

240 PROJECT REPORT

Pelin ZEYTUN

214051791

My question is “Is there any effect of the scores of the players for draft selection?” my data files are basketball_players.csv and basketball_draft.csv. firstly, i read data and i found their columns. After index the main columns, I grouped them with the draft selection and found the summation. Then players join draft selection. Then I used .describe() function for calculation mean, standard deviation, min. ,max. %25,%50%75 values.

```
In [60]: players = players.set_index(['playerID', 'tmID'])
```

```
In [61]: draft_Selection = draft.groupby(['playerID', 'tmID'])['draftSelection'].sum()
```

```
In [62]: players = players.join(draft_Selection)
```

```
In [90]:
```

```
In [91]: players.describe()
```

Out[91]:

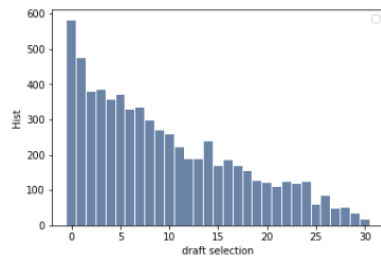
	year	stint	GP	GS	minutes	points	oRebounds	dRebounds	rebounds	assists	...	Po
count	23751.000000	23751.000000	23751.000000	23751.000000	23751.000000	23751.000000	23751.000000	23751.000000	23751.000000	23751.000000	...	2375
mean	1982.914235	1.034651	47.964338	0.935624	1097.296661	492.130689	50.382594	112.825271	209.064208	107.060376	...	
std	19.996255	0.440119	27.818925	6.935694	993.764138	503.053318	69.506573	148.146626	238.276611	135.377884	...	
min	1937.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	
25%	1970.000000	1.000000	22.000000	0.000000	137.000000	81.000000	0.000000	0.000000	23.000000	11.000000	...	
50%	1986.000000	1.000000	53.000000	0.000000	866.000000	329.000000	20.000000	52.000000	133.000000	58.000000	...	
75%	2000.000000	1.000000	75.000000	0.000000	1898.000000	758.500000	75.000000	173.000000	308.000000	152.000000	...	
max	2011.000000	5.000000	84.000000	82.000000	3882.000000	4029.000000	895.000000	1111.000000	2149.000000	1164.000000	...	9

8 rows × 39 columns

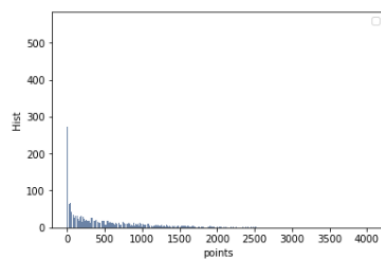


Then i found histogram, pmf, cdf for points and draft selection.

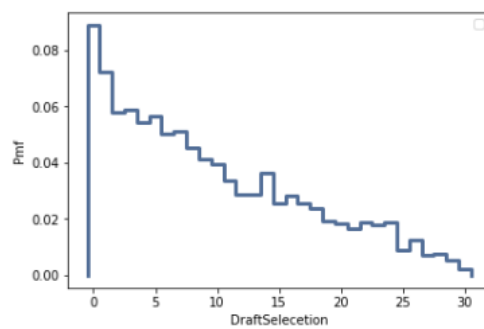
```
In [119]: histds= thinkstats2.Hist(players.draftSelection)
          thinkplot.Hist(histds)
          thinkplot.Config(xlabel= 'draft selection', ylabel='Hist')
```



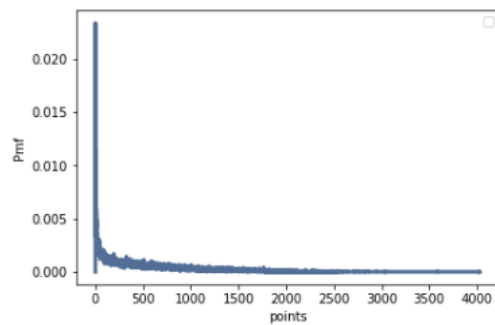
```
In [117]: histpoint= thinkstats2.Hist(players.points)
          thinkplot.Hist(histpoint)
          thinkplot.Config(xlabel= 'points', ylabel='Hist')
```



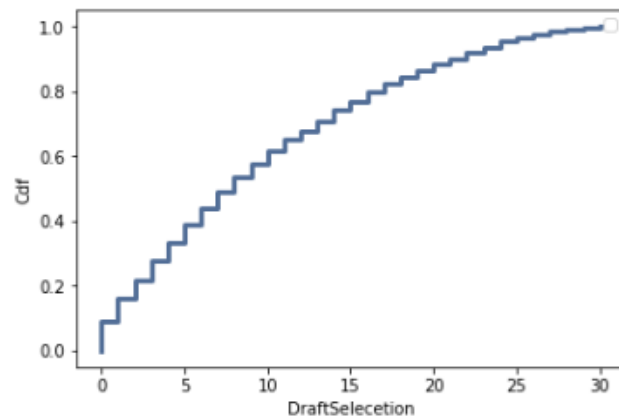
```
In [118]: pmfds = thinkstats2.Pmf(players.draftSelection)
          thinkplot.Pmf(pmfds)
          thinkplot.Config(xlabel='DraftSelecection', ylabel='Pmf')
```



