P8130_hw3_zl2974

Jeffrey Liang

10/13/2020

Problem 1 (30p)

A study was conducted over a six-month period at a local ambulatory virology clinic. The goal was to test the effect of a structured exercise program for overweight/obese, virally suppressed HIV positive subjects on different parameters. A total of 36 individuals agreed to participate in the intervention group (group 1) and another group of 36 individuals were selected as controls (group 0). The table below shows descriptive statistics: mean(SD), median(Q1, Q3) to summarize the Systolic Blood Pressure (SBP) variable by groups at baseline (pre), at 6 months follow-up (post) and also the absolute changes (Δ=Post-Pre). We want to perform some tests to assess changes in SBP for the two groups (within and between).

For each question, make sure to state the formulae for hypotheses, test-statistics, decision rules/p-values, and provide interpretations in the context of the problem. Use a type I error of 0.05 for all tests

```
##
##
## |month    |                   |          0 (N=72)          |          1 (N=72)          |
## |:--------|:------------------|:--------------------------:|:--------------------------:|
## |6_month  |systolic           |                            |                            |
## |         |-  Mean (SD)       |      130.139 (14.345)      |      125.056 (15.444)      |
## |         |-  Median (Q1, Q3) | 127.500 (120.000, 140.000) | 124.000 (116.750, 135.000) |
## |         |systolic_diff      |                            |                            |
## |         |-  Mean (SD)       |      -3.333 (14.813)       |      -8.583 (17.169)       |
## |         |-  Median (Q1, Q3) |  -3.500 (-12.250, 8.250)   |  -5.500 (-23.000, 3.000)   |
## |baseline |systolic           |                            |                            |
## |         |-  Mean (SD)       |      133.472 (15.938)      |      133.639 (15.108)      |
## |         |-  Median (Q1, Q3) | 131.000 (122.500, 143.500) | 134.000 (121.500, 144.000) |
## |         |systolic_diff      |                            |                            |
## |         |-  Mean (SD)       |      -3.333 (14.813)       |      -8.583 (17.169)       |
## |         |-  Median (Q1, Q3) |  -3.500 (-12.250, 8.250)   |  -5.500 (-23.000, 3.000)   |
```

a) Perform appropriate tests to assess if the Systolic BP at 6 months is significantly different from the baseline values for each of the groups:

   i) Intervention group (5p)
   ii) Control group (5p)

b) Now perform a test and provide the 95% confidence interval to assess the Systolic BP absolute changes between the two groups. (12p)

c) What are the main underlying assumptions for the tests performed in parts a) and b)? (3p)

   i) Use graphical displays to check the normality assumption and discuss the findings. (3p)
   ii) If normality is questionable, how does this affect the tests validity and what are some possible remedies? (2p)

PROOF

1) We use paired test for testing systolic BP difference within group of two measure point. Because the population variance is unknown, and we have no prior knowledge of how's systolic BP going to change in 6-month period so two-side paired t-test is used for this problem

a.

H_0: the difference between systolic value of baseline and 6 month later is the equal to 0(no difference) in intervention groups

H_1: the difference between systolic value of baseline and 6 month later is not the equal to 0(there's difference) in intervention groups

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = -8.583$$
$$s_d = \sqrt{\sum_{i=1}^{n} (d_i - \bar{d})^2/(n-1)} = 17.169$$
$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = -3 \sim t_{36-1}$$

With critical value $t_{36-1,1-\alpha/2} = 2.03$ and $t_{36-1,\alpha/2} = -2.03$, we reject the Null hypothesis and conclude that there's difference between baseline systolic BP value and 6-month later in intervention group.

a.

H_0: the difference between systolic value of baseline and 6 month later is the equal to 0(no difference) in control groups

H_1: the difference between systolic value of baseline and 6 month later is not the equal to 0(there's difference) in control groups

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = -3.333$$
$$s_d = \sqrt{\sum_{i=1}^{n} (d_i - \bar{d})^2/(n-1)} = 14.813$$
$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = -1.35 \sim t_{36-1}$$

With critical value $t_{36-1,1-\alpha/2} = 2.03$ and $t_{36-1,\alpha/2} = -2.03$, we can not reject the Null hypothesis and conclude that there's no difference between baseline systolic BP value and 6-month later in control groups.

