STAT1378-Assignment II

Yunbae Chae

Contents

Overview

This document includes answers for all Exercises 1, 2 and 3. I have included the assign_q2.zip, but please expect the Github links in Exercise 2. I have fully prepared all the Github contents for the higher grade. Please carefully inspect every single aspect of this document because I have prepared it in a way that I might achieve the full mark. Thank you for your time for marking. Please enjoy!

Exercise 1

I am displaying the plots only.

Plot 1

How the three lap, with no shortcut world record develop over time
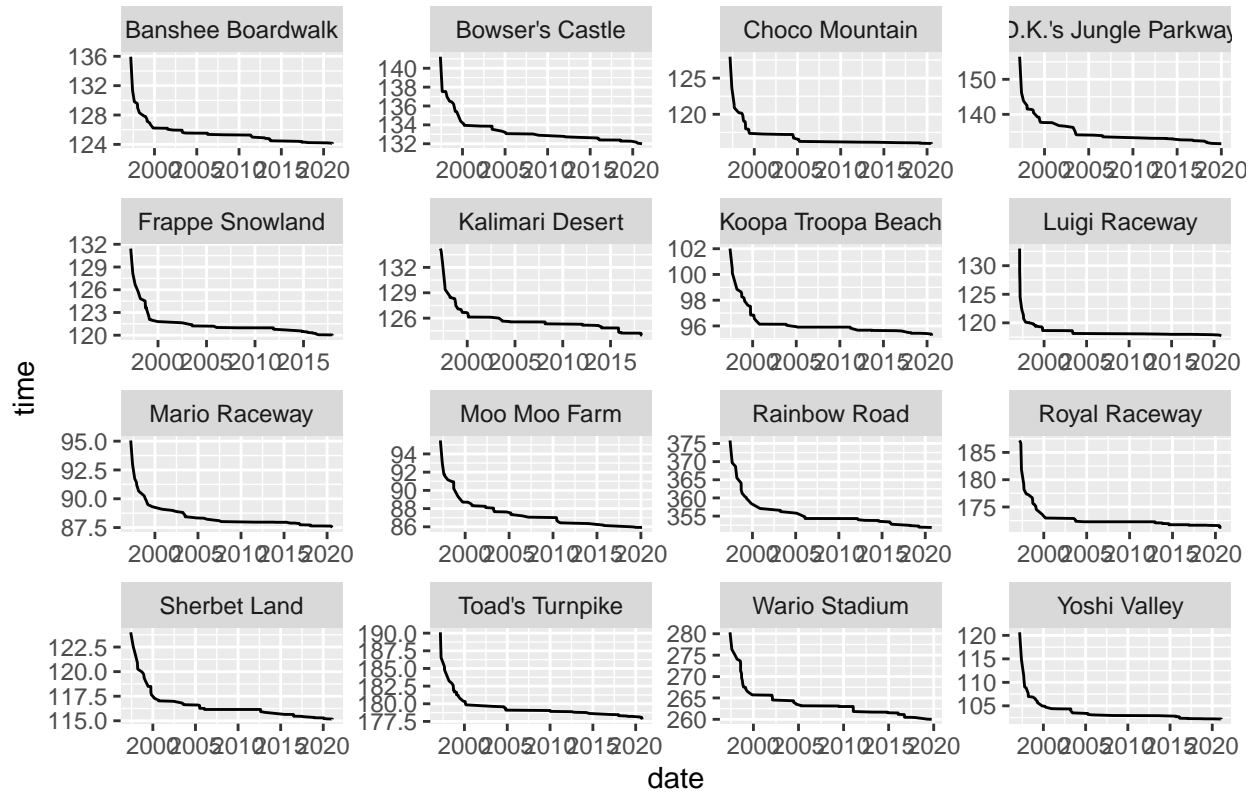
Plot 2

