Qualitative variables

Harpreet

2024-11-13

Qualitative Variables- Focus on the categorical variables model.

```
weight=read.csv("D:\\RData603\\Project\\weight_change_dataset.csv")
print(colnames(weight))
```

```
##  [1] "Participant.ID"                "Age"
##  [3] "Gender"                        "Current.Weight..lbs."
##  [5] "BMR..Calories."                "Daily.Calories.Consumed"
##  [7] "Daily.Caloric.Surplus.Deficit" "Weight.Change..lbs."
##  [9] "Duration..weeks."              "Physical.Activity.Level"
## [11] "Sleep.Quality"                 "Stress.Level"
## [13] "Final.Weight..lbs."
```

```
print(head(weight))
```

```
##   Participant.ID Age Gender Current.Weight..lbs. BMR..Calories.
## 1              1  56      M                228.4         3102.3
## 2              2  46      F                165.4         2275.5
## 3              3  32      F                142.8         2119.4
## 4              4  25      F                145.5         2181.3
## 5              5  38      M                155.5         2463.8
## 6              6  56      F                152.9         2100.6
##   Daily.Calories.Consumed Daily.Caloric.Surplus.Deficit Weight.Change..lbs.
## 1                  3916.0                         813.7              0.2000
## 2                  3823.0                        1547.5              2.4000
## 3                  2785.4                         666.0              1.4000
## 4                  2587.3                         406.0              0.8000
## 5                  3312.8                         849.0              2.0000
## 6                  2262.4                         161.9            -12.5135
##   Duration..weeks. Physical.Activity.Level Sleep.Quality Stress.Level
## 1                1               Sedentary     Excellent            6
## 2                6             Very Active     Excellent            6
## 3                7               Sedentary          Good            3
## 4                8               Sedentary          Fair            2
## 5               10          Lightly Active          Good            1
## 6                9               Sedentary          Poor            6
##   Final.Weight..lbs.
## 1              228.6
## 2              167.8
```

```
## 3              144.2
## 4              146.3
## 5              157.5
## 6              140.4
```

starting down with Only categorical variables

1. Gender- M or F
2. Duration..weeks.
3. Physical.Activity.Level
4. Sleep.Quality
5. Stress.Level

```
Weight_full_model= lm(Weight.Change..lbs.~ factor(Gender)+ factor(Duration..weeks.)+factor(Physical.Acti
print(summary(Weight_full_model))
```

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Gender) + factor(Duration..weeks.) +
##     factor(Physical.Activity.Level) + factor(Sleep.Quality) +
##     factor(Stress.Level), data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.3526  -2.0865   0.1295   2.0104   8.0747
##
## Coefficients:
##                                                  Estimate Std. Error t value
## (Intercept)                                        5.3669     2.3187   2.315
## factor(Gender)M                                   -0.6926     0.9109  -0.760
## factor(Duration..weeks.)2                         -3.3352     2.2929  -1.455
## factor(Duration..weeks.)3                         -3.6956     1.9807  -1.866
## factor(Duration..weeks.)4                         -4.2244     2.1575  -1.958
## factor(Duration..weeks.)5                         -3.1102     2.0140  -1.544
## factor(Duration..weeks.)6                         -2.8804     1.8205  -1.582
## factor(Duration..weeks.)7                         -6.6863     2.2507  -2.971
## factor(Duration..weeks.)8                         -5.8073     2.0801  -2.792
## factor(Duration..weeks.)9                         -5.7899     2.0024  -2.892
## factor(Duration..weeks.)10                        -5.9925     1.9313  -3.103
## factor(Duration..weeks.)11                       -11.6345     2.3980  -4.852
## factor(Duration..weeks.)12                        -4.3041     1.7098  -2.517
## factor(Physical.Activity.Level)Moderately Active  -0.7642     1.2094  -0.632
## factor(Physical.Activity.Level)Sedentary          -0.9274     1.1750  -0.789
## factor(Physical.Activity.Level)Very Active         2.2698     1.1445   1.983
## factor(Sleep.Quality)Fair                          1.8439     1.4179   1.300
## factor(Sleep.Quality)Good                          2.7397     1.4514   1.888
## factor(Sleep.Quality)Poor                         -7.2450     1.2534  -5.780
## factor(Stress.Level)2                             -0.2645     1.7727  -0.149
## factor(Stress.Level)3                              1.3717     1.7292   0.793
## factor(Stress.Level)4                              1.2301     1.9617   0.627
## factor(Stress.Level)5                              2.1266     1.9266   1.104
## factor(Stress.Level)6                             -1.3727     1.6634  -0.825
```

