Final Work

Name of the course: Introduction to R

Name of the lecturer: Dr. Nitzan Shahar

Teaching Assistants: Ido Ben Artzi, Gili Katab

Date of submission: 19/3/2025

Name of submitter: Asaf Hoory

Project Link in GitHub:

https://github.com/AsafHoory/Final_Work_R#

אוניברסיטת תל אביב

ב ע ק ב ו ת   ה ל א   נ ו ד ע

Dataset[1]

The dataset I chose is from an experiment I conducted using PsychoPy in Professor Liad Modrik's lab, under the supervision of PhD student Maor Shreiber. In the experiment, we aimed to examine whether participants have the ability to process stimuli unconsciously. My dataset includes 20 participants who were sampled using convenience sampling. Participants 5 and 21 were excluded from the data analysis, so their data is not included. Due to the nature of the experiment, it is highly likely that there are identical rows in the code (some of which even appear consecutively). Therefore, addressing duplicates is not very meaningful in this case, and I have avoided doing so. The dataset contains many variables, most of which were removed during the preprocessing stage. In the end, I kept the most relevant data, which includes a limited number of columns:

- Participant number
- Task success/failure
- PAS grade
- Session number (out of the two sessions the participants attended in the lab)
- Contrast level of the stimulus presented.
- Contrast factor- I divided the contrast numeric values into four-factor levels (add later on in the descriptive part)

I chose this dataset because it is particularly close to my heart. It is the result of sixty hours spent conducting an experiment in the lab, in addition to many hours spent recruiting participants. Therefore, I am glad I had the opportunity to analyze it as part of my R programming studies.

Exploratory Analysis[2]

First, we calculated the mean and standard error (SE) of the PAS grade, accuracy, and contrast across all subjects and stored the results in the variable 'summary_stat'. We then presented these results in a table using the variable 'summary_table'. Additionally, we calculated the accuracy rate in the main task for each PAS level, stored it in the variable 'pas_summary', and presented it using Kable. The outputs are attached.

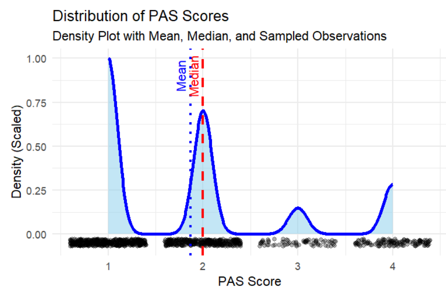
Variable	Mean	SE
PAS Score	1.865	0.009
WM Accuracy	0.571	0.004
Contrast	0.062	0.001

PASkey.keys	WM_Mean	WM_SE
1	0.510	0.006
2	0.616	0.007
3	0.695	0.015
4	0.609	0.011

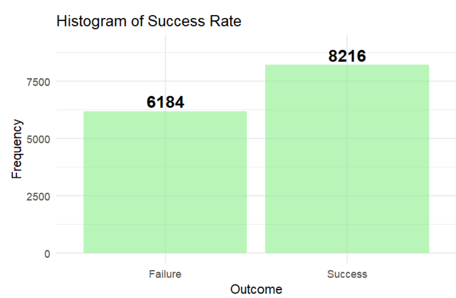
[1] Question A.1
[2] Question A.2

After doing so, we created a series of five exploratory plots using the ggplot package.

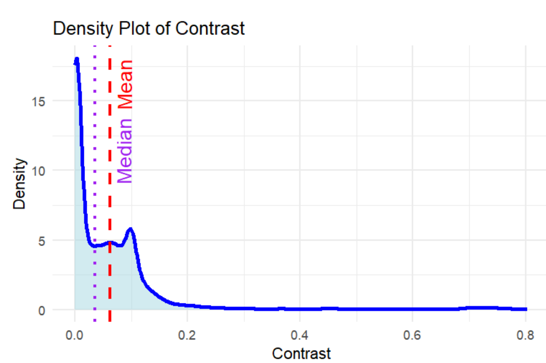
1. **<u>PAS plot:</u>** This density plot illustrates the distribution of 14,400 observations, including the mean and median. Unlike other parts of this plot that display results based on the entire sample, for aesthetic purposes, 1,000 observations were randomly sampled and visualized (the black dots) using *geom_jitter* to depict individual data points within the plot.



