DS102-08-08

Lesson 8 Practice Hands-On

Heather Walker

2022-08-17

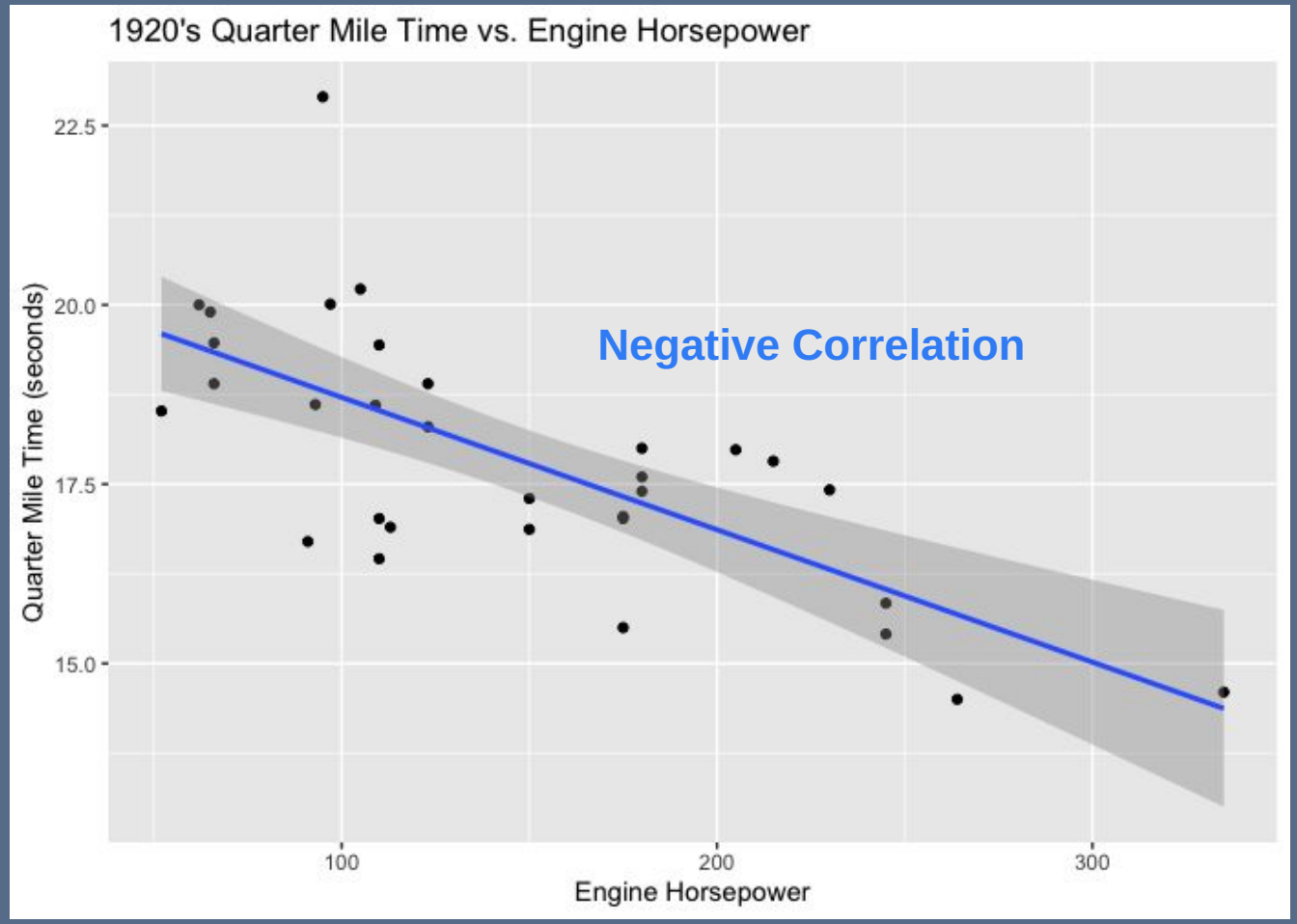
1.

Relationship between quarter mile time and horsepower for dataset `mtcars`.

SCATTER PLOT

```
# Plot relationship
between hp and qsec
d <- ggplot(mtcars, aes(x
= hp, y = qsec))

# Add title and improve
axis labels:
d + geom_point() +
geom_smooth(method=lm,
se=TRUE)
+
ggtitle("1920's Quarter
Mile Time vs. Engine
Horsepower")
+
xlab("Engine Horsepower")
+ ylab("Quarter Mile Time
(seconds)")
```



LINEAR REGRESSION

```
# Compute linear regression with function `lm()`
lin_reg <- lm(hp ~ qsec, mtcars)
print(lin_reg)

# To see all the information stored in linear
regression object:
summary(lin_reg)

# this is the equation it's calculating:
# y = mx + b
# y = -27.17x + 631.70
```

R-SQUARED VALUE

```
Adjusted R-squared:  0.485
```

Is this what you would expect?

- The Adjusted R-squared value explains 48.5% of the variability of the data
- The remaining (51.5%) variability of the data is from other variables
- This makes sense -- while the power to get a car from a standing start to racing a quarter mile is based on engine horsepower, the weight of the car is a huge factor in how much power is needed to fight the inertia of a standing start.

