#Amulya\_geereddy

#U62067273

**Preprocessing:**

library(rio)

library(car)

my\_data =import("6304 Module 8 Assignment Data.xlsx")

my\_data1=my\_data[(my\_data$cylinders==4)|(my\_data$cylinders==6)|(my\_data$cylinders==8),]

my\_data2 =my\_data1[(my\_data1$fuel=="gas")|(my\_data1$fuel=="diesel"),]

set.seed(62067273)

illinois\_data=subset(my\_data2, region=="champaign urbana, IL"|

region=="chicago, IL"|

region=="danville, IL"|

region=="peoria, IL"|

region=="quad cities, IA/IL"|

region=="rockford, IL"|

region=="southern illinois, IL"|

region=="springfield, IL")

texas\_data =subset(my\_data2, region =="amarillo, TX"|

region== "austin, TX"|

region== "brownsville, TX"|

region=="college station, TX"|

region=="corpus christi, TX"|

region=="dallas/fort worth"|

region=="el paso, TX"|

region== "galveston, TX"|

region=="houston, TX"|

region=="lubbock, TX"|

region=="odessa/midland"|

region=="tyler/east, TX"|

region=="waco, TX")

nc\_data =subset(my\_data2 ,region=="asheville, NC"|

region=="boone, NC"|

region=="charlotte, NC"|

region=="eastern, NC"|

region=="fayetteville, NC"|

region=="greensboro, NC"|

region=="wilmington, NC"|

region=="winston-salem, NC")

illinois\_data$state=rep('ill',nrow(illinois\_data))

texas\_data$state =rep("texas",nrow(texas\_data))

nc\_data$state =rep("nc",nrow(nc\_data))

illinois\_sample=illinois\_data[sample(1:nrow(illinois\_data),150,replace=FALSE),]

texas\_sample=texas\_data[sample(1:nrow(texas\_data),150,replace=FALSE),]

nc\_sample=nc\_data[sample(1:nrow(nc\_data),150,replace=FALSE),]

totalsample=rbind(illinois\_sample,texas\_sample,nc\_sample)

colnames(totalsample) = tolower(make.names(colnames(totalsample)))

totalsample$state=as.factor(totalsample $state)

**Analysis1**

leveneTest(asking.price~state,data=total\_sample)

**Output:**

> leveneTest(asking.price~state,data=totalsample)

Levene's Test for Homogeneity of Variance (center = median)

Df F value Pr(>F)

group 2 7.2256 0.0008159 \*\*\*

447

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**From the above, P-value is 0.0008** which is less than 5% and it is significant, therefore we can reject the null hypothesis. So, **we can say that there is at least one variance that is different**.

**Analysis2**

anov\_test= aov(asking.price~state,data=totalsample)

summary(anov\_test)

tur\_test=TukeyHSD(anov\_test)

tur\_test

par(mar=c(5.1,8,4.1,2.1))

plot(tur\_test,las=1.5,cex.axis=.8)

par(mar=c(5.1,4.1,4.1,2.1))

**Output:**

> anov\_test= aov(asking.price~state,data=totalsample)

> summary(anov\_test)

Df Sum Sq Mean Sq F value Pr(>F)

state 2 9.938e+08 496896973 4.845 0.00829 \*\*

Residuals 447 4.584e+10 102558545

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

From the above values we can say that the Null hypothesis states that the mean asking.price for all states are equal. However, the p-value of ANOVA test gives a **significant value of 0.008**. Therefore, **we do have enough evidence to reject the Null Hypothesis.**

> tur\_test=TukeyHSD(anov\_test)

> tur\_test

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = asking.price ~ state, data = totalsample)

$state

diff lwr upr p adj

nc-ill -12.88667 -2762.7350 2736.962 0.9999330

texas-ill 3145.98667 396.1383 5895.835 0.0202128

texas-nc 3158.87333 409.0250 5908.722 0.0195888

>

> par(mar=c(5.1,8,4.1,2.1))

> plot(tur\_test,las=1.5,cex.axis=.8)

> par(mar=c(5.1,4.1,4.1,2.1))

Chart, box and whisker chart

Description automatically generated

The p-values from the results of TuckeyHSD clearly states that Texas-Illinois comparisons are statistically significant.

