Tour de France Module

The Tour de France is the biggest cycling race in the world, composed of 21 stages across terrain all over France. Cyclists compete for 3 weeks to see who can secure the Yellow Jersey or Maillot Jaune and win the General Classification of the race. The General Classification (GC) is the overall competition of the Tour de France where riders compete to see who can finish all 21 stages with the least cumulative time. Riders battle through the mountains and hills of France to gain time on rivals and earn the privilege to wear the yellow jersey. Because the Tour has such prestige and difficulty, cheating has always been a factor in the race. The winner of the first Tour de France, Maurice Garin, was disqualified from the 2nd edition of the race after being found to have used trains and cars to aid him reaching the finish line. In fact, 11 riders were disqualified along with Garin in 1904, and the subsequent winner Henri Cornet also being accused of hitching a ride in cars. This was not the only form of cheating in the Tour, as there have been several different eras of performance enhancement and other changes in the sport. Testing at the tour is done after each stage, where the stage winner, stage runner-up, overall leader, and 2 random racers are all tested.

1903-1950: Amphetamines

During this period, there was no restriction on performance enhancing substances as riders would commonly use amphetamines and strychnine to aid their performance. Performance enhancement was not restricted in any way by the race organizers.

1950-1965: Transisiton

The tour began to take an anti-doping stance and began to discourage the use of performance enhancement. No testing or restrictions were used at this time.

1965-1990: Steroids

The tour began to test riders for performance enhancers in 1965 but did not hand out very severe punishments for violations. Tests for amphetamines became far more effective at this time so people began to stray away from them. The most common performance enhancers used at this time were steroids. Several Tour winners tested positive for steroids during this time but did not have results rescinded. Riders were given a time penalty if they were caught doping instead of receiving a ban. Some champions later admitted to doping during the race, but some were never caught.

1990-2010: EPO

As other drugs became detectable by testing, Erythropoietin also known as EPO, became widely used. EPO is used by people suffering from anemia to increase red blood cell production, which is beneficial to athletic performance. EPO was used as testing improved as the drug is not able to be detected, only the results of increased red blood cells. Testing during this period shifted to being more frequent and happening during training, not only the Tour itself.

2010-Present: Clean

Since Alberto Contador was caught doping in his 2010 victory, later given to Andy Schleck, no winner of the Tour has been caught doping. (Chris Froome, 4x Tour Champ was caught using salbutamol but was cleared of any doping use). The Tour is beginning to move past the days of Lance Armstrong and doping controversies, or riders are learning to hide their doping better.

Write the null and alternative hypothesis for if the distance between eras of the Tour de France.

Null: There is no significant difference in the mean distance of the Tour between the different eras of doping in the Tour de France

Alternative: There is a significant difference between mean distance of the Tour in the different eras of doping in the Tour de France.

Based on the boxplots below, do you think we reject or fail to reject the null hypothesis?

We should reject the null hypothesis because the means appear to be different between eras.



Below is output for an Anova test, do the results support your prediction in the previous question?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum Sq** | **Mean Sq** | **F-Value** | **P-Value** |
| Era | 4 | 25436546 | 6359136 | 24.11 | 5.13e-14 |
| Resuduals | 101 | 26645153 | 263813 |  |  |
| Total | 105 | 52081699 |  |  |  |

Response = Distance

We reject the null hypothesis because the p-value is less than 0.05, meaning that it is significant and at least one of the mean distances is significantly different from another mean distance.

We reject the null hypothesis because the p-value is less than 0.05, meaning that it is significant and the means are not all equal.

Answer the following question regardless of your above answer:

Pairwise t Test:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1. Amphetamines | 2. Transition | 3. Steroids | 4. EPO |
| 2. Transition | 0.0518 | ---- | ---- | ---- |
| 3. Steroids | 4.6e-08 | 0.0034 | ---- | ---- |
| 4. EPO | 5.2e-12 | 7.0e-06 | 0.0424 | ---- |
| 5. Clean | 2.1e-10 | 5.6e-06 | 0.0117 | 0.3766 |

Using the output from the pairwise t test above, determine what eras the amphetamine doping era is significantly different from.

The amphetamine doping era is significantly different from the Steroid, EPO, and Clean era, while having a p-value of just over 0.05 between Amphetamines and Transition.

Is the Steroid era distance significantly different from the EPO era distance? The Clean era?

Yes, the mean distance of the steroid era is significantly different from the EPO and Clean eras.



This plot shows the average speed of the Tour de France era over the different eras, what are some reasons that riders in the clean era are going as fast or faster than other eras.

Distance is one reason; modern tour routes are much shorter than the original tour routes. Road surface is another, paved roads are significantly quicker. Biking technology is also much more advanced now than it was in previous years. The peloton is also more cohesive now, riders stay together and draft more efficiently. Nutrition is another big advancement, where cars bring water and carbohydrates for riders throughout each stage as opposed to riders carrying their own nutrition.