```
## factor(Stress.Level)7                                0.3535     1.7899   0.198
## factor(Stress.Level)8                              -10.6627     1.8948  -5.627
## factor(Stress.Level)9                               -9.8607     1.8368  -5.368
##                                                    Pr(>|t|)
## (Intercept)                                         0.02345 *
## factor(Gender)M                                     0.44951
## factor(Duration..weeks.)2                           0.15008
## factor(Duration..weeks.)3                           0.06609 .
## factor(Duration..weeks.)4                           0.05405 .
## factor(Duration..weeks.)5                           0.12684
## factor(Duration..weeks.)6                           0.11793
## factor(Duration..weeks.)7                           0.00402 **
## factor(Duration..weeks.)8                           0.00668 **
## factor(Duration..weeks.)9                           0.00505 **
## factor(Duration..weeks.)10                          0.00273 **
## factor(Duration..weeks.)11                         6.74e-06 ***
## factor(Duration..weeks.)12                          0.01402 *
## factor(Physical.Activity.Level)Moderately Active    0.52945
## factor(Physical.Activity.Level)Sedentary            0.43249
## factor(Physical.Activity.Level)Very Active          0.05110 .
## factor(Sleep.Quality)Fair                           0.19753
## factor(Sleep.Quality)Good                           0.06305 .
## factor(Sleep.Quality)Poor                          1.72e-07 ***
## factor(Stress.Level)2                               0.88180
## factor(Stress.Level)3                               0.43018
## factor(Stress.Level)4                               0.53256
## factor(Stress.Level)5                               0.27331
## factor(Stress.Level)6                               0.41192
## factor(Stress.Level)7                               0.84397
## factor(Stress.Level)8                              3.19e-07 ***
## factor(Stress.Level)9                              9.01e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.738 on 73 degrees of freedom
## Multiple R-squared:  0.814,  Adjusted R-squared:  0.7478
## F-statistic: 12.29 on 26 and 73 DF,  p-value: < 2.2e-16
```

```
weight_model_take2=lm(Weight.Change..lbs.~factor(Gender)+factor(Physical.Activity.Level)+factor(Sleep.Q
print(summary(weight_model_take2))
```

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Gender) + factor(Physical.Activity.Level) +
##     factor(Sleep.Quality) + factor(Stress.Level), data = weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.313  -1.369  -0.074   2.093  12.709
##
## Coefficients:
##                                                  Estimate Std. Error t value
## (Intercept)                                       1.7189     1.8081   0.951
## factor(Gender)M                                  -0.8365     0.9712  -0.861
```

```
## factor(Physical.Activity.Level)Moderately Active   0.0540     1.2103   0.045
## factor(Physical.Activity.Level)Sedentary          -0.6371     1.2509  -0.509
## factor(Physical.Activity.Level)Very Active         1.4788     1.2152   1.217
## factor(Sleep.Quality)Fair                          0.5978     1.3964   0.428
## factor(Sleep.Quality)Good                          1.1141     1.4241   0.782
## factor(Sleep.Quality)Poor                         -8.1191     1.2927  -6.281
## factor(Stress.Level)2                              0.2977     1.7632   0.169
## factor(Stress.Level)3                              1.2671     1.7525   0.723
## factor(Stress.Level)4                              1.5876     2.0526   0.773
## factor(Stress.Level)5                              1.8938     1.8458   1.026
## factor(Stress.Level)6                             -0.2817     1.7490  -0.161
## factor(Stress.Level)7                             -0.2628     1.8777  -0.140
## factor(Stress.Level)8                            -10.8574     1.8308  -5.930
## factor(Stress.Level)9                             -9.1820     1.8864  -4.867
##                                                  Pr(>|t|)
## (Intercept)                                         0.344
## factor(Gender)M                                     0.392
## factor(Physical.Activity.Level)Moderately Active    0.965
## factor(Physical.Activity.Level)Sedentary            0.612
## factor(Physical.Activity.Level)Very Active          0.227
## factor(Sleep.Quality)Fair                           0.670
## factor(Sleep.Quality)Good                           0.436
## factor(Sleep.Quality)Poor                       1.42e-08 ***
## factor(Stress.Level)2                               0.866
## factor(Stress.Level)3                               0.472
## factor(Stress.Level)4                               0.441
## factor(Stress.Level)5                               0.308
## factor(Stress.Level)6                               0.872
## factor(Stress.Level)7                               0.889
## factor(Stress.Level)8                           6.50e-08 ***
## factor(Stress.Level)9                           5.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.159 on 84 degrees of freedom
## Multiple R-squared:  0.7351, Adjusted R-squared:  0.6878
## F-statistic: 15.54 on 15 and 84 DF,  p-value: < 2.2e-16
```

```
weight_model_take3=lm(Weight.Change..lbs.~factor(Sleep.Quality)+factor(Stress.Level), data=weight)
print(summary(weight_model_take2))
```

