PSTAT 120C HW 5

TJ Sipin

2022-09-09

Reading

In your own words, compare and contrast parametric and nonparametric statistcs.

Parametric statistics are used when general assumptions hold. Nonparametric statistics are used in situations in which specific assumptions do not hold or are not particularly certain to be held. They can also be useful for when studying the effects on data that aren't intuitively mathematical, such as those that are subjective.

Additionally, parametric statistics are used to apply to problems in which the the sample distributions are specified, except for the values of a finite number of parameters. Nonparametric statistics is used when they are not specified.

What are two reasons why researchers might choose a nonparametric test over its parametric equivalent?

It may be useful if assumptions do not hold or may not hold, such as constant variance. They are likely to be more powerful in detecting population differences when the assumptions are not satisfied.

Additionally, nonparametric statistics can be used for observations that are difficult to quantify.

List the nonparametric tests we've discussed that are appropriate for paired samples.

- Sign test
- Wilcoxon signed-rank test

List the nonparametric tests we've discussed that are appropriate for independent samples.

- Sign test
- Wilcoxon signed-rank test
- Mann-Whitney U
- Kruskal-Wallis test
- Friedman test

What assumptions do most nonparametric tests appear to make?

- Independence
- Number of measurements in the sample from population i must be greater than or equal to 5.

Practice

1. The following data record the number of ear infections each of 7 swimmers had before (X_i) taking a medication designed to prevent infections and after (Y_i) taking the medication. Using the data provided, answer the following questions:

Swimmer	X_i	Y_i	D_i	Sign
A	3	2		
B	0	1		
C	5	4		
D	4	0		
E	2	1		
F	4	3		
G	3	1		

a. Consider $D_i = X_i - Y_i$. Fill in the D_i column. Then fill in the "Sign" column with the sign of the differences.

Swimmer	X_i	Y_i	D_i	Sign
A	3	2	1	+

B	0	1	-1	-
C	5	4	1	+
D	4	0	4	+
E	2	1	1	+
F	4	3	1	+
G	3	1	2	+

b. Use the *sign test* to conduct a hypothesis test assessing whether there is a difference between the population distributions of X and Y.

- Write the null and alternative hypotheses in words and in symbols.;

  $H_0 : p = 0.50$, where $p = P(X > Y)$, or it is equally likely for X to be greater than Y as it is for Y to be greater than X.

- Calculate a test statistic M;

  $M = $ # of positive differences where $D_i = X_i - Y_i$.

$$M = 6$$

- Calculate the relevant binomial probability;

$$P(M \geq 6) = \binom{7}{6}0.5^6 0.5^1 + \binom{7}{7}0.5^7 0.5^0$$
$$= 7(0.5^7) + 1(0.5^7)$$
$$= 0.0625$$
$$= \frac{1}{16}$$
$$P(M \leq 1) = \binom{7}{1}0.5^1 0.5^6 + \binom{7}{0}0.5^7$$
$$= 7(0.5^7) + 0.5^7$$
$$= 0.0625$$
$$= \frac{1}{16}$$

So the two-tailed p-value for $M = 6, n = 7$ is $\frac{1}{8} = 0.125$.

- Can you reject the null hypothesis, assuming $\alpha = 0.05$?

  Since $p = 0.125$, we would fail to reject since $p > 0.05$.

c. Use the Wilcoxon signed-rank test to conduct a hypothesis test assessing whether there is a difference between the population distribution X and Y.

- Calculate a test statistic T;

$$T^+ = 1.2 + 1.2 + 7 + 1.2 + 1.2 + 6 = 17.8$$
$$T^- = 1.2$$
$$T = \min(17.8, 1.2) = 1.2$$

It could also be that

$$T^+ = 3 + 3 + 7 + 3 + 3 + 6 = 25$$
$$T^- = 3$$
$$T = \min(17.8, 3) = 3$$

- Determine the appropriate critical value for $\alpha = 0.05$ (or as close as possible to 0.05);

  Since $n = 7$, our critical value is 2.

- Can you reject the null hypothesis?

  We reject if $T \leq T_0 = 2$. In the case that $T = 1.2$, then we reject the null hypothesis and conclude that there is a difference in ear infections among swimmers before taking medication and after. In the case that $T = 3$, we fail to reject the null hypothesis.

d. Were you able to reject the null with one test, both tests, or neither? If you rejected with one test and not the other, what do you think contributed to this difference?

We rejected the null hypothesis for the Wilcoxon signed-rank test. For the sign test, we are looking purely at the sign of the difference, rather than the ranks, so there is a sense of generality and less information for the former. This may have led to less inferential power, leading us to fail to reject since we lacked evidence to reject, as opposed to the Wilcoxon signed-rank test.

2. The coded values for a measure of brightness in paper (light reflectivity), prepared by two different processes, are as shown in the accompanying table for samples of size **9** drawn randomly from each of the two processes. Do the data present sufficient evidence to indicate a difference in locations of brightness measurements for the two processes? Give the attained significance level.

