Tugas 03 Solutions

Roni Yunis

4/23/2021

Load Package

```
library (dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
library (lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##      date, intersect, setdiff, union
```

```
library (ggplot2)
library (caTools)
library (lubridate)
```

Import Dataset dan summary data

```
garment <- read.csv("data/garments_worker_productivity.csv")
head(garment)
```

```
##        date  quarter department        day team targeted_productivity   smv  wip
## 1 1/1/2015 Quarter1     sweing Thursday    8                  0.80 26.16 1108
## 2 1/1/2015 Quarter1  finishing  Thursday    1                  0.75  3.94   NA
## 3 1/1/2015 Quarter1     sweing Thursday   11                  0.80 11.41  968
## 4 1/1/2015 Quarter1     sweing Thursday   12                  0.80 11.41  968
## 5 1/1/2015 Quarter1     sweing Thursday    6                  0.80 25.90 1170
## 6 1/1/2015 Quarter1     sweing Thursday    7                  0.80 25.90  984
##   over_time incentive idle_time idle_men no_of_style_change no_of_workers
## 1      7080        98         0        0                  0          59.0
## 2       960         0         0        0                  0           8.0
## 3      3660        50         0        0                  0          30.5
## 4      3660        50         0        0                  0          30.5
## 5      1920        50         0        0                  0          56.0
## 6      6720        38         0        0                  0          56.0
##   actual_productivity
## 1           0.9407254
## 2           0.8865000
## 3           0.8005705
## 4           0.8005705
## 5           0.8003819
## 6           0.8001250
```

summary(garment)

```
##      date              quarter           department             day
##  Length:1197        Length:1197        Length:1197        Length:1197
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##       team        targeted_productivity      smv              wip
##  Min.   : 1.000   Min.   :0.0700        Min.   : 2.90   Min.   :     7.0
##  1st Qu.: 3.000   1st Qu.:0.7000        1st Qu.: 3.94   1st Qu.:  774.5
##  Median : 6.000   Median :0.7500        Median :15.26   Median : 1039.0
##  Mean   : 6.427   Mean   :0.7296        Mean   :15.06   Mean   : 1190.5
##  3rd Qu.: 9.000   3rd Qu.:0.8000        3rd Qu.:24.26   3rd Qu.: 1252.5
##  Max.   :12.000   Max.   :0.8000        Max.   :54.56   Max.   :23122.0
##                                                         NA's   :506
##    over_time       incentive          idle_time          idle_men
##  Min.   :    0   Min.   :   0.00   Min.   :  0.0000   Min.   : 0.0000
##  1st Qu.: 1440   1st Qu.:   0.00   1st Qu.:  0.0000   1st Qu.: 0.0000
##  Median : 3960   Median :   0.00   Median :  0.0000   Median : 0.0000
##  Mean   : 4567   Mean   :  38.21   Mean   :  0.7302   Mean   : 0.3693
##  3rd Qu.: 6960   3rd Qu.:  50.00   3rd Qu.:  0.0000   3rd Qu.: 0.0000
##  Max.   :25920   Max.   :3600.00   Max.   :300.0000   Max.   :45.0000
##
## no_of_style_change no_of_workers   actual_productivity
##  Min.   :0.0000     Min.   : 2.00   Min.   :0.2337
##  1st Qu.:0.0000     1st Qu.: 9.00   1st Qu.:0.6503
##  Median :0.0000     Median :34.00   Median :0.7733
##  Mean   :0.1504     Mean   :34.61   Mean   :0.7351
##  3rd Qu.:0.0000     3rd Qu.:57.00   3rd Qu.:0.8503
```

```
## Max.    :2.0000     Max.    :89.00   Max.    :1.1204
##
```

Bisa dilihat bahwa pada variabel `wip` ada 506 data NA's atau missing value.

