The Correlation Between the Long Jump and 100 meter dash
STAT 213 Elements Prob and Math

Cole Goodman, Ismael Kabore, Linh Nguyen, Taha Taher

Abstract:

Long jump is one of the important categories in athletics, athletes of this long jump and run are required to have a good foundation of health, strong muscles, great endurance, and the ability to respond and be flexible. A long jump is considered successful, the athlete must run to gain momentum, stop at the right jump line, and perform the jump with the correct technique, with the distance as far as possible. With the research on distance running being relatively new and unexplored, in this project, we aim to explore the relationship between the 100 meter long run and the 100 meter long jump to further research the relationship between the two sets. If we can predict their long jump based on how fast they can run. After receiving a sample collected from various high schools and colleges across the US in 2022 and summarizing the results for the long jump and running using univariate and two-variable summary statistics with Studio, a two-sample t-test showed a p-value is greater than the significance level and, therefore, there is sufficient evidence to suggest that there was a correlation between the 100 meter dash and long jump.

Background:

Track and field, also known as athletics, is a sport that involves a variety of running, jumping, and throwing events. Therefore, our area of research will focus on whether one's top speed in the 100 meter dash can predict their success in the long jump. The 100 meter dash is a track and field event that involves sprinting a distance of 100 meters (about 109 yards) as quickly as possible. The long jump is a field event in which athletes compete to jump the greatest distance from a take-off point. In the long jump, athletes run down a runway and then take off form a marked line or board, trying to land as far as possible in a sand pit. Both the 100-meter dash and long jump are popular events at track and field competitions, including the Olympic Games. The 100 meter dash and long jump require different abilities and training approaches. There has been debate among coaches and researchers about whether an athlete's top speed in the 100 meter dash can predict their success in the long jump. Some research suggests that there may be a relationship between these two events, as both require explosive power and speed, but more investigation is required. According to the Academy of Sport Speed Australia, through much research, sport scientists agree the most important training factor contributing to an increase in stride length is force(pwsadmin, 2019). Because force is relied on heavily in both events, one's ability in one event may transfer over to the other event.

The relationship between the 100 meter dash speed and long jump performance is interesting because it has implications for training and talent identification. If a strong relationship does exist, it may be beneficial for coaches to focus on improving an athlete's top speed in the 100 meter dash in order to improve their jump performance. On the other hand, if the relationship is not as strong, coaches may need to adopt a more multifaceted approach to training for the long jump that focuses on other factors such as technique, strength, and power. The null hypothesis is that the slope, $\widehat{\beta_1}$, does not have a linear relation: $H_0 = \widehat{\beta_1} = 0$. The alternative hypothesis is that there is a linear relationship, or $H_a: \widehat{\beta_1} \neq 0$

Methods:

To explore the potential relationship between the 100 meter dash and long jump, we curated our own dataset. The personal bests of 100 high school and college track & field athletes who compete in both the 100 meter dash and the long jump were randomly sampled, with all of their performances coming from the Outdoor 2022 Season. The data used in this study were from the TRFFS(Track & Field Results Reporting Service) and Athletic.net. 50 of the athletes were competing at the high school level, in various states across the country. The other 50 athletes compete at the collegiate level, with 25 competing in Division 1 and the other 25 competing in Division 3. There are no potential biases with the data, as all track & field events results are recorded and posted publicly, which eliminates the potential for non-response biases. Our dataset represents a QQ situation, with the first quantitative variable being the athlete's personal record marks in Long Jump, and the second quantitative variable being the athlete's personal best marks in the 100 meter dash.

To analyze this data, we will produce 2 histograms and a linear regression plot to describe the (bivariate) data. The histograms display the frequency distributions of the personal best marks, in each respective event. The scatterplot plot produces the line of best fit, and describes the relationship between the two quantitative variables.

Results:

The total sample size of the study was 100 respondents, with all of them being male. The median 100 meter personal record was 11.09 seconds and the median long jump personal record was 7.185 meters. The best 100 meter personal record was 10.27 seconds, with the worst being 12.99. The best long jump personal record was 5.82, with the worst being 8.35.

The shape of the histograms varies. The shape of the 100 Meter PR histogram is slightly skewed to the right and unimodal, while the shape of the Long Jump PR histogram is relatively symmetrical and unimodal.

To test the hypothesis, we used a linear model to calculate the linear regression line in the scatterplot. We found that the model had an intercept of 15.08754 and a slope of -0.70519 and a standard error of 0.06953. With this, we can manually calculate the t-value to be $t_{TS} = \frac{-0.70519}{0.06953} = -10.1422408$. This can also be shown in the summary of the linear model, after the Std. Error column in Figure 1.1.

From here, we can find the p-value. Since our alternate hypothesis is to see if our slope, or $\widehat{\beta_1}$, is not equal to 0, or $\widehat{\beta_1} \neq 0$, we can use the R function $2 * (1 - pt(abs(-10.1422408), 98))$ to give us the p-value of 0. This implies that our α is less than the given significance level of 0.05, and there is sufficient evidence to reject the null hypothesis and conclude that there is a linear relationship between someone's personal record in the 100 meter dash and their personal record in the Long Jump.

