HW 7

Joseph

2025-10-14

Pre-Class Assignment 7 - Crazy Correlations Fall 25

Instructions: This assignment is related to your Chapter 5 (Salkind) material. It is not timed but assignment must be submitted by the specified due. Individual assignments that are not submitted will receive a grade of "0". This short assignment is worth 10 points.

Two Inter Miami fans are arguing about soccer. One fan is upset with the coach and staff because she thinks that the older players are being played for less time than the younger players, and that the younger players are given more time on the field. She feels this is unfair to the older players.

They decide to settle the argument with statistics. They want to know if there is a correlation between **age and total minutes played for the 2024 soccer season.** They are testing whether older players played for less time than the younger players.

Player	Age (x)	Minutes Played (for the whole season; y)
L. Messi	37	1485
L. Suárez	37	1920
L. Campana	24	1275
R. Taylor	30	1660
J. Alba	35	2231
B. Cremaschi	19	1139
M. Rojas	29	804
D. Gómez	21	1359
I. Fray	22	653
L. Frugis Afonso	23	240
F. Redondo	21	1121
S. Busquets	36	2484
J. Gressel	31	2366
S. Kryvtsov	33	1388
D. Ruiz	20	1075
N. Allen	26	918
T. Avilés	20	2207
S. Borgelin	23	133
Y. Bright	23	1403
D. Callender	27	2880
N. Freire	30	740
D. Martínez	26	282
F. Negri	29	791
R. Sailor	26	310
L. Sunderland	23	192
M. Weigandt	25	1997

```
data <- read_tsv("
  Player    Age Minutes Played
  L. Messi  37  1485
  L. Suárez 37  1920
  L. Campana    24  1275
  R. Taylor 30  1660
  J. Alba   35  2231
  B. Cremaschi  19  1139
  M. Rojas  29  804
  D. Gómez  21  1359
  I. Fray   22  653
  L. Frugis Afonso  23  240
  F. Redondo    21  1121
  S. Busquets   36  2484
  J. Gressel    31  2366
  S. Kryvtsov   33  1388
  D. Ruiz   20  1075
  N. Allen  26  918
  T. Avilés 20  2207
  S. Borgelin   23  133
  Y. Bright 23  1403
  D. Callender  27  2880
  N. Freire 30  740
  D. Martínez   26  282
  F. Negri  29  791
  R. Sailor 26  310
  L. Sunderland 23  192
  M. Weigandt   25  1997
  "
)
```

```
## Rows: 26 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (1): Player
## dbl (2): Age, Minutes Played
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
data %>% head()
```

```
## # A tibble: 6 x 3
##   Player         Age `Minutes Played`
##   <chr>        <dbl>            <dbl>
## 1 L. Messi        37             1485
## 2 L. Suárez       37             1920
## 3 L. Campana      24             1275
## 4 R. Taylor       30             1660
## 5 J. Alba         35             2231
## 6 B. Cremaschi    19             1139
```

Problem 1

$$r_{xy} = \frac{n\sum XY - \sum x \sum y}{\sqrt{\left[n\sum X^2 - (\sum X)^2\right]\left[n\sum Y^2 - (\sum Y)^2\right]}}$$

```
nrow(data) # 26
```

```
## [1] 26
```

```
sum(data$Age * data$`Minutes Played`)        # 927,435
```

```
## [1] 927435
```

```
sum(data$Age) * sum(data$`Minutes Played`)   # 23,004,888
```

```
## [1] 23004888
```

```
sum(data$Age ^ 2)                            # 19,392
```

```
## [1] 19392
```

```
sum(data$Age) ^ 2                            # 484,416
```

```
## [1] 484416
```

```
sum(data$`Minutes Played` ^ 2)               # 56,999,709
```

```
## [1] 56999709
```

```
sum(data$`Minutes Played`) ^ 2               # 1,092,500,809
```

```
## [1] 1092500809
```

$n = 26$

$\sum XY = 927,435$

$\sum x \sum y = 23,004,888$

$\sum x^2 = 19,392$

$(\sum x)^2 = 484,416$

$\sum y^2 = 56,999,709$

$(\sum y)^2 = 1,092,500,809$

```
(26 * 927435 - 23004888) / sqrt((26 * 19392 - 484416) * (26 * 56999709 - 1092500809))
```

```
## [1] 0.3993805
```

```
cor(data$Age, data$`Minutes Played`) # to double check
```

```
## [1] 0.3993805
```