Module 2:
Inference for a Normal Population

Different flavours of t tests

Hypothesis testing for means using t tests Agenda

1. **Why do we use Student t-tests instead of Z scores?**
2. **What are the three types of t-tests**

- **One sample t tests**
  - Assumptions
  - When assumptions not met, use median and rank → **Signed test**
- **Paired t test**
  - Assumptions
- **Two sample t test**
  - Assumptions
  - When variances aren't equal → **Welch's approximate t test**
  - Other assumptions not met: median and rank → **Mann Whitney U test**

Four general ways to address violations:

- **Ignore**
  - Sometimes you can use a method even if assumptions are violated
  - Thank the Central Limit Theorem for robustness
    - *means of large samples are normally distributed*
    - **tests that are based on** comparing sample **means** will be robust when they have sufficient sample size
    - **this doesn't help with F-test etc.**
  - Especially true if sample sizes are large ($n_i >> 50$) and violations are not extreme
    - **sample size must increase to accommodate how extreme violations are between groups**
      - especially if two samples both differ in opposite directions
  - Even with CLT we can't always ignore:
    - **outliers**
    - **frequency distribution between groups is very different**

Four general ways to address violations:

- Ignore
    - sometimes you can use a method even if assumptions are violated
    - especially true if sample sizes are large an violations are not extreme
- **Transform**
    - attempt to force normality and other assumptions onto data
    - We will investigate various tools
        - Usually boils down to: ***take the log of the data***
        - *Changes each measurement in the same way (1 to 1 correspondence) so that you can **transform back to get original data without ambiguity***
            - **monotonic relationship with original values**
            - remember to transform back the upper and lower limits for a Confidence Interval
    - work that does not always pay off but you at least maintain power of the test that you are using (since nonparametric tests usually reduce power)

Four general ways to address violations:

- Ignore
  - sometimes you can use a method even if assumptions are violated
  - especially true if sample sizes are large an violations are not extreme
- Transform
  - attempt to force normality and other assumptions onto data
  - We will investigate various tools
  - work that does not always pay off
- **Use Non-parametric method**
  - classes of methods that do not require assumption of normality
  - not cost free! Often lose power etc.

Power = 1-P[FTR Ho|Ho is incorrect] = P[reject|Ho is incorrect]

Four general ways to address violations:

- Ignore
  - sometimes you can use a method even if assumptions are violated
  - especially true if sample sizes are large an violations are not extreme
- Transform
  - attempt to force normality and other assumptions onto data
  - We will investigate various tools
  - work that does not always pay off
- Use Non-parameter method
  - classes of methods that do not require assumption of normality
  - not cost free! Often lose power etc.
- **Computationally Intensive Methods** (module LP2_Module5)
  - Simulation
  - Bootstrap
  - randomization/permutation test

Mann-Whitney U-Test

- Compares the central tendencies of two groups using ranks
  - uses ranks of measurements to test whether frequency distribution of two groups are the same
  - Small samples lead to little power
  - All group samples are random samples
  - Distribution of the variable has the same shape in every population

- Nonparametric version of two-sample t-test

Mann-Whitney U-Test

Method:

1. Declare hypotheses
2. Rank all individuals from both groups together in order (smallest to largest)
3. Sum ranks for all individuals in each group
4. Calculate test statistic, U

Mann-Whitney U-Test

Method:

1. Declare hypotheses
2. Rank all individuals from both groups together in order (smallest to largest): R_1, R_2, ...R_n
3. Sum ranks for all individuals in each group
4. Calculate test statistic, U

$$\boxed{U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1}$$

$$\boxed{U_2 = n_1 n_2 - U_1}$$

U_1 is the number of times an individual from pop 1 has a lower rank than an individual from pop 2 out of all pairwise comparisons

Mann-Whitney U-Test

Method:

1. Declare hypotheses
2. Rank all individuals from both groups together in order (smallest to largest): R_1, R_2, ...R_n
3. Sum ranks for all individuals in each group
4. Calculate test statistic, U
5. U is the larger of U_1 or U_2
6. Determine the P-value by comparing the observed U with the critical value of the null distribution for U (https://real-statistics.com/statistics-tables/mann-whitney-table/)
7. If both samples have n > 10, use Z transformation

$$Z = \frac{2U - n_1 n_2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/3}}$$

You will normally do this using R or other software. To emulate that, you can use this online WMU calculator:
https://www.socscistatistics.com/tests/mannwhitney/

Mann-Whitney U-Test

How to deal with tied ranks:

- Determine the ranks that the values would have been assigned if they were slightly different

- Average these ranks, and assign **that** average to each tied individual

- Count all those individuals when deciding the rank for the next largest individual

Intuitive way to understanding MWU test

- View the MWU test as the 'extremeness' of the distributions of the two populations

Example: two populations each have three data points:

4 5 6 0 1 2

The higher ranks/values are all in one sample, and all the lower ranks/values are in the other sample

Remember there are 20 ways of selecting 3 items from 6 (for group 1)

$$\binom{6}{3} = \frac{6*5*4*3*2*1}{3*2*1(3*2*1)} = 20$$

Mann-Whitney U-Test

Assumptions:

- Both samples are random samples
- Both populations have the same shape of distribution
  - **They can both be skewed but it must be in the same direction**
  - **Same variance**

Mann-Whitney U-Test

When intruding lions take over a pride of females, they often kill most or all the infants in the group to reduce the time until females are again sexually receptive. A long-term study measured the time to reproduction of female lions after losing cubs to infanticide and compared this to the time to reproduction of females who had lost their cubs to accidents. The data are not normally distributed within the two groups, and we have been unable to find a transformation that makes them normal.

