Math 328 Chapter 8 HW Part 2.Rmd

David Oniani

April 19, 2021

## Exercise 8.38

Music Time: Descriptive.

Participants were asked to judge when 45 seconds had passed under three conditions:

* Silence
* Listening to an upbeat song
* Listening to a calm song

The order in which each subject experienced these three conditions was randomized.

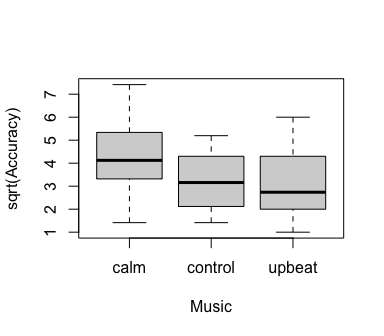
The response is Accuracy, which is the absolute difference between the reported time and 45 seconds. Use the square root transformation of Accuracy.

### 8.38 (a) Boxplots

library (Stat2Data)  
data ("MusicTime")  
str (MusicTime)

## 'data.frame': 60 obs. of 6 variables:  
## $ MusicBg : Factor w/ 2 levels "no","yes": 2 1 1 2 2 2 2 2 1 1 ...  
## $ Subject : Factor w/ 20 levels "subj1","subj10",..: 1 12 14 15 16 17 18 19 20 2 ...  
## $ Sex : Factor w/ 2 levels "f","m": 2 1 1 1 2 2 1 2 2 2 ...  
## $ TimeGuess: int 43 18 68 26 40 47 29 38 29 41 ...  
## $ Music : Factor w/ 3 levels "calm","control",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ Accuracy : int 2 27 23 19 5 2 16 7 16 4 ...

boxplot (sqrt(Accuracy) ~ Music, data=MusicTime)



Comments on the plot: It seems like the sqrt(Accuracy) is, on average, higher when listening to a calm song. The median value seems to be the biggest in this case. sqrt(Accuracy) looks approximately the same when listening in silence (control) or upbeat music. That being said, it does seem like listening in silence (control) has slightly bigger sqrt(Accuracy) value. The median value of listening in silence (control) seems to be bigger than that of listening to upbeat as well. Hence, the median values show the following descending order: calm, control, upbeat.

### 8.38 (b)

The null hypothesis is: 45 seconds passed at the same time with all three types of music (i.e., accuracy was the same in all three cases).

In symbols, H\_0: acc\_1 = acc\_2 = acc\_3 or H\_0: u\_1 = u\_2 = u\_3.

### 8.38 (c)

Answer: The boxplots do provide the visual evidence against the null hypothesis. This is the case since, clearly, sqrt(Accuracy) was the largest when listening to the calm music. Furthermore, sqrt(Accuracy) was noticeably larger when listening to a calm music than listening to any other type of music (silence/control or upbeat).

## Exercise 8.40

Conduct a repeated measures ANOVA.

mt1 = aov (sqrt(Accuracy) ~ Music + Error (Subject), data=MusicTime)  
summary (mt1)

##   
## Error: Subject  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Residuals 19 66.7 3.511   
##   
## Error: Within  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Music 2 18.03 9.016 6.52 0.00368 \*\*  
## Residuals 38 52.55 1.383   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

library (emmeans)  
summary (emmeans (mt1, pairwise ~ Music), infer=c(T,T))

## Note: re-fitting model with sum-to-zero contrasts

## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates

## $emmeans  
## Music emmean SE df lower.CL upper.CL t.ratio p.value  
## calm 4.32 0.323 46.4 3.67 4.97 13.358 <.0001   
## control 3.20 0.323 46.4 2.55 3.85 9.897 <.0001   
## upbeat 3.12 0.323 46.4 2.47 3.77 9.642 <.0001   
##   
## Warning: EMMs are biased unless design is perfectly balanced   
## Results are given on the sqrt (not the response) scale.   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## calm - control 1.1195 0.372 38 0.213 2.026 3.010 0.0125   
## calm - upbeat 1.2020 0.372 38 0.295 2.109 3.232 0.0070   
## control - upbeat 0.0825 0.372 38 -0.824 0.989 0.222 0.9733   
##   
## Note: contrasts are still on the sqrt scale   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 3 estimates   
## P value adjustment: tukey method for comparing a family of 3 estimates

Conclusions: from the very get-go, we can see that the p-value (approximately, 0.004, which is less than the 0.05 cutoff) from the summary of aov is significant meaning that there is a significant evidence against the null hypothesis that average accuracy for detecting when 45 seconds have passed was the same for all three types of music. Further analysis using emmeans functions shows that the difference between calm and upbeat has the smallest p-value of approximately 0.007 and therefore, is significant, which was expected based on our ordering in exercise 8.38 (decreasing order: calm, control, upbeat). An estimated difference calm - upbeat is approximately 1.202. The difference between calm and control is also significant with the p-value of 0.01 and an estimate of approximately 1.120. Finally, the difference between control and upbeat is not significant (an observation that we made in exercise 8.38) with the p-value of 0.973 and an estimate of approximately 0.083.

