North Carolina Senior Games Participation Analysis

Created by Thomas Grover

Motivating Questions:

1. What trends can we see in pickleball participation? Is the sport thriving or declining?
2. What is the demographic of pickleball players? How old are most players?

Pickleball Trends Analysis:

The first step in the analysis was subsetting the participation data supplied by North Carolina Senior Games. We needed just records in which EventType was Pickleball. I created a grouping variable to separate the years into "Pre-Covid" and "Post-Covid" categories. This is useful in analysis, and makes logical sense given the real world implications caused by covid and quarantining period.

Frequencies of Pickleball Participation By Year

(Separated by Pre/Post-Covid Groups)

Pre|Post Covid	Event Year	Count
Pre-Covid	2016	**466**
	2017	**467**
	2018	**552**
	2019	**563**
Post-Covid	2022	**501**
	2023	**719**
	2024	**864**

Count represents the number of pickleball participants for the respective year. We can see the more recent years have higher frequencies. This may lead one to assume that pickleball is clearing trending upwards. But what does the data say?

To determine if there were any statistically significant differences in pickleball participation over the years, we conduct a T-test. Our target parameter is essentially the difference between the average participation before and after covid. (We can think of this as a small sample hypothesis test for a target parameter $\Theta = \mu_1 - \mu_2$).

It is important to note that this method assumes our observations come from a normal distribution.

Table 1: Summary Statistics of the Number of Pickleball Participants per Year by Era

Era	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Post-Covid		3	694.7	182.7	105.5	501.0	864.0
Pre-Covid		4	512.0	52.7320	26.3660	466.0	563.0
Diff (1-2)	Pooled		182.7	122.6	93.6129		
Diff (1-2)	Satterthwaite		182.7		108.7		

Table 2: T-Test Results

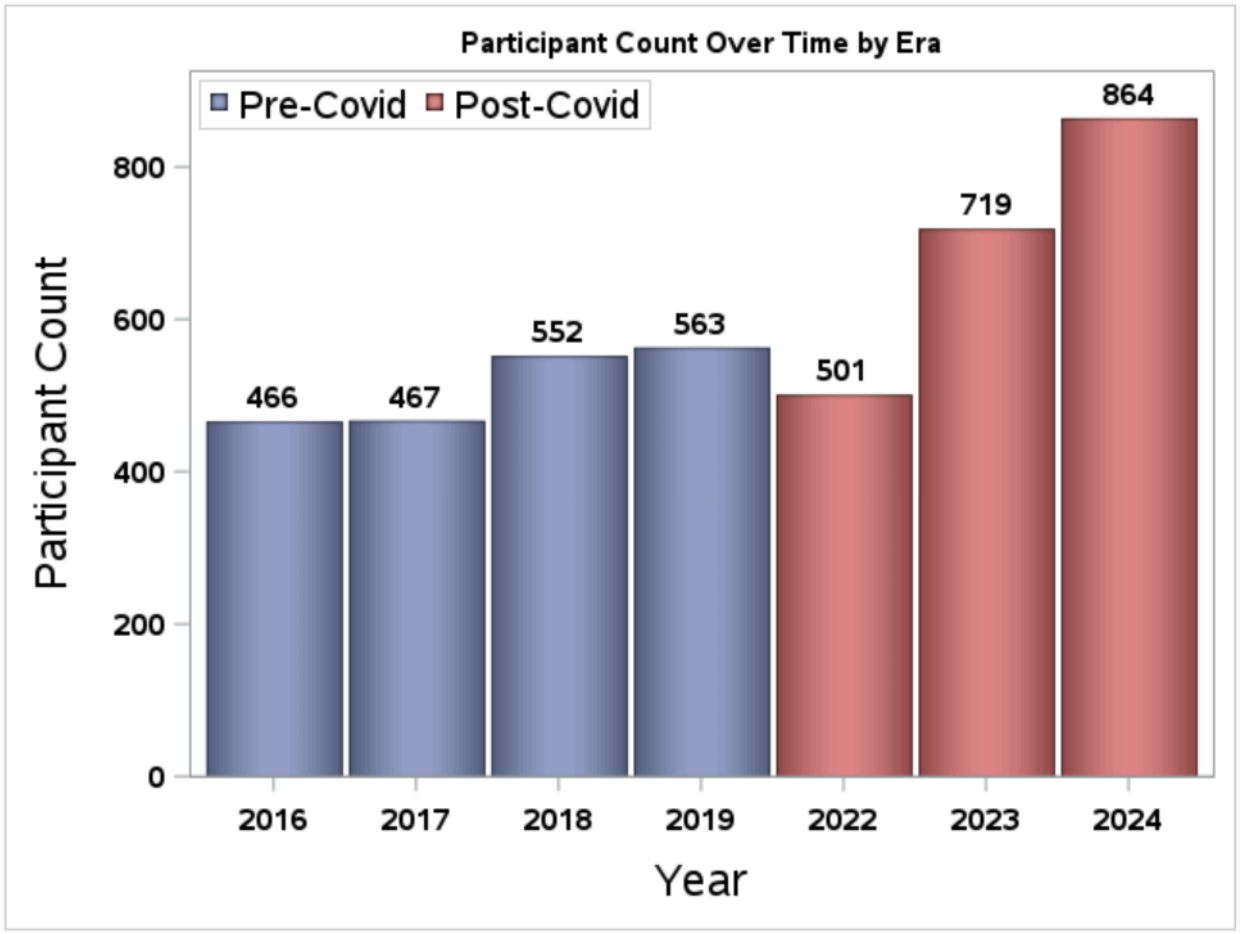
Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	5	1.95	0.1085
Satterthwaite	Unequal	2.2518	1.68	0.2209