**Analysis3**

leveneTest(odometer~state,data=totalsample)

anov\_test2= aov(odometer~state,data=totalsample)

summary(anov\_test2)

tur\_test2=TukeyHSD(anov\_test2)

tur\_test2

par(mar=c(5.1,8,4.1,2.1))

plot(tur\_test2,las=1.5,cex.axis=.8)

par(mar=c(5.1,4.1,4.1,2.1))

**Output:**

> leveneTest(odometer~state,data=totalsample)

Levene's Test for Homogeneity of Variance (center = median)

Df F value Pr(>F)

group 2 0.6844 0.5049

447

>

> anov\_test2= aov(odometer~state,data=totalsample)

> summary(anov\_test2)

Df Sum Sq Mean Sq F value Pr(>F)

state 2 1.174e+10 5.872e+09 1.469 0.231

Residuals 447 1.787e+12 3.997e+09

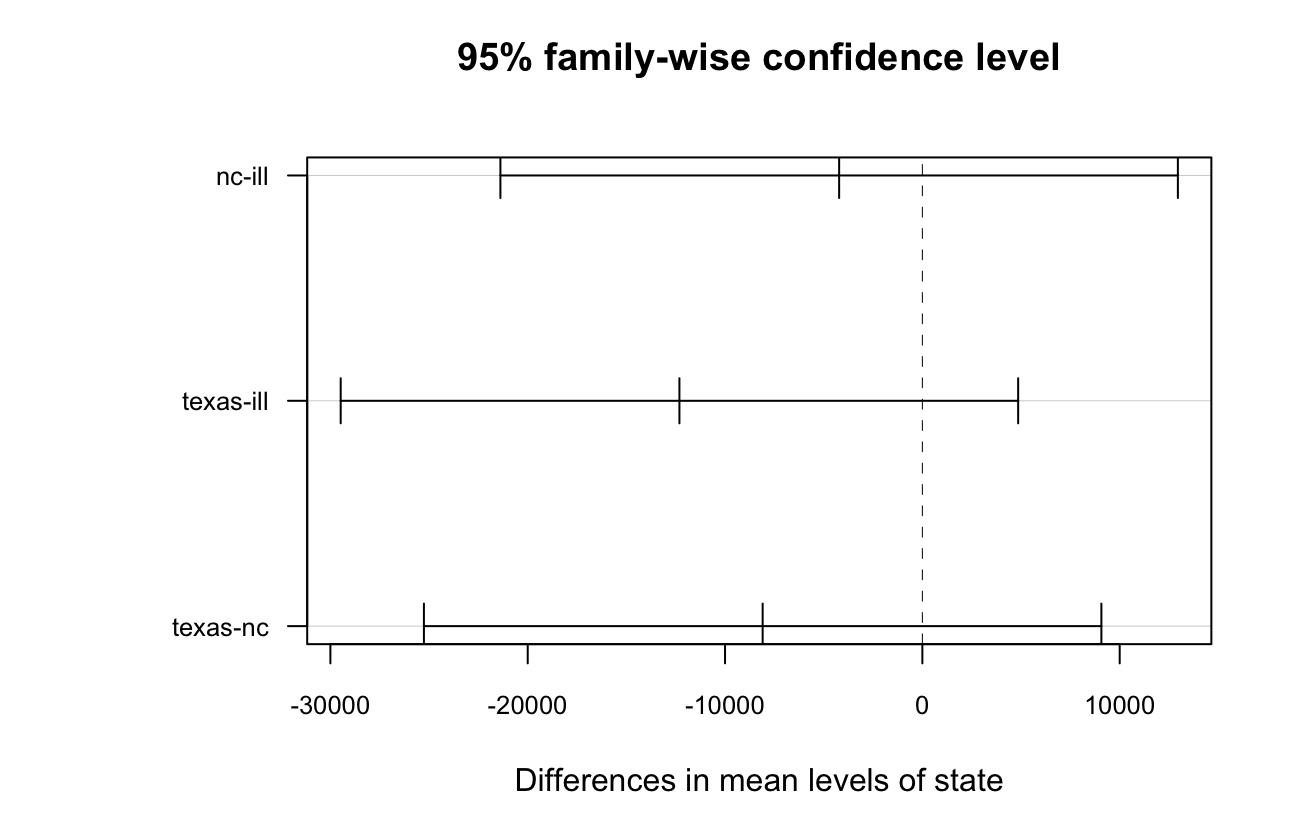
From the above values we can state that the Null Hypothesis states that the mean odometer for all states is equal. However, the p-value of the ANOVA test gives a **non-significant value of 0.23**. Therefore, we **do not have enough evidence to reject the Null Hypothesis**.