2) Now that we compare two groups, and the population variance is unknown, we first test if the standard diviation of two groups is equal:

H_0: the variance between systolic difference of intervention group and control is the equal(no difference)

H_1: the variance between systolic value of intervention group and control is not the equal(there's difference)

$$s_{d_1} = \sqrt{\sum_{i=1}^{n_1} (d_i - \bar{d_1})^2/(n_1 - 1)} = 17.169$$
$$s_{d_0} = \sqrt{\sum_{j=1}^{n_0} (d_i - \bar{d_0})^2/(n_0 - 1)} = 14.813$$
$$F = s_1^2/s_0^2 = 1.343 \sim F_{n_1-1,n_0-1}$$

With F-test critical value of 1.961, we cannot reject the null hypothesis that two sample variance is different. And we have no idea how's the intervention going to change the systolic value between groups, so two-side equal variance t-test is used.

H_0: the difference between systolic difference of intervention group and control is the equal to 0(no difference)

H_1: the difference between systolic value of intervention group and control is not the equal to 0(there's difference)

$$\bar{X_1} - \bar{X_0} = -8.583 - -3.333 = -5.25$$

$$\begin{aligned} s_{pool} &= \frac{(n_1 - 1)s_1 + (n_0 - 1)s_0}{n_1 + n_0 - 2} \\ &= \frac{(36 - 1) * 17 + (36 - 1) * 15}{36 + 36 - 2} \\ &= 15.991 \end{aligned}$$

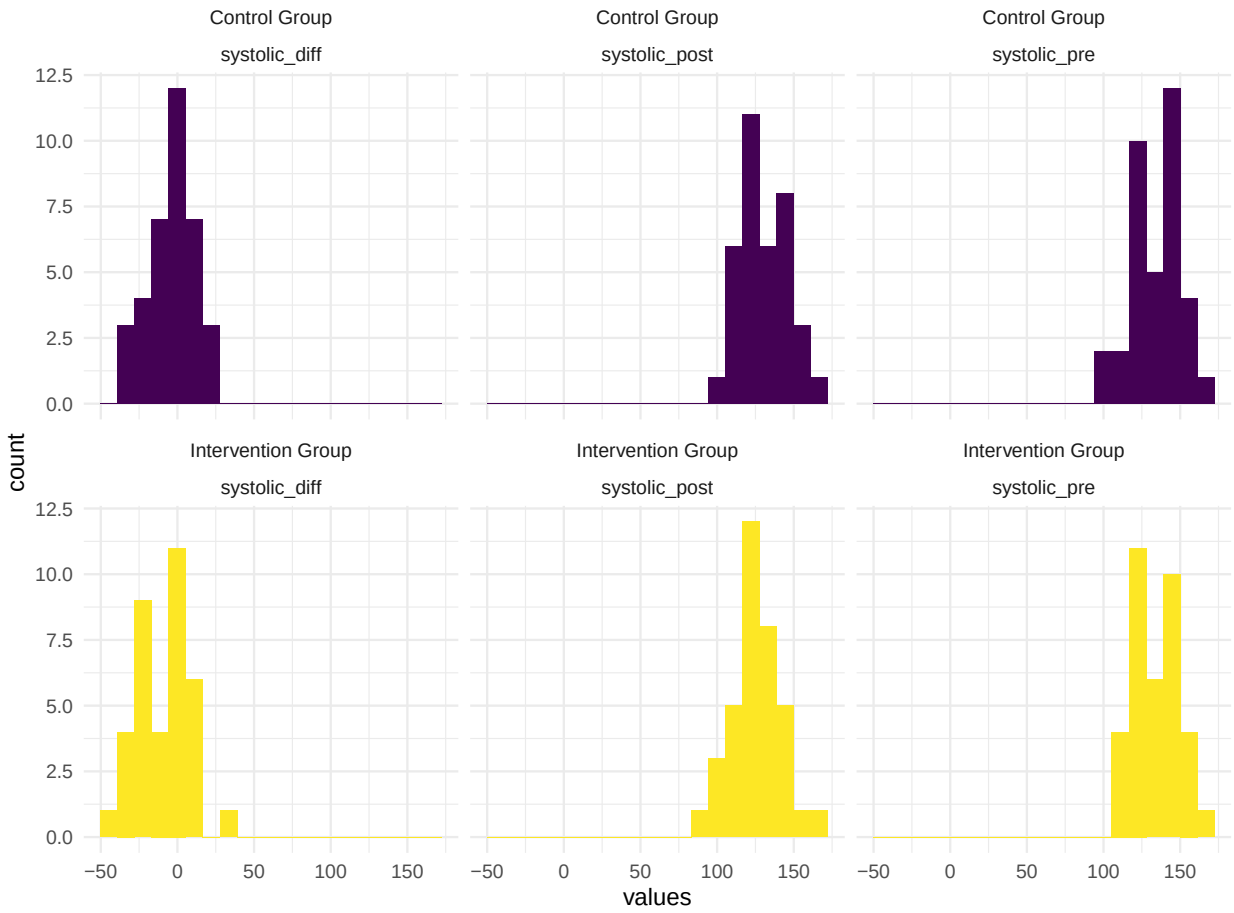
$$\begin{aligned} t &= \frac{\bar{X_1} - \bar{X_0}}{s_{pool} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_0}}} \\ &= -1.389 \sim t_{36+36-2} \end{aligned}$$

