Conducting and Interpreting t-Tests

Alternative tests for failed assumptions

Jenine Harris
Brown School

Import and clean the data

```
# import nhanes 2015-2016
nhanes.2016 <- read.csv("/Users/harrisj/Box/teaching/Teaching/Fall2020/d

# add labels to sex and rename variables
nhanes.2016.cleaned <- nhanes.2016 %>%
  mutate(RIAGENDR = recode_factor(.x = RIAGENDR,
                                  `1` = 'Male',
                                  `2` = 'Female')) %>%
  rename(sex = RIAGENDR) %>%
  rename(systolic = BPXSY1) %>%
  rename(systolic2 = BPXSY2) %>%
  mutate(diff.syst = systolic - systolic2)

# check the data
summary(object = nhanes.2016.cleaned)
```

```
##       SEQN           cycle              SDDSRVYR     RIDSTATR      sex
##  Min.   :83732   Length:9544        Min.   :9    Min.   :2    Male  :4676
##  1st Qu.:86222   Class :character   1st Qu.:9    1st Qu.:2    Female:4868
##  Median :88726   Mode  :character   Median :9    Median :2
##  Mean   :88720                      Mean   :9    Mean   :2
##  3rd Qu.:91210                      3rd Qu.:9    3rd Qu.:2
##  Max.   :93702                      Max.   :9    Max.   :2
##
##     RIDAGEYR        RIDAGEMN        RIDRETH1        RIDRETH3        RIDEXMON
##  Min.   : 0.00   Min.   : 0.00   Min.   :1.00    Min.   :1.000   Min.   :1.0
##  1st Qu.: 9.00   1st Qu.: 5.00   1st Qu.:2.00    1st Qu.:2.000   1st Qu.:1.0
```

Identifying and using alternate tests for when t-test assumptions are not met

The alternate tests to use when assumptions are failed:

- one-sample t-test -> Sign test
- paired-samples t-test -> Wilcoxon Signed-Rank Test
- independent-samples t-test -> Mann-Whitney U or Kolmogorov-Smirnoff

Alternative to one-sample t-test failing assumptions: Sign Test

- When the data failed the assumption of normality for a one-sample t-test, the median could be examined rather than the mean, just like for descriptive statistics when the variable is not normally distributed.

- The **Sign Test** tests whether the **median** of a variable is equal to some hypothesized value.

- Before conducting a Sign Test, conduct EDA to examine the median value of the `systolic` variable.

```
# examine median for systolic variable
median(nhanes.2016.cleaned$systolic, na.rm = TRUE)
```

```
## [1] 118
```

NHST Step 1: Write the null and alternate hypotheses

- H0: The median systolic blood pressure in the US population is 120.
- HA: The median systolic blood pressure in the US population is not 120.
- The median systolic blood pressure is 118.
- This is close but a little lower than the 120 hypothesized to be the median value.

NHST Step 2: Compute the test statistic

- The *Sign Test* is conducted using the `SIGN.test()` function from the BSDA package.
- Add the `md = 120` argument to the code since the `SIGN.test()` can be used in other settings aside from one-sample median tests.
- The `md =` option in the `SIGN.test()` indicates the hypothesized value to test.

```
# compare observed median SBP to 120
BSDA::SIGN.test(nhanes.2016.cleaned$systolic, md = 120)
```

```
##
##      One-sample Sign-Test
##
## data:  nhanes.2016.cleaned$systolic
## s = 3004, p-value < 2.2e-16
## alternative hypothesis: true median is not equal to 120
## 95 percent confidence interval:
##  116 118
## sample estimates:
## median of x
##         118
##
## Achieved and Interpolated Confidence Intervals:
```

NHST Step 3: Compute the probability for the test statistic (p-value)

- The p-value is shown in scientific notation in the output as `< 2.2e-16` which is well below .05.