> tur\_test2=TukeyHSD(anov\_test2)

> tur\_test2

Tukey multiple comparisons of means

95% family-wise confidence level



The p-values from the results of TuckeyHSD clearly states that all possible comparisons are statistically not significant.

**Analysis4**

anov\_test3= aov(`asking price`~region,data=texas\_sample)

summary(anov\_test3)

tur\_test3=TukeyHSD(anov\_test3)

tur\_test3

par(mar=c(2.1,12,4.1,2.1))

plot(tur\_test3,las=1,cex.axis=0.6)

par(mar=c(5.1,4.1,4.1,2.1))

**Output:**

> anov\_test3= aov(`asking price`~region,data=texas\_sample)

> summary(anov\_test3)

Df Sum Sq Mean Sq F value Pr(>F)

region 8 2.778e+09 347207314 2.096 0.0399 \*

Residuals 141 2.335e+10 165626924

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Null Hypothesis states that the mean asking.price for all regions are equal. However, the p-value of the ANOVA test gives a **significant value of 0.039**. Therefore, **we do have enough evidence to reject the Null Hypothesis**.

> tur\_test3=TukeyHSD(anov\_test3)

> tur\_test3

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = `asking price` ~ region, data = texas\_sample)

$region

diff lwr upr p adj

austin, TX-amarillo, TX -517.1429 -16771.974 15737.6882 1.0000000

brownsville, TX-amarillo, TX -1235.0000 -24350.160 21880.1598 1.0000000

college station, TX-amarillo, TX -4952.3333 -32402.560 22497.8937 0.9997228

corpus christi, TX-amarillo, TX 7623.1250 -12650.211 27896.4613 0.9585369

el paso, TX-amarillo, TX 8808.4231 -7584.749 25201.5956 0.7499153

houston, TX-amarillo, TX 3216.2632 -12556.121 18988.6472 0.9993148

lubbock, TX-amarillo, TX -3316.2500 -20873.474 14240.9742 0.9996101

waco, TX-amarillo, TX -2766.1111 -19995.134 14462.9118 0.9998846

brownsville, TX-austin, TX -717.8571 -20403.435 18967.7209 1.0000000

college station, TX-austin, TX -4435.1905 -29067.006 20196.6251 0.9997268

corpus christi, TX-austin, TX 8140.2679 -8114.563 24395.0989 0.8152493

el paso, TX-austin, TX 9325.5659 -1717.417 20368.5491 0.1718598

houston, TX-austin, TX 3733.4060 -6365.082 13831.8945 0.9624662

lubbock, TX-austin, TX -2799.1071 -15506.093 9907.8788 0.9988000

waco, TX-austin, TX -2248.9683 -14498.483 10000.5460 0.9996841

college station, TX-brownsville, TX -3717.3333 -33328.436 25893.7696 0.9999825

corpus christi, TX-brownsville, TX 8858.1250 -14257.035 31973.2848 0.9537090

el paso, TX-brownsville, TX 10043.4231 -9756.540 29843.3866 0.8043779

houston, TX-brownsville, TX 4451.2632 -14837.867 23740.3935 0.9983255

lubbock, TX-brownsville, TX -2081.2500 -22855.238 18692.7379 0.9999970

waco, TX-brownsville, TX -1531.1111 -22028.469 18966.2467 0.9999997

corpus christi, TX-college station, TX 12575.4583 -14874.769 40025.6854 0.8788789