According to the critical value we choose(-1.994), we cannot reject H_0 that there's no difference between the difference of baseline systolic BP and 6 month later in intervention group and control group.

3) The testing in 1),2) are based on that sample is draw from a normal distributed population or the distributions fit normality. And thus, if we used these test method upon a skewed(not normal) sample, the result of our interpretation can be wrong. As ploted as follow, the sample distribution satisfied normality requirement for testing.

   But if the sample is skewed, there're:

   - By central limit theorem, if our sample size is large enough, we will have a sample distribution approximate normal distribution
   - Use normal transformation: we have methods(eg. logorithm or square root) to transform our sample distribution to approximate normal.
   - Use bayesian inference or other statistics inference methods that don't rely on Normality assumption of the distribution.



Problem 2 (25p)

We have discussed the fact that we are not guaranteed to make the correct decision by the process of hypothesis testing and there is always some level of uncertainty in statistics. The two main errors that we are trying to minimize/control are type I and type II. A type I error occurs when we reject the null hypothesis H_0 , when H_0 is true. When we set the significance level at 5%, we are saying that we will allow ourselves to make a type I error less than 5% of the time. In practice we can only calculate this probability using a series of "what if" calculations, because we do not really know the truth. In this exercise you learn how to create your own 'true' scenario, simulate corresponding data, and quantify the type I error over many repetitions.

Scenario: The average IQ score of Ivy League colleges is 120. We will assume this to be the null hypothesis (true mean is 120) with a standard deviation of 15 and a significance level of 5%. For the alternative hypothesis we will consider that the 'true mean is less than 120'.

Most of the time (95%) when we generate a sample from the underlying true distribution, we should fail to reject the null hypothesis since the null hypothesis is true. Let us test it!

a) Generate one random sample of size n=20 from the underlying (null) true distribution. Calculate the test statistic, compare to the critical value and report the conclusion: 1, if you reject H_0 or 0, if you fail to rejected H_0 . (5p) Hint: use rnorm(20, mean = 120, sd = 15)

b) Now generate 100 random samples of size n = 20 from the underlying (null) true distribution and repeat the process in part (a) for each sample (calculate the test statistic, compare to the critical value, and record 1 or 0 based on criteria above). Report the percentage of 1s and 0s respectively across the 100 samples. The percentage of 1s represents the type I error. (7.5p) Suggestions: use a for loop to

loop over the 100 samples and create a variable using the function ifelse() to keep track of your 1's and 0's.

c) Now generate 1000 random samples of size n = 20 from the underlying (null) true distribution, repeat the same process, and report the percentage of 1s and 0s across the 1000 samples. (7.5p)

d) Final conclusions: compare the type I errors (percentage of 1s) from part b) and c). How do they compare to the level that we initially imposed (i.e. 0.05)? Comment on your findings. (5p)