Process A	Process B
6.1	9.1
9.2	8.2
8.7	8.6
8.9	6.9
7.6	7.5
7.1	7.9
9.5	8.3
8.3	7.8
9.0	8.9

a. Answer the question by using the **Mann-Whitney U test**.

Doing this by hand would be slightly tedious, so instead we work in R. However, We use the following equation for our test statistic U:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W$$
$$= 9(9) + \frac{9(10)}{2} - W$$
$$= 81 + 45 - W$$

```
proc_a <- c(6.1, 9.2, 8.7, 8.9, 7.6,
            7.1, 9.5, 8.3, 9.0)
proc_b <- c(9.1, 8.2, 8.6, 6.9, 7.5,
            7.9, 8.3, 7.8, 8.9)
total <- c(proc_a, proc_b)
sort(total)
```

```
##  [1] 6.1 6.9 7.1 7.5 7.6 7.8 7.9 8.2 8.3 8.3 8.6 8.7 8.9 8.9 9.0 9.1 9.2 9.5
```

Using the `sort()` function, we find that $W = 1 + 3 + 5 + 9.5 + 12 + 13.5 + 15 + 17 + 18 = 94$, so $U = 32$. Using Table 8, Appendix 3 in the book, we find that $U_0 = 21$. Since $U = 32 \nleq U_0 = 21$, we fail to reject the null using $\alpha = 0.10$.

```
wilcox.test(proc_b, proc_a)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  proc_b and proc_a
## W = 32, p-value = 0.4795
## alternative hypothesis: true location shift is not equal to 0
```

This is in line with the `wilcox.test()` function. Our p-value is higher than 0.10 at 0.4795, so we fail to reject the null. Thus, we conclude that there is not sufficient evidence to indicate no difference in locations of brightness measurements for the two processes.

b. Answer the question by using the independent samples $t - test$.

We calculate the standard deviation S_p for the independent sample t-test by using the formula:

$$S_p = \sqrt{\frac{\sum(X_{1i} - \bar{X}_1)^2 + \sum(X_{2i} - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

```
sse_a <- 0
for (i in proc_a) {
  sse_a <- sse_a + (i - mean(proc_a))^2
}
sse_b <- 0
for (i in proc_b) {
  sse_b <- sse_b + (i - mean(proc_b))^2
}
```

We find that

$$S_p = \sqrt{\frac{10.02 + 3.86}{9 + 9 - 2}} = 0.9313968.$$

We calculate the value of the independent sample t-test by using the formula

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p}\sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 0.304.$$

At the 0.10 significance level for this two-tailed t-test, our critical value is 1.746 with 16 degrees of freedom. Since $t = 0.304 < t_0 = 1.746$, we fail to reject the null hypothesis. We can check this in R:

```
t.test(proc_a, proc_b)
```

```
##
##   Welch Two Sample t-test
##
## data:  proc_a and proc_b
## t = 0.30368, df = 13.367, p-value = 0.7661
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -0.8125664  1.0792331
## sample estimates:
## mean of x mean of y
##   8.266667  8.133333
```

Since our p-value is 0.766, which is higher than 0.10, we fail to reject the null hypothesis that there is no difference in locations of brightness measurements for the two processes.

    c. Provide the null and alternative hypotheses for parts *a* and *b*, along with any assumptions made for each test.

H_0 : there is no difference in locations of brightness measurements for the two processes.

H_a : there is a difference in locations of brightness measurements for the two processes.

The assumptions made for Mann-Whitney U are that we need a continuous or ordianl outcome variable, one independent variable with two groups, data that are similar in shape across groups, random sampling, independence, and sufficient sample size.

The assumptions made for the independent sample t-test are that the data must be moderately normal and that n_1 and n_2 must be equal (or nearly equal).

    3. Consider the Friedman statistic

$$F_r = \frac{12b}{k(k+1)}\sum_{i=1}^{k}(\bar{R}_i - \bar{R})^2.$$

Square each term in the sum, and show that an alternative form of F_r is

$$F_r = \frac{12}{bk(k+1)}\sum_{i=1}^{k}R_i^2 - 3b(k+1).$$

Hint: Recall that $\bar{R}_i = \frac{R_i}{b}$ and $\bar{R} = \frac{k+1}{2}$, and note that $\sum_{i=1}^{k}R_i = \frac{bk(k+1)}{2}$.

$$F_r = \frac{12b}{k(k+1)}\sum_{i=1}^{k}(\bar{R}_i - \bar{R})^2$$

$$= \frac{12b}{k(k+1)}\sum_{i=1}^{k}(\bar{R}_i^2 - 2\bar{R}_i\bar{R} + \bar{R}^2)$$

$$= \frac{12b}{k(k+1)}\left[\sum_{i=1}^{k}\frac{R_i^2}{b^2} - 2\frac{(k+1)}{2}\sum_{i=1}^{k}\bar{R}_i + \sum_{i=1}^{k}\bar{R}^2\right]$$

$$= \frac{12b}{k(k+1)}\sum_{i=1}^{k}R_i^2 - \frac{12b}{k(k+1)}\frac{(k+1)}{b}\frac{bk(k+1)}{2} + \frac{12b}{k(k+1)}\frac{k(k+1)(k+1)}{4}$$

