

Daniela Cruz Perez

DSS445: HW6 Data Mining

11/9/2025

Q1

- a. The independent variables are the number of patients seen (*PatientsSeen*), patient wait time (*AvgWaitTime*), the number of nurses on shift (*NursesOnShift*), the number of doctors on shift (*DoctorsOnShift*), whether day was during a weekend (*Weekend*), and the severity of the case (*EmergencyLevel*). These variables are used to predict the response variable, average patient satisfaction score (*SatisfactionScore*)
- b. The numeric predictor variables are *PatientsSeen*, *AvgWaitTime*, *NursesOnShift*, and *DoctorsOnshift*. The categorical predictor variables are *Weekend* and *EmergencyLevel*.

Q2

a.

```
8 satisfaction$WeekendDummy <- ifelse(satisfaction$Weekend == "Yes", 1, 0)
9 satisfaction$EmergencyLvlDummy <- ifelse(satisfaction$EmergencyLevel == "High", 3,
10                                         ifelse(satisfaction$EmergencyLevel == "Medium", 2, 1))
11
12 model <- lm(SatisfactionScore ~ PatientsSeen + AvgWaitTime + NursesOnShift + DoctorsOnShift + WeekendDummy
13               + EmergencyLvlDummy, data = satisfaction)
14 summary(model)
15 |
```

```

call:
lm(formula = SatisfactionScore ~ PatientsSeen + AvgWaitTime +
    NursesOnShift + Doctorsonshift + WeekendDummy + EmergencyLvlDummy,
  data = satisfaction)

Residuals:
    Min      1Q  Median      3Q     Max 
-11.0184 -2.5150  0.1971  2.7606 13.8144 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 83.002172  1.960046 42.347 < 2e-16 ***
PatientsSeen -0.035348  0.004502 -7.852 2.76e-13 ***
AvgWaitTime  -0.286632  0.034179 -8.386 1.04e-14 ***
NursesOnshift 0.726408  0.073472  9.887 < 2e-16 ***
Doctorsonshift 0.137554  0.190636  0.722  0.47144  
WeekendDummy   -2.113767  0.683870 -3.091 0.00229 ** 
EmergencyLvlDummy -3.369232  0.455445 -7.398 4.14e-12 *** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.481 on 193 degrees of freedom
Multiple R-squared:  0.6038,    Adjusted R-squared:  0.5915 
F-statistic: 49.02 on 6 and 193 DF,  p-value: < 2.2e-16

```

- b. The predictors that were statistically significant include PatientsSeen, AvgWaitTime, NursesOnShift, Weekend, and EmergencyLevel.
- c. The numeric predictor average wait time has a coefficient of -.287, therefore as average wait time increases, the patient satisfaction score decreases by .287 units. In addition, the categorical predictor, level of emergency has a coefficient of -.37, therefore as the emergency level increases, the patient satisfaction score decreases by 3.7 units.
- d. The model's adjusted r-squared value is .5915, therefore the model explains 59.15% of the variability of the data.

Q3

a.

```

> #Q2
> model2 <- lm(SatisfactionScore ~ PatientsSeen + AvgWaitTime + NursesOnShift + WeekendDummy
+                 + EmergencyLvlDummy, data = satisfaction)
> summary(model2)

Call:
lm(formula = SatisfactionScore ~ PatientsSeen + AvgWaitTime +
    NursesOnShift + WeekendDummy + EmergencyLvlDummy, data = satisfaction)

Residuals:
    Min      1Q  Median      3Q     Max 
-10.6248 -2.5559  0.1162  2.8373 13.8766 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 83.574986  1.789866 46.693   < 2e-16 ***
PatientsSeen -0.035403  0.004495 -7.875 2.36e-13 ***
AvgWaitTime  -0.285383  0.034093 -8.371 1.12e-14 ***
NursesOnShift 0.724517  0.073334  9.880   < 2e-16 ***
WeekendDummy -2.134512  0.682421 -3.128  0.00203 ** 
EmergencyLvlDummy -3.356045  0.454516 -7.384 4.43e-12 *** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.476 on 194 degrees of freedom
Multiple R-squared:  0.6027, Adjusted R-squared:  0.5925 
F-statistic: 58.87 on 5 and 194 DF,  p-value: < 2.2e-16

```

- b. The adjusted R-squared for the second model is .5924, therefore it is slightly better than the first model of .5915.
- c. It is better to use the second model as it removed variables that are not statistically significant and explains 59.24% of the variability of the data.

