

# Take Home Exam

PSYC 2020 / 6022

Fall 2025

**Instructions:** This exam is **optional for PSYC 2020** and **required for PSYC 6022**. It will primarily focus on using R to analyze data, but you will need to synthesize information you have learned in both lecture and lab to complete it properly. If you complete this exam in PSYC 2020, it will replace your lowest exam grade (unless all of your other exam scores are higher than your score on this one).

You must complete this exam using RMarkdown. Your final submission should include both a .rmd file as well as a knitted word, PDF, or html file. You should be able to complete the entire exam using RMarkdown.

You may use any resources at your disposal to figure out how to get the answers, including lecture notes, lab materials, textbooks, and internet searches. **Be careful using ChatGPT for assistance.** It can help with certain coding problems, but it often makes mistakes, and you must demonstrate that you understand the procedures you are using and how to properly interpret them.

**You must complete this on your own, and the final product must be your work and your work alone.** You cannot seek help from classmates or other people besides course instructors (Dingjing Shi, Alina Tran, Jess Helmer). Any Ed Discussion questions about the content of this exam that are not strictly procedural (e.g., due date) should be asked privately to instructors only.

There are three sections each with 9 required questions. In total, each section is worth 32 points. That means the total score for the required questions is only worth 96 points.

**However,** each section has one additional four-point bonus question. These questions challenge you to go above and beyond what we have shown you in class. These questions are not required. But to get a score of 100, you must answer at least one bonus question. If you do all of them, your score can be as high as 108, in which case you will receive extra credit towards your final grade.

For all questions involving statistical significance, assume  $\alpha = .05$

# Part 1

## About the data

This section uses one of Mark Himmelstein’s datasets, based on a study that was conducted in the end of 2020. In this study, participants were recruited to forecast several different questions about economic and geopolitical events. An example of the types of questions forecasted is shown below.

One way of measuring forecasting accuracy is called the “Brier Score”. Brier Scores range from 0 to 2, and lower scores refer to more accurate forecasts. In our study, we had people make forecasts about several different questions on multiple occasions several weeks apart, and then looked at their mean Brier Scores.

We also wanted to determine if we could assess how well people reasoned about the laws of probability, which we called “coherence”. To do this, we used a scale that included items such as this (see next page)

In this example, a coherent response would correctly identify that the probability of an all-time low temperature in Chicago within the next five years would be at least as large as an all-time low temperature in Chicago in three years (since every time the latter is true, the former is also true), and the probability of an all-time low temperature within three years would be at least as large as within the next year.

We also collected several other variables related to people’s reasoning ability, which are described below (research papers describing each of these scales are included with the exam, but are optional reading). The variables in the dataset `forecasting_data.csv` are described below:

```
library(tidyverse)
info <- tibble(variable = c("ID", "brier1", "brier2", "bn_score", "crt_score", "ns_score", "cfs_score"),
               description = c("Participant identifier", "Average Brier Score (lower = more accurate) at Time 1",
                              "Average Brier Score (lower = more accurate) at Time 2",
                              "Berlin Numeracy Score (Measure of Numerical Reasoning, higher = better)",
                              "Cognitive Reflection Test Score (A different measure of reasoning, higher = better)",
                              "Number Series Score (Numerical Pattern Matching, higher = better)",
                              "Coherence Score (Ability to follow the logic of probability, higher = better)"))
knitr::kable(info)
```

variable	description
ID	Participant identifier
brier1	Average Brier Score (lower = more accurate) at Time 1
brier2	Average Brier Score (lower = more accurate) at Time 2
bn_score	Berlin Numeracy Score (Measure of Numerical Reasoning, higher = better)
crt_score	Cognitive Reflection Test Score (A different measure of reasoning, higher = better)
ns_score	Number Series Score (Numerical Pattern Matching, higher = better)
cfs_score	Coherence Score (Ability to follow the logic of probability, higher = better)

You can also view our study **pre-registration** on the [Open Science Foundation website](#) for more info, as well as two papers titled *The wisdom of many in few: Finding individuals who are as wise as the crowd* and *Measuring probabilistic coherence to identify superior forecasters* on to Canvas (also purely optional reading).

**Research Question 1:** We want to understand how related our new measure of coherence about probability (`cfs_score`) is to other variables that involve reasoning ability.

1. Create a scatterplot between `cfs_score` and `crt_score`. **(4 points)**
2. Use the `cor()` function to find the correlation coefficient between `cfs_score` and `crt_score` to confirm your answer to the previous question. **(3 points)**
3. Use the `cor.test()` function to test whether the correlation is significant. **(3 points)**

4. Generate a **correlation matrix** between `cfs_score`, `bn_score`, `crt_score`, and `ns_score`. What do you conclude? **(4 points)**
5. Conduct a simple linear regression using `crt_score` to predict `cfs_score` and show the summary table. **(3 points)**
6. Create two new variables: one representing `cfs_score` converted to  $z$ -scores, and another representing `crt_score` converted to  $z$ -scores. **(4 points)**

## Part 2

### About the data

This section uses the same dataset as Part 1

**Research Question 2:** Forecasts often tend to become easier as time passes. For example, forecasting the weather a week from today is relatively difficult. But if we wait six days, that same forecast will become about the weather tomorrow, and will be much easier. So we want to determine if people's forecasting accuracy at Time 2 is **different** than at Time 1.

1. If we want to determine whether the difference between people's Brier scores at Time 1 and Time 2 are significantly **different** from zero. State the null and alternative hypothesis
2. What statistical procedure would we use? Include the type of test you would conduct, and whether the test would be one or two-tailed **(3 points)**
3. Calculate the mean of `brier1`, the mean of `brier2`, and the difference between the means **(4 points)**
4. Create a new variable in the dataset that is the difference of each person's scores for `brier1` and `brier2`. Call it `brier_diff` **(3 points)**
5. Calculate the mean of `brier_diff` **(3 points)**
6. Calculate the standard deviation of `brier_diff` using the `sd()` function. If you did Question 5 this should validate your result **(3 points)**
7. Calculate the standard error of `brier_diff` **(4 points)**
8. Determine whether people's Brier scores at Time 1 and Time 2 are significantly different from one another. **(4 points)**