In [1]:
```
# DSC530-T302
# Stephen Smitshoek
# Week07
# Exercise 9-1
```

In [2]:
```
import first
import thinkstats2
import numpy as np
```

In [3]:
```
class DiffMeansPermute(thinkstats2.HypothesisTest):
    def TestStatistic(self, data):
        group1, group2 = data
        test_stat = abs(group1.mean() - group2.mean())
        return test_stat

    def MakeModel(self):
        group1, group2 = self.data
        self.n, self.m = len(group1), len(group2)
        self.pool = np.hstack((group1, group2))

    def RunModel(self):
        np.random.shuffle(self.pool)
        data = self.pool[:self.n], self.pool[self.n:]
        return data
```

In [4]:
```
class CorrelationPermute(thinkstats2.HypothesisTest):
    def TestStatistic(self, data):
        xs, ys = data
        test_stat = abs(thinkstats2.Corr(xs, ys))
        return test_stat

    def RunModel(self):
        xs, ys = self.data
        xs = np.random.permutation(xs)
        return xs, ys
```

In [5]:
```
class PregLengthTest(thinkstats2.HypothesisTest):
    def MakeModel(self):
        firsts, others = self.data
        self.n = len(firsts)
        self.pool = np.hstack((firsts, others))

        pmf = thinkstats2.Pmf(self.pool)
        self.values = range(35, 44)
        self.expected_probs = np.array(pmf.Probs(self.values))

    def RunModel(self):
        np.random.shuffle(self.pool)
        data = self.pool[:self.n], self.pool[self.n:]
        return data

    def TestStatistic(self, data):
        firsts, others = data
        stat = self.ChiSquared(firsts) + self.ChiSquared(others)
        return stat

    def ChiSquared(self, lengths):
```

```
        hist = thinkstats2.Hist(lengths)
        observed = np.array(hist.Freqs(self.values))
        expected = self.expected_probs * len(lengths)
        stat = sum((observed - expected)**2 / expected)
        return stat
```

In [6]:
```
def diff_in_means(live):
    firsts = live[live.birthord == 1]
    others = live[live.birthord != 1]

    data = firsts.prglngth.values, others.prglngth.values
    ht = DiffMeansPermute(data)
    birthord_pvalue = ht.PValue()

    data = firsts.totalwgt_lb.values, others.totalwgt_lb.values
    ht = DiffMeansPermute(data)
    totalwgt_lb_pvalue = ht.PValue()

    return birthord_pvalue, totalwgt_lb_pvalue
```

In [7]:
```
def test_corr(live):
    live = live.dropna(subset=['agepreg', 'totalwgt_lb'])
    data = live.agepreg.values, live.totalwgt_lb.values
    ht = CorrelationPermute(data)
    pvalue = ht.PValue()

    return pvalue
```

In [8]:
```
def chi_squared(live):
    data = firsts.prglngth.values, others.prglngth.values
    ht = PregLengthTest(data)
    pvalue = ht.PValue()

    return pvalue
```

In [9]:
```
live, firsts, others = first.MakeFrames()
```

In [11]:
```
i = 0
n = len(live)
while i < 10:
    sample = thinkstats2.SampleRows(live, n)
    birthord_pvalue, totalwgt_lb_pvalue = diff_in_means(sample)
    corr_pvalue = test_corr(sample)
    chi_pvalue = chi_squared(sample)

    print("n = {}".format(n))
    print("Birth Order Mean Diff P-Value = {}".format(birthord_pvalue))
    print("Total Weight Mean Diff P-Value = {}".format(totalwgt_lb_pvalue))
    print("Age vs Weight Corr P-Value = {}".format(corr_pvalue))
    print("Chi_Squared Preg Length P-Value = {}".format(chi_pvalue))
    print()

    n //= 2
    i += 1
```

```
n = 9148
Birth Order Mean Diff P-Value = 0.198
Total Weight Mean Diff P-Value = 0.0
Age vs Weight Corr P-Value = 0.0
Chi_Squared Preg Length P-Value = 0.0

n = 4574
Birth Order Mean Diff P-Value = 0.351
Total Weight Mean Diff P-Value = 0.0
Age vs Weight Corr P-Value = 0.0
Chi_Squared Preg Length P-Value = 0.0

n = 2287
Birth Order Mean Diff P-Value = 0.42
Total Weight Mean Diff P-Value = 0.0
Age vs Weight Corr P-Value = 0.0
Chi_Squared Preg Length P-Value = 0.0

n = 1143
Birth Order Mean Diff P-Value = 0.022
Total Weight Mean Diff P-Value = 0.0
Age vs Weight Corr P-Value = 0.013
Chi_Squared Preg Length P-Value = 0.0

n = 571
Birth Order Mean Diff P-Value = 0.43
Total Weight Mean Diff P-Value = 0.0
Age vs Weight Corr P-Value = 0.073
Chi_Squared Preg Length P-Value = 0.0

n = 285
Birth Order Mean Diff P-Value = 0.761
Total Weight Mean Diff P-Value = 0.0
Age vs Weight Corr P-Value = 0.003
Chi_Squared Preg Length P-Value = 0.0

n = 142
Birth Order Mean Diff P-Value = 0.747
Total Weight Mean Diff P-Value = 0.0
Age vs Weight Corr P-Value = 0.206
Chi_Squared Preg Length P-Value = 0.0

n = 71
Birth Order Mean Diff P-Value = 0.372
Total Weight Mean Diff P-Value = 0.266
Age vs Weight Corr P-Value = 0.259
Chi_Squared Preg Length P-Value = 0.0

n = 35
Birth Order Mean Diff P-Value = 0.506
Total Weight Mean Diff P-Value = 0.077
Age vs Weight Corr P-Value = 0.782
Chi_Squared Preg Length P-Value = 0.0

n = 17
Birth Order Mean Diff P-Value = 0.349
Total Weight Mean Diff P-Value = 0.333
Age vs Weight Corr P-Value = 0.645
Chi_Squared Preg Length P-Value = 0.0
```