**Name:**

**Classwork 9**

**Home Court Advantage?**

We’re going to look at a dataframe called **MiamiHeat**. Here are the variables in this dataframe:

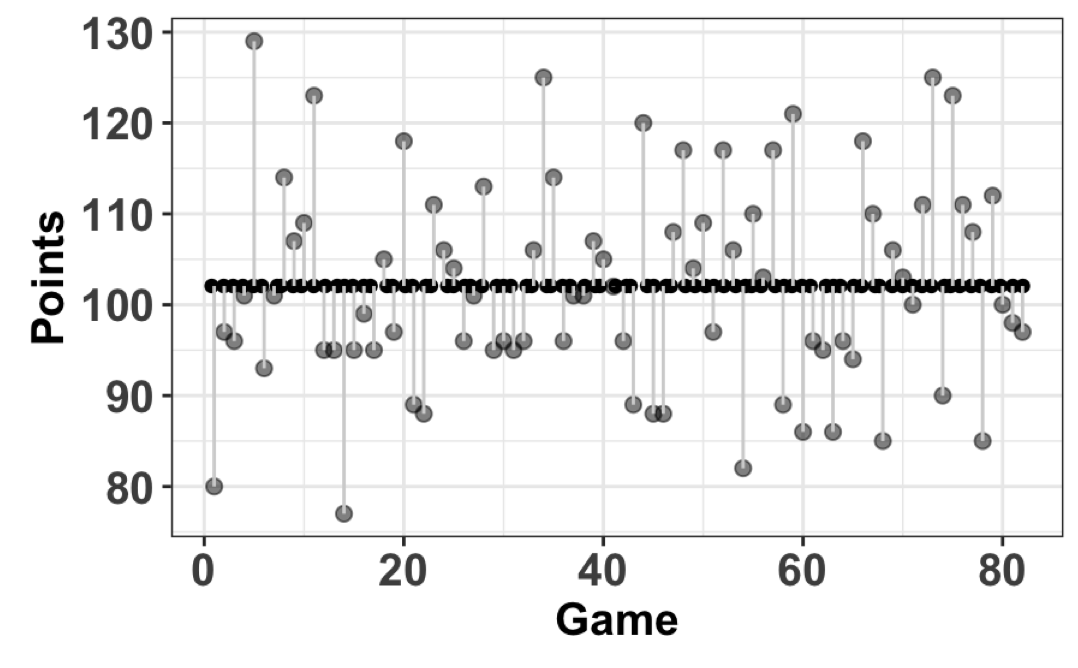
A data frame with 82 observations on the following 33 variables.

* MDY Date the game was played as a date object
* Date Date the game was played as a character string
* Location Away or Home
* Opp Opponent tream
* Win Game result: L or W
* FG Field goals made
* FGA Field goals attempted
* FG3 Three-point field goals made
* FG3A Three-point field goals attempted
* FT Free throws made
* FTA Free throws attempted
* Rebounds Total rebounds
* OffReb Offensive rebounds
* Assists Number of assists
* Steals Number of steals
* Blocks Number of shots blocked
* Trunovers Number of turnovers
* Fouls Number of fouls
* Points Number of points scored
* OppFG Opponet's field goals made
* OppFGA Opponent's Field goals attempted
* OppFG3 Opponent's Three-point field goals made
* OppFG3A Opponent's Three-point field goals attempted
* OppFT Opponent's Free throws made
* OppFTA Opponent's Free throws attempted
* OppOffReb Opponent's Total rebounds
* OppRebounds Opponent's Offensive rebounds
* OppAssists Opponent's assists
* OppSteals Opponent's steals
* OppBlocks Opponent's shots blocked
* OppTurnovers Opponent's turnovers
* OppFouls Opponent's fouls
* OppPoints Opponent's points scored

