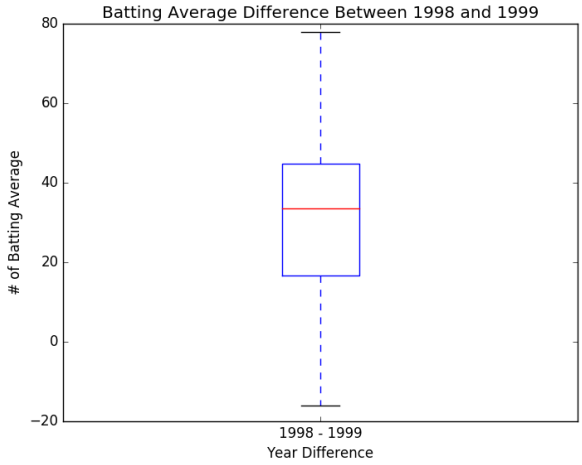
Jin Young Jeon

Lab 1

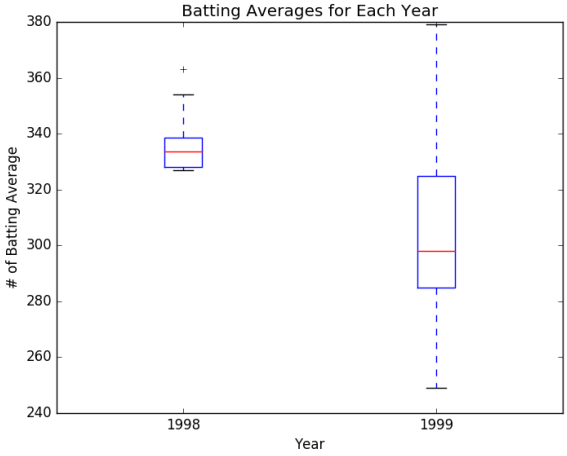
1. Batting Average Difference Between Year 1998 and 1999



Median: 33
Q1: 12
Q3: 47

[See lab1-code.py for details]

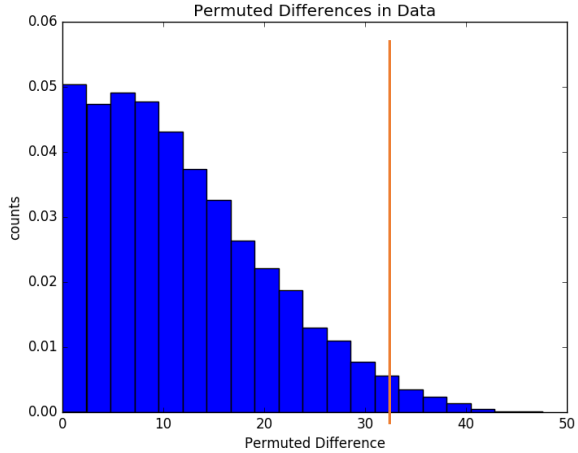
2. Batting Averages for Each Year



Year 1998:
Median: 333
IQR: 11

Year 1999:
Median: 298
IQR: 51

3. Histogram of Permuted Difference (Actual Mean Difference ‘d’ = 33.40)



The p-value is approximately 0.01 which is significantly less than 0.05. Because p-value is small, we can reject the null hypothesis that states the batting averages of the two years are not significantly different. We can assume that there is statistical difference in batting averages of two years.

[Appendix]
Include your Python implementation for this function in the report (appendix) [1pt]

```
datGrpA = np.array([363,354,339,337,336,331,328,328,327,327])
datGrpB = np.array([379,298,342,281,249,298,297,303,257,332])

def permtest(datGrpA, datGrpB, nperm):
        d = np.abs(np.mean(datGrpA) - np.mean(datGrpB))
        i = 0
        n = len(datGrpA)
        sets = np.concatenate([datGrpA, datGrpB])
        lst_permdiff = []
        for k in range(nperm):
                np.random.shuffle(sets)
                permdiff = np.abs(np.mean(sets[:n]) - np.mean(sets[n:]))
                i += d < permdiff
                lst_permdiff.append(permdiff)
        pvalue = i / nperm

        #Histogram from the collected 10000 # of nperm for permdiff
        plt.hist(lst_permdiff, bins = 20, normed=True)
        plt.title('Permuted Differences in Data')
        plt.xlabel('Permuted Difference')
        plt.ylabel('counts')
        plt.show()
        return pvalue, permdiff, d

permtest(datGrpA, datGrpB, 10000)
# pvalue < 0.05
# d = 33.40
```