Rodriguez_Felipe_DSC530_Exercise3.2

December 18, 2022

Author: Felipe Rodriguez

0.1 Assignment 1-1

[253]:

```
# Carried over from assignment to download urls
from os.path import basename, exists


def download(url):
    filename = basename(url)
    if not exists(filename):
        from urllib.request import urlretrieve

        local, _ = urlretrieve(url, filename)
        print("Downloaded " + local)
```

[254]:

```
preg.birthord.value_counts().sort_index()
```

[254]:

```
1.0     4413
2.0     2874
3.0     1234
4.0      421
5.0      126
6.0       50
7.0       20
8.0        7
9.0        2
10.0       1
Name: birthord, dtype: int64
```

[255]:

```
preg.birthord.isnull().sum()
```

[255]:

```
4445
```

[256]:

```
preg.prglngth.value_counts().sort_index()
```

[256]:

```
0       15
1        9
```

```
2        78
3       151
4       412
5       181
6       543
7       175
8       409
9       594
10      137
11      202
12      170
13      446
14       29
15       39
16       44
17      253
18       17
19       34
20       18
21       37
22      147
23       12
24       31
25       15
26      117
27        8
28       38
29       23
30      198
31       29
32      122
33       50
34       60
35      357
36      329
37      457
38      609
39     4744
40     1120
41      591
42      328
43      148
44       46
45       10
46        1
47        1
48        7
```

```
50      2
Name: prglngth, dtype: int64
```

[257]:
```
preg.totalwgt_lb.mean()
```

[257]:
```
7.265628457623368
```

[258]:
```
preg['totalwgt_kg'] = preg.totalwgt_lb / 2.205
preg.totalwgt_kg.mean()
```

[258]:
```
3.2950695952940463
```

[259]:
```
download("https://github.com/AllenDowney/ThinkStats2/raw/master/code/
 ↪2002FemResp.dct")
download("https://github.com/AllenDowney/ThinkStats2/raw/master/code/
 ↪2002FemResp.dat.gz")
```

[260]:
```
resp = nsfg.ReadFemResp()
```

[261]:
```
resp.head()
```

[261]:
```
   caseid  rscrinf  rdormres  rostscrn  rscreenhisp  rscreenrace  age_a  \
0    2298        1         5         5            1          5.0     27
1    5012        1         5         1            5          5.0     42
2   11586        1         5         1            5          5.0     43
3    6794        5         5         4            1          5.0     15
4     616        1         5         4            1          5.0     20

   age_r  cmbirth  agescrn  …  pubassis_i     basewgt  adj_mod_basewgt  \
0     27      902       27  …           0  3247.916977      5123.759559
1     42      718       42  …           0  2335.279149      2846.799490
2     43      708       43  …           0  2335.279149      2846.799490
3     15     1042       15  …           0  3783.152221      5071.464231
4     20      991       20  …           0  5341.329968      6437.335772

      finalwgt  secu_r  sest  cmintvw  cmlstyr  screentime  intvlngth
0  5556.717241       2    18     1234     1222    18:26:36  110.492667
1  4744.191350       2    18     1233     1221    16:30:59   64.294000
2  4744.191350       2    18     1234     1222    18:19:09   75.149167
3  5923.977368       2    18     1234     1222    15:54:43   28.642833
4  7229.128072       2    18     1233     1221    14:19:44   69.502667

[5 rows x 3087 columns]
```

