

SAS[®] GLOBAL FORUM 2018

USERS PROGRAM

How Effective is Change of Pace Bowling in Cricket?

Deven Chakraborty
Imperial College London

April 8 - 11 | Denver, CO
#SASGF

How Effective is Change of Pace Bowling in Cricket?

Deven Chakraborty

Imperial College London

Abstract

Cricket, similar to baseball in the usage of a bat and ball, is one of the most popular sports in the world, especially in India. At a whopping 411 million unique TV viewers in 2017 (Ahluwalia 2017), the Indian Premier League (IPL) cricket tournament has a follower base 25% larger than the entire population of the United States. T20 cricket, the format of the IPL, is dominated by batsmen, meaning bowlers (pitchers) struggle to keep swashbuckling pinch-hitters from smashing them out of the park. In the past few years, however, many bowlers have developed a strategy where a ball comes out much slower than their normal pace in an effort to deceive the batsman. But, no public data on bowling speeds exist. So, I recorded data from approximately 800 balls from last year's IPL to analyze the effectiveness of these slower balls. It turns out the slower balls are effective because these get batsmen out a statistically significant higher portion of the time than a normal pace ball. But, there are many more interesting questions that can be asked: Who has the best slower ball? Is there such thing as too slow of a slower ball? When are bowlers most likely to bowl a slower ball? All of these questions and more can be answered by my data set and can be used by cricket teams around the world to improve their standing in the extremely competitive industry of sport.

Methods

A majority of the work for this project was not actually working with the data because unfortunately, none of it existed beforehand. There is no public database where I could scrape bowling speeds off individual balls, so I had to watch and record the data for each of the 818 balls by hand. To select which balls to analyze, I picked the one fast bowler from each of the 8 teams who had played at least 10 matches in the season and recorded as many observations (as time permitted) for each one as possible. After that I used Base SAS 9.4 to organize and represent the data.