Exploratory Data Analysis dan Visualisasi

Membersihkan data yang missing value

data yang sudah hapus missing value di simpan pada objek `clean_garment`

```
clean_garment <- na.omit(garment)
summary(clean_garment)
```

```
##      date              quarter           department            day
##  Length:691         Length:691         Length:691         Length:691
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      team         targeted_productivity      smv              wip
##  Min.   : 1.000   Min.   :0.070        Min.   :10.05   Min.   :     7.0
##  1st Qu.: 4.000   1st Qu.:0.700        1st Qu.:18.79   1st Qu.:  774.5
##  Median : 6.000   Median :0.750        Median :22.52   Median : 1039.0
##  Mean   : 6.521   Mean   :0.724        Mean   :23.25   Mean   : 1190.5
##  3rd Qu.:10.000   3rd Qu.:0.800        3rd Qu.:28.08   3rd Qu.: 1252.5
##  Max.   :12.000   Max.   :0.800        Max.   :54.56   Max.   :23122.0
##    over_time         incentive        idle_time          idle_men
##  Min.   :    0    Min.   :  0.00    Min.   :  0.000   Min.   : 0.0000
##  1st Qu.: 4560    1st Qu.: 30.00    1st Qu.:  0.000   1st Qu.: 0.0000
##  Median : 6840    Median : 45.00    Median :  0.000   Median : 0.0000
##  Mean   : 6508    Mean   : 44.48    Mean   :  1.265   Mean   : 0.6397
##  3rd Qu.: 7200    3rd Qu.: 60.00    3rd Qu.:  0.000   3rd Qu.: 0.0000
##  Max.   :25920    Max.   :138.00    Max.   :300.000   Max.   :45.0000
##  no_of_style_change no_of_workers   actual_productivity
##  Min.   :0.0000     Min.   :26.00   Min.   :0.2337
##  1st Qu.:0.0000     1st Qu.:52.00   1st Qu.:0.6615
##  Median :0.0000     Median :57.00   Median :0.7506
##  Mean   :0.2605     Mean   :52.45   Mean   :0.7220
##  3rd Qu.:0.0000     3rd Qu.:58.00   3rd Qu.:0.8004
##  Max.   :2.0000     Max.   :89.00   Max.   :1.1005
```

Bisa dilihat tidak ada lagi data yang missing value (NA's)

Melihat stuktur data dari objek `clean_garment`

```
glimpse(clean_garment)
```

```
## Rows: 691
## Columns: 15
## $ date                  <chr> "1/1/2015", "1/1/2015", "1/1/2015", "1/1/2015...
## $ quarter               <chr> "Quarter1", "Quarter1", "Quarter1", "Quarter1...
## $ department            <chr> "sweing", "sweing", "sweing", "sweing", "swei...
## $ day                   <chr> "Thursday", "Thursday", "Thursday", "Thursday...
## $ team                  <int> 8, 11, 12, 6, 7, 3, 2, 1, 9, 10, 5, 4, 1, 3, ...
## $ targeted_productivity <dbl> 0.80, 0.80, 0.80, 0.80, 0.80, 0.75, 0.75, 0.7...
## $ smv                   <dbl> 26.16, 11.41, 11.41, 25.90, 25.90, 28.08, 19....
## $ wip                   <int> 1108, 968, 968, 1170, 984, 795, 733, 681, 872...
## $ over_time             <int> 7080, 3660, 3660, 1920, 6720, 6900, 6000, 690...
## $ incentive             <int> 98, 50, 50, 50, 38, 45, 34, 45, 44, 45, 50, 0...
## $ idle_time             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ idle_men              <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ no_of_style_change    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ no_of_workers         <dbl> 59.0, 30.5, 30.5, 56.0, 56.0, 57.5, 55.0, 57....
## $ actual_productivity   <dbl> 0.9407254, 0.8005705, 0.8005705, 0.8003819, 0...
```

Bisa dilihat bahwa data yang sudah bersih dari missing value tinggal 691 observasi baris dengan 15 kolom/variabel

