STAT 165/265: Forecasting

Discussion 9: Independence Assumptions

4/4/2025

You will work through this sheet in class on Friday, 4/4 in groups of 2-3. We will explore independence assumptions and how forecasts change when independence assumptions are removed.

Please submit the Discussion sheet for participation points on Gradescope under the "Discussion 9" assignment. Submit what you have at the end of the hour – it's ok if you didn't finish every question thoroughly!

For students who have contacted us about absences this Friday, please spend an hour on this sheet and submit to the same Gradescope assignment.

Part I. Earthquakes

For this problem, we will consider earthquakes only of at least 6.0 magnitude for the region of California or the surrounding area (defined as the region of 31.0 to 43.0 degrees North and -126.0 to -114.0 degrees West).

Suppose the historical monthly base rate of an earthquake occurring in this region was **56/936 ≈ 5.98%** from the years 1933 (inclusive) through 2010 (inclusive).

1.1. Assuming independence of earthquakes across months, calculate the probability of at least one earthquake occurring in this region over a year.
1 - P(no earthquake in a year) = 1 - P(no earthquake in a month)^12 = 1 - (1-P(earthquake in a month))^12 = 1-(1-0.0598)^12 = **52.29%**

1.2 Removing the independence assumption, what do *you* think is the probability of at least one earthquake occurring in this region over a year. In other words, how many years from 1933 to 2010 do *you* think had at least one earthquake? How does this relate to the answer from 1.1?
This is subjective, but consider how earthquakes may be correlated and how the absence of an earthquake in this region in a particular month might affect the presence in the following month(s).

True Probability: 31 unique years / 78 total years -> **39.74%**

Part II. Voting Patterns

For this problem, we will consider the proportion of U.S. states voting for the Republican Party in U.S. presidential elections since 1856.

Historically, the probability that a randomly chosen state votes for the Republican Party is **≈54.73%**.

2.1. Assuming independence, calculate the probability that at least 40/50 states vote for the Republican Party in a given election. Hint: Use a Gaussian approximation.

Let R denote the number of states voting Republican in a given election. R ~ Bin(n=50, p=0.5473). R can be approximated with a Normal(mean=27.365, var=12.388). P(R > 40) = **0.02%.**

2.2. Removing the independence assumption, what do *you* think is the probability of at least 80% of states voting for the Republican party in a given U.S. presidential election? In other words, how many elections from 1856 to 2020 do *you* think had at least 80% of states voting for the Republican Party? How does this relate to the answer from 2.1?
This is subjective, but consider how states in certain regions may be correlated and vote similarly and how this may affect the true probability.

True probability: **19.05%**

Part III. Disease Outbreaks

For this problem, we will consider "Disease outbreaks in the United States," according to Wikipedia.

Suppose the historical annual base rate of disease outbreak in the U.S. was **50/286, or 17.48%** from the years 1738 (inclusive) through 2023 (inclusive).

3.1. Assuming independence of disease outbreaks across years, calculate the probability of at least one disease outbreak occurring in this region over a decade.
1 - P(no outbreak in a decade) = 1 - P(no outbreak in a year)^10 = 1 - (1-P(outbreak in a year))^10 = 1-(1-0.1748)^10 = **85.35%**

3.2 Removing the independence assumption, what do *you* think is the probability of at least one disease outbreak occurring in this region over a decade. In other words, how many decades from 1738 to 2024 do *you* think had at least one disease outbreak? How does this relate to the answer from 3.1?
This is subjective, but consider how outbreaks may be correlated and how the absence or presence of an outbreak in the U.S. in a particular month might affect the presence of a new one in the following month(s). Some considerations include immunity, advances in public health systems/responses, etc. Perhaps even more significant is that records of disease outbreaks may be incomplete / missing from less recent time periods.
True probability: **60.00%**