Statistical Inference

Assignment 2

Sahil Goyal

2020326

1.

Hypothesis:

Let μ_0 be the average bread height

Let:

- H_0 = The bakers claim is wrong ($\mu \leq 15$)
- H_1 = The bakers claim is correct ($\mu > 15$)

Therefore, our hypothesis becomes:

$$H_0: \mu \leq \mu_0 \text{ v/s } H_1: \mu > \mu_0$$

Here $\mu_0 = 15$

Testing:

We will use a one tailed test in order to test our hypothesis. We will make our inferences using 2 ways, test statistic and p value.

$$Z_{test} = ((\text{sample\_mean} - \mu_0) * \sqrt{n})/\sigma \sim N(0,1)$$

Here, n is the number of samples (sample size) and σ is the standard deviation.

The corresponding p value is: $P(Z > |z|)$

As this is a one tailed test, we will only get one critical value.

Result:

We get the results:

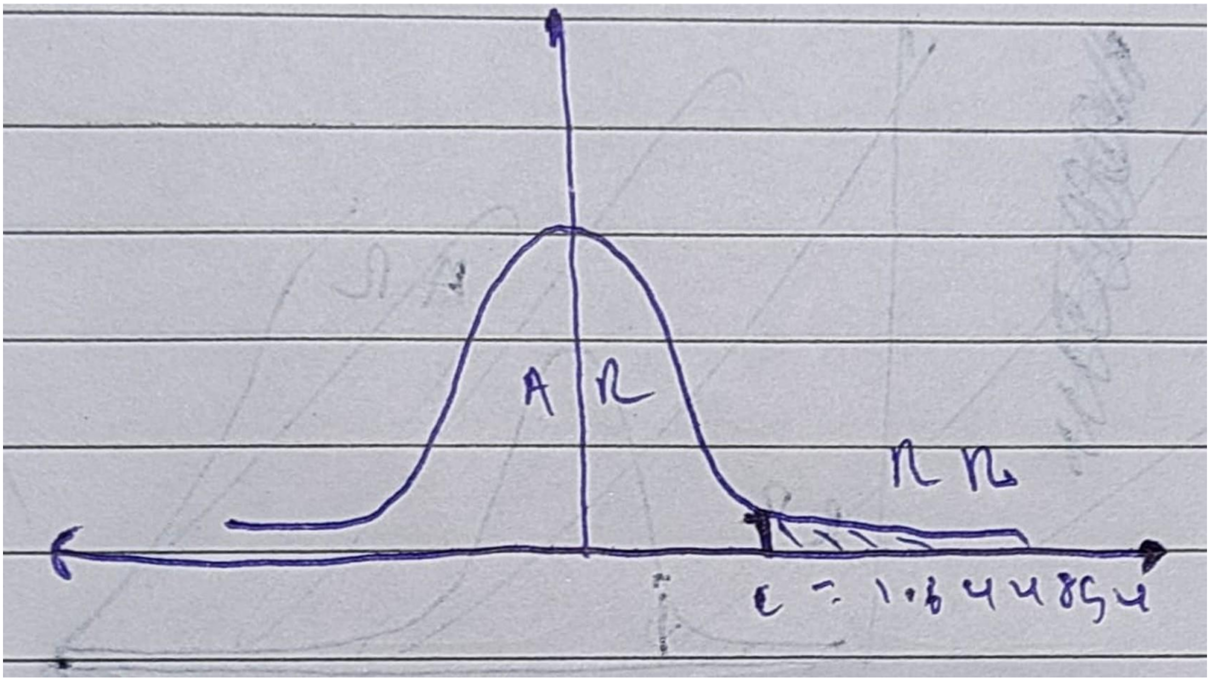
- The test statistic is: 12.64911
- The critical value is: 1.644854
- The P value is: 5.657419e-37

Inference:

We can draw the following inferences:

- As we can see, the test statistic lies in the rejection region (inferred from critical value).
- We can also see that the p value is much less than the significance level (α).
- Therefore, we can reject the Hypothesis H_0, that the bakers claim is wrong.
- Therefore, we can conclude that the bakers claim is correct and that his bread height is more than 15 cm on average.

Graph:



2.

Hypothesis:

Let μ_0 be the mean time on death row

Let:

- H_0 = Population mean time on death row is 15 years ($\mu = 15$)
- H_1 = Population mean time on death row is not 15 years ($\mu \neq 15$)

Therefore, our hypothesis becomes:

$H_0: \mu = \mu_0$ v/s $H_1: \mu \neq \mu_0$

Here $\mu_0 = 15$

Testing:

We will use a two tailed test in order to test our hypothesis. We will make our inferences using 2 ways, test statistic, and p value.