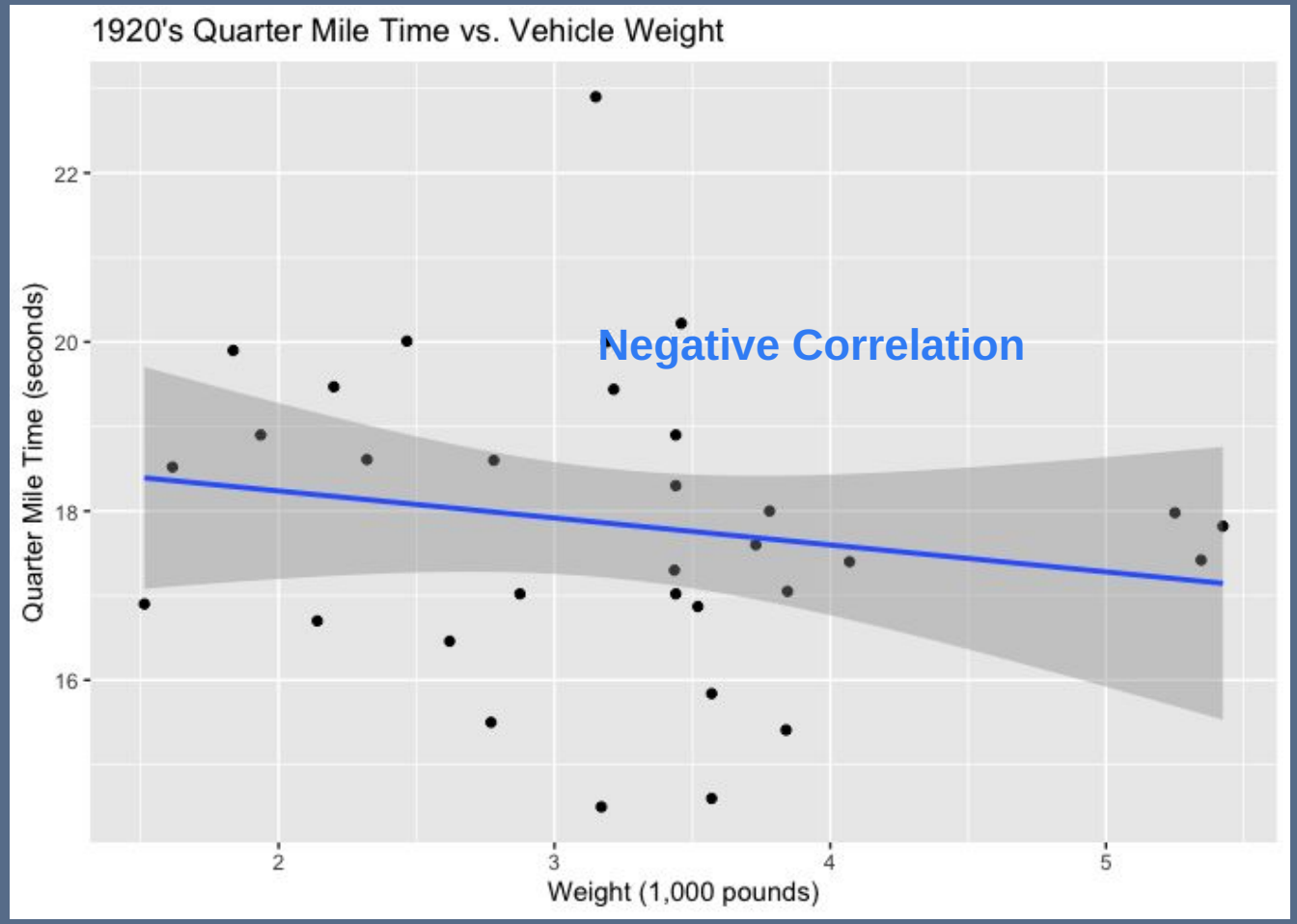
2.

Relationship between quarter mile time and vehicle weight for dataset `mtcars`.

SCATTER PLOT

```
# Plot relationship
between hp and qsec
d <- ggplot(mtcars, aes(x
= wt, y = qsec))

# Add title and improve
axis labels:
d + geom_point() +
geom_smooth(method=lm,
se=TRUE)
+
ggtitle("1920's Quarter
Mile Time vs. Vehicle
Weight")
+
xlab("Weight (1,000
pounds)")+ ylab("Quarter
Mile Time (seconds)")
```



LINEAR REGRESSION

```
# Compute linear regression with function `lm()`
lin_reg <- lm(wt ~ qsec, mtcars)
print(lin_reg)

# To see all the information stored in linear
regression object:
summary(lin_reg)

# this is the equation it's calculating:
# y = mx + b
# y = -0.09567x + 4.92479
```

R-SQUARED VALUE

```
Adjusted R-squared:  -0.00179
```

Is this what you would expect?

- The Adjusted R-squared value is -0.00179
- The negative R-squared value threw me -- it was not what I expected.
- But I consulted the archives (~Google~) and there were some explanations:
  R^2 compares the fit of the chosen model with that of a horizontal straight line (the null hypothesis). If the chosen model fits worse than a horizontal line, then R^2 is negative...
  Bottom Line: a negative R^2 is not a mathematical impossibility or the sign of a computer bug. It simply means that the chosen model (with its constraints) fits the data really poorly.
- **Source :** https://stats.stackexchange.com/questions/12900/when-is-r-squared-negative
- This then makes sense, as the data points on the scatter plot do NOT follow the line.

“

Thank you!

”