How the WR for Rainbow Road develop over time
With shortcuts, it is quicker to finish a 3 lap race than completing a single lap!
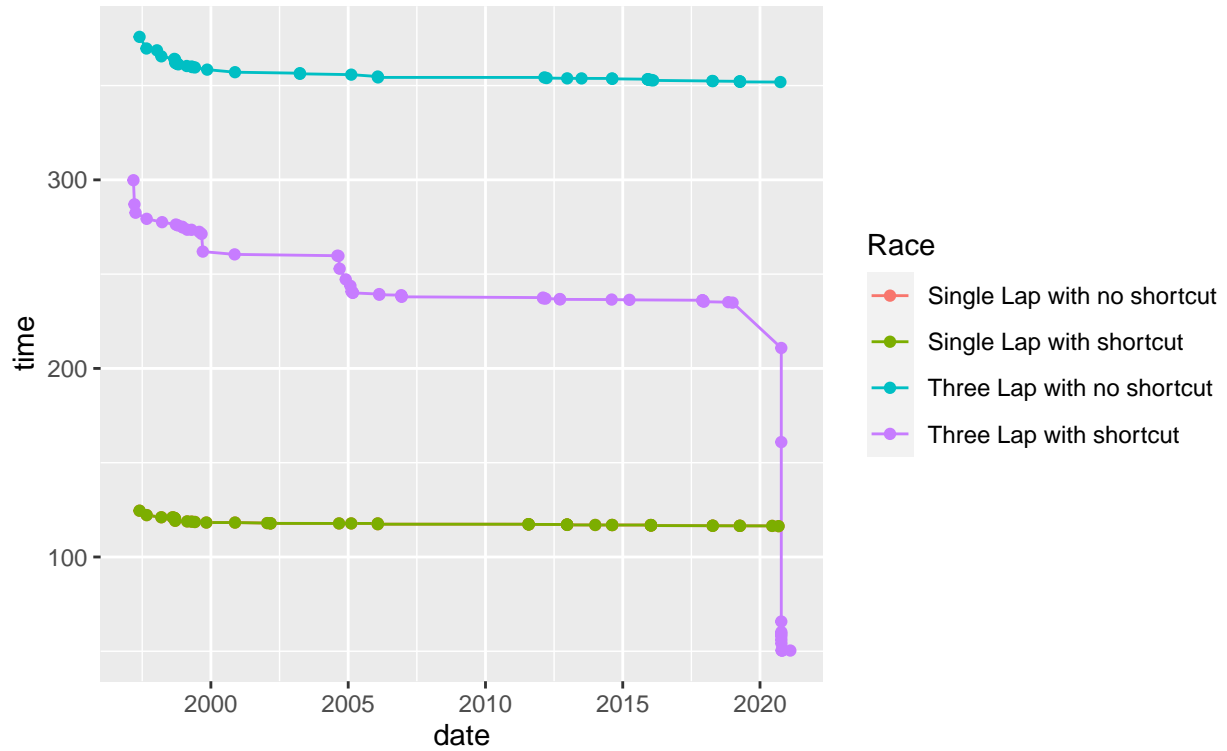
Plot 3

Comparing the Original and Current WR for three lap and no shortcut races
All lines are relatively parallel.
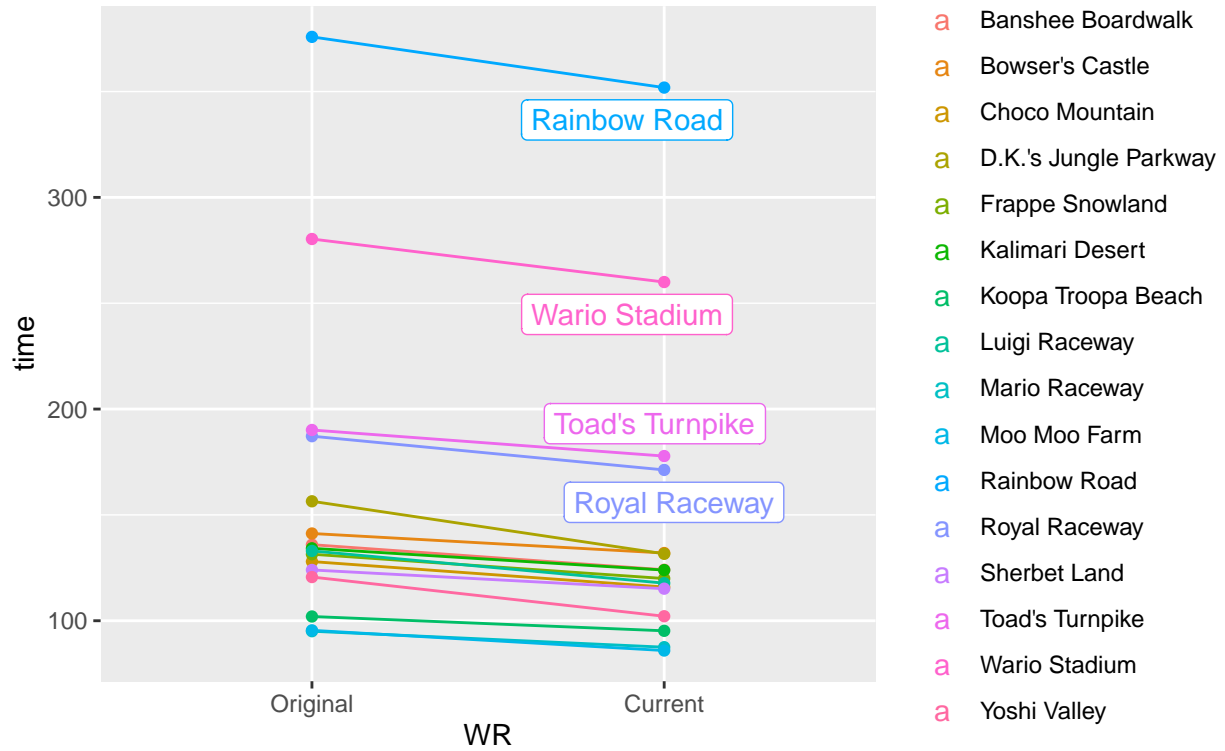
Plot 4

