Assignment 3

```r
AAA <- read.delim("/Users/loansaori/Downloads/AAA.dat", sep=" ")
```

PART A:

1. Using a linear model what is the interpretation of the estimated slope parameter?

```r
AAA.linear <- lm(Demand ~ Price, data = AAA)
summary(AAA.linear)
##
## Call:
## lm(formula = Demand ~ Price, data = AAA)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -85.42 -48.93 -15.40  48.30 165.54
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9666.81     111.68   86.56   <2e-16 ***
## Price       -1770.12      31.79  -55.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 62.38 on 40 degrees of freedom
## Multiple R-squared:  0.9873, Adjusted R-squared:  0.9869
## F-statistic:  3100 on 1 and 40 DF,  p-value: < 2.2e-16
```

In linear model, if the price a gallon of gasoline increase 1 dollar, the number of sold per gasoline station decrease 1770.12 gallons.

---

2. Using a Log-linear model what is the interpretation of the estimated slope parameter?

```r
AAA["LnDemand"] <- log(AAA$Demand)
AAAlog.lin <- lm(LnDemand ~ Price, data = AAA)
summary(AAAlog.lin)
##
## Call:
## lm(formula = LnDemand ~ Price, data = AAA)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.012949 -0.005648 -0.001026  0.004325  0.021548
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.92407    0.01451   684.0   <2e-16 ***
## Price       -0.50964    0.00413  -123.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008104 on 40 degrees of freedom
## Multiple R-squared:  0.9974, Adjusted R-squared:  0.9973
## F-statistic: 1.523e+04 on 1 and 40 DF,  p-value: < 2.2e-16
exp(-0.50964)
## [1] 0.6007118
```

In log-lin model, 1 dollar increase in the price a gallon of gasoline corresponds to approximately an expected decrease in the number of sold per gasoline station of 0.6

---

3. Using a Log-Log model what is the interpretation of the estimated slope parameter?

```r
AAA["LnPrice"] <- log(AAA$Price)
AAA["LnDemand"] <- log(AAA$Demand)
AAAln.lm <- lm(LnDemand ~ LnPrice, data = AAA)
summary(AAAln.lm)
```

```
##
## Call:
## lm(formula = LnDemand ~ LnPrice, data = AAA)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.009590 -0.004378  0.000003  0.004344  0.009639
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.358176   0.011827   875.8   <2e-16 ***
## LnPrice     -1.775692   0.009446  -188.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.005324 on 40 degrees of freedom
## Multiple R-squared:  0.9989, Adjusted R-squared:  0.9988
## F-statistic: 3.533e+04 on 1 and 40 DF,  p-value: < 2.2e-16
```

In log-log model,1.775692 is the expected percentage decrease in the number of sold per gasoline station for a 1 % increase in the price a gallon of gasoline.

---

4. Which of these three models would you expect to provide the best model for predicting the relationship between price and demand? Why?

Consider the Residual standard error in each model, RMSE of linear model = 62.38, RMSE of log-lin model = 0.008104 and RMSE of log-log model = 0.005324. We can see log-log model predicts the relationship between price and demand with smallest errors, therefore, this is the best model.

---

5. If price of gasoline is $3.oo per gallon, what is the expected demand under a linear model?

```
9666.81 -1770.12*3
## [1] 4356.45
```

Linear model equation: demand = 9666.81 -1770.12*price With price of gasoline is $3.oo per gallon, expected demand is 4356.45

---

6. If price of gasoline is $3.oo per gallon, what is the expected demand under a Log-linear model?

```
exp (9.92407 -0.50964*3)
## [1] 4425.551
```

Linear model equation: log (demand) = 9.92407 -0.50964*price With price of gasoline is $3.oo per gallon, expected demand is 4425.551

---

7. If price of gasoline is $3.oo per gallon, what is the expected demand under a Log-Log model?

```
exp(10.358176 - 1.775692*log(3) )
## [1] 4480.003
```

Linear model equation: log (demand) = 10.358176 - 1.775692*log(price) With price of gasoline is $3.oo per gallon, expected demand is 4480.003

PART B:

1. Show that e^(B_1)= lambda_B / lambda_A .

x = 0 for treatment A, so log (lambda_A) = B_0 + B_1*0 = B_0 x = 1 for treatment B, so log (lambda_B) = B_0 + B_1*1 = B_0 + B_1

By subtraction: B_1 = log (lambda_B) - log (lambda_A) = log (lambda_B/(lambda_A)
Since B_1 = log (lambda_B / lambda_A), therefore e^(B_1)= lambda_B / lambda_A

---

2. Write down the estimated model and interpret B_1.

```
treatment_a <- c(8, 7, 6, 6, 3, 4, 7, 2, 3, 4)
treatment_b <- c(9, 9, 8, 14, 8, 13, 11, 5, 7, 6)
y <- c (treatment_a, treatment_b)
y
## [1]  8  7  6  6  3  4  7  2  3  4  9  9  8 14  8 13 11  5  7  6
X <- c(rep(0,10),rep (1,10))
X
## [1] 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
poisson_model <- glm(y ~ X, family = poisson(link = log))
summary(poisson_model)
##
## Call:
## glm(formula = y ~ X, family = poisson(link = log))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5280  -0.7622  -0.1699   0.6938   1.5399
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.6094     0.1414  11.380  < 2e-16 ***
## X             0.5878     0.1764   3.332 0.000861 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 27.857  on 19  degrees of freedom
## Residual deviance: 16.268  on 18  degrees of freedom
## AIC: 94.349
##
## Number of Fisher Scoring iterations: 4
```

The poisson regression equation is log (lambda) = 1.6094 + 0.5878*X

Interpret B_1: log (lambda_B/(lambda_A) = 0.5878 -> lambda_B = e^(0.5878)* lambda_A = 1.8* lambda_A Conclusion: the estimated average number of imperfections using treatment B is about 1.8 times that using treatment A

---

3. Carry out a test for H_0: lambda_A = lambda_B against H_1: lambda_A ≠ lambda_B.

Using the result from Question (2) above: lambda_B = 1.8* lambda_A. Therefore, lambda_A = lambda_B if and only if B_1 = 0. Therefore, Test for H_0: lambda_A = lambda_B is as same as test H_0: B_1 = 0 However, the p-value of B_1 in Question (2) outcome is 0.000861 < 0.05, so we reject the null hypothesis H_0: B_1 = 0 in favour of the alternative Ha: B_1 ≠ 0 Equivalently we reject H_0: lambda_A = lambda_B in favour of H_1: lambda_A ≠ lambda_B

---

4. Construct an approximate 95% confidence interval for lambda_B / lambda_A .

95% CI for B_1 is 0.588 ± 1.96*0.176 = (0.242, 0.934) 95% CI for lambda_B / lambda_A is exp(0.242, 0.934) = (1.27, 2.54).
