Basic Statistical Concepts – Part II - Exercises

1. A corporation has 272 accounts receivable in a particular category. A random sample of 50 of them was selected. The sample mean was \$492.36, and the sample standard deviation was \$149.92.

    a. Find a 99% confidence interval for the population mean value of these accounts receivable.

    b. Find a 95% confidence interval for the total value of these accounts receivable.

    c. Without doing the calculations, state whether a 90% confidence interval for the population total would be wider or narrower than the interval found in part (b).

2. A corporation employs many sales representatives. A random sample of 60 of them was taken, and it was found that, for 36 of the sample members, the volume of orders taken this month was higher than for the same month last year.

    a. Find a 95% confidence interval for the population proportion of sales representatives with a higher volume of orders this month than in the same month last year. Use both methods and compare the results.

    b. Use R to define a function that produces the more accurate confidence interval.

3. A machine being used for packaging seedless raisins has been set so that on the average 15 ounces of raisins will be packaged per box. The quality control engineer wishes to test the machine settings and selects a sample of 30 consecutive raisin packages filled during the production process. Their weights are recorded below:

15.2	15.3	15.1	15.7	15.3	15.0	15.1	14.3	14.6	14.5
15.0	15.2	15.4	15.6	15.7	15.4	15.3	14.9	14.8	14.6
14.3	14.4	15.5	15.4	15.2	15.5	15.6	15.1	15.3	15.1

    Assume that the weight per box is normally distributed.

    a. Is there evidence that the mean weight per box is different from 15 ounces? (use $\alpha = 0.05$).

    b. Is there evidence that the standard deviation of weight per box is different from 0.25 ounces? (use $\alpha = 0.05$).

4. Stocks on the National Association of Security Dealers (NASD) system were analyzed in *Financial Analysts Journal* (Jan/Feb 1993). The annualized monthly returns (%) for a sample of 13 large-firm NASD stocks were computed and are summarized as follows: $\bar{x} = 13.5\%$, $s = 23.84$. Conduct a test of hypothesis to determine whether the mean annualized monthly return for large-firm NASD stocks exceeds 10%. Use $\alpha = 0.05$.

5. The management of the Tiger baseball team decided to sell only low alcohol beer in their ballpark to help combat rowdy fan conduct. They claimed that more than 40% of the fans would approve of this decision. Let *p* equal the proportion of Tiger fans on opening day that approved of this decision. Knowing that 550

fans out of a sample of 1278 said that they approved of this new policy what can you conclude?

6. A sample of 45 sales receipts from the university bookstore has $\bar{x} = 73.5$ and $s = 12.4$. Assume that sales receipts follow a normal distribution

   a. Use these values to perform a test of $H_0 : \mu = 80$ against $H_1 : \mu < 80$ with $\alpha = 0.05$. Calculate the p-value.

   b. Test $H_0 : \mu = 80$ against $H_1 : \mu \neq 80$ with $\alpha = 0.05$. Calculate the p-value. Why does the p-value change from the previous one.

   c. Define a 95% confidence interval for μ.

7. Suppose you want to compare the mean daily sales of two restaurants located in the same city. If you were to record the restaurants' total sales for each of 12 randomly selected days during a 6 months period, the results might appear as shown in file restaurants.csv.

   a. Assuming independent samples, do these data provide evidence of a difference between the mean daily sales of the two restaurants?

   b. Now consider that we have a paired sample. Is your conclusion the same?

   c. Discuss which solution is more adequate.

8. The US Office Management and Budget (OMB) requires government agencies to produce annual performance and accounting reports (PARS) each year. A research team at George Mason University evaluated the quality of the PARS for 24 government agencies (*The Public Manager*, Summer 2008) where evaluation scores ranged from 12 (lowest) to 60 (highest). The file pars.csv contains evaluations for all 24 agencies for 2 consecutive years. Data for a random sample of 5 of these agencies are shown in the following table

Agency	Year 1 score	Year 2 score
GSA	34	40
Agriculture	33	35
Social Security	33	33
USAID	32	42
Defense	17	32

   Suppose you want to conduct a paired difference test to determine whether the true mean evaluation score of government agencies in year 2 exceeds the true mean evaluation in year 1.

   a. Explain why the data should be analyzed using a paired difference test.

   b. Compute the test using a calculator for the five observations and conclude.

   c. Compute the test using the sample presented in file pars.csv and conclude.

9. The accompanying table shows percentage changes in the Dow-Jones index over the five first trading days of each year (x) of thirteen years, and also the corresponding percentage changes (y) in the index over the whole year.

x	1.5	0.2	-0.1	2.8	2.2	-1.6	-1.3	5.6	-1.4	1.4	1.5	-4.7	1.1
y	14.9	-9.2	19.6	20.3	-3.7	27.7	22.6	2.3	11.9	27.0	-4.3	20.3	4.2

   a. Calculate the sample correlation coefficient (Pearson)

   b. Test, against a two-sided alternative, the null hypothesis that the population correlation is 0

10. File wage.csv present information about wages and other characteristics of a random samples of workers:

   - Wage – monthly earnings
   - Hours – average weekly hours
   - IQ – IQ score
   - Educ – number of years of education
   - Exper – years of work experience
   - Age – age in years
   - Married - =1 if married, 0 otherwise

   a. Test if the mean salary in the population is greater for married workers than for non-married in the population.

   b. Test if the correlation coefficient between wage and Age is positive in the population.

Answers

1- a) (453.54; 549.18) b) (122332.9; 145511.0) c) narrower

2- a) Usual: (0.476; 0.724) Alternative: (0.474; 0.714) without corr b) …

3- a) p-value=0.1369 do not reject H0 or $T_{obs} = 1.5298$, $c = \pm 2.045$

   b) p-value=0.0000 reject H0 or $Q_{obs} = 76.395$, $q_1 = 16.047$, $q_2 = 45.722$

4- p-value=0.3031 do not reject H0 or $T_{obs} = 0.5293$, $c = 1.7823$ (one side test)

5- p-value=0.01336 Reject H0

6- a) p-value=0.00051 b) p-value=0.001028 c) (69.7746; 77.2254)

7- a) p-value=0.7047 b) p-value=0.0000 c) …

8- a) … b) p-value=0.0359 c) p-value=0.0414

9- a) *-0.4066* b) p-value=0.1679

10- a) p-value=0.0000 NO b) p-value=0.0000