Otto Friedrich University Bamberg



EESYS-DAE-M: Data Analytics in der Energieinformatik

**Assignment - 01**

**Submitted By:**

**Kummari Daniel Raj - 1973448**

**Shashidhar Reddy Nimmagari - 1985736**

**Manasa Reddy Bandari - 1942692**

**January 14, 2021**

**Exercise a)**

**Question**: Use this formula to check whether real-time feedback significantly decreases the average energy consumption per shower.

**H0 : Real-time feedback significantly decreases energy consumption per shower?**

**H1 : Real-time feedback significantly increases energy consumption per shower?**

**r code---**

|  |
| --- |
| # split in Baseline und no Baseline Phase  data\_with\_baseline <- shower\_data %>%dplyr::filter(Shower <= 10)  data\_no\_baseline <- shower\_data %>%dplyr::filter(Shower > 10)  # built two dataframes: First dataframe = Control group, Second dataframe = Experimental group  data\_control <- data\_no\_baseline %>% filter(Showertime <= 20)  data\_feedback <- data\_no\_baseline %>% filter(Showertime > 20)  # group by Hh\_id for control and feedback group  data\_control\_Energy <- data\_control %>% group\_by(Shower) %>%  summarise(SumEnergy = sum(Energy),  avgEnergy = mean(Energy))  data\_feedback\_Energy <- data\_feedback %>% group\_by(Shower) %>%  summarise(SumEnergy = sum(Energy),  avgEnergy = mean(Energy))  # two-sample and two-sided t-test of average energy of each shower in the according showertime  t.test(data\_control\_Energy$avgEnergy, data\_feedback\_Energy$avgEnergy, mu = 0, conf.level = .95)  # Hence Reject Null Hypothesis |

**Answer:** The joined showers table shows, that the average energy consumption,calculated using given formula.

Performed to check, average energy consumption is decrease based on the feedback for each shower. We don’t have enough confidence to accept null hypothesis. Hence we reject null hypothesis.

**Exercise b)**

**Question:** It often seems that young people are more aware of the environmental impact of their actions than older people and therefore may consume fewer resources by default: For this reason, check whether the baseline energy consumption of young people (20-29 years old) is statistically significantly different from the baseline consumption of older people. Describe the results.

**H0: People aged 20-29 do not consume less than older people during the baseline phase**

**H1: People aged 20-29 do consume less than older people during the baseline phase**

**r code---**

|  |
| --- |
| # combine the Shower- and Survey-dataset by a left\_join |
|  |
| # split in Baseline und no Baseline Phase |
| data\_with\_baseline <- shower\_data %>%dplyr::filter(Shower <= 10) |
| #data\_no\_baseline <- data %>%dplyr::filter(Shower > 10) |
| combined\_dataset <- dplyr::left\_join(data\_with\_baseline, survey\_data) |
|  |
| # use this dataset and group it by Hh\_ID, age, then summarise the average Volume |
| age\_participants <- combined\_dataset %>% dplyr::group\_by(Hh\_ID, alter) %>% |
| dplyr::summarise(avgEnergy = mean(Energy)) |
|  |
| # filter dataset by people aged 20-29 and older and save it in according new dataset |
| age\_20\_29 <- age\_participants%>% dplyr::filter(alter == "20-29") |
|  |
| age\_older\_20\_29 <- age\_participants%>%filter(alter != "20-29") |
| #age\_20\_29$avgEnergy |
| #two-sample and two-sided t-test of people aged 20-29 and older and compare their average Volume |
| t.test(age\_20\_29$avgEnergy, age\_older\_20\_29$avgEnergy, mu = 0, conf.level = .95) |
| # combine the Shower- and Survey-dataset by a left\_join |
|  |
| # split in Baseline und no Baseline Phase |

**Answer: Started to explore lower age group people have high concern to save resources and influence on consumption, performed statistical t-test to check the two tailed confidence, no enough confidence to accept nor reject null hypothesis , we need to perform more statistical measure to understand.**

**Exercise c)**

**Question:** Does the gender of the participants influence the average shower volume of the users during the baseline phase (column “gesl” of the survey data)?