Merubah type data tanggal dari char menjadi date

```
garment <- clean_garment %>%
  mutate (date = dmy(date)) %>%
  arrange(date)
```

```
## Warning: Problem with `mutate()` input `date`.
## i  325 failed to parse.
## i Input `date` is `dmy(date)`.
```

```
## Warning: 325 failed to parse.
```

```
glimpse(garment)
```

```
## Rows: 691
## Columns: 15
## $ date                  <date> 2015-01-01, 2015-01-01, 2015-01-01, 2015-01-...
## $ quarter               <chr> "Quarter1", "Quarter1", "Quarter1", "Quarter1...
## $ department            <chr> "sweing", "sweing", "sweing", "sweing", "swei...
## $ day                   <chr> "Thursday", "Thursday", "Thursday", "Thursday...
## $ team                  <int> 8, 11, 12, 6, 7, 3, 2, 1, 9, 10, 5, 4, 2, 1, ...
## $ targeted_productivity <dbl> 0.80, 0.80, 0.80, 0.80, 0.80, 0.75, 0.75, 0.7...
## $ smv                   <dbl> 26.16, 11.41, 11.41, 25.90, 25.90, 28.08, 19....
## $ wip                   <int> 1108, 968, 968, 1170, 984, 795, 733, 681, 872...
## $ over_time             <int> 7080, 3660, 3660, 1920, 6720, 6900, 6000, 690...
## $ incentive             <int> 98, 50, 50, 50, 38, 45, 34, 45, 44, 45, 50, 0...
## $ idle_time             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ idle_men              <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ no_of_style_change    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ no_of_workers         <dbl> 59.0, 30.5, 30.5, 56.0, 56.0, 57.5, 55.0, 57....
## $ actual_productivity   <dbl> 0.9407254, 0.8005705, 0.8005705, 0.8003819, 0...
```

Menampilkan banyaknya transaksi berdasarkan variabel quarter

```
garment %>%
  group_by(quarter) %>%
  count() %>%
  arrange(-n)
```

```
## # A tibble: 5 x 2
## # Groups:   quarter [5]
##   quarter       n
##   <chr>     <int>
## 1 Quarter1    211
## 2 Quarter2    188
## 3 Quarter4    140
## 4 Quarter3    129
## 5 Quarter5     23
```

Hasil analisis menunjukkan bahwa quarter dengan jumlah transaksi terbanyak adalah pada Quarter 1 se-
banyak 211 jumlah transaksi.

Menampilkan jumlah actual productivity berdasarkan tanggal

```
garment %>%
  group_by(date) %>%
  summarise(
    jmlActPro = sum(actual_productivity)
    )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 32 x 2
##    date       jmlActPro
##    <date>         <dbl>
##  1 2015-01-01      9.06
##  2 2015-01-02      8.38
##  3 2015-01-03      8.46
##  4 2015-02-02      8.56
##  5 2015-02-03      8.40
##  6 2015-03-01      9.34
##  7 2015-03-02      9.20
##  8 2015-03-03      8.71
##  9 2015-04-01      9.44
## 10 2015-04-02      9.35
## # ... with 22 more rows
```

Memfilter nilai actual productity berdasarkan quarter 1

```
garmentquarter1 <- filter(garment, quarter == "Quarter1")
```

```
daily_productivity_q1 <- garmentquarter1 %>%
  group_by(date) %>%
  summarise(
    productivity = sum(actual_productivity)
    )
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
head(arrange(daily_productivity_q1, (-productivity)))
```

```
## # A tibble: 6 x 2
##   date       productivity
##   <date>            <dbl>
## 1 2015-06-01         9.45
## 2 2015-04-01         9.44
## 3 2015-04-02         9.35
## 4 2015-05-01         9.35
## 5 2015-03-01         9.34
## 6 2015-07-01         9.27
```

Visualisasi hasil dari daily_productivity_q1

```
daily_productivity_q1 %>%
  ggplot(aes(x=date, y=productivity)) +
  geom_line(color = "tomato3") +
  labs(title = "Daily Productivity Quarter 1",x = "Date",y = "Productivity") +
  theme_minimal()
```

