

STAT 165/265 HW 3

January 31, 2024

Submit to Gradescope by Tuesday, February 6 at 11:59pm

Deliberate Practice: Zeroth and First Order Forecasting

Expected completion time: 60 minutes

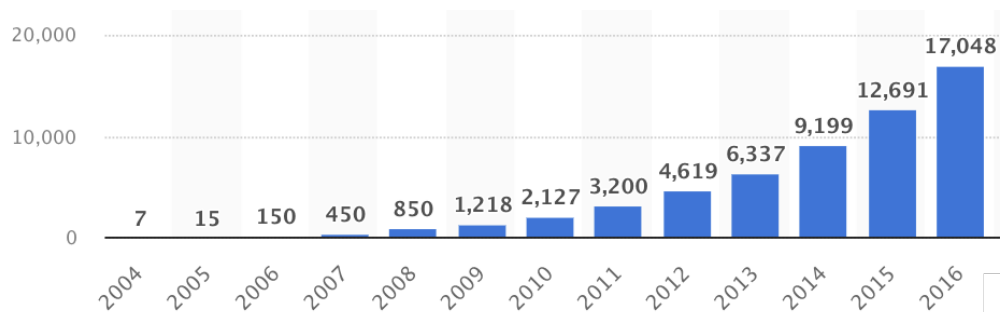
Graded on completion & using the correct relative error formula

For each of the following time series, estimate the value of the next few points without looking things up, then look at your answers and report your relative errors. Try to be accurate, even if you don't use a zeroth/first order forecast for each question. Just write down your answer and relative error for each question (no need to show work). When you have answered all the questions, write 1-2 paragraphs summarizing which strategies worked well and which strategies worked poorly.

1. United States energy usage

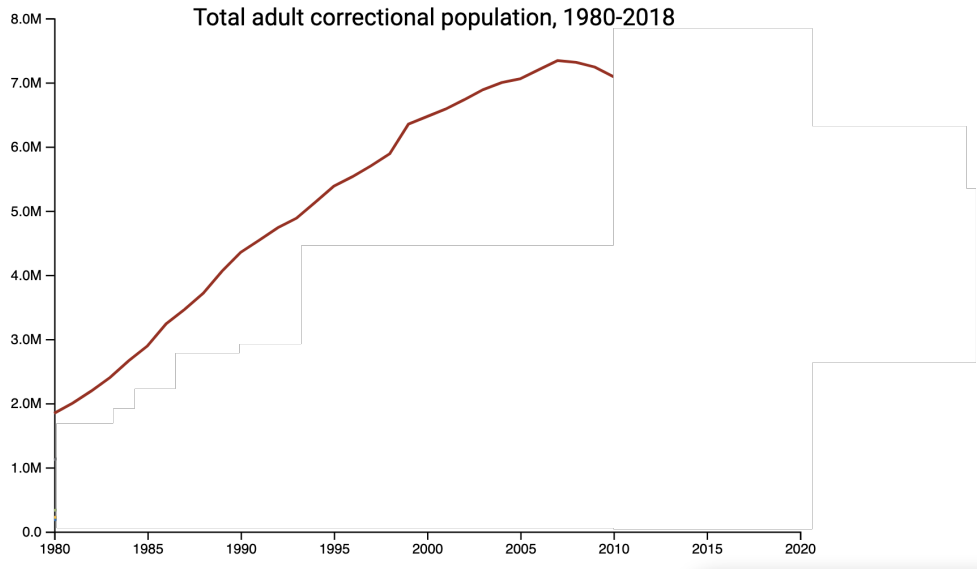
- Find the graph by going to [this link](#). In “Chart options” and “Date Range Type” select “Custom”, in “Date Range” enter “01/01/2024 - 01/21/2024”, and switch the time zone to Pacific.
- Predict what the energy usage was on 01/26/2024 at 4pm Pacific time.
- After you've made your prediction, extend the date range to find the true answer.

2. Number of full-time Facebook/Meta employees until 2016



- Predict the number of full-time employees in 2018 and 2020.
- After you've made your prediction, check your answers at [this link](#). Report your relative error for each of the years requested.