**H0: Gender of participants does not influence the average shower volume during the baseline phase**

**H1: Gender of participants does influence the average shower volume during the baseline phase**

**r code---**

|  |
| --- |
| # use combined dataset and compute average Volume for every Household and the according gender  gender <- combined\_dataset %>% group\_by(Hh\_ID, gesl) %>% summarise(avgVolume=mean(Volume))  # filter the new dataset gender by sexuality (männlich & is not männlich = weiblich)  weiblich <- gender%>% filter(gesl == "weiblich")  mannlich <- gender%>%filter(gesl != "weiblich")  #two-sample & two-sided t-test, compare the average volume of each gender and the according Household ID's on a confidence level of 95 %  t.test(mannlich$avgVolume, weiblich$avgVolume, mu = 0, conf.level = .95) |
|  |

**Answer: Started to explore gender influence on consumption, performed statistical t-test to check the two tailed confidence, no appropriate confidence and means are not equal to accept, rejecting null hypothesis.**

**Exercise d)**

**Question: Political decision-makers may find it very interesting to see which demographic groups have consumption when targeted by (real-time) feedback. Test whether study participants with a monthly income of 9000 Swiss francs or more show similar water consumption changes (possibly savings) as less wealthy ones. Remove participants who have not indicated their salary from the test. What do you observe?**

**H0: participants with an income <9000 show similar water consumption changes than the ones with a higher income**

**H1: participants with an income <9000 do not show similar water consumption changes than the ones with a higher income**

**r code---**

|  |
| --- |
| #  combined\_dataset\_no\_baseline <- dplyr::left\_join(data\_no\_baseline, survey\_data)  c\_d\_baseline <- combined\_dataset %>% dplyr::group\_by(Hh\_ID, einkommen) %>%  summarise(avgVolume = mean(Volume))  c\_d\_no\_baseline <- combined\_dataset\_no\_baseline %>% group\_by(Hh\_ID, einkommen) %>%  summarise(avgVolume = mean(Volume))  cd\_bllower\_9000 <- c\_d\_baseline %>% filter(einkommen %in% c("< 8000 Fr.", "8000 - 8999 Fr."))  cd\_blabove\_9000 <- c\_d\_baseline %>% filter(einkommen %in% c("10000 - 11999 Fr.", "12000 - 14999 Fr.", "9000 - 9999 Fr.",  "Mehr als 15000 Fr."))  na.omit(cd\_bllower\_9000)  na.omit(cd\_blabove\_9000)  cd\_Iplower9000 <- c\_d\_no\_baseline %>% filter(einkommen %in% c("< 8000 Fr.", "8000 - 8999 Fr."))  cd\_Ipabove9000 <- c\_d\_no\_baseline %>% filter(einkommen %in% c("10000 - 11999 Fr.", "12000 - 14999 Fr.", "9000 - 9999 Fr.",  "Mehr als 15000 Fr."))  comparison\_bl\_above <- c(cd\_blabove\_9000$avgVolume - cd\_Ipabove9000$avgVolume)  diff\_Volume\_above <- cd\_Ipabove9000  diff\_Volume\_above$Avarage\_BL <- cd\_blabove\_9000$avgVolume  diff\_Volume\_above$diff <- c(comparison\_bl\_above)  # Setting names  names(diff\_Volume\_above) <- c("Hh\_ID", "einkommen", "Avarage\_volume\_IP", "Avarage\_volume\_BL", "diff\_above")  comparison\_below <- cd\_bllower\_9000$avgVolume - cd\_Iplower9000$avgVolume  diff\_volume\_Below <- cd\_bllower\_9000  diff\_volume\_Below$Avarage\_IP <- cd\_Iplower9000$avgVolume  diff\_volume\_Below$diff <- c(comparison\_below)  # Setting names  names(diff\_volume\_Below) <- c("Hh\_ID", "einkommen", "Avarage\_volume\_BL", "Avarage\_volume\_IP", "diff\_below")  t.test(diff\_volume\_Below$diff\_below, diff\_Volume\_above$diff\_above, mu = 0, conf.level = .95)  # Reject Null Hypothesis |

