Age groups

In [1]:

```
import pandas as pd
import numpy as np
from scipy.stats import chi2_contingency
from scipy.stats import fisher_exact

data = pd.read_csv(filepath_or_buffer='../../../Archive/HTWTempRatios.csv')
```

In [2]:

```
#Compute a contingency table for age groups hitting the wall.
data["HTW"] = (data['DoS15km'] >= 0.25) | (data['DoS20km'] >= 0.25)
data["AgeGroup"] = "None"

data.loc[data["Age"].between(17,29, inclusive='both'), 'AgeGroup'] = "17-29" #remove any with missing age
data.loc[data["Age"].between(30,39, inclusive='both'), 'AgeGroup'] = "30-39"
data.loc[data["Age"].between(40,49, inclusive='both'), 'AgeGroup'] = "40-49"
data.loc[data["Age"].between(50,59, inclusive='both'), 'AgeGroup'] = "50-59"
data.loc[data["Age"].between(60,99, inclusive='both'), 'AgeGroup'] = "60+" #remove any unrealisic outlier

# Show number of male/female runner per age group
mf_tab = pd.crosstab(data["AgeGroup"], data['Gender'])
mf_tab
```

Out[2]:

Gender	F	M
AgeGroup		
17-29	39300	49731
30-39	41320	84164
40-49	38938	85337
50-59	16488	45773
60+	3331	15941
None	1032	2141

Runners Hitting the Wall per age group

In [3]:

```
df = data.loc[data["AgeGroup"] != "None"] # drop datapoints with missing or wrong age.
htw_tab = pd.crosstab(df["AgeGroup"], df['HTW'])
f_htw_tab = pd.crosstab((df.loc[df["Gender"] == "F"])["AgeGroup"], df['HTW'])
m_htw_tab = pd.crosstab((df.loc[df["Gender"] == "M"])["AgeGroup"], df['HTW'])
```

In [4]:

```
# Number of females hitting the wall per age group
f_htw_tab
```

Out[4]:

HTW	False	True
AgeGroup		
17-29	36549	2751
30-39	39363	1957
40-49	37290	1648
50-59	15576	912
60+	3152	179

In [5]:

```
c, p, dof, expected = chi2_contingency(f_htw_tab)
print("Chi-squared females HTW by age groups, p: ", p)
```

```
Chi-squared females HTW by age groups, p:  3.145193918645074e-72
```

In [6]:

```
# Number of males hitting the wall per age group
m_htw_tab
```

Out[6]:

HTW	False	True
AgeGroup		
17-29	43284	6447
30-39	75567	8597
40-49	77699	7638
50-59	41528	4245
60+	14362	1579

In [7]:

```
c, p, dof, expected = chi2_contingency(m_htw_tab)
print("Chi-squared males HTW by age group, p: ", p)
```

```
Chi-squared males HTW by age group, p:  6.021907258238046e-130
```

In [8]:

```
# Overall runners hitting the wall per age group
htw_tab
```

Out[8]:

HTW	False	True
AgeGroup		
17-29	79833	9198
30-39	114930	10554
40-49	114989	9286
50-59	57104	5157
60+	17514	1758

In [9]:

```
c, p, dof, expected = chi2_contingency(htw_tab)
print("Chi-squared (all) by age group, p: ", p)
```

```
Chi-squared (all) by age group, p:  1.955975541098763e-120
```