$$= \frac{12\sum_{i=1}^{k}R_i^2}{bk(k+1)} - 6b(k+1) + 3b(k+1)$$

$$= \frac{12}{bk(k+1)}\sum_{i=1}^{k}R_i^2 - 3b(k+1)$$

    4. A quality control chart has been maintained for a measurable characteristic of items taken from a conveyor belt at a fixed point in a production line. The measurements obtained today, in order of time from top left to bottom right (68.2 the earliest recorded and 70.1 the latest), are as follows:

$$68.2 \quad 71.6 \quad 69.3 \quad 71.6 \quad 70.4 \quad 65.0 \quad 63.6 \quad 64.7$$
$$65.3 \quad 64.2 \quad 67.6 \quad 68.6 \quad 66.8 \quad 68.9 \quad 66.8 \quad 70.1$$

(a) Classify the measurements in this time series according to whether each is above or below the sample mean and determine, using the **runs test**, whether there are runs of high or low measurements, suggesting a lack of stability in the production process.

The sample mean is 67.66875. This means that our sample can be represented as

$$H \quad H \quad H \quad H \quad H \quad L \quad L \quad L$$
$$L \quad L \quad L \quad H \quad L \quad H \quad L \quad H$$

Our test statistic R is the number of runs, so $R = 7$. n_1 is the number of high elements, so $n_1 = 8$ and n_2, the number of low elements is also 8. There are $Y_1 = 4$ runs of highs and $Y_2 = 3$ runs of lows. The maximum possible number of runs m is $2n_1 = 16$ since $n_1 = n_2$. Getting as close to $\alpha = 0.05$ as possible without exceeding it, we find that our critical values are 4 and 13 since $\mathbb{P}(R \leq 3) = 0.009$ and $\mathbb{P}(R \leq 13) = 0.991$. We would reject the hypothesis that the production process randomly produces high and low measurements if $R \leq 3$ or $R \geq 14$ at the $\alpha = 0.018$ significance level. Since $R = 7$, we fail to reject the null hypothesis, so we conclude that there is not enough evidence to say that there is a lack of stability in the production process.

b. Divide the time period into two equal parts and compare the means of each, using the *t-test*, with $\alpha = 0.05$. Do the data provide evidence of a shift in the mean level of the quality characteristic? Explain.

We divide the time period by splitting between the two rows. We call R_1 the first row and R_2 the second. So the means $\bar{R}_1 = 68.05$ and $\bar{R}_2 = 67.2875$.

We find S_p:

$$S_p = \sqrt{\frac{\sum(X_{1i} - \bar{X}_1)^2 + \sum(X_{2i} - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$
$$= 2.66$$

```
R_1 <- c(68.2, 71.6, 69.3, 71.6, 70.4, 65.0, 63.6, 64.7)
R_2 <- c(65.3, 64.2, 67.6, 68.6, 66.8, 68.9, 66.8, 70.1)
sum_R_1 <- 0
sum_R_2 <- 0
for (i in R_1) {
  sum_R_1 = sum_R_1 + (i - mean(R_1))^2
}
for (i in R_2) {
  sum_R_2 = sum_R_2 + (i - mean(R_2))^2
}
S_p = sqrt((sum_R_1 + sum_R_2)/(8 + 8 - 2))
```

We find the test statistic t:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p}\sqrt{\frac{n_1 n_2}{n_1 + n_2}} = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.7625}{2.658259\sqrt{\frac{2}{8}}} = 0.5736838.$$

At the $\alpha = 0.05$ significance level, our critical value $T_0 = 2.145$, which is greater than our test statistic, so we fail to reject the null hypothesis, so there is a shift in the mean level of the quality characteristic.

```r
knitr::opts_chunk$set(echo = TRUE, warning = FALSE, message = FALSE, eval = T)

library(dplyr)
library(knitr)
library(kableExtra)
proc_a <- c(6.1, 9.2, 8.7, 8.9, 7.6,
            7.1, 9.5, 8.3, 9.0)
proc_b <- c(9.1, 8.2, 8.6, 6.9, 7.5,
            7.9, 8.3, 7.8, 8.9)
total <- c(proc_a, proc_b)
sort(total)
wilcox.test(proc_b, proc_a)
sse_a <- 0
for (i in proc_a) {
  sse_a <- sse_a + (i - mean(proc_a))^2
}
sse_b <- 0
for (i in proc_b) {
  sse_b <- sse_b + (i - mean(proc_b))^2
}
t.test(proc_a, proc_b)
R_1 <- c(68.2, 71.6, 69.3, 71.6, 70.4, 65.0, 63.6, 64.7)
R_2 <- c(65.3, 64.2, 67.6, 68.6, 66.8, 68.9, 66.8, 70.1)
sum_R_1 <- 0
sum_R_2 <- 0
for (i in R_1) {
  sum_R_1 = sum_R_1 + (i - mean(R_1))^2
}
for (i in R_2) {
  sum_R_2 = sum_R_2 + (i - mean(R_2))^2
}
S_p = sqrt((sum_R_1 + sum_R_2)/(8 + 8 - 2))
```