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Gender) + factor(Physical.Activity.Level) +
##     factor(Sleep.Quality) + factor(Stress.Level), data = weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.313  -1.369  -0.074   2.093  12.709
##
## Coefficients:
##                                                  Estimate Std. Error t value
## (Intercept)                                       1.7189     1.8081   0.951
## factor(Gender)M                                  -0.8365     0.9712  -0.861
```

```
## factor(Physical.Activity.Level)Moderately Active   0.0540     1.2103   0.045
## factor(Physical.Activity.Level)Sedentary          -0.6371     1.2509  -0.509
## factor(Physical.Activity.Level)Very Active         1.4788     1.2152   1.217
## factor(Sleep.Quality)Fair                          0.5978     1.3964   0.428
## factor(Sleep.Quality)Good                          1.1141     1.4241   0.782
## factor(Sleep.Quality)Poor                         -8.1191     1.2927  -6.281
## factor(Stress.Level)2                              0.2977     1.7632   0.169
## factor(Stress.Level)3                              1.2671     1.7525   0.723
## factor(Stress.Level)4                              1.5876     2.0526   0.773
## factor(Stress.Level)5                              1.8938     1.8458   1.026
## factor(Stress.Level)6                             -0.2817     1.7490  -0.161
## factor(Stress.Level)7                             -0.2628     1.8777  -0.140
## factor(Stress.Level)8                            -10.8574     1.8308  -5.930
## factor(Stress.Level)9                             -9.1820     1.8864  -4.867
##                                                  Pr(>|t|)
## (Intercept)                                         0.344
## factor(Gender)M                                     0.392
## factor(Physical.Activity.Level)Moderately Active    0.965
## factor(Physical.Activity.Level)Sedentary            0.612
## factor(Physical.Activity.Level)Very Active          0.227
## factor(Sleep.Quality)Fair                           0.670
## factor(Sleep.Quality)Good                           0.436
## factor(Sleep.Quality)Poor                       1.42e-08 ***
## factor(Stress.Level)2                               0.866
## factor(Stress.Level)3                               0.472
## factor(Stress.Level)4                               0.441
## factor(Stress.Level)5                               0.308
## factor(Stress.Level)6                               0.872
## factor(Stress.Level)7                               0.889
## factor(Stress.Level)8                           6.50e-08 ***
## factor(Stress.Level)9                           5.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.159 on 84 degrees of freedom
## Multiple R-squared:  0.7351, Adjusted R-squared:  0.6878
## F-statistic: 15.54 on 15 and 84 DF,  p-value: < 2.2e-16
```

