[Race walking](https://en.wikipedia.org/wiki/Racewalking) is a competitive individual sport with a similar layout to cross country running, except instead of running, competitors walk. Athletes utilize a swift walking motion which involves a lot of movement of the hips and arms, enabling them to move faster. Just like in running the first participant to cross the finish line wins. According to [Olympic rules](https://olympics.com/en/news/all-you-need-to-know-about-race-walking). The main rule in racewalking is that racewalkers **must** maintain contact with the ground at **all times**, unlike in running when the stride often involves both feet being off the ground. Additionally, racewalkers cannot bend their leading knee when they are walking. Judges are positioned throughout courses to ensure these rules are withheld. If an athlete violates either of the rules, judges show them a "warning paddle". Three warning paddles result in a red paddle which means the athlete is disqualified from the event. These rules provide an added challenge to the sport. The World Athletics Race Walking Team Championships are an international race-walking event hosted annually in different cities. After events are completed, total medals are tallied for each country competing, which is why they are considered the "team" championships. In addition to winning medals for 1st, 2nd, and 3rd place respectively, athletes can also set new records for both themselves as well as others. The [2024 championship](%5D(https:/worldathletics.org/competitions/world-athletics-race-walking-team-championshi/world-athletics-race-walking-team-championshi-7199325) was hosted in Antalya, Turkey and consisted of four events, a women's 10 and 20k, and a men's 10 and 20k. Analyzing podium finishes, records, and speeds in long (20k) and shorter (10k) distance races can provide insight on patterns in either event, such as which countries dominate. Additionally, comparing the speeds between the distances can show if there is a difference in pacing approaches in a long race versus a shorter race. Assessing the number of records can further discern whether or not certain approaches are more successful than others.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| POS | The position the athlete finished in. |
| COUNTRY | The country the athlete is from. |
| ATHLETE | The name of the athlete in `first name, last name` order. |
| RECORD | The record the athlete made if any. The record types are as follows: Personal Best (`PB`), Season's Best (`SB`), National Record (`NR`), Area Under 20 Record (`AU20R`), National Under 20 Record (`NU20R`) |
| DISTANCE | The distance of the race the athlete competed in, either `10000`m or `20000`m. |
| GENDER | The gender of the athlete, either `Man` or `Woman`. |
| TIME | The total time it took the athlete to complete the race, measured in minutes. |
| SPEED | The average speed of the athlete for the race, measured in meters/minute. |
| REC | If the athlete made a record in the race, binary variable where `1` means they did and `0` means they did not. |
| PODIUM | If the athlete had a podium finish in the race, binary variable where `1` means they did and `0` means they did not. A podium finish means the athletes `POS` was 1st, 2nd, or 3rd. |

1. The below histogram shows SPEED faceted by DISTANCE. Compare the distribution of SPEED for the 10k races with the distribution for the 20k races

A graph of a number of bars

Description automatically generated with medium confidence

1. Perform a t.test to determine if there is a significant difference in the mean speed for each distance.

H0: Ha:

1. Based on the below boxplot would you say that there is a significant difference in the mean speed for each distance and for each gender?

A graph with blue and black squares

Description automatically generated

1. Use a two-way ANOVA test to determine if there is a significant difference in the mean speed for each distance and for each gender?

Factor A: H0: Ha:

Factor B: H0: Ha:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **D.f.** | **Sum of Squares** | **Mean Square** | **F-value** | **P-value** |
| **DISTANCE** |  |  |  |  |  |
| **GENDER** |  |  |  |  |  |
| **Residual** |  |  |  |
| **Total** |  |  |

* 1. Fit the model:
  2. A female athlete walked in the 10k with a podium finish and a speed of 221 meters per minute. Calculate the log(odds), odds, and predicted odds of her making a record.
     1. log(odds):
     2. odds:

* + 1. :

1. Perform a g-test to assess the overall fit of the multiple logistic regression model:

H0: Ha:

1. After performing all of these tests on different models, are there any noticeable differences in the different races the athletes competed in?