Examples

Regular

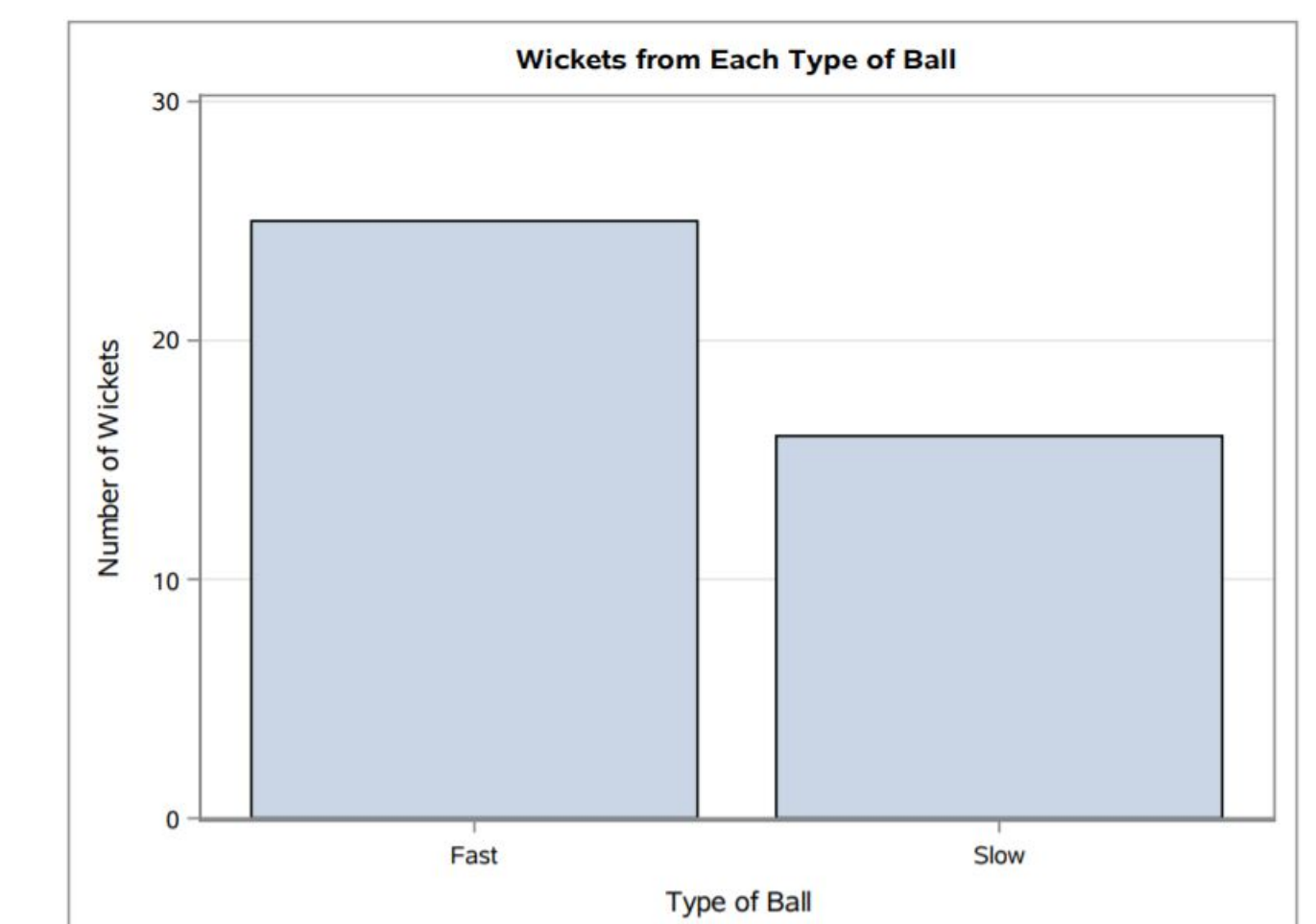
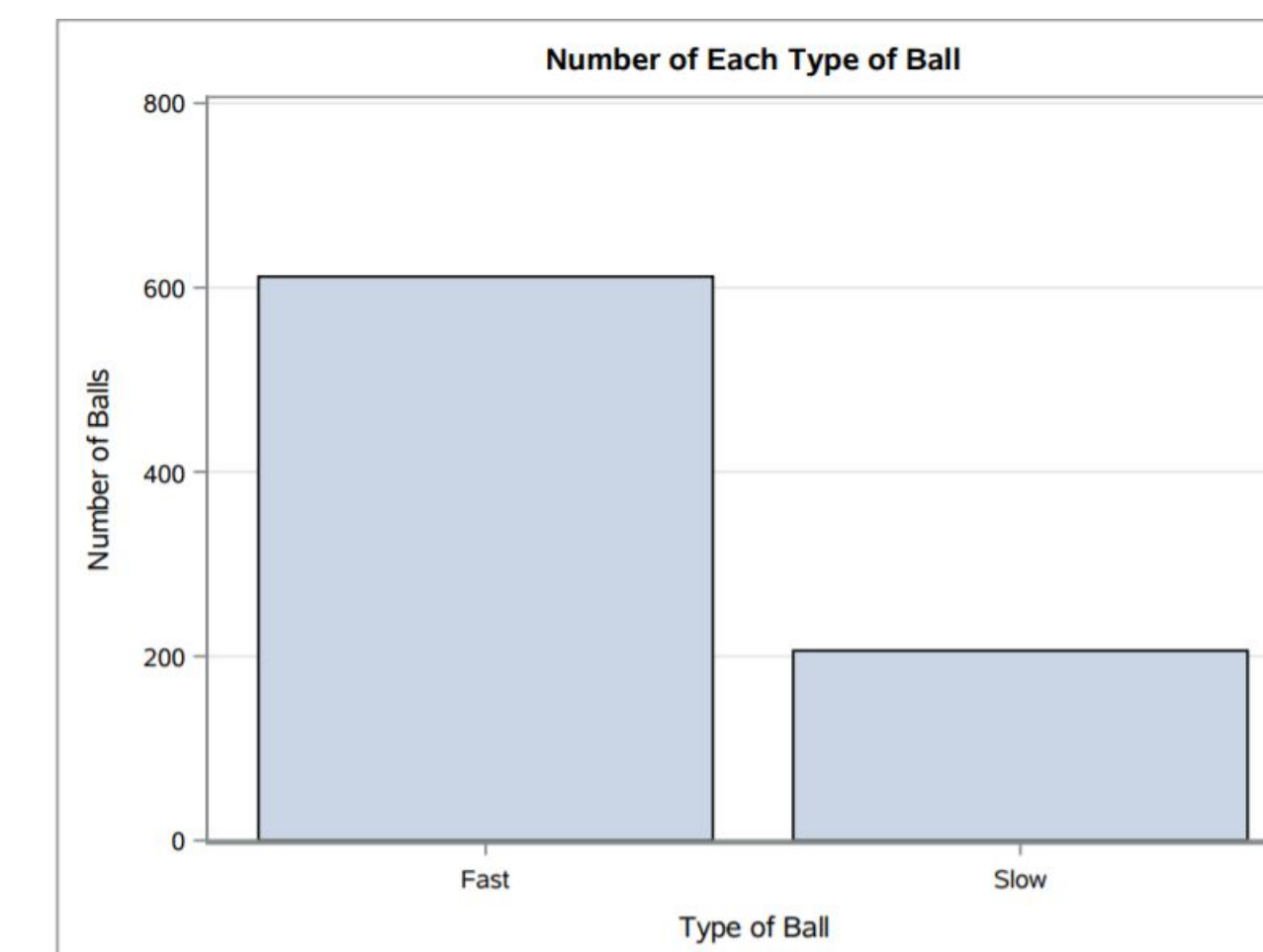


Slower



Is the Slower Ball an Effective Tactic?

