Dmitriy Strelkov

Dr. Eric Corwin

PHYS 581 – Design of Experiments

December 9th, 2016

**Author’s note:** This report is meant to be read simultaneously with the attached commented code. I have deliberately included pauses in the code so that various elements and figures can be talked about without needing to take up space in this document. When executing my code in R, I advise you not to proceed to the next step in the code until I mention to do so in this paper. This will improve readability of both the report and the code. Pressing the Enter key in the R console advances to the next step in the code.

It’s Tea Time! Elucidation of the Relationship Between Taste Preferences, Brewing Conditions, and Perceived Tea Quality

The core concept behind Bayesian modeling is the reassignment of probability across possibilities given data. Beginning with prior assumptions as to the underlying mechanism behind a phenomenon, we collect data and assess the validity of our prior assumptions, updating our beliefs about the system to accurately represent the data available. One of the most powerful parts of Bayesian modeling is that even if very little is known about the system prior to data collection, Bayesian statistics provides us with tools to quantifiably verify certain characteristics of the system in question. It also provides us with the tools to uncover results that are not apparent at first glance.

To illustrate these concepts concretely, I will be presenting an analysis of tea tasting data that has been collected by the other Design of Experiments class. Twelve observers with varying taste preferences and tea knowledge were given 15 cups of tea to taste and rate on a scale of one to ten, ten being the highest possible rating. The different tea cups had some combination of four different parameters: Green/Black tea, Sweetened/Non-sweetened, Long/Short brew time, and High/Low leaf mass while brewing. In this paper, I will investigate the significance of these factors on tea scores. I will also investigate the significance of observer taste preferences, whether preferences are more significant than brewing factors, and finally the differences in rating distributions across all three taste preferences. From these distributions I will also posit the most likely tea score a cup of tea will have given the observer preferences and brewing factors.

The first thing to do when analyzing a set of data is to figure out how it is distributed. Sourcing the code provides us with a histogram of the total data set(tea scores), and it looks like a normal distribution. Let’s check to make sure that it is indeed a normal distribution. Using the next algorithm, the standard deviation for the tea score data is computed to be 1.858. In order to test whether or not our data is normally distributed or not, we will generate a synthetic set of normally distributed data with the same mean and standard deviation as our tea score data and then apply an algorithm to both data sets that returns the probability of the input being normally distributed. If the values are similar, then the data truly is normally distributed.

The next graphic shows the histogram of the synthetic data and the tea score data side by side. Proceeding to the next step in the code, a few outputs are seen in the console. The most important output in this step are the two numbers, one near 0.25 and one near -366. These two numbers represent the scaled probability of the distribution being normal and the exponential scaling factor that allows us to compare the likelihood of any particular scaled probability to any other scaled probability. The scaling factor allows for the assessment of probability of extremely small values. Back to the numbers, we can see that they are separated at most by a value on the order of 1. To put this in perspective, if data containing a model other than a normal distribution were to be entered into this algorithm, the difference in scaling factors would be on the order of 1000. The similarity in scaled probability and scaling factor indicate that the data truly is normally distributed.

Now that we know(to a high degree of certainty) the distribution of the totality of the data, we can begin to look at the subsections within the data for deviations from this behavior. We will begin by making simple observations, then we will make simple comparisons, and from those simple comparisons we will arrive at a basis for determining the underlying rating distribution. The first thing we notice by simply dividing up the data into black tea and green tea scores is that on average, people prefer green tea. Additionally, the highest possible score is exclusively reserved for green tea and the lowest possible score is exclusively reserved for black tea. Let’s investigate whether this holds true for all taste preferences, that is, whether or not the prior assumptions(black tea is bitter, those with bitter preferences will like black tea better) have an effect on the score distribution.

As seen on the bar plot, most people have no preference in taste, and there are slightly more who prefer sweet tasting tea to bitter tasting tea. Another thing to note is that tea knowledge correlates nearly exactly with taste preference, with those who are minimally knowledgeable preferring sweet tastes, those who have intermediate knowledge having no taste preference, and those who are well versed in the art of tea preferring bitter tea. This information could be useful later when interpreting results.

We see that, on average, those with no taste preference enjoy tea the most, those with bitter taste preferences enjoy tea slightly less, and those with sweet taste preferences enjoy tea significantly less than the other two groups. The distribution of the bitter preference group is distinctly different from the other two distributions, which look to be similar. We won’t worry about distributions for now; our goal at the moment is to simply make observations and comparisons that can give us some insight into our data. Next we see that in the no preference group green tea is heavily favored over black tea. At the same time, black tea has a better reception among this group than across the entire population. The distributions still seem to reflect normal distributions.

Apparent deviation from the normal distribution arises in those who prefer bitter tasting tea; this will be addressed later. What is interesting to note here is that the group who purports to prefer bitter tasting tea has a lower mean score for black tea and a higher mean score for green tea than the group who has no taste preferences. This is unexpected given that black tea is generally quite bitter and green tea quite mild. I expected a higher mean score for black tea and a lower mean score for green tea.

Moving on to the group with sweet taste preferences, we see a significant drop in tea scores across the board, and a return to a normal-looking distribution. This is unsurprising, as tea is generally not a beverage known for its sugar content, green tea and black tea especially. Given that minimal tea knowledge correlates to sweet taste preference, there could be an unfamiliarity with the beverage that affects perceived quality in the observer. Our final observations about the influence of the tea type on score concludes with a blanket statement that black tea is generally less popular than green tea, and those who prefer sweet tasting tea enjoy tea significantly less than those with different flavor preferences. We will now move on to examining the effect of adding sweetener on tea score.

