Your Task

You have taste and chemical data about beer brewing. Your task is to construct a predictive relationship between the mean preference rating of a beer and its measurements on the chemical variables. Describe the identified relationship and comment on its anticpated performance characteristics including both its strengths and weaknesses. Use the predictive relationship to construct predictions for the mean preference rating of each of the beers for which chemical data are provided in your file 2. Your work will be partially assessed by comparing your predictions with the observed mean preference ratings for the additional beers.

In addition to your predictions you should carefully describe what analyses you did and what methods you used. You should include a short executive summary that could be read by any literate person.

The Data

There are two data files (file 1, the 91 beers; file 2, the 4 extra beers). There are missing values for some records and these are indicated by "NA"s in the single-line files. The variables are as described in Tables 3, 4, and 5.

BEER CHEMISTRY
AND
CANADIANS' BEER PREFERENCES

Beer companies want to understand the relationship between the chemical characteristics of beer and the preferences for beer exhibited by consumers. Two data sets are provided to you in this study. The first set consists of chemical measurements on 91 beers and preference measurements on the same beers collected from beer consumers in blind taste tests. You are asked to use these data to develop a statistical model relating beer chemistry and consumer preferences for beer. The second data set consists of chemical measurements on a holdout sample of 4 beers. You are asked to employ your statistical model to predict consumer preferences for the beers in this holdout sample.

STUDY DESCRIPTION

1 Background

Beer is a popular beverage with Canadians, as evidenced by the fact that they purchase almost as much beer as fluid milk on an annual basis. In 1987, for example, beer sales in Canada were just over two billion liters, which amounted to 113 liters for each Canadian over 19 years of age that year (Statistics Canada, 1989).

Competition among beer companies for the loyalty of beer drinkers is strong. Advertising and promotion play a vital role in establishing and maintaining the competitive positions of brands, but the flavor of the beer is an important long-term competitive element. A poor quality beer or one that is out of vogue with current taste will eventually erode a brand's competitive position, no matter how cleverly it is promoted and advertised. For this reason, breweries conduct blind taste tests of beers to measure the acceptability of their own brands, competitors' brands, and new product formulations.

Brewers and beer marketers are interested in understanding how the physical characteristics of beer as measured in the laboratory correlate with the expressed degree of consumer liking in taste tests. If the relationship of beer chemistry to consumer preference can be established analytically, then brewers and marketers can adjust the characteristics of a brand to give it maximum appeal to a target market segment. But the task is not a simple one. Beers contain hundreds of chemical compounds, including volatiles, amino acids, sugars, nucleotides, and bittering compounds, and no mathematical formula is known that relates these compounds to consumers' preferences. Moreover, consumers' taste preferences are not static. Preferences also vary from one market segment to another, depending on consumers' characteristics such as gender, age, education, consumption level, and regular brand. Finally, even within a homogeneous group of consumers, preferences will vary widely from one individual to another, and from one occasion to another for the same individual.

In this study, we are asking you to examine consumer-taste-test and chemical data in order to estimate the underlying relationship, if any, between beer chemistry and Canadians' beer preferences.

3.89	3.59	36.80	1.00786	2.02	3.82	11.48	5.02	0.03
145	13.0	20	0.65	2.71	25	19	125	1.0
5.700	12.028	32.408	25.655	0.396	0.136	0.045	97.463	17.709
0.152	0.109	3.562	0.387	0.228	0.172	2.165	0.012	
11731	36	87	6.18					
3.90	2.68	36.40	1.0076	1.96	3.76	11.39	5.00	0.02
143	14.3	20	0.7	2.76	20	27	122	1.0
8.765	10.164	25.049	21.00	0.140	0.074	0.016	79.012	17.176
0.124	0.101	2.841	0.31	0.236	0.177	0.919	0.015	
11732	36	88	5.82					

