Syed Abidi, DS-530, Pgae- 25, Exercise 2-4

Question: Using the variable totalwgt_lb, investigate whether first babies are lighter or heavier than others…

In [1]:
```
from __future__ import print_function, division

import numpy as np
import sys

import thinkplot
import math
import nsfg
import thinkstats2
```

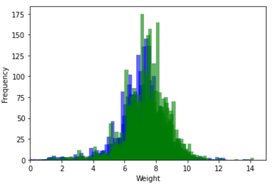
In [2]:
```
#  Read and selected records for live births by applying the boolean funcion
preg = nsfg.ReadFemPreg()
live = preg[preg.outcome == 1]
```

In [3]:
```
#  Compared the distribution of pregnancy lengths for first babies and others
firsts = live[live.birthord == 1]
others = live[live.birthord != 1]
```

Analysis to conclude if first babies are lighter or heavier than others.

In [4]:
```
# Create the Histogram Plot of the Total Weight of babies by using on one axis

first_hist = thinkstats2.Hist(firsts.totalwgt_lb)
other_hist = thinkstats2.Hist(others.totalwgt_lb)
width = 0.25
thinkplot.PrePlot(2)
thinkplot.Hist(first_hist, align='right', width=width, color = 'blue')
thinkplot.Hist(other_hist, align='left', width=width, color = 'green')
thinkplot.Show(xlabel='Weight', ylabel='Frequency', xlim=[0,15])
```




```
<Figure size 576x432 with 0 Axes>
```

In [5]:
```
# First we calculate the Mean of total weight of first babies
meanfirst_lb = round(firsts.totalwgt_lb.mean(), 2)
print(f'The mean of total weight of first babies is:: {meanfirst_lb} pounds')

meanothers_lb = round(others.totalwgt_lb.mean(), 2)
print(f'The mean of total weight of other babies is:: {meanothers_lb} pounds')


# We also calculate the Variance of total weight of first babies
varfirst_lb = round(firsts.totalwgt_lb.var(),2)
print(f'\nThe Variance of total weight of first babies is:: {varfirst_lb} pounds')
varothers_lb = round(others.totalwgt_lb.var(),2)
print(f'The Variance of total weight of first babies is:: {varothers_lb} pounds')



# We also calculate the Standard Deviation of total weight of first babies
stdfirst_lb = round(firsts.totalwgt_lb.std(),2)
print(f'\nThe Standard Deviation of total weight of first babies is:: {stdfirst_lb} pounds')
stdothers_lb = round(others.totalwgt_lb.std(),2)
print(f'The Standard Deviation of total weight of first babies is:: {stdothers_lb} pounds')
```

```
The mean of total weight of first babies is:: 7.2 pounds
The mean of total weight of other babies is:: 7.33 pounds

The Variance of total weight of first babies is:: 2.02 pounds
The Variance of total weight of first babies is:: 1.94 pounds

The Standard Deviation of total weight of first babies is:: 1.42 pounds
The Standard Deviation of total weight of first babies is:: 1.39 pounds
```

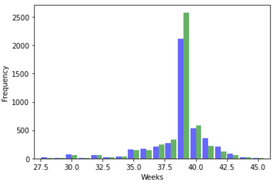
Conclusion: Based on the above Histogram, it is hard to decide if all first babies (blue bars) are lighter than others (green bars). As we can see in above histogram, there are fewer "First Babies" than "others". so some of the apparent diffeernces in the bistorgrams are due to sample sizes. some bars shows first babies are lighter than others and vice versa. So these are not the best choices for comparing two distributions.

Analysis to conclude if first babies are lighter or heavier than others.