Daily Productivity Quarter 1
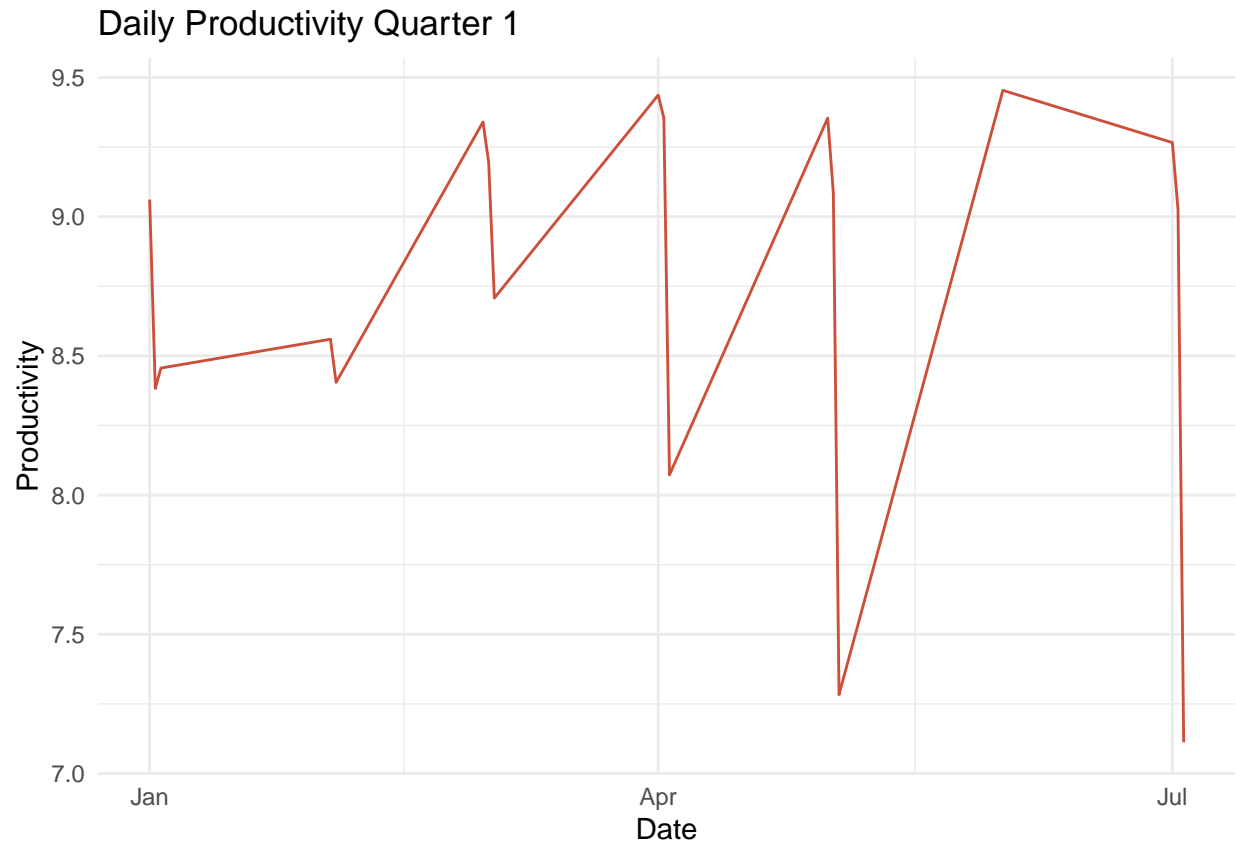
Membagi data menjadi data training dan data testing

```
splitdata <- sample.split(garment$actual_productivity, SplitRatio = 0.7)
trainingset <- subset(garment, splitdata == TRUE)
testingset <- subset(garment, splitdata == FALSE)
```

```
dim(trainingset)
```

```
## [1] 483  15
```

```
dim(testingset)
```

```
## [1] 208  15
```