#### Q4

a.

```

> model1 <- lm(SatisfactionScore ~ PatientsSeen + AvgWaitTime + NursesOnShift + Doctorsonshift + WeekendDummy
+                 + EmergencyLvlDummy + Nurses_Weekend, data = satisfaction)
> summary(model1)

Call:
lm(formula = SatisfactionScore ~ PatientsSeen + AvgWaitTime +
    NursesOnShift + Doctorsonshift + WeekendDummy + EmergencyLvlDummy +
    Nurses_Weekend, data = satisfaction)

Residuals:
    Min      1Q  Median      3Q     Max 
-10.8999 -2.3772  0.1525  2.3866 13.7900 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 84.436195  2.110463 40.008   < 2e-16 ***
PatientsSeen -0.034881  0.004485 -7.778 4.40e-13 ***
AvgWaitTime  -0.293846  0.034234 -8.584 3.10e-15 ***
NursesOnShift 0.634569  0.089571  7.085 2.58e-11 ***
Doctorsonshift 0.100990  0.190706  0.530  0.59703  
Weekenddummy -5.650318  2.107824 -2.681  0.00799 ** 
EmergencyLvlDummy -3.347133  0.453110 -7.387 4.47e-12 *** 
Nurses_Weekend 0.278820  0.157292  1.773  0.07788 . 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.456 on 192 degrees of freedom
Multiple R-squared:  0.6102, Adjusted R-squared:  0.596 
F-statistic: 42.93 on 7 and 192 DF,  p-value: < 2.2e-16

```

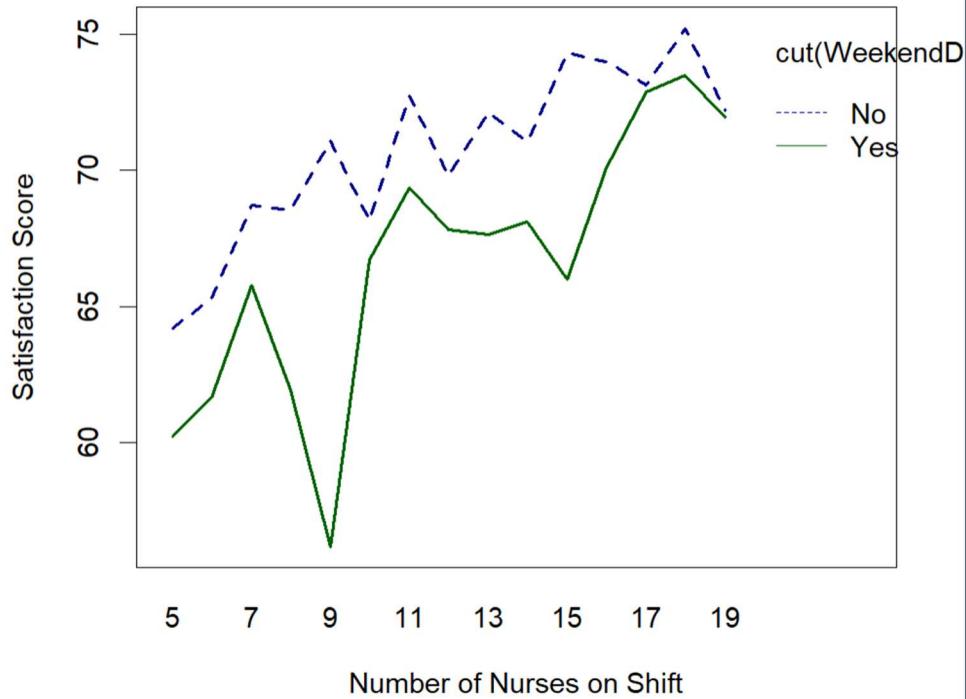
- b. The interaction variable between NursesOnShift and Weekend is not statistically significant.
- c. The coefficient sign is positive, therefore as the number of nurses on a weekend increases, the patient satisfaction score increases by .279 units.

Q5

a.

```
+   xlab = 'Miles per gallon')
> with(satisfaction,
+   interaction.plot(x.factor = Nurses_cat,
+     trace.factor = cut(WeekendDummy, 2, labels = c("No", "Yes")),
+     response = SatisfactionScore,
+     main = "Interaction Plot: Weekend x Number of Nurses on Satisfaction Score",
+     xlab = "Number of Nurses on shift",
+     ylab = "Satisfaction Score",
+     col = c("darkblue", "darkgreen"),
+     lwd = 2))
```

**Interaction Plot: Weekend x Number of Nurses on Satisfaction**



- b. The interaction plot shows there is variation in satisfaction score for whether it was or wasn't a weekend and the number of nurses on shift. The satisfaction score of weekend

days and weekend days tend to be lower than when there are fewer nurses on shift. The satisfaction score is higher on the weekdays than on the weekend.

Q6

- a. The predictor variables that most affect the response variable were the weekend and emergency level as they were both statistically significant and have a higher coefficient values, thus affecting the satisfaction score in higher units negatively. Followed by nurses on shift, which had a positive relationship to the satisfaction score.
- b. If I were to adjust one variable to raise satisfaction, I would recommend increasing the number of nurses on shift as this positively impacts the satisfaction score.
- c. Another variable that could be introduced to improve model accuracy is the average amount of time spent with patients. This may reflect how more amount of time spent between the patient and health care provider may positively affect the satisfaction score.