When data reflects predictions it is always a nice warm feeling. Take a look at the distributions of scores in the no preference group and rejoice in how the means and distributions hardly change whether the tea is sweetened or unsweetened. This is in complete contrast to the effect of sweetener on the scores of the bitter preference group; the mean shifts by about 25% in either direction depending on whether the tea contains sweetener or not. However, once again the data reflects our prior ideas about expected values: adding sweetener reduces tea scores in those who prefer bitter tea, and omitting it raises tea scores in the same group. Finally, the group who prefers sweet tastes also follows the trend of producing data in accord with priors, but the effect is diminished when compared to the bitter preference group. It was unexpected to me how the effect of this parameter varied across different flavor profiles. In those with bitter preferences this factor is more significant than the type of tea being tasted! At the same time, this factor has next to no impact on those who have no flavor preference. In those with sweet preferences, this factor had similar significance to the type of tea being tested. Next up: brew time.

Longer brew times result in denser tea flavor in the actual beverage. I didn’t expect this factor to have much of an impact on scores; I was correct in some respects and mistaken in others. The mean scores of the no preference group do not change significantly with brew time, but the bitter preference group once again shows a different story. The mean of the bitter preference group shifts by 10% in either direction, increasing with a long brew time and decreasing with a short one. One possible explanation is that, as experienced tea drinkers, the bitter preference group can discern a large difference in flavor profiles with varying brewing time, while those with no preference don’t care, and those with sweet preferences don’t have the necessary experience to tell the difference.

The last brewing condition we will examine is tea leaf mass. When brewing tea, the amount of leaves used determines the flavor ceiling. More leaves lead to a denser flavor, similarly to a long brew time. I expected this factor to be the least significant overall, and I was correct. The effect on the mean tea score was not especially large for any of the flavor profiles. However, this factor was the most significant factor besides tea type in the no preference group which was surprising.

Phew. That was a lot of text but not very much in-depth analysis. That’s okay, because by separating the observers into distinct groups and seeing how each parameter changed the average tea score in each group, we have formulated an intuition on which brewing conditions to prioritize for each group when creating a probability distribution of their tea ratings. We have determined that the significant factors are as follows for each group: no preference tea scores are most affected by tea type and leaf mass, sweet preference tea scores are most affected by tea type and sweetener, and bitter preference tea scores are most affected by tea type, sweetener, and brew time. We have also determined that tea type is the most significant factor across the board, and that black tea is less popular than green tea. Using these observations, we will now create multivariate probability distributions of tea scores for each of the three groups.

First we will examine the no preference group. The significant brewing conditions for this group are tea type and leaf mass. The way the resulting contour plots are created in the code are as follows: We take the histograms of the significant factors and use the count function(in the ‘plyr’ library) to turn them into a data frame that contains the occurrences of each score from 1 to 10, filling in missing values as having 0 occurrences. Then we matrix multiply these counts by each other to create a likelihood surface plot. This likelihood surface plot has a maximum, and the scores associated with that maximum are then multiplied together and subsequently rooted to the nth power, where n is how many scores are associated with the maximum likelihood. This provides us with the most likely score value for tea brewed with those specific parameters tasted by that group of observers.

The first four contour plots seen are the no preference likelihood contour plots. It is interesting to note that although the mean preference for high mass is higher than the mean preference for low mass, the most likely score is higher in the low leaf mass. This is due to the lower sample size of data with the low mass parameter. In the low mass data, lower values have a larger effect on the mean compared to low values in the high mass data. Additionally, a larger proportion of the low mass data is at the maximum number of occurrences: 45% vs 21%. We still see the continued trend of black tea being significantly less popular than green tea reflected in these plots.

The next four contour plots show the sweet preference likelihoods. Here we once again see strange outcomes. The most likely score for green tea is lower than black tea, a phenomenon we see nowhere else. Ironically, the brewing conditions that result in the highest projected score for the group with sweet taste preferences result in bitter tea. This reinforces the earlier idea that inexperienced tea drinkers (sweet preference) can’t really tell the difference between different kinds of tea and don’t know what kind of tea they really like.

Finally, we arrive at the bitter preference likelihoods. Because the bitter preference group has three different factors that go into determination of most likely score, it is a bit trickier to visualize the score probability distribution on a surface plot. The problem of extracting the actual maximum likelihood was solved by creating a 3D array, similar to the matrix created for likelihood values in previous groups, but with an added dimension for the third parameter. The contour plots that are displayed in the code represent the score distribution at the plane of most likely value. That is, a slice of the 3D array is taken at the plane where the highest likelihood occurs and is projected onto a surface plot. In doing this we cannot visualize the full contribution from brew time, but we can still extract the most likely scores given certain conditions.

The first four plots in this group show all the possible permutations of the brewing conditions for green tea. Among those who prefer bitter tastes, the highest most probable score goes to the unsweetened long brew green tea. This is expected. Continuing to the four plots for black tea we reach the same conclusions. This reinforces our idea that those with experience drinking tea (bitter preference) know what they like and can discern between tea flavors.

In the end, what do we have? We have the normalized probability distributions for the tea scores for three different flavor profiles. We can make a wealth of associations between different brewing conditions and perceived effect on tea. These are exactly the types of conclusions talked about in the first paragraph. By performing Bayesian data analysis on this set of tea score data we went from unquantifiable assumptions about how brewing conditions and taste preference affects perceived tea quality, to the observation that the totality of the data is normally distributed, to obtaining a predictive mathematical model for future experiments. And now, I think I’m going to go brew myself some tea.