## Exercise 8.62

These data are from three car manufacturers. The task is to model Price as a function of car type and either age or mileage.

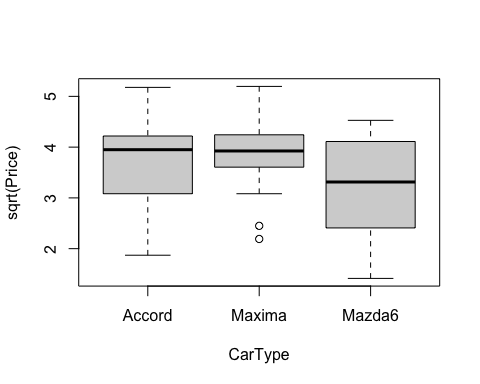
### 8.62 (a)

ANOVA for Price vs car type.

data ("ThreeCars2017")  
str (ThreeCars2017)

## 'data.frame': 90 obs. of 7 variables:  
## $ CarType: Factor w/ 3 levels "Accord","Maxima",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ Age : int 3 2 1 2 2 1 2 3 3 4 ...  
## $ Price : num 15.9 16.4 18.9 16.9 20.5 19 17.5 18 13.6 12 ...  
## $ Mileage: num 17.8 19 20.9 24 24 24.2 30.1 32 34.8 35.7 ...  
## $ Mazda6 : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ Accord : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Maxima : int 0 0 0 0 0 0 0 0 0 0 ...

boxplot (sqrt(Price) ~ CarType, data=ThreeCars2017)



cars1 = lm (sqrt(Price) ~ CarType, data=ThreeCars2017)  
summary (cars1)

##   
## Call:  
## lm(formula = sqrt(Price) ~ CarType, data = ThreeCars2017)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.85980 -0.44900 0.06441 0.57569 1.48415   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.6927 0.1435 25.725 <2e-16 \*\*\*  
## CarTypeMaxima 0.1925 0.2030 0.948 0.3456   
## CarTypeMazda6 -0.4187 0.2030 -2.063 0.0421 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7862 on 87 degrees of freedom  
## Multiple R-squared: 0.09825, Adjusted R-squared: 0.07752   
## F-statistic: 4.74 on 2 and 87 DF, p-value: 0.01112

anova (cars1)

## Analysis of Variance Table  
##   
## Response: sqrt(Price)  
## Df Sum Sq Mean Sq F value Pr(>F)   
## CarType 2 5.860 2.92999 4.7397 0.01112 \*  
## Residuals 87 53.782 0.61819   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Findings: The boxplot shows that there is a difference in mean prices in the three models. More specifically, it looks like Mazda6 has the lower price, on average, than either Accord or Maxima. The median value of Mazda6 is below those of Accord and Maxima. Median price of Accord and Maxima seem to be approximately the same, but the mean value should be higher for Maxima as its box and lower whisker are clearly above the Accord boxplot. After performing anova on the fitted model, we also got a significant p-value of approximately 0.01 (< 0.05) for CarType, meaning that there is likely a significant difference between at least two car types.

summary (emmeans (cars1, pairwise ~ CarType), infer=c(T,T))

## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates

## $emmeans  
## CarType emmean SE df lower.CL upper.CL t.ratio p.value  
## Accord 3.69 0.144 87 3.41 3.98 25.725 <.0001   
## Maxima 3.89 0.144 87 3.60 4.17 27.066 <.0001   
## Mazda6 3.27 0.144 87 2.99 3.56 22.808 <.0001   
##   
## Results are given on the sqrt (not the response) scale.   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## Accord - Maxima -0.193 0.203 87 -0.6766 0.292 -0.948 0.6113   
## Accord - Mazda6 0.419 0.203 87 -0.0654 0.903 2.063 0.1037   
## Maxima - Mazda6 0.611 0.203 87 0.1272 1.095 3.011 0.0095   
##   
## Note: contrasts are still on the sqrt scale   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 3 estimates   
## P value adjustment: tukey method for comparing a family of 3 estimates

Findings: We see that the difference between Maxima and Mazda6 is significant with the p-value of approximately 0.010 and an estimate of 0.611. Differences Accord - Maxima and Accord - Mazda6 are not statistically significant with the p-values of approximately 0.611 and 0.104 respectively.