$T_{test} = ((\text{sample\_mean} - \mu_0) * \sqrt{n})/\sigma \sim t_{(n-1)}$

Here, n is the number of samples (sample size) and σ is the standard deviation.

The corresponding P value is: $P(T > |t|) = 2*P(T > t)$

As this is a two tailed test, we will get two critical values.

Result:

We get the following results:

- The test statistic is: 3.299144

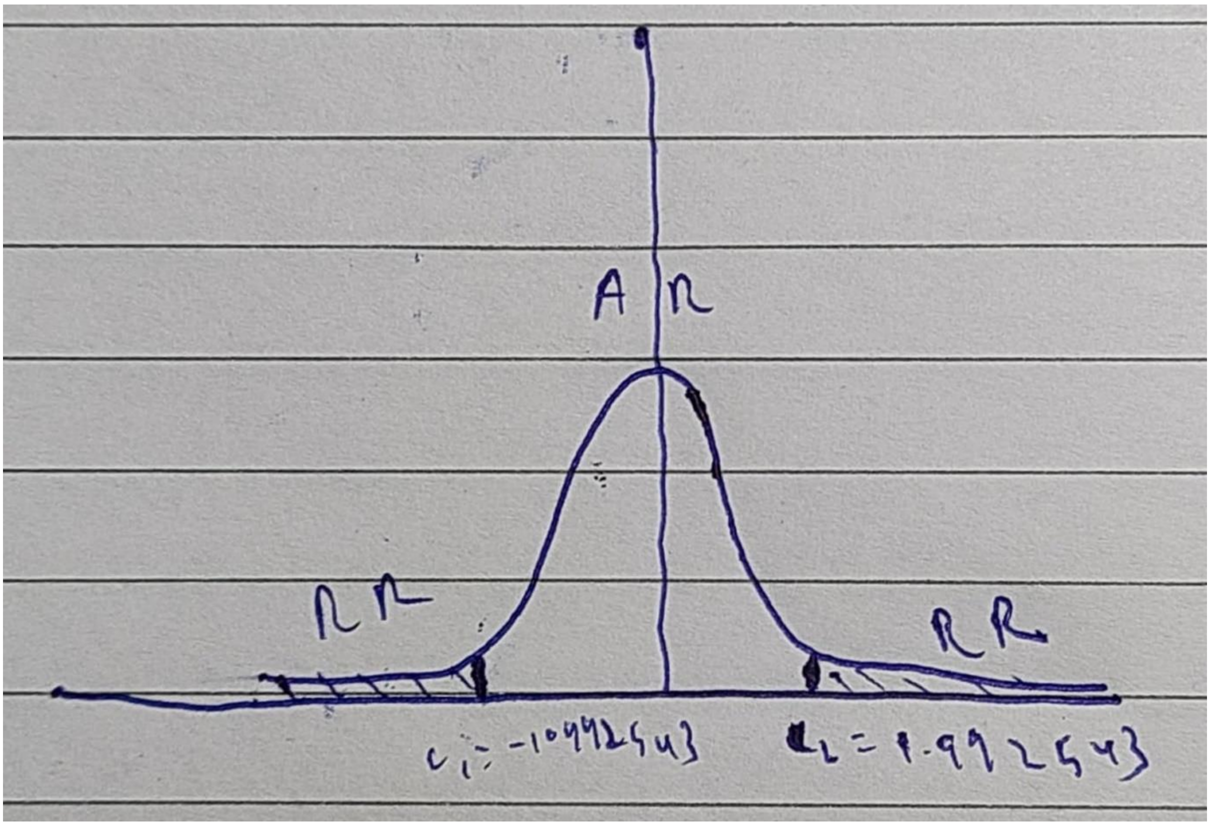
- The first critical value is: -1.992543
- The second critical value is: 1.992543
- The P value is: 0.001493164

Inference:

We can draw the following inferences:

- As we can see, the test statistic lies in the rejection region (inferred from critical value).
- We can also see that the p value is less than the significance level (α).
- Therefore, we can reject the Hypothesis H_0, that the population mean time on death row is 15 years.
- Therefore, we can conclude that the mean population time on death row likely isn't 15 years.

Graph:



3.

Assumptions:

We are assuming that the distribution of the yield of green gram is normal distribution

Hypothesis:

Let μ_0 be the average yield (in quintals per hectare)

Let:

- H_0 = The new variety of green gram has an average yield of 12 quintals/hectare ($\mu = 12$)
- H_1 = The new variety of green gram does not have an average yield of 12 quintals/hectare ($\mu \neq 12$)

Therefore, our hypothesis becomes:

$H_0: \mu = \mu_0$ v/s $H_1: \mu \neq \mu_0$

Here $\mu_0 = 12$

Testing:

We will use a two tailed test in order to test our hypothesis. We will make our inferences using 2 ways, test statistic, and p value.