NHST Steps 4 & 5: Interpret the probability and write a conclusion

- The probability is extremely low of finding a median systolic blood pressure of 118 in the sample if the population had a median systolic blood pressure of 120.
- The output also includes a 95% confidence interval of 116 to 118, which means that this sample likely came from a population where the median systolic blood pressure was between 116 - 118.
- The median in the sample is 118 and the median in the population is likely between 116 - 118.
- Full interpretation: The median systolic blood pressure for NHANES participants was 118. A Sign Test comparing the median to a hypothesized median of 120 had a statistically significant ($s = 3004$; $p < .05$) result. The sample with a median systolic blood pressure of 118 was unlikely to have come from a population with a median systolic blood pressure of 120. The 95% confidence interval indicates this sample likely came from a population where the median systolic blood pressure was between 116 and 118. This suggests that the median systolic blood pressure in the US population is between 116 and 118.

Alternative when the paired-samples t-test fails assumptions: Wilcoxon Signed-Ranks Test

- The Wilcoxon Signed-Rank Test is an alternative to the *paired-samples t-test* when the continuous variable is not normally distributed.
- The Wilcoxon test determines if the differences between paired values of two related samples is symmetrical around zero.
- That is, instead of comparing the mean difference to zero, the test compares the distribution of the differences around zero.
- It appeared relatively evenly distributed around zero.

The Wilcoxon signed-ranks test

- The steps used to compute the test statistic:
  - Step 1: Find the differences between the two paired measures (measure 1 - measure 2)
  - Step 2: Put the *absolute values* of the differences in order from smallest to largest and give each one a rank
  - Step 3: Sum the ranks for all the *positive* differences
  - Step 4: Sum the ranks for the *negative* differences
- The test statistic is the smaller of the *Step 3* and *Step 4* values. If the sum of the ranks of all the *positive* differences is smaller, that sum is W.
- If the sum of the ranks of the *negative* values is smaller, that sum is W.
- The distribution of W is approximately normal when the sample is more than 20, which is much of the time.
- Because it approximates a normal distribution, a z-statistic is used to test whether the W is statistically significant.
  - The z-statistic has the standard cutoff values of -1.96 and 1.96 for statistical significance ($\alpha = .05$; $p < .05$).

NHST Step 1: Write the null and alternate hypotheses

H0: The distribution of the difference between the systolic blood pressure measures taken at time 1 and time 2 in the US population is symmetric around zero.

HA: The distribution of the difference between the systolic blood pressure measures taken at time 1 and time 2 in the US population is not symmetric around zero.

NHST Step 2: Calculate the test statistic

- The `wilcox.test()` function in base R can be used to test this null hypothesis.
- Include the `paired = TRUE` argument in the code.

```
# test the distribution of SBP by time period
wilcox.test(x = nhanes.2016.cleaned$systolic,
            y = nhanes.2016.cleaned$systolic2,
            paired = TRUE)
```

```
##
##      Wilcoxon signed rank test with continuity correction
##
## data:  nhanes.2016.cleaned$systolic and nhanes.2016.cleaned$systolic2
## V = 9549959, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

- The `V` statistic that R returns in the output is the sum of the *ranks of positive differences* rather than the W, which would have been the smaller of the two sums, positive differences or negative differences.
- V would be the same as W when the sum of the ranks of positive differences was lowest, but different from W when the sum of the ranks for negative differences was lowest.

NHST Step 3: Compute the probability for the test statistic (p-value)

The p-value is shown in scientific notation in the output as `< 2.2e-16` which is well below .05.

NHST Steps 4 & 5: Interpret the probability and write a conclusion

- The resulting output is similar to the output for many of the tests. The interpretation:

- Interpretation: We used a Wilcoxon Signed-Ranks Test to determine whether the distribution of the difference in systolic blood pressure measured at time 1 and time 2 was symmetrical around zero. The resulting test statistic and p-value indicated that the sample came likely from a population where the differences were not symmetrical around zero ($p < .05$). That is, we found a significant difference between the first and second blood pressure measures.

Alternative when the independent-samples t-test normality assumption fails: Mann-Whitney U test

- The Mann-Whitney U test is an alternative to the independent-samples t-test when the continuous variable is not normally distributed.
- The U test also relaxes the variable type assumption and can be used for ordinal variables in addition to continuous variables.
- Similar to the Wilcoxon Signed-Rank Test, the Mann-Whitney U test puts all the values for the continuous (or ordinal) variable in order, assigns each value a rank, and compares ranks across the two groups of the categorical variable.
- The test statistic is computed using the sums of the ranks for each group.
- The distribution for the test statistic approximates normality as long as the sample size is greater than 20 and a z-score is used to determine the corresponding p-value.

Step 1: Write the null and alternate hypotheses

H0: There is no difference in ranked systolic blood pressure values for males and females in the US population.

HA: There is a difference in ranked systolic blood pressure values for males and females in the US population.

Step 2: Compute the test statistic

- As if the similar process was not confusing enough, this test is also called the **Wilcoxon Rank-Sum Test**---which is not the same as the **Wilcoxon Signed-Ranks Test**.

- The same R function used in the previous test can be used here with two changes, use of `formula =` instead of `x =` and `y =` and use of `paired = FALSE`:

