Logo, company name

Description automatically generated

**SanFrancisco Salaries 2014**

**Final Project Report**

**DSCI 5180 Section 002– Introduction to Business Decision Process(8W1)**

**Manisha Reddy Kotha**

Performed the operations from the module 1 to module 4 that includes normal distribution, confidence interval estimation, inferences from data (hypotheses testing), more Inferences from data (Multiple samples) using statistical techniques.

**Module 1**

**Assuming the BasePay in San Francisco in 2014 is distributed normally. The BasePay in San Francisco are known to have a mean of 91173.994 and a standard deviation of 33217.39.What is the probability that the BasePay, on average, less than 95000?**

|  |
| --- |
| μ=91173.994, σ=33217.39, x = 95000  The probability that the BasePay on average, less than 95000is given by  P(X<95000) |
|  |

P((x−μ)/σ < (95000-91173.994)/33217.39) = P (z < 0.1151) = **0.5458**

Hence, the probability that the BasePay on average, less than 95000 is 0.5458.

**Module 2**

**Construct 95% confidence interval for the mean TotalPayBenefits of San Francisco.**

The confidence interval for a population mean is given by

A picture containing arrow

Description automatically generated

where, sample mean x‾ : 138528.4128

σ : 47879.22197

n : 22334

the margin of error of a confidence interval for a population mean is given by

A picture containing diagram

Description automatically generated

where zα/2 is the critical value for the level of confidence, c=1−α

σ is the population standard deviation, and n is the sample size.

c=1- α

α= 1-0.95 = 0.05

α/2 = 0.025

Using the table of standard normal critical values or technology, we find that the critical value is zα/2 = 1.96.

x‾±E = 138528.4128 ± 1.96(47879.22197/ √22334)

So, we are 95% confident that the true mean TotalPayBenefits of San Francisco is between **1,37,900.47 and 1,39,156.35**

**Module 3**

**The reporter data claims that the TotalPay in the year 2014 is greater than 103000. The mean is 103505.76 and the standard deviation is 40722.929, Assume level of significance is 0.01 will be used. n= 22334.**

**a) Test the below hypothesis.**

**H0 μ <= 103000**

**Ha μ > 103000**

**b) Determine If there is sufficient evidence to conclude that the average TotalPay is different from what the reporter has indicated?**

**Step1:** Hypothesis:

H0 μ <= 103000

Ha μ > 103000

**Step2:** Test statistic:

**Diagram

Description automatically generated with low confidence**

Using technology, z = **1.856**

**Step3:** t Critical value:

Since the symbol used in the alternative hypotheses is ">", we know that the test is **right-tailed**, and we will reject the null hypothesis, H0 if t>tα. Using the technology, we find that the critical value, tα for the level of significance, α=0.01and the degrees of freedom, df= (22334-1) = 22333 is **2.326**. Thus, we will reject the null hypothesis, H0, if t > tα.

**Step4:** Conclusion:

As **1.856 < 2.326**, we will fail to reject the null hypothesis H0, hence there is **not sufficient** evidence to conclude that the average TotalPay is different from what the reporter has indicated.

**Module 4**

**Doing a Hypothesis testing to check the difference between the mean BasePay of HSA Social Worker(μ1) & mean BasePay of Hospital Eligibility Worker(μ2). Test the hypotheses** **H0: μ1 >= μ2 , Ha: μ1 < μ2. Assume 0.05 level of significance. Assuming that the population variances are equal and that the two populations are normally distributed.**

**Step1**: Hypothesis:

H0: μ1 >= μ2

Ha: μ1 < μ2

**Step2**: t Test Statistic:

A picture containing text, clock

Description automatically generated

(μ1−μ2)  is the hypothesized difference between the two population means μ1 and μ2 (the value for this difference is typically assumed to be zero, i.e., μ1−μ2=0)

x1 = 64562.195

s1 =7669.2573

n1 = 108

n2 = 72

x2 = 66820.582

s2 = 9221.9081

t test statistic = **-1.783**

**Step3:** t critical value:

df = (n1+n2-2): (108 + 72 -2) = 178

Level of significance: 0.05, this is a left tailed test.

Reject H0 if t < **-1.653**

**Step4:** Conclusion:

Since t is less than critical value, we **reject the null hypotheses** i.e., μ1 >= μ2. Hence, we can conclude that the mean BasePay of HSA Social Worker is less than mean BasePay of Hospital Eligibility Worker.

**References:**

<https://www.hawkeslearning.com/Statistics/excel.html#t-test>

<https://learn.hawkeslearning.com>

**Appendix:**

* Calculations performed using Excel are attached in separate sheets in the dataset file for each module question with the module number.
* The results are highlighted in yellow in each module sheets in the excel attached “San\_Francisco\_Salaries\_2014.xlsx”.