TOPIC 1

MOVIE

Name: Akash Mandal

Roll No. : 16MA20007

Course: CS40003

Processing data

1. Converted imdb\_score to numeric data
2. Removed rows with no year information
3. Divided data into pre2016 and in2016 parts.

Population mean

1. Calculated population mean
2. µ0 = 6.442138
3. Calculated population standard deviation
4. σ = 1.113503

Code Snippet

> mu <-mean(MOVIE$imdb\_score)

> mu

[1] 6.442138

Code Snippet

> mu0 <- mean(pre2016$imdb\_score)

> mu0

[1] 6.417995

> sigma <- sd(pre2016$imdb\_score)

> sigma

[1] 1.113503

Hypothesis Testing

Case I: Population standard deviation is known

1. We will consider an upper tailed test as follows

H0 : µ = 6.442138

H1 : µ > 6.442138

1. Since standard deviation is known we shall apply the z-test

We decide the rejection region based on α = 0.05, and get the critical z0 value

zcrit = 1.644854

1. We collect sample, calculate sample mean

xbar = 6.399057

1. We calculate the z-score, and compare with zcrit

zscore = -0.1751116

Since zscore < zcrit, we fail to reject the null hypothesis.

1. Thus we cannot say that the mean IMDb score increased in the year 2016.

Code Snippet

> #population sd known, z test

>

> #critical value, alpha = 0.05

> alpha = 0.05

> zcrit = qnorm(1 - alpha)

> zcrit

[1] 1.644854

> #sample mean

> xbar <- mean(in2016$imdb\_score)

> xbar

[1] 6.399057

> #calculate zstat

> n <- length(in2016$imdb\_score)

> z = (xbar−mu0)/(sigma/sqrt(n))

> z

[1] -0.1751116