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	2	3	12.01	0.0740

From the analysis in SAS, we can see that with a pooled variance test we get a p-value of p=0.1085, and when we don't assume equal variances we get a p-value of p=0.2209. We

can also see the p-value for equality of variances p=0.0740, meaning according to the data it is extremely unlikely that the two eras have the same variances.

We need to consider this in the context of the real world though, and especially with how limited our data is (test between samples of 3 and 4 observations respectively). I would argue that assuming equal variances actually makes sense here (same populations, same organizations data, same event), but clearly based on the very low p-value when testing for equality of variances, the data doesn't support this. So what can we say about the trends of pickleball participation? A visual will support our insights.



This graph shows the frequencies we calculated earlier in a visual manner, with Era's separated by color. The uptrend is visually significant, but our data doesn't provide clear statistical evidence of a difference between pre and post-covid years. However, in my

best judgement with all of this information taken into account, and the understanding of national and global pickleball trends (growing rapidly), ***I would tell NCSG that pickleball is one of their events trending upwards, and that they should market towards it more and invest more resources into it.***

Pickleball Demographic Analysis:

The next goal was gaining insight into the demographic of pickleball players. The first steps of analyzing Age involved getting basic summary statistics about the distribution of the variable. First, acquiring the basic measures of **center and variability**. And second, by taking a look at **quantile data** and **extreme observations** to get a better understanding of the context of the data.