Table 1: A fragment of the data from File 1

3.92	3.58	36.65	1.00770	1.97	3.78	11.45	5.02	0.11
144	14.6	30	0.63	2.81	38	19	131	3.7
3.664	15.180	18.469	12.204	0.043	0.160	0.008	69.734	10.070
0.040	0.125	1.575	0.118	0.198	0.155	0.287	0.000	
11904	59	9						
3.94	4.23	36.7	1.00808	2.07	3.84	11.35	4.92	0.00
143	16.6	2	0.81	2.79	32	13	143	3.0
2.607	18.329	18.88	13.681	0.036	0.160	0.009	74.543	10.633
0.047	0.117	1.71	0.135	0.209	0.201	0.304	0.00	
11905	59	9						

Table 2: A fragment of the data from one File 2

2 The Data

The data and background information for this study have been provided by Molson Breweries. The data consist of two files. File 1 contains observations on four taste-test variables and measurements on 35 chemical variables for each of 91 beers. File 2 contains observations on three taste-test variables and measurements on the same chemical variables for 4 additional beers. Each of you has four different beers in File 2. A fragment of the data for Files 1 and 2 appear in Tables 1 and 2, respectively. The data fragments each include observation sets for two beers.

Tables 3 and 4 list the 35 chemical variables measured for each test liquid. The variables are of two basic types. Variables 1 to 18 are analytical measurements obtained by laboratory procedures that are standard in the beer industry; these are given in Table 3. A brief description of the analytical variables is also presented in Table 3. Variables 19 to 35 are beer volatiles measured by gas chromatography; these are given in Table 4. All measurements were made in the same laboratory, using reasonably consistent procedures. Some measurements are missing. Reasons for missing observations include, among others, inadequate sample liquid provided to the laboratory for analysis, unknown values (for example, the age of a competitor's beer), and incomplete laboratory work.

Only File 1 has consumer preference data. The data were gathered in consumer taste tests, which we now briefly describe. Some details are omitted to protect the proprietary interests of Molson Breweries.

A quota sample of adults who have drunk a specified minimum amount of beer in a recent time interval is recruited for each taste test. The recruiting is done by telephone solicitation. The quotas relate to target proportions of subjects in specified gender, age, and education classes. The personal profiles of subjects vary somewhat among tests, as the marketing needs of taste tests differ from one test to another. The tests included in this study, however, have been selected so that their subject profiles are reasonably uniform. The time of year and location of the test vary from test to test. The taste tests are generally held at central locations (such as hotels) in major Canadian cities. Subjects are given general instructions about the purpose of the test but are not told the name of the sponsoring brewery. A small monetary incentive is given to each subject for participating in the test.

Each subject tastes and evaluates a sequence of test beers. The different sequences are assigned randomly to subjects. The experimental design for each test is a replicated latin-square design with three factors: subject, trial order, and beer. Replicates are incomplete on occasion because of unusable or missing response data. The beers are served to subjects without any brand or brewery identification, i.e., the tasting is done on a blind basis. For each product tasted, a small amount of the beer, chilled to a temperature of about 6°C, is presented to the subject in a clear glass. Crackers and water are provided so the subject can clear his or her palate between tastings. The beers tested include Molson brands, Molson developmental products, and competitors' products purchased from retail outlets.

Each subject is given a questionnaire, which is completed without communication with other tasters. The environment of the tasting room is quiet, bright, and fresh. No smoking is permitted during the test. The aim is to have subjects make their evaluations without distraction or contamination from extraneous factors. The questionnaire asks for background information about the subject, such as age, education, regular beer brand, and beer consumption. It also contains a battery of opinion questions about the aroma, flavor, taste, appearance, and other characteristics of each beer being tasted. Among the opinion questions for each beer is one that asks the subject to give a preference rating using a nine-point rating scale having semantic labels ranging from "dislike extremely" (rating 1) to "like extremely" (rating 9).