### 8.62 (b)

ANCOVA with car type and age.

cars2 = lm (sqrt(Price) ~ CarType + Age, data=ThreeCars2017)  
summary (cars2)

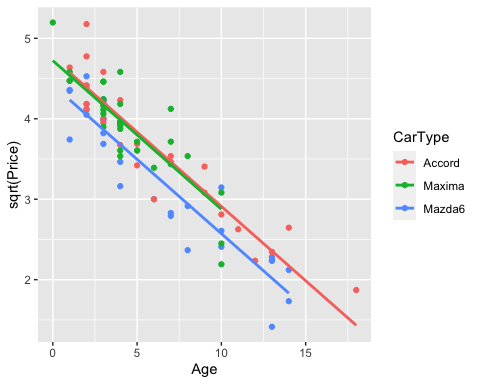
##   
## Call:  
## lm(formula = sqrt(Price) ~ CarType + Age, data = ThreeCars2017)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.68449 -0.19349 -0.01224 0.21052 0.78832   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.75802 0.07452 63.848 < 2e-16 \*\*\*  
## CarTypeMaxima -0.03531 0.08041 -0.439 0.662   
## CarTypeMazda6 -0.33866 0.07982 -4.243 5.55e-05 \*\*\*  
## Age -0.18473 0.00845 -21.862 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3088 on 86 degrees of freedom  
## Multiple R-squared: 0.8625, Adjusted R-squared: 0.8577   
## F-statistic: 179.8 on 3 and 86 DF, p-value: < 2.2e-16

anova (cars2)

## Analysis of Variance Table  
##   
## Response: sqrt(Price)  
## Df Sum Sq Mean Sq F value Pr(>F)   
## CarType 2 5.860 2.930 30.723 8.554e-11 \*\*\*  
## Age 1 45.581 45.581 477.946 < 2.2e-16 \*\*\*  
## Residuals 86 8.202 0.095   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Results: From the summary, we can see that the intercept is significant with the p-value less than 2e-16. Mazda6 car type is also significant with the p-value of 5.55e-05. Age is significant with the p-value of less than 2e-16. Model Maxima is not significant with the p-value of 0.662. The ANCOVA model shows that both CarType and Age are statistically significant with the p-values of 8.554e-11 and 2.2e-16 (in fact, less than 2.2e-16, so Age is very significant) respectively.

library (ggplot2)  
qplot (Age, sqrt(Price), data=ThreeCars2017, color=CarType) +  
 geom\_line (aes(y=predict (cars2)), lwd=1)



summary (emmeans (cars2, pairwise ~ CarType), infer=c(T,T))

## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates

## $emmeans  
## CarType emmean SE df lower.CL upper.CL t.ratio p.value  
## Accord 3.74 0.0564 86 3.63 3.85 66.316 <.0001   
## Maxima 3.71 0.0570 86 3.59 3.82 65.063 <.0001   
## Mazda6 3.40 0.0567 86 3.29 3.52 60.033 <.0001   
##   
## Results are given on the sqrt (not the response) scale.   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## Accord - Maxima 0.0353 0.0804 86 -0.156 0.227 0.439 0.8993   
## Accord - Mazda6 0.3387 0.0798 86 0.148 0.529 4.243 0.0002   
## Maxima - Mazda6 0.3033 0.0810 86 0.110 0.496 3.746 0.0009   
##   
## Note: contrasts are still on the sqrt scale   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 3 estimates   
## P value adjustment: tukey method for comparing a family of 3 estimates

Conclusions: The qplot shows a reasonable pattern suggesting that the older the car, regardless of the car type, the lower the price. That being said, it seems like Maxima car type has a lot of younger cars with some of them being fresh and thus, the prices are high. The oldest Maxima in the dataset is also 10 years old. Car type Accord has some of the oldest cars (> 15). Mazda6 is similar to Accord in terms of younger cars, but does not have cars as old as Accord. Contrasts do not show any significant differences. emmeans shows that the emmeans for Accord, Maxima, and Mazda6 are 3.74, 3.71, and 3.40 respectively with the p-value less than 0.0001 in all cases.