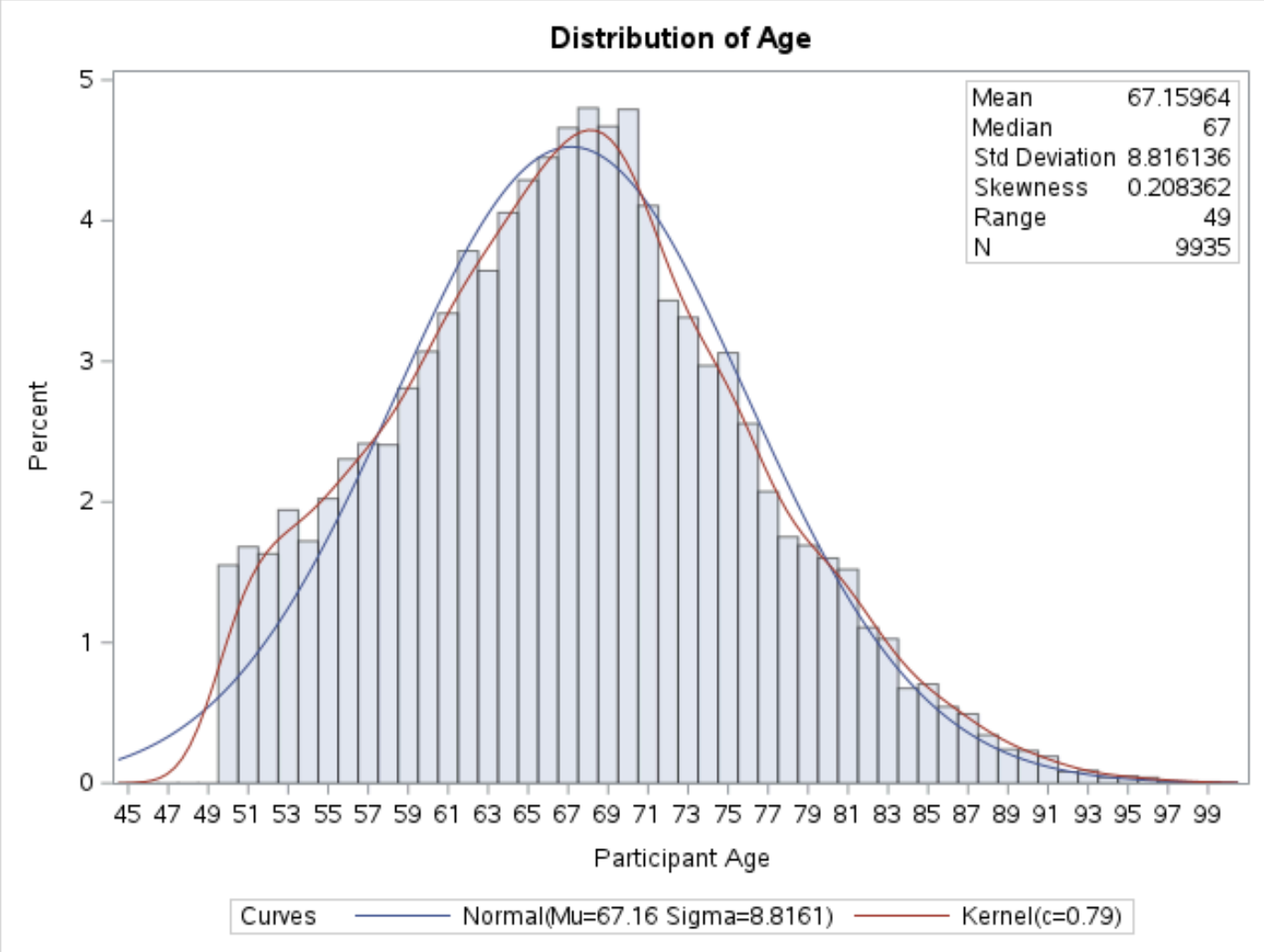
Basic Statistical Measures			
Location		Variability	
Mean	67.15964	Std Deviation	8.81614
Median	67.00000	Variance	77.72426
Mode	68.00000	Range	49.00000
		Interquartile Range	12.00000

Quantiles (Definition 5)	
Level	Quantile
100% Max	99
99%	88
95%	82
90%	79
75% Q3	73
50% Median	67
25% Q1	61
10%	55
5%	53
1%	50
0% Min	50

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
50	9883	96	6864
50	9877	96	9019
50	9838	97	5426
50	9818	98	7301
50	9815	99	7170

The next step I took in this was creating a visual representation of the overall distribution of players, to get a feel for who was actually playing pickleball. Here we have a histogram of the distribution of age, where **each bar represents a single age**. The Y-Axis represents the percentage of all pickleball players that specific age makes up. We also overlay a normal distribution with the **same mean and standard deviation we observed**, as well as a kernel distribution, for comparison.

Notice the **skewness value** in the legend, indicating we have **right skewed data with its positive value**.



So what can we conclude about our demographic?

1. All measures of center indicate that the middle of our **age distribution is somewhere between 67-68 years of age**.
2. Our **standard deviation of 8.816** indicates that about **68% of total pickleball players are 58-76 years old**.
3. When we take into account the minimum age of 50 for participation, it appears that given our large sample, the age of participants converges in distribution to a Normal distribution with mean mu = 67.16 and variance = 8.816^2.
   (Let Y_i = age of participant, $Y_i \sim N(67.16, 8.816^2)$ approximately)