In [10]:

```
# Effect sizes between men and women, within age group
df1 = df.loc[df["AgeGroup"] == "17-29"]
df2 = df.loc[df["AgeGroup"] == "30-39"]
df3 = df.loc[df["AgeGroup"] == "40-49"]
df4 = df.loc[df["AgeGroup"] == "50-59"]
df5 = df.loc[df["AgeGroup"] == "60+"]
oddsr1, p1 = fisher_exact(pd.crosstab(df1['Gender'], df1['HTW']))
oddsr2, p2 = fisher_exact(pd.crosstab(df2['Gender'], df2['HTW']))
oddsr3, p3 = fisher_exact(pd.crosstab(df3['Gender'], df3['HTW']))
oddsr4, p4 = fisher_exact(pd.crosstab(df4['Gender'], df4['HTW']))
oddsr5, p5 = fisher_exact(pd.crosstab(df5['Gender'], df5['HTW']))
#print("Difference between Male/Female within age group: ")
print("Effect size for HTW between M/F within each age group")
print ("Age Group 17-29 M vs. F:\n p: ",p1, " OR: ", oddsr1)
print ("Age Group 30-39 M vs. F:\n p: ",p2, " OR: ", oddsr2)
print ("Age Group 40-49 M vs. F:\n p: ",p3, " OR: ", oddsr3)
print ("Age Group 50-59 M vs. F:\n p: ",p4, " OR: ", oddsr4)
print ("Age Group 60+ M vs. F:\n p: ",p5, " OR: ", oddsr5)
```

```
Effect size for HTW between M/F within each age group
Age Group 17-29 M vs. F:
 p:  1.6474952537552732e-191   OR:  1.978860548932631
Age Group 30-39 M vs. F:
 p:  2.7750442515647507e-259   OR:  2.2882955190897842
Age Group 40-49 M vs. F:
 p:  5.189780685309522e-208   OR:  2.2243309287758075
Age Group 50-59 M vs. F:
 p:  2.5021991745757493e-54   OR:  1.7458132243052247
Age Group 60+ M vs. F:
 p:  2.9065017297764597e-18   OR:  1.9359778559031087
```

In [11]:

```
#Effect sizes between successive age groups
g1 = df.loc[(df["AgeGroup"] == "17-29") | (df["AgeGroup"] == "30-39") ]
g2 = df.loc[(df["AgeGroup"] == "30-39") | (df["AgeGroup"] == "40-49") ]
g3 = df.loc[(df["AgeGroup"] == "40-49") | (df["AgeGroup"] == "50-59") ]
g4 = df.loc[(df["AgeGroup"] == "50-59") | (df["AgeGroup"] == "60+") ]

oddsr1, p1 = fisher_exact(pd.crosstab(g1["AgeGroup"],g1["HTW"]))
oddsr2, p2 = fisher_exact(pd.crosstab(g2["AgeGroup"],g2["HTW"]))
oddsr3, p3 = fisher_exact(pd.crosstab(g3["AgeGroup"],g3["HTW"]))
oddsr4, p4 = fisher_exact(pd.crosstab(g4["AgeGroup"],g4["HTW"]))
print("Effect size for HTW between consequtive age groups (F+M).")
print ("Age Group 17-29 vs. 30-39:\n p: ",p1, " OR: ", oddsr1)
print ("Age Group 30-39 vs. 40-49:\n p: ",p2, " OR: ", oddsr2)
print ("Age Group 40-49 vs. 50-59:\n p: ",p3, " OR: ", oddsr3)
print ("Age Group 50-59 vs. 60+:\n p: ",p4, " OR: ", oddsr4)
```

```
Effect size for HTW between consequtive age groups (F+M).
Age Group 17-29 vs. 30-39:
 p:  1.8442832507110868e-51   OR:  0.797026438112674
Age Group 30-39 vs. 40-49:
 p:  4.157398679773083e-18   OR:  0.8794045312221599
Age Group 40-49 vs. 50-59:
 p:  7.933325469560573e-10   OR:  1.1182997263359846
Age Group 50-59 vs. 60+:
 p:  0.00028888051472587795   OR:  1.1114832558452532
```

