Data Analytics

SESSION 3: STATISTICS (CONTD.)

Assignment 2

Data Analytics

Table of Contents

1.Introduction .............................................................................................................................................3

2.Objective ..................................................................................................................................................3

3.Prerequisites ............................................................................................................................................3

4.Associated Data Files ...............................................................................................................................3

5.Problem Statement ..................................................................................................................................3

6.Expected Output ......................................................................................................................................3

7.Approximate Time to Complete Task ......................................................................................................3

Data Analytics

**1. Introduction**

This assignment will help you understand the concepts learnt in the session.

**2. Objective**

This assignment will test your skills on the concepts of statistics.

**3. Prerequisites**

Not applicable.

**4. Associated Data Files**

Not applicable.

**5. Problem Statement**

**Practical Application of CLT**

1. Engineers must consider the breadths of male heads when designing motorcycle helmets for men. Men have head breadths that are normally distributed with a mean of 6.0 inches and a standard deviation of 1.0 inch

Ans: Write down the given information: > mu = 6 > sigma = 1

1. If one male is randomly selected, what is the likelihood that his head breadth is less than 6.2 inches?

Ans: Find P(x < 6.2) using the normal distribution and the given parameters: > p = pnorm ( 6 . 2 , mean = mu, sd = sigma ) > s i g n i f ( p , 3 ) [ 1 ] 0. 5 7 9

Actual mean we have is 6.0. New breadth is 6.2. Difference between the two mean is (6.2-6.0)/1.0=0.2

Now let’s check for the probability of head breadth is less than 6.2

P(x<6.2)=P(z<0.2) => 0.5793.

1. The Safeguard Helmet Company plans an initial production run of 100 helmets. How likely is it that 100 randomly selected men have a mean head breath of less than 6.2 inches?

Ans: - Find P (¯x < 6.2) using the normal distribution for the sampling distribution of ¯x (since the CLT applies). The standard deviation will be the standard error:

Now n=no or randomly selected men or n=100

Formula for standard error = sigma/sqrt(n) or 1.0/ = 0.1

Now let us find the value of mean

(6.2-6.0)/0.1=0.2/0.1 =2

Now let us check for probability

P(z<2.0)= 0.9772

1. The production manager sees the result in part b and reasons that all helmets should be made for men with head breadths of less than 6.2 inches, because they would fit all but a few men. What is wrong with that reasoning?

Ans:- The Probability of concerning means does not apply to individuals. First Part (a) is relevant since helmets will be worn by one man at a time.

Probability of an individual has a head breadth greater than 6.2)= 1-0.5793 =0.4207.

Hence, it is clear from above equation that nearby 42 men out of 100 would not find a helmet that fits.

**Two-tailed Test Of Population Mean With Known Variance**

1. Suppose the mean weight of King Penguins found in an Antarctic colony last year was 15.4 kg. In a sample of 35 penguins same time this year in the same colony, the mean penguin weight is 14.6 kg. Assume the population standard deviation is 2.5 kg. At .05 significance level, can we reject the null hypothesis that the mean penguin weight does not differ from last year?

Ans: - The null hypothesis of the two-tailed test of the population mean can be expressed as follows:

Where is a hypothesized value of the true population mean μ.

Let us define the test statistic t in terms of the sample mean, the sample size and the sample standard deviation s: t= (x-)/(s/)

Then the null hypothesis of the two-tailed test is to be rejected if t ≤−tα∕2 or t ≥ tα∕2, where tα∕2 is the 100(1 − α) percentile of the Student t distribution with n − 1 degrees of freedom.

Now let’s us apply null hypothesis on our problem for penguins

The null hypothesis is that μ = 15.4. We begin with computing the test statistic.

The null hypothesis is that μ = 15.4. We begin with computing the test statistic.

> xbar = 14.6            # sample mean   
> mu0 = 15.4             # hypothesized value   
> s = 2.5                # sample standard deviation   
> n = 35                 # sample size   
> t = (xbar−mu0)/(s/sqrt(n))   
> t                  # test statistic   
[1] −1.8931

We then compute the critical values at .05 significance level.

> alpha = .05   
> t.half.alpha = qt(1−alpha/2, df=n−1)   
> c(−t.half.alpha, t.half.alpha)   
[1] −2.0322  2.0322

Based on above equation we can say that the test statistic -1.8931 lies between the critical values -2.0322, and 2.0322. Hence, at 0.05 significance level, we do not reject the null hypothesis that the mean penguin weight does not differ from last year.

**6. Expected Output**

N/A Data Analytics

4

**7. Approximate Time to Complete Task**