

Homework Assignment 2

Problem 1

A credit card company collects data on 10,000 users. The data contained two variables: an indicator of the costumer status, i.e., current ($def = 0$) or in default ($def = 1$) and a measure of their loan balance relative to income, i.e., low ($bal = 1$), medium ($bal = 2$) and high ($bal = 3$). The data is in the following table:

		def	
		0	1
bal			
1	8,940	64	
2	651	136	
3	76	133	

1. Compute the estimated marginal distribution of costumer status
2. What is the conditional distribution of bal , given $def = 1$?
3. Make a prediction for the status of a costumer with a high balance.

Problem 2

In a recent episode of Mythbusters, Jamie and Adam (the show's hosts) wanted to determine whether women are better multitaskers than men. To test this theory, they had 10 men and 10 women perform a set of tasks that required multitasking in order to have sufficient time to complete all of the tasks. They use a scoring system that produces scores between 0 and 100.

The women ended up with an average of 72 with a standard deviation of 5, while the men averaged 64 with a standard deviation of 9. In the show, The Mythbusters concluded that this 8 point difference confirms the myth that women are better multitaskers. Based on the results from the experiment, do you agree with their conclusion? Why?

Problem 3

During a recent breakout of the flu, 850 out of 6,224 people diagnosed with the virus presented severe symptoms. During the same flu season, an experimental anti-virus drug was being tested. The drug was given to 238 people with the flu and only 6 of them developed severe symptoms. Based only on this information, can you conclude, for sure, that the drug is a success?

Now, it turns out that the people who received this drug were all MBA students. Can you infer any causal connection between the drug and the lack of severe symptoms? What are some potential confounding variables that may influence whether someone develops severe symptoms or not?

Problem 4

Time to revisit the slides... rework the following examples:

- Oracle vs. SAP
- Gender gap in the Chicago banking industry
- Google's new search algorithm

Make sure you understand the computation of the standard errors and that you can understand the progression of the examples and their conclusions.

Problem 5

In 1960, census results indicated that the age at which American men first married had a mean of 23.3 years. It is widely suspected that young people today are waiting longer to get married. We want to find out if the mean age at first marriage has increased during the past 50 years. We plan to test our hypothesis by selecting a random sample of 40 men who married for the first time last year. The men in our sample married at an average age of 24.2 years, with a standard deviation of 5.3 years.

1. Based on a 99% confidence interval, what do you conclude?
2. Now, use a t-stat... explain your conclusion

Problem 6

A SurveyUSA poll conducted on March 1, 2011 asked randomly sampled Los Angeles residents about their views on American vs. foreign-made products. One of the questions on the survey was “If an American-made product cost slightly more than a foreign-made product, which would you be more likely to buy?”

81 out of the 166 respondents between the ages of 18 and 34, and 248 out of the 334 respondents 35 years and older said they would prefer the American-made product. We are interested to see if younger people are less likely to choose American-made products. Test this hypothesis at the 5% level.

Problem 7

Your dad claims that he shoots better than 50% from the three point line. You bet him \$100 that doesn't, based on a challenge where he attempts 10 shots. He makes 7 of his shots and says that you owe him \$100. You say that you do not, because you cannot reject the null hypothesis that he shoots exactly 50%. Who is right, here?

Problem 8

You have collected a dataset containing rents for a random sample of 67 two bedroom apartments in West Campus. Some summary statistics of the data are below:

n	mean	sd
67	981.3	182.6

1. What is your best estimate for the average rent among all two bedroom apartments in West Campus?
2. Give a 95% confidence interval for the average rent. Is \$1000/month a plausible value for the average rent? What about \$1100?

Your dataset includes a number of other characteristics of each apartment, including whether they have laundry machines in-unit. The table below summarizes each subset of the data (units with and without in-unit laundry)

laundry_in_unit	n	mean	sd
No	19	847.3	89.5
Yes	48	1034.4	183.4

3. What is your best guess for the *difference* in average rents for units with laundry and without laundry?
4. Compute the t-stat for testing whether the difference is zero. What do you conclude?
5. Give a 95% confidence interval for the difference. Is zero a plausible value? Is \$150?
6. (Optional) Suppose you own a two bedroom rental in West Campus which does not have in-unit laundry. You have the option to add laundry machines to your unit. You assess that you'll need to charge an additional \$150 a month in rent to make it worth your while. Based on your answers above, is adding laundry to your unit a good or bad idea? Do you need any more information to decide?