Final_Project_Essig_Wesley

Wesley Essig

12/10/2021

Introduction

A 100-mph fastball takes ~400 milliseconds to reach the plate. This is roughly the same amount of time it takes to blink. Batters are relying on razor-thin margins of reaction time and bat speed to make contact with the ball. It's intuitive that velocity should be advantageous for the pitcher – the greater the pitch velocity the less time a batter has to react and make a decision to swing. Striking out a batter depends on a lot more than just velocity however.

The question I asked is this: Ignoring other variables, how much of strikeout rate can be attributed solely to velocity. And is it different for different types of pitches?

The data set

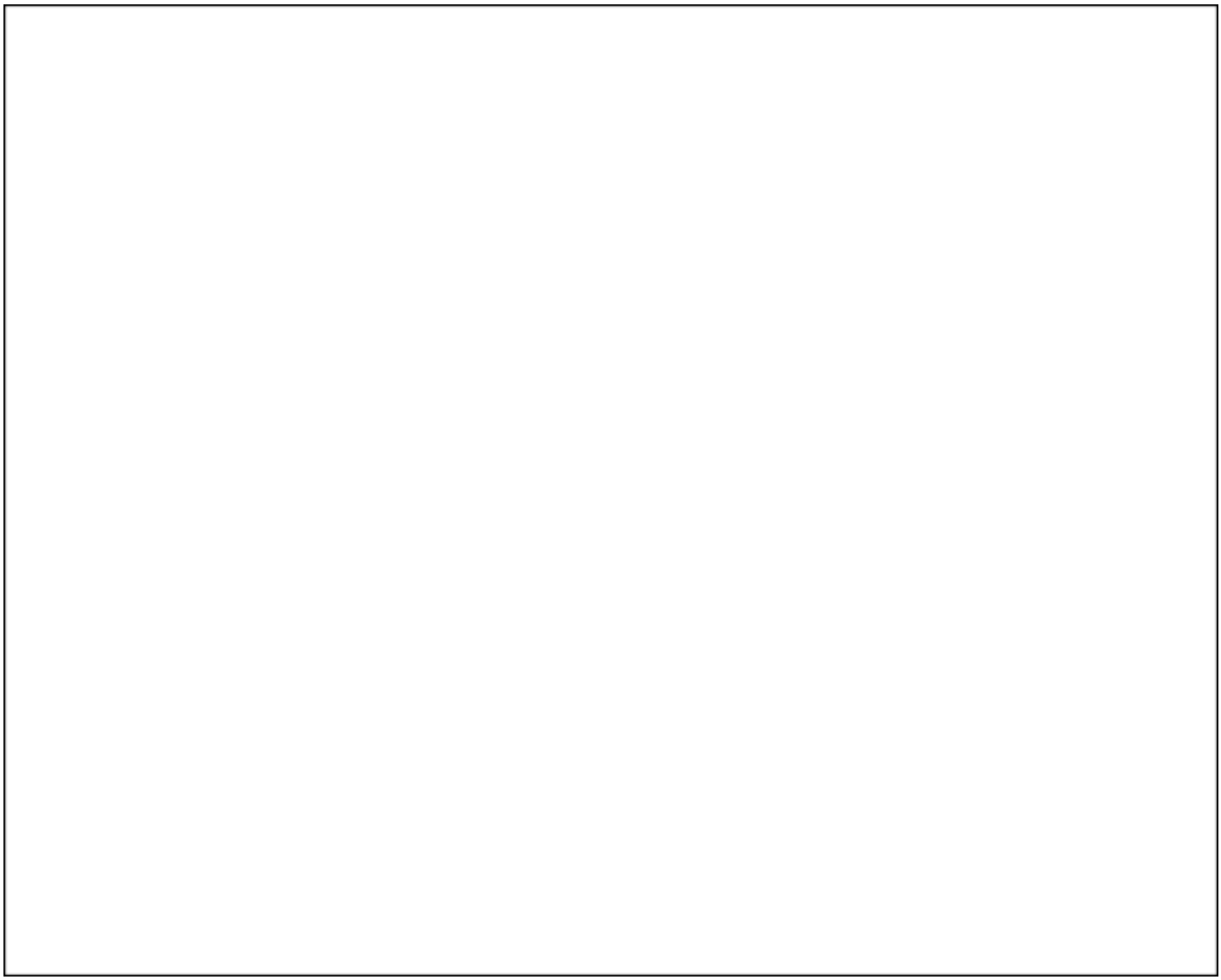
I used fangraphs.com to obtain my data set. Specifically, I chose the leader boards for qualified pitchers from 2021. To qualify, a pitcher must have pitched at least 160 innings. Limiting the data to qualified pitchers meant each pitcher threw anywhere from 2500 to 3200 total pitches. To further remove any small sample size effects, I manually removed any data indicating the pitcher threw the pitch less than 2% of the time. This ensured that every pitcher whose average velocity I incorporated threw that pitch at least 50 times.

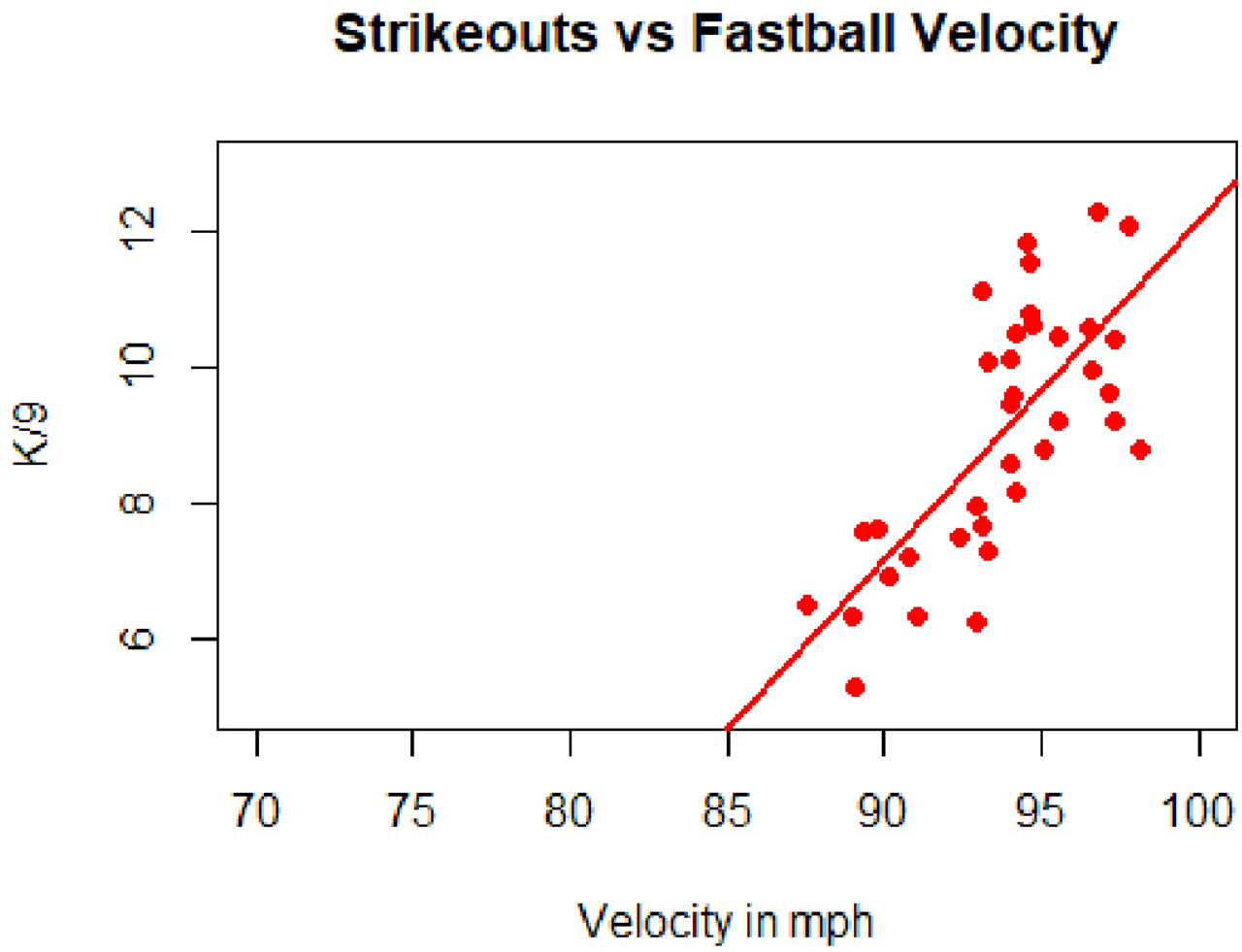
I also had restricted the data collected to 5 pitch types to ensure a large enough sample of pitchers. Only fastballs (36 pitchers), sinkers (27 pitchers), changeups (33 pitchers), curveballs (30 pitchers), and sliders (30 pitchers) were used by a majority of pitchers. Pitch velocity units are miles per hour. Strikeouts are measured per 9 innings pitched; using a rate was necessary for normalization because not every pitcher pitched the same number of innings.

The methods

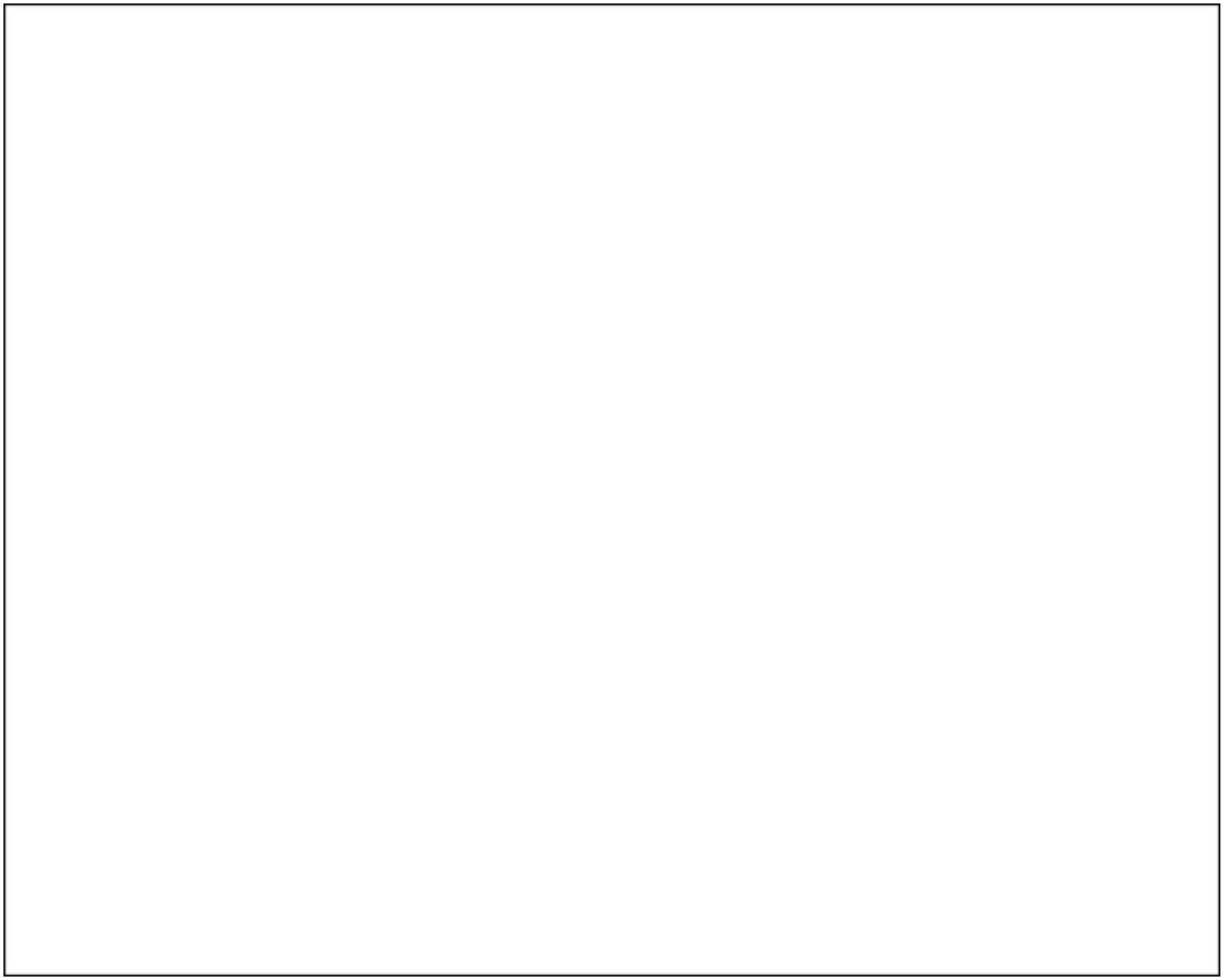
I used ordinary least squares regression to explore the relation between velocity and strikeout rate for each of the 5 pitch types. The null hypothesis was there is no relation between velocity and strikeout rate. In the appendix, I assess the assumptions of OLS regression: homoscedasticity, linearity, and normality of the residuals.