Time Difference between Original and Current WR for three
lap and no shortcut races

D.K.'s Jungle Parkway has the biggest reduction even it's not the longest
track.

X-axis: Time Difference between Original and Current WR (in s)
Y-axis: Track
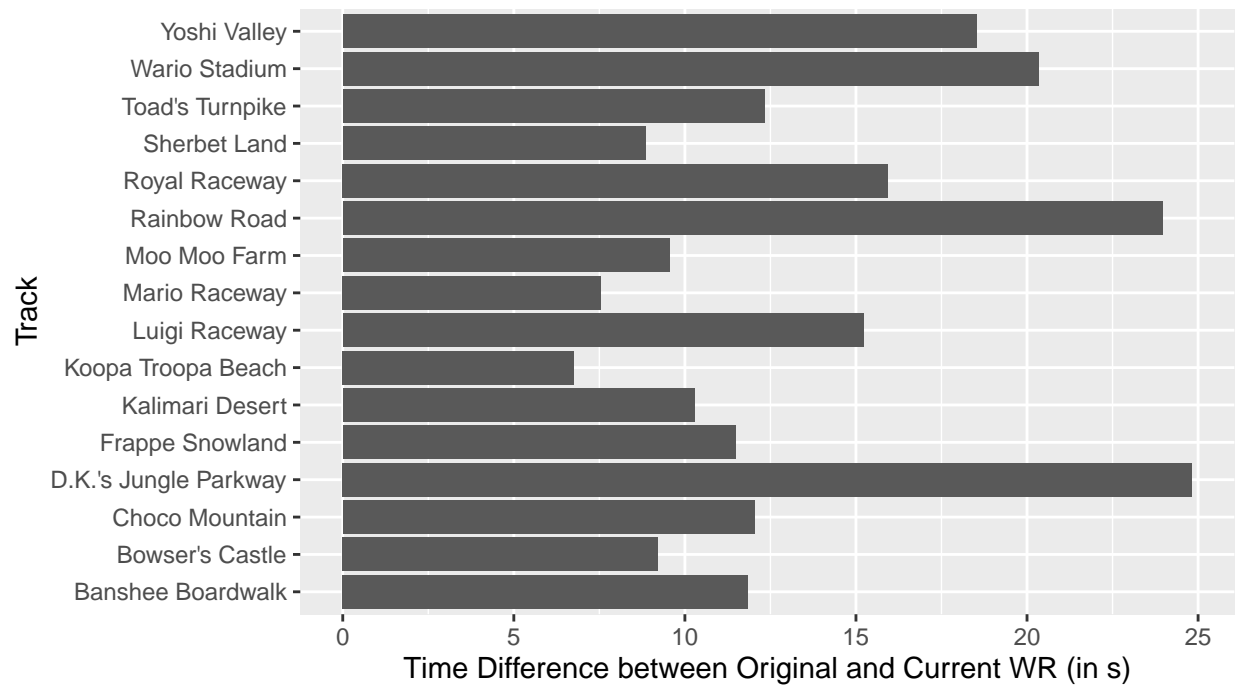
Exercise 2

Please use the following links to inspect my contents for this exercise!

