

# Problem Set 2

Alex, Micah, and Scott

12/07/2020

```
library(data.table)
library(sandwich)
library(lmtest)

library(ggplot2)
library(knitr)
```

## 2. Sports Cards

In this experiment, the experimenters invited consumers at a sports card trading show to bid against one other bidder for a pair trading cards. We abstract from the multi-unit-auction details here, and simply state that the treatment auction format was theoretically predicted to produce lower bids than the control auction format. We provide you a relevant subset of data from the experiment.

In this question, we are asking you to produce p-values and confidence intervals in three different ways:

1. Using a `t.test`;
2. Using a regression; and,
3. Using randomization inference.

```
d <- fread('../data/list_data_2019.csv')
```

1. Using a `t.test`, compute a 95% confidence interval for the difference between the treatment mean and the control mean. After you conduct your test, write a narrative statement, using inline code evaluation that describes what your tests find, and how you interpret these results. (You should be able to look into `str(t_test_cards)` to find the pieces that you want to pull to include in your written results.)

```
t_test_cards <- t.test(d[uniform_price_auction==0,bid], d[uniform_price_auction==1,bid]) # this should
str(t_test_cards)
```

```
## List of 10
## $ statistic : Named num 2.82
##   .. attr(*, "names")= chr "t"
## $ parameter : Named num 62
##   .. attr(*, "names")= chr "df"
## $ p.value    : num 0.00642
## $ conf.int   : num [1:2] 3.56 20.85
##   .. attr(*, "conf.level")= num 0.95
## $ estimate   : Named num [1:2] 28.8 16.6
##   .. attr(*, "names")= chr [1:2] "mean of x" "mean of y"
## $ null.value : Named num 0
##   .. attr(*, "names")= chr "difference in means"
## $ stderr     : num 4.33
```

```
## $ alternative: chr "two.sided"
## $ method      : chr "Welch Two Sample t-test"
## $ data.name   : chr "d[uniform_price_auction == 0, bid] and d[uniform_price_auction == 1, bid]"
## - attr(*, "class")= chr "htest"
```

Answer: The 95% confidence interval is between 3.56 and 20.85. And we have the p-value of 0.006421, which is small enough to reject the null hypothesis that the bid price for treatment auction format is equal to the control auction format

2. In plain language, what does this confidence interval mean?

Answer: Confidence interval gives a range of values for an unknown parameter (for example, a population mean). For a given estimation in a given sample, using a higher confidence level generates a wider (i.e., less precise) confidence interval. In general terms, a confidence interval for an unknown parameter is based on sampling the distribution of a corresponding estimator. It means were this procedure to be repeated on numerous samples, the fraction of calculated confidence intervals (which would differ for each sample) that encompass the true population parameter would tend toward 95%

3. Conduct a randomization inference process, with `n_ri_loops = 1000`, using an estimator that you write by hand (i.e. in the same way as earlier questions). On the sharp-null distribution that this process creates, compute the 2.5% quantile and the 97.5% quantile using the function `quantile` with the appropriate vector passed to the `probs` argument. This is the randomization-based uncertainty that is generated by your design. After you conduct your test, write a narrative statement of your test results.

```
## first, do you work for the randomization inference

simulation_q3 <- function(d, num_simulations) {
  simulated_ate_list <- c()
  for (i in 1:num_simulations){
    d[ , simulation := sample(c(0,1), replace=TRUE, size=.N) ]
    simulated_ate <- d[ , .(mean_group = mean(bid)), keyby=simulation][ , diff(mean_group)]
    simulated_ate_list <- c(simulated_ate_list, simulated_ate)
  }
  return(simulated_ate_list)
}

n_ri_loops <- 1000

cards_ate <- d[ , .(mean_group = mean(bid)), keyby=uniform_price_auction][ , diff(mean_group)]
cards_ri_distribution <- simulation_q3(d, n_ri_loops) # numeric vector of length equal
# to your number of RI permutations
cards_ri_quantiles <- quantile(cards_ri_distribution, probs=c(2.5/100, 97.5/100)) # there's a built-in
cards_ri_p_value <- mean(abs(cards_ri_distribution) >= abs(cards_ate))
cards_ate

## [1] -12.20588
cards_ri_quantiles

##      2.5%      97.5%
## -9.513839  8.512125
cards_ri_p_value

## [1] 0.009
```

**Answer:** Because p-value of 0.009 is less than 0.05. We reject the Sharp Null Hypothesis that the treatment effect is zero for all subjects or observations. Thus, with a 95% confidence, the experiment's ATE of -12.2058824 is not likely to be a random result, and it is likely to be different from zero.

4. Do you learn anything different if you regress the outcome on a binary treatment variable? To answer this question, regress bid on a binary variable equal to 0 for the control auction and 1 for the treatment auction and then calculate the 95% confidence interval using *classical standard errors* (in a moment you will calculate with *robust standard errors*). There are two ways to do this – you can code them by hand; or use a built-in, `confint`. After you conduct your test, write a narrative statement of your test results.

```
mod <- lm(bid ~ uniform_price_auction, data = d) # this should be a model object, class = 'lm'.
confint(mod, level=0.95)
```

```
##                2.5 %    97.5 %
## (Intercept)      22.71534 34.931716
## uniform_price_auction -20.84416 -3.567603
```

**Answer:** The OLS based confidence interval produces the same values of confidence interval. Since zero is not included in the confidence interval, we can say that with 95% confidence the true ATE is located within this range (confidence interval).

5. Calculate the 95% confidence interval using robust standard errors, using the `sandwich` package. There is a function in `lmtest` called `coefci` that can help with this. It is also possible to do this work by hand. After you conduct your test, write a narrative statement of your test results.

```
library(sandwich)
library(lmtest)

m <- lm(bid ~ uniform_price_auction, data = d)
cards_robust_ci <- coefci(mod, delev = 0.95, vcov = vcovHC)
```

**Answer:** Using the `sandwich` package, we also see that the robust standard errors is very close to what we had calculated using other prior methods 6. Characterize what you learn from each of these different methods – are the results contingent on the method of analysis that you choose? **Answer:** Compared to part 3 above, there is a slight deviation in confidence interval around zero when using a randomization inference based method. However, the overall approach does not lead to any material difference in test results.