In [12]:

```
#Effect sizes between successive age groups, female only
f_df = (df.loc[df["Gender"] == "F"])
g1 = f_df.loc[(f_df["AgeGroup"] == "17-29") | (f_df["AgeGroup"] == "30-39") ]
g2 = f_df.loc[(f_df["AgeGroup"] == "30-39") | (f_df["AgeGroup"] == "40-49") ]
g3 = f_df.loc[(f_df["AgeGroup"] == "40-49") | (f_df["AgeGroup"] == "50-59") ]
g4 = f_df.loc[(f_df["AgeGroup"] == "50-59") | (f_df["AgeGroup"] == "60+") ]

oddsr1, p1 = fisher_exact(pd.crosstab(g1["AgeGroup"],g1["HTW"]))
oddsr2, p2 = fisher_exact(pd.crosstab(g2["AgeGroup"],g2["HTW"]))
oddsr3, p3 = fisher_exact(pd.crosstab(g3["AgeGroup"],g3["HTW"]))
oddsr4, p4 = fisher_exact(pd.crosstab(g4["AgeGroup"],g4["HTW"]))
print("Effect size for HTW between consequtive age groups (F only).")
print ("Age Group 17-29 vs. 30-39:\n p: ",p1, " OR: ", oddsr1)
print ("Age Group 30-39 vs. 40-49:\n p: ",p2, " OR: ", oddsr2)
print ("Age Group 40-49 vs. 50-59:\n p: ",p3, " OR: ", oddsr3)
print ("Age Group 50-59 vs. 60+:\n p: ",p4, " OR: ", oddsr4)
```

```
Effect size for HTW between consequtive age groups (F only).
Age Group 17-29 vs. 30-39:
 p:  8.604064034396752e-43   OR:  0.6605223904972399
Age Group 30-39 vs. 40-49:
 p:  0.0005722777015625179   OR:  0.8889189990261253
Age Group 40-49 vs. 50-59:
 p:  6.18639820700976e-11   OR:  1.3248724699687346
Age Group 50-59 vs. 60+:
 p:  0.7391330695303394   OR:  0.9699021506812717
```

In [13]:

```
# Effect size between 17-29 and 40-49 groups for females, where we have largest differences.
g = f_df.loc[(f_df["AgeGroup"] == "17-29") | (f_df["AgeGroup"] == "40-49") ]
oddsr1, p1 = fisher_exact(pd.crosstab(g["AgeGroup"],g["HTW"]))
print("Female 17-29 vs 40-49:\n")
print("p: ",p1)
print("OR: ",oddsr1)
```

```
Female 17-29 vs 40-49:

p:  5.5167562182057243e-64
OR:  0.58715090219515
```

In [14]:

```
#Effect sizes between successive age groups, male only
m_df = (df.loc[df["Gender"] == "M"])
g1 = m_df.loc[(m_df["AgeGroup"] == "17-29") | (m_df["AgeGroup"] == "30-39") ]
g2 = m_df.loc[(m_df["AgeGroup"] == "30-39") | (m_df["AgeGroup"] == "40-49") ]
g3 = m_df.loc[(m_df["AgeGroup"] == "40-49") | (m_df["AgeGroup"] == "50-59") ]
g4 = m_df.loc[(m_df["AgeGroup"] == "50-59") | (m_df["AgeGroup"] == "60+") ]

oddsr1, p1 = fisher_exact(pd.crosstab(g1["AgeGroup"],g1["HTW"]))
oddsr2, p2 = fisher_exact(pd.crosstab(g2["AgeGroup"],g2["HTW"]))
oddsr3, p3 = fisher_exact(pd.crosstab(g3["AgeGroup"],g3["HTW"]))
oddsr4, p4 = fisher_exact(pd.crosstab(g4["AgeGroup"],g4["HTW"]))

print("Effect size for HTW between consequtive age groups (M only).")
print ("Age Group 17-29 vs. 30-39:\n p: ",p1, " OR: ", oddsr1)
print ("Age Group 30-39 vs. 40-49:\n p: ",p2, " OR: ", oddsr2)
print ("Age Group 40-49 vs. 50-59:\n p: ",p3, " OR: ", oddsr3)
print ("Age Group 50-59 vs. 60+:\n p: ",p4, " OR: ", oddsr4)
```

```
Effect size for HTW between consequtive age groups (M only).
Age Group 17-29 vs. 30-39:
 p:  1.27518532304456e-52   OR:  0.7638084589884682
Age Group 30-39 vs. 40-49:
 p:  9.370911033589651e-19   OR:  0.8640710984290758
Age Group 40-49 vs. 50-59:
 p:  0.05270217948188681   OR:  1.0398542090417837
Age Group 50-59 vs. 60+:
 p:  0.01990636204899734   OR:  1.075549812528776
```

