NFL_Rushing_QBs

December 6, 2019

0.0.1 Question: Do rushing/scrambling quarter backs have shorter careers than pocket passers?

Standard math package and plotting imports:

[1]:
```
import math
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

Load in our data:

[2]:
```
df = pd.read_csv('U:\\Python\\NFL\\qb_rushing.csv')
```

Look at top 10 QBs, ordered by total rushing yards:

[3]:
```
df = df.set_index('player_rank').sort_index()
df.head(10)
```

[3]:

player_rank	player	first_year	last_year	draft	team	league	games_played	games_started	rushing_attempts
1	Michael Vick	2001	2015	1-1	TOT	NFL	143.0	115.0	873
2	Randall Cunningham	1985	2001	2-37	TOT	NFL	161.0	135.0	775
3	Cam Newton	2011	2019	1-1	CAR	NFL	125.0	124.0	934
4	Tom Matte	1961	1972	1-7 5-37AFL	CLT	NFL	142.0	91.0	1200
5	Steve Young	1985	1999	1-1	TOT	NFL	169.0	143.0	722

```
6              NFL         123.0          123.0              698
7              NFL         246.0          239.0              675
8              NFL         161.0          153.0              669
9              NFL          99.0           76.0              687
10             NFL         167.0          161.0              616
```

```
             total_yards  yards_per_attempt  rushing_td  yards_per_game
player_rank
1                   6109               7.00          36            42.7
2                   4928               6.36          35            30.6
3                   4806               5.15          58            38.4
4                   4646               3.87          45            32.7
5                   4239               5.87          43            25.1
6                   3922               5.62          19            31.9
7                   3674               5.44          32            14.9
8                   3590               5.37          37            22.3
9                   3506               5.10          23            35.4
10                  3459               5.62          29            20.7
```

[4]:
```
print('Total QBs: {}'.format(df.shape[0]))
```

```
Total QBs: 985
```

Cut down list to only QBs who were in the league for at least 1 year, started 10+ games, since the 1966-67 season (first Super Bowl):

[5]:
```
df = df[(df['first_year'] < df['last_year']) &
        (df['games_started'] >= 10) &
        (df['first_year'] > 1966)]
print('QBs who played >1 year, started 10+ games, since 1966-67 season: {}'\
      .format(df.shape[0]))
```

```
QBs who played >1 year, started 10+ games, since 1966-67 season: 320
```

Look at summary statistics for games played, games started, rushing attempts and yards per attempt for these 320 QBs. Note the average career of these QBs was ~88 games and 3.5 rushing yards per attempt:

[6]:
```
categories = ['games_played', 'games_started', 'rushing_attempts',
              'yards_per_attempt']
print(df[categories].describe())
quantiles = [0.25, 0.5, 0.75]
gp_quant = np.quantile(df['games_played'], quantiles)
ra_quant = np.quantile(df['rushing_attempts'], quantiles)
ypa_quant = np.nanquantile(df['yards_per_attempt'], quantiles)
```

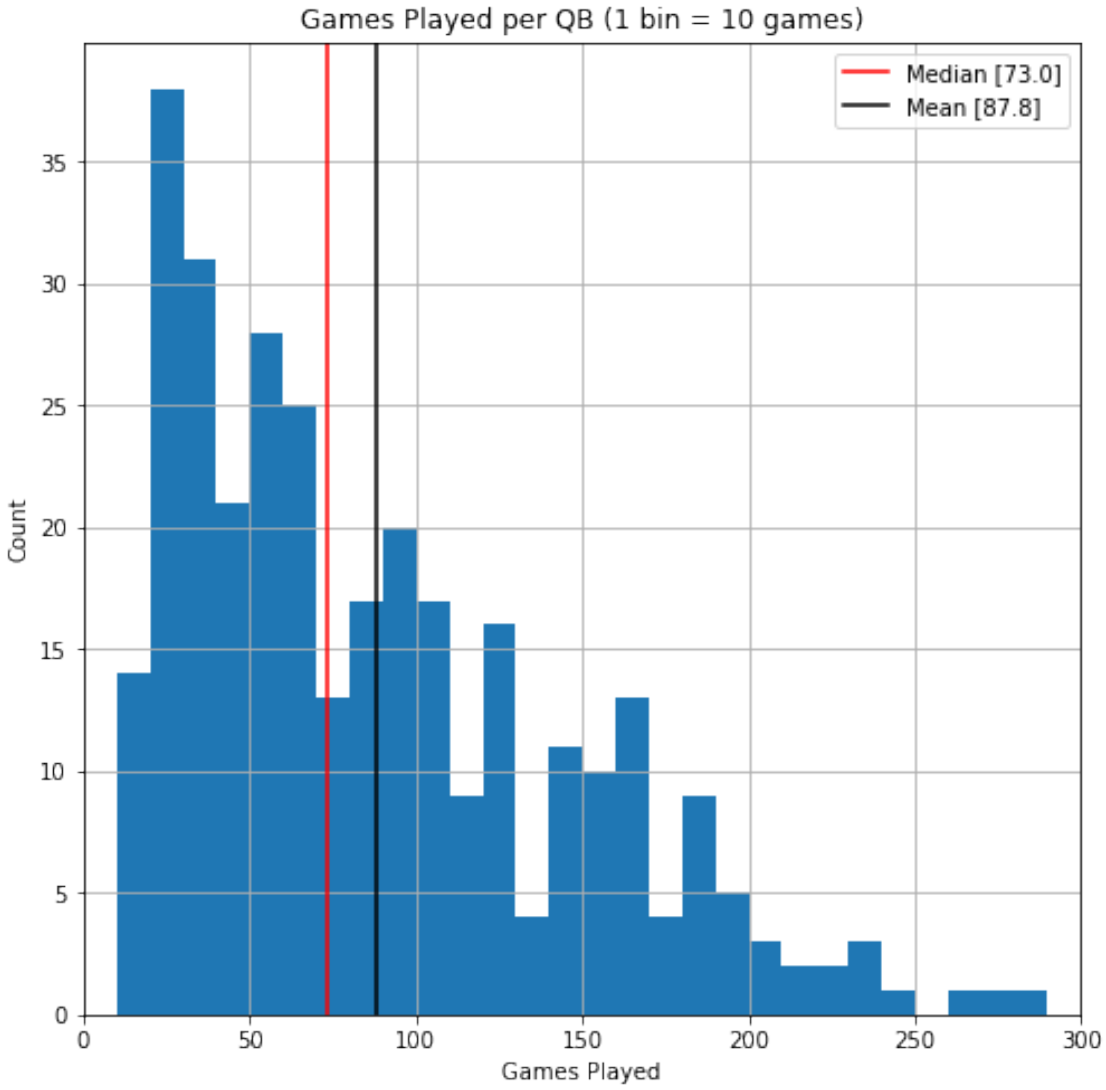
```
       games_played  games_started  rushing_attempts  yards_per_attempt
count    320.000000     320.000000        320.000000         320.000000
mean      87.793750      65.984375        183.625000           3.475625
```

std	59.252194	58.508658	165.786555	1.560323
min	12.000000	10.000000	10.000000	-0.030000
25%	38.000000	20.000000	59.750000	2.330000
50%	73.000000	44.000000	129.500000	3.470000
75%	125.000000	97.250000	248.000000	4.595000
max	302.000000	298.000000	934.000000	7.250000

Plot a histogram of games played per QB. Each bin in histogram represents 10 games:

[7]:
```
games_max = int(max(df['games_played']))
games_mean = round(df['games_played'].mean(), 1)
games_median = df['games_played'].median()