```
weight_interaction_model= lm(Weight.Change..lbs.~(factor(Sleep.Quality)+factor(Stress.Level))^2, data=w
print(summary(weight_interaction_model))
```

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ (factor(Sleep.Quality) + factor(Stress.Level))^2,
##     data = weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.509  -1.190   0.000   1.251  16.775
##
## Coefficients: (2 not defined because of singularities)
##                                                 Estimate Std. Error t value
## (Intercept)                                      0.90000    4.36708   0.206
## factor(Sleep.Quality)Fair                        2.00000    5.04267   0.397
```

```
## factor(Sleep.Quality)Good                            0.96667    5.04267   0.192
## factor(Sleep.Quality)Poor                           -6.86346    4.88254  -1.406
## factor(Stress.Level)2                                1.20000    5.04267   0.238
## factor(Stress.Level)3                               -0.30000    5.04267  -0.059
## factor(Stress.Level)4                                4.15260    3.78200   1.098
## factor(Stress.Level)5                                0.85000    5.34856   0.159
## factor(Stress.Level)6                                1.50000    4.88254   0.307
## factor(Stress.Level)7                               -0.40000    5.34856  -0.075
## factor(Stress.Level)8                               -4.57874    6.17598  -0.741
## factor(Stress.Level)9                              -12.20581    3.08799  -3.953
## factor(Sleep.Quality)Fair:factor(Stress.Level)2     -3.14000    5.96657  -0.526
## factor(Sleep.Quality)Good:factor(Stress.Level)2     -0.60667    5.96657  -0.102
## factor(Sleep.Quality)Poor:factor(Stress.Level)2      2.14010    7.01910   0.305
## factor(Sleep.Quality)Fair:factor(Stress.Level)3      1.80000    7.13141   0.252
## factor(Sleep.Quality)Good:factor(Stress.Level)3      0.36667    6.17598   0.059
## factor(Sleep.Quality)Poor:factor(Stress.Level)3      1.89297    5.73767   0.330
## factor(Sleep.Quality)Fair:factor(Stress.Level)4     -4.25260    5.49512  -0.774
## factor(Sleep.Quality)Good:factor(Stress.Level)4     -4.11927    5.19786  -0.792
## factor(Sleep.Quality)Poor:factor(Stress.Level)4           NA         NA      NA
## factor(Sleep.Quality)Fair:factor(Stress.Level)5     -2.15000    6.67082  -0.322
## factor(Sleep.Quality)Good:factor(Stress.Level)5      1.68333    7.35089   0.229
## factor(Sleep.Quality)Poor:factor(Stress.Level)5      1.46086    6.04595   0.242
## factor(Sleep.Quality)Fair:factor(Stress.Level)6     -2.46667    6.04595  -0.408
## factor(Sleep.Quality)Good:factor(Stress.Level)6     -1.56667    6.30334  -0.249
## factor(Sleep.Quality)Poor:factor(Stress.Level)6     -2.00745    5.77710  -0.347
## factor(Sleep.Quality)Fair:factor(Stress.Level)7      0.06667    6.42817   0.010
## factor(Sleep.Quality)Good:factor(Stress.Level)7      3.33333    7.35089   0.453
## factor(Sleep.Quality)Poor:factor(Stress.Level)7     -0.37889    6.17598  -0.061
## factor(Sleep.Quality)Fair:factor(Stress.Level)8     -6.58104    7.35089  -0.895
## factor(Sleep.Quality)Good:factor(Stress.Level)8     -2.98419    7.35089  -0.406
## factor(Sleep.Quality)Poor:factor(Stress.Level)8     -9.31800    6.78890  -1.373
## factor(Sleep.Quality)Fair:factor(Stress.Level)9      4.63575    4.71699   0.983
## factor(Sleep.Quality)Good:factor(Stress.Level)9      5.04842    5.04267   1.001
## factor(Sleep.Quality)Poor:factor(Stress.Level)9           NA         NA      NA
##                                                    Pr(>|t|)
## (Intercept)                                        0.837357
## factor(Sleep.Quality)Fair                          0.692931
## factor(Sleep.Quality)Good                          0.848568
## factor(Sleep.Quality)Poor                          0.164499
## factor(Stress.Level)2                              0.812642
## factor(Stress.Level)3                              0.952740
## factor(Stress.Level)4                              0.276198
## factor(Stress.Level)5                              0.874216
## factor(Stress.Level)6                              