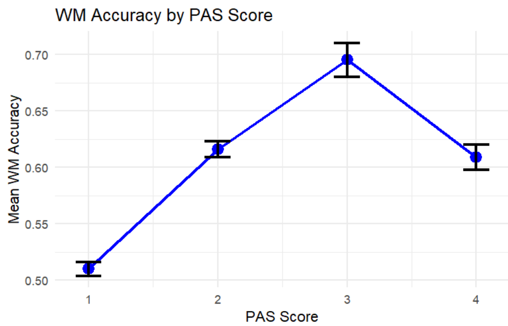
2. **<u>Succussion plot:</u>** A histogram plot shows the number of successes and failures in the main task.



3. **<u>Contrast plot:</u>** A density plot for contrast showing the mean and median.

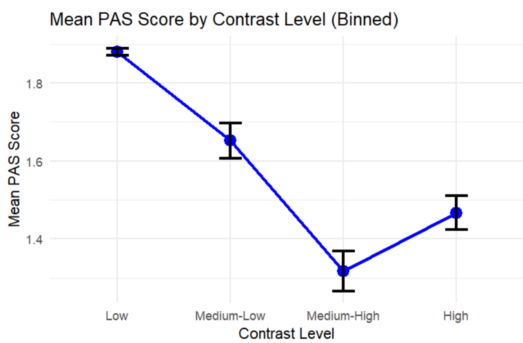


4. **<u>Succussion in each PAS plot:</u>** A plot illustrating the success rate and standard error (SE) at each PAS level. This plot is based on the variable *pas_summary*, which stores the mean and SE of accuracy for each PAS level.



5. **<u>PAS in every contrast level plot:</u>** In this plot, we have divided the contrast variable that was numeric, into four levels of factor:
    a. Low
    b. Medium-low
    c. Medium-high
    d. High

    The mean and SE of PAS are shown in each factor level of contrast.



Brief Academic Background and Research Questions:

[3]Although in recent years some studies have found evidence suggesting that working memory (WM) may operate on unconscious perceptual contents, demonstrations of unconscious WM are lacking. **Our study aims to investigate whether working memory can function independently of consciousness. Additionally, we seek to examine whether working memory is influenced by awareness or remains unaffected by it. Finally, by a multiple linear regression model, we aimed to determine whether the level of contrast, working memory, and the interaction between them is in relationship with stimulus awareness.**[4]

"We predicted that WM could operate independently of awareness and that the two would not be positively correlated. Additionally, we expected that greater stimulus contrast would lead to higher awareness. Finally, we did not expect a significant interaction effect between contrast and working memory in predicting PAS."

We measured awareness by Perceptual Awareness Scale (PAS) ((Ramsøy & Overgaard, 2004). This subjective measure asks the participant to indicate his awareness of prime stimulus on a scale between one to four, while a higher grade indicates higher awareness. measured WM, Our main task was under the influence of (Soto et al., 2011). In this task, subjects were exposed to a short Gabur stimulation (16.67 ms), and following a mask, they were asked to indicate if the second Gabur (target) tends to the right or the left compared to the short Gabur (prime). Above chance succussion in the main task indicates the accurate operation of WM. Additionally, we operated the contrast that a higher quantification indicates lower contrast.
We postulated an above-chance intercept in the logistic regression of performance while predicting performance from PAS, moreover, we hypothesized a null correlation between PAS and performance in the main task. Finally, we estimated that PAS would be negatively correlated with the contrast level (as mentioned earlier, we quantified contrast so that higher contrast quantification indicates a less contrasted prime).

Variable Defenition:[5]

Logistic Regression:
independent variable:
PAS report (1/2/3/4)
dependent variable:
performance in the main task (accurate/inaccurate)
Linear Regression:
independent variable:
1. contrast level (0 - 0.803).
2. performance in the main task (accurate/inaccurate)
dependent variable:
PAS report (1/2/3/4)

[3] Question A.3
[4] Research questions
[5] Question B.1

Pre-Processing[6]