Accidental: 110, 117, 133, 135, 140, 168, 171, 238, 255
Infanticide: 211, 232, 246, 251, 275

Mann-Whitney U-Test

When intruding lions take over a pride of females, they often kill most or all the infants in the group to reduce the time until females are again sexually receptive. A long-term study measured the time to reproduction of female lions after losing cubs to infanticide and compared this to the time to reproduction of females who had lost their cubs to accidents. The data are not normally distributed within the two groups, and we have been unable to find a transformation that makes them normal.

Accidental: 110, 117, 133, 135, 140, 168, 171, 238, 255

Infanticide: 211, 232, 246, 251, 275

1. H_0: $\mu_{accidental} = \mu_{infanticide}$

   H_A: $\mu_{accidental} \neq \mu_{infanticide}$

In words: the mean time to reproduction of female lions who have lost cubs due to accident or infanticide is the same.

Mann-Whitney U-Test

When intruding lions take over a pride of females, they often kill most or all the infants in the group to reduce the time until females are again sexually receptive. A long-term study measured the time to reproduction of female lions after losing cubs to infanticide and compared this to the time to reproduction of females who had lost their cubs to accidents. The data are not normally distributed within the two groups, and we have been unable to find a transformation that makes them normal.

Accidental: 110, 117, 133, 135, 140, 168, 171, 238, 255

Infanticide: 211, 232, 246, 251, 275

1. H_0: $\mu_{accidental} = \mu_{infanticide}$

   H_A: $\mu_{accidental} \neq \mu_{infanticide}$

2. Rank all individuals from both groups together

110, 117, 133, 135, 140, 168, 171, 211, 232, 238, 246, 251, 255, 275

3. Sum ranks for all individuals in each group

Accidental (**R_1**): 1+2+3+4+5+6+7+10+13 = 51

Infanticide **(R_2)** : 8+9+11+12+14 = 54

Mann-Whitney U-Test

When intruding lions take over a pride of females, they often kill most or all the infants in the group to reduce the time until females are again sexually receptive. A long-term study measured the time to reproduction of female lions after losing cubs to infanticide and compared this to the time to reproduction of females who had lost their cubs to accidents. The data are not normally distributed within the two groups, and we have been unable to find a transformation that makes them normal.

Accidental: 110, 117, 133, 135, 140, 168, 171, 238, 255

Infanticide: 211, 232, 246, 251, 275

1. H_0: $\mu_{accidental} = \mu_{infanticide}$

   H_A: $\mu_{accidental} \neq \mu_{infanticide}$

2. Rank all individuals from both groups together

110, 117, 133, 135, 140, 168, 171, 211, 232, 238, 246, 251, 255, 275

3. Sum ranks for all individuals in each group

Accidental (R_1): 1+2+3+4+5+6+7+10+13 = 51

Infanticide (R_2): 8+9+11+12+14 = 54

4. Calculate test statistic, U (NB: U is larger of U_1 or U_2):

$U_1 = n_1 n_2 + n_1(n_1+1)/2 - R_1 = (9)(5) + (9)(10)/2 - 51$

$= 45+45-51 = 39$

$U_2 = n_1 n_2 - U_1 = 45 - 39 = 6$

$\boxed{U = 39}$

Mann-Whitney U-Test

When intruding lions take over a pride of females, they often kill most or all the infants in the group to reduce the time until females are again sexually receptive. A long-term study measured the time to reproduction of female lions after losing cubs to infanticide and compared this to the time to reproduction of females who had lost their cubs to accidents. The data are not normally distributed within the two groups, and we have been unable to find a transformation that makes them normal.

Accidental: 110, 117, 133, 135, 140, 168, 171, 238, 255

Infanticide: 211, 232, 246, 251, 275

1. H_0: $\mu_{accidental} = \mu_{infanticide}$

   H_A: $\mu_{accidental} \neq \mu_{infanticide}$

2. Rank all individuals from both groups together

110, 117, 133, 135, 140, 168, 171, 211, 232, 238, 246, 251, 255, 275

3. Sum ranks for all individuals in each group

Accidental: 1+2+3+4+5+6+7+10+13 = 51

Infanticide: 8+9+11+12+14 = 54

4. Calculate test statistic, U (NB: U is larger of U_1 or U_2): **U = 39**

5. Determine critical value from Table E

Critical Value is 38 for $n_1 = 9$, $n_2 = 5$ for alpha =0.05

Our value of **U** is > Critical value so we can reject the null hypothesis that females have equal times to reproduction regardless of how their previous cub died.