# Histogram of games played
plt.figure(figsize = (8, 8))
plt.hist(df['games_played'], [i for i in range(0, games_max, 10)])
plt.axvline(games_median,
            color = 'red',
            label = 'Median [{}]'.format(games_median))
plt.axvline(games_mean,
            color = 'black',
            label = 'Mean [{}]'.format(games_mean))
plt.title('Games Played per QB (1 bin = 10 games)')
plt.xlabel('Games Played')
plt.ylabel('Count')
plt.xlim(0, 10 * math.floor(games_max/10))
plt.grid(which = 'major')
plt.legend(loc = 'best')
plt.show()
```

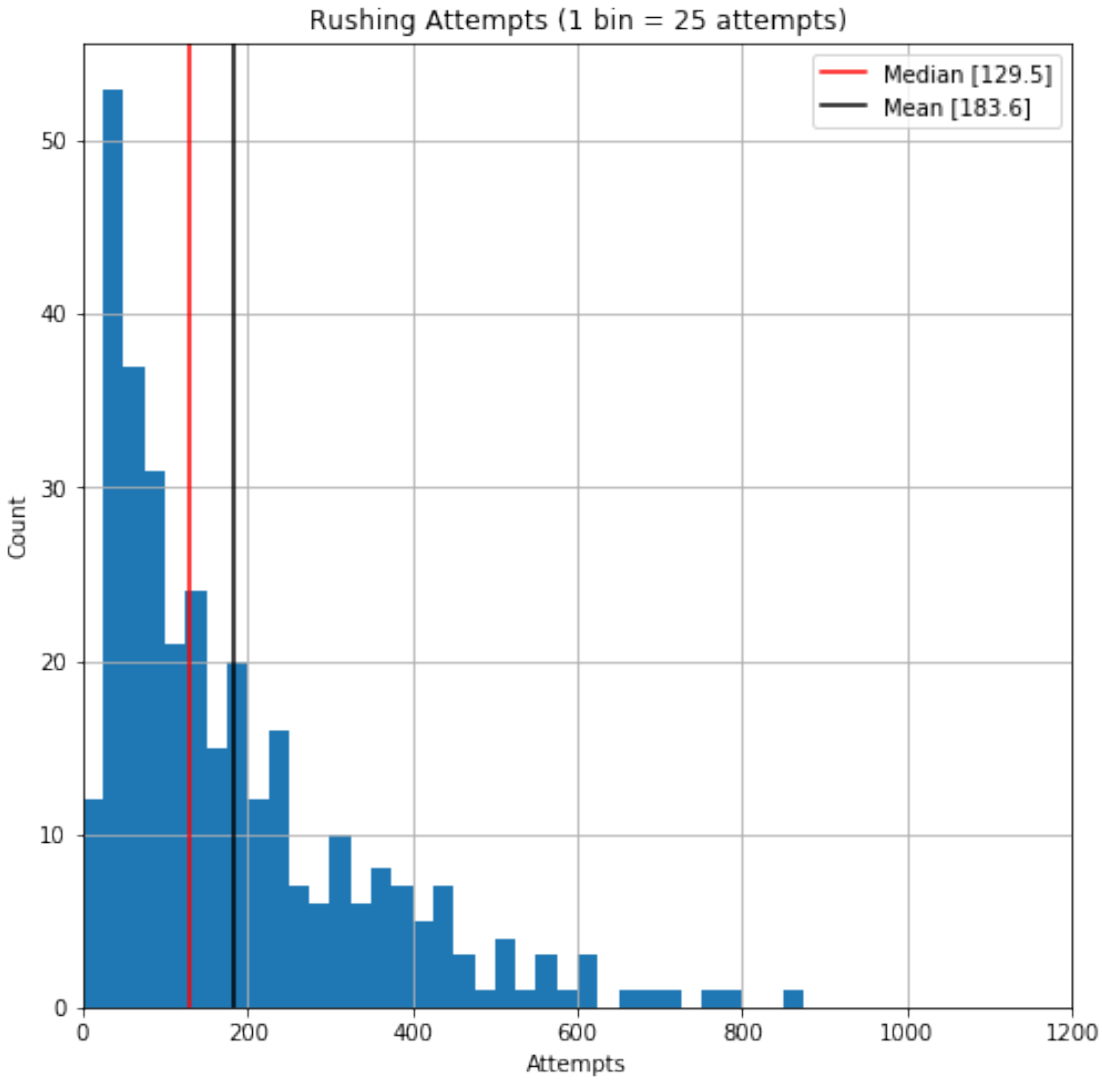


Histogram of rushing attempts per QB. Each bin represents 25 rushing attempts:

[8]:
```
attempts_max = int(max(df['rushing_attempts']))
attempts_mean = round(df['rushing_attempts'].mean(), 1)
attempts_median = round(df['rushing_attempts'].median(), 1)

