Forecasting Revisited

Earlier in the semester, you made forecasts using probabilities. You assigned probabilities to "binary" events. Binary events have two possible outcomes: they occur or they do not occur. In that exercise, we set the **settlement date** to **November 14**. Now that date is here, so we can revisit your forecasts and compare them to what happened.

The twelve events that you stated probabilities for are as follows.

1HOG5D	HOG stock closed 5%, or more, lower on the settlement date compared to the first trading day of the current calendar year.	
2HOG1D	HOG stock closed 1%, or more, lower on the settlement date compared to the first trading day of the current calendar year.	
3HOG5U	HOG stock closed 5%, or more, higher on the settlement date compared to the first trading day of the current calendar year.	
4HOG1U	HOG stock closed 1%, or more, higher on the settlement date compared to the first trading day of the current calendar year.	
5HOG5M	HOG stock closed more than 5% lower and less than 5% higher on the settlement date compared to the first trading day of the current calendar year.	
6YAHOO	The CEO of Yahoo! is different on the settlement date from who it was on the first day of the calendar year.	
7ALLSP	Allstate Corp (stock ticker ALL) is part of the S&P 500 on the settlement date.	
8TWTNB	Taylor Swift (@taylorswift13) has fewer Twitter followers than Barack Obama (@BarackObama) on the settlement date, according to https://twitter.com/taylorswift13 and https://twitter.com/barackobama.	
9USDEUR	On the settlement date, you can get 0.78 or more EUR for 1 USD, according the Currency Converter at www.oanda.com.	
10BUFFS	The CU Buffs football team has a current season winning percentage of greater than 25% as of the settlement date.	
11SNOW	The cumulative snowfall since this past July (as reported on the National Weather Service website) is equal to or greater than the normal cumulative level. Note: the statement of this event should have specified some geographic bounds.	
12CO68	Colorado Amendment 68 was approved by voters in the 2014 election.	

Part 1: Looking at Consistency of Each Person's Forecasts

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

1) Of the people who made forecasts, how many had a probability for 4HOG1U that was greater than or equal to their probability for 3HOG5U?

2) Answer these questions:
    a. TRUE or FALSE? For each person making a forecast, the probability (forecast) for 2HOG1D should be greater than or equal to the probability for 1HOG5D.

    b. TRUE or FALSE? For each person making a forecast, the probability for 1HOG5D should be equal to the probability for 3HOG5U.

    c. TRUE or FALSE? For each person making a forecast, the probability for 1HOG5D *plus* the probability for 3HOG5U *plus* the probability for 5HOG5M should equal 1. (For this question, assume that HOG is still being traded on the settlement date.)

You should be able to identify correct logical relationships among the first five events, or similarly related sets of events.

Part 2: Determining Outcomes

We will now determine which of the twelve events occurred and which did not. Using the table below, and the best information you can find, determine whether or not your assigned event occurred.

Assigned Events	
Range of First Letter of Last Name	**Find out whether this event occurred**
A-C	6YAHOO
D-G	7ALLSP
H-K	8TWTNB
L-N	10BUFFS
O-S	11SNOW
T-Z	12CO68

In class, we will discuss the events in that table along with the first five events (about HOG stock) and event 9USDEUR.

In your copy of the spreadsheet on the Occurrence tab, record events that occurred with a 1 and events that did not occur with a 0.

Part 3: How Did We Do?

We can compare individual forecasting performance by looking at the **total squared difference** for each person making a forecast. In this context, the *difference* for a given 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, respectively. The *squared difference* is the square of that difference. The *total squared difference* is the sum of the twelve squared differences (one for each of the twelve 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 Response ID 103 would be written $x_{103,8}$ (and in this data set, that value is 0.85).
      - Denote the outcome of event j as y_j. For example, the outcome of event twelve was "did not occur," which we represent with a 0. Thus, $y_{12} = 0$.
        Using that notation, write the expression for the total squared difference T_i for person i.

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

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

You should be able to answer questions about the total squared differences, for example, what percent of them were below a certain level, or which Section (1, 2, or 3) had the best average.

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 .5 and higher to 1; round probabilities of under .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 twelve events? How many people predicted 9 or more events (using the rounded probabilities)?

You should be able to answer questions about the rounded probabilities, for example, which event had the most people with a rounded probability that matched the outcome.

Part 4: Calibration

In this part, we will work with the probabilities as entered, not the rounded probabilities.

1) 0% and 100%
    a. How many forecasts were exactly 0 (i.e., entered as 0, not rounded to 0)?

    b. Of those forecasts, for how many of them was the outcome 0 (did not occur)? For how many was the outcome 1?

    c. How many forecasts were exactly 1 (i.e., entered as 1, not rounded to 1)?

    d. Of those forecasts, for how many of them was the outcome 1 (did occur)? For how many was the outcome 0?