**Answer:**

**No enough and sufficient information to accept null hypothesis.**

**Exercise e)**

**Question: Energy consumption can be reduced by reducing the time per shower, the flow rate, by showering the hot water temperature or by stopping the shower when applying shampoo. What seems to be the preferred strategy of the users?**

**r code---**

|  |
| --- |
| data\_with\_baseline <- shower\_data %>%dplyr::filter(Shower <= 10)  data\_no\_baseline <- shower\_data %>%dplyr::filter(Shower > 10)  #data\_control\_IP\_no <- data\_no\_baseline %>% filter(group <= 2)  data\_feedback\_group <- shower\_data %>% filter(group > 2) %>% group\_by(Hh\_ID, Shower <= 10) %>%  summarise(avgShower = mean(Showertime),  avgTemp = mean(Avgtemperature),  avgFlowrate =mean(Flowrate),  avgBreaktime = mean(Breaktime))  #avgShower  #data\_feedback\_group  data\_feedback\_group\_1 <- data\_feedback\_group %>% filter(`Shower <= 10`== FALSE)  data\_feedback\_group\_2 <- data\_feedback\_group %>% filter(`Shower <= 10` == TRUE)  # combine datasets  b <- right\_join(data\_feedback\_group\_1, data\_feedback\_group\_2, by = "Hh\_ID")  dim(b)  #get the differences of the phases  attach(b)  diff\_showertime <- avgShower.x - avgShower.y  diff\_Breaktime <- avgBreaktime.x - avgBreaktime.y  diff\_flowrate <- avgFlowrate.x - avgFlowrate.y  diff\_temperature <- avgTemp.x - avgTemp.y  detach(b)  #diff\_showertime  #get the differences of the phases  dim(data\_feedback\_group)  data\_feedback\_group\_1$diff\_showertime <- diff\_showertime  data\_feedback\_group\_1$diff\_Breaktime <- diff\_Breaktime  data\_feedback\_group\_1$diff\_flowrate <- diff\_flowrate  data\_feedback\_group\_1$diff\_temperature <- diff\_temperature  #  data\_feedback\_group\_2$diff\_showertime <- diff\_showertime  data\_feedback\_group\_2$diff\_Breaktime <- diff\_Breaktime  data\_feedback\_group\_2$diff\_flowrate <- diff\_flowrate  data\_feedback\_group\_2$diff\_temperature <- diff\_temperature  data\_feedback\_group = rbind(data\_feedback\_group\_1,data\_feedback\_group\_2)  avgShower\_boolean <- case\_when(data\_feedback\_group[,7] < 0 ~ TRUE,  data\_feedback\_group[,7] >= 0 ~ FALSE)  avgBreaktime\_boolean <- case\_when(data\_feedback\_group[,8] < 0 ~ FALSE,  data\_feedback\_group[,8] >= 0 ~ TRUE)  avgFlowrate\_boolean <- case\_when(data\_feedback\_group[,9] < 0 ~ TRUE,  data\_feedback\_group[,9] >= 0 ~ FALSE)  avgTemperature\_boolean <- case\_when(data\_feedback\_group[,10] < 0 ~ TRUE,  data\_feedback\_group[,10] >= 0 ~ FALSE)  windows()  layout(matrix(1:4,nrow=2,byrow=TRUE))  barplot(table(avgBreaktime\_boolean), col = "orchid3", main = "Avarage Breaktime")  barplot(table(avgTemperature\_boolean), col = "cornflowerblue", main = "Avarage Temperature")  barplot(table(avgFlowrate\_boolean), col = "tomato", main = "Avarage Flowrate")  barplot(table(avgShower\_boolean), col = "forestgreen", main = "Avarage Showertime") |

**Answer:**

**Based on below graphs, we understand flow rate should be higher and shower time to be lower, this could be potential strategy.**

**Chart, bar chart

Description automatically generated**