T_{test} = ((sample_mean - μ_0) * √n)/σ ~ $t_{(n-1)}$

Here, n is the number of samples (sample size) and σ is the standard deviation.

The corresponding P value is: $P(T > |t|) = 2*P(T > t)$

As this is a two tailed test, we will get two critical values.

Result:

We get the following results:

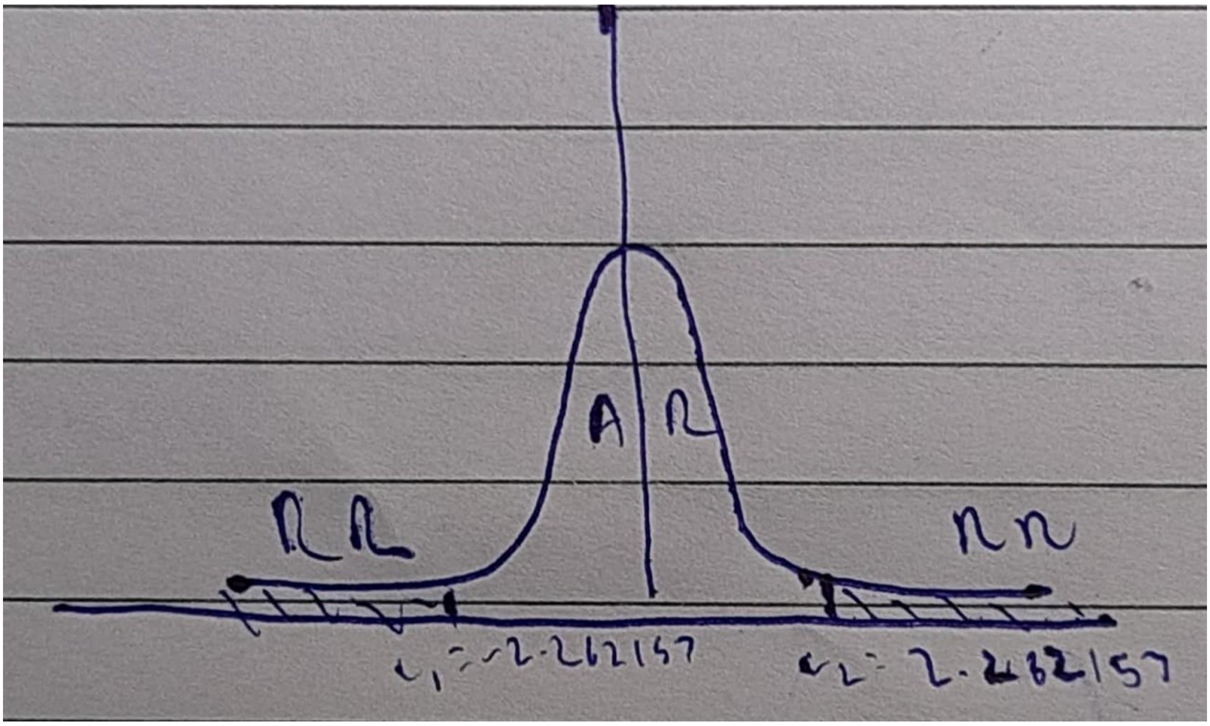
- The test statistic is: 1.835644
- The first critical value is: -2.262157
- The second critical value is: 2.262157
- The P value is: 0.09959876

Inference:

We can draw the following inferences:

- As we can see, the test statistic lies in the acceptance region (inferred from critical value).
- We can also see that the p value is greater than the significance level (α).
- Therefore, we can accept the Hypothesis H_0, that the new variety of green gram has an average yield of 12 quintals/hectare.
- Therefore, we can conclude that the new variety of green gram could likely have an average yield of 12 quintals/hectare.

Graph:



4.

Hypothesis:

Let μ_1 and μ_2 be the average amount of time boys and girls spend playing sports every day respectively.

Let:

- H_0 = The average amount of time boys and girls ages 7 through 11 spend playing sports each day is believed to be the same. ($\mu_1 = \mu_2$)
- H_1 = The average amount of time boys and girls ages 7 through 11 spend playing sports each day is believed to be different. ($\mu_1 \neq \mu_2$)

Therefore, our hypothesis becomes:

H_0: $\mu_1 = \mu_2$ v/s H_1: $\mu_1 \neq \mu_2$

Testing:

We will use a two sample two tailed test in order to test our hypothesis. We will make our inferences using 2 ways: test statistic, and p value.

