Homework 4

Shun-Lung Chang, Dilip Hiremath

1. From your previous analysis you know that HomeOwnership is an important predictor for getting a credit card approved. You hence limit your analysis in this homework to homeowning applicants only. Create a subset of the data just including the homeowners.

(a) (half a point) For this subset, compute the median income.

(b) (half a point) For this subset, compute the standard deviation of income.

(c) (half a point) How many applicants are in this subset?

(d) (half a point) How many of the applicants in the subset are married?

(e) (half a point) How many applicants in the subset own a home?

2. On a typical working day, your team is able to process 120 credit card applications. To simulate this situation you draw a random sample of size 120 from the data subset generated in Question 1. (In order to make the results reproducible, use set.seed(201803) prior to drawing the sample.) Based on this sample of size 120, you want to test the null hypothesis that the mean income in the population is equal to 4750 USD. [hint: use the command t.test to perform a one-sample t-test to answer this question.]

(a) (1 point) Based on the result obtained, do you conclude to reject the null hypothesis of the true population mean being equal to 4750 USD?

(b) (half a point) How large is the test-statistic?

(c) (half a point) How large is the corresponding p-value?

(d) (half a point) Does the 95%-confidence interval contain the score 4.75?

3. To check whether R actually computes the right thing, you decide to double check.

(a) (1 point) You first compute the mean and standard deviation of income in your sample and report these numbers.

(b) (half a point) Next you compute the standard error of the mean by dividing the standard deviation of your sample by the square root of the sample size.

(c) (1 point) Finally, you compute the test statistic t which is the ratio of the difference between sample mean and hypothetical value and the standard error of the mean.

4. Now, you compare the empirical results with the corresponding theoretical distribution.

(a) (1 point) Compute the 2.5% quantile and the 97.5% quantile of the t-distribution with 119 degrees of freedom. Does the test statistic fall inside this range?

(b) (1.5 points) Compute the probability that a random variable that follows a t-distribution with 119 degrees of freedom takes on values that are in absolute values larger than the observed test-statistic i.e. P ($T \geq |2.1488|$).

5. Now, you simulate a full years work of your team, by drawing a total of 220

10. (2.5 points) Visualise the previous results. Draw a plot for the pdf of the t-distribution with the adequate number of degrees of freedom for the test statistic t. Color the areas under the pdf for all values smaller than t and larger than t.