The most important question is of course, *is the slower ball effective?* The answer is **yes**, and it can be shown by a one-sided significance test of proportions. 16 out of 206 slower balls managed to take a wicket, whereas 25 out of 616 regular balls managed to take a wicket. I consider H_0 to be $p_{\text{slow}} \leq p_{\text{normal}}$ and H_1 to be $p_{\text{slow}} > p_{\text{normal}}$ with a significance level of 0.05. The test yielded a p-value of 0.018, well below the significance level. Therefore, we can say slower balls take significantly more wickets than regular ones. These bar charts illustrate the significance:



Another viable significance test for the data was a Chi-Square test of association between slower balls and wickets. Running the test (where $DF = 1$) yielded a Chi-Square value of 4.39 and a p-value of 0.04, which, at a significance level of 0.05, shows that the 2 variables are in fact related. Furthermore, the proportion of wicket-taking regular balls was only 4.08%, whereas the proportion of wicket-taking slower balls was 7.77%, almost twice that of the regular ball. The corresponding output from a "Proc Freq" is shown here:

Wicket vs. Slower Ball Association

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of slow by wicket		
		wicket(wicket)	
	0	1	Total
slow(slow)			
0	587 71.76 95.92 75.55	25 3.06 4.08 60.98	612 74.82
1	190 23.23 92.23 24.45	16 1.96 7.77 39.02	206 25.18
Total	777 94.99	41 5.01	818 100.00

Statistics for Table of slow by wicket

Statistic	DF	Value	Prob
Chi-Square	1	4.3887	0.0362
Likelihood Ratio Chi-Square	1	4.0092	0.0453
Continuity Adj. Chi-Square	1	3.6494	0.0561
Mantel-Haenszel Chi-Square	1	4.3834	0.0363
Phi Coefficient		0.0732	
Contingency Coefficient		0.0731	
Cramer's V		0.0732	