All of the relevant code used to obtain these results can be found in the Appendix.

```
> summary(model)

Call:
lm(formula = df$Long_Jump_PR ~ df$X100_M_PR)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.79823 -0.26421  0.00479  0.20741  0.77184 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  15.08754    0.77536   19.46   <2e-16 ***
df$X100_M_PR -0.70519    0.06953  -10.14   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3441 on 98 degrees of freedom
Multiple R-squared:  0.5121,	Adjusted R-squared:  0.5071 
F-statistic: 102.9 on 1 and 98 DF,  p-value: < 2.2e-16
```

Figure 1.1: a summary of the generated linear model

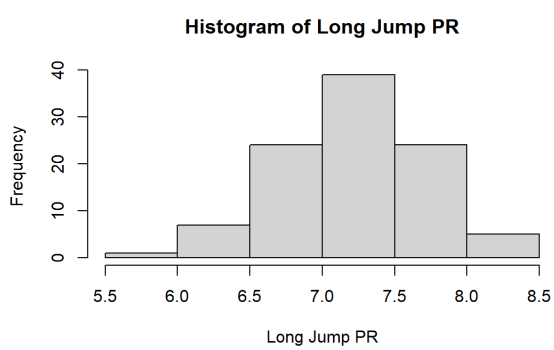

Figure 1.2: The histogram displaying the frequency distributions of the Long Jump PR

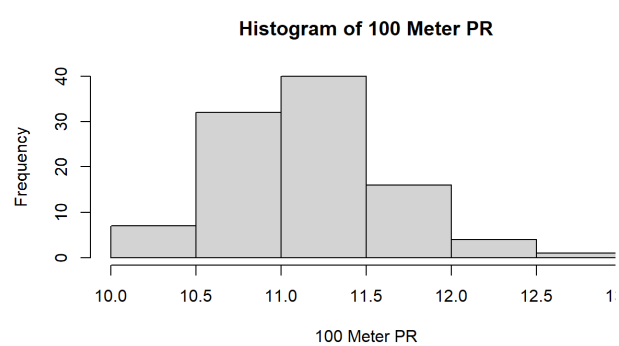

Figure 1.2 and 1.3: The histogram displaying the frequency distributions of the 100 meter PR

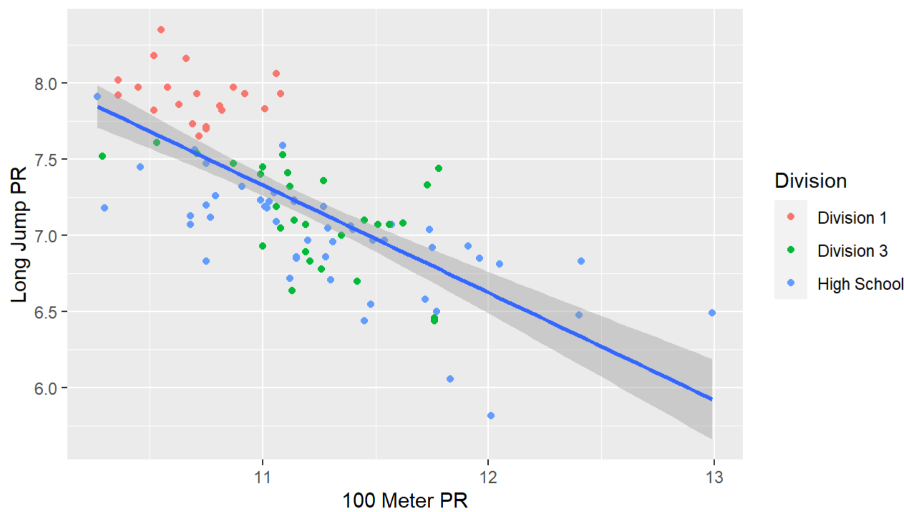

Figure 1.4: Scatterplot of the relationship between the PR in the 100 Meter Dash, and the PR in the Long Jump. Linear regression line is shown in blue, and the standard error throughout the line is shaded in gray.

Discussion:

In short, the objective of this study was to determine if there was a correlation between the 100 meter dash and long jump. For example, a study published in the journal of Sports and Medicine found that top speed was significantly correlated with long jump performance in elite male sprinters (Zanone et al. 2012). The null hypothesis, therefore, is that the same in the meantime on the dash and long jump is equal to 0. The alternative hypothesis is that the difference in the meantime on the dash and long jump is not equal to 0. We determined that, assuming equal variances, the test-statistic is -10.1422408 and the p-value is 0. Since 0 is larger than the significance level of 0.05, we fail to reject the null hypothesis. There is sufficient evidence to suggest that there was correlation between the 100 meter dash and long jump. All our participants were male so it might be different for females. The result might not translate to women and more research is needed to conclude.

References:

Pwsadmin. “Components of Success in the 100m Sprint.” *Academy of Sport Speed Australia*,26 July 2019,https://academyofsportspeed.com/2019/07/components-of-success-in-the-100m-sprint-track/.

Appendix:

```
df <- read.csv('long_jump_data.csv')
View(df)
data_plot <- plot(df$X100_M_PR, df$Long_Jump_PR, xlab = '100 Meter PR', ylab='Long Jump PR')
model <- lm(df$Long_Jump_PR~df$X100_M_PR)
abline(model)
summary(model)

PR100M <- df$X100_M_PR
PRLJ <- df$Long_Jump_PR
Division <- df$Division

predict(model, new_data=my_distances, interval='confidence')
library(ggplot2)
ggplot(data_plot, aes(x=PR100M, y=PRLJ))+
        geom_point(aes(col=Division))+
        geom_smooth(method='lm')+
        xlab("100 Meter PR")+
        ylab("Long Jump PR")
model
hist(PR100M, main="Histogram of 100 Meter PR", xlab='100 Meter PR')
hist(PRLJ, main="Histogram of Long Jump PR", xlab='Long Jump PR')
summary(PR100M)
median(PR100M)
hist(PRLJ)
2*(1-pt(abs(-10.1422408), 98))
```