8.2. [20 points] Analysis Part I¶

Find 95% lower (one-sided) and two-sided confidence intervals for the reduction in risk corresponding to the primary endpoint (data “Through day 29”), using method 3 and also using the cruder conservative approach via simultaneous Bonferroni confidence bounds for N·1 and N1· described in the notes on causal inference. (For the Bonferroni approach to two-sided intervals, use Sterne’s method for the underlying hypergeometric confidence intervals. Feel free to re-use your own code from the previous problem set.)

In [ ]:
```
from cibin.permute_utils import hypergeom_conf_interval
from cibin.utils import *
```

Data from Regeneron Press

In [2]:
```
n=753
m=752
N=n+m
n01 = 59
n11 = 11
n00 = m-n01
n10 = n-n11
alpha=0.05
n11, n10, n01, n00
```

Out[2]:
```
(11, 742, 59, 693)
```

One-sided confidence intervals

Upper one-sided

In [3]:
```
N_dot_1 = hypergeom_conf_interval(n11*N/n, n11, N, 1-alpha, alternative=
'upper')
N_1_dot = hypergeom_conf_interval(n01*N/m, n01, N, 1-alpha, alternative=
'lower')

upper = (N_dot_1[1] - N_1_dot[0])/N

ci = [-0.5, upper]
N_dot_1[1] - N_1_dot[0], ci
```

Out[3]:
```
(397, [-0.5, 0.2637873754152824])
```

Lower one-sided

In [4]:
```
N_dot_1 = hypergeom_conf_interval(n11*N/n, n11, N, 1-alpha, alternative=
'lower')
N_1_dot = hypergeom_conf_interval(n01*N/m, n01, N, 1-alpha, alternative=
'upper')

lower = (N_dot_1[0] - N_1_dot[1])/N

ci = [lower, 0.5]
N_1_dot[1] - N_dot_1[0], ci
```

Out[4]:
```
(395, [-0.26245847176079734, 0.5])
```

In [ ]:
```
tau_lower_oneside(n11, n10, n01, n00, alpha=0.05, nperm=1) # result arou
nd (-0.15, 0.42), span 858
```

Two-sided

In [5]:
```
N_dot_1 = hypergeom_conf_interval(n11*N/n, n11, N, 1-alpha/2, alternativ
e='two-sided')
N_1_dot = hypergeom_conf_interval(n01*N/m, n01, N, 1-alpha/2, alternativ
e='two-sided')
lower = (N_dot_1[0] - N_1_dot[1])/N
upper = (N_dot_1[1] - N_1_dot[0])/N
ci = [lower, upper]
N_1_dot[1] - N_dot_1[0], N_dot_1[1] - N_1_dot[0], ci
```

Out[5]:
```
(515, 516, [-0.34219269102990035, 0.34285714285714286])
```

In [6]:
```
two_side = hypergeom_conf_interval(n, n01, N, 1-alpha/2, alternative='tw
o-sided', method='Sterne')
```

In [7]:
```
two_side[1]-two_side[0]
```

Out[7]:
```
49
```

In [8]:
```
(two_side[0]/N, two_side[1]/N)
```

Out[8]:
```
(0.06312292358803986, 0.09568106312292358)
```

In [ ]:
```
tau_twosided_ci(n11, n10, n01, n00, 0.05, exact=False, max_combinations=
10**5, reps=100) # result around (-0.15, 0.42), span 858
```

- Discuss the differences between the two sets of confidence intervals.

According to Sterne's method, one-sided interval is wider than the two-sided one. On the other hand, for Li and Ding's method, the two-sided interval is wider than the one-sided interval.

Overall, the interval according to Sterne's method is wider as Sterne is more conservative.

- Is it statistically legitimate to use one-sided confidence intervals? Why or why not?

I think it is statistically legitimate to use one-sided confidence intervals if we support the Regeneron study's statement. This is because one-sided confidence interval favor one bound,upper or lower. The study claims that subjects with their antibody cocktail are less likely to be infected. He wants to show that the antibody cocktail can make the infection rate reduce certain percentage with high confidence. Hence, one-sided confidence interval works.

However, I would choose two-sided confidence intervals if we would like to coduct a study in general. Two-sided confidence interval will value both upper and lower bound. As a result, if we want to study the changes in infection rate of COVID-19, using two-sided confidence intervals can be more statistically legitimate.

- Are the 2-sided confidence intervals preferable to the one-sided intervals? Why or why not?

In general, I think it is preferable to use a two-sided confidence interval as one-sided can accept more extreme outcomes. Nonetheless, it also depends on the assumptions for the data. In this situation, a one-sided confidence interval works better because we assume that the sample risk is less than the population risk. In a more general case, we may prefer a two-sided confidence interval as I mentioned above. We can be able to understand the increase or decrease and overall changes using two-sided interval.