To tell how good a set of forecasts is, we will use the idea of *calibration*. A set of forecasts is *well-calibrated*, if, of all the times it forecasts 40% chance, the outcome did occur about 40% of the time. Can we look at the calibration for a forecast of 32%? Probably not: that particular value is unlikely to have been assigned enough times to have a good sized sample. Therefore, we will group the forecasts into buckets.

2) Calibration Table
    a. Find the values in this table:

Forecast probability range—Note which endpoint is included in which range	**Number of forecasts in this range**	**Percent of all the forecasts that are in this range**	**Percent of time the forecasted event occurred for forecasts in this range**
[0%, 10%]	110	12.56%	
(10%, 20%]	58	6.62%	
(20%, 30%]	75		
(30%, 40%]			
(40%, 50%]		15.64%	
(50%, 60%]	70		
(60%, 70%]	97	11.07%	
(70%, 80%]			
(80%, 90%]			
(90%, 100%]	88	10.05%	

The second column should add up to the total number of forecasts in our data set. The third column logically sums to 100%.

Note that the fourth column does **not** logically sum to 100%. If we are *perfectly* calibrated, the numbers in the third column would be 5%, 15%, 25%, …, 95%. That would mean that of all the times people assigned a probability between 30% and 40%, the event occurred 35% of the time.

Even the best weather forecasters in the world are not *perfectly* calibrated, so we shouldn't feel bad if we are not. **If we are well-calibrated, then the numbers in the third column are at least increasing.**

b. Create a histogram showing the relative frequency of each bucket (i.e., the third column in the table above). Use informative labels at the base of your columns.

c. Create a column or bar chart showing the calibration percentages (i.e., the fourth column in the table above). Use informative labels at the base of your columns.

d. Discussion:
    i. One common problem is that people overuse very low probabilities. Is the first column or bar from the chart in part c much higher than 5%?

    ii. Are there other systematic deviations from perfect calibration?

    iii. Overall, is the class well calibrated?

You should be able to find the counts and percentages of forecasts within any given range. You should be able to determine counts of forecasts within a range for events that actually did occur (or did not). You should be able to interpret calibration information.

Selected Answers

Part 1

1) 60 (out of 73 people who made forecasts)

Part 3

1b) 0 is the lowest possible value for total squared difference. The total squared difference would be 0 if someone assigned probabilities of 0 or 1 for each event and was perfectly correct.

2a) 372 of the rounded probabilities were 0

2c) 504 of the rounded probabilities were 1

Note that 372 + 504 = 876 = 73 people x 12 forecasts/person

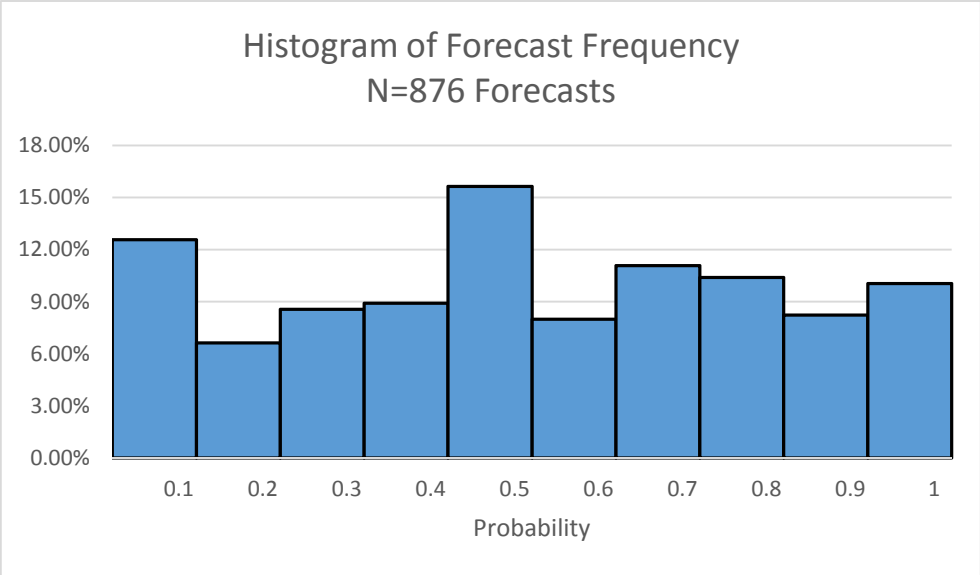
(We did not include the forecasts of people who only answered for some, but not all, of the events.)

Part 4

1a) 36 of the 876 forecasts were exactly 0

1c) 56 of the 876 forecasts were exactly 1

2b)



For many of the questions, we can't provide solutions in this document because the outcomes of the events have not yet been determined.