ECON626_Lab 6

Qingwen Wang

Problem 0: In class warm-up

1. if power = 0.8, we need 156,972 visitors per group, totally 313944 visitors. Depends on the possible variation between days, we need 2 weeks or 11 days to run the test.

```
# Example output
pwr_2p(p1=0.505, p2=0.50, sig_level=0.05, power = 0.8, verbose=True)
```

```
     Difference of proportion power calculation for binomial distribution (arcsine

               h = 0.01000017
               n = 156972
       sig.level = 0.05
           power = 0.8
     alternative = two.sided

NOTE: same sample sizes
```

if power =0.95, we need some 259886 visitors per group, totally 519772 visitors. Depends on the possible variation between days, we need 3 weeks or 19 days.

```
pwr_2p(p1=0.505, p2=0.50, sig_level=0.05, power = 0.95, verbose=True)
```

```
     Difference of proportion power calculation for binomial distribution (arcsine trans

               h = 0.01000017
               n = 259885.5
       sig.level = 0.05
           power = 0.95
     alternative = two.sided

NOTE: same sample sizes
```

2. if our site has strong weekly variation in conversion rates, we need to run 2 weeks when power = 0.8; and 3 weeks when power = 0.95.

Problem 1: Power analysis with expected utility

In class foundation

1. no, from the picture, we can see that since utility of doing the experiment is always below 0.

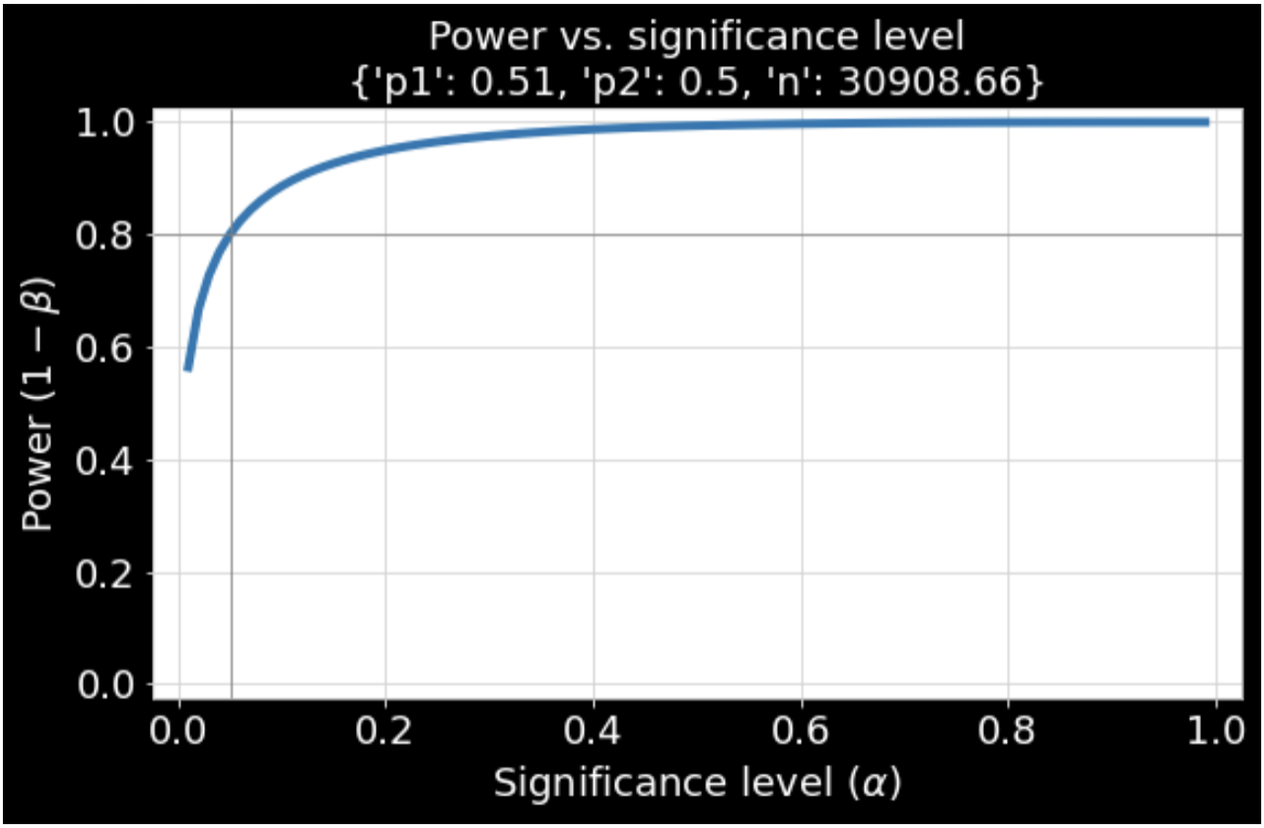
2. sample size is 61818 (30909 per group), that power can reach 0.5 with alpha = 0.05.

```
pwr_2p(p1=0.51, p2=0.50, sig_level=0.05, power = 0.8, verbose=True, alternative="greater")
```