[262]:
```
resp.age_r.value_counts().sort_index()
```

```
[262]: 15    217
       16    223
       17    234
       18    235
       19    241
       20    258
       21    267
       22    287
       23    282
       24    269
       25    267
       26    260
       27    255
       28    252
       29    262
       30    292
       31    278
       32    273
       33    257
       34    255
       35    262
       36    266
       37    271
       38    256
       39    215
       40    256
       41    250
       42    215
       43    253
       44    235
       Name: age_r, dtype: int64
```

```
[263]: resp[resp.caseid==2298]
```

```
[263]:    caseid  rscrinf  rdormres  rostscrn  rscreenhisp  rscreenrace  age_a  \
       0    2298        1         5         5            1          5.0     27

          age_r  cmbirth  agescrn  …  pubassis_i      basewgt  adj_mod_basewgt  \
       0     27      902       27  …           0  3247.916977      5123.759559

             finalwgt  secu_r  sest  cmintvw  cmlstyr  screentime   intvlngth
       0  5556.717241       2    18     1234     1222    18:26:36  110.492667

       [1 rows x 3087 columns]
```

```
[264]: preg[preg.caseid==2298]
```

```
[264]:       caseid  pregordr  howpreg_n  howpreg_p  moscurrp  nowprgdk  pregend1  \
       2610    2298         1        NaN        NaN       NaN       NaN       6.0
       2611    2298         2        NaN        NaN       NaN       NaN       6.0
       2612    2298         3        NaN        NaN       NaN       NaN       6.0
       2613    2298         4        NaN        NaN       NaN       NaN       6.0

             pregend2  nbrnaliv  multbrth  …  religion_i  metro_i      basewgt  \
       2610       NaN       1.0       NaN  …           0        0  3247.916977
       2611       NaN       1.0       NaN  …           0        0  3247.916977
       2612       NaN       1.0       NaN  …           0        0  3247.916977
       2613       NaN       1.0       NaN  …           0        0  3247.916977

             adj_mod_basewgt     finalwgt  secu_p  sest  cmintvw  totalwgt_lb  \
       2610      5123.759559  5556.717241       2    18      NaN       6.8750
       2611      5123.759559  5556.717241       2    18      NaN       5.5000
       2612      5123.759559  5556.717241       2    18      NaN       4.1875
       2613      5123.759559  5556.717241       2    18      NaN       6.8750

             totalwgt_kg
       2610     3.117914
       2611     2.494331
       2612     1.899093
       2613     3.117914

       [4 rows x 245 columns]
```

```
[265]: resp[resp.caseid==1].age_r
```

```
[265]: 1069    44
       Name: age_r, dtype: int64
```

```
[266]: preg[preg.caseid==2298].prglngth
```

```
[266]: 2610    40
       2611    36
       2612    30
       2613    40
       Name: prglngth, dtype: int64
```

```
[267]: preg[preg.caseid==5012].birthwgt_lb
```

```
[267]: 5515    6.0
       Name: birthwgt_lb, dtype: float64
```

0.2 Assignment 1-2

[268]:

```python
# Import utilities for code
import numpy as np
import sys

import nsfg
import thinkstats2


# Read in file
def ReadFemResp(dct_file='2002FemResp.dct',
                dat_file='2002FemResp.dat.gz',
                nrows=None):
    dct = thinkstats2.ReadStataDct(dct_file)
    df = dct.ReadFixedWidth(dat_file, compression='gzip', nrows=nrows)
    CleanFemResp(df)
    return df


def CleanFemResp(df):
    pass


def ValidatePregnum(resp):
    # Resp File count
    resp_pregnum_count = resp.pregnum.count()
    print('Resp File Count:', resp_pregnum_count)

    # Read in preg
    preg = nsfg.ReadFemPreg()

    # Preg File count
    preg_pregnum_count = preg.pregnum.count()
    print('Preg File Count:', preg_pregnum_count)

    # Validation
    if preg_pregnum_count == resp_pregnum_count:
        print('Value counts match!')
    else:
        print('Values do not match!')

    return


# Main function
def main(script):
    resp = ReadFemResp()
    ValidatePregnum(resp)


if __name__ == '__main__':
    main(sys.argv)
```

```
Resp File Count: 7643
Preg File Count: 13593
Values do not match!
```

0.3 Assignment 2-1

The data for when first babies arrive late show that, on average, when weeks are above 40 most are first born. However, the important thing to note is that the majority of pregnancies are within 39-40 weeks. As weeks progress to 44, there are slightly more first borns.

The summary statistics I would use to get an evening summary would be the same as an axious patient. The data shows that most pregnancies last around 39-40 weeks but there is a chance that that first borns will slightly last longer but there are few pregnancies that last that long.

Do first babies arrive late?

Pregnancies typically last 39-40 weeks, once they are above that timeframe they are considered "late." In the histrogram created in chapter two, you can see that there are more babies that are first that are late than others. Also, 39 weeks is when most babies are born and you can see that first born babies do are born the most in this category but babies other than first born are typicall born here more often. Between 37.5 weeks other babies are born more commonly than first borns.