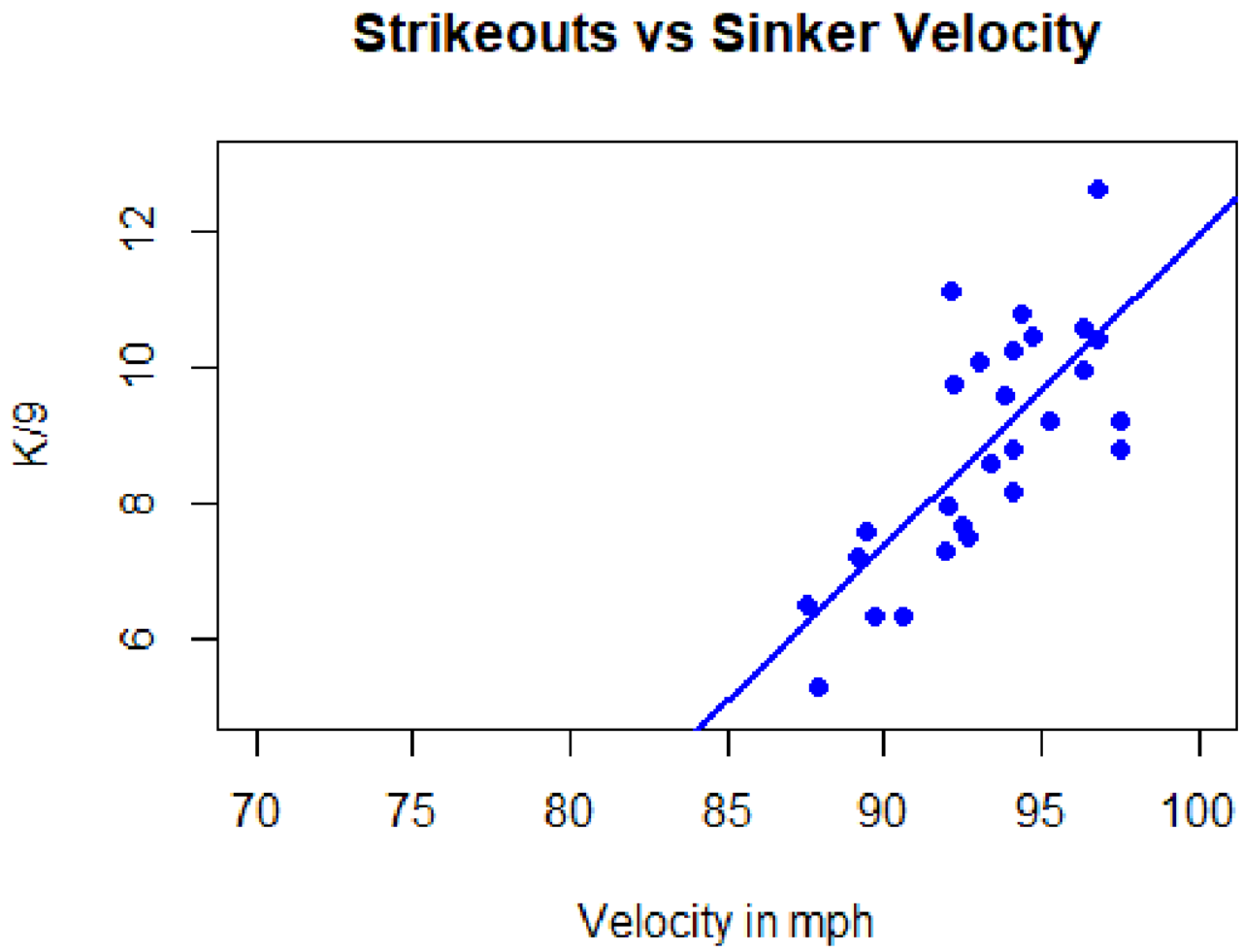
Results



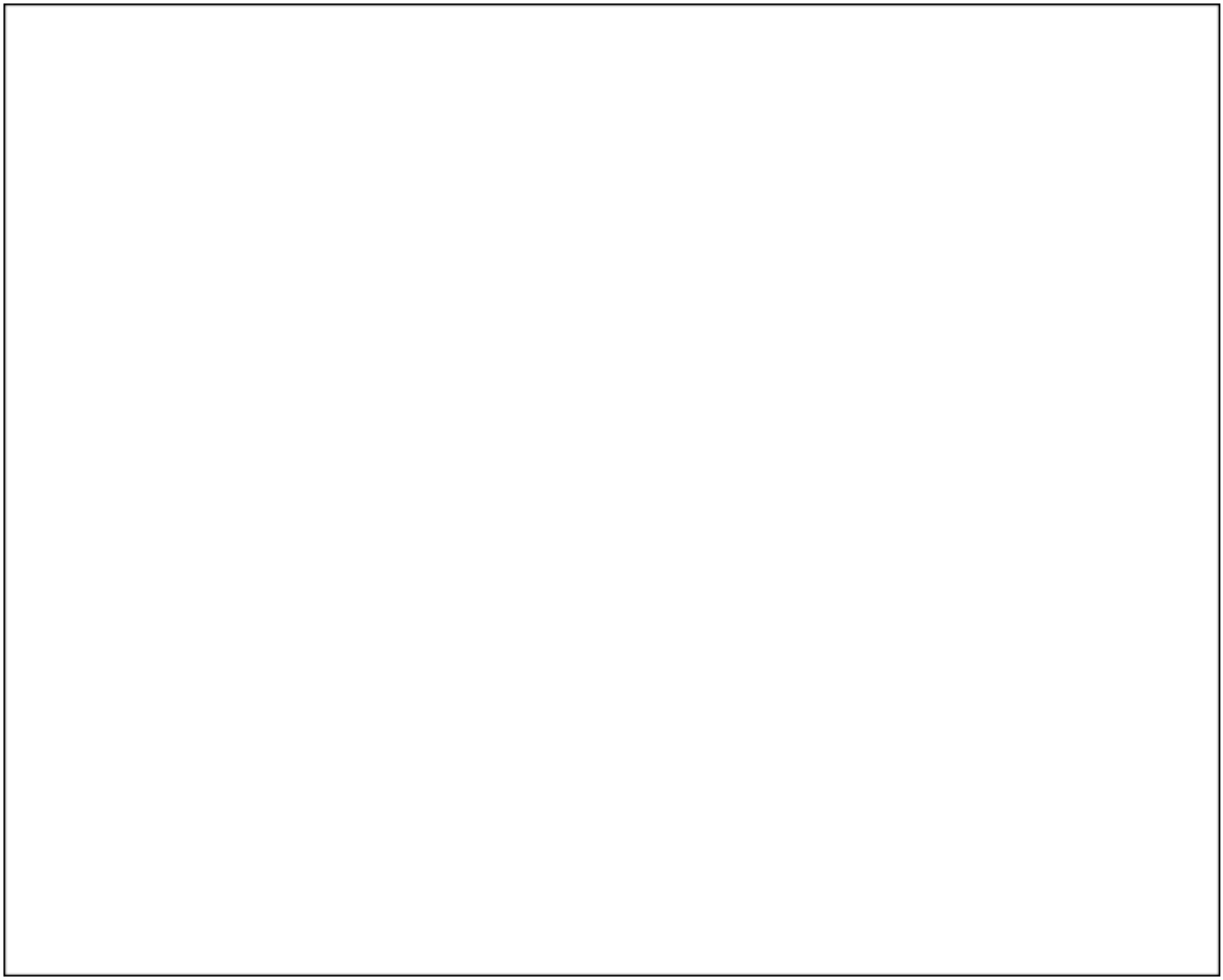


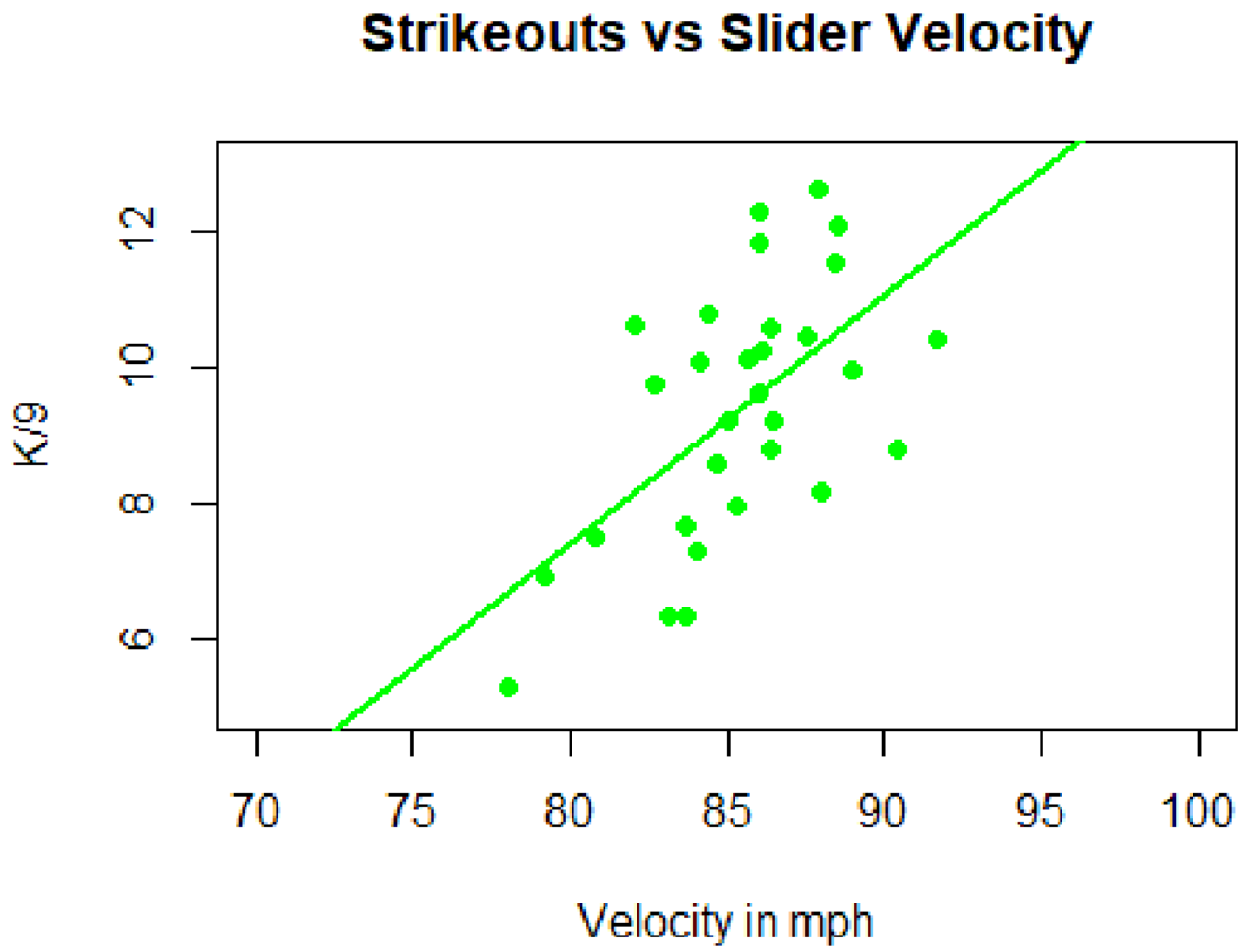
This a plot of strikeout rate vs fastball velocity (in mph) with a linear regression line superimposed. For every 1-mph increase in fastball velocity, a pitcher increases their strikeout rate (strikeouts per 9 innings pitched) by an average of 0.498. 95% CI: 0.33906 to 0.6567. p value : We can reject the null hypothesis here – a clear association exists between fastball velocity and strikeout rate.