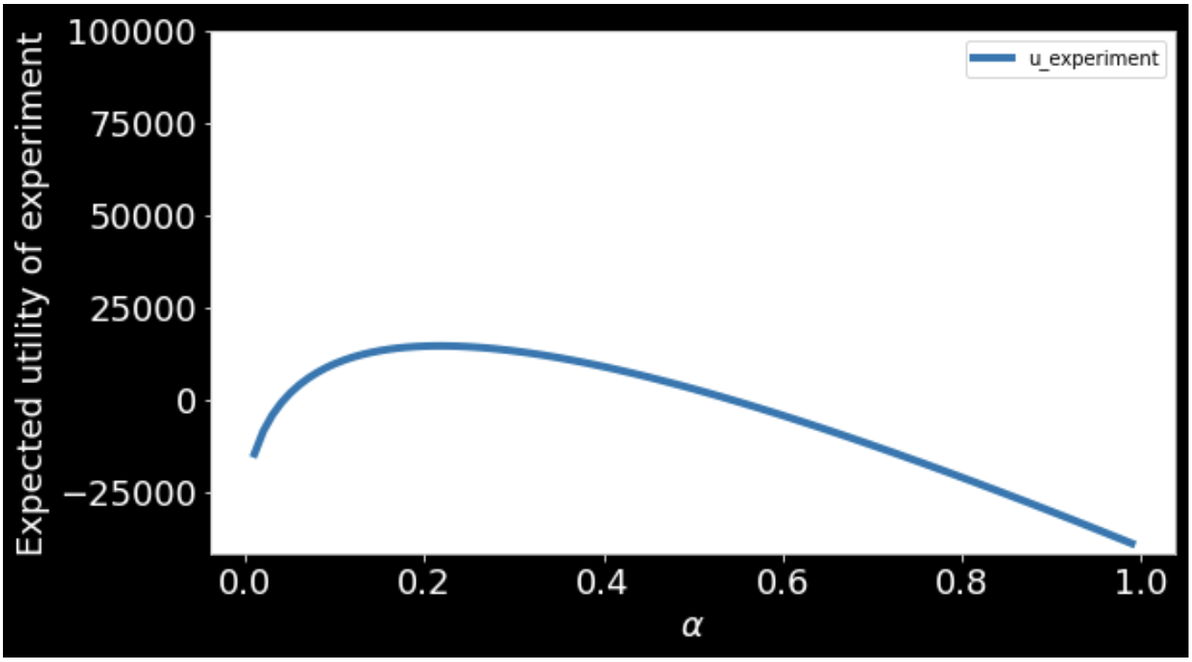
```
     Difference of proportion power calculation for binomial distribution (arcsine transformati

              h = 0.02000133
              n = 30908.66
      sig.level = 0.05
          power = 0.8
    alternative = greater

NOTE: same sample sizes
```