Model Regresi Sederhana

Liner regresi sederhana untuk mengukur pengaruh variabel `incentive` terhadap `actual_productivity`

```
linermodel.s <- lm(actual_productivity ~ incentive,
                  data = trainingset)
summary(linermodel.s)
```

```
##
```

```
## Call:
## lm(formula = actual_productivity ~ incentive, data = trainingset)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32722 -0.03125  0.00190  0.05525  0.28747
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.513511   0.007954   64.56   <2e-16 ***
## incentive   0.004627   0.000153   30.25   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0911 on 481 degrees of freedom
## Multiple R-squared:  0.6554, Adjusted R-squared:  0.6547
## F-statistic: 914.9 on 1 and 481 DF,  p-value: < 2.2e-16
```

Kalau dilihat dari model diatas bisa dijelaskan bahwa insentive punya hubungan signifikan terhadap produktivity aktual, artinya bisa disimpulkan bahwa besaran insentive sangat berpengaruh pada peningkatan produktivitas aktual. Persamaan regresi dari model diatas adalah $actual_productity = 0.5182 + 0.0045 * incentive$

## Prediksi Model Regresi dengan Data Testing

```
yprediksi.s <- predict(linermodel.s, newdata = testingset)
summary(yprediksi.s)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.5135  0.6523  0.7333  0.7205  0.7911  1.0641
```

Visualisasi Hasil Data Training

```
trainingset %>%
  ggplot ()+
  geom_point(aes(x = incentive, y = actual_productivity)) +
  geom_line(aes(x = incentive,
                y = predict(linermodel.s, newdata = trainingset)), colour = 'blue') +
  labs(title = "Pengaruh Insentive terhadap Produktivitas Aktual - Data Training",
       x = "Insentif",
       y = "Produktivitas Aktual" ) +
  theme_minimal()
```

Pengaruh Insentive terhadap Produktivitas Aktual – Data Training
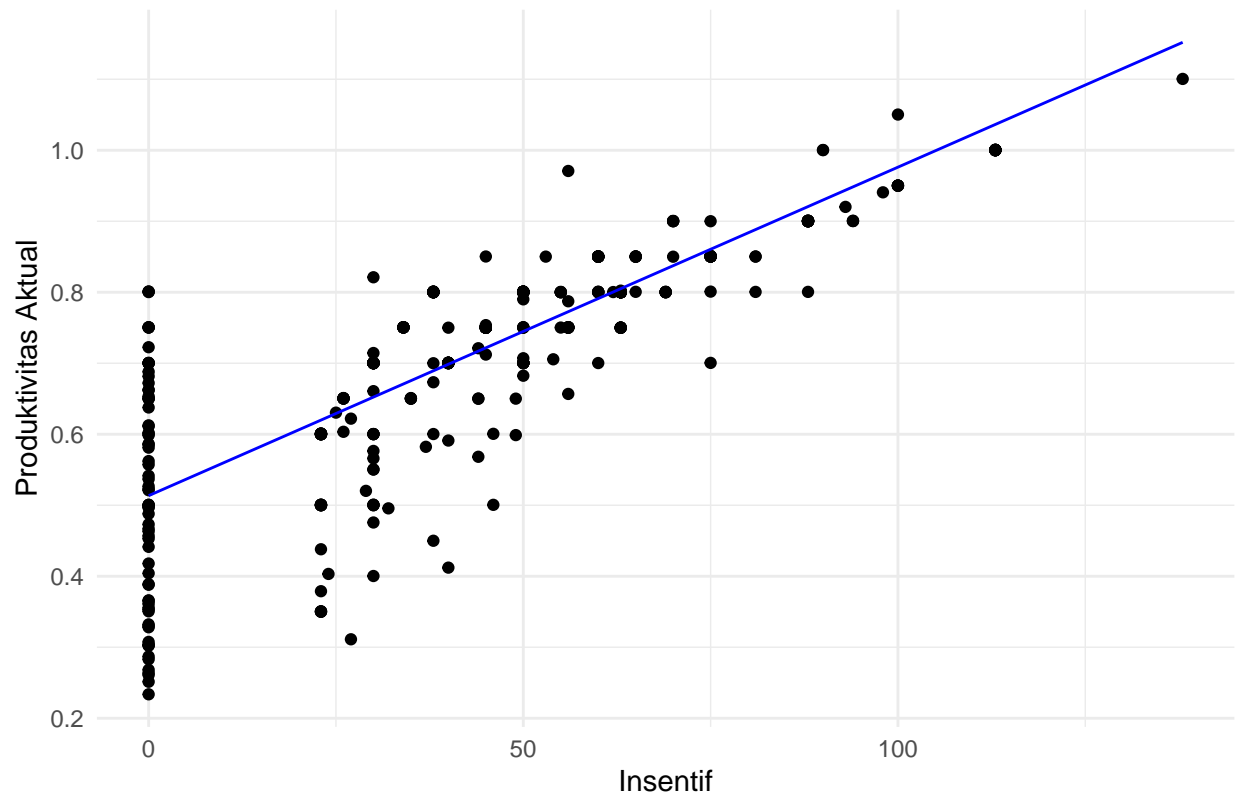
Visualisasi Hasil Data Testing

```
testingset %>%
  ggplot ()+
  geom_point(aes(x = incentive, y = actual_productivity)) +
  geom_line(aes(x = incentive,
                y = predict(linermodel.s, newdata = testingset)), colour = 'blue') +
  labs(title = "Pengaruh Insentive terhadap Produktivitas Aktual - Data Testing",
       x = "Insentif",
       y = "Produktivitas Aktual" ) +
  theme_minimal()
```