In [6]:
```
# Plotting the Pregnency Length Histograms on one axis

first_hist = thinkstats2.Hist(firsts.prglngth)
other_hist = thinkstats2.Hist(others.prglngth)

width = 0.45
thinkplot.PrePlot(2)
thinkplot.Hist(first_hist, align='right', width=width, color='blue')
thinkplot.Hist(other_hist, align='left', width=width, color='green')
thinkplot.Show(xlabel='Weeks', ylabel='Frequency', xlim=[27,46])
```




```
<Figure size 576x432 with 0 Axes>
```

In [7]:
```
# Calculate Mean of Pregnancy length of first babies
meanfprg = round(firsts.prglngth.mean(),2)
mean_others_prg = round(others.prglngth.mean(),2)
print(f'The mean of first pregency length of first babies is:: {meanfprg} weeks')
print(f'The mean of first pregency length of other babies is:: {mean_others_prg} weeks')

# Calculate Variance of Pregnancy length of first babies
varfprg = round(firsts.prglngth.var(),2)
var_others_prg = round(others.prglngth.var(),2)
print(f'\nThe variance of first pregency length of first babies is:: {varfprg} weeks')
print(f'The variance of first pregency length of other babies is:: {var_others_prg} weeks')


# Calculate Standard Deviation of Pregnancy length of first babies
stdfprg = round(firsts.prglngth.std())
std_others_prg = round(others.prglngth.std(), 2)
print(f'\nThe Standard Deviation of first pregency length of first babies is:: {stdfprg} weeks')
print(f'The Standard Deviation of first pregency length of other babies is:: {std_others_prg} weeks')
```

```
The mean of first pregency length of first babies is:: 38.6 weeks
The mean of first pregency length of other babies is:: 38.52 weeks

The variance of first pregency length of first babies is:: 7.79 weeks
The variance of first pregency length of other babies is:: 6.84 weeks

The Standard Deviation of first pregency length of first babies is:: 3 weeks
The Standard Deviation of first pregency length of other babies is:: 2.62 weeks
```

In [8]:
```
# Define function to calculate Cohen's d

def CohenEffectSize(group1, group2):
    diff = group1.mean() - group2.mean()
    var1 = group1.var()
    var2 = group2.var()
    n1, n2 = len(group1), len(group2)
    pooled_var = (n1 * var1 + n2 * var2) / (n1 + n2)
    d = diff / math.sqrt(pooled_var)
    return d
```

In [12]:
```
# Calculate Cohen Effect Size for Total Weight of First and Other babies

Cohen_Effect_Size_Weight = round(CohenEffectSize(firsts.totalwgt_lb, others.totalwgt_lb),3)
print(f' The Cohen Effect Size of the total Weight of first and other babies:: {Cohen_Effect_Size_Weigh
t}')
```

```
 The Cohen Effect Size of the total Weight of first and other babies:: -0.089
```

Conclusion: Whether first babies are lighter or heavier than others? Based on the mean, varirance, standar devivation, Histogram and Effect size of less than 0.2, we see that fewer "First Babies" than "others". so some of the apparent diffeernces in the Historgrams are due to the sample sizes. some bars shows first babies are lighter than others and vice versa. So these are not the best choices for comparing two distributions.

In [13]:
```
# Calculate Cohen Effect Size for Length of Pregnancy of First and Other babies

Cohen_Effect_Size_Length = round(CohenEffectSize(firsts.prglngth, others.prglngth),3)
print(f' The Cohen Effect Size of the total pregancy length of first and other babies:: {Cohen_Effect_S
ize_Length}')
```

```
 The Cohen Effect Size of the total pregancy length of first and other babies:: 0.029
```

Conclusion: To summarize whether first baies arrive late and per the mean, varirance, standar devivation, Histogram and Effect size of less than 0.2. As we can see in above histogram, there are fewer "First Babies" than "others". so some of the apparent diffeernces in the bistorgrams are due to sample sizes. some bars shows first babies are lighter than others and vice versa. So these are not the best choices for comparing two distributions.

In [ ]:
```

```

CohenEffectSize of CohenEffectSize of totalwgt_lb is -0.089 and prglngth is 0.029 and . This implies that Standard deviation of Total Weight is almost 3 times more than Standard deviation of difference of Pregnancy Lenghth.