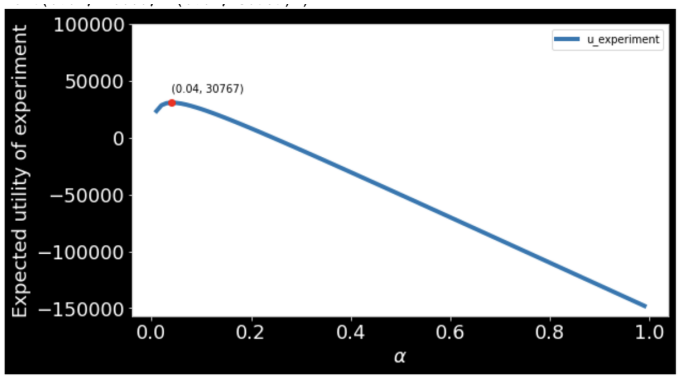


3. we can increase our sample size to have higher power with lower false positive rate; we can do pilot first to lower the 'p_null_hypo' (probability that null hypothesis is true); Besides, if we still have extra money, we can try to develop a more effective version of the new product, which can deliver higher conversion rate.
4. when P(null hypothesis) = 0.1, utility can be larger than zero, we will become willing to do the experiment.



Homework extension

1. Set p_null_hypo = 0.2.
   When alpha = 0.04, power = 0.86, we get the maximum utility = 30767

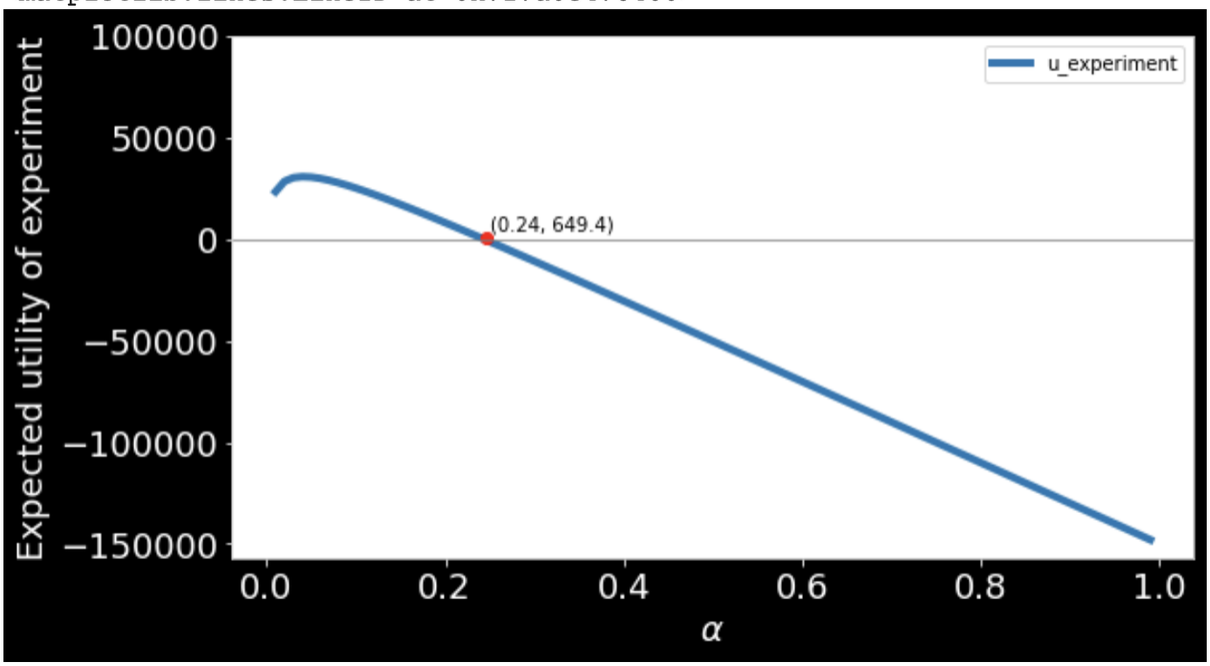



```
df_utility
```

	u_experiment	alpha	power
0	23396.761688	0.01	0.692460
1	28476.646451	0.02	0.780958
2	30282.727611	0.03	0.828534
3	30767.498587	0.04	0.859594
4	30548.674602	0.05	0.881858

2. when alpha is between 0.01 to 0.24 , the experiment utility is larger than zero that we will be willing to run the experiment.

```
<matplotlib.lines.Line2D at 0x7f7a0e476400>
```



	u_experiment	alpha	power
23	649.441478	0.24	0.983118
24	-1247.437252	0.25	0.984407

Bonus problem: Notebook section "Multiple treatments"