In [15]:

```
# Effect size between 17-29 and 40-49 for males, where we have largest differences.
g = m_df.loc[(m_df["AgeGroup"] == "17-29") | (m_df["AgeGroup"] == "40-49") ]
oddsr1, p1 = fisher_exact(pd.crosstab(g["AgeGroup"],g["HTW"]))
print("Male 17-29 vs 40-49:\n")
print("p: ",p1)
print("OR: ",oddsr1)
```

```
Male 17-29 vs 40-49:

p:  4.94315278868589e-117
OR:  0.6599848141475854
```

Analysis of runner pacing well: running a negative or equal split.

In [16]:

```
splits_tab = pd.crosstab(df['AgeGroup'], df['SplitRatio'] <= 1)
f_splits_tab = pd.crosstab((df.loc[df["Gender"] == "F"])["AgeGroup"], df['SplitRatio'] <= 1)
m_splits_tab = pd.crosstab((df.loc[df["Gender"] == "M"])["AgeGroup"], df['SplitRatio'] <= 1)
```

In [17]:

```
# Female negative splits per age group
f_splits_tab
```

Out[17]:

SplitRatio	False	True
AgeGroup		
17-29	34451	4849
30-39	37147	4173
40-49	35924	3014
50-59	15718	770
60+	3232	99

In [18]:

```
c, p, dof, expected = chi2_contingency(f_splits_tab)
print("Chi-squared female negatice splits per age group, p: ", p)
```

```
Chi-squared female negatice splits per age group, p:  8.558705866172767e-249
```

In [19]:

```
# Male negative splits per age group
m_splits_tab
```

Out[19]:

SplitRatio	False	True
AgeGroup		
17-29	42409	7322
30-39	74372	9792
40-49	77684	7653
50-59	42892	2881
60+	15241	700

In [20]:

```
c, p, dof, expected = chi2_contingency(m_splits_tab)
print("Chi-squared male negative splits per age group, p: ",p)
```

```
Chi-squared male negative splits per age group, p:  0.0
```

In [21]:

```
# Overall negative splits per age group
splits_tab
```

Out[21]:

SplitRatio	False	True
AgeGroup		
17-29	76860	12171
30-39	111519	13965
40-49	113608	10667
50-59	58610	3651
60+	18473	799

In [22]:

```
c, p, dof, expected = chi2_contingency(splits_tab)
print("Chi-squared all runners negative splits per age group, p: ",p)
```

```
Chi-squared all runners negative splits per age group, p:  0.0
```