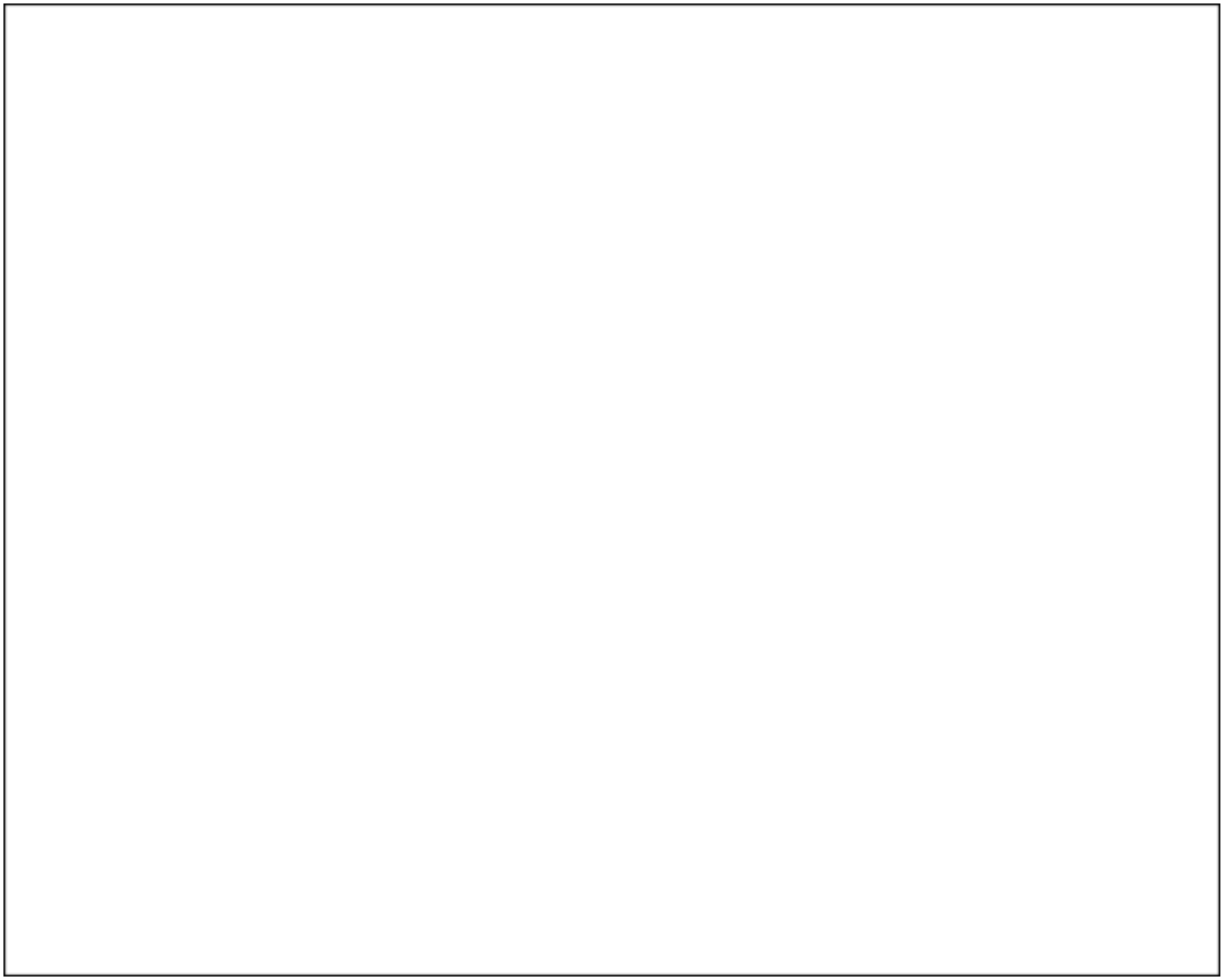


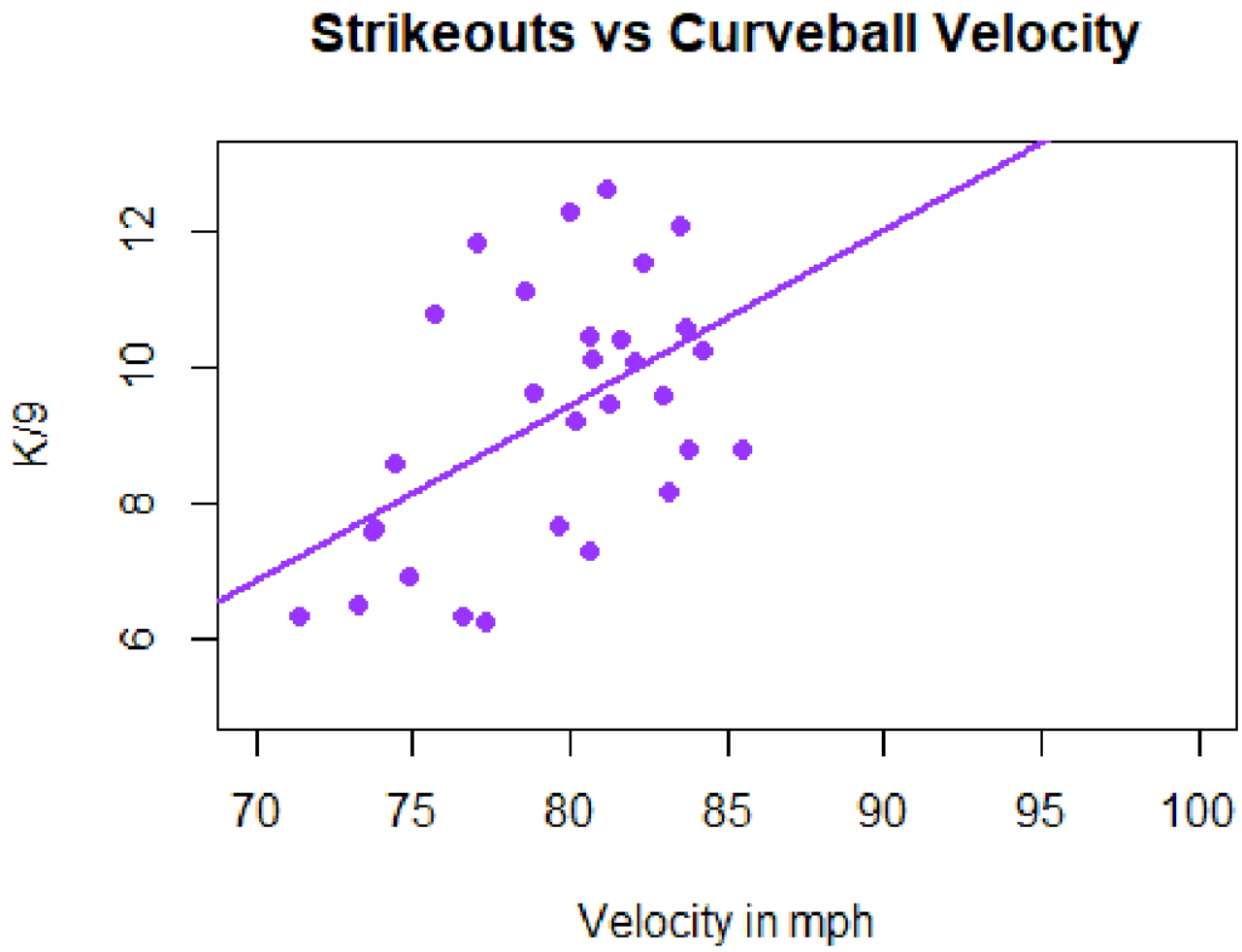
The results for sinkers look very similar to those for fastballs. For every 1-mph increase in sinker velocity, a pitcher increases their strikeout rate (strikeouts per 9 innings pitched) by an average of 0.458. 95% CI: 0.29243 to 0.62327. p value : We can reject the null hypothesis here as well as a clear association between sinker velocity and strikeout rate exists.