0.759646
## factor(Stress.Level)7                              0.940611
## factor(Stress.Level)8                              0.461094
## factor(Stress.Level)9                              0.000191 ***
## factor(Sleep.Quality)Fair:factor(Stress.Level)2    0.600468
## factor(Sleep.Quality)Good:factor(Stress.Level)2    0.919321
## factor(Sleep.Quality)Poor:factor(Stress.Level)2    0.761404
## factor(Sleep.Quality)Fair:factor(Stress.Level)3    0.801513
## factor(Sleep.Quality)Good:factor(Stress.Level)3    0.952837
## factor(Sleep.Quality)Poor:factor(Stress.Level)3    0.742506
## factor(Sleep.Quality)Fair:factor(Stress.Level)4    0.441761
```

```
## factor(Sleep.Quality)Good:factor(Stress.Level)4 0.430913
## factor(Sleep.Quality)Poor:factor(Stress.Level)4       NA
## factor(Sleep.Quality)Fair:factor(Stress.Level)5 0.748245
## factor(Sleep.Quality)Good:factor(Stress.Level)5 0.819579
## factor(Sleep.Quality)Poor:factor(Stress.Level)5 0.809819
## factor(Sleep.Quality)Fair:factor(Stress.Level)6 0.684604
## factor(Sleep.Quality)Good:factor(Stress.Level)6 0.804484
## factor(Sleep.Quality)Poor:factor(Stress.Level)6 0.729334
## factor(Sleep.Quality)Fair:factor(Stress.Level)7 0.991757
## factor(Sleep.Quality)Good:factor(Stress.Level)7 0.651705
## factor(Sleep.Quality)Poor:factor(Stress.Level)7 0.951266
## factor(Sleep.Quality)Fair:factor(Stress.Level)8 0.373896
## factor(Sleep.Quality)Good:factor(Stress.Level)8 0.686083
## factor(Sleep.Quality)Poor:factor(Stress.Level)8 0.174544
## factor(Sleep.Quality)Fair:factor(Stress.Level)9 0.329306
## factor(Sleep.Quality)Good:factor(Stress.Level)9 0.320415
## factor(Sleep.Quality)Poor:factor(Stress.Level)9       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.367 on 66 degrees of freedom
## Multiple R-squared:  0.7705, Adjusted R-squared:  0.6558
## F-statistic: 6.716 on 33 and 66 DF,  p-value: 3.128e-11
```

```
weight_interaction_model2= lm(Weight.Change..lbs.~factor(Stress.Level), data=weight)
print(summary(weight_interaction_model2))
```

```
##
## Call:
## lm(formula = Weight.Change..lbs. ~ factor(Stress.Level), data = weight)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.8705  -2.4223   0.9663   2.9295  12.0838
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -0.7867     1.7421  -0.452 0.652631
## factor(Stress.Level)2    2.2708     2.3279   0.975 0.331925
## factor(Stress.Level)3   -0.5414     2.3279  -0.233 0.816622
## factor(Stress.Level)4    1.8836     2.7935   0.674 0.501841
## factor(Stress.Level)5   -0.1965     2.4636  -0.080 0.936596
## factor(Stress.Level)6    0.2572     2.3670   0.109 0.913710
## factor(Stress.Level)7   -0.5602     2.5245  -0.222 0.824874
## factor(Stress.Level)8 -12.9180     2.4636  -5.243 1.02e-06 ***
## factor(Stress.Level)9 -10.0209     2.5969  -3.859 0.000213 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.778 on 91 degrees of freedom
## Multiple R-squared:  0.4462, Adjusted R-squared:  0.3975
## F-statistic: 9.165 on 8 and 91 DF,  p-value: 3.611e-09
```

```r
print(anova(weight_interaction_model2,weight_model_take3))
```

```
## Analysis of Variance Table
##
## Model 1: Weight.Change..lbs. ~ factor(Stress.Level)
## Model 2: Weight.Change..lbs. ~ factor(Sleep.Quality) + factor(Stress.Level)
##   Res.Df    RSS Df Sum of Sq     F  Pr(>F)
## 1     91 3037.8
## 2     88 1511.8  3      1526 29.61 2.5e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Significant reduction in RSS suggests that the addition of Sleep.Quality improved the model fit. Pvalue- : Extremely small, indicating that the addition of Sleep.Quality is highly statistically significant. so the model 2 is best. Which indicates that the Weight change is highly influenced by the Sleep quality and Stress level after observation on the 100 participants data.

retake- Now check it with the stepwise model

```r
#Weight_full_model= lm(Weight.Change..lbs.~ factor(Gender)+ factor(Duration..weeks.)+factor(Physical.Ac
#print(summary(Weight_full_model))
#stepwise_model= ols_step_both_p(Weight_full_model,p_enter = 0.05, p_remove = 0.06, details=TRUE)
#print(summary(stepwise_model$model))
```