STAT 165/265 HW 2

Jan 24, 2024

Submit to Gradescope by Tuesday, January 30 at 11:59pm

Deliberate Practice: Calibration

Expected completion time: 20 minutes

Go to https://www.openphilanthropy.org/calibration and complete 20 questions under the Confidence Intervals (level 60%) category.

At the end, go to the "Results" tab and take screenshots showing that you completed at least 20 questions and your calibration performance (the chart that appears below the results).

Deliberate Practice: Estimation

Expected completion time: 60 minutes

For each of the following questions, estimate the answer without looking things up, then look up the answer and calculate the relative error of your estimate[1]. For each quantity, we provided one link with a reasonable-seeming answer, which you should use as the "official" answer when you calculate the error. We recommend spending around 5 minutes on each estimation question.

In addition to the questions below, devise two estimation questions of your own, for quantities you are interested in. Again, estimate the answer without looking things up, then look up a reasonable answer and compute your relative error.

1. How many cattle are in the world as of April 2023?

2. What's the length of the SF Bay Bridge, in meters?

3. How heavy, in tons, is the Titanic?

---

[1]Note that some of these questions still leave some ambiguity depending on specific definitions, and different Google search results might give slightly different answers!

4. How many paid subscribers of Netflix are there as of Q3 2023?

5. How many statistics and biostatistics degrees (Bachelors) were given in the U.S. in 2021?

6. What was the GDP of Mexico in 2022?

7. How many musicians are employed in the U.S.?

8. What's the depth of the deepest hole, in meters, that humans have dug?

Here are our links that contain answers:

1. https://www.statista.com/statistics/263979/global-cattle-population-since-1990/

2. https://en.wikipedia.org/wiki/San_Francisco%E2%80%93Oakland_Bay_Bridge

3. https://en.wikipedia.org/wiki/Titanic

4. https://www.demandsage.com/netflix-subscribers/

5. https://magazine.amstat.org/blog/2022/12/01/statsbiostatsdegree/

6. https://countryeconomy.com/gdp/mexico

7. https://www.zippia.com/musician-jobs/demographics/

8. https://en.wikipedia.org/wiki/Kola_Superdeep_Borehole

On Gradescope, for each of the **10 questions (8 provided plus 2 created by you)**, please submit your estimate, your relative error, and a brief explanation of your reasoning (e.g. what quantities you used as part of the decomposition and how you combined them). Your explanation can be brief and does not have to be complete sentences. **The purpose of these questions is to get practice making estimates, rather than to get all the answers right. You will not be graded on accuracy, but instead on whether you made a reasonable attempt and learned from your mistakes.**

Lab

Expected completion time: 90 minutes

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.

Predictions

Expected completion time: 60 minutes

Make and submit predictions to the questions on this Google Form:
https://forms.gle/7KimRTxXtZXvSdAT8.

Be sure to follow the format described on the front page of the form. For each question, you will submit a mean and inclusive 80% confidence interval (or a probability for question 3) as well as an explanation of your reasoning (1-2 paragraphs). For questions 1-3, your prediction (but not the explanation) will appear on the public leaderboard. Question 0 will remain private and not count towards the leaderboard.

Please include a copy of your Google Form responses with your Gradescope submission. Ideally, just include legible screenshots of your email confirmation receipt showing your predictions and reasoning.

[STAT 265 only] Strictly Proper Scoring Rules

Expected completion time: 90 minutes

This question is optional for STAT 165 students. You should tag pages for this question if and only if you are enrolled in STAT 265.

1. Here we will explore a scoring rule that is neither Brier nor log-loss.

$$S(\mathbf{p}, i) = \frac{p_i}{\|\mathbf{p}\|_2} = \frac{p_i}{\sqrt{p_1^2 + \cdots + p_c^2}}$$

   where $\mathbf{p} = \begin{bmatrix} p_1 & \cdots & p_c \end{bmatrix}^\top$ is a probability vector representing the forecast probabilities, and i is the index of the actual outcome in a categorical setting with c possible outcomes.

   (a) Demonstrate that the scoring rule $S(\mathbf{p}, i)$ as defined above is proper relative to the set of all probability distributions over c outcomes. You are not required to do the second derivative test here.
   Hint: Show that for any true probability distribution Q, the expected score under Q is maximized when the forecasted probability distribution $\mathbf{p}$ is equal to Q.

   (b) True or False: Is the scoring rule above strictly proper? No proof is required for this part. Recall that a scoring rule is strictly proper if the equality $\mathbf{S}(Q, Q) = \mathbf{S}(\mathbf{p}, Q)$ holds if and only if $\mathbf{p} = Q$.
   Note that $\mathbf{S}(Q, Q)$ represents the vector $[S(Q, 1), S(Q, 2), \ldots, S(Q, c)]$, where each component $S(Q, j)$ is the score obtained when the actual outcome is j and the forecasted distribution is Q.

2. Now we will examine symmetric vs asymmetric distributions:

   (a) Let f be any strictly convex function. Define a scoring rule:

   $$F(p, i) = f(i) - f(p) - f'(p) \cdot (i - p).$$

   Prove that F is strictly proper.

   (b) Show that $f(x) = x^2$ gives the Brier score, while $f(x) = x \cdot \log(x)$ gives the log loss.

   (c) Consider $f(x) = \alpha x^3$ for some $\alpha > 0$. Discuss how this scoring rule behaves differently from symmetric scoring rules like Brier score or log loss, especially in contexts where certain outcomes (like '1') are more critical or have higher stakes compared to others (like '0').