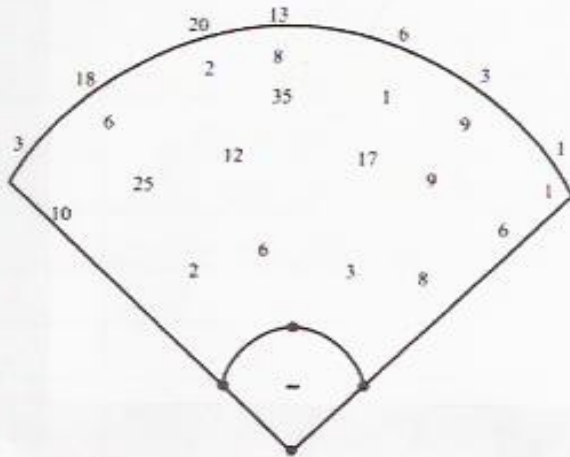


7. The figure below shows the numbers and locations of homeruns (indicated by numbers outside the park) and other hits making it to the outfield for Mark McGuire in 1999. The table to the right has been constructed using this information, categorizing hits by left field, center field, and right field.



Location	Type of Hit		Total
	Homeruns	Outfield	
Left	21	43	64
Center	39	84	123
Right	4	33	37
Total	64	160	224

Is it reasonable to say that the variables LOCATION & TYPE OF HIT are independent? Defend your answer with appropriate graphs, tables, and/or calculations.

ANSWER

In order to check for 2 categorical variables to be independent, we need to see if conditional distributions "match" (for the most part) each other as well as the respective marginal distributions. If they match fairly well, then the 2 variables may be determined as independent -- if the distributions don't match well enough, then the variables are

deemed as not independent. We do not call them dependent in this later case, because calling 2 variables "dependent" implies a more structured form of being "not independent" than could possibly be gleaned from a 2-way-table. So we phrase the result in clumsy English, as not independent.

Any way, we must compute marginals & conditionals of, lets say, LOCATION by HIT.

1st we find marginal columns on the table, including total-total in lower right corner

column marginals

	Type of Hit		
	Homeruns	Outfield	Total
Location			
Left	21	43	64
Center	39	84	123
Right	4	33	37
Total	64	160	224

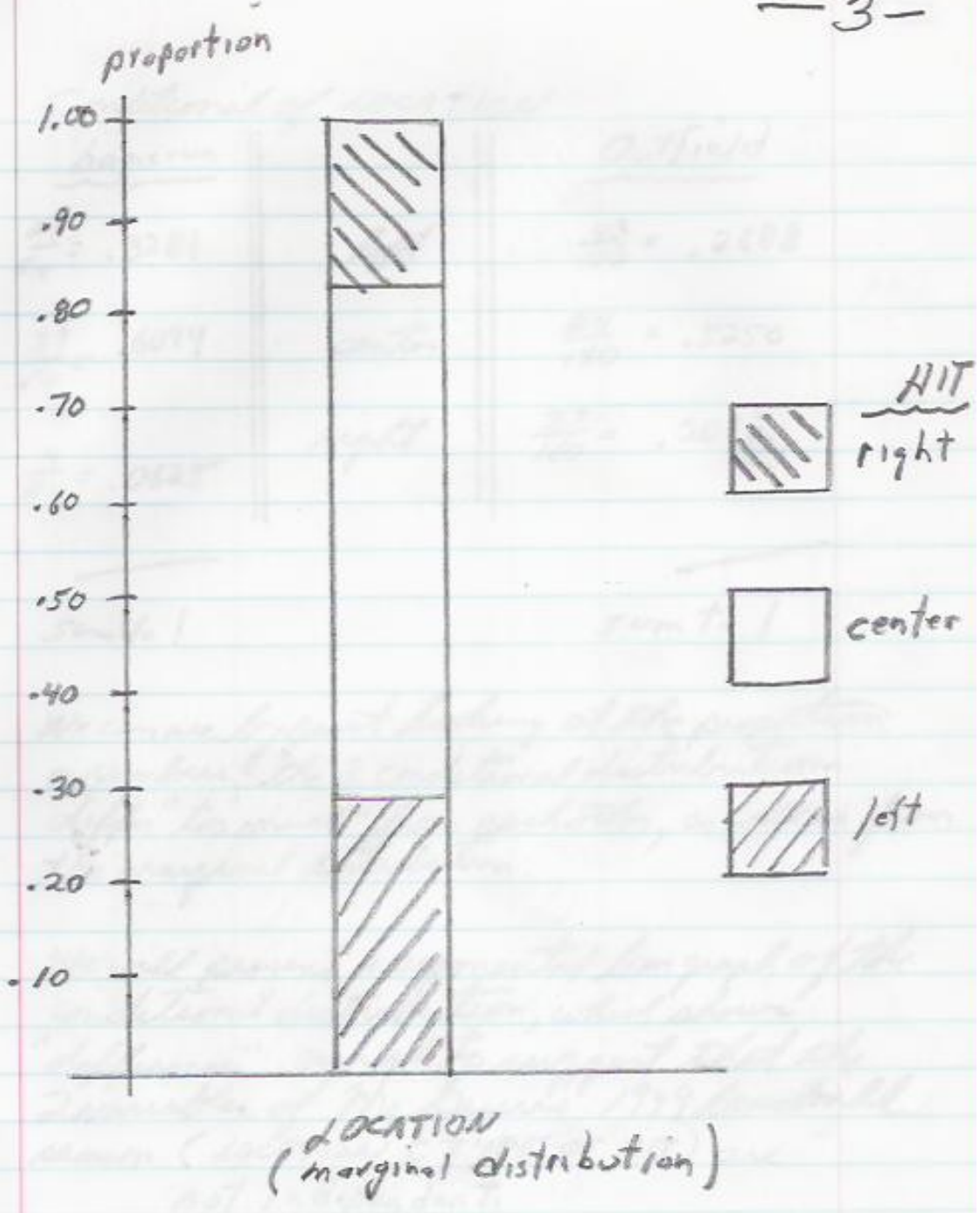
row marginal

Total-total

Marginal of LOCATION

left $64/224 \approx .2857$
 center $123/224 \approx .5491$
 right $37/224 \approx .1652$
 sum to 1

the segmented paragraph of this is shown next



Conditional of LOCATION

<u>home run</u>		<u>Outfield</u>
$\frac{21}{64} = .3281$	left	$\frac{43}{160} = .2688$
$\frac{39}{64} = .6094$	center	$\frac{84}{160} = .5250$
$\frac{4}{64} = .0625$	right	$\frac{33}{160} = .2062$
sum to 1		sum to 1

We can see by just looking at the proportion numbers, the 2 conditional distributions differ "too much" from each other, as well as from the marginal distribution.

We will present a segmented bar graph of the conditional distribution, which shows "difference" enough to suggest that the 2 variables of Mc Geuer's 1999 baseball season (LOCATION & TYPE OF HIT) are not independent.

Conditional distribution shown next.