# Histogram of rushing attempts
plt.figure(figsize = (8, 8))
plt.hist(df['rushing_attempts'], [i for i in range(0, attempts_max, 25)])
plt.axvline(attempts_median,
            color = 'red',
            label = 'Median [{}]'.format(attempts_median))
plt.axvline(attempts_mean,
            color = 'black',
```

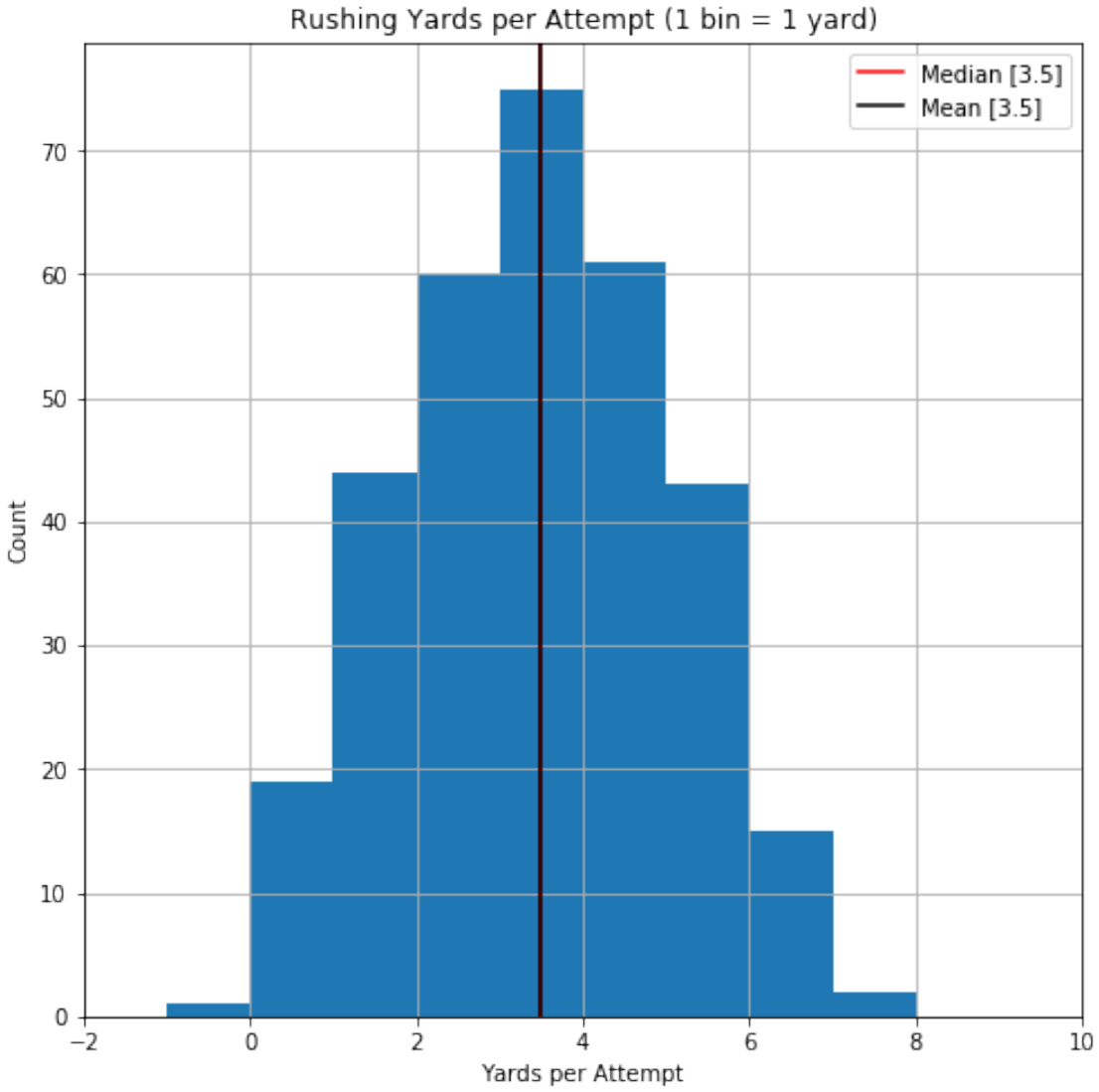
```
            label = 'Mean [{}]'.format(attempts_mean))
plt.xlim(0, 1200)
plt.title('Rushing Attempts (1 bin = 25 attempts)')
plt.xlabel('Attempts')
plt.ylabel('Count')
plt.grid(which = 'major')
plt.legend(loc = 'best')
plt.show()
```



Histogram of rushing yards per attempt. Each bin represents 1 yard:

[9]:
```
yards_max = int(max(df['yards_per_attempt']))
yards_mean = round(df['yards_per_attempt'].mean(), 1)
yards_median = round(df['yards_per_attempt'].median(), 1)
```