Pengaruh Insentive terhadap Produktivitas Aktual – Data Testing
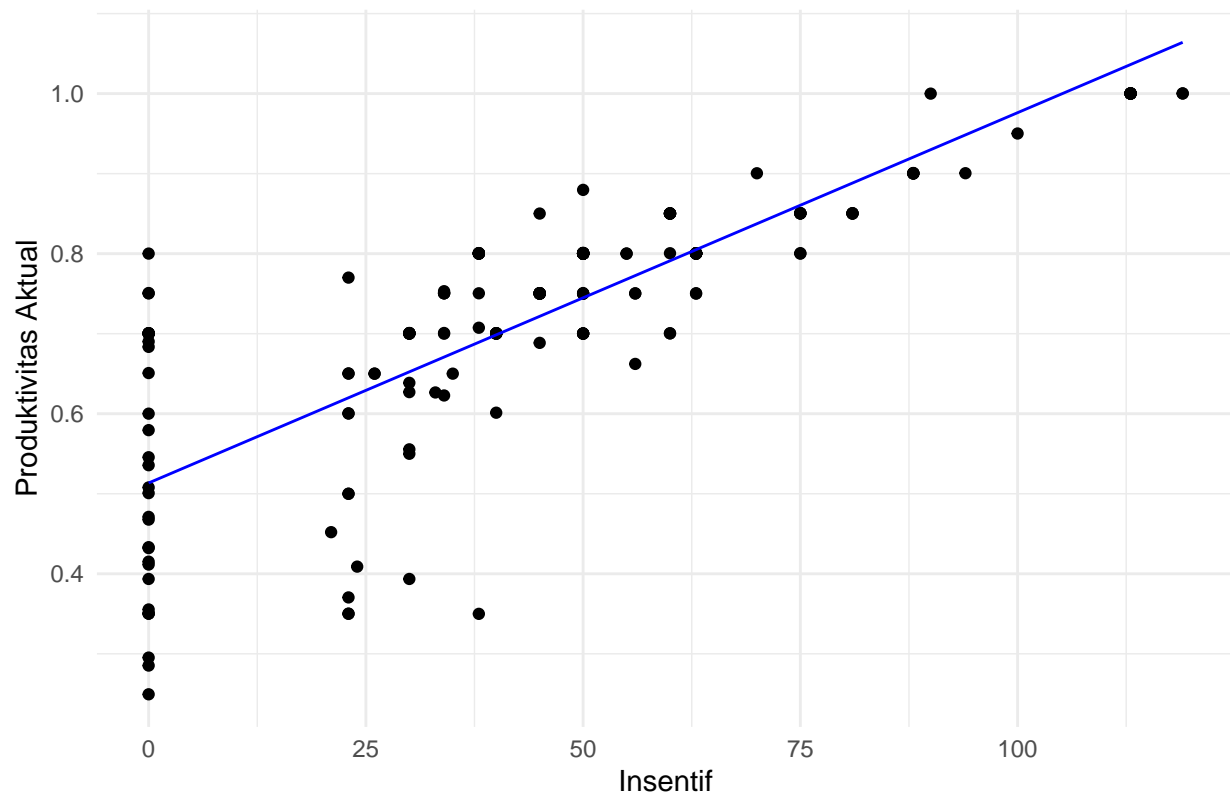
Berdasarkan visualisasi diatas bisa dilihat bahwa model regresi liner yang dihasilkan dapat memprediksi data testing

