could add: *“Label each output section clearly: ‘Descriptive Statistics’, ‘Variance’, ‘Visualizations’.”* |
| **Depth & Originality** | **3** | Gave standard analysis only. Did not interpret results (e.g., noting which test scores were most variable). | Prompt: *“Provide brief interpretation of the results (e.g., which test had highest variance).”* |
| **Stylistic Appropriateness** | **4** | Professional and technical. Code was clean, but explanations were minimal. | Ask: *“Add 1–2 sentences summarizing findings in plain language for a non-technical reader.”* |
| **Ethical & Cultural Considerations** | **5** | Neutral and appropriate. No ethical/cultural concerns in dataset analysis. | Maintain neutral, professional tone in statistical reporting. |

**🏆 Overall Takeaways**

* **Strengths:** ChatGPT excelled in **accuracy**, **clarity**, and producing runnable code with stats + plots.
* **Weaknesses:** Lacked **depth** (no interpretation) and minor issues with **completeness** (sometimes omitted variance initially).
* **Next Steps:** Refine prompt to require (1) *all statistics explicitly*, (2) *interpretive summary*, and (3) *organized output sections*.

**📊 Combined Scorecard – ChatGPT/OpenAI**

| **Dimension** | **Exercise 1 (Correlation Code)** | **Exercise 2 (R Debugging)** | **Exercise 3 (Descriptive Stats + Viz)** |
| --- | --- | --- | --- |
| **Accuracy & Factuality** | **5** – All formulas/codes correct across Excel, R, Python, SAS, Stata. | **4** – Flagged most errors (missing library, summarise misuse, write.xlsx issue), but missed nuance with table(). | **5** – Produced correct descriptive stats + plots; results matched dataset. |
| **Completeness & Relevance** | **4** – Provided multi-language outputs but no definition/interpretation of correlation. | **4** – Gave fixes, but initial output lacked full corrected script. | **4** – Delivered stats + plots, but sometimes omitted variance. |
| **Clarity & Coherence** | **5** – Organized, code labeled by language. Easy to follow. | **5** – Clear bullet-pointed fixes and explanations. | **5** – Well-structured, code + explanations separated, runnable. |
| **Depth & Originality** | **3** – Only Pearson correlation shown; no alternatives (e.g., Spearman). | **3** – Fixed syntax but didn’t suggest alternative methods (e.g., count() vs summarise()). | **3** – Standard stats only; no interpretation of findings. |
| **Stylistic Appropriateness** | **4** – Technical and professional, but dry. | **4** – Clear but lacked inline code comments. | **4** – Professional but minimal narrative (no plain-language summary). |
| **Ethical & Cultural Considerations** | **5** – Neutral, technical only. | **5** – Neutral, technical only. | **5** – Neutral, appropriate. |

**🏆 Overall Observations Across Exercises**

* **Strengths:**
  + **Accuracy** and **Clarity**: ChatGPT consistently delivers correct, structured, and readable outputs.
  + **Reliability**: Code was almost always runnable with minor adjustments.
* **Weaknesses:**
  + **Depth**: Tends to give “bare minimum” technical answers without offering alternatives, interpretations, or deeper analysis.
  + **Completeness**: Sometimes omits requested details (e.g., variance, full corrected script).
  + **Stylistic**: Technically sound but not always polished for reports (e.g., missing plain-language summaries).
* **Improvement Strategy:**  
  Use prompts that explicitly require:
  + **All requested outputs** (stats, code, tables, plots).
  + **Explanations + interpretations** (not just numbers).
  + **Alternative methods or approaches** (to show depth).
  + **Polished style** (plain-language summaries + structured sections).

**\_\_\_**

**For exercise 2, here is a sample code you can start from with errors:**

**Sample Code in R (with errors)**

# Loading necessary libraries

library(readxl)

library(dplyr)

# Reading the Excel file

data <- read\_excel("path/to/your/excel\_file.xlsx")

# Inspecting the first few rows

head(data)

# Clean the data (remove rows with NA values)

cleaned\_data <- na.omit(data)

# Filter rows where the 'age' column is greater than 18

adults <- filter(cleaned\_data, age > 18)

# Calculate the mean of the 'income' column, assuming it's numeric

mean\_income <- mean(adults$income, na.rm = TRUE)

# Create a summary of the 'education' column

education\_summary <- summarise(adults, education\_summary = table(education))

# Check the distribution of the 'age' column

hist(adults$age)

# Save the cleaned data to a new Excel file

write.xlsx(adults, "path/to/cleaned\_data.xlsx")

**Answer Key**

The errors are:

* Error in filter() function
* Error in summarise() function
* Error in write.xlsx() function
* Potential issue with missing libraries

For exercise 3, here is some sample data not on data.gov you could also use:

**Sample Data:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student ID** | **Test 1** | **Test 2** | **Test 3** | **Test 4** |
| A | 94 | 99 | 68 | 70 |
| B | 99 | 93 | 61 | 99 |
| C | 79 | 76 | 76 | 89 |
| D | 69 | 84 | 98 | 87 |
| E | 86 | 91 | 75 | 94 |
| F | 78 | 96 | 63 | 94 |
| G | 63 | 73 | 86 | 67 |
| H | 64 | 89 | 65 | 62 |
| I | 63 | 89 | 88 | 66 |
| J | 82 | 78 | 69 | 94 |
| K | 96 | 75 | 63 | 70 |
| L | 88 | 94 | 97 | 67 |
| M | 80 | 63 | 100 | 88 |
| N | 96 | 73 | 83 | 64 |
| O | 95 | 91 | 67 | 87 |
| P | 95 | 70 | 72 | 63 |
| Q | 98 | 79 | 60 | 63 |
| R | 65 | 75 | 82 | 60 |
| S | 81 | 65 | 93 | 81 |
| T | 72 | 62 | 88 | 98 |
| U | 86 | 69 | 92 | 67 |
| V | 92 | 86 | 80 | 74 |
| W | 72 | 95 | 75 | 99 |
| X | 76 | 63 | 86 | 71 |
| Y | 89 | 70 | 64 | 87 |
| Z | 69 | 96 | 79 | 62 |
| AA | 81 | 93 | 60 | 96 |
| BB | 96 | 80 | 78 | 88 |
| CC | 93 | 62 | 89 | 87 |
| DD | 91 | 67 | 99 | 74 |