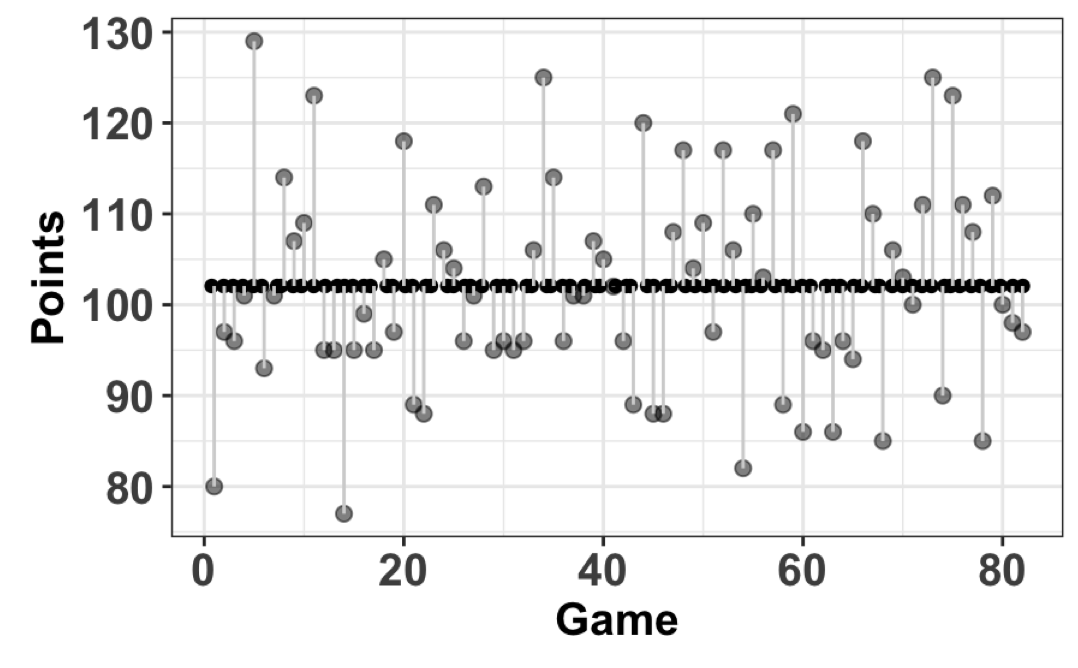
1. What are the cases in this data frame?
2. Do you think there is a home court advantage? Why or why not?
3. If there is a home court advantage, knowing whether the game was played away or home might help us predict the number of points. Write this possible DGP as a word equation.
4. One of the faceted histograms below is the empirical sample. Can you tell which one it is? What are you looking for that is different in the empirical sample?

|  |  |  |
| --- | --- | --- |
| 1. | 2. | 3. |
| 4. | 5. | 6. |
| 7. | 8. | 9. |

1. Make a visualization of the empirical sample. Now can you figure out which is the empirical data in the faceted histograms above? (Just a note: The bin sizes are different from the default.)
2. Do you think the likelihood of getting a pattern of data like the empirical sample from a random process is high? Low? Medium? Explain your reasoning.
3. Let’s go back to our “whole thing” diagram.
4. If the likelihood of getting a sample like the empirical sample from a random process is high, what theory of the DGP would that support? What theory would it rule out? *Explain your reasoning*.
5. If I knew that the next game was being played at home, what should I use to predict the points that will be scored by **MiamiHeat**? Should I use the mean of just the home games? Will that help me make a better prediction? Or would that be just as good/bad as using the mean of all the games? Are there any reasons why we might want to use the mean of all the games?
6. Here I’ve had R draw all the residuals from the mean of all the games for you. Which looked like the “average residual”? How big is the average residual (just by eyeballing from this visualization)?

