

Chapter 13: ***Handling violations of assumptions***

1

Assumptions of t -tests

- The sampling units are randomly sampled from the population
- Numerical variable(s) have an approximately normal distribution
- Two-sample t -test: standard deviations (and variances) are the same in the two populations

2

4 potential options

- Ignore the violations
- Transform the data
- Use a nonparametric method
- Use a permutation test

3

4 potential options

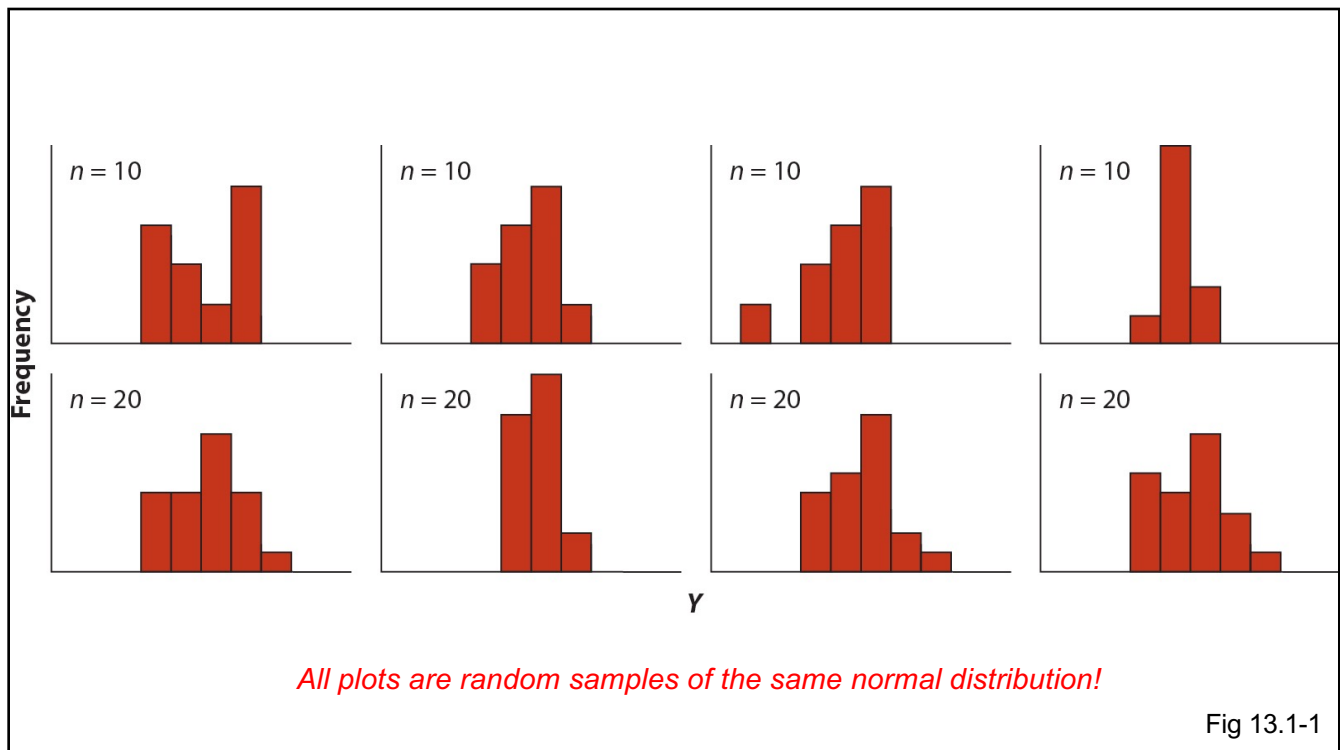
- Ignore the violations
- Transform the data
- Use a nonparametric method
- Use a permutation test

4

Detecting deviations from normality

- Use graphs
 - Frequency distribution histogram
 - Box plot
 - Strip plot
- But a sample of a normal distribution may look noisy due to sampling error, particularly when sample sizes are small

