Mathematical Statistics. Home Assignment 2.

Hypotheses Testing.

MDI, Fall 2022.

1. A random sample of ten students found the following figures, in hours, for time spent studying in the week before the final exams:

   $$28 \quad 57 \quad 42 \quad 35 \quad 61 \quad 39 \quad 55 \quad 46 \quad 49 \quad 38.$$

   Assume that the population distribution is normal.

   (a) Find the sample mean and sample standard deviation.

   (b) Test at the 5% significance level the null hypothesis that the population mean is 40 against the alternative that it is higher.

2. If you live in California, the decision to purchase earthquake insurance is a critical one. An article in the Annals of the Association of American Geographers (June 1992) investigated many factors that California residents consider when purchasing earthquake insurance. The survey revealed that only 133 of 337 randomly selected residences in Los Angeles County were protected by earthquake insurance.

   (a) What are the appropriate null and alternative hypotheses to test the claim that less than 40% of the residents of Los Angeles County were protected by earthquake insurance?

   (b) Does the data provide sufficient evidence to support the null hypothesis? (Use $\alpha = 0.10$)

3. The data in the following table show the numbers of daily parking offences in two areas of a city. The day identifications are unknown and the recordings were not necessarily made on the same days. Is there evidence that the areas experience different mean numbers of offences? State the needed assumptions and perform hypotheses test.

Area A	Area B
38	32
38	38
29	22
45	30
42	34
33	28
27	32
32	34
32	24
34	no data

4. Random sample of 1562 undergraduates enrolled in marketing courses were asked to respond on a scale from one to seven to the proposition 'Advertising helps to raise our standard of living.' The sample mean response was 4.27 and the sample standard deviation was 1.32. Test at the $\alpha = 0.01$ level, against a two-sided alternative, the null hypothesis that the population mean is 4.

5. The American Hospital Association reports in Hospital Statistics that the mean cost to general community hospitals per patient per day in U.S. hospitals was \$951 in 1998. In that same year, a random sample of 30 daily costs in New York City hospitals yielded a mean of \$1185. Assuming a population standard deviation of \$333 for New York City hospitals, do the data provide sufficient evidence to conclude that in 1998 the mean cost in NYC hospitals exceeded the national mean of \$951? Perform the required hypothesis test at the 5% significance level.

6. Suppose a student is asked to measure the content of the iron in the water in the lake Brown, and test the hypothesis $H_0 : \mu = 10$ against the alternative $H_1 : \mu > 10$. He made 6 measurements and has found that P-value of the test is 9%. By mistake he used z-statistic instead of t-statistic. What is the actual P-value of the test?

7. Two surveys were evaluated in Moscow and Tver. From sample of 200 persons in Moscow 125 were against smoking at restaurants. In Tver 52 from sample of 100 were against smoking in restaurants. Let p_1 and p_2 are population proportions of persons who are against smoking in Tver and Moscow, respectively.

   (a) Construct two-sided 95% confidence interval for $p_2 - p_1$.

   (b) At 5% significance level test null hypothesis $H_0 : p_1 = p_2$ against $H_1 : p_2 > p_1$

   (c) At 2.5% significance level test null hypothesis $H_0 : p_2 = 0.55$ against $H_1 : p_2 > 0.55$.

Bonus

How much do seat belts help? To answer this, a study was undertaken of cars that had had been equipped with seat belts (lap-and-shoulder belts) and that had subsequently been involved in accidents. A random sample of $10,000$ occupants showed the following injury rates (reconstructed from U.S. *Department of Transportation*, 1981):

	Seat Belt Worn		
Severe or Fatal injury	Yes	No	Totals
Yes	3	119	122
No	829	9049	9878
Totals	832	9168	10000

1. State H_0 in words and symbols. How would you check the positive effect of using seat belts?

2. Construct appropriate confidence interval. Explore behaviour of the boundaries with different confidence levels: 90%, 95%, 99%. What inferences can you make?

3. Perform the hypotheses test with different significance levels $\alpha = 0.1, 0.05, 0.01$. Comment on the results.