The taste-test data in File 1 consist of observations on four variables for each liquid. These variables are listed in Table 5 and include: a liquid identification number (variable 36), a test number (variable 37), the number of subjects who rate the beer (variable 38), and the mean preference rating of the subjects (variable 39). Beers tasted together in the same test have the same test number. Response data at the level of the individual subject are not made available, because they are too numerous. Data concerning the personal profiles of subjects participating in each test were not made available for proprietary reasons.

3 The Analyst's Task

You are asked to perform two tasks. First, you are to construct a predictive relationship between the mean preference rating of a beer and its measurements on the 35 chemical variables. You are also asked to describe the identified relationship and to comment on its anticipated performance characteristics, including both strengths and weaknesses. Next, the predictive relationship is to be used to predict the mean preference rating of each of the your four beers for which chemical data were provided in File 2. The success of your predictions will be assessed by comparing the predictions with the actual mean preference ratings for the 4 liquids.

1. pH: Hydrogen-ion concentration. A measure of acidity of the beer. Values from 0 to 7 represent acidity. Pure water has a pH of 7.
2. COLOR: Spectrophotometric color. A measure of beer color intensity based on the light absorbance of the beer. Expressed in degrees Standard Reference Method (SRM).
3. RI: Refractive index. A measure of the refraction of light passing through the liquid.
4. SP GR: Specific gravity. The weight ratio of a given volume of beer to the same volume of water.
5. AE: Apparent extract. The apparent amount of sucrose in the beer as measured by a saccharometer. Expressed as a percentage by weight (degrees Plato).
6. RE: Real extract. The actual amount of extract (solids) in the beer. Expressed as a percentage by weight (degrees Plato).
7. OE: Extract of original wort. (Wort is the liquid which is fermented into beer.) The calculated original amount of solids in the wort from which the beer was brewed. Expressed as a percentage by weight (degrees Plato).
8. ALC: Alcohol. The alcohol content of the beer. Expressed as a percentage by volume.
9. RFE: Remaining flavor extract. The amount of yeast fermentable extract remaining in the beer. Expressed as a percentage by weight.
10. CAL: Calories. The number of calories in 341 ml of beer (the volume of one standard bottle).
11. BU: Bitterness units. The amount of bitter substance in the beer as measured by the international method, which involves a spectrophotometric comparison of prepared beer with an isooctane-octyl alcohol standard.
12. VDK: Vicinal diketones. The concentration of vicinal diketones, consisting mainly of diacetyl, in the beer. Expressed in micrograms per liter of beer.
13. AIR: Air. The volume of air in 341 ml of beer (the volume of one standard bottle). Expressed in milliliters.
14. CO_2: Carbon dioxide. The volume of dissolved carbon dioxide in the beer. Expressed as the number of volumes of gas per volume of liquid.
15. FTU: Formazin turbidity units. The degree of turbidity or haziness of the beer measured in comparison with a formazin standard. A measure of 0 represents perfect clarity; a measure of 200 represents a distinctly hazy liquid.
16. AGE: Age. Expressed in days since bottled.
17. FOAM: Foam collapse rate. Measured by the sigma value method. Expressed in sigma units.
18. SO_2: Sulphur dioxide. The quantity of sulphites in the beer. Expressed in terms of their SO_2 equivalent, in milligrams per liter of beer.

Table 3: Chemical Variables - Analytical

19. Isobutyraldehyde
20. N-Propanol
21. Ethyl Acetate
22. Iso-Butanol
23. N-Butanol
24. Ethyl Propionate
25. Propyl Acetate
26. 3-Methyl-1-Butanol
27. 2-Methyl-1-Butanol
28. Iso-Butyl Acetate
29. Ethyl Butyrate
30. 3-Methyl Butyl Acetate
31. 2-Methyl Butyl Acetate
32. Ethyl Hexanoate
33. Ethyl Octanoate
34. 2-Phenylethyl Acetate
35. Ethyl Decanoate

Table 4: Chemical Variables - Volatiles (Units are Parts per Billion)

36. Liquid identification number
37. Test number
38. Number of subjects rating the beer
39. Mean preference rating (1,2, . . . , 9: low to high)

Table 5: Taste-test variables