- https://github.com/MQ-STAT1378/assignment2-question2-david-yunbae
- https://david-yunbae.github.io/assignment2-question2-david-yunbae

Exercise 3

As requested in the prompt, the contents in this exercise aim for high distinction. I have prepared the contents in a way so that the explanations heavily rely on examples. I have made sure to include and cite all essential statistical concepts.

Introduction: Pearson's χ^2 test (Goodness of fit)

Lining with Greenwood and Nikulin (1996), a χ^2 test is when there is a hypothesis that a data should follow a particular probability distribution.

For example, you can google the populations in China, Japan and South Korea. They are 1.4, 0.126, 0.052 billion respectively. You can sample a group of 1,000 East-Asians in Sydney. You think the number of people residing overseas is perfectly proportional to the population of their home countries. In this way, you expect them to be:

E_1 : 887 (China), E_2 : 80 (Japan), E_3 : 33 (South Korea).

All E_1 :, E_2 : and E_3 : must be greater than 5 (to be consistent with normal approximation). Say you have gained the data from the 1,000 people (they are only dummy values):

O_1 : 500 (China), O_2 : 100 (Japan), O_3 : 400 (South Korea).

Intuitively, this is a critically flawed hypothesis. In terms of statistical concepts,

The test starts with the general test statistic:

$$\tau = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i}$$

Then the p-value must be calculated:

$$P - value = P(\chi^2_{g-1} \geq \tau_{obs})$$
$$P - value = P(\chi^2_2 \geq 4255.3337775)$$

The p-value is less than 2.2×10^{-16}.

Generally, the p-value of less than 0.05 is too small. If p-value is too small, the observed ratio deviates from the expected ratios. If the p-value is greater than 0.05, then you could conclude the observed ratio lies within the expected ratios. In the population case, the p-value was less than 2.2×10^{-16}: There is not enough evidence that the population of the countries affect the proportion of overseas residents.

Now, what if you want to continue the research, and ask the 1,000 if they consider themselves to be financially stable or not? Your findings would definitely improve your knowledge about inter-cultural believes in East-Asian countries on how financial confidence leads to the choice of extending their lives overseas.

Contingency table (Test of independence)

Extending from the last example, you could make a new hypothesis that in all the three countries, the observations would show that the majority of people residing in Sydney are those that are financially confident. If this is the case, then you would see that the proportions of financially confident people would be levelly similar in all three countries. This is a 2 X 3 table test of independence.

The next example is a more general 2 by 2 table. There is a case when the "members of a sample are doubly classified (i.e., classified in two separate ways)" Stefanescu, Berger, and Hershberger (2005). Then the results form a rectangular table (contingency table) like this:

	Men	Women
Studying	3	27
Not-studying	33	9

The example above is testing whether the proportion of studying individuals is the same for both genders or not (equivalent to the following):

H_0: **the two variables are independent against each other.** H_1: **not H_0.**

This is the main discussion in this assignment. Basically, the principle of this statistical test is to compare the observations data with an arbitrary idea of the experimenter. If the observed data does not match with the forecast pattern of distribution, then the experimenter should realise there is a flaw in his projection. He/ she would investigate how to adjust the prediction or conclude that there is no relationship between the observed data and the hypothesis.

The above is an example of the contingency table. It is the case when the individuals in a sample are classified according to two separate characteristics. This means every single participants has either of the first characteristic, and also either of the second characteristic. Each individual in the sample is either male of female. Also, they are either studying or not studying. An experimenter can be curious as to whether these two characteristics have a relationship between them. In other words, if female individuals have more studiers relative to males, for instance.

In fact, the above example predicts that the proportion p of male studiers should be same as the proportion p of female studiers, because gender should not affect one's willingness to study. The χ^2 test is a statistical point of view to this problem.