3. Total adult correctional population until 2010

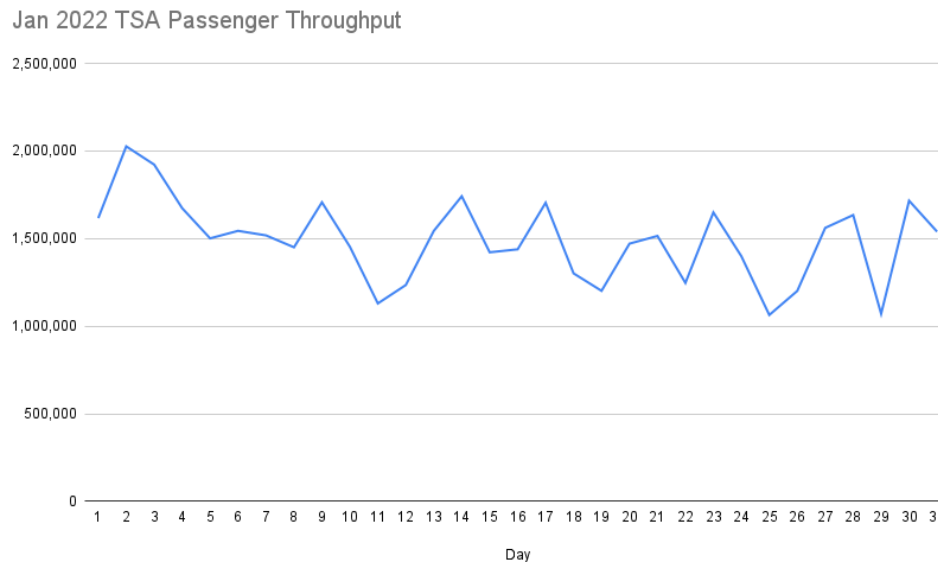


- Predict the adult correctional population in 2014 and 2018.
- After you've made your prediction, check your answers at [this link](#). Report your relative error for each of the years requested.

4. World GDP until 1950

- Find the graph by going to [this link](#).
- Predict World GDP in 1960, 1970, and 1990.
- After you've made your predictions, move the slider at the bottom to reveal the correct answers. Report your relative error for each of the years requested.

5. TSA passenger throughput



- This graph shows TSA checkpoint throughput for January 2022.
- Predict the passenger throughput number for Feb 7, 2022.
- Check your answers at [this link](#).

6. Per capita CO2 emissions in the US

- Find the graph by going to [this link](#).
- Predict the US per capita CO2 emissions in 2015, 2025, and 2030.
- Extend the slider to check your answer for 2015. No need to provide relative error for 2025 and 2030 since they are in the future!

7. US military expenditure as a percentage of GDP until 2012

- Find the graph by going to [this link](#).
- Predict US military expenditure as a percentage of GDP in 2015, 2018, and 2019.
- After you've made your predictions, move the slider at the bottom to reveal the correct answers. Report your relative error for each of the years requested.

8. Worldwide total fertility rate until 1978

- Find the graph by going to [this link](#).
- Predict the worldwide total fertility rate in 1988, 1998, and 2003.
- After you've made your predictions, move the slider at the bottom to reveal the correct answers. Report your relative error for each of the years requested.

9. US Child mortality rate until 1955

- Find the graph by going to [this link](#).
- Predict the child mortality rate in 1960, 1990, and 2010.

- After you've made your predictions, move the slider at the bottom to reveal the correct answers. Report your relative error for each of the years requested.

On Gradescope, for each of the **9 questions**, submit your estimates and relative errors. Then, include 1-2 paragraphs of reasoning on which strategies worked well and which strategies didn't. Please also submit the time it took to complete this exercise.

Deliberate Practice: Estimation and Calibration

Expected completion time: 45 minutes

Graded on completion

For each of the following questions, estimate an inclusive 80% confidence interval for the answer without looking things up. For each quantity, we provided one link with a reasonable-seeming answer. We recommend spending around 5-10 minutes on each estimation question. Write down your reasoning for one of the questions.