Model Regresi Berganda

Liner regresi berganda untuk mengukur pengaruh variabel `over_time`, `incentive`, `idle_time`, `idle_men` terhadap `actual_productivity`

```
linermodel.b <- lm(actual_productivity ~ idle_men + idle_time + incentive + smv + no_of_workers, data =
summary(linermodel.b)
```

```
##
## Call:
## lm(formula = actual_productivity ~ idle_men + idle_time + incentive +
##     smv + no_of_workers, data = garment)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.36342 -0.02946  0.00646  0.04988  0.28318
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.5779614  0.0198200  29.161  < 2e-16 ***
## idle_men    -0.0045694  0.0009773  -4.676 3.53e-06 ***
## idle_time    0.0004105  0.0002468   1.664  0.09667 .
```

10

```
## incentive       0.0043655  0.0001289  33.874  < 2e-16 ***
## smv            -0.0016363  0.0006098  -2.683  0.00747 **
## no_of_workers -0.0001850  0.0004522  -0.409  0.68267
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08992 on 685 degrees of freedom
## Multiple R-squared:  0.665,  Adjusted R-squared:  0.6625
## F-statistic: 271.9 on 5 and 685 DF,  p-value: < 2.2e-16
```

Berdasarkan hasil diatas bisa dilihat bahwa variabel bebas (idle_men, incentive, smv) berhubungan secara signifikan dengan actual_productivity. Sementara Variabel idle_time dan no_of_workers tidak berpengaruh pada actual_productivity. Persaman regresi liner dari model diatas adalah sebagai berikut: $actual\_productivity = 0{,}577 - 0{,}0045 idle\_men + 0{,}0043 incentive - 0.0016*smv$

Berdasarkan persamaan diatas cobalah hitung nilai actual_productivity, kalau diketahui: smv = 26.82 idle_men = 2 incentive = 26

```
actual_productivity = 0.577 - 0.0045*2 + 0.0043*26 - 0.0016*26.82
actual_productivity
```

```
## [1] 0.636888
```

Confident Interval

```
confint(linermodel.b)
```

```
##                      2.5 %         97.5 %
## (Intercept)    5.390462e-01   0.6168765170
## idle_men      -6.488197e-03  -0.0026505614
## idle_time     -7.401176e-05   0.0008950206
## incentive      4.112448e-03   0.0046185276
## smv           -2.833686e-03  -0.0004389081
## no_of_workers -1.072828e-03   0.0007029264
```

Akurasi Model

```
sigma(linermodel.b)/mean(clean_garment$actual_productivity)
```

```
## [1] 0.1245397
```

Jadi bisa dilihat bahwa tingkat kesalahan (error rate) dari model yang sudah dihasilkan adalah sebesar 12 % dengan akurasi sebesar 88%