**ASSIGNMENT 6**

**BU ID: U55-32-1699**

Reading the data into the file:

> setwd("C:/Users/Lenovo/Downloads")

> measurements <- read.csv("Assgn6.csv")

Summary statistics – first step to analysing any data  
  
temp sex heartrate

Min. : 96.30 Min. :1.0 Min. :57.00

1st Qu.: 97.80 1st Qu.:1.0 1st Qu.:69.00

Median : 98.30 Median :1.5 Median :74.00

Mean : 98.25 Mean :1.5 Mean :73.76

3rd Qu.: 98.70 3rd Qu.:2.0 3rd Qu.:79.00

Max. :100.80 Max. :2.0 Max. :89.00

**1. To create a variable and assign the constraints given:**  
> measurements$temp\_level <- ifelse(measurements$temp >=98.6 , 1, 0)

> View(measurements)

**2. Body temperature level by sex:**

> aggregate(measurements$temp, by=list(measurements$sex), summary)

Group.1 x.Min. x.1st Qu. x.Median x.Mean x.3rd Qu. x.Max.

1 1 96.30000 97.50000 98.00000 97.98923 98.40000 99.50000

2 2 96.40000 98.20000 98.60000 98.50923 98.80000 100.80000

Where 1- males and 2- females

> table(measurements$temp\_level,measurements$sex)

1 2

0 51 30

1 14 35

0- the temperature being less than 98.6  
1- the temperature being more than 98.6

**3. Proportion of males and females involved in the test:**

|  |
| --- |
| > prop.table(table(measurements$temp\_level,measurements$sex))    1 2  0 0.3923077 0.2307692  1 0.1076923 0.2692308  Step 1: Setting up the hypothesis  H0 : p1=p2 (There is no difference in proportion of women and men having temperatures  Above 98.6.) H1 : p1 ≠ p2 (There is a difference in proportion of women and men having temperatures  Above 98.6.)  α =0.05  Step 2: Selecting the test statistic z=p^1−p^2 / sqrt( p^(1−p^)⋅(1n1+1n2))  α/2 = 0.05/2 = 0.025 Reject H0 if z >= 1.960  Otherwise , do not reject H0.  Computing the test statistic: male – success = 14, failure = 51  Female – success = 35, failure = 30  P1 = 14/65 = 0.215  P2 = 35/65 = 0.538  Step2: Select the appropriate test statistic  We use the two sample z-statistic, which is calculated by dividing the difference in  the sample proportions by the standard error of the difference in sample proportions,  where the standard error is calculated under the assumption of null hyphothesis.  Step3: State the decision rule   * Determine the appropriate critical value from the [standard normal distribution](https://onlinecampus.bu.edu/bbcswebdav/pid-5329611-dt-content-rid-18891459_1/courses/17fallmetcs555_b1/documents/Table_A_Standard_Normal_Probabilities.pdf)  associated with a right hand tail probability of α/2=0.05/2=0.025. * Decision Rule: Reject H0 if |z|≥1.960 * Otherwise, do not reject   Step4: Compute the test statistic and the associated p value:  z = 0.21-0.53/SQRT(49/130)(1-49/130)(2/65)  Which is equal to -45.76, and the absolute value is 45.76  Step5: Conclusion  We reject H0 since 45.76 is greater than 1.960. We have significant evidence that the  proportion of people with higher body temperature is not the same across men and women.  The risk difference is p1(hat)-p2(hat), which is 0.53-0.21,  i.e females are 32% more likely to have a higher body temperature than men.  **4. To formally test:**   1. Set up the hypotheses and select the alpha level    * H0:β1= or OR=1 (there is no association between body temperature levels and   sex)   * + H1:β1≠0or OR≠0 (there is an association between body temperature levels and   sex)   * + α=0.05  1. Select the appropriate test statistic    * z=β1/SEβ̂ 1 2. State the decision rule    * Determine the appropriate value from the standard normal distribution  associated with a right hand tail probability of α/2=0.05/2=0.025    * Using the table, zα/2=1.960    * Decision Rule: Reject H0 if |z|≥1.960 or  Reject H0 if p ≤ α    * Otherwise, do not reject H0 3. Compute the test statistic   > m4 <- glm(measurements$temp\_level ~ measurements$sex)  > summary(m4)  Call:  glm(formula = measurements$temp\_level ~ measurements$sex)  Deviance Residuals:  Min 1Q Median 3Q Max  -0.5385 -0.2154 -0.2154 0.4615 0.7846  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) -0.10769 0.12771 -0.843 0.400648  measurements$sex 0.32308 0.08077 4.000 0.000106 \*\*\*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for gaussian family taken to be 0.2120192)  Null deviance: 30.531 on 129 degrees of freedom  Residual deviance: 27.138 on 128 degrees of freedom  AIC: 171.27  Number of Fisher Scoring iterations: 2  Odds ratio for the sex and temperature association: 1.2403  C statistic: > roc( measurements$temp\_level ~ measurements$prob)  Call:  roc.formula(formula = measurements$temp\_level ~ measurements$prob)  Data: measurements$prob in 81 controls (measurements$temp\_level 0) < 49 cases (measurements$temp\_level 1).  Area under the curve: 0.672 is the C statistic   1. Conclusion   We reject the null hypothesis as we find evident data that there is an association  between the sex and the temperature variable.  **5. Multiple logistic regression:**  Call:  glm(formula = measurements$temp\_level ~ measurements$sex + measurements$heartrate,  family = binomial)  Deviance Residuals:  Min 1Q Median 3Q Max  -1.6524 -0.8639 -0.6103 1.0489 2.0480  Coefficients:  Estimate Std. Error z value Pr(>|z|)  (Intercept) -7.34755 2.21406 -3.319 0.000905 \*\*\*  measurements$sex 1.38919 0.39868 3.484 0.000493 \*\*\*  measurements$heartrate 0.06337 0.02850 2.223 0.026195 \*  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  (Dispersion parameter for binomial family taken to be 1)  Null deviance: 172.26 on 129 degrees of freedom  Residual deviance: 152.24 on 127 degrees of freedom  AIC: 158.24  Number of Fisher Scoring iterations: 4  Odds ratio for sex an heart rate for a 10 unit increase: > exp(m2$coefficients[2]\*10)  measurements$sex  1079345  6. The multiple logistic regression is more efficient than the former as this gives us  Proper estimates of the significance of the relationship of the paramters in the model.  From the ,model we can see that both sex and heart rate have a p value that is very  Much smaller than 0.01, which shows high significance of the data points.  Curve:  > m2<-glm(measurements$temp\_level ~ measurements$sex + measurements$heartrate,  family = binomial)  library(pROC)  measurements$prob <-predict(m2, type=c("response"))  n <- roc( measurements$temp\_level ~ measurements$prob)  plot(n) |
|  |
| |  | | --- | |  | |