The χ^2 test approach is to find the τ_{obs} first. The τ_{obs} for a two-sided test is:

$$\tau = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}, \text{ under } H_0.$$

$$P - value = P(\chi^2_1 \geq 32.9142857)$$
$$P - value = 3.8401057 \times 10^{-8}$$

Solely based on the sample data, the P-value is small: we have evidence against H_0, i.e. there is strong evidence to suggest that the proportion of studying individuals is not the same for both genders.

Fisher's exact test

The following is a similar test to the Ronald Fisher's "lady tasting tea" experiment. It was based on a 2 by 2 contingency table. In order to understand the test, you should set up the skeleton table first (using the last example):

	Men	Women	Total
Studying	A	B	A + B
Not-studying	C	D	C + D
Total	A + C	B + D	A + B + C + D (=n)

An hypergeometric distribution gives the probability of one particular distribution:

$$p = \frac{\binom{A+B}{A}\binom{C+D}{C}}{\binom{n}{A+C}} = \frac{(A+B)!(C+D)!(A+C)!(B+D)!}{A!B!C!D!n!}$$

Using this on the contingency table from the last example:

	Men	Women	Total
Studying	3	27	30
Not-studying	33	9	42
Total	36	36	72

One particular probability for this distribution, assuming H_0 (the proportion of studying individuals is the same for both male and female):

$$p = \frac{\binom{30}{3}\binom{42}{33}}{\binom{72}{36}} = \frac{30!42!36!36!}{3!27!33!9!72!}$$

$$p = 4.0910044 \times 10^{-9}$$

According to Fisher, the significance level can be calculated conditioned on the fixed marginal totals (30, 42, 36 and 36). You should account for all the arrangements that are as extreme as the observed case. Adding up all the probabilities yields the p-value.

It is apparent that the calculations involved in the Fisher's tests require some computing power. The p-value is found using an r package called "stats." For simplicity, the test should be two-sided. The p-value is 8.4177433×10^{-9}. P-value is small. H_0 is rejected. Once again, there is strong evidence to suggest that the proportion of studying individuals is not the same for both genders.

Simulation

As it was set out in the assignment prompt, I have run 10,000 simulations in four different settings. Based on the results alone, all four results had one thing in common: In every settings, Fisher's exact test's rejection rates were significantly lower than those of the χ^2 test.

Setting 1 (Small sample size, small p):

Significance Level	Prop: Fisher	Prop: Chi
0.01	0.0013	0.4604
0.05	0.0141	0.4604
0.10	0.0380	0.4604

Setting 2 (Small sample size, normal p):

Significance Level	Prop: Fisher	Prop: Chi
0.01	0.0049	0.3634
0.05	0.0211	0.3634
0.10	0.0444	0.3634

Setting 3 (Large sample size, small p):

Significance Level	Prop: Fisher	Prop: Chi
0.01	0.0065	0.2141
0.05	0.0346	0.3442
0.10	0.0733	0.3448

Setting 4 (Large sample size, normal p):

Significance Level	Prop: Fisher	Prop: Chi
0.01	0.0069	0.1678
0.05	0.0415	0.2793
0.10	0.0762	0.3839

Conclusion

A personal remark is that the statement *"Despite the fact that Fisher's test gives exact p-values, some authors have argued that is it conservative, i.e. that its actual rejection rate is below the nominal significance level."* is credible.

The simulation size was 10,000, so I generated 40,000 contingency tables in total. This seems reasonable, as 40,000 should be a large enough size of tests. The settings were set out to cover all possible test environments. Most importantly, χ^2 test results were at least 5 times and up to 500 times greater than the Fisher's tests'.

Hence, based on the simulation results, it seems reasonable to conclude that the Fisher's exact test is conservative. It has a low rejection rate, below the nominal significance level.

References

Greenwood, P. E., and M. S. Nikulin. 1996. *A Guide to Chi-Squared Testing.* Wiley Series in Probability and Statistics. Wiley. https://books.google.com.au/books?id=bc8zfQSKOwIC.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Stefanescu, Catalina, Vance W Berger, and Scott Hershberger. 2005. "Yates's Continuity Correction." *Encyclopaedia of Statistics in Behavioural Science* 4: 2127–29.

Wickham, Hadley. 2021. *Tidyverse: Easily Install and Load the Tidyverse.* https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.