0.4 Assignment 2-4

[269]:

```
# Import Utilities
import sys

import thinkstats2
import nsfg
import thinkplot
```

[270]:

```
# Weight Calculations
def WeightInvestigation():
    # Define first born and other born
    firsts = live[live.birthord == 1]
    others = live[live.birthord != 1]

    # Means for both values
    mean_first = firsts.totalwgt_lb.mean()
    mean_other = others.totalwgt_lb.mean()
    print('Mean Weight')
    print('First Born:', mean_first)
    print('Others:', mean_other)

    # Cohen D Calculation
    difference = (mean_first - mean_other)
    print('Difference in pounds:', difference)
    d_weight = thinkstats2.CohenEffectSize(firsts.totalwgt_lb, others.
 ↪totalwgt_lb)
    print('Cohen d:', d_weight)
    return
```

[271]:

```
# Length Calculations
def LengthInvestigation():
    # Define first born and other born
    firsts = live[live.birthord == 1]
    others = live[live.birthord != 1]

    # Means for both values
    mean_first = firsts.prglngth.mean()
    mean_other = others.prglngth.mean()
    print('Mean Length')
    print('First Born:', mean_first)
    print('Others:', mean_other)

    # Cohen D Calculation
    difference = (mean_first - mean_other)
    print('Difference in Pregnancy Length:', difference)
    d_length = thinkstats2.CohenEffectSize(firsts.prglngth, others.prglngth)
    print('Cohen d:', d_length)
    return
```

[272]:

```
# Histogram for Weight
def histogramshowweight():
    # Bring in preg
    preg = nsfg.ReadFemPreg()
    # Define live births
    live = preg[preg.outcome == 1]
    firsts = live[live.birthord == 1]
    others = live[live.birthord != 1]

    # Create histogram
    first_hist = thinkstats2.Hist(firsts.totalwgt_lb, label='First')
    other_hist = thinkstats2.Hist(others.totalwgt_lb, label='Other')

    # Display histogram
    width = 0.01
    thinkplot.PrePlot(2)
    thinkplot.Hist(first_hist, align='right', width=width)
    thinkplot.Hist(other_hist, align='left', width=width)
    thinkplot.show(xlabel='Total Weight Pounds', ylabel='Frequency',␣
 ↪xlim=[0,15])
    return
```

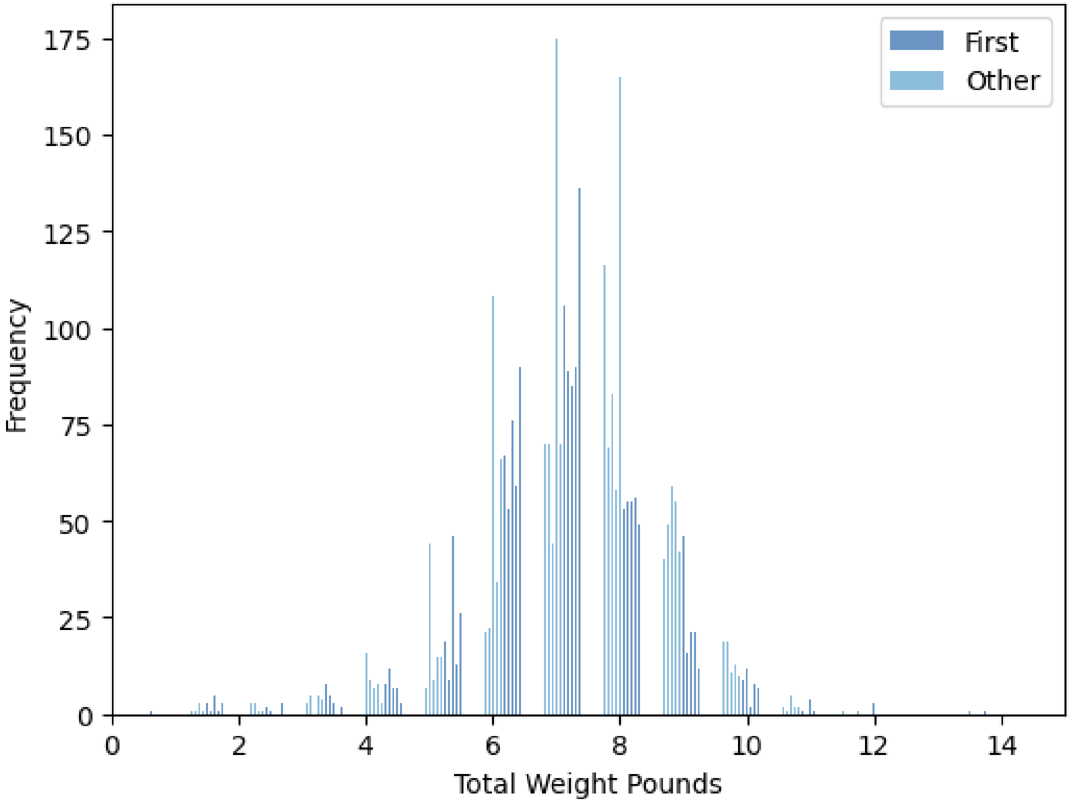
[273]:

```
# Histrogram for Length
def histogramshowlength():
    # Bring in preg
    preg = nsfg.ReadFemPreg()
    # Define live births
    live = preg[preg.outcome == 1]
    firsts = live[live.birthord == 1]
    others = live[live.birthord != 1]

    # Create histogram
    first_hist = thinkstats2.Hist(firsts.prglngth, label='First')
    other_hist = thinkstats2.Hist(others.prglngth, label='Other')

    # Display histrogram
    width = 0.45
    thinkplot.PrePlot(2)
    thinkplot.Hist(first_hist, align='right', width=width)
    thinkplot.Hist(other_hist, align='left', width=width)
    thinkplot.show(xlabel='Total Pregnancy Length', ylabel='Frequency',␣
 ↪xlim=[27,46])
    return
```