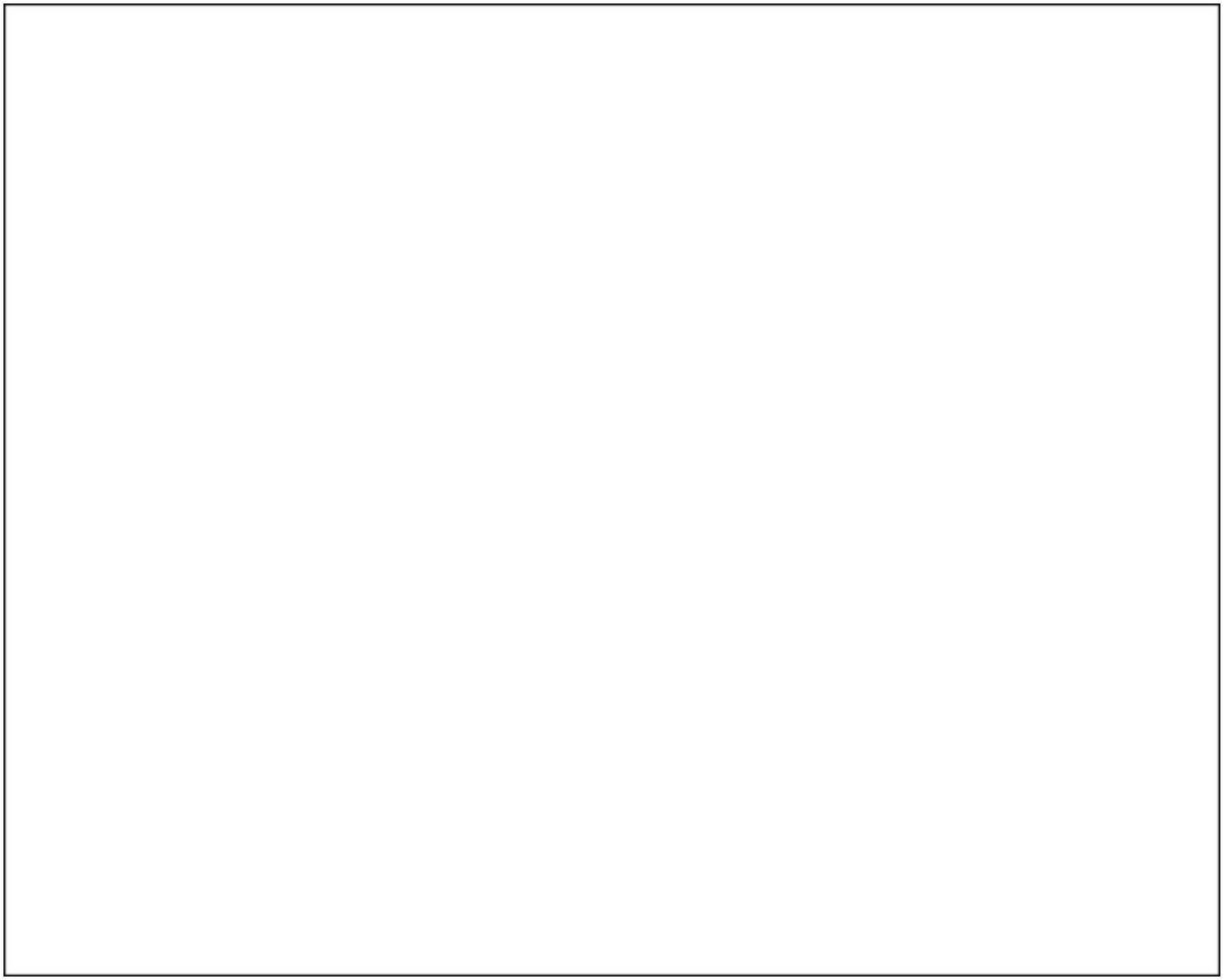


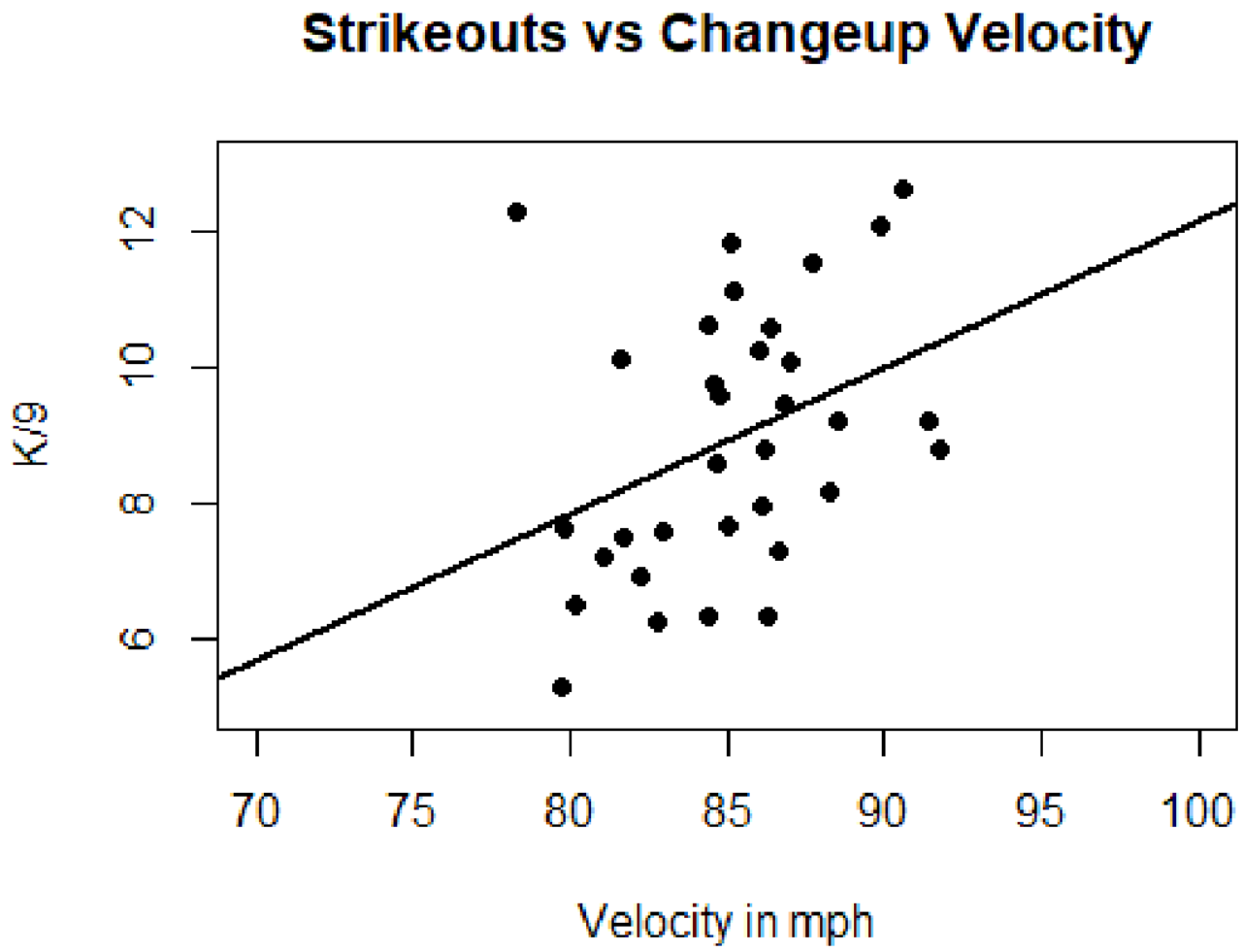
For every 1-mph increase in slider velocity, a pitcher increases their strikeout rate (strikeouts per 9 innings pitched) by an average of 0.367. 95% CI: 0.17675 to 0.55739. p value : We can still reject the null hypothesis here as well, but slider velocity is less correlated to strikeout rate than first 2 pitches.





For every 1-mph increase in curveball velocity, a pitcher increases their strikeout rate (strikeouts per 9 innings pitched) by an average of 0.258. 95% CI: 0.09387 to 0.42195. p value: We can still reject the null hypothesis here, but curveball velocity is even less correlated to strikeout rate than slider velocity is.





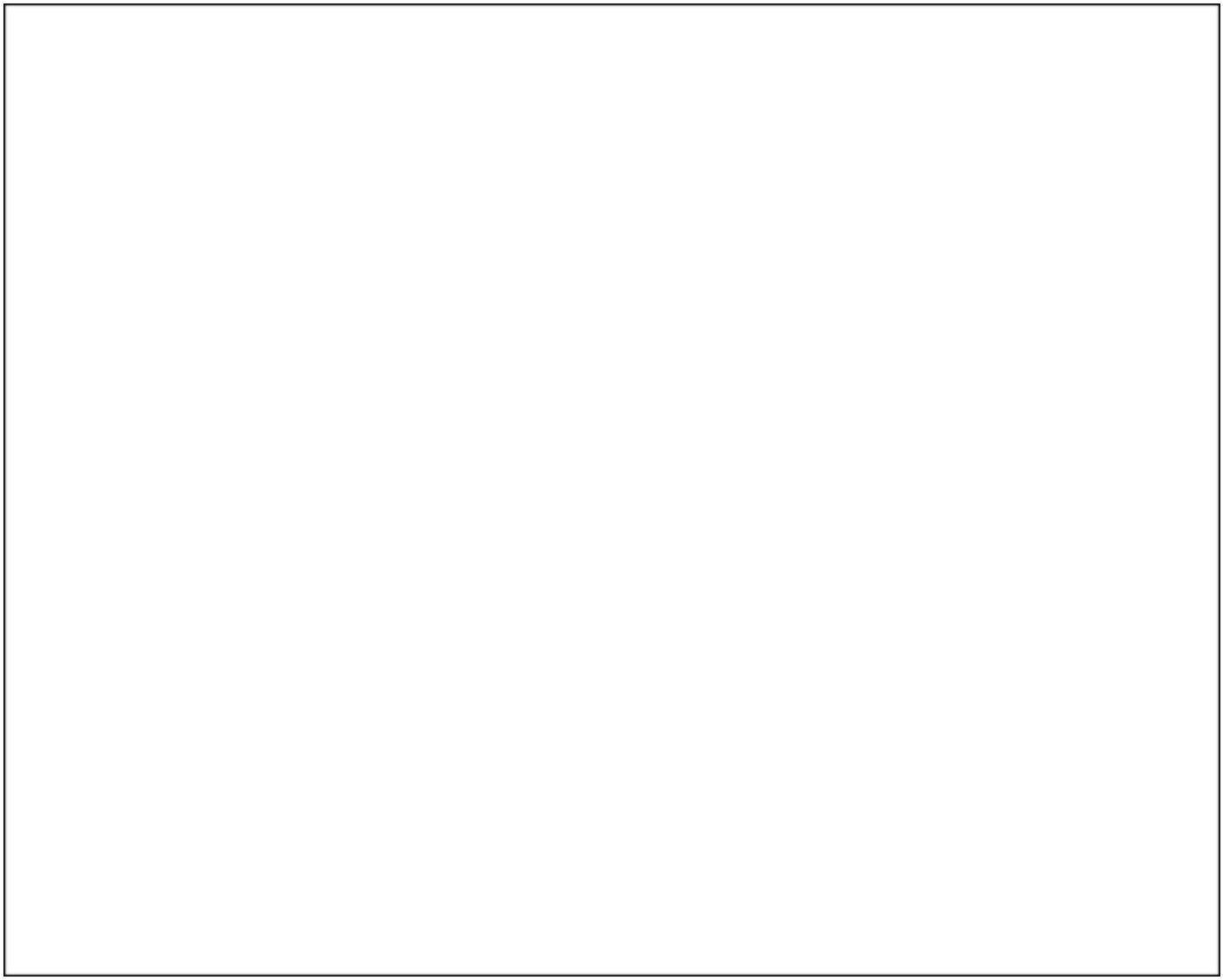
For every -mph increase in changeup velocity, a pitcher increases their strikeout rate (strikeouts per 9 innings pitched) by an average of 0.216. 95% CI: 0.022 to 0.4104. p value: We can just barely reject the null hypothesis for changeups and the relation appears to be weak.