```
# Histogram of yards per attempt
plt.figure(figsize = (8, 8))
plt.hist(df['yards_per_attempt'], bins = [i for i in range(-2, 11)])
plt.axvline(yards_median,
            color = 'red',
            label = 'Median [{}]'.format(yards_median))
plt.axvline(yards_mean,
            color = 'black',
            label = 'Mean [{}]'.format(yards_mean))
plt.xlim(-2, 10)
plt.title('Rushing Yards per Attempt (1 bin = 1 yard)')
plt.xlabel('Yards per Attempt')
plt.ylabel('Count')
plt.grid(which = 'major')
plt.legend(loc = 'best')
plt.show()
```



First, look at career length of top 25% vs bottom 25% of QBs based on rushing attempts. This is interesting, but doesn't actually give us much information to go on since longer careers equals more rushing opportunities.

[10]:
```
att_top25 = df[df['rushing_attempts'] > ra_quant[2]]
att_bot25 = df[df['rushing_attempts'] < ra_quant[0]]
att_top25_mean_games = round(att_top25['games_played'].mean(), 1)
att_top25_median_games = round(att_top25['games_played'].median(), 1)
att_bot25_mean_games = round(att_bot25['games_played'].mean(), 1)
att_bot25_median_games = round(att_bot25['games_played'].median(), 1)
print('Mean games played of top 25% of QBs [count: {}] based on rushing '
      'attempts: {}'.format(att_top25.shape[0], att_top25_mean_games))
print('Mean games played of bottom 25% of QBs [count: {}] based on rushing '
```

```
      'attempts: {}'.format(att_bot25.shape[0], att_bot25_mean_games))
```

```
Mean games played of top 25% of QBs [count: 79] based on rushing attempts: 156.4
Mean games played of bottom 25% of QBs [count: 80] based on rushing attempts:
41.5
```

Instead, it's more insightful to look at yards per rushing attempt. Here I compare the length of careers of the top 25% of QBs - based on career yards per attempt - to the bottom 25%. QBs in the bottom quartile for rushing yards per attempt have an average career length of 11 more games (more than half a season) than the bottom 25%.

[11]:
```
ypa_top25 = df[df['yards_per_attempt'] > ypa_quant[2]]
ypa_bot25 = df[df['yards_per_attempt'] < ypa_quant[0]]
ypa_top25_mean_games = round(ypa_top25['games_played'].mean(), 1)
ypa_top25_median_games = round(ypa_top25['games_played'].median(), 1)
ypa_bot25_mean_games = round(ypa_bot25['games_played'].mean(), 1)
ypa_bot25_median_games = round(ypa_bot25['games_played'].median(), 1)
print('Mean games played, top 25% of QBs [count: {}] based on yards per '
      'attempt: \n---> {} <---'.format(ypa_top25.shape[0],
                                       ypa_top25_mean_games))
print('Mean games played, bottom 25% of QBs [count: {}] based on yards per '
      'attempt: \n---> {} <---'.format(ypa_bot25.shape[0],
                                       ypa_bot25_mean_games))
```

```
Mean games played, top 25% of QBs [count: 80] based on yards per attempt:
---> 84.2 <---
Mean games played, bottom 25% of QBs [count: 79] based on yards per attempt:
---> 95.3 <---
```