5



6

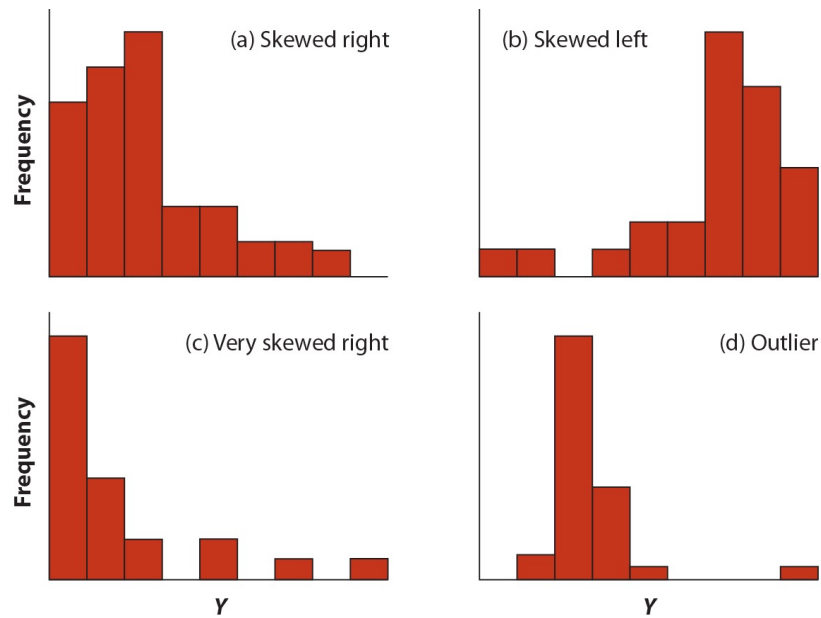


Fig 13.1-2

7

Normal quantile plot

- The **normal quantile plot** compares each observation in the sample with its quantile expected from the standard normal distribution. Points fall roughly along a straight line if the data come from a normal distribution
- https://www.youtube.com/watch_popup?v=X9_ISJ0YpGw

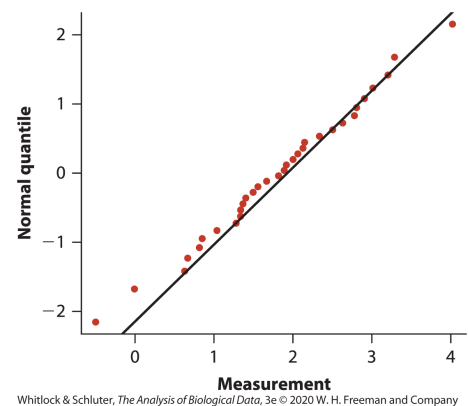


Fig 13.1-3

8

Ex 13.1: The benefits of marine reserves



- Interested if marine reserves have a measurable impact on biodiversity
- Total biomass per unit area in reserve
- Total biomass per unit area in control
 - Either area before it was reserve or closely matched non-reserve area
- Ratio: reserve/control
- What test should you run?

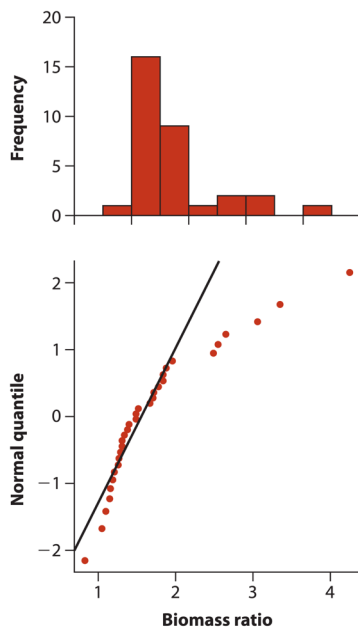
Ex 13.1: The benefits of marine reserves



- One-sample t -test
- H_0 : mean biomass ratio is unaffected by reserve protection ($\mu = 1$)
- H_A : mean biomass ratio is affected by reserve protection ($\mu \neq 1$)
- But first you should check if your sample meets test assumptions!

11

Sample of 32 biomass ratios



Whitlock & Schluter, The Analysis of Biological Data, 3e © 2020 W. H. Freeman and Company

- Test assumes biomass-ratio data are drawn from a population having a normal distribution
- Does the sample appear normal?

Fig 13.1-4

12

Formal tests of normality

- H_0 : sample has normal distribution
- H_A : sample does not have normal distribution
- Should be used with caution
 - Small sample sizes lack power to reject a false null (Type II error)
 - Large sample sizes can reject null when the departure from normality is minimal and would not affect methods that assume normality

13

Shapiro-Wilk test

- A **Shapiro-Wilk test** evaluates the goodness of fit of a normal distribution to a set of data randomly sampled from a population
- Most commonly used formal test of normality
- Estimates mean and standard deviation using sample data
- Tests goodness-of-fit between sample data and normal distribution (with mean, sd of the sample)

14

15

4 potential options

- Ignore the violations
- Transform the data
- Use a nonparametric method

16

Ignore violations of normality?