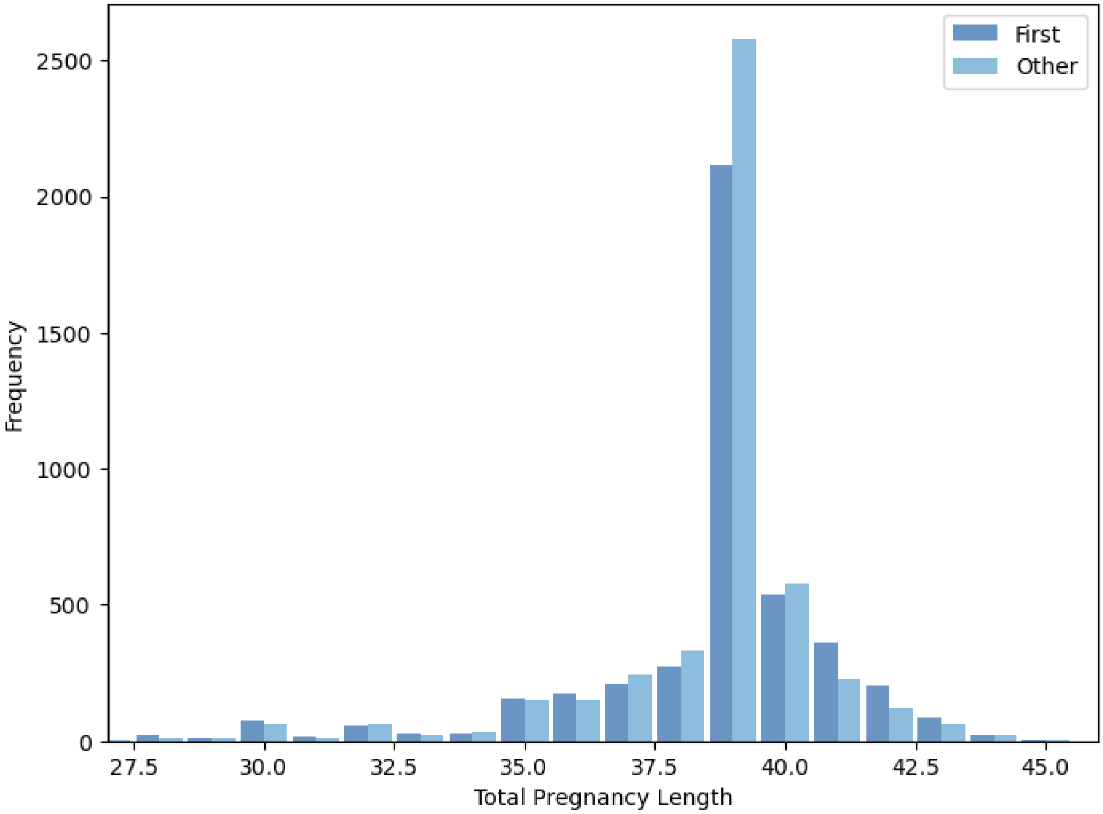
[274]:

```
# Main function
def main(script):
    WeightInvestigation()
    histogramshowweight()
    LengthInvestigation()
    histogramshowlength()

if __name__ == '__main__':
    main(sys.argv)
```

```
Mean Weight
First Born: 7.201094430437772
Others: 7.325855614973262
Difference in pounds: -0.12476118453549034
Cohen d: -0.088672927072602
```



```
Mean Length
First Born: 38.60095173351461
Others: 38.52291446673706
Difference in Pregnancy Length: 0.07803726677754952
Cohen d: 0.028879044654449883
```



```
<Figure size 800x600 with 0 Axes>
```

The differences in first born babies weight versus others is about 0.12lbs or about 1.77%. The difference in pregnancy length is about 0.2%. Cohen D, or the difference in means in standard deviations, for both baby weight and pregnancy length are both very small and can supply a trend of what is seen with the data.