Statistical Inference Assignment 2

Pankhuri Kasliwal

April 12, 2020

1 Problem 1

The sample size is n = 90, number of favourable cases is 50. hence we find the sample proportion i.e. p-hat = 0.5556. (as given in the code). We check if the conditions for binomial distribution are satisfied or not. We find n*p = 63 and n*q = 27. Both these values are greater than 5 hence conditions are satisfied. (here p is the proportion of friends or acquaintances having names ending in ”ko”).

- **Null and Alternative Hypothesis** : $H_o : p \geq 0.70$ v/s $H_a : p < 0.70$.
- **Rejection Region** : We consider the significance level or alpha to be 0.05. We know that Z_α here follows standard normal distribution, so we use the function qnorm(), which gives us critical z-values corresponding to a lower tailed area. Thus we get Z_{cv} as -1.644854. Based on this information we get our rejection region as R = $\{z : z < -1.644\}$.
- **Test Statistic** : We use the formula given in the code to compute the value for our z-statistic. We get the value of Z_{stat} to be -2.990284.
- **Inference** : Since $Z_{stat} < Z_{cv}$, we reject the null hypothesis.
- **Conclusion** : The null hypothesis Ho is rejected. Therefore, there is enough evidence to claim that the population proportion p is less than p_0.

alpha	0.05
fav_cases	50
n	90
np	63
nq	27
p_hat	0.555555555555556
p_val	0.7
z_cv	-1.64485362695147
z_val	-2.99028409048353
z.alpha	-1.64485362695147

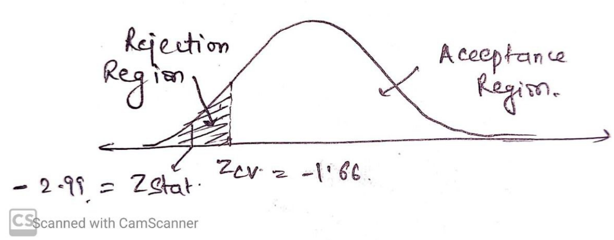

Figure 1: Computed values

Hence we can say that there is enough evidence to conclude that proportion of friends or acquaintances having names ending in ”ko” is less than 0.70 and that the proportion has dropped for this generation.

2 Problem 2

For the given data, we have sample means as 7.8 and 5.7 respectively for sample1 and sample2. Also the corresponding variances are 2.4 and 1.12. Thus we find the difference between observations and get the mean value of the difference for the population of all pairs as 2.1. The standard deviation of difference for the paired sample data is 1.287. We choose our random variable X_d to be the mean difference in work times on days when eating breakfast and on days when not eating breakfast. From the null and alternative hypothesis we can infer that it is a right-tailed test. All the assumptions i.e. samples are dependent, are SRS, pairs of values have differences that are from a population having approx normal are satisfied. The t-stat t distribution with n-1 degrees of freedom.

- **Null and Alternative Hypothesis** : $H_o : \mu_d = 2$ or $\mu_d \leq 2$ v/s $H_a : \mu_d > 2$
- **Rejection Region** : We consider the significance level or alpha to be 0.05. We know that the dof is n-1 i.e. 9, so we use the function qt(), which gives us critical t-values. Thus we get t_{cv} as 1.833113. Based on this information we get our rejection region as $R = \{t : t > 1.833\}$.
- **Test Statistic** : We use the formula given in the code to compute the value for our t-statistic. We get the value of t_{stat} to be 0.2457696.
- **Inference** : Since $t_{stat} < t_{cv}$, we accept the null hypothesis.
- **Conclusion** : The null hypothesis Ho is accepted. Therefore, there is enough evidence to claim that the difference between population mean of both the samples never greater than 2, it may be equal to 2 or less than two.

d	2
data1	num [1:10] 8 7 9 5 9 8 10 7 …
data2	num [1:10] 6 5 5 4 7 7 7 5 6…
diff	num [1:10] 2 2 4 1 2 1 3 2 0…
mu_d	2.1
mu1	7.8
mu2	5.7
n1	10L
n2	10L
sigma_1	2.4
sigma_2	1.12222222222222
sigma_d	1.28668393770792
t_cv	1.83311293265624
t_val	0.245769576155712

Figure 2: Computed values

Hence we can say that there is enough evidence to conclude that the mean difference in work times on days when eating breakfast and on days when not eating breakfast is never greater than 2 hours.