The preprocessing part of the code loads 20 CSV files from a specified folder in our computer, combining them into a single data frame (df) while adding a file_name column to track the source of each row. It verifies that essential variables—contrast (stimulus contrast level), WMresp_acc (accuracy in the working memory task), PASkey.keys (participant's perceptual awareness score), session (experimental session number), and participant (participant number)—are present, stopping execution if any are missing. The code selects only these columns, removes rows containing NA values, and standardizes the participant column by converting it to numeric to ensure correct ordering when using arrange(). Additionally, it converts WMresp_acc, PASkey.keys, and contrast to numeric to facilitate statistical analysis, explicitly noting that PASkey.keys is temporarily numeric for mean and standard error calculations. After preprocessing, the dataset contains 14,400 observations — 720 trials per participant — and preserving naturally occurring duplicate rows that reflect the experimental trial structure rather than redundant data.

Functions:[7]

This code included two functions

1. The function calculates the mean PAS, success rate, and contrast for each participant after verifying that the required columns exist in the dataset. If any columns are missing, it returns an error message; otherwise, it returns a table with the rounded mean values. The first three lines from the output are attached.

```
 participant mean_PAS mean_success mean_contrast

1             1     1.49         0.756          0.061
2             2     1.58         0.581          0.11
3             3     1.57         0.525          0.058
```

2. Bonus question: The function calculates the standard error (SE) for PAS and contrast for each participant using the data.table package. It divides the standard deviation by the square root of the sample size for each participant and rounds the values to three decimal places. The function then renames the computed columns and returns a table with SE values for each participant. The first three lines from the output are attached.

```
participant se_PAS se_contrast

1:            1  0.026       0.003
2:            2  0.031       0.003
3:            3  0.026       0.002
```

[6] Question B.2. dyplr functions used in this part of the code: mutate, select, filter, group_by, and summarise.

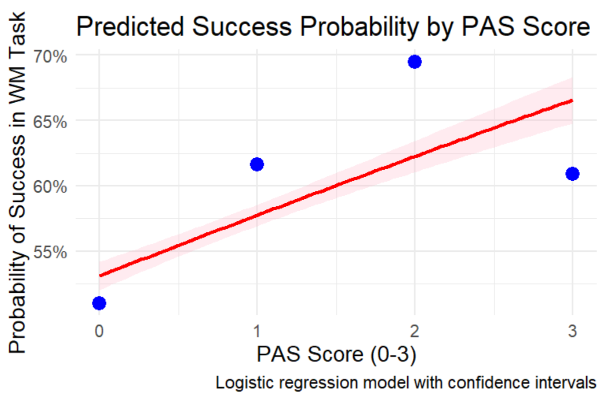
[7] Questions B.3, B.4

Regression: [8]

Logistic Regression [9]

$y = 0.12 + 0.19 * PAS$

We performed a frequentist logistic regression to predict task success from PAS score. The model included PAS as a predictor, with task success (1 = success, 0 = failure) as the dependent variable. The intercept ($\beta = 0.12$, 95% CI = [0.08, 0.18], $p < .001$) was significant, with an odds ratio of 1.13 (95% CI = [1.08, 1.18]). This represents the baseline odds of success when PAS = 1, meaning that even at the lowest PAS level, the probability of success remains above 50%. The results further showed that PAS had a significant positive effect on the probability of success ($\beta = 0.19$, 95% CI = [0.17, 0.25], $p < .001$). This suggests that higher PAS scores are associated with an increased likelihood of success. The odds ratio (OR) for PAS was 1.21 (95% CI = [1.17, 1.25]), indicating that for each additional unit in PAS, the odds of success increased by 21%. A Pearson correlation analysis between PAS and success revealed a weak but significant positive correlation ($r = 0.09$, $t = 11.25$, $p < 2.2 \times 10 ^ {-16}$), indicating that although the correlation is small, it is still statistically significant. Hence, our data supported the first hypothesis, demonstrating that participants showed above-chance success while reporting they were not aware of the stimulus. In contrast, our data did not support the second hypothesis, demonstrating a positive significant correlation between PAS and success in the main task.



Logistic Regression: Predicted Success Probability by PAS Score: The x-axis represents the PAS score (which ranged from 1 to 4 in the actual experiment), while the y-axis represents the probability of success in the main task. The blue point represents the mean performance for each PAS.