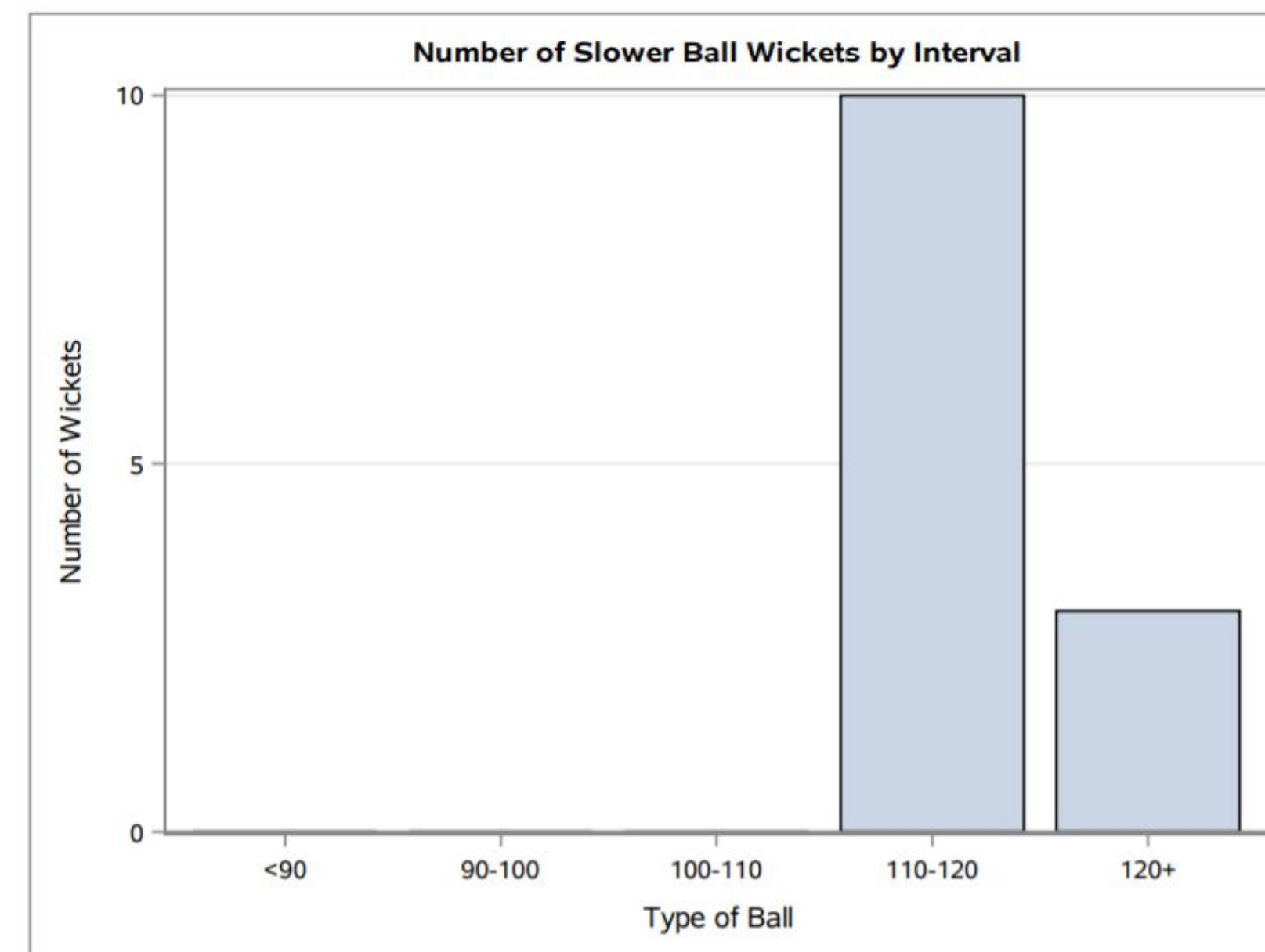
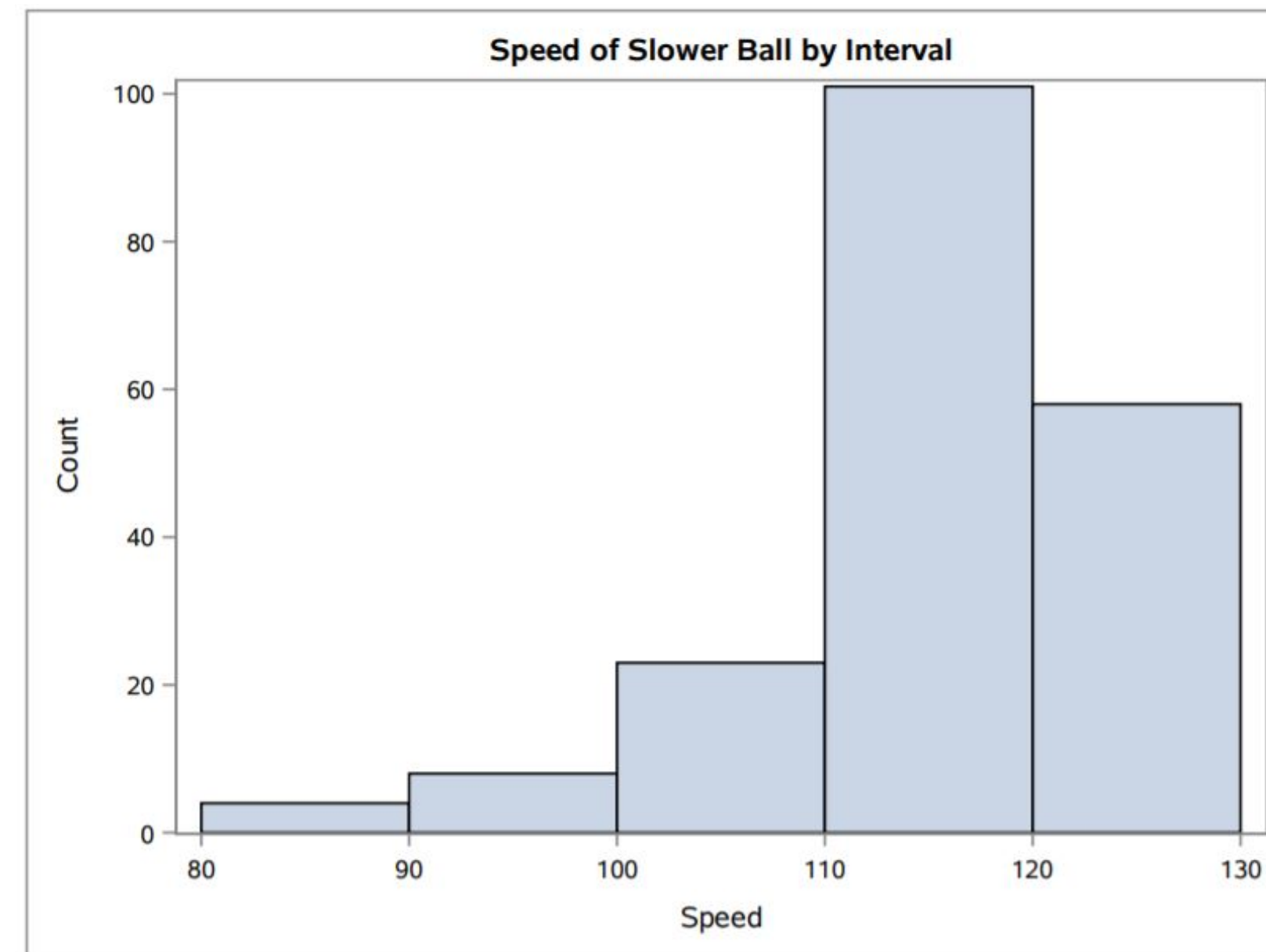
How Effective is Change of Pace Bowling in Cricket?

Deven Chakraborty

Imperial College London

What is the Optimal Speed for a Slower Ball?