T_{test} = (((sample_mean_boys – sample_mean_girls)-(μ_1 - μ_2)))/($\sigma_1^2/n_1 + \sigma_2^2/n_2$) ~ $t_{(min(n1-1,n2-1))}$

Here σ_1 and σ_2 are the standard deviations for boys and girls respectively and n_1 and n_2 are the number of samples (sample size) for boys and girls respectively.

The corresponding P value is: $P(T > |t|) = 2*P(T > t)$

As this is a two tailed test, we will get two critical values.

Result:

We will get the following results:

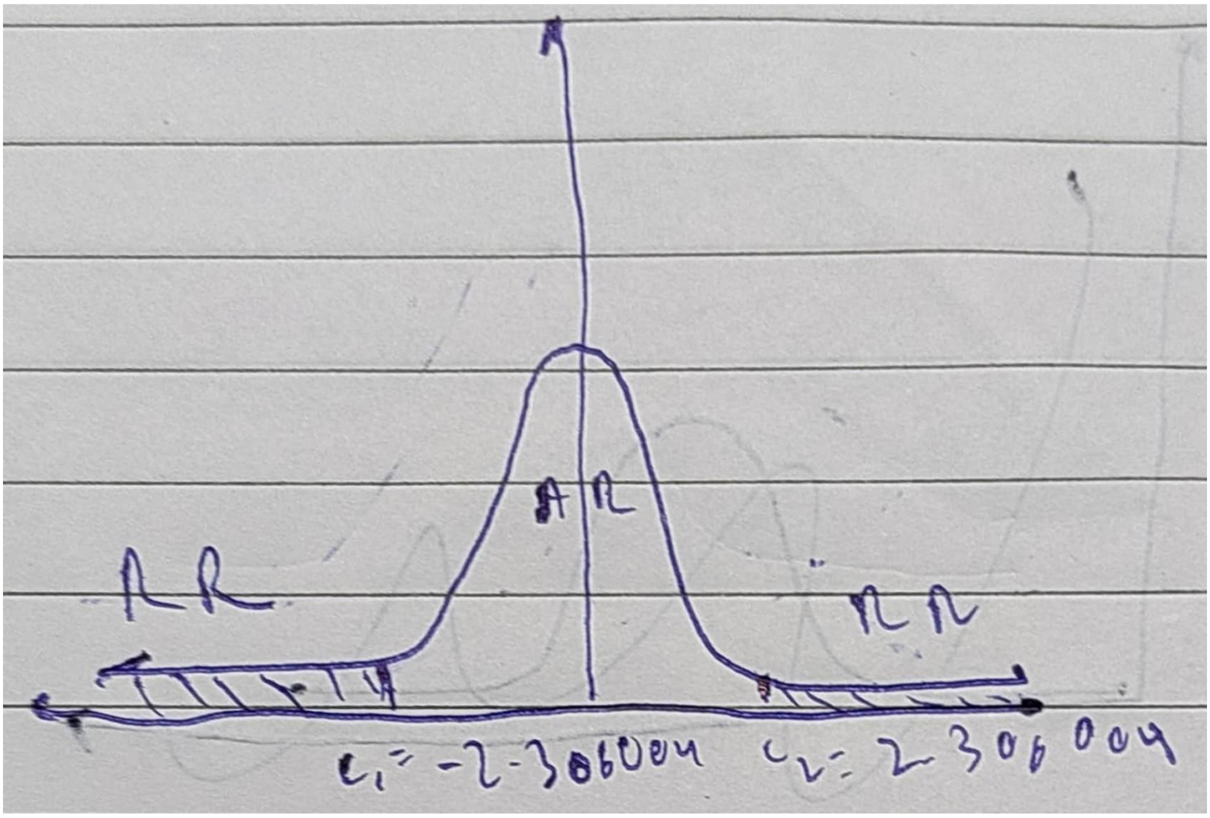
- The test statistic is: 3.142338
- The first critical value is: -2.306004
- The second critical value is: 2.306004
- The P value is: 0.01375661

Inference:

We can draw the following inferences:

- As we can see, the test statistic lies in the rejection region (inferred from critical value).
- We can also see that the p value is less than the significance level (α).
- Therefore, we can reject the Hypothesis H_0, that the average amount of time boys and girls ages 7 through 11 spend playing sports each day is the same.
- Therefore, we can conclude that the average amount of time boys and girls ages 7 through 11 spend playing sports each day is different.

Graph:



5.