The logistic regression model's area under the curve (AUC) was 0.5598, indicating weak discriminative ability in predicting success in the main task based on PAS scores. The AIC value of 19,551.79 is relatively high, suggesting that the model may not be optimal and could potentially be improved with additional predictors or alternative modeling approaches. These

---

[8] Quesion C.1, C2, C3, and C4

[9] "We wanted b0 to represent the success rate when PAS = 1, so we subtracted 1 from all PAS values, so that PAS scores range from 0 to 3 instead of 1 to 4 in the figures."

results imply that while PAS has some predictive value, its overall contribution to predicting task success is limited. The ROC plot is visually shown below.

ROC plot: The curve shows the model's ability to distinguish success from failure, with the blue line representing the model performance and the red dashed line indicates the chance level.

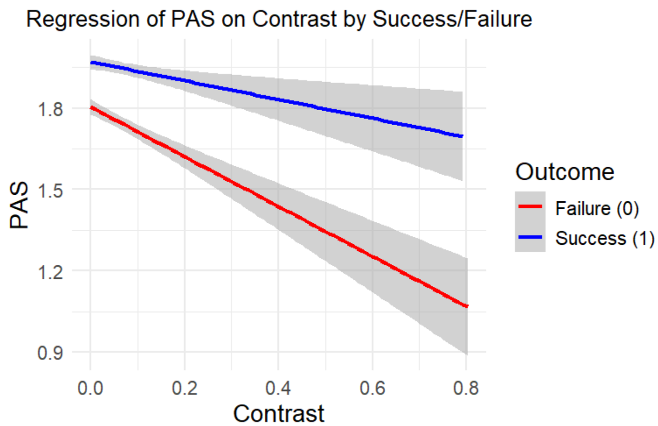
Multiple linear regression

$y = 0.8 - 0.92 * contrast + 0.16 * accuracy + 0.57 * contrast * Accuracy$

We performed a frequentist linear regression to predict PAS score from contrast and WM response accuracy (WMresp_acc), including an interaction term for contrast × WMresp_acc to test whether the relationship between contrast and PAS is moderated by WM response accuracy.[10] As mentioned before, contrast was quantified such that higher values indicate a less contrasted prime (i.e., the first Gabur). The results showed that for each one-unit increase in contrast, PAS increased by 0.92 on average (Estimate = -0.92, $t = -7.427$, $p < 1.18 \times 10^{-13}$). Similarly, for correct WM responses (WMresp_acc = 1) compared to incorrect responses (WMresp_acc = 0), PAS increased on average by 0.17 (Estimate = 0.17, $t = 8.324$, $p < 2 \times 10^{-16}$). The interaction term (contrast × WMresp_acc) had a positive effect (Estimate = 0.57, $t = 3.421$, $p = 0.000625$), indicating that the effect of contrast on PAS depended on WM response accuracy. Specifically, when the WM response was correct, the effect of contrast on PAS was weaker compared to when the response was incorrect.

These findings suggest that both contrast and WM response accuracy are significant predictors of PAS. As expected, higher contrast led to increased stimulus awareness, aligning with our hypothesis (as mentioned, in our quantification of contrast, lower numerical values indicate higher actual contrast of the prime stimulus.). However, contrary to our predictions, we found a positive relationship between WM response accuracy and PAS. Additionally, while we did not anticipate a significant interaction effect, the results indicate that contrast has a stronger influence on PAS when WM responses are incorrect. The AIC value of 41,391.94 suggests that the model does not fully capture the variance in PAS, indicating that additional factors may contribute to stimulus awareness. The results of the linear regression are visually presented below.

[10] In the actual experiment, our goal was not to predict PAS from performance in the main task. This relationship was examined solely to meet the requirement of performing multiple linear regression instead of simple linear regression.



Multiple linear regression: The X-axis represents contrast (lower values indicate higher prime contrast), and the Y-axis represents PAS. Blue and red lines show success and failure, with shaded area s representing standard error (SE).

Bibliography

1.

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, *3*(1), 1–23. https://doi.org/10.1023/B:PHEN.0000041900.30172.e8https://doi.org/10.1016/j.cub.2011.09.049

Soto, D., Mäntylä, T., & Silvanto, J. (2011). Working memory without consciousness. *Current Biology*, *21*(22), R912–R913. https://doi.org/10.1016/j.cub.2011.09.049