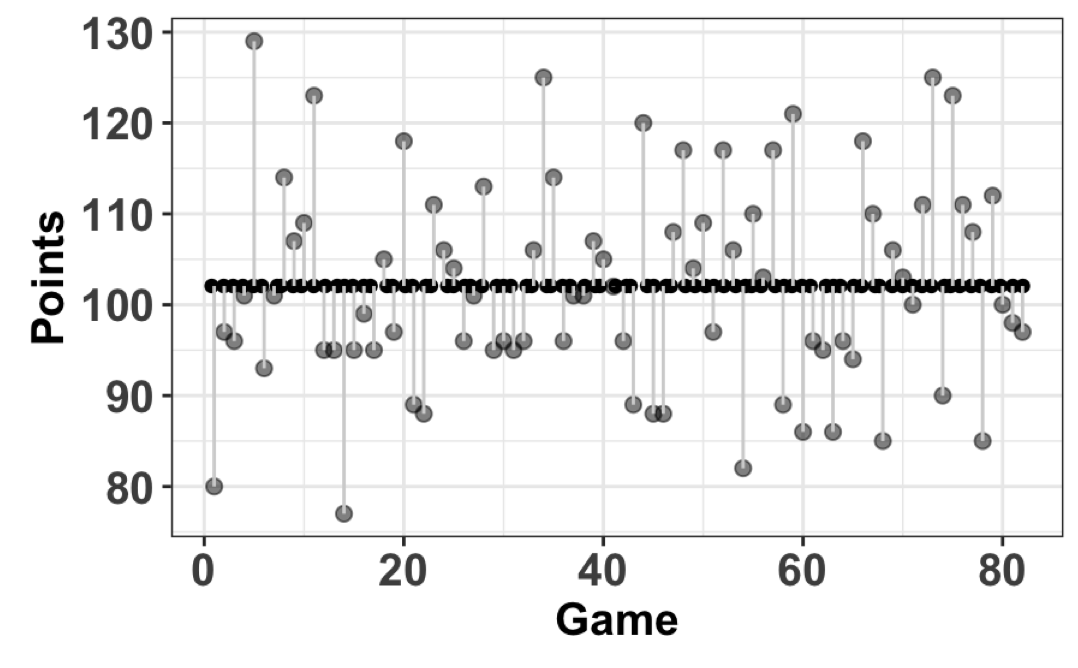


1. Which statistic calculates the average residual for you? Use R to calculate it and see if it is similar to your estimate from the previous question.
2. This is just a copy of the same plot. What would the variance look like on this plot? What unit is variance in? Does the number for variance seem about right?