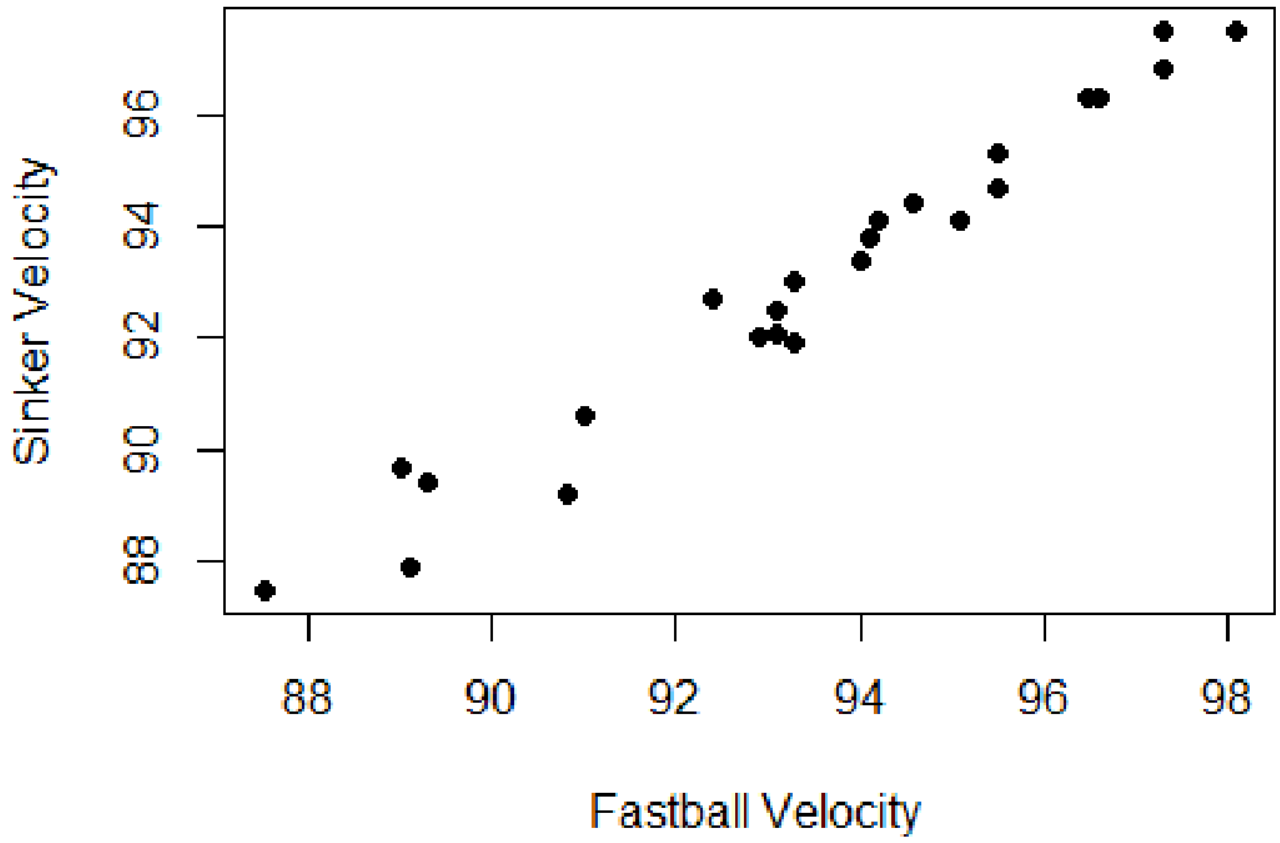
Summary Table

```
##           slope   p value
## fastball  0.498 3.86e-07
## sinker    0.458 9.39e-06
## slider    0.367 6.15e-04
## curveball 0.258 3.92e-03
## changeup  0.216 3.34e-02
```

Fastballs

For both fastballs there is a very similar and significant positive relation between strikeouts and velocity. It shouldn't be surprising that fastballs (aka "4-seam fastballs") and sinkers (aka "2-seam fastballs") have similar results. They are both thrown with essentially the same grip and arm motion; the main difference between them is a 90-degree rotation in the ball. This changes the way the seams create drag as the ball moves through the air. A plot of sinker velocity vs. fastball velocity shows just how strong the relation is between the two pitches.





Breaking balls

It's also natural to group curveballs and sliders together. These two pitches also have the same grip and a continuum exists between them depending on the amount of side spin (more of a slider) vs. topspin (more of a curveball). Statcast draws a line and divides them two into binary categories. It's worth mentioning that both pitches have much more horizontal and vertical movement than either type of fastball, which makes it more difficult for pitchers to place them accurately. I suspect that pitch location plays a larger role than velocity here; however, velocity still correlates significantly to strikeouts, although to a lesser degree than for fastballs.

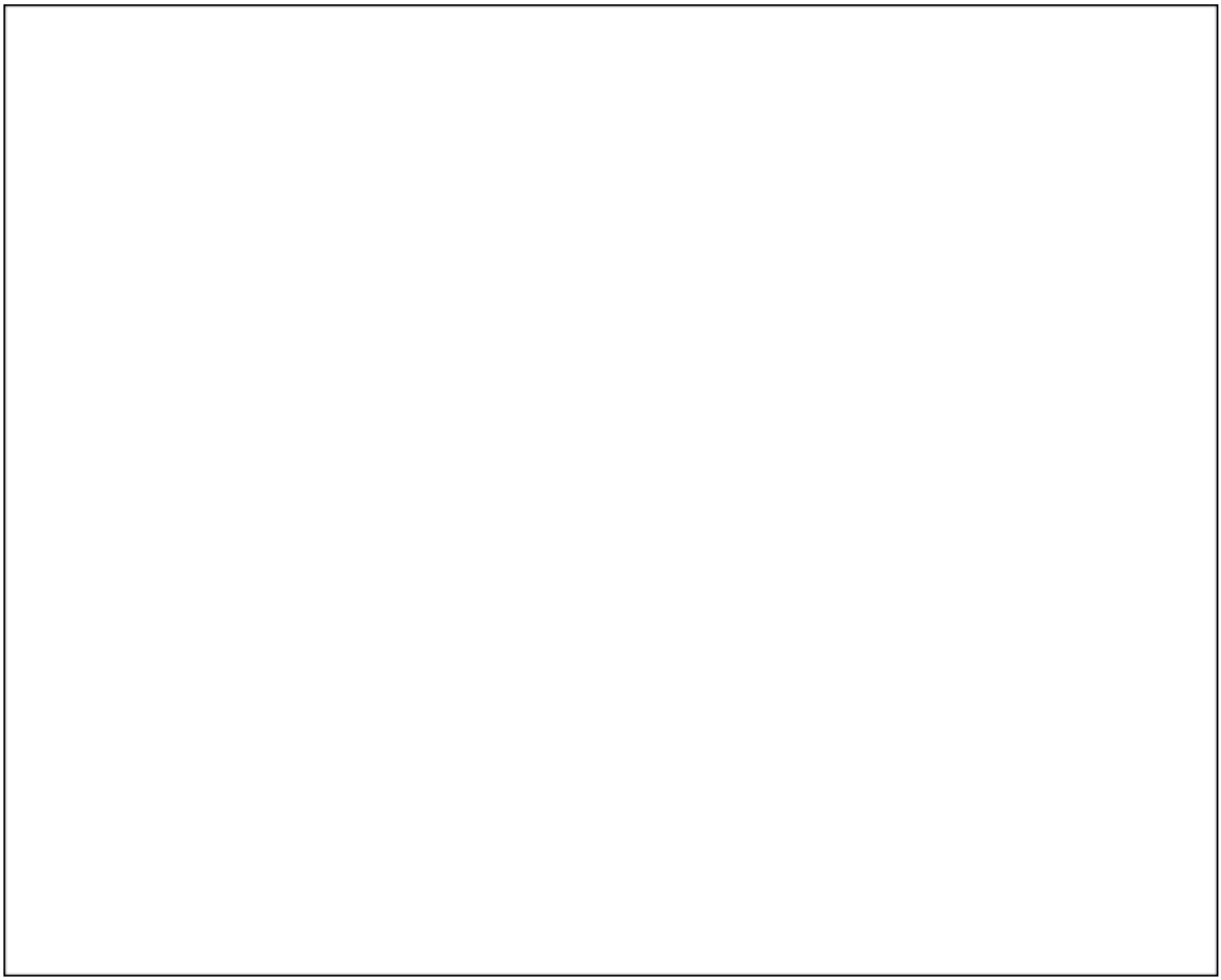
Changeups and Dylan Cease

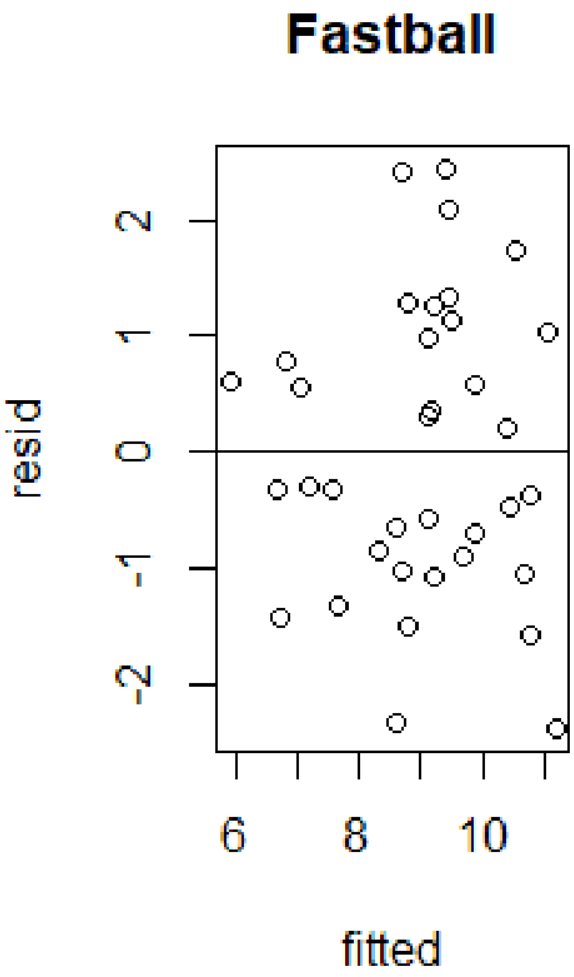
Changeups are in a category of their own. They are the only so-called "deception" pitch. They are thrown with the same motion and grip as a fastball, but the pitcher holds the ball

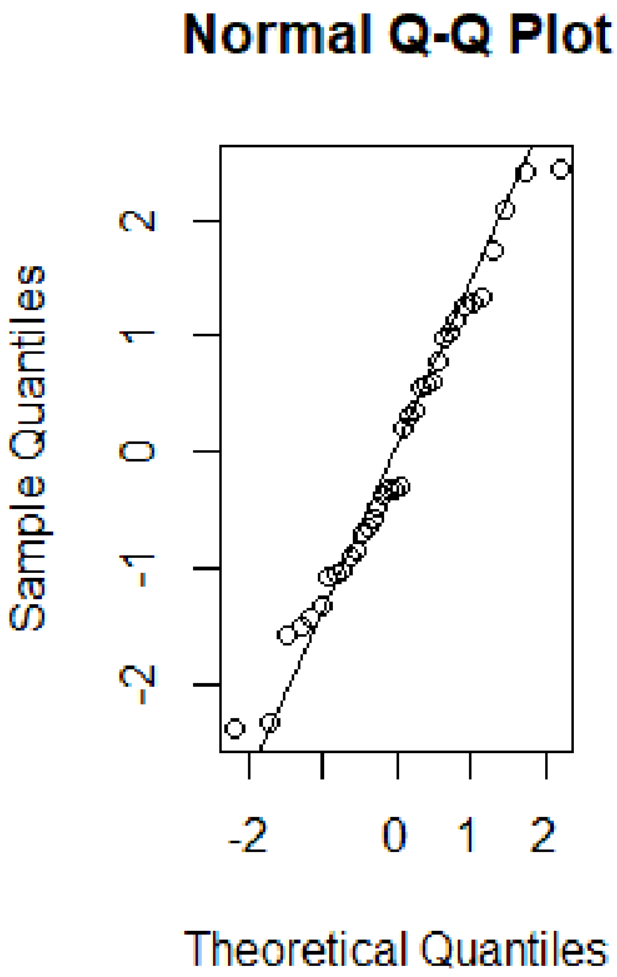
closer to his palm, causing the ball to come out slower than the fastball. The idea is to trick the batter into thinking it's a fastball. What makes a pitch deceptive is not something easily quantifiable. A barely significant correlation exists between velocity and strikeout rate for changeups.