We've seen slower balls are in fact effective in the game. But what is the best speed for them? I have highlighted some specific speed ranges in the graphs below:

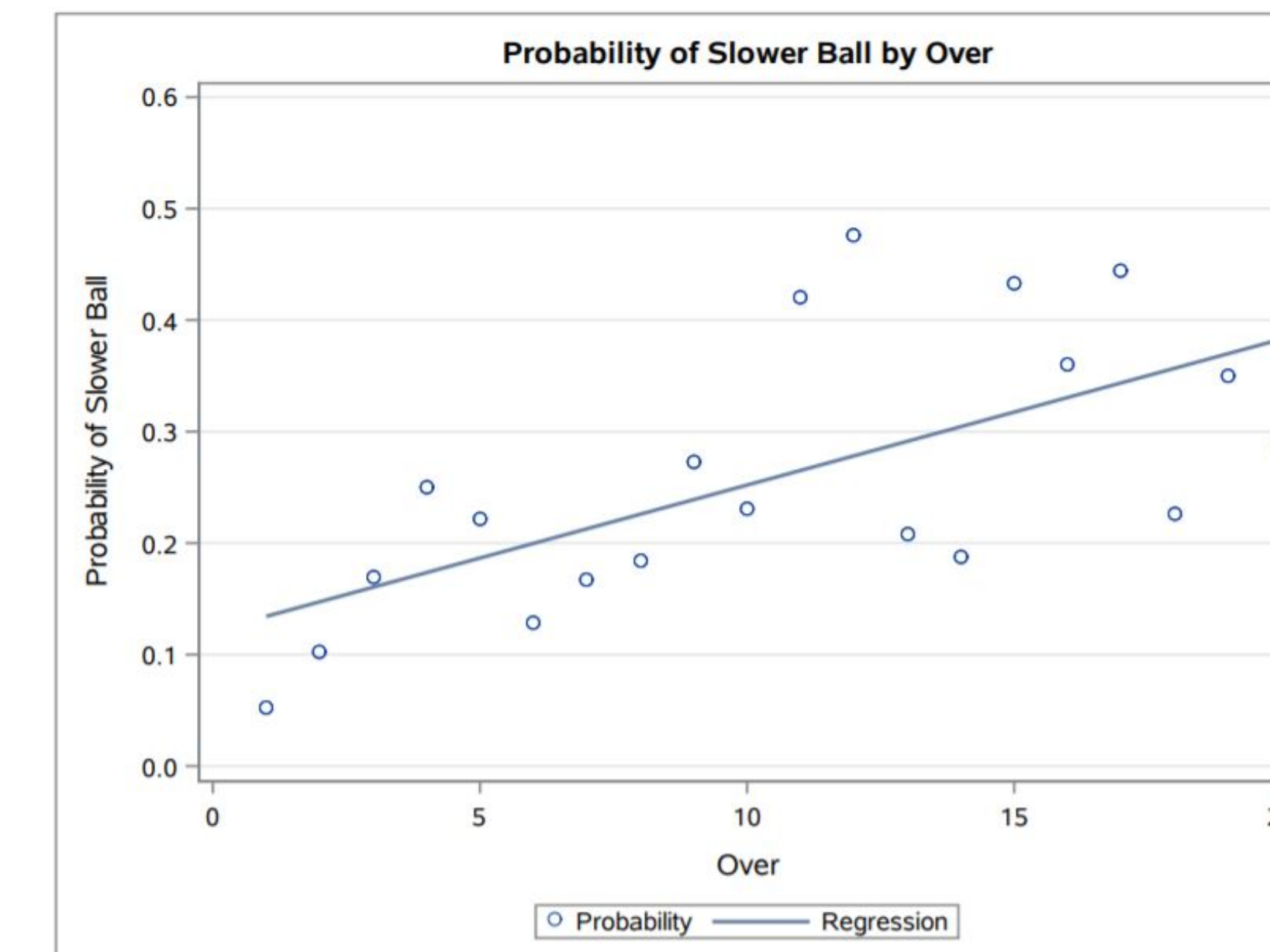


It should be noted that for bowlers with a higher average pace such as Pat Cummins, their slower variation can occur in the 120's (kilometers per hour). It should also be noted that only 196 of the observed 206 slower balls and 13 of the 16 slower ball wickets appear here. This is due to the fact that sometimes the ball would not receive a broadcasted speed, but I could tell from replays that it was an attempted slower ball.

It appears as if the 110-120kph range is the most effective. To test this, I used another z-test of proportions where $H_0 = p_{110-120} \leq p_{rest}$ and $H_A = p_{110-120} > p_{rest}$ in terms of wickets per ball (at a significance level of 0.05). This test yielded a p-value of 0.013, confirming that the 110-120kph range is the ideal one for slower balls for these professional bowlers.

When are Bowlers Most Likely to Bowl a Slower Ball?