el paso, TX-college station, TX 13760.7564 -10962.571 38484.0836 0.7121806

houston, TX-college station, TX 8168.5965 -16147.548 32484.7414 0.9790286

lubbock, TX-college station, TX 1636.0833 -23873.972 27146.1387 0.9999999

waco, TX-college station, TX 2186.2222 -23099.071 27471.5150 0.9999990

el paso, TX-corpus christi, TX 1185.2981 -15207.874 17578.4706 0.9999998

houston, TX-corpus christi, TX -4406.8618 -20179.246 11365.5222 0.9936756

lubbock, TX-corpus christi, TX -10939.3750 -28496.599 6617.8492 0.5715325

waco, TX-corpus christi, TX -10389.2361 -27618.259 6839.7868 0.6150160

houston, TX-el paso, TX -5592.1599 -15911.852 4727.5320 0.7411462

lubbock, TX-el paso, TX -12124.6731 -25008.153 758.8073 0.0825756

waco, TX-el paso, TX -11574.5342 -24007.039 857.9705 0.0896400

lubbock, TX-houston, TX -6532.5132 -18616.224 5551.1973 0.7436285

waco, TX-houston, TX -5982.3743 -17584.062 5619.3137 0.7897729

waco, TX-lubbock, TX 550.1389 -13381.379 14481.6567 1.0000000

>

> par(mar=c(2.1,12,4.1,2.1))

> plot(tur\_test3,las=1,cex.axis=0.6)

> par(mar=c(5.1,4.1,4.1,2.1))

Chart

Description automatically generated

The p-values from the results of TuckeyHSD conveys that all possible comparisons are statistically not significant.

**Analysis 5**

anov\_test4=aov(asking.price~fuel+condition,data=totalsample)

summary(anov\_test4)

tur\_test4=TukeyHSD(anov\_test4)

tur\_test4

par(mfrow=c(1,2))

par(mar=c(5.1,8,4.1,2.1))

plot(tur\_test4,las=1.5,cex.axis=.8)

par(mfrow=c(1,1))

par(mar=c(5.1,4.1,4.1,2.1))

**Output:**

> anov\_test4=aov(asking.price~fuel+condition,data=totalsample)

> summary(anov\_test4)

Df Sum Sq Mean Sq F value Pr(>F)

fuel 1 4.594e+09 4.594e+09 56.71 2.85e-13 \*\*\*

condition 5 6.356e+09 1.271e+09 15.69 3.13e-14 \*\*\*

Residuals 443 3.589e+10 8.101e+07

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

P value is very much less than 0.05. Therefore, **we do have enough evidence to reject the Null Hypothesis.**

> tur\_test4=TukeyHSD(anov\_test4)

> tur\_test4

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = asking.price ~ fuel + condition, data = totalsample)

$fuel

diff lwr upr p adj

gas-diesel -12809.41 -16152.33 -9466.495 0

$condition

diff lwr upr p adj

fair-excellent -10303.737 -16442.4299 -4165.044 0.0000313

good-excellent -6185.993 -8968.7542 -3403.231 0.0000000

like new-excellent 3432.633 -417.0991 7282.366 0.1116161

new-excellent -5129.421 -30944.5151 20685.672 0.9929793

salvage-excellent -12628.421 -38443.5151 13186.672 0.7269696

good-fair 4117.745 -2200.4605 10435.950 0.4250101

like new-fair 13736.370 6880.9789 20591.762 0.0000003

new-fair 5174.316 -21256.6324 31605.264 0.9934481

salvage-fair -2324.684 -28755.6324 24106.264 0.9998622

like new-good 9618.626 5488.6614 13748.590 0.0000000

new-good 1056.571 -24801.7975 26914.940 0.9999969

salvage-good -6442.429 -32300.7975 19415.940 0.9803276

new-like new -8562.055 -34556.8985 17432.789 0.9351381

salvage-like new -16061.055 -42055.8985 9933.789 0.4873182

salvage-new -7499.000 -43931.5478 28933.548 0.9917330

>

> par(mfrow=c(1,2))

> par(mar=c(5.1,8,4.1,2.1))

> plot(tur\_test4,las=1.5,cex.axis=.8)

> par(mfrow=c(1,1))

> par(mar=c(5.1, 4.1,4.1,2.1))Chart, box and whisker chart

Description automatically generated