# Test 2

Izzi Grasso 4/17/2020

### Load libraries

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.2

library(tidyverse)

## ── Attaching packages ─────────────────── tidyverse 1.2.1 ──

## ✓ tibble  2.1.3     ✓ purrr   0.3.2
## ✓ tidyr   0.8.1     ✓ dplyr   0.7.5
## ✓ readr   1.1.1     ✓ stringr 1.4.0
## ✓ tibble  2.1.3     ✓ forcats 0.4.0

## Warning: package 'tibble' was built under R version 3.5.2

## Warning: package 'purrr' was built under R version 3.5.2

## Warning: package 'stringr' was built under R version 3.5.2

## Warning: package 'forcats' was built under R version 3.5.2

## ── Conflicts ────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()

library(PMCMR)

## PMCMR is superseded by PMCMRplus and will be no longer maintained. You may
wish to install PMCMRplus instead.
```

## Problem 1

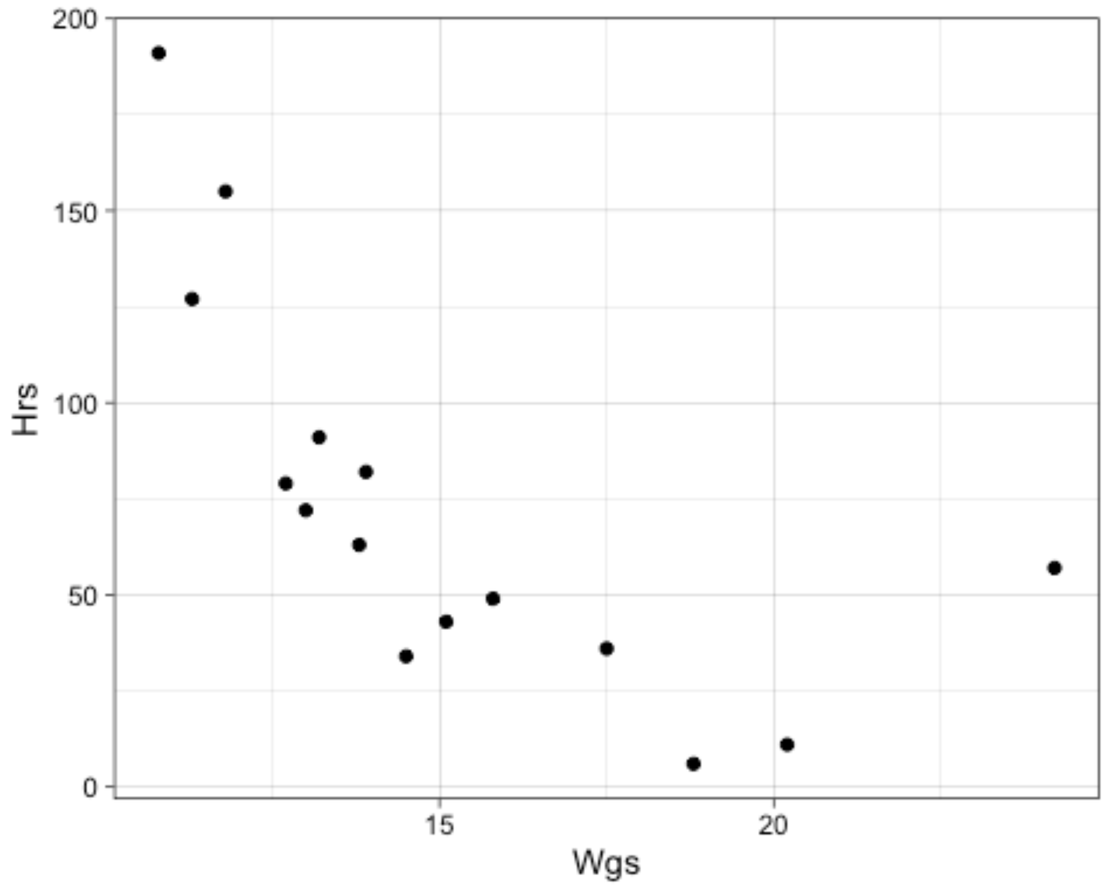
### Data