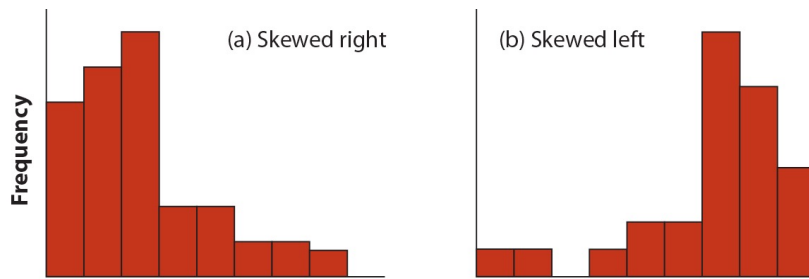
- t -tests assume data are drawn from a population having a normal distribution
- But can sometimes be used when data are not normal
 - Recall central limit theorem
- When sample sizes are large the sampling distribution of means behaves roughly as assumed by t -distribution

17

Ignore violations of normality?

- But what is a “large” sample size?
- Depends on the shape of the distribution
- If distributions of two groups being compared are **skewed in different directions**, then avoid t -tests even for large samples
- If distributions are similarly skewed then there is more leeway

18



a vs a: sample size of ~30 in each group ok

a vs b: need sample size of 500 or more in each group

Fig 13.1-2

19

Ignore violations of equal standard deviations?

- Two-sample t -tests assume equal standard deviations in the two populations
- If sample sizes are > 30 in each group AND sample sizes in two groups are even, then even up to a 3x difference can be ok
- Otherwise, use Welch's t -test

20

4 potential options

- Ignore the violations
- Transform the data
- Use a nonparametric method

21

Data transformations

- A **data transformation** changes each measurement by the same mathematical formula
- Can make standard deviations more similar and improve fit of the normal distribution to the data
- All observations must be transformed
- If two samples then they both must be transformed in same way
- If used then usually best to back transform confidence intervals

22

Log transformation

- Data are transformed by taking the natural log (or sometimes log base-10) of each measurement

$$Y' = \ln[Y]$$

- Common uses:
 - Measurements are ratios or products
 - Frequency distribution skewed to the right
 - Group having larger mean also has larger standard deviation
 - Data span several orders of magnitude

23

Log transformation

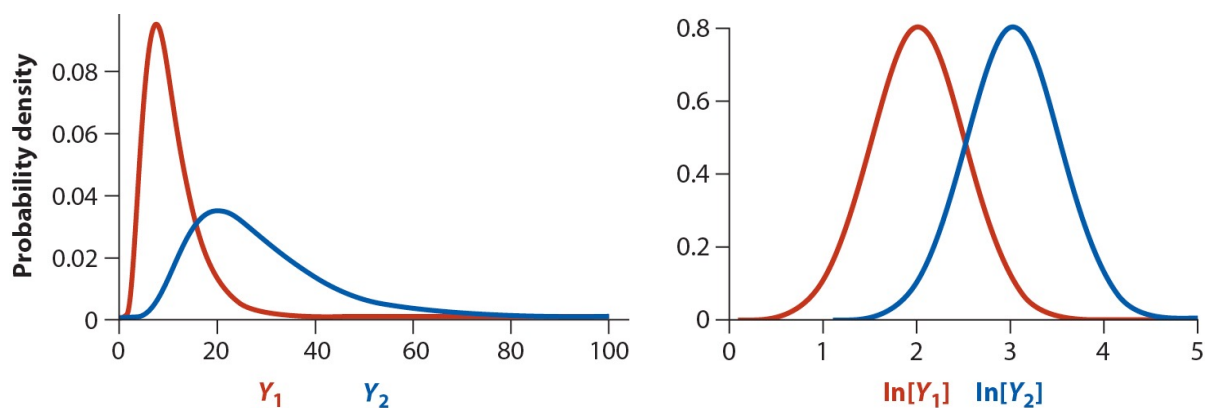


Fig 13.3-1

24

Other transformations

- Arcsine (best use: proportions)
- Square-root (counts)
- Square (skewed left)
- Antilog (skewed left)
- Reciprocal (skewed right)

25

Ex 13.1: The benefits of marine reserves



- One-sample t -