Perhaps the most useful insight this dataset can give is when a fast bowler is most likely to deliver a slower ball. The following graph and "Proc Reg" output models the probability of a bowler delivering a slower ball in the given over:



Probability of Slower Ball by Over

The REG Procedure
Model: MODEL1
Dependent Variable: Probability

Number of Observations Read	20
Number of Observations Used	20

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.11390	0.11390	12.71	0.0022
Error	18	0.16126	0.00896		
Corrected Total	19	0.27516			

Root MSE	0.09465	R-Square	0.4139
Dependent Mean	0.25860	Adj R-Sq	0.3814
Coeff Var	36.60201		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.12118	0.04397	2.76	0.0130
over	1	0.01309	0.00367	3.57	0.0022

The overall regression model is statistically significant at 5% level ($F=12.71$, $p\text{-value}=0.0022$) and the explained variance is about 41.4%. The coefficient of the over shows that as number of overs increases, the probability of a slow ball increases by 1.31%.

The data's upward trend shows that as the game goes on, bowlers are more likely to bowl slower balls, which most cricket fans would expect. At the latter stage of an innings, batsmen are trying to get as many runs as possible, so bowlers try to negate this by putting off their rhythm with slower balls.

Although these matches will never extend beyond 20 overs, this analysis is important since it gives a better estimate of the probability that a slower ball will be delivered in a specific over than just the mean of the probabilities.