1. How many employment-based Green Card applications did US Citizenship and Immigration Services process in 2021? Note that the question is about the number of applications *processed*, not *approved*. [Link to Answer](#)
2. How many acres of US land were burnt by wildfires in 2022? [Link to Answer](#)
3. How many records have the Beatles sold, in terms of worldwide certified (not claimed) sales, as reported by Wikipedia? Records include singles and full-length albums. [Link to Answer](#)
4. How many metric tons of crude oil were produced in North America in 2022? [Link to Answer \(scroll to bottom\)](#)

On Gradescope, submit your estimate for each of the **4 questions**, with your reasoning for one of the questions. Please also submit the time it took to complete this exercise.

Deliberate Practice: Reference Classes

Expected completion time: 60 minutes

Graded on completion

For the following questions, describe 3 reference classes you would use to answer it. To do this, you can look up information about the reference classes, but not the answer itself. For at least 1 and at most 2 of the 4 questions below, use a large language model (such as ChatGPT) to help you brainstorm ideas for reference classes. You can find links to different large language models below:

- [ChatGPT](#)

- [Other GPT models](#) (for example, you can try `text-davinci-003` or `davinci`)
- [GPT-J 6B and GPT-NeoX 20B](#) (smaller, open-source models)

Once you have come up with reference classes for all of the questions, look up the answers, and discuss which types of reference classes tended to work well, which worked poorly, and modifications you would make in the future.

1. How much did the movie *Harry Potter and the Deathly Hallows - Part 2* gross worldwide? [Link to Answer](#)
2. A diplomatic boycott of a competition is when a country does not send high-ranking officials to attend the competition as official representatives, but still sends athletes. How many countries issued a diplomatic boycott of the 2022 Winter Olympics in China by January 31? [Link to Answer](#)
3. What was Delta Air Lines' [operating revenue](#) in 2020? [Link to Answer](#)
4. How many years were there between the beginning and end of the construction of the [Sydney Opera House](#)? [Link to Answer](#)

On Gradescope, submit your reference classes for each of the **4 questions**. Then, also include 1-2 paragraphs total of discussion on which kinds of reference classes worked and which didn't. 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/vGDhi2Gy4TDqh98i6>.

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 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.

Submit your reasoning for each question to Gradescope.

[STAT 265 only] Succession Events

Expected completion time: 90 minutes

Graded on accuracy

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.

Suppose we are forecasting whether a particular event A will occur within the next 12 months. Each month, the event occurs with probability p (does not occur with probability $1 - p$).

Let x_i denote whether event A happened in month i . We have that $x_i \sim \text{Bernoulli}(p)$, where p is a random variable. After each month, we update our belief about p based on the information from the previous month(s).

Recall the density of the Beta distribution:

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

where $\Gamma(\cdot)$ denotes the [Gamma function](#). Recall that $\Gamma(n) = (n-1)!$ for integer $n > 0$.

1. Suppose that *a priori*, we have no information as to whether event A is likely, so we take a $p \sim \text{Uniform}(0, 1)$ prior.
 - (a) Suppose we observe that event A does not occur in any of the first r months. We now update our belief about p . Show that the posterior $p \mid x_1 = 0, x_2 = 0, \dots, x_r = 0 \sim \text{Beta}(1, r + 1)$.
 - (b) Show that $\mathbb{P}(x_r = 0 \mid x_1 = 0, \dots, x_{r-1} = 0) = \frac{r}{r+1}$
 - (c) Use what you found above to find the probability that the event does not happen in the next k months (i.e. find $\mathbb{P}(x_1 = 0, \dots, x_k = 0)$). (Hint: x_i and x_j are *dependent*)
2. Suppose instead that *a priori* we *do* have some information about how likely event A is to occur each month, which we encode into our prior $p \sim \text{Beta}(\alpha, \beta)$.
 - (a) Suppose we observe that event A does not occur in any of the first r months. We now update our belief about p . Show that the posterior $p \mid x_1 = 0, x_2 = 0, \dots, x_r = 0 \sim \text{Beta}(\alpha, r + \beta)$.
 - (b) Find $\mathbb{P}(x_r = 0 \mid x_1 = 0, \dots, x_{r-1} = 0)$ in terms of r, α, β .
 - (c) Set $\alpha = 1, \beta = 2$, and use what you found above to find the probability that the event does not happen in the next 12 months (i.e. find $\mathbb{P}(x_1 = 0, \dots, x_{12} = 0)$). (Hint: x_i and x_j are *dependent*)