Module 9 – Forecasting

Background

- In this module, you will check other students' forecasts to see how good they were at predicting the future. A group of students assigned probabilities for 10 "binary" events. (Binary events have two possible outcomes: they occur or they do not occur.) They recorded their probabilities **prior** to the events occurring. Now that the time of the events has passed, we can analyze how accurate the students were with their forecasts. The Excel workbook for this module consists of 3 tabs.

    1. The "Snowfall" tab contains hypothetical raw data of forecasts made to a question regarding the probability of snowfall at a future date.
    2. The "Responses" tab contains the raw data of forecasts made to 10 events. Each row contains the forecasts for a single student
    3. The "Occurrence" tab contains a list of the events forecasted into which the outcome of 0 = "didn't occur" or 1 = "occurred" has been recorded.

Part 0: Basics of probability and summation notation

Although in this module you will be relying primarily on subjective probabilities, it is important to recall the basic concepts of computing probabilities. Additionally, when summing a long list of values, the long-hand notation for the sum can be impractical or even impossible to display. In mathematics, there is a compact notation to represent a sum of an indexed list of values.

1) To begin this Module, view the following screencasts:
   Probability Introduction
   Summation Notation

Part 1: Consistency

In making predictions about the likelihood of a future event, it is important to make sure those probabilities are consistent. Consider the following list of questions:

i. What is the probability that tomorrow we will get 1 inch or more of snowfall?
ii. What is the probability that tomorrow we will get 3 inches or more of snowfall?
iii. What is the probability that tomorrow we will get 5 inches or more of snowfall?
iv. What is the probability that it will snow less than 1 inch or not at all tomorrow?

1) Please consider the relationships among the probabilities assigned to the questions above.
    a. True or False: The probability assigned to question i) has to be greater than or equal to the probability assigned to question ii). Explain.

    b. True or False: The probability assigned to question ii) has to be greater than or equal to the probability assigned to question iii) if the same time range is used for both (e.g., tomorrow or some other range. Explain.

    c. What can we say about the relationship between the probability assigned to i) and iv) if the forecasts made are consistent?

2) Use the data on the "Snowfall" tab to answer the following questions:
    a. How many individuals had consistent forecasts when just looking at questions i) and ii)?

    b. How many individuals had consistent forecasts when just looking at questions i) and iv)?

c. How many individuals had consistent forecasts for all questions?

Part 2: Looking at Consistency of Each Person's Forecasts

In this part, you are going to look at some forecasts of stock prices and other things that were made prior to the event actually occurring. This is actual data related to forecasts made regarding these events. The first 5 events relate to the percent change in trading price of Amazon stock and the actual occurrences are listed on the "occurrence" tab. Here is a list of the events and whether or not they occurred:

Event 1	What is the probability that AMZN stock will close 5%, or more, lower (that is, less than or equal to -5%) on the settlement date compared to the first trading day of the current calendar year?	0
Event 2	What is the probability that AMZN stock will close 1%, or more, lower (that is, less than or equal to -1%) on the settlement date compared to the first trading day of the current calendar year?	0
Event 3	What is the probability that AMZN stock will close 5%, or more, higher (that is, greater than or equal to 5%) on the settlement date compared to the first trading day of the current calendar year?	1
Event 4	What is the probability that AMZN stock will close 1%, or more, higher (that is, greater than or equal to 1%) on the settlement date compared to the first trading day of the current calendar year?	1
Event 5	What is the probability that AMZN stock will close between 5% lower and 5% higher (that is, between -5% and 5%, not inclusive) on the settlement date compared to the first trading day of the current calendar year?	0
Event 6	What is the probability of snow in Boulder on the settlement date (as recorded by the National Weather Service www.weather.gov)?	0
Event 7	What is the probability that Allstate Corp (stock ticker ALL) will be part of the S&P 500 on the settlement date?	1
Event 8	What is the probability that gold (GC=F) will be worth more than $1150 per ounce on the settlement date? Finance.yahoo.com, "last price."	1

Event 9	What is the probability that on the settlement date, you can get 0.90 or more EUR for 1 USD, according the Currency Converter at www.oanda.com?	0
Event 10	What is the probability that light sweet crude oil(CLH18.NYM) will be trading at more than $47 per barrel on the settlement date (finance.yahoo.com)?	1

Events 1-5 are related to each other because they are all about AMZN stock.

1) Of the people who made forecasts, how many had a probability for Event 4 that were greater than or equal to their probability for Event 3?