```
HoursMissed <- c(49, 36, 127, 91, 72, 34, 155, 11, 191, 6, 63, 79, 43, 57,
82)
AnnualWages <- c(15.8, 17.5, 11.3, 13.2, 13.0, 14.5, 11.8, 20.2, 10.8, 18.8,
13.8, 12.7, 15.1, 24.2, 13.9)

df <- data.frame(
  Hrs = HoursMissed,
  Wgs = AnnualWages
)
```

### Visualization

```r
ggplot(df, aes(x = Wgs, y = Hrs)) +
  geom_point() +
  theme_linedraw()
```

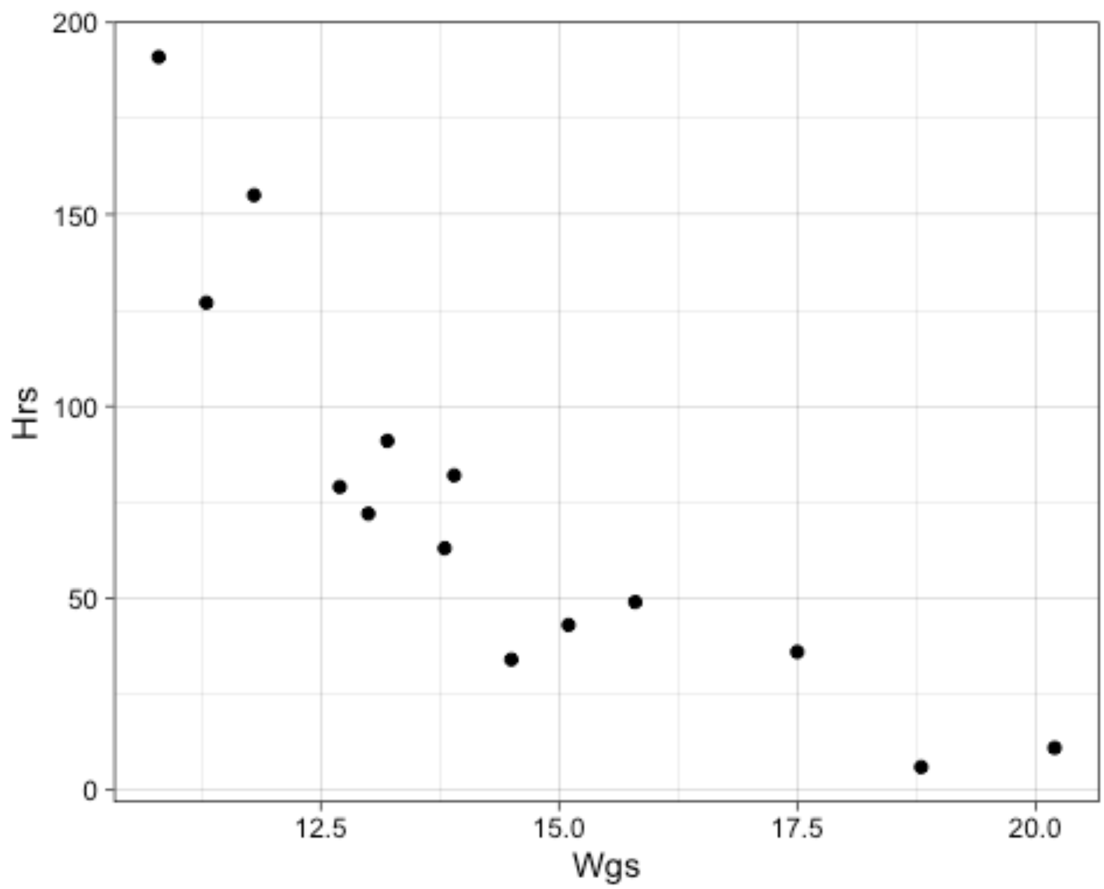



### Delete Possible Outlier

```r
df2 <- df[-14,]
```

### Re-plot

```r
ggplot(df2, aes(x = Wgs, y = Hrs)) +
  geom_point() +
  theme_linedraw()
```




### Spearman's rank correlation (df w/ possible outlier)

Null hypothesis: $\rho = 0$ Alternative hyp: $\rho \neq 0$

```
cor.test(df$Wgs, df$Hrs, method = "spearman")

##
##  Spearman's rank correlation rho
##
## data:  df$Wgs and df$Hrs
## S = 1038, p-value = 2.39e-05
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.8535714
```

### Spearman's rank correlation (df w/o possible outlier)

Null hypothesis: $\rho = 0$ Alternative hyp: $\rho \neq 0$