Hypothesis:

Let μ_1 and μ_2 be the average change in weight in children for food A and B respectively. Let μ_d be the mean value of the difference for the population of the pairs of the data.

Let:

- H_0 = There is no average change in the weight of children due to food B ($\mu_d = 0$)
- H_1 = There is average change in the weight of children due to food B ($\mu_d \neq 0$)

Therefore, our hypothesis becomes:

H_0: $\mu_d = 0$ v/s H_1: $\mu_d \neq 0$

Testing:

As we need to check whether there is any average change in weight of children between food A and food B, we can see that the two samples are dependant.

We will use a two sample two tailed test in order to test our hypothesis. We will make our inferences using 2 ways, test statistic, and p value.

T_{test} = ((sample_mean_of_differences – μ_d) * √n)/σ ~ $t_{(n-1)}$

Here, n is the number of samples (sample size) and σ is the standard deviation, and sample_mean_of_differences is the mean of the differences of the two samples.

The corresponding P value is: P(T > |t|) = 2*P(T > t)

As this is a two tailed test, we will get two critical values.

Result:

We get the following results:

- The test statistic is: -4.320494
- The first critical value is: -2.364624
- The second critical value is: 2.364624
- The P value is: 0.003478084

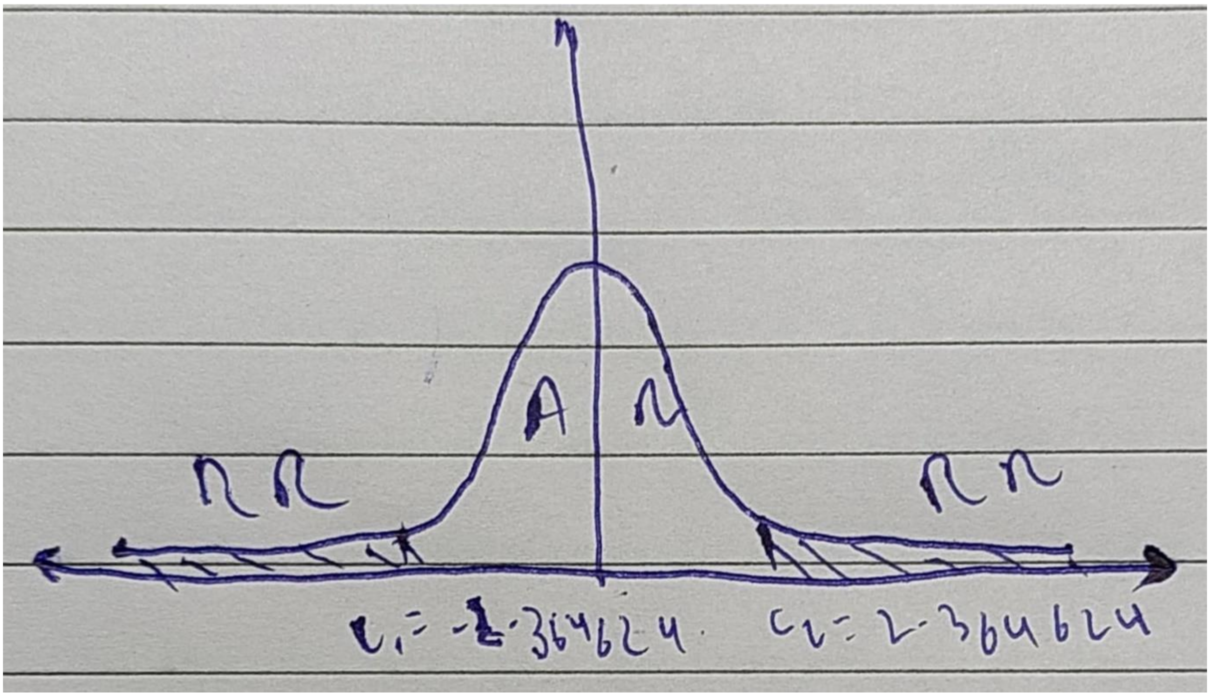
```
data:  la and lb
t = -4.3205, df = 7, p-value = 0.003478
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -3.0946083 -0.9053917
sample estimates:
mean difference
             -2
```

Inference:

We can draw the following inferences:

- As we can see, the test statistic lies in the rejection region (inferred from critical value).
- We can also see that the p value is less than the significance level (α).
- Therefore, we can reject the Hypothesis H_0, that there is no average change in the weight of children due to food B
- Therefore, we can conclude there is average change in the weight of children due to food B.

Graph:


RR
A
R
RR
$t_1 = -2.364624$
$t_2 = 2.364624$