In [23]:

```
# Effect sizes between men and women, within age group

df1 = df.loc[df["AgeGroup"] == "17-29"]
df2 = df.loc[df["AgeGroup"] == "30-39"]
df3 = df.loc[df["AgeGroup"] == "40-49"]
df4 = df.loc[df["AgeGroup"] == "50-59"]
df5 = df.loc[df["AgeGroup"] == "60+"]
oddsr1, p1 = fisher_exact(pd.crosstab(df1['Gender'], df1['SplitRatio'] <= 1))
oddsr2, p2 = fisher_exact(pd.crosstab(df2['Gender'], df2['SplitRatio'] <= 1))
oddsr3, p3 = fisher_exact(pd.crosstab(df3['Gender'], df3['SplitRatio'] <= 1))
oddsr4, p4 = fisher_exact(pd.crosstab(df4['Gender'], df4['SplitRatio'] <= 1))
oddsr5, p5 = fisher_exact(pd.crosstab(df5['Gender'], df5['SplitRatio'] <= 1))
print("Effect size for Negative Split between M/F within each age group")
print ("Age Group 17-29 M vs. F:\n p: ",p1, " OR: ", oddsr1)
print ("Age Group 30-39 M vs. F:\n p: ",p2, " OR: ", oddsr2)
print ("Age Group 40-49 M vs. F:\n p: ",p3, " OR: ", oddsr3)
print ("Age Group 50-59 M vs. F:\n p: ",p4, " OR: ", oddsr4)
print ("Age Group 60+ M vs. F:\n p: ",p5, " OR: ", oddsr5)
```

```
Effect size for Negative Split between M/F within each age group
Age Group 17-29 M vs. F:
 p:  6.327592315403746e-25   OR:  1.2266519146322405
Age Group 30-39 M vs. F:
 p:  2.8885770635229154e-16   OR:  1.172026159671495
Age Group 40-49 M vs. F:
 p:  5.267760069681919e-13   OR:  1.1741986412412397
Age Group 50-59 M vs. F:
 p:  8.311889480682368e-15   OR:  1.3711138576987687
Age Group 60+ M vs. F:
 p:  0.0001275621518623354   OR:  1.4994111444475593
```

In [24]:

```
# Effect sizes between age groups
g1 = df.loc[(df["AgeGroup"] == "17-29") | (df["AgeGroup"] == "30-39") ]
g2 = df.loc[(df["AgeGroup"] == "30-39") | (df["AgeGroup"] == "40-49") ]
g3 = df.loc[(df["AgeGroup"] == "40-49") | (df["AgeGroup"] == "50-59") ]
g4 = df.loc[(df["AgeGroup"] == "50-59") | (df["AgeGroup"] == "60+") ]
htw_tab1 = pd.crosstab(g4["AgeGroup"], df['HTW'])

oddsr1, p1 = fisher_exact(pd.crosstab(g1["AgeGroup"],g1['SplitRatio'] <= 1))
oddsr2, p2 = fisher_exact(pd.crosstab(g2["AgeGroup"],g2['SplitRatio'] <= 1))
oddsr3, p3 = fisher_exact(pd.crosstab(g3["AgeGroup"],g3['SplitRatio'] <= 1))
oddsr4, p4 = fisher_exact(pd.crosstab(g4["AgeGroup"],g4['SplitRatio'] <= 1))
print("Effect size for Negative Split between consequtive age groups (F+M).")
print ("Age Group 17-29 vs. 30-39:\n p: ",p1, " OR: ", oddsr1)
print ("Age Group 30-39 vs. 40-49:\n p: ",p2, " OR: ", oddsr2)
print ("Age Group 40-49 vs. 50-59:\n p: ",p3, " OR: ", oddsr3)
print ("Age Group 50-59 vs. 60+:\n p: ",p4, " OR: ", oddsr4)
```

```
Effect size for Negative Split between consequtive age groups (F+M).
Age Group 17-29 vs. 30-39:
 p:  9.04047385903589e-70   OR:  0.7907991454275962
Age Group 30-39 vs. 40-49:
 p:  3.0403979217925753e-101   OR:  0.7497928713511726
Age Group 40-49 vs. 50-59:
 p:  3.6240509276245927e-100   OR:  0.6634477581294234
Age Group 50-59 vs. 60+:
 p:  3.916675395737675e-21   OR:  0.6943352874759751
```

