MSIS 5303 – Statistics for Data Science – Fall 2021 - Assignment 7

Solution

1) It is claimed that the distance of fly balls hit to the outfield (in baseball) is normally distributed with a mean of 250 feet and a standard deviation of 50 feet. Suppose you believe that the mean distance is greater than 250 feet. You randomly sample 49 fly balls and find that the average distance for your sample was 258 feet. You wish to test your belief at $\alpha = 0.05$ level.
    a. State the null and alternate hypotheses clearly.
    b. Conduct the hypothesis test based on the *test statistic* and *critical value(s)*. Clearly indicate each.
    c. What is the **p-value**? Use the **p-value** to conduct the same test
    d. Report your conclusion in words, in the context of the problem.

Solution:

a. *Ho: $\mu = 250$ feet, Ha: $\mu > 250$ feet.*
b. Because the population is normally distributed with known standard deviation, the sampling distribution of $\overline{X}$ ~ Normal (**250**,**50/√49**) under the null hypothesis.
Test statistic: $z = (258 - 250)/(50/\sqrt{49}) = 1.12$
Critical Value: $z_{0.05} = 1.645$
Since the Text-statistic < Critical Value, we ***fail to reject the null hypothesis.***
c. The p-value is the **area to the right** of the test-statistic =1.12 on a standard normal distribution = 0.131. Since $p > (\alpha = 0.05)$, *we* ***fail to reject the null hypothesis.***
d. ***We fail to reject the null hypothesis that the mean distance of fly balls hit to the outfield is 250 feet.***

2) Suppose it is claimed that the mean weight of a bag of the same brand of candies is 0.13 ounces. You wish to show that it is not 0.13 ounces and wish to test the claim at $\alpha = 0.01$ level. You collected a sample of 16 small bags of the same brand of candies. The weight of each bag was then recorded. The mean weight was two ounces with a standard deviation of 0.12 ounces. Assume that the population distribution of bag weights is normal with a known population standard deviation of 0.1 ounce.
    a. State the null and alternate hypotheses clearly.
    b. Conduct the hypothesis test based on the *test statistic* and *critical value(s)*. Clearly indicate each.
    c. What is the **p-value**? Use the **p-value** to conduct the same test
    d. Report your conclusion in words, in the context of the problem.

Solution:

a. *Ho: $\mu = 0.13$ ounces, Ha: $\mu \neq 0.13$ ounces.*
b. Because the population is normally distributed with known standard deviation, the sampling distribution of $\overline{X}$ ~ Normal (**2**,**0.12/√16**) under the null hypothesis.

Test statistic: z = (2 – 0.13)/(0.1/√16) = 74.8
Critical Values: $z_{0.005}$ = (-2.576, 2.576) for the two-tailed test.
*Since the |Text-statistic| > |Critical Value|, we **reject the null hypothesis.***

c. The p-value is the **area to the left** of -2.8 **+ area to the right** of +2.8 (the test statistic) on a standard normal distribution = 0.000. Since p < (α = 0.05), *we **reject the null hypothesis.***
d. ***We reject the null hypothesis that the mean weight of a bag of the same brand of candies is 0.13 ounces.***

3) A survey of the mean number of cents off that coupons give was conducted by randomly surveying one coupon per page from the coupon sections of a recent San Jose Mercury News. The following data were collected: 20¢; 75¢; 50¢; 65¢; 30¢; 55¢; 40¢; 40¢; 30¢; 55¢; $1.50; 40¢; 65¢; 40¢. Assume the underlying distribution is approximately normal. You wish to conduct a hypothesis test (α = 0.05 level) to determine if the mean cents off for coupons is less than 50¢.
   a. State the null and alternate hypotheses clearly.
   b. Conduct the hypothesis test based on the *test statistic* and *critical value(s)*. Clearly indicate each.
   c. What is the **p-value**? Use the **p-value** to conduct the same test
   d. Report your conclusion in words, in the context of the problem.

