

# STAT 165/265 HW 5

Feb 14, 2024

**Due Tuesday, February 20 at 11:59pm**

## Deliberate Practice: Combining Forecasts

*Expected completion time: 90 minutes*

*Graded on completion*

For the following questions, look up and provide links to at least 3 data sources that could be good reference classes.

1. Will this year's NBA MVP be a guard?
2. Will OpenAI announce GPT-5 before the end of 2024?

For each reference class, comment on:

- How closely related it is to the question of interest
- Whether or not you expect it to systematically over/under-estimate the answer
- How much error the reference class's finite sample size induces

Use these considerations to assign relative weights to each reference class, and compute a weighted average. Some of the 3 data sources can be external predictions (e.g. on odds on a betting site), but at least one of the three should be an actual dataset. The dataset doesn't need to be large, but it should come from a reliable source.

It is OK if two data sources overlap, but the conditions for the same data source to be two reference classes should be different (e.g: the same Wikipedia table but different conditions for including a row in the reference class). You could also try using ChatGPT to generate reference classes and comment on how useful they are.

On Gradescope, please also submit the time it took to complete this exercise.

## Lab

*Expected completion time: 120 minutes*

*Graded on accuracy*

[Link to Jupyter notebook.](#)

Please follow the instructions in the notebook to print out your code and answers and submit to Gradescope. You may use languages other than Python, although we will generally be providing starter code in Python.

On Gradescope, please also submit the time it took to complete this exercise.

## Predictions

*Expected completion time: 60 minutes*

*Graded on accuracy as part of the class forecasting competition*

Make and submit predictions to the questions on this Google Form:

<https://forms.gle/QUgtmKSYE8dgXV3z7>.

Be sure to follow the format described at the top of the form. For each question, you will submit a mean and inclusive 80% confidence interval or a probability (whichever the question asks for). We provide cells on the Google form for you to type out your reasoning (1-2 paragraphs), which you should submit to Gradescope with the rest of this assignment. For questions 1-3, your prediction (but not the explanation) will appear on the public leaderboard.

**Remember to submit your reasoning for each question to Gradescope.**

## [STAT 265 only] Combining Forecasts

*Expected completion time: 90 minutes*

*Graded on accuracy*

Recommended reading: [Chapter 10.1 of Prof. Steinhardt's \*Forecasting\*](#)

We will borrow the setup and notation from the textbook chapter linked above. Suppose we want to forecast a binary event whose true probability of occurring is  $\mu$ . Consider a set of signals  $X_1, X_2, \dots, X_n$  related to a binary event, where each signal  $X_i$  is a probability estimate for the event's occurrence. Suppose we have that the signals are unbiased ( $\mathbb{E}[X_i] = \mu$ ) and each has some variance  $\text{Var}[X_i] = \sigma_i^2$ .

Define the expected Brier score as  $\text{EBS}(X) = \mathbb{E}[(X - \mu)^2 + \mu(1 - \mu)]$ .

1. Show  $\mu(X - 1)^2 + (1 - \mu)X^2 = (X - \mu)^2 + \mu(1 - \mu)$ .
2. Find the average of the expected Brier scores of each of the signals.
3. Suppose instead of taking any particular signal as our forecast, we take the average of the signals as our forecast. Formally, our forecast is  $X = \frac{1}{n} \sum_{i=1}^n X_i$ . Find the expected Brier score of this forecast, in terms of  $n, \sigma_i$ , and covariance terms between the signals.
4. When is the expected Brier score of the average of the signals better than the average expected Brier score of the signals? Give a concrete example of when the score resulting from averaging the signals performs better than the average of the scores of the signals. Note that under the definition used in this problem, “better” Brier scores are ones that are lower (see [chapter 3.1](#) for details on this).

**Hint:** Consider the covariance between the signals.