PROOF

a) I have write 2 function to approach this question, the first function *null_sim* is function that take sample size and repeated times to generate n times the n size random sample draw from the $\mathcal{N}(120, 15)$. And the second function I wrote is *z_test*, which take the parameter of list of value generate from *null_sim* and then produce test statistics and Hypothesis test result.

```
# FUNCTION 1
null_sim = #generate random sample under H0
  function(sample_size = 20,
           repeat_time = 10) {
    #collect random sample
    return_value = list()
    repeat {
      return_value =
        append(return_value,
               list(rnorm(sample_size, 120, 15))) #random sample generate
      if (length(return_value) >= repeat_time) {
        break
      }
    }
    #return collect random sample as list
    return(return_value)
  }

# FUNCTION 2
z_test = function(x = NA,
                  #take the list from null_sim
                  mu = 0,
                  sample_variance = NA,
                  method = "less",
                  alpha = 0.05) {
  x_mean = c()
  x_se = c()
  x_zscore = c()
  x_p = c()
  x_H0 = c()
  if (is.list(x)) {
    #for multi-simulation
    for (i in seq(1, length(x), 1)) {
      x_mean = append(x_mean, mean(x[[i]]))
      x_se = append(x_se, sample_variance / sqrt(length(x[[i]])))
      x_zscore =
        append(x_zscore, (x_mean[[i]] - mu) / x_se[[i]])
      x_p = append(x_p, 1 - pnorm(abs(x_zscore[[i]])))
      x_H0 = append(x_H0, x_zscore[[i]] < qnorm(alpha))
    }
```

```
  } else{
    #for one row vector
    x_mean = mean(x)
    x_se = sample_variance / sqrt(length(x))
    x_zscore = (x_mean - mu) / x_se
    x_p = 1 - pnorm(abs(x_zscore))
    x_H0 = x_zscore < qnorm(alpha)
  }
  #return the test result as tibble for better statistics summary
  return(tibble(zscore = x_zscore,
                #p_value = x_p, #Future features
                Reject_H0 = x_H0))
}
```

Noted that the population mean and variance are known, and see that the alternative hypothesis is "the true mean is less than 120", so One-side test is perform.

H_0: the mean the equal to 120 (no difference)

H_1: the mean is not the equal to 120(there's difference)

The test-statistics is calculate via:

$$z\ score = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

we have chosen the Type-I error rate of 0.05, so the critical value for the test statistics is -1.645. Any test statistics fall below this value and we reject the Null Hypothesis and return *TRUE* value to columns name *Reject_H0*.

Using the two function, we have generate 1 sample of sample size 20 and conducted hypothesis testing, the result to the problem:

```
samp = null_sim(repeat_time = 1)

print(samp)
```

```
## [[1]]
##  [1] 101.9 124.2 136.3  84.8 126.4 127.6 111.4 111.8 111.5 106.6 112.8 105.0
## [13] 108.4 121.0 134.4 118.3 112.3 106.3 107.4 156.2
```

```
tibble(value = as.vector(samp[[1]])) %>%
  summarise(mean = mean(value),
            standard_deviate = sd(value))
```

```
## # A tibble: 1 x 2
##    mean standard_deviate
##   <dbl>            <dbl>
## 1  116.             15.2
```

```
sim_1 = z_test(samp, mu = 120, sample_variance = 15)

sim_1 %>% janitor::tabyl(Reject_H0)
```

```
##  Reject_H0 n percent
##      FALSE 1       1
```

```
## We fail to reject H0 because the test statistics calculated is
##  -1  and fall above critical value
```

Using the same method, I generate the result for the following question:

b)

```
##  Reject_H0  n percent
##      FALSE 94    0.94
##       TRUE  6    0.06
```

So under the 100 times simulation, we make 6 times mistakes for rejecting H0 while the H0 is indeed correct.

$P(Type\ I\ error \mid H_0) = \frac{number\ of\ time\ we\ reject\ H_0\ while\ H_0\ is\ TRUE}{number\ of\ trails\ under\ H_0} = 0.06$

c)

```
##  Reject_H0   n percent
##      FALSE 953   0.953
##       TRUE  47   0.047
```

So under the 1000 times simulation, we make 47 times mistakes for rejecting H0 while the H0 is indeed correct.

$P(Type\ I\ error \mid H_0) = \frac{number\ of\ time\ we\ reject\ H_0\ while\ H_0\ is\ TRUE}{number\ of\ trails\ under\ H_0} = 0.047$

d) After the simulation from above, we have $P(Type\ I\ error \mid H_0)$ of 0.06 and 0.047 for 100 and 1000 times simulation. Both figures are not exactly equal but close to the set alpha rate of the setting, and the result from 1000 times simulation is closer compared to 100 times.

If we set the sample size to 100 and do a 10000 times simulation we still have 0.052 of $P(Type\ I\ error \mid H_0)$, which is still around α but not to 0. In fact, the $P(Type\ I\ error \mid H_0)$ in these simulation is evaluate as the $\widehat{p}$ of proportion testing, so as the simulation times(the sample size to estimate α) increase, the standard error of the estimate decrease and the estimate is closer to the true value(α).

So the only method to lower $P(Type\ I\ error \mid H_0)$ is to set the α rate to a smaller value before testing.

Eg. 1000 simulation of 20 sample size with alpha = 0.01

```
##  Reject_H0   n percent
##      FALSE 986   0.986
##       TRUE  14   0.014
```