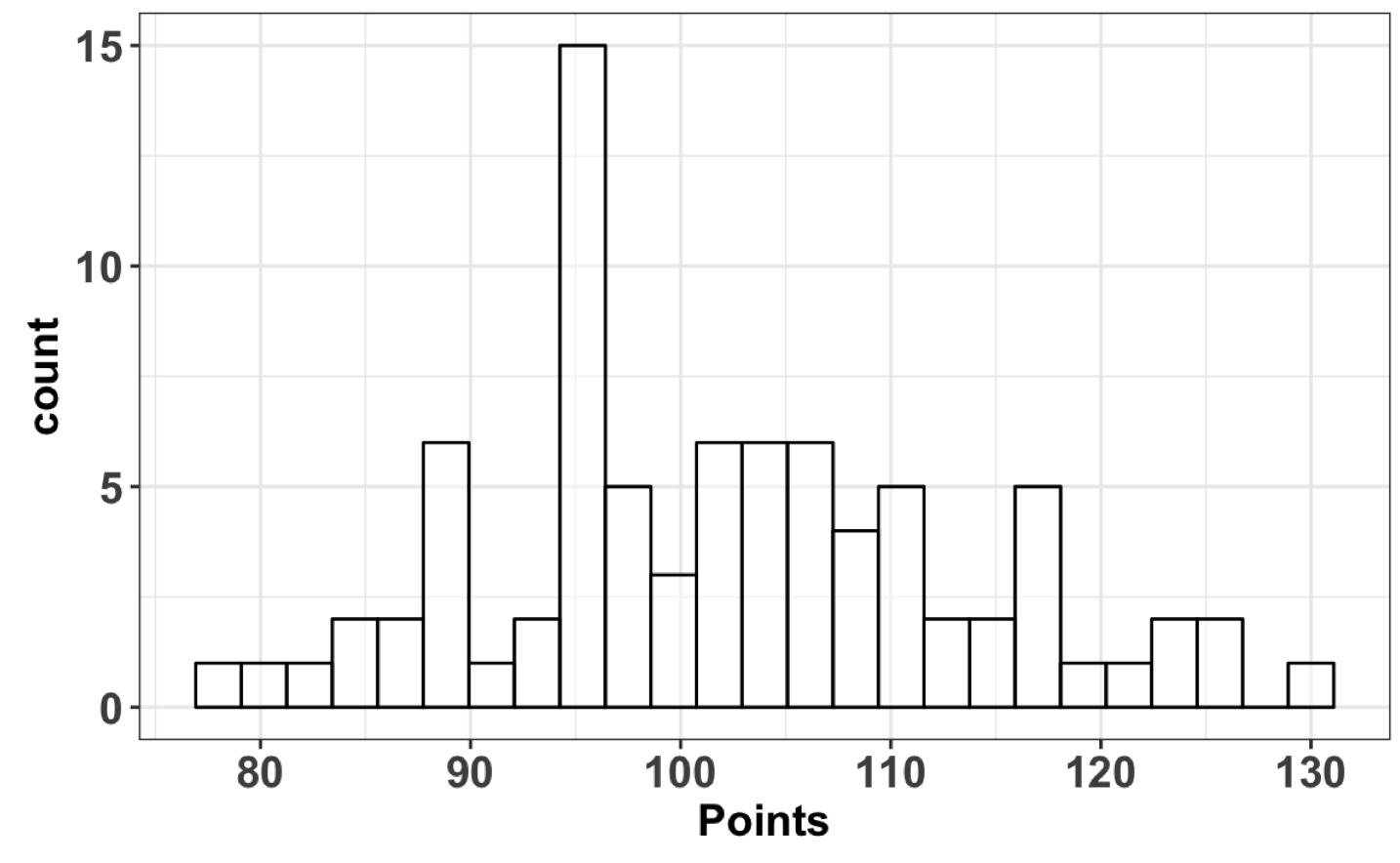
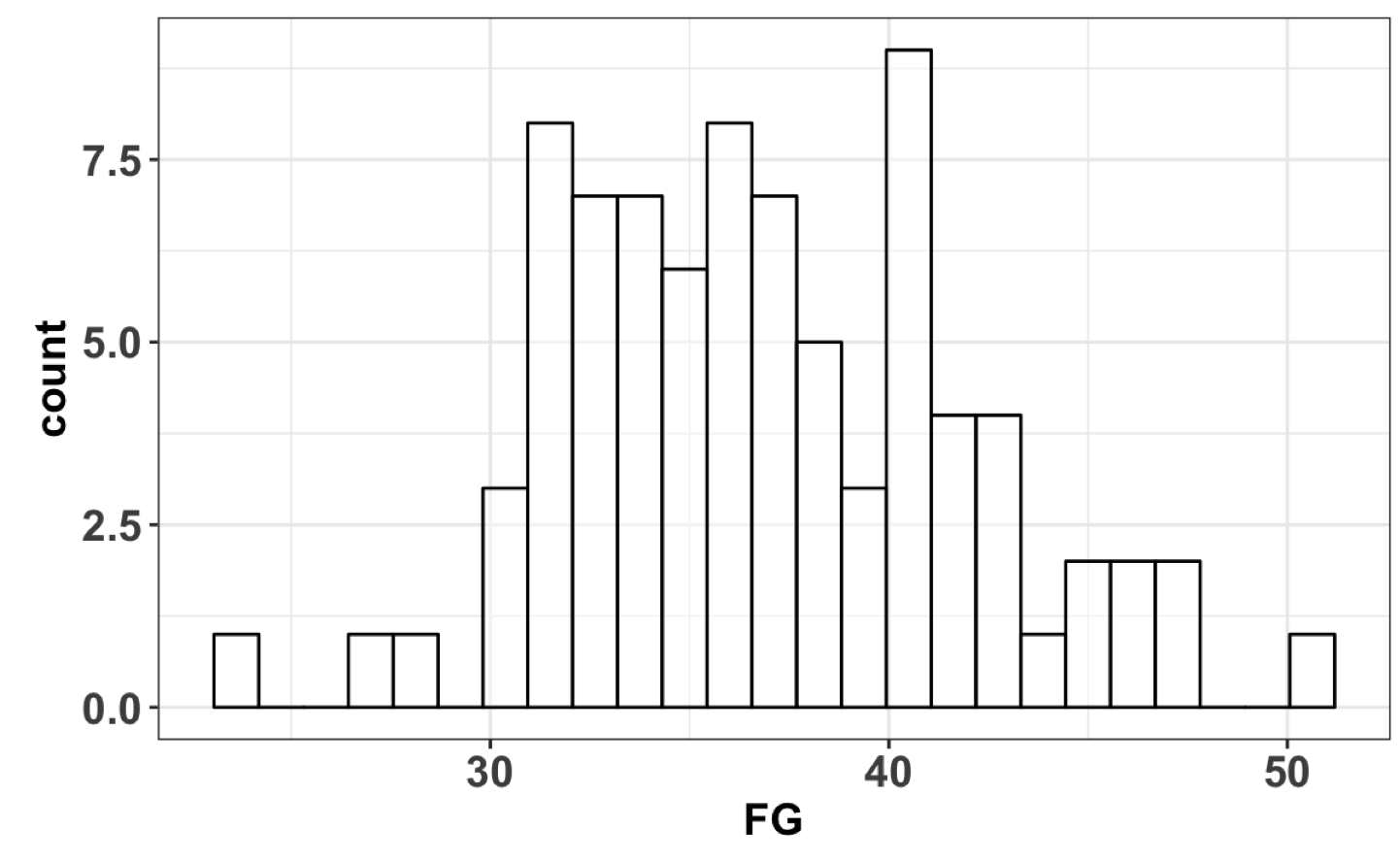


1. Here is the formula for variance. Label on the plot above, which parts are the ? ? ? ?

14. This is just a copy of the same plot. What would the sum of squares look like on this plot?



1. This is the same data (Points) on a histogram. How might we draw the standard deviation here? This is the variable FG (field goals) on a histogram. Try to draw the standard deviation first, estimate it, and then use R to see if your estimate was pretty close.

1. Describe your strategy for estimating standard deviation on a histogram here.