STAT8801 March 5, 2014

Today:

Review your case with others who will be playing the same role as you. What questions do you think the statistician will ask? Prepare your answers. You may need to make them up, if they aren't given in the prompt. You can choose to have different potential answers than others with the same case.

Next:

1) With your partner, role-play each consultation (one where you are the client, and one where your partner is). Video record them. I expect these consultations to be between 15-30 minutes long; a real client would be longer.
   Cameras can be checked out from the library; see https://www.lib.umn.edu/media/equipment.
2) Watch the video where you were the consultant. Find two segments of particular interest; preferably one where you think you did well and one where you think you could have done better.
3) Either post your videos online (you can do so privately, so that one has to have the URL to access it), or talk with me about other options (I'll look into some in the meantime).
4) Reflect on your experience, and write a page about it. Specifically comment on the segments of particular interest, and specify at what time of the video they occurred.

Due date: by the end of spring break, though hopefully you can get it done before the break.

PS. The next writing assignment will be based on these, so pay close attention!

You are a traffic engineer with the local transit authority. Students at the local university have complained about buses being “bunched” together. You’ve decided to test this claim and have done some data collection, and now need it analyzed.

You’d like the statistician to do the analysis for you and write a report with the results. You have some statistical background, as you took a class or two in college, but it’s been several years since you’ve had to do any statistics.

Your overall goal is to improve bus service through the University corridor, and to do so, first want to identify what problems exist, if any.

First, you want to know how often bunching occurs. You’re also interested in how long students usually have to wait, what that may depend on, and if it can be predicted. (For prediction, note that the On, Off, and Cars variables from the previous bus should be used.)

Data was collected during one key interval, the weekday afternoon rush. Several engineers and interns were involved in data collection; each took a different two-hour block during afternoon rush hours, 3-7 PM, Monday-Friday. Each intern sat at the same designated point, noting the time of every arrival and other information related to the bus.

You have 207 observations, from 11 different days over a three week period. The variables are as following, and a sample day is on the back.

You're willing to define “bunching” as when a bus comes within three minutes of the last one. You'd prefer the result be something like “bunching only occurs X% of the time.”

Full data at this link; excerpt follows.

http://users.stat.umn.edu/~rend0020/Teaching/STAT8801-2014Spring/buses.csv

Sample Data

Day: Day of the Week, M, T, W, H, F
Date: Day of the Month
Temp: recorded temperature for that session, in Fahrenheit
Rain: 1 if raining, 0 if not
Hour: Hour the bus arrived
Minute: Minute the bus arrived
Time: Time in minutes since last bus (a computed variable; -1 if the first bus)
Bunch: 1 if time since last bunch is 3 or more minutes, 0 if less (another computed variable; -1 if the first bus)
On: Number of people who got on the bus
Off: Number of people who got off the bus
Cars: Number of cars at the next red light after the bus left

Day	Date	Temp	Rain	Hour	Minute	Time	Bunch	On	Off	Cars
T	26	78	0	3	27	-1	-1	0	0	1
T	26	78	0	3	33	6	0	0	0	4
T	26	78	0	3	43	10	0	1	0	3
T	26	78	0	3	48	5	0	0	1	1
T	26	78	0	4	9	21	0	0	0	3
T	26	78	1	4	12	3	0	0	1	4
T	26	78	1	4	12	0	1	0	0	4
T	26	78	0	4	21	9	0	0	0	6
T	26	78	0	4	29	8	0	0	0	3
T	26	78	0	4	40	11	0	0	4	9
T	26	78	0	4	44	4	0	0	1	8
T	26	78	0	4	59	15	0	0	0	2

You are a market analyst for a local supermarket chain, and are studying drinking water preference among university students. You went to the recent Spring Festival on campus and set up a booth, and had students who stopped by sample water. They received a $2 coupon for purchases made at your supermarket in exchange.

You did two different experiments during the day; it's possible some students were in both experiments.

For the first, each subject was given four identical cups from which to taste: three contained different brands of bottled water (Fiji, Aquafina, and your store brand) and a fourth contained tap water. The subjects did not know which type of water was in each cup. For double-blinding of the experiment the water was poured from four jugs labeled only as A, B, C, and D. (You are proud of yourself for thinking to blind the water jugs in this way. For the second, each subject was given three cup, poured from jugs from Fiji, Aquafina, and your store brand, but all three jugs actually contained tap water. The students in this experiment were shown and told which jug the water came from. In each experiment, you recorded the order of preference for each subject. You also recorded gender, age, and class, as well as what kind of water they usually drink (tap, bottled, or filtered) and the subject's favorite brand of bottled water, if any.

You want to know which brand is preferred, and whether it is due to the taste or the brand name. You'd also like to know if this depends on gender, age, or the other covariates you recorded.

You'd like the statistician to do the analysis for you and write a report with the results. You have some statistical background, as you took a class or two in college, but it's been several years since you've had to do any statistics.

Full data at these links; excerpt follows.

http://users.stat.umn.edu/~rend0020/Teaching/STAT8801-2014Spring/water1.csv

http://users.stat.umn.edu/~rend0020/Teaching/STAT8801-2014Spring/water2.csv

Variables in the analysis:

Gender - Gender of subject recorded as M or F for Male or Female, respectively.

Age – Age of subject recorded in integer values.

Class – Academic rank of subject recorded as F, SO, J, SR, O for freshman, sophomore, junior, senior, and other, respectively.

UsuallyDrink - Type of water the subject usually consumes recorded as B, F, T, NA for bottled, filtered, tap, and not applicable, respectively. See the Special Notes below regarding the NA values.

FavBotWatBrand - Favorite bottled water brand of the subject. Bottled water brands are spelled with all capital letters (e.g., AQUAFINA, DEER PARK, NONE). Note that some have blanks in them so do not delimitate the data file using blanks.

Preference (first data set) - List of ordered preferences from the first taste test recorded as four-letter strings containing the letters A, B, C, and D. For example if Preference = BCAD, then the subject chose water type B as his first preference, C as his second, A as his third, and D as his fourth, or least preferred, water type. For decoding purposes A, B, C, and D correspond to Store Brand, Aquafina, Fiji, and Tap water.

Preference (second data set) - List of ordered preferences from the second taste test recorded as three-letter strings containing the letters A, F, and S.

Sample from first:

Gender	Age	Class	UsuallyDrink	FavBotWatBrand	Preference
F	18	F	F	STORE	CABD
F	18	F	T	NONE	CABD
F	18	F	T	FIJI	CADB
F	18	F	T	AQUAFINA	CBDA

Sample from second:

Gender	Age	Class	UsuallyDrink	FavBotWatBrand	Preference
F	9	O	B	AQUAFINA	AFS
F	18	F	F	AQUAFINA	FSA
F	18	F	T	AQUAFINA	FSA
F	19	SO	F	FIJI	FSA