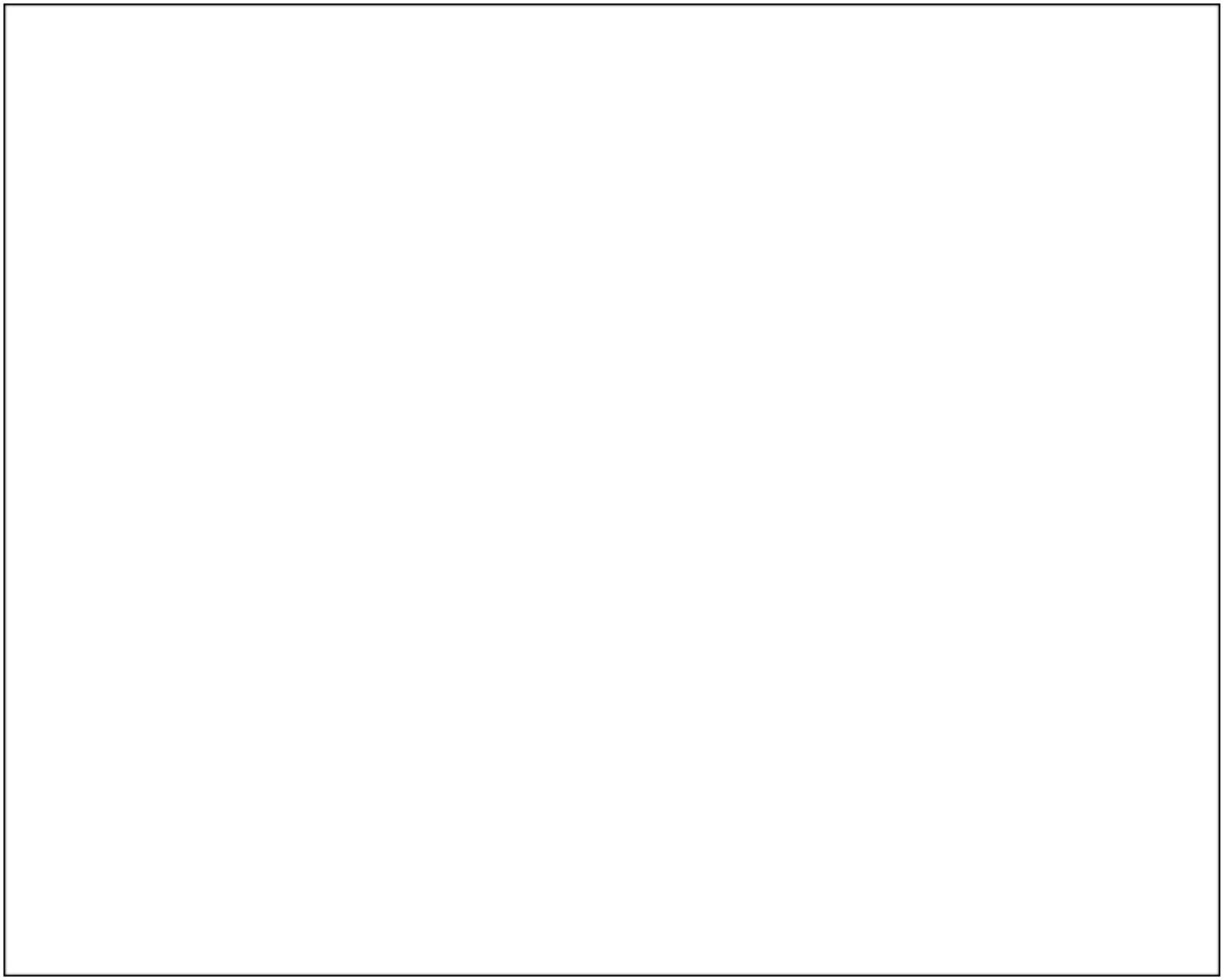
I investigated the possibility of a relation between the difference in the average velocities and strikeout rate because a batter who is expecting a fastball but gets a changeup may have their timing thrown off by the larger velocity difference between the two pitches. And, indeed, least squares regression gave a slope of 0.511 and a p value of . Here is where we get to the case of Dylan Cease and his 18.5-mph differential. Every other pitcher fell within the range [2.7, 12.4]. Removing Dylan Cease caused the slope to become 0.145 with p value of 0.398, not even significant enough to reject the null hypothesis. I still suspect the velocity difference matters, but clearly not in a way that is consistent for every pitcher.

