STAT1006 Week 2 Cheat Sheet

Lisa Luff

8/13/2020

Contents

Revision

Paired t-test

- Applies when observations are made about the same person or object
- Usually before or after some experimental treatment has been applied, or on matched pairs (twins, siblings, etc)
- The samples are not "independent" because the measurements are made on the same person/object, and are therefore linked
- In this case we use paired data to test the difference between the two population means
- The variable of interest is then $X_{difference} = (X\_1 - X\_2)$
- And $H_0 : \mu_{diff} = 0; H_A : \mu_{diff} \neq ($ or $< 0,$ or $> 0);$ where $\mu_{diff} = \mu_1 - \mu_2$
- This is conceptually the same as testing one population

Steps of a paired t-test

1. Define H_0 and H_A in terms of μ_{diff}
2. Calculate the test statistic

- $t = \frac{\bar{x}_{diff} - \mu_{diff}}{\frac{s_{diff}}{\sqrt{n}}}$

3. Find the p-value

- To find the p-value, use Student's t distribution with $n - 1$ degrees of freedom

4. Decision and 5. Conclusion

- Make a decision based on the p-value and reword this as a conclusion to made about the data in terms of the original question

Assumptions for Paired t-test

1. The observations are differences from paired or matched samples (dependent)
2. The differences can be considered a random sample from a population of differences
3. The sample mean difference is Normally distributed

- Either the underlying data is Normal
- Or the sample size is large enough for the Central Limit Theorem to apply

New Topic

The Wilcoxon Signed Rank Test for Matched Pairs Steps

1. Collect Sample

- Draw and SRS of size n from a population for a matched pairs study

2. Set up hypotheses

- Suppose we test $H_0 : \eta_d = 0$ where $\eta_d = \eta_1 - \eta_2$ and η_1 and η_2 are the medians

3. Calculate X_d

- Take the differences $X_d = X_1 - X_2$ in responses within pairs

4. Remove ties

- Ties are dealt with the same as for Wilcoxon Signed Rank single sample tests
- Remove any with a difference of 0 ($X_d = 0$)

5. Rank

- Rank the absolute values of these differences

6. Find the test statistic

- The sum W^+ of the ranks for the positive differences is the Wilcoxon Signed Rank statistic
- W_- is the sum of the ranks of the negative differences
- $W = min(W^+, W_-)$

7. Calculate the p-value

- If there are no ties at any point:
  - Use the exact method psignrank or wilcox.test(*Sample1*, *Sample2*, paired = TRUE, alternative = "*x*", exact=TRUE)
- Otherwise:
  - Calculate the p-value for the Wilcoxon Signed Rank statistic using the Normal approximation with the continuity correction.
- If the distribution of the responses is not affected by the different treatments within pairs, then:
  - $\mu_{W+} = \frac{n*(n+1)}{4}$
  - $\sigma_{W+} = \sqrt{\frac{n*(n+1)*(2*n+1)}{24}}$
  - Under H_0, the sampling distribution of $z = \frac{(W^+ - \mu_{W+})}{\sigma_{W+}}$ is approximately Standard Normal
  - In R: wilcox.test(*Sample1*, *Sample2*, paired = TRUE, alternative = "*x*", exact = FALSE)
- R calculates the p-value from the exact distribution under two conditions:
  1. n is less than 50
  2. AND there are no ties at any point
  - Otherwise, Normal approximation with the continuity correction is being used

8. Decision and 9. Conclusion

- The Wilcoxon Signed Rank test rejects the hypothesis that there are no systematic differences within pairs when the rank sum W^+ is far from its mean

Functions of the Wilcoxon Signed Rank Statistic in R

Function	What it does
dsignrank(x, n, log=FALSE)	Density for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size n
psignrank(q, n, lower.tail=TRUE, log.p=FALSE)	Distribution function for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size n
qsignrank(p, n, lower.tail=TRUE, log.p=FALSE)	Quantile function for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size n
rsignrank(nn, n)	Random generation function for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size n