<u>**Solution:**</u>

a. *Ho: μ = 50 cents, Ha: μ < 50 cents.*
Because the population is normally distributed, but with unknown standard deviation, the sampling distribution of $\bar{X}$ **~ t(13 df),** <u>under the null hypothesis</u>, **because the sample size is small (< 30)**
b. *Test statistic: z = (53.93 – 50)/(* 31.63/√14) = 0.464
Critical Values:- $t_{(13, 0.05)}$ = -1.771
*Since the Text-statistic > Critical Value, we **fail to reject the null hypothesis for the lower-tailed test.***
c. The p-value is the **area to the left** of 0.464 on a t- distribution with 13 df = 0.6750 (from tables we know it will be between 0.6 and 0.75). Since p > (α = 0.05), *we **fail to reject the null hypothesis.***
d. ***We fail to reject the null hypothesis that the mean number of cents off coupons is (no more than) 50 cents.***

4) Researchers are concerned about the impact of students working while they are enrolled in classes, and they'd like to know if students work too much and therefore are spending less time on their classes than they should be. The researchers think that the students work more than 7 hours on average. They know from previous studies that the standard deviation of this variable is about 5 hours. A survey of 200 students provides a sample mean of 7.10 hours worked. You wish to test the researchers' belief at α = 0.05 level.
   a. State the null and alternate hypotheses clearly.
   b. Conduct the hypothesis test based on the *test statistic* and *critical value(s)*. Clearly indicate each.
   c. What is the **p-value**? Use the **p-value** to conduct the same test
   d. Report your conclusion in words, in the context of the problem.

e. Suppose it is desired to make the hypothesis test sensitive enough to detect an effect size of 0.1 with a power of 0.9. What sample size would be needed?

Solution:

a. *Ho: μ = 7 hours, Ha: μ > 7 hours.*
b. We have an unknown population distribution, with known standard deviation, and sample size > 30, so we can use the Central Limit Theorem for the sampling distribution of $\overline{X}$ ~ Normal (**7**, **5/√200**) under the null hypothesis.
Test statistic: z = (7.10 – 7)/(5/√200) = 0.283
Critical Value: $z_{0.05}$ = 1.645
*Since the Text-statistic < Critical Value, we **fail to reject the null hypothesis.***
c. The p-value is the **area to the right** of the test-statistic =0.283 on a standard normal distribution = 0.389. Since p > (α = 0.05), *we **fail to reject the null hypothesis.***
d. ***We fail to reject the null hypothesis that students work (no more than) 7 hours a week.***

5) Public Policy Polling recently conducted a survey asking adults across the U.S. about music preferences. When asked, 80 of the 571 participants admitted that they have illegally downloaded music. You believe that less than 15% of adults across the US download music illegally. Conduct the test at α = 0.01 level.
   a. State the null and alternate hypotheses clearly.
   b. Conduct the hypothesis test based on the *test statistic* and *critical value(s)*. Clearly indicate each.
   c. What is the **p-value**? Use the **p-value** to conduct the same test
   d. Report your conclusion in words, in the context of the problem.

Solution:

a. *Ho: p = 0.15, Ha: p < 0.15.*
b. We use the normal approximation to the Binomial for proportions; p' ~ Normal (**0.15**, **√(0.15)(0.85)/571**) under the null hypothesis.
Test statistic: z = (80/571 – 0.15) / (√(0.15)(0.85)/571) = -0.66
Critical Value: $z_{0.05}$ = -2.231
*Since the Text-statistic > Critical Value, **we fail to reject the null hypothesis.***
c. The p-value is the **area to the left** of the test-statistic =-0.67 on a standard normal distribution = 0. P-value = 0.251. Since p > (α = 0.01), *we **fail to reject the null hypothesis.***
d. ***We fail to reject the null hypothesis that 15% of adults across the US download music illegally.***