```
# test the distribution of systolic by sex
u.syst.by.sex <- wilcox.test(formula = nhanes.2016.cleaned$systolic ~ nh
            paired = FALSE)
u.syst.by.sex
```

```
##
##      Wilcoxon rank sum test with continuity correction
##
## data:  nhanes.2016.cleaned$systolic by nhanes.2016.cleaned$sex
## W = 7186882, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

Step 3: Calculate the probability that your test statistic is at least as big as it is if there is no relationship (i.e., the null is true)

The p-value is shown in scientific notation in the output as $< 2.2e - 16$ which is well below .05.

Steps 4 & 5: Interpret the probability and write a conclusion

- A Mann-Whitney U test comparing systolic blood pressure for males and females in the US found a *statistically significant difference* between the two groups ($p < .05$).

Effect size for Mann-Whitney U

- One effect size that can be used with Mann-Whitney U is r, which is computed using the z-statistic from the Mann-Whitney U test and dividing by the square root of the sample size:

$$r = \frac{z}{\sqrt{n}}$$

Calculating the effect size for U

- For the current analysis, the effect size for the comparison of male and female systolic blood pressure can be determined using the z from the U test.
- `qnorm()` finds the z-statistic that corresponds to a given p-value.
- Add the p-value from `u.syst.by.sex` to the `qnorm()` function to get the z-statistic.

```
# use qnorm to find z from p-value
qnorm(u.syst.by.sex$p.value)
```

```
## [1] -9.206125
```

- The z-statistic is negative and large.
- Because effect size is about the size or strength and not the direction (positive or negative) of a relationship, the absolute value can be used to get the effect size r with the sample size of 7,145.

$$r = \frac{9.206125}{\sqrt{7145}} = .109$$

Interpreting the effect size

- Consistent with the effect size from the t-test comparing males and females, this is a pretty small effect size, with r effects being classified as:
  - $r = .1$ to $r < .3$ is small
  - $r = .3$ to $r < .5$ is medium
  - $r \geq .5$ is large

Getting the r effect size in R

- The R code is a little tricky this time because the value of n used in the R calculations is the entire n without dropping the missing values, which is inconsistent with the calculation of the U stat
- To remove the `NA`, create a new data frame without the `NA` values for systolic, and use the new data frame with the `wilcoxonR()` command to get the r.

```
# new data frame with no NA
nhanes.2016.cleaned.noNA <- nhanes.2016.cleaned %>%
  drop_na(systolic)

# use new data frame to get r
rcompanion::wilcoxonR(x = nhanes.2016.cleaned.noNA$systolic,
                      g = nhanes.2016.cleaned.noNA$sex)
```

```
##    r
## 0.11
```

- Add effect size to interpretation: A Mann-Whitney U test comparing systolic blood pressure for males and females in the US found a statistically significant difference between the two groups ($p < .05$). Histograms demonstrated the differences with notably more females with systolic blood pressure below 100 compared to males along with some other differences. The effect size was small, $r = .11$, indicating a weak but statistically significant relationship between sex and systolic blood pressure.