```
cor.test(df2$Wgs, df2$Hrs, method = "spearman")

##
##  Spearman's rank correlation rho
```

```
##
## data:  df2$Wgs and df2$Hrs
## S = 870, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##        rho
## -0.9120879
```

At $\alpha = 0.05$, for both correlation analyses we reject the null hypothesis and conclude there is a strong, negative correlation between annual wages and hours missed.

# Problem 2

### Data

```
Day1 <- c( 5.0, 4.8, 5.1, 5.1, 4.8, 5.1, 4.8, 4.8, 5.0, 5.2, 4.9,4.9, 5.0 )
Day2 <- c( 5.8, 4.7, 4.7, 4.9,5.1, 4.9, 5.4, 5.3, 5.3, 4.8,5.7, 5.1, 5.7 )
```

### Variance test

Null hypothesis: $\sigma_1^2 = \sigma_2^2$ Alternative hyp: $\sigma_1^2 < \sigma_2^2$

```
var.test(x = Day1, y = Day2, ratio = 1,
         alternative = "less", conf.level = 0.99)

##
##  F test to compare two variances
##
## data:  Day1 and Day2
## F = 0.12987, num df = 12, denom df = 12, p-value = 0.0006359
## alternative hypothesis: true ratio of variances is less than 1
## 99 percent confidence interval:
##  0.0000000 0.5396439
## sample estimates:
## ratio of variances
##          0.1298701
```

At $\alpha = 0.01$, we reject the null hypothesis. There is significant evidence that the variability of the process is greater on the second day than on the first.

# Problem 3

### Data

Null Hypothesis: $s^2 = \sigma^2$ Alternative hyp: $s^2 \neq \sigma^2$

```
SSquared = 0.0002
SigmaSquared = 0.0003
alpha = 0.05
n = 10

TestStatistic <- ((n-1)*SSquared)/SigmaSquared
```

```r
ChiSq <- qchisq((1-alpha), df = (n-1))

TestStatistic > ChiSq

## [1] FALSE
```

At $\alpha = 0.05$, we fail to reject the null hypothesis. There is not significant evidence to refute the company's claim.

## Problem 4

### Data

```r
VaccineMatrix <- matrix(c(24, 9, 13, 289, 100, 565),
                 ncol = 3, byrow = TRUE)
colnames(VaccineMatrix) <- c("NoVaccine", "OneShot", "TwoShots")
rownames(VaccineMatrix) <- c("Flu", "NoFlu")
VaccineTable <- as.table(VaccineMatrix)
```

### Chi squared test for independence

Null hypothesis: Frequency of flu is independent of Vaccines. Alternative hypothesis: Not null hypthesis.

```r
chisq.test(VaccineTable)

##
##  Pearson's Chi-squared test
##
## data:  VaccineTable
## X-squared = 17.313, df = 2, p-value = 0.000174
```

At $\alpha = 0.05$, we reject the null hypothesis. The data presents significant evidence to indicate the number of flu cases is not independent of vaccine. Post hoc analysis is required to determine whether or not the vaccine is successful and which vaccination process is best.

## Problem 5

### Data

```r
df3 <- data.frame(
  Value = c(8001, 7910, 8111, 7802, 7500,
            8025, 7932, 8101, 7820, 7601,
            8100, 7900, 8201, 7904, 7702,
            8055, 7990, 8175, 7850, 7633,
            7991, 7892, 8102, 7819, 7600,
            8007, 7922, 8235, 8100, 7561),
  MachineNum = rep(1:5, times = 6),
  EnvNum = rep(1:6, times = 1, each = 5)
)
```

**Since this is a non-parametric exam, we can assume non-normality**

### Friedman Test, RBD

Null hypothesis: Distributions are identical Alternative hypothesis: Not null hypothesis.

```
friedman.test(df3$Value, df3$EnvNum, df3$MachineNum)

##
##  Friedman rank sum test
##
## data:  df3$Value, df3$EnvNum and df3$MachineNum
## Friedman chi-squared = 12.886, df = 5, p-value = 0.02447
```

At $\alpha = 0.10$, we reject the null hypothesis. There is significant evidence that optical mark readers operate differently in different environments.