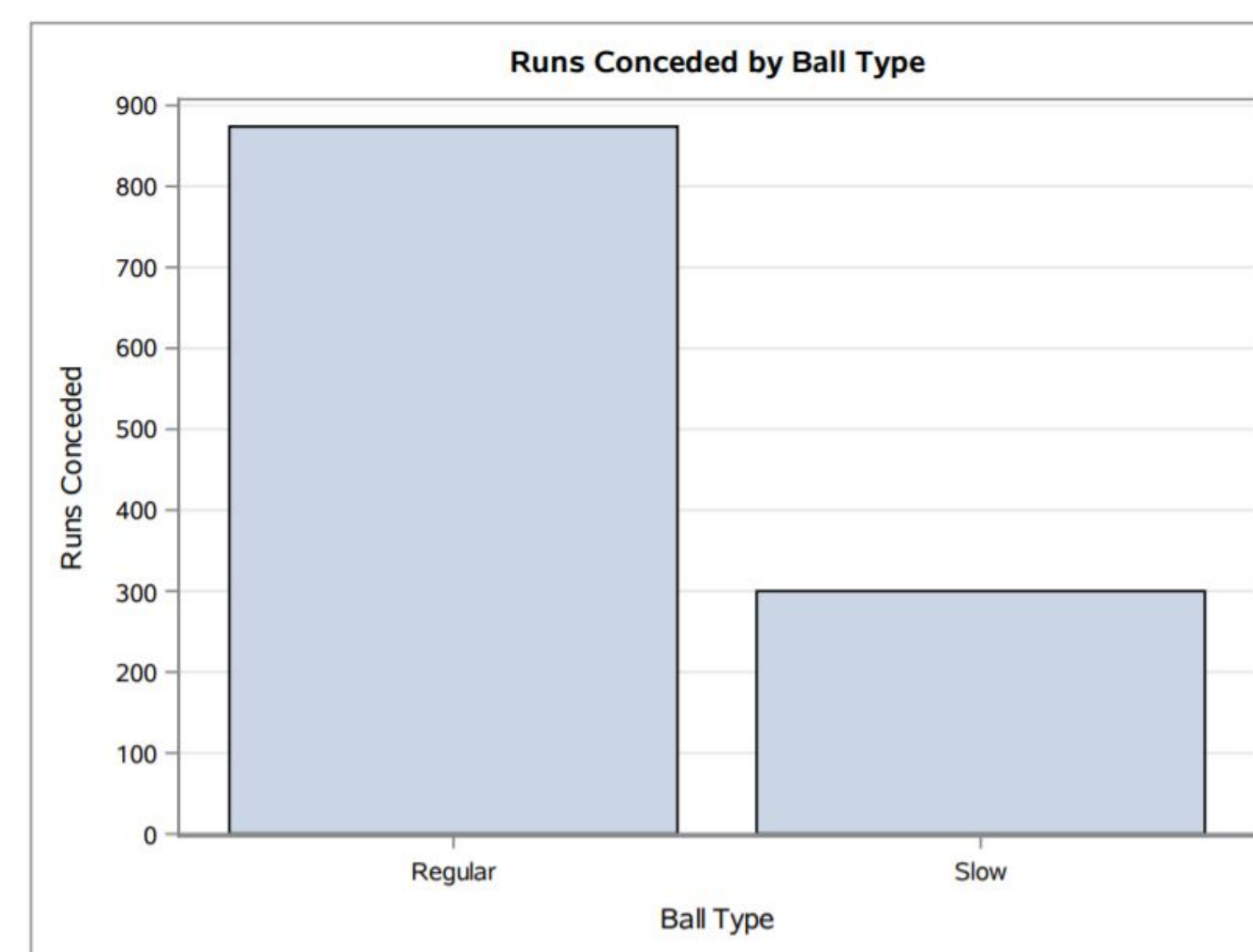
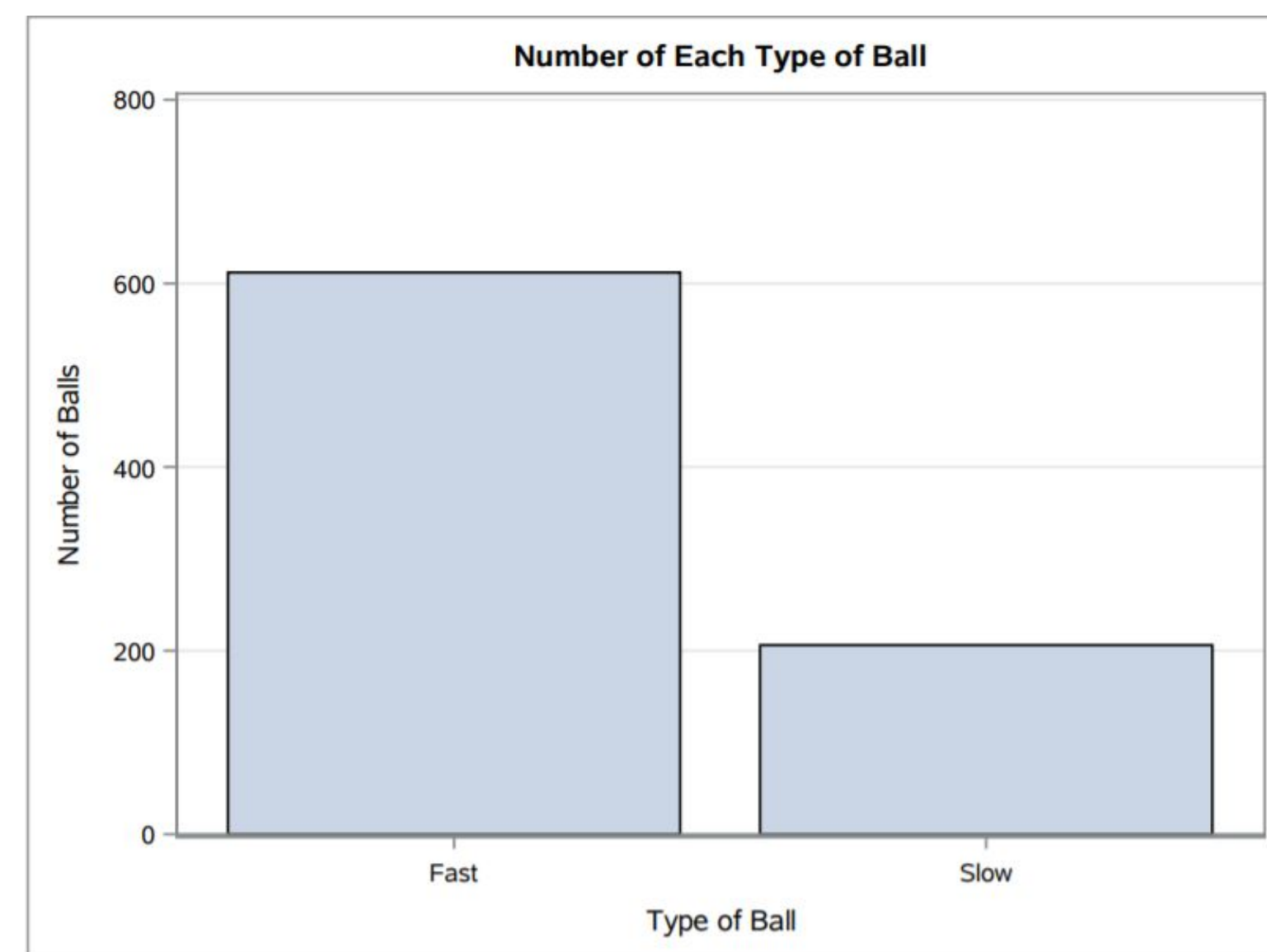
How Effective is Change of Pace Bowling in Cricket?

Deven Chakraborty

Imperial College London

Does the Slower Ball Give Away More Runs?

We've seen how the slower ball is better at taking wickets than a normal one, proving its validity as a strategy. But does this come at an additional price? We've already seen a representation of the quantities of slower and regular balls against the number of wickets they took, now I chose to compare the quantities against the runs they conceded:



As you can see, the ratio looks quite similar, so I decided to test for a difference between the average number of runs given away per ball using a t-test of means. I chose H_0 to be $\bar{x}_{\text{slower}} = \bar{x}_{\text{regular}}$ and H_A to be $\bar{x}_{\text{slower}} \neq \bar{x}_{\text{regular}}$. The test yielded a p-value of 0.78, implying the null hypothesis cannot be rejected. So it is unclear as to whether or not the slower ball gives away more runs, which may come as a surprise to many cricket fans who believe it to be much more risky than any other type of ball.