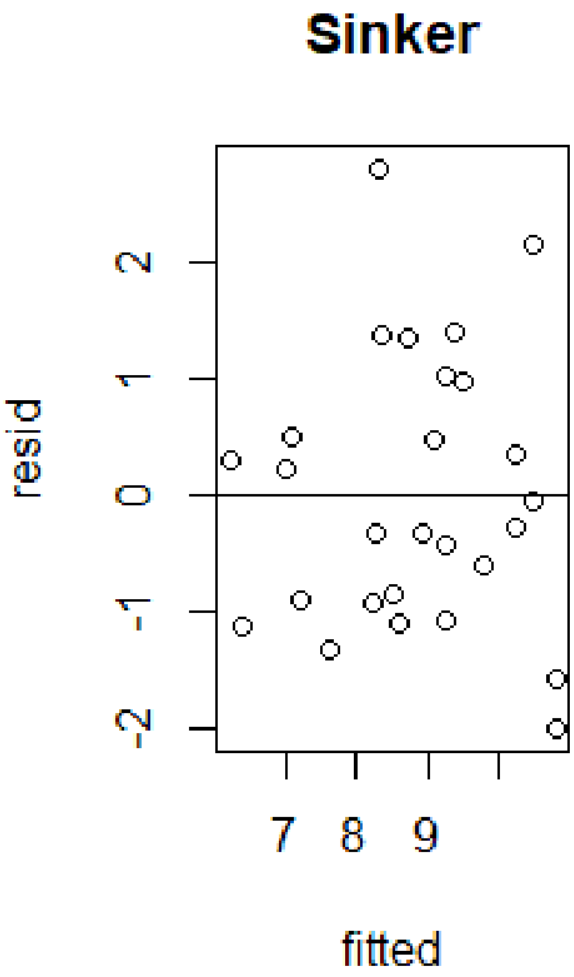
Assessing the assumptions

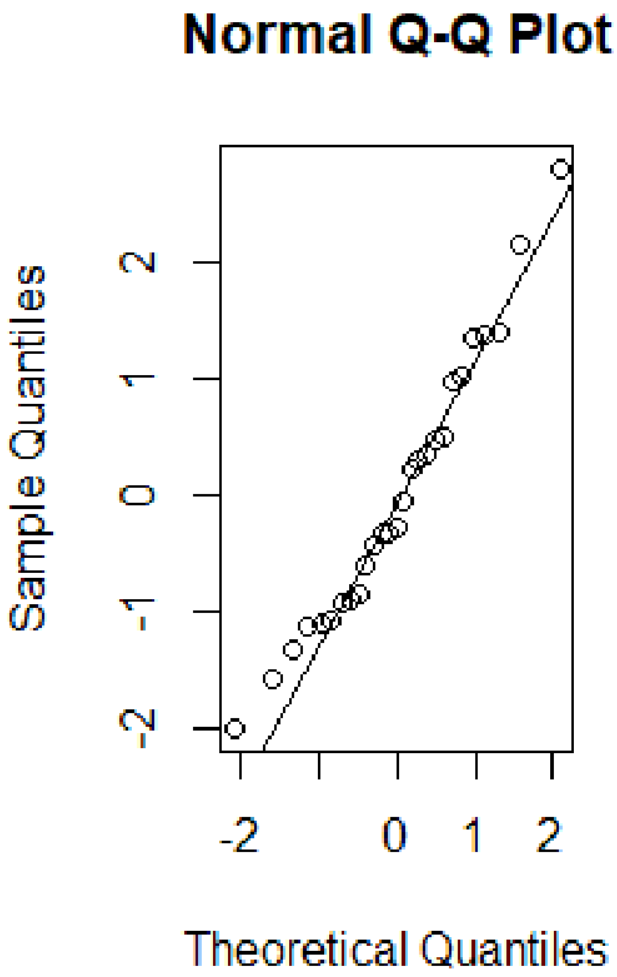


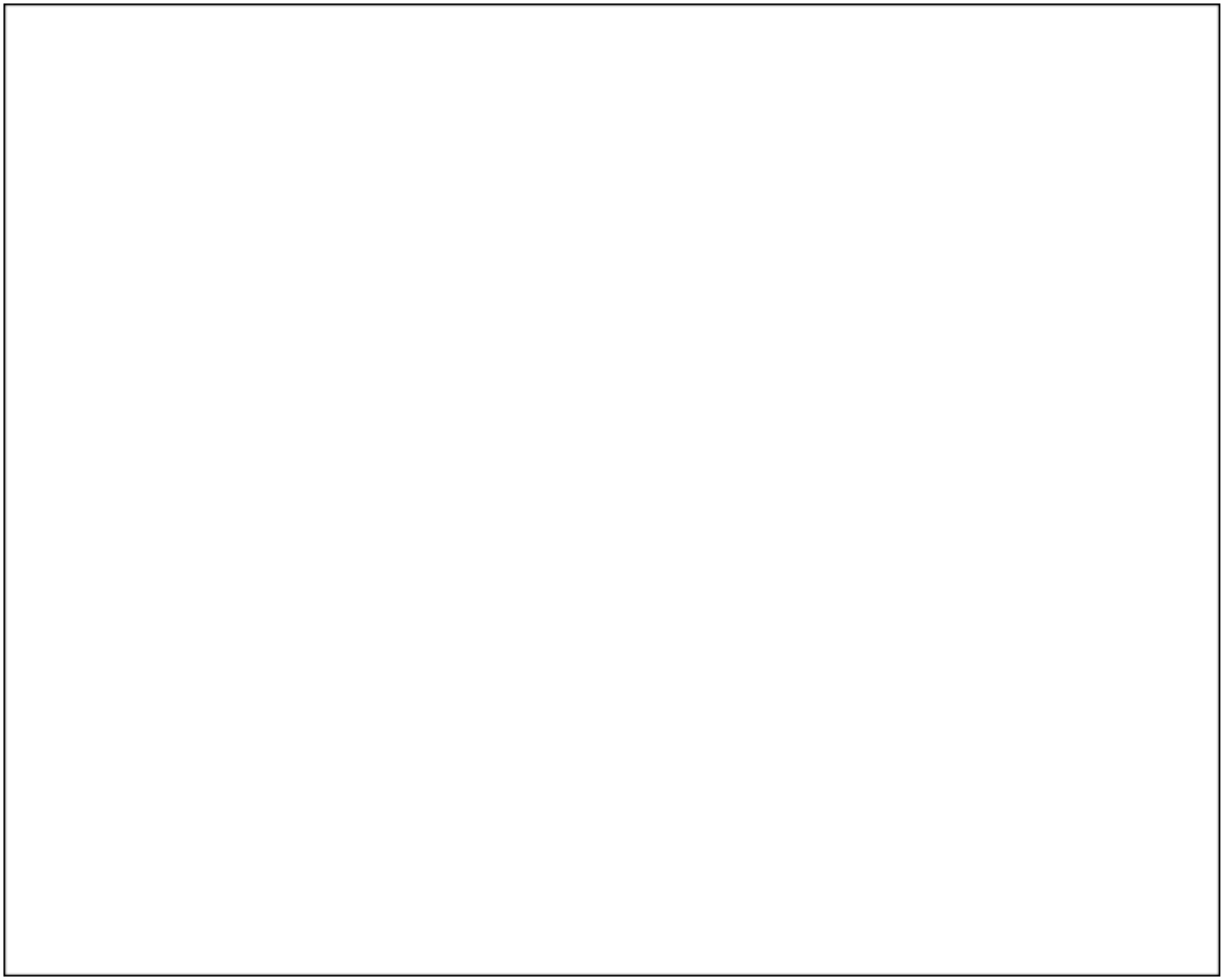

Fastball
resid
fitted
6
8
10
-2
-1
0
1
2

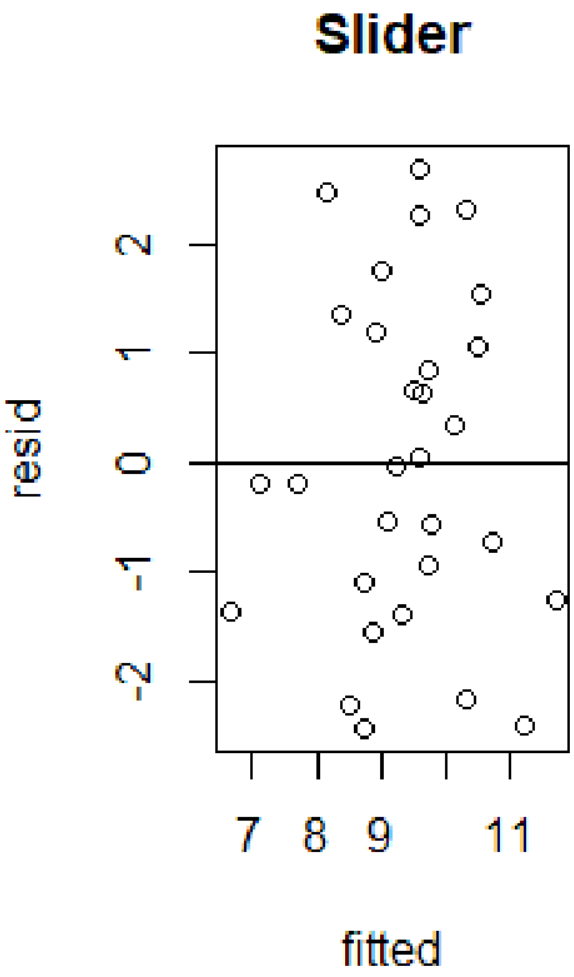

Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-2
0
1
2

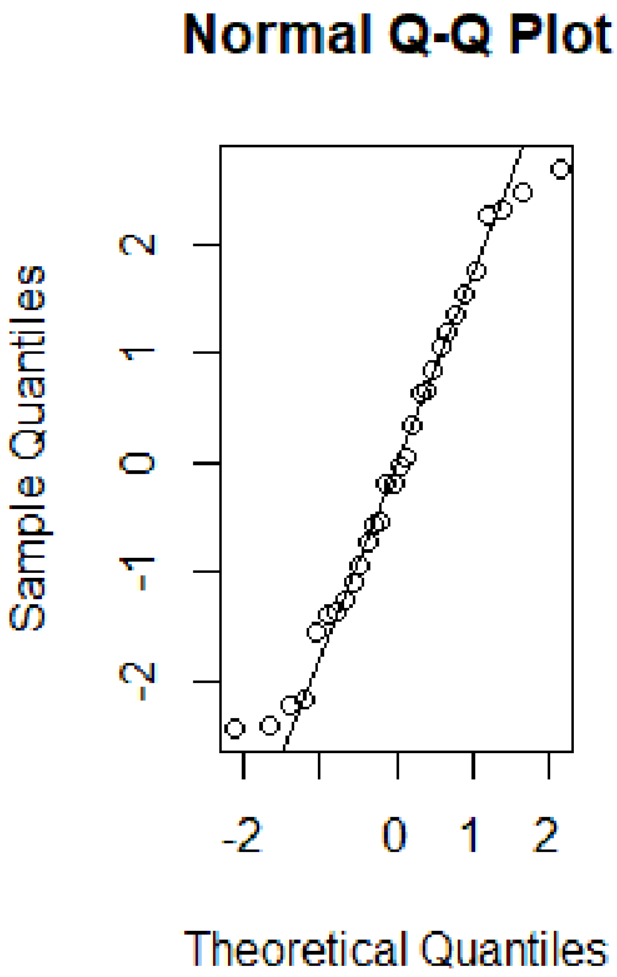


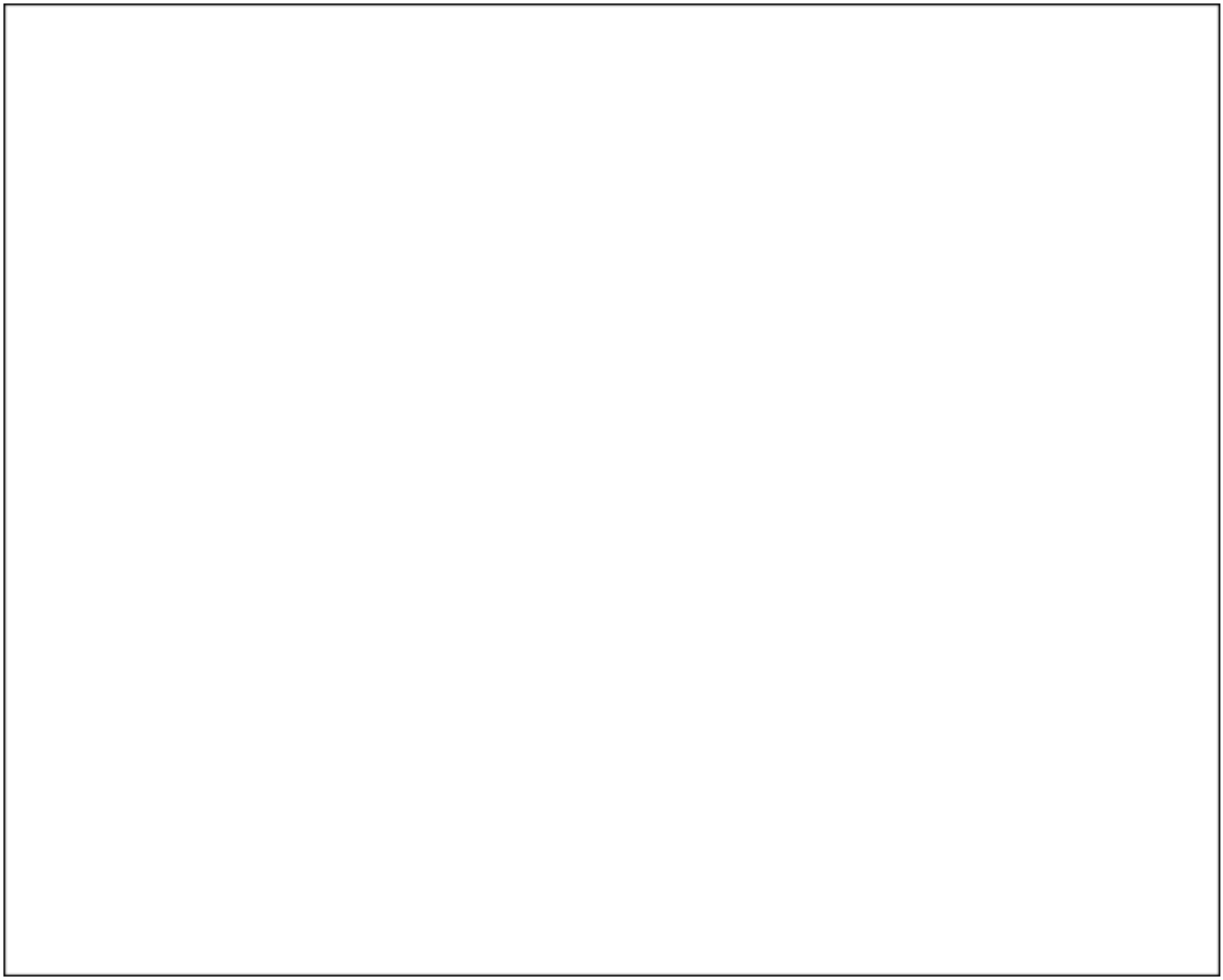

Sinker
resid
fitted
-2
-1
0
1
2
7
8
9

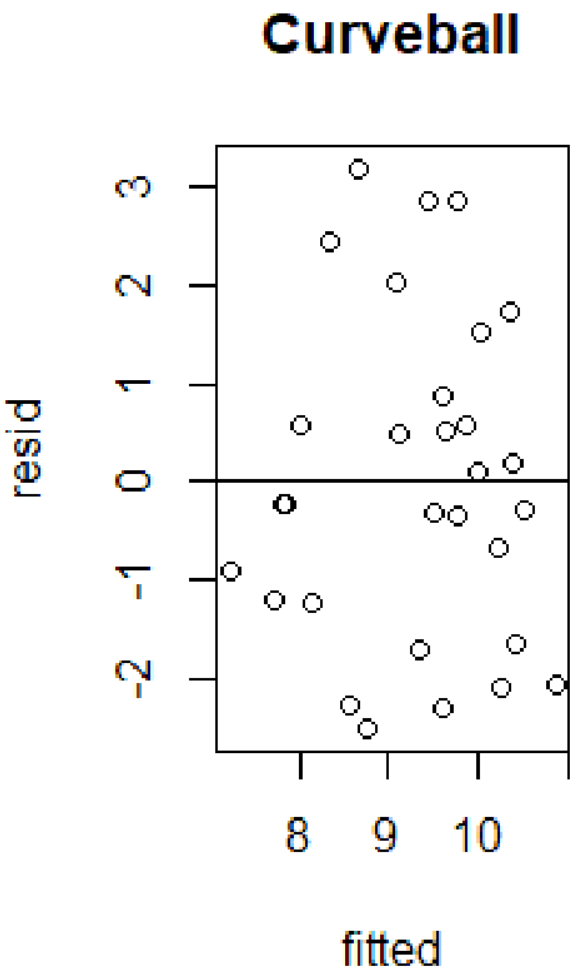

Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-2
-1
0
1
2

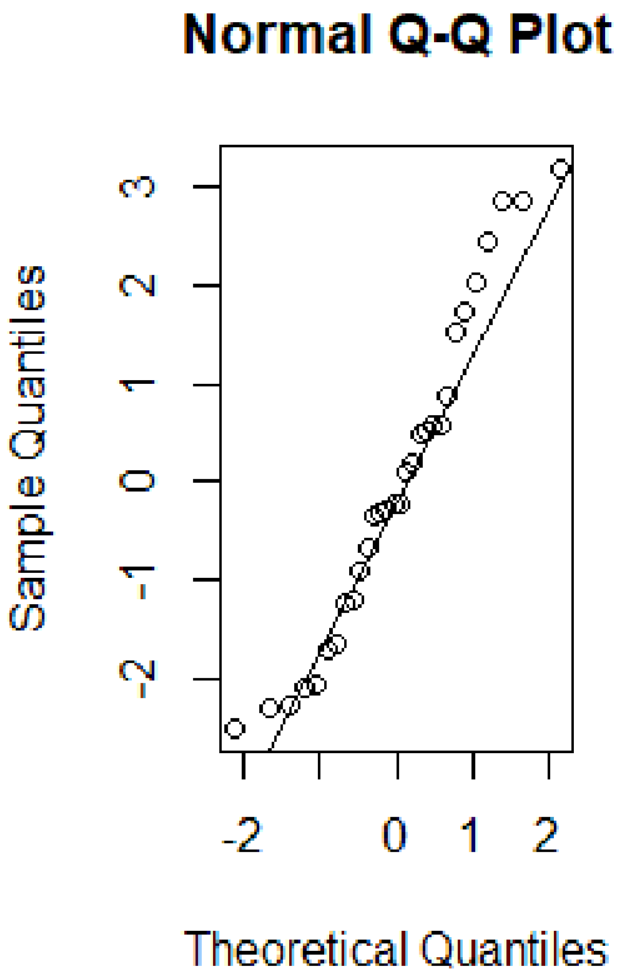


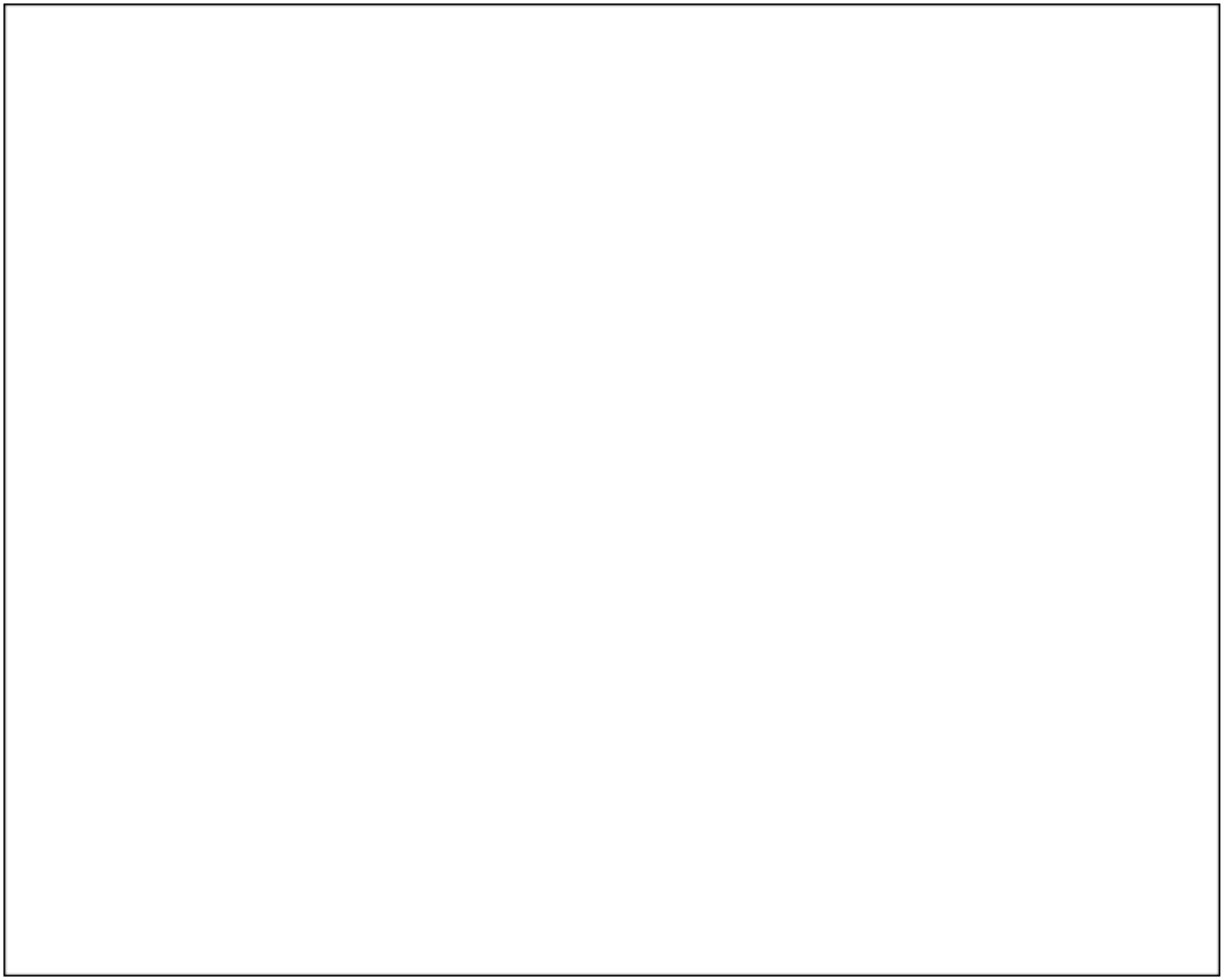

Slider
resid
fitted
-2
-1
0
1
2
7
8
9
11

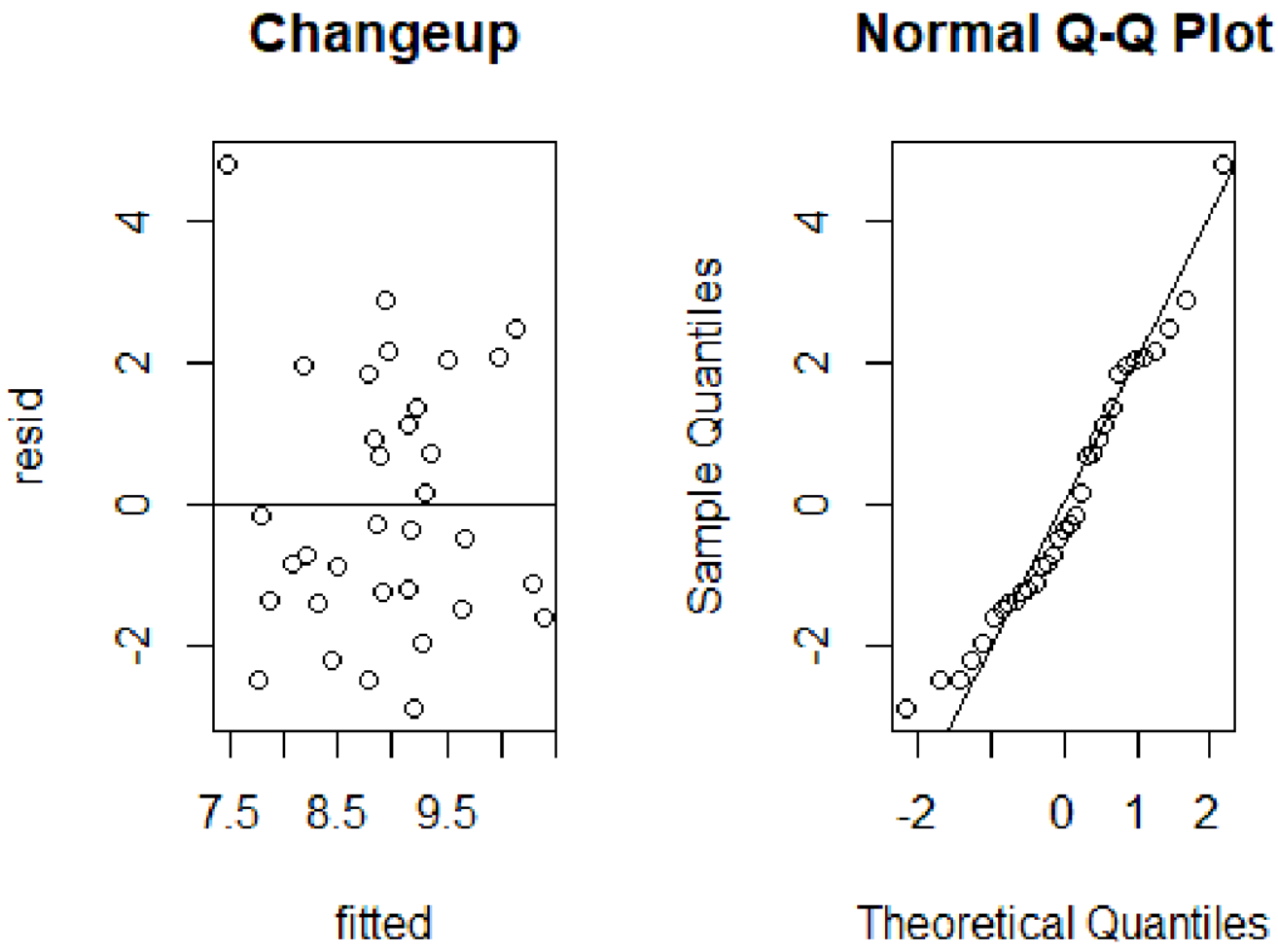

Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-2
-1
0
1
2
-2
0
1
2




Curveball
resid
fitted
-2
-1
0
1
2
3
8
9
10


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-2
-1
0
1
2
3
-2
0
1
2





All of the residuals look homoscedastic, with no clear non-linearity. The residuals do appear to have slightly "light tails," but the effect appears small. It is also possible that statcast (the high-speed camera technology) used to collect these data has a built-in bias. One factor that helps protect against this bias is the use of data from the entire league. If the cameras are consistently biased in every ballpark, then this is a problem. However, if the bias is inconsistent among the ballparks, the result is additional noise, no systemic bias. Finally, I do believe it is worth mentioning again that the sample of pitchers is not the same for each pitch, which likely increases the uncertainty of the analysis slightly.