Psignrank Function

- psignrank(q, n, lower.tail=TRUE, log.p=FALSE)
  - q - The vector of quantitiles (or W)
  - n - Number(s) of observations in the sample(s). A positive integer, or a vector of such integers
  - lower.tail: logical - If TRUE (default), probabilities are P[X ≤ x], otherwise, P[X > X]
  - log.p: logical - If TRUE, probabilities p are given as $\log(p)$
- The range of W is between 0 and n*(n + 1)/2

Wilcoxon Rank Sum Test

- The Wilcoxon Rank Sum test compares any two continuous distributions, regardless of distribution, by testing:
  - H_0: The two distributions are the same
  - H_A: One has values that are systematically larger
- The hypotheses are non-parametric because they do not involve an specific parameter such as mean or median
- If the two distributions have the same shape, the general hypotheses reduce to comparing medians
- The Wilcoxon Rank Sum test rejects the hypothesis that the two populations have identical distributions when the rank sum W is far from its mean
  - If two distributions are identical, then samples of the same size, should have roughly the same number of small, medium and large values. Thus the sum of the ranks for each sample, should be roughly the same
  - Instead, if the sample sizes differ, then the average of the ranks should be similar for the two samples

Wilcoxon Rank Sum Test Steps

1. Collect samples

- Draw an SRS of size n_1 from one population and draw an independent SRS of size n_2 from a second population.
- There are N observations in all, where $N = n_1 + n_2$

2. Set hypotheses

H_0	H_A	Compute
$\tilde{\mu}_1 = \tilde{\mu}_2$	$\tilde{\mu}_1 < \tilde{\mu}_2$	μ_1
$\tilde{\mu}_1 = \tilde{\mu}_2$	$\tilde{\mu}_1 > \tilde{\mu}_2$	μ_2
$\tilde{\mu}_1 = \tilde{\mu}_2$	$\tilde{\mu}_1 \neq \tilde{\mu}_2$	μ

3. Rank

- Rank all N observations, keeping track of which samples the data value comes from

4. Find the test statistic

- The sum of the ranks for the first sample is the test statisitic w_1
- $u_1 = w_1 - \frac{n_1*(n_1+1)}{2}$
- This is applied to the second sample to get the test statistic w_2
- To find W, find the minimum of the two

5. W mean and standard deviation **(not needed)**

- If the two populations have the same continuous distribution (under H_0), then W has:
  - $\mu_W = \frac{n_1*(N+1)}{2}$
  - $\sigma_W = \sqrt{\frac{n_1*n_2(N+1)}{12}}$

6. Exact p-value

- If no ties and N is small, use the exact method in R
  - Either wilcox.test(*variable1*, *variable2*, mu = 0, alternative = *insert*, paired = FALSE, exact = TRUE) or pwilcox(W, n_1, n_2)

7. Approximate p-value

- To calculate the p-value, we need to know the sampling distribution of the rank sum W when the null hypothesis is true

i) P-values for the Wilcoxon test are often based on the fact that the rank sum statistic W becomes approximately Normal as the two sample sizes increase. To a good approximation, Z has a standard Normal distribution when the null hypothesis is true and the two sample sizes are relatively large *
$z = \frac{W - \mu_W}{\sigma_W} = \frac{W - n_1*(N+1)/2}{\sqrt{n_1*n_2*(N+1)/12}}$

ii) A better approximation for the p-value is using a continuity correction

- **OR** just use wilcox.test(*variable1*, *variable2*, mu = 0, alternative = *insert*, paired = FALSE, exact = FALSE)

Summary

Test	Exact - R only, no manual calculation	Approximate (with ties) - R or manually
Wilcoxon Signed Rank	psignrank() - discrete	Calculate the p-value using Normal approximation with continutity correction
(One sample)	wilcox.test($\ldots$exact=TRUE) - no ties AND n less than 50	wilcox.test($\ldots$exact=FALSE)
Wilcoxon Rank Sum	pwilcox() - discrete	Calculate the p-value using Normal approximation with continuity correction
(Two sample or Paired)	wilcox.test($\ldots$exact=TRUE) - no ties AND n less than 50	wilcox.test($\ldots$exact=FALSE)