Which of the Venues is Best for Slower Balls?

Now if you haven't watched cricket before, this may not make sense to you. 'Why would the stadium impact how good the players are?' you may ask. Well in cricket, since the ball bounces for every pitch, the strip where the ball lands (known also as a pitch) influences every ball. If the pitch is soft, the ball may not bounce much. If it's hard, it may bounce a lot. Since groundskeepers often have different theories on how to maintain the pitch, the efficacy of a slower ball can be greatly affected by the location of the match. A table detailing the slower ball statistics is shown below:

Venue (Team)	Number of Matches	Slower Balls	Slower Ball Wickets	Wickets per Ball
Mohali (KXIP)	3	6	1	0.167
Hyderabad (SRH)	7	48	6	0.125
Kolkata (KKR)	6	18	2	0.111
Mumbai (MI)	6	32	3	.094
Delhi (DD)	4	14	1	0.071
Pune (RPS)	5	15	1	.067
Bangalore (RCB)	5	38	2	0.053
Rajkot (GL)	4	35	0	0
Total	40	206	16	.078

To see which ground was the most conducive, I ran a few more tests of proportion (significance level of 0.05) for Mohali and for Hyderabad. Both tests were inconclusive (p-values of 0.20 and 0.08). Despite not having any slower ball wickets in Rajkot and thereby making any inference using it potentially unreliable, I ran a test to see if Rajkot was in fact the worst ground to bowl a slower ball, and a p-value of 0.03 showed that Rajkot, the home of the Gujarat Lions, is in fact the worst place to bowl a slower ball. This wouldn't have much effect next year however since as of 2018, Gujarat Lions are a defunct franchise.

How Effective is Change of Pace Bowling in Cricket?

Deven Chakraborty

Imperial College London

Conclusion

In summary, the slower ball is an excellent tactic for any fast bowler, as it takes more wickets than a regular ball without giving away more runs. But beyond the slower ball, cricket is one of the biggest sports in the world and is sadly under-analyzed compared to many other sports. I have asked just a few of the countless questions data can answer in cricket. In addition, although my project did not yield any outright shocking results, I had some outcomes which may be slightly surprising, and I believe the quantification of the wide-held beliefs about slower balls of many cricket fans is important nonetheless. Finally, for those of you who are not cricket fans, I hope I have sparked your interest in the game, and the next time it's in the news, you'll know a little bit more about what's going on.

References

- Ahluwalia, H. (2017). *IPL viewership jumped 22.5% in 2017: BARC*. [online] <http://www.livemint.com/>. Available at: <http://www.livemint.com/Consumer/Y8J13d9IObcYS7xFFcUXNI/IPL-viewership-jumped-225-in-2017-BARC.html> [Accessed 21 Sep. 2017].

Sample Code

```
*Full dataset creation;

libname bowling 'D:\Documents\IPL 2017 Project\bowlingData.xlsx'; *Creating library from Excel sheet;

data sasuser.all; *Creating comprehensive dataset;
length name $ 8;
set bowling.'bkumar$'n bowling.'bthampi$'n bowling.'saravind$'n bowling.'pcummins$'n
    bowling.'cwoakes$'n bowling.'jbumrah$'n bowling.'ssharma$'n bowling.'junadkat$'n;
if name ^= ' ' and slow ^= ' '; *Ignoring incomplete records;
totalBalls + 1; *Creating counters for slower and regular balls and wickets;
totalWickets + wicket;
if slow = 1 then do;
    slowWickets + wicket;
    slowBalls + 1;
end;
else do;
    fastWickets + wicket;
    fastBalls + 1;
end;
format name $8.;
run;
```

Sample Code (Cont.)

```
*Creating dataset for analyzing cost of a slower ball;

data sasuser.slowCost;
set sasuser.all;
if slow = 1 then do; *Slower ball stats;
    slowRuns + runs + extra; *Include 'extras' values;
    slowExtras + extra;
end;
else if slow = 0 then do; *Regular ball stats;
    fastRuns + runs + extra;
    fastExtras + extra;
end;
drop opponent venue over score wickets totalBalls
    totalWickets fastWickets fastBalls slowWickets slowBalls;
run;

*Chi-Square test of association between slower balls and wickets;

ods pdf file='X2.pdf'; *Output to pdf;
proc freq data=sasuser.all;
title 'Wicket vs. Slower Ball Association';
tables slow * wicket /
    nocum
    chisq
    scores=table
    alpha=0.05; *Tables statement with options;
run;
ods pdf close;

*Linear regression of chance of
slower ball in any given over;

ods pdf file= 'procReg.pdf';
*Special dataset created from datalines;
proc reg data=sasuser.slowProbOvers;
*Model over against chance of slower ball;
model Probability = over;
run;
ods pdf close;
```



SAS[®] GLOBAL FORUM 2018

April 8 - 11 | Denver, CO
Colorado Convention Center

#SASGF