**There are 6 questions (27 points).**

* ***No collaboration is allowed.***
* **Your solutions should be compiled in this word file, and be submitted to Canvas by the due time.**
* For multiple choice questions, there is only one correct answer. Choose the one that is closest to your answer.
* You may receive partial credits by showing your work even if your final answer is incorrect.

1. (6 pts) Bags of a certain brand of tortilla chips claim to have a net weight of 14 oz. Net weights actually vary slightly from bag to bag. A representative of a consumer advocate group wishes to see if there is any evidence that the mean net weight is less than advertised and so intends to test the hypotheses *H*0: ** 14, *H*a: ** > 14. To do this, he selects 25 bags of tortilla chips of this brand at random and determines the net weight of each. Data can be found in the Excel sheet with name “***tortilla\_chips***” of the exam data file.

(1) Make a histogram for the weights of the 25 bags. Use intervals: [13.50-13.59], [13.60-13.69], [13.70-13.79], …, [14.30-14.39]. (2pts)

(2) What is the value of the test statistic (keep 4 decimal places)? (2pts)

(3) What is the p-value of the test (keep 4 decimal places)? (2pts)

2. (3 pts) The profit in merging two companies often lies in eliminating redundant staff. The merger of two pharmaceutical companies (call them A and B) allows senior management to eliminate one of their sales forces. Which one should the merged company retain? To help reach a decision, we can compare the average level of sales obtained during a recent period by the two sales forces. Both sales forces market similar products and were organized into 20 comparable geographical districts. For each district, the data (***sales\_force***) give the average dollar sales per representative per day in that district. Data can be found in the Excel sheet with name “***sales\_force***” of the exam data file.

(1) Which of the following is the appropriate procedure to compare the performance between the two sales forces? (1pts)

A. One sample z-test

B. One sample t-test

C. Independent two-sample t-test

D. Matched-pairs t-test

E. Chi-squared test of independence

F. Chi-squared test of goodness of fit

G. One-way ANOVA

(2) What is the p-value for the hypothesis test ? Here, is the mean sales of sales force A, and is the mean sales of sales force B. (2pts)

3. (6 pts) Fashion buyers for department stores try to anticipate which colors will be popular during the next season. Getting the reaction of a sample of consumers can help. A department store has sampled customers from its operations in the East. For logistical reasons, it has a separate sample of customers from its western operations. Each customer was shown two designs for the coming fall season. One design features red fabrics, the other features violet. Then the customer was asked which color he/she preferred. If customers in the two regions have similar preferences, buyers can order the same line of clothing for both districts. Otherwise, the buyers will have to do a special order for each district. Data can be found in the Excel sheet with name “***Colors\_Regions***” of the exam data file.

(1) (1 pt) To study whether customer’s preferred color depends on the region of the customer, what is the appropriate model to be used? \_\_\_\_\_\_\_\_\_\_\_

A. One sample z-test

B. One sample t-test

C. Independent two-sample t-test

D. Matched-pairs t-test

E. Chi-squared test of independence

F. Chi-squared test of goodness of fit

G. One-way ANOVA

(2) State the null hypothesis of the test. (1 pt)

(3) What is the value of the test statistic? (2 pts)

(4) What is the p-value of the test? (2 pts)

4. (4 pts) Movie studios often release films into selected markets and use the reactions of audiences to plan further promotions. In the data (***movie\_reviews***), viewers rate the film on a scale that assigns a score from 0 (dislike) to 100 (great) to the movie. The viewers are located in one of three test markets: urban, rural, and suburban. The groups vary in size. Data can be found in the Excel sheet with name “***movie\_reviews***” of the exam data file.

(1) What is the null hypothesis for testing if there is any association between movie ratings and markets? (2 pts)

(2) What is the p-value of the test? At a significance level of 0.05, what is the conclusion of the test? (2 pts)

5. (4 pts) The data set (***Softdrink)*** includes the brands of soft drinks sold at a convenience store. Do you think that sales are equally divided among those brands? Use a hypothesis test to answer the question. State the null and alternative hypotheses, and calculate the test statistic and the p-value. Data can be found in the Excel sheet with name “***Softdrink***” of the exam data file.

6. (4 pts) Managers collected a sample of invoices. The counts of four leading digits (1,2,3,and 4) are summarized in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Leading Digit** | | | |
| 1 | 2 | 3 | 4 |
| Count | 323 | 157 | 131 | 92 |

The managers want to test whether the distribution of the four leading digits in the sample follows Benford’s Law as follows.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Leading Digit (Benford’s Law)** | | | |
| 1 | 2 | 3 | 4 |
| Proportion | 0.431 | 0.252 | 0.179 | 0.139 |

(1) Calculate the test statistic. (2 pts)

(2) What is the p-value of the test? What is the conclusion of the test with ? (2 pts)