2) Answer these questions:
    a. TRUE or FALSE? For each person making a forecast, the probability (forecast) for Event 2 should be greater than or equal to the probability for Event 1.

    b. TRUE or FALSE? For each person making a forecast, the probability for Event 1 should be equal to the probability for Event 3.

    c. TRUE or FALSE? For each person making a forecast, the probability for Event 1 *plus* the probability for Event 3 *plus* the probability for Event 5 should equal 1. (For this question, assume that AMZN is still being traded on the settlement date.)

    d. What other pairs of events do you need to examine for consistency?

Part 3: How Did They Do?

We can compare individual forecasting performance by looking at the **total squared difference** for each person making a forecast. (Recall how you use the total squared difference in the OJ module.) In this context, the *difference* for that person for a given event is the probability assigned for that event minus the outcome of that event. The probability assigned is a number between 0 and 1. The outcome is either a 1 or 0, meaning the event occurred or did not occur, respectively. The *squared difference* is the square of that difference. The *total squared difference* is the sum of the ten squared differences (one for each of the ten events).

1) Total Squared Difference

   a. Write a mathematical expression for the total squared difference. To do that, use this notation:
      - Denote the probability assigned by person i for event j as $x_{i,j}$. For example, the probability for Event 8 by the person with ResponseID 103 would be written $x_{103,8}$.
      - Denote the outcome of event j as y_j. For example, if the outcome of event three was "did occur," then we represent it with a 1 as $y_3 = 1$.
        Using that notation, write the expression for the total squared difference T_i for person i.

Note: Response ID 1 has a total squared difference of 1.6016

   b. What is the lowest possible value that one of these total squared differences could take on?

   c. What is the highest possible value?

   d. What percentage of the events did occur? What value would the total squared difference be for a forecaster who assigned this percentage as their probability forecast for all the events?

   e. Who (by Response ID) has the lowest total squared difference?
      What is that lowest total squared difference?

   f. Why do you think you are asked to square the differences before you add them up (rather than just adding up the differences)?

Another way to examine performance is to round the probabilities assigned and to see how well the rounded values match the outcomes.

2) Rounding Probabilities
   Round all the probabilities in the data set to 0 or 1. Use the usual rounding rule: round probabilities of 0.5 and higher to 1; round probabilities of under 0.5 to 0.
   a. After the rounding, how many probabilities are 0?

   b. For the probabilities that were rounded down to 0, what percentage of those forecasts had an outcome of 1 (did occur)?

   c. After the rounding, how many probabilities are 1?

   d. For the probabilities that were rounded up to 1, what percentage of those forecasts had an outcome of 1 (did occur)?

   e. If we were perfect forecasters, what would the answers to questions b and d be?

   f. How many people were perfect forecasters based on their rounded probabilities? That is, once the probabilities are rounded, how many people predicted all ten events?

Selected Answers

Part 1

1) a) True: The probability assigned to i) has to be greater than or equal to the probability assigned to ii) because if it snows 3 inches or more, it also snowed 1 inch or more.
b) True: The probability assigned to ii) has to be greater than or equal to the probability assigned to iii) because if it snows 5 inches or more, it also snowed 3 inches or more.
c) They should sum to 100%.

2) a) 10 b) 7, c)5

Part 2

1) 702 (out of 851 people who made forecasts)

2a) TRUE. If the stock is down 5% or more, it must also be down 1% or more. So, Event 1 is a subset of Event 2.
2b) FALSE.
2c) TRUE. Assuming that the stock was still being traded on the settlement date, there is no option remaining for the percent change in the close cost.

Part 3

1a)

$$T_i = \sum_{j=1}^{10} \left(x_{i,j} - y_j\right)^2$$

1b) 0 is the lowest possible value for total squared difference. The total squared difference would be 0 if someone only assigned probabilities of 0 or 1 for each event and was perfectly correct.
1c) 10 is the highest possible value for the total squared difference. The total squared difference would be 10 if someone only assigned probabilities of 0 or 1 for each event and was incorrect for every event.
1e) Multiple Response IDs have the lowest total squared difference of 0.

2a) 4451 of the rounded probabilities are 0.
2b) 37.81%
2c) 4059 of the rounded probabilities are 1.
2d) 63.37%
2e) If we were perfect forecasters, then for b) 0% of the forecasts rounded to 0 would have occurred and for d) 100% of the forecasts rounded to 1 would have occurred.
2f) There were 23 perfect forecasters based on their rounded probabilities. (Response ID: Multiple Response IDs)

The creation of this work "Mathematical Analysis in Business" was supported by Open CU Boulder 2020-2021, a grant funded by the Colorado Department of Higher Education with additional support from the CU Office of the President, CU Office of Academic Affairs, CU Boulder Office of the Provost, and CU Boulder University Libraries.