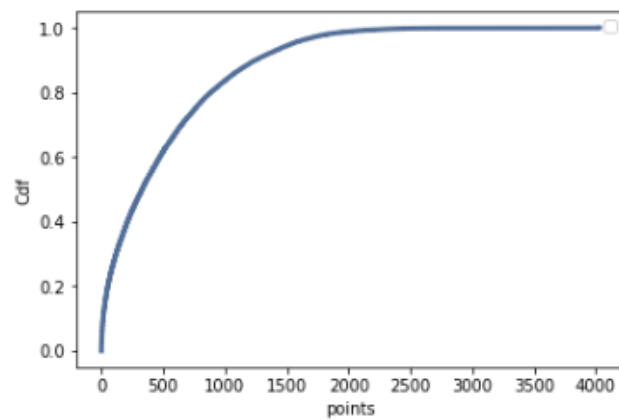
```
In [120]: pmfpoint = thinkstats2.Pmf(players.points)
          thinkplot.Pmf(pmfpoint)
          thinkplot.Config(xlabel='points', ylabel='Pmf')
```



```
In [121]: cdfds = thinkstats2.Cdf(players.draftSelection)
thinkplot.Cdf(cdfds)
thinkplot.Config(xlabel='DraftSelecetion', ylabel='Cdf')
```



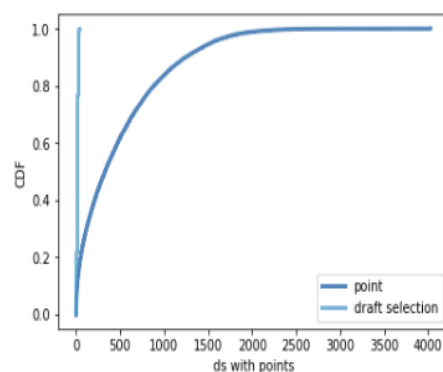
```
In [122]: cdfpoint = thinkstats2.Cdf(players.points)
thinkplot.Cdf(cdfpoint)
thinkplot.Config(xlabel='points', ylabel='Cdf')
```



In this part, i found cdf points and draft selection together. And, i found their p value.

```
48]: pointcdf = thinkstats2.Cdf(players.points, label='point')
     dscdf = thinkstats2.Cdf(players.draftSelection, label='draft selection')

     thinkplot.PrePlot(2)
     thinkplot.Cdfs([pointcdf, dscdf])
     thinkplot.Config(xlabel='ds with points', ylabel='CDF')
```



```
51]: pvalue = pointcdf[0]
     pvalue
```

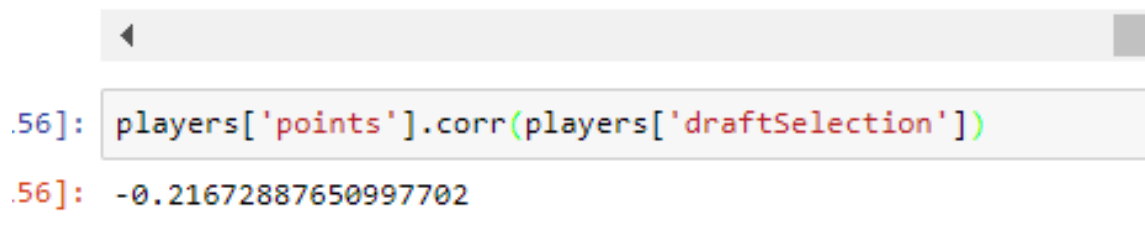
```
51]: 0.02336743716054061
```

```
52]: pvalue = dscdf[0]
     pvalue
```

```
52]: 0.08889228886168911
```

If $p\text{-value} \leq 0.05$ you reject the null hypothesis. If $p\text{-value} > 0.05$ indicates a null hypothesis. Draft selection's p value more than 0.05. so my project's hypotesis is a null hypotesis. As a result, there is an effect of the scores of the players for draft selection.

Then i found correlation. It ranges from -1.0 to +1.0. When r is 0, there is no relationship between the variables. When r is positive, one variable gets larger the other gets larger. When r is negative, one gets larger, the other gets smaller. In my project, r is negative. So, it is inverse correlation.

A screenshot of a Jupyter Notebook interface. At the top, there is a grey header bar with a left-pointing arrow on the left and a small grey square on the right. Below this, a code cell is shown with a light grey background. The prompt ".56]:" is in blue. The code "players['points'].corr(players['draftSelection'])" is in a monospaced font, with "players" in blue, "points" in red, "corr" in green, and "draftSelection" in red. Below the code, the output ".56]:" is in red, followed by the value "-0.21672887650997702" in black. A thin green horizontal line is positioned below the output.

```
.56]: players['points'].corr(players['draftSelection'])
```

```
.56]: -0.21672887650997702
```