Alternative when the independent-samples t-test variance assumption fails: The Kolmogorov-Smirnov (or K-S) test

- The K-S test is used when the variances in the two groups are unequal, the homogeneity of variances assumption is *not met*, whether or not the normality assumption is met.
- When variances are unequal, the larger variance has a bigger influence on the size of the t-statistic, so one group is dominating the t-statistic calculations.
- The K-S test compares the the distributions of the groups.

NHST Step 1: Write the null and alternate hypotheses

H0: The distribution of systolic blood pressure for males and females is the same in the US population.

HA: The distribution of systolic blood pressure for males and females is not the same in the US population.

NHST Step 2: Compute the test statistic

- The `ks.test()` command is used to test the null hypothesis.
- Unfortunately, `ks.test()` takes two vectors as arguments, one vector for each group.
- A little additional data management is needed to separate the males from the females.
- The `pull()` function is useful for getting a single variable out of a data frame as a stand alone vector.
- With `pull()`, use the `var =` argument with the name of the variable to pull out into its own vector:

```
# get vectors for male and female systolic
males.systolic<- nhanes.2016.cleaned %>%
  filter(sex == "Male") %>%
  pull(var = systolic)

females.systolic <- nhanes.2016.cleaned %>%
  filter(sex == "Female") %>%
  pull(var = systolic)
```

- Compare `males.systolic` and `females.systolic` using `ks.test()`.

Step 3: Calculate the probability that your test statistic is at least as big as it is if there is no relationship (i.e., the null is true)

The p-value is shown in scientific notation in the output as `< 2.2e-16` which is well below .05.

Steps 4 & 5: Interpret the probability and write a conclusion

- The K-S test compared the distribution of systolic blood pressure for males and females in the US and found a *statistically significant difference* between the two groups ($D = .11$; $p < .05$).
- The test statistic, D, is the maximum distance between the two empirical cumulative distribution functions (ECDF), which are a special type of probability distribution showing the cumulative probability of the values of a variable.
- To examine the difference between the ECDF for systolic blood pressure of males and females in the sample, graph the two ECDF curves.

```
nhanes.2016.cleaned %>%
  ggplot(aes(x = systolic, color = sex)) +
  stat_ecdf(size = 1) +
  theme_minimal(base_size = 16) +
  labs(x="Systolic blood pressure (mmHg)",
       y="Cumulative probability",
       title = "ECDF of systolic blood pressure in mmHg by sex\nfor 2015
  scale_color_manual(values=c("gray", "#7463AC"), name = "Sex")
```

Steps 4 & 5: Interpret the probability and write a conclusion

- The K-S test compared the distribution of systolic blood pressure for males and females in the US and found a *statistically significant difference* between the two groups ($D = .11$; $p < .05$).

- The test statistic, D, is the maximum distance between the two empirical cumulative distribution functions (ECDF), which are a special type of probability distribution showing the cumulative probability of the values of a variable.

- To examine the difference between the ECDF for systolic blood pressure of males and females in the sample, graph the two ECDF curves.

Interpreting the ECDF plot

- At the widest gap between these two curves, males and females were .11 apart, giving a test statistic of $D = .11$.
- The probability of getting a test statistic this large or larger is determined by examining the K-S probability distribution.
- In this case, the probability of .11 difference between the two was very tiny ($p < .05$) if the null hypothesis were true, so the difference between the distributions for males and females would be reported as statistically significant.
- Interpreting the results: A K-S test comparing systolic blood pressure for males and females found a statistically significant difference between the two groups ($D = .11$; $p < .05$). This sample likely came from a population where the distribution of systolic blood pressure was different for males and females.