In [25]:

```
# Effect sizes between age groups female only.
f_df = (df.loc[df["Gender"] == "F"])
g1 = f_df.loc[(f_df["AgeGroup"] == "17-29") | (f_df["AgeGroup"] == "30-39") ]
g2 = f_df.loc[(f_df["AgeGroup"] == "30-39") | (f_df["AgeGroup"] == "40-49") ]
g3 = f_df.loc[(f_df["AgeGroup"] == "40-49") | (f_df["AgeGroup"] == "50-59") ]
g4 = f_df.loc[(f_df["AgeGroup"] == "50-59") | (f_df["AgeGroup"] == "60+") ]

oddsr1, p1 = fisher_exact(pd.crosstab(g1["AgeGroup"],g1['SplitRatio'] <= 1))
oddsr2, p2 = fisher_exact(pd.crosstab(g2["AgeGroup"],g2['SplitRatio'] <= 1))
oddsr3, p3 = fisher_exact(pd.crosstab(g3["AgeGroup"],g3['SplitRatio'] <= 1))
oddsr4, p4 = fisher_exact(pd.crosstab(g4["AgeGroup"],g4['SplitRatio'] <= 1))

print("Effect size for Negative Split between consequtive age groups (F only).")
print ("Age Group 17-29 vs. 30-39:\n p: ",p1, " OR: ", oddsr1)
print ("Age Group 30-39 vs. 40-49:\n p: ",p2, " OR: ", oddsr2)
print ("Age Group 40-49 vs. 50-59:\n p: ",p3, " OR: ", oddsr3)
print ("Age Group 50-59 vs. 60+:\n p: ",p4, " OR: ", oddsr4)
```

```
Effect size for Negative Split between consequtive age groups (F only).
Age Group 17-29 vs. 30-39:
 p:  7.049468683382514e-24   OR:  0.798131198338086
Age Group 30-39 vs. 40-49:
 p:  1.0651855290240096e-31   OR:  0.7468509217731616
Age Group 40-49 vs. 50-59:
 p:  9.146643315814892e-42   OR:  0.5838951669154245
Age Group 50-59 vs. 60+:
 p:  6.3866926537615255e-06   OR:  0.6252740452616691
```

In [26]:

```
# Effect sizes between age groups male only
m_df = (df.loc[df["Gender"] == "M"])
g1 = m_df.loc[(m_df["AgeGroup"] == "17-29") | (m_df["AgeGroup"] == "30-39") ]
g2 = m_df.loc[(m_df["AgeGroup"] == "30-39") | (m_df["AgeGroup"] == "40-49") ]
g3 = m_df.loc[(m_df["AgeGroup"] == "40-49") | (m_df["AgeGroup"] == "50-59") ]
g4 = m_df.loc[(m_df["AgeGroup"] == "50-59") | (m_df["AgeGroup"] == "60+") ]

oddsr1, p1 = fisher_exact(pd.crosstab(g1["AgeGroup"],g1['SplitRatio'] <= 1))
oddsr2, p2 = fisher_exact(pd.crosstab(g2["AgeGroup"],g2['SplitRatio'] <= 1))
oddsr3, p3 = fisher_exact(pd.crosstab(g3["AgeGroup"],g3['SplitRatio'] <= 1))
oddsr4, p4 = fisher_exact(pd.crosstab(g4["AgeGroup"],g4['SplitRatio'] <= 1))

print("Effect size for Negative Split between consequtive age groups (M only).")
print ("Age Group 17-29 vs. 30-39:\n p: ",p1, " OR: ", oddsr1)
print ("Age Group 30-39 vs. 40-49:\n p: ",p2, " OR: ", oddsr2)
print ("Age Group 40-49 vs. 50-59:\n p: ",p3, " OR: ", oddsr3)
print ("Age Group 50-59 vs. 60+:\n p: ",p4, " OR: ", oddsr4)
```

```
Effect size for Negative Split between consequtive age groups (M only).
Age Group 17-29 vs. 30-39:
 p:  3.173761322558781e-59   OR:  0.7625884997559754
Age Group 30-39 vs. 40-49:
 p:  4.5356937185736234e-73   OR:  0.7482352934866341
Age Group 40-49 vs. 50-59:
 p:  6.641758363870426e-67   OR:  0.6818154328255547
Age Group 50-59 vs. 60+:
 p:  1.0258060794659224e-19   OR:  0.6837819241158386
```

In [ ]: