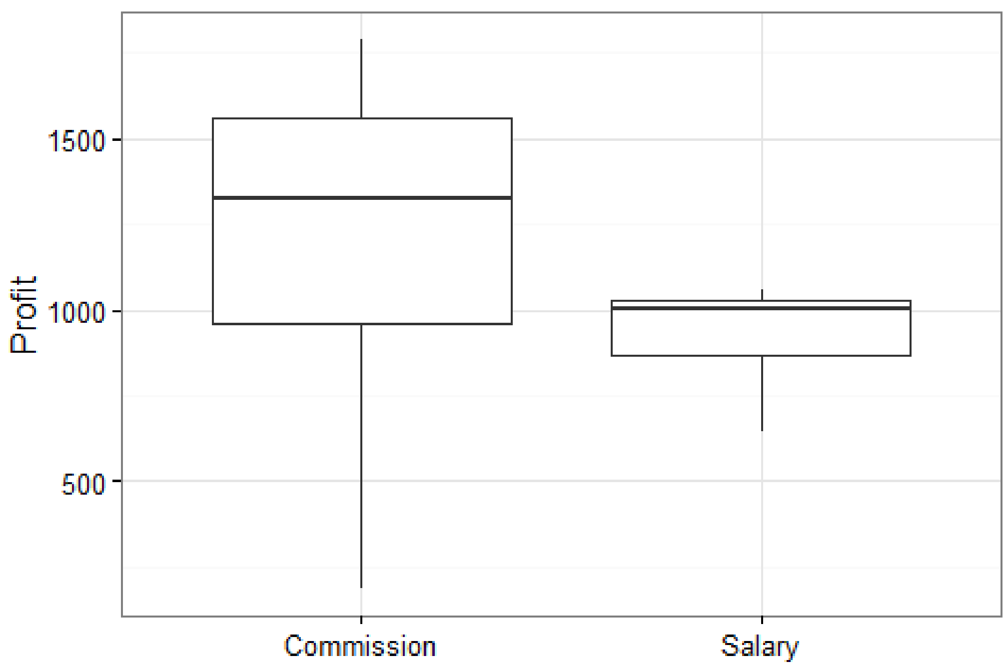
In order to help managers to solve this question, we need to construct a linear regression model so that we can see how significant of each variable. Based on our requirement, we need to construct a single linear regression model.

My regression model is as follows: Profit = B0+Commison*X1 +Outlets*X2+ Commission*Outlets*X3+Area*X4+Population*X5+Area*Population*X6+e.

In order to show the result, graph is one of the most important aspects. I made a box-plot to show how commission and salary affect profit.



As we can see, Salary has relatively low median than Commission. 25% and 75% percentile shows that Salary has very low variance and Commission has relatively high variance. Also, from the R code result, the p value for Commission is 0.00045. This is a binary string. X1 will take value 1 if it is Salary and 0 for Commission. The P value is too small so that we can say that we reject the null hypothesis. Compensation status play a role in the amount of profits.

Then, we look at our model again, the p value for interaction term (Commission*Outlets) is 3.90e-06. It shows that there is an effect of the compensation status depend upon the number of outlets being served. We take the value 0 for commission-only. Our regression model become Profit = B0+ +Outlets*X1+Area*X2+Population*X3+Area*Population*X4+e. Based on the estimate value it has positive effect on the productive.

Manager also believe that larger districts most likely have an effect on the productivity of the representative. We intuitively think that large area will present same result with large population.

The slope of this react differently for commission. I believe that we need to add a weight function so that we prove managers’ assumption.

Appendix

```
library(ggplot2)

url <- "C:/Users/shiy/Desktop/2015/SalesReps.csv"

sale.reps <- read.table(url, header=TRUE, sep=",", na.strings="?",

                stringsAsFactors=FALSE)



commis <- sale.reps$Commission

outlets <- sale.reps$Outlets

prof <- sale.reps$Profit

a <- sale.reps$Area

pop <- (sale.reps$Popn)



fit <- lm(prof ~ commis + commis:outlets + outlets + a + pop + a:pop)

summary(fit)



# Construct Plot

ggplot(data=sale.reps, mapping=aes(x=Commission, y=Profit)) +

  geom_boxplot() +
```

```
  labs(x="", y="Profit") +
  theme_bw(16)

ggplot(data=sale.reps,
       mapping=aes(x=Popn, y=Profit, colour=Commission)) +
  geom_point(size=3) +
  labs(x="population",
       y="Profit of representative",
       colour="Commission") +
  theme_bw(16) +
  theme(legend.position="bottom")

ggplot(data=sale.reps,
       mapping=aes(x=Area, y=Profit, colour=Commission)) +
  geom_point(size=3) +
  labs(x="Area of District",
       y="Profit of representative",
       colour="Commission") +
```

```
  theme_bw(16) +

  theme(legend.position="bottom")
```

```
fit1 <- lm(prof ~ a + pop + outlets)
```

```
Call:
lm(formula = prof ~ commis + commis:outlets + outlets + a + pop +
    a:pop)

Residuals:
    Min      1Q  Median      3Q     Max
-347.16  -46.45   15.56   73.71  284.93

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           202.562    234.888   0.862  0.39316
commisSalary         1239.568    326.768   3.793  0.00045 ***
outlets                 5.654      1.094   5.170 5.49e-06 ***
a                     -13.598      6.165  -2.206  0.03268 *
pop                   101.726     43.857   2.319  0.02507 *
commisSalary:outlets   -9.175      1.740  -5.273 3.90e-06 ***
a:pop                  -1.979      1.881  -1.052  0.29864
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 134.2 on 44 degrees of freedom
Multiple R-squared:  0.8767,	Adjusted R-squared:  0.8599
F-statistic: 52.16 on 6 and 44 DF,  p-value: < 2.2e-16
```