equal_digit_analysis

September 15, 2016

1 Equal digit analysis

1.1 Coulter count data

1.1.1 Aaron Stern

Summary: this notebook finds overall consistency with the the authors' report. The only exceptions are minute differences (e.g. number of equal digit pairs = 636 vs 644) presumably due to some unreported filtering.

In [2]:
```
import numpy as np
import scipy.stats as stats
from __future__ import division, generators, absolute_import, print_functio
from openpyxl import load_workbook

import matplotlib.pyplot as plt

%matplotlib inline

DATA_DIR = '../data/'
```

RTS

In [13]:
```
DATA_DIR = '../data/'
wb = load_workbook(DATA_DIR + 'Bishayee Coulter Counts.10.20.97-7.16.01.xl

ws = wb.get_sheet_by_name('Sheet1')
rawData = []

for i in range(3, 1732):
    for col in ['C','D','E']:
        try:
            rawData.append(float(ws[col + str(i)].value))
        except:
            # if empty cell, report location in spreadsheet
            print(col,i)
```

```
        eq = 0
        for count in rawData:
            if (count - count % 10)/10 % 10  == count % 10:
                eq += 1

        1 - stats.binom.cdf(eq,len(rawData),0.1)
```

```
E 542
E 1252
```

Out[13]: 8.1076654279144122e-09

1.1.2 Other investigators in lab

```
In [15]: wb = load_workbook(DATA_DIR + 'Other Investigators in Lab.Coulter Counts.4

        ws = wb.get_sheet_by_name('Sheet1')
        rawData = []

        for i in range(3, 1010):
            for col in ['C','D','E']:
                try:
                    rawData.append(float(ws[col + str(i)].value))
                except:
                    # if empty cell, report location in spreadsheet
                    print(col,i)

        eq = 0
        for count in rawData:
            if (count - count % 10)/10 % 10  == count % 10:
                eq += 1
```

```
E 61
E 62
E 63
E 64
E 66
E 67
E 68
E 69
E 70
E 71
E 72
E 73
E 74
E 75
E 76
```

E 77
E 78
E 79
E 80
E 81
E 82
E 83
E 84
E 85
E 86
E 87
E 88
E 99
E 100
E 101
E 102
E 103
E 104
E 105
E 106
E 107
E 108
E 109
E 110
E 111
E 112
E 113
E 114
E 115
E 116
E 117
E 118
E 133
E 134
E 135
E 136
E 137
E 138
E 314
E 315
E 316
E 317
E 318
E 319
E 323
E 324
E 325
E 327

```
E 463
C 673
E 696
D 704
E 704
E 719
E 724
E 725
E 726
E 727
E 728
E 781
E 814
E 867
E 970
E 972
```

```
Out[15]: 0.56274856153791086
```

```
In [16]: 1 - stats.binom.cdf(eq,len(rawData),0.1)
```

```
Out[16]: 0.56274856153791086
```

1.1.3 Outside labs

```
In [17]: wb = load_workbook(DATA_DIR + 'Outside Lab 1.Coulter Counts.6.7.91-4.9.99.

         ws = wb.get_sheet_by_name('Sheet1')
         rawData = []

         for i in range(2, 111):
             for col in ['B','C','D']:
                 try:
                     rawData.append(float(ws[col + str(i)].value))
                 except:
                     # if empty cell, report location in spreadsheet
                     print(col,i)

         eq = 0
         for count in rawData:
             if (count - count % 10)/10 % 10  == count % 10:
                 eq += 1
```

```
D 41
D 42
D 43
D 44
D 45
```

```
D 46
D 47
D 48
D 49
D 50
D 51
D 52
```

In [18]:
```
1 - stats.binom.cdf(eq,len(rawData),0.1)
```

Out[18]:
```
0.41608665899966624
```

In [19]:
```
wb = load_workbook(DATA_DIR + 'Outside Lab 2.Coulter Counts.6.6.08-7.7.08.
rawData = []
ws = wb.get_sheet_by_name('Sheet1')

for i in range(3, 123):
    for col in ['B','C','D']:
        try:
            rawData.append(float(ws[col + str(i)].value))
        except:
            # if empty cell, report location in spreadsheet
            print(col,i)
```

In [20]:
```
1 - stats.binom.cdf(eq,len(rawData),0.1)
```

Out[20]:
```
0.72578603959485988
```

In [ ]: