Behavioral Econ Final Question 2

Wendy Shi

2023-11-30

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

In our analysis, we conducted a class experiment following in the format of the experiments conducted in the paper "Overconfidence and Excess Entry" by Colin Camerer and Dan Lovallo.

We will recreate the data displays table 2, 3, 4, 5 from the paper in our analysis.

First table 3 gives a general description of our experiment.

Sample	n	Selection_procedure	Rank_order
Duke, undergraduates	14	class students	R/S

Then we will analyze industry profit data by recreating tables 2 and 4. First we enter in data noting the round, capacity for each round, and actual number of entrants in the random round and the skill rounds. The first data is displayed here.

round	capacity	entrants_r	entrants_s
1	2	4	7
2	4	5	10
3	8	12	13
4	6	9	10
5	4	5	4
6	2	6	5
7	8	12	13
8	6	5	10
9	4	7	10
10	6	9	10
11	8	10	13
12	2	3	6

We then mutate to create variables in our data noting industry profit for random and skilled rounds separately. The variables are notated as "ind_profit_r" and "ind_profit_s".

round	capacity	entrants_r	entrants_s	ind_profit_r	ind_profit_s
1	2	4	7	30	0
2	4	5	10	40	-10
3	8	12	13	10	0
4	6	9	10	20	10
5	4	5	4	40	50
6	2	6	5	10	20
7	8	12	13	10	0
8	6	5	10	50	10
9	4	7	10	20	-10
10	6	9	10	20	10
11	8	10	13	30	0
12	2	3	6	40	10

In our analysis we found the following:

```
## [1] "industry profit is positive in 100% of the random rounds and 80% of the skilled rounds"
```

```
## [1] "average industry profit across rounds in random rounds is 26.67"
```

```
## [1] "average industry profit across rounds in skilled rounds is 7.5"
```

From these calculations, we conclude similar results as in the paper, there is indeed more entry (and lower industry profit) when people are betting on their own relative skill rather than on a random device. In the all of the random rank rounds, industry profit was positive, and it was positive in a smaller proportion, 80% of the skill ranked rounds. We also found consistent results with the paper of a difference in average industry profit of about 26.67 - 7.5 = 19.17, which means about 2 extra entrants per round in the skill conditions.

Even though our sample size is much smaller overall than in the paper (not enough to apply rule of 30 in CLT), we believe the differences follow a distribution not too far from the normal, and conduct a paired 2 tail t test. Note because of the smaller sample sizes, this result should be taken with caution.

The matched-pair t-test using these comparisons yields t = -3.53, p<0.01, which means we reject the null hypothesis that the industry profits are the same, and conclude there is evidence for a difference in the industry profits.

estimate	statistic	p.value	parameter	conf.low	conf.high	method	alternative
-19.16667	-3.529927	0.0047154	11	-31.11749	-7.21584	Paired t-test	two.sided

We also replicate table 5 in the paper, which tests for an alternative explanation of the blind spots hypothesis. Which suggests that excessive entry in skill conditions may be due to players underforecasting how many others will enter. We test this hypothesis by similar method to the paper, and look at the subject's forecast for how many entrants there will be and calculate the expected profit from entering. We then look at the difference in expected average profits in random rounds and the same statistic in skilled rounds using only the rounds in which a subject entered. A more detailed description of the method can be found in the paper.

In order to do this, we first enter in all of the data. The expected number of entrants in random rounds is denoted as "exp_num_entrants_r", and the expected number of entrants in skill rounds is denoted as "exp_num_entrants_s", there are two indicator variables denoting if the subject actually entered in that round noted as "entered_r" and "entered_s". An example of the entered dataframe is provided for person 4.

exp_num_entrants_r	entered_r	exp_num_entrants_s	entered_s
6	0	5	1
8	1	8	0

exp_num_entrants_r	entered_r	exp_num_entrants_s	entered_s
10	1	12	1
10	1	10	1
7	1	8	0
4	1	6	0
12	1	12	1
10	0	10	1
6	1	8	1
8	1	8	1
12	0	13	1
5	0	5	0

For each dataframe i, we calculate the expected profit the subject i thinks the average entrant will earn in each round for random and skill rounds and note that as the variable "exp_profit_r" and "exp_profit_s", we also calculate average expected profit for each subject for random vs skilled rounds (averaged for only rounds they entered). Another example of such a finished dataframe is provided, along with the dataframe showing the list of expected profit in random rounds and skilled rounds for the subjects.

exp_num_entrants_r	entered_r	exp_num_entrants_s	entered_s	capacity	exp_profit_r	exp_profit_s
6	0	5	1	2	1.67	4.00
8	1	8	0	4	1.25	1.25
10	1	12	1	8	3.00	0.83
10	1	10	1	6	1.00	1.00
7	1	8	0	4	2.86	1.25
4	1	6	0	2	7.50	1.67
12	1	12	1	8	0.83	0.83
10	0	10	1	6	1.00	1.00
6	1	8	1	4	5.00	1.25
8	1	8	1	6	3.75	3.75
12	0	13	1	8	0.83	0.00
5	0	5	0	2	4.00	4.00

expected_avg_profit_ifentering_r	expected_avg_profit_ifentering_s
1.56	1.28
0.85	1.34
2.90	1.94
3.15	1.58
3.16	1.71
0.00	0.00
5.62	1.67
5.71	2.50
2.39	1.95
3.08	0.73
1.61	1.36
3.05	1.50
3.29	2.54
3.12	2.21

We calculate the table 5 values, the first value being the mean difference between the average expected profit in random rounds and skilled rounds, and also conduct a t test on this value. The p value is less than 0.01, meaning we are able to reject the null that there is no difference between the average expected profit in random rounds and skilled rounds at the 0.05 significance level. Again, as above, because of smaller sample size, we take this result with caution.

```
## [1] 1.227143
```

estimate	statistic	p.value	parameter	conf.low	conf.high	method	alternative
1.227143	3.683135	0.0027573	13	0.5073537	1.946932	Paired t-test	two.sided

We also calculate the number of subjects for which the average random expected profit greater than average skilled expected profit, and calculate the percentage. We found this nubmer to be 12 out of 14, which is 86%. Lastly, we calculated the number of subjects with average skilled expected profit being negative, and found this was 0, and therefore 0%.

```
## [1] 12
```

```
## [1] 0.8571429
```

```
## [1] 0
```

```
## [1] 0
```

Our results are mostly consistent with the results of the papers, people are entering the market more during the skilled rounds even when they believe the expected profits from entering are lower than in the random rounds. Suggesting evidence that entrants are indeed overconfident in their skill.

The results of this paper are interesting, people tend to be overconfident in areas where they are more in control of their outcomes. This is seen in other world situations, for example people often think that driving is much safer than it actually is because we are in control ourselves (in an episode of freakonomics), however aggregate data suggests we are overconfident in our abilities to control the car.

We do note however, this is simply a class experiment in a limited setting, and our results, similar to as noted in the paper, need to be applied to real world market entry with a grain of salt.