### 8.62 (c)

ANCOVA with car type and Mileage.

cars3 = lm (sqrt(Price) ~ CarType + Mileage, data=ThreeCars2017)  
summary (cars3)

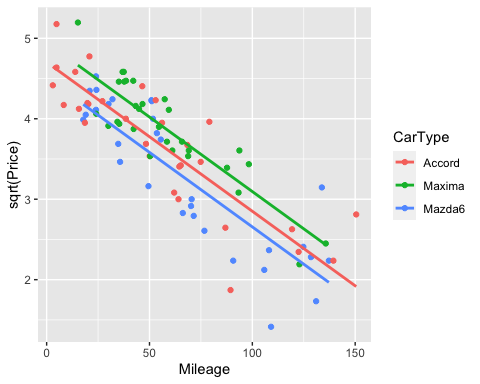
##   
## Call:  
## lm(formula = sqrt(Price) ~ CarType + Mileage, data = ThreeCars2017)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.17580 -0.29061 0.00874 0.26803 1.11687   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.702937 0.097572 48.200 <2e-16 \*\*\*  
## CarTypeMaxima 0.244898 0.103842 2.358 0.0206 \*   
## CarTypeMazda6 -0.194472 0.104765 -1.856 0.0668 .   
## Mileage -0.018527 0.001179 -15.711 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.402 on 86 degrees of freedom  
## Multiple R-squared: 0.767, Adjusted R-squared: 0.7589   
## F-statistic: 94.37 on 3 and 86 DF, p-value: < 2.2e-16

anova (cars3)

## Analysis of Variance Table  
##   
## Response: sqrt(Price)  
## Df Sum Sq Mean Sq F value Pr(>F)   
## CarType 2 5.860 2.930 18.133 2.687e-07 \*\*\*  
## Mileage 1 39.886 39.886 246.849 < 2.2e-16 \*\*\*  
## Residuals 86 13.896 0.162   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Results: The summary of the model cars3 shows that the Intercept and Mileage are both significant with the p-values less than 2e-16 in both cases. ANOVA models shows that both CarType and Mileage are significant with the p-values of 2.687e-07 and 2.2e-16 (less than 2.2e-16) respectively.

qplot (Mileage, sqrt(Price), data=ThreeCars2017, color=CarType) +  
 geom\_line (aes(y=predict (cars3)), lwd=1)



summary (emmeans (cars3, pairwise ~ CarType), infer=c(T,T))

## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates

## $emmeans  
## CarType emmean SE df lower.CL upper.CL t.ratio p.value  
## Accord 3.60 0.0736 86 3.45 3.75 48.904 <.0001   
## Maxima 3.85 0.0734 86 3.70 3.99 52.366 <.0001   
## Mazda6 3.41 0.0739 86 3.26 3.55 46.109 <.0001   
##   
## Results are given on the sqrt (not the response) scale.   
## Confidence level used: 0.95   
##   
## $contrasts  
## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## Accord - Maxima -0.245 0.104 86 -0.4926 0.00276 -2.358 0.0533   
## Accord - Mazda6 0.194 0.105 86 -0.0554 0.44433 1.856 0.1578   
## Maxima - Mazda6 0.439 0.104 86 0.1905 0.68827 4.210 0.0002   
##   
## Note: contrasts are still on the sqrt scale   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 3 estimates   
## P value adjustment: tukey method for comparing a family of 3 estimates

Conclusions: The qplot shows that Accord has some cars with the mileage that is less than that of any car of either Maxima or Mazda6. Accord also has some cars with the most mileage. That being said, Accord model contains all mileages from both Maxima and Mazda6 and hence, we do not expect very significant differences in contrasts. Maxima and Mazda6 look similar in terms of mileage, but Maxima, on average, does seem to have cars with smaller mileage than Mazda6 (i.e., Maxima line is Mazda6 line moved up across y-axis and then shifted left across x-axis). The contrasts show that emmean for Accord, Maxima, and Mazda6 are 3.60, 3.85, and 3.41 respectively with all of the p-values less than 0.0001. The contrasts show that Accord - Maxima difference is marginally significant with the p-value of approximately 0.05 and an estimate of -0.245. Accord - Mazda6 difference is not significant as its p-value is approximately 0.16. Maxima - Mazda6 contrast, however, is significant (as expected) with the p-value of 0.0002 and an estimate of 0.439.

### 8.62 (d)

Your choice of final model and why?

I would go with cars3 model. This is due to the fact that mileage might be more important metric of assessing comparing cars than age. One might have a car which is 10 years old but with no mileage as well as a car that is 1 year old but with a lot of mileage. Due to the same reason, the car mileage also speaks a lot to the condition of the car and from a more practical perspective, tells us that some car models are able to withstand more miles (e.g., while Maxima model had some of the youngest cars, Accord had some cars whose mileage were clearly bigger than any other car in Maxima or Mazda6). Hence, mileage can also help us make decisions if purchasing a particular car model (e.g., maybe Accord is better as we have some evidence that its cars could drive a lot longer, more miles, etc).