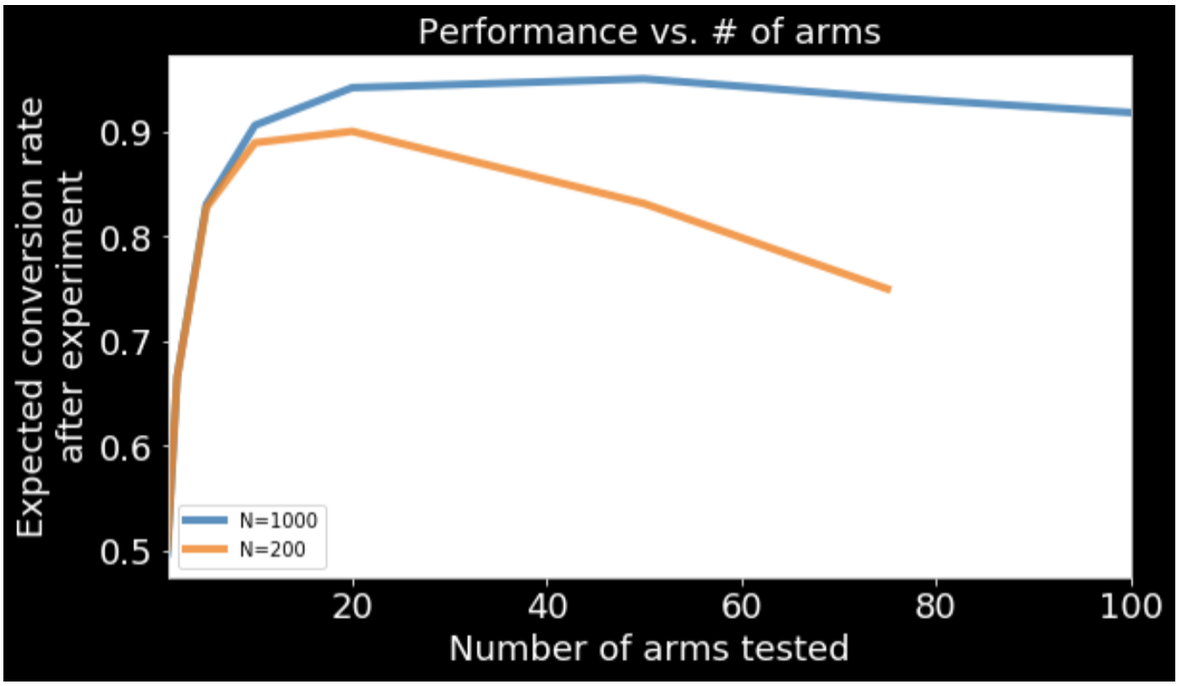
1. set 'alternative'=less: the hypothesis that the left column (control) has a lower probability than the right (treatment)

```
from rpy2.robjects import IntVector, default_converter
from rpy2.robjects.packages import importr
stats = importr('stats')
def stats_prop_test(e1, e2, n1, n2, alternative):
  x = IntVector([e1, e2])
  ns = [n1, n2]
  n = IntVector(ns)
  n_array = np.array([n1, n2])
```

```
  # Recommended to not use continuity correction for large sam
ple sizes:
  # https://en.wikipedia.org/wiki/Yates%27s_correction_for_con
tinuity
  result = stats.prop_test(x=x, n=n, alternative=less, correct
=False)
  z = result.rx('statistic')[0][0] ** 0.5
  r_estimate = result.rx('estimate')[0]
  estimates = np.array([r_estimate[0], r_estimate[1]])
  var = np.array(list(map(lambda p: p * (1 - p), estimates)))

  se = np.sqrt(np.divide(var, n_array))
  difference = estimates[1] - estimates[0]
  p_value = result.rx('p.value')[0][0]
  return {
      'estimates': estimates,
      'difference': difference,
      'se': se,
      'difference_se': np.sqrt((se ** 2).sum()),
      'z': z,
      'p_value': p_value
  }
```

2. The best conversion rate is higher than the results in the notebook, since when choosing the arms, it is more strict. For different N, the more arms tested, the better the best one will be. Splitting sample size across multiple arm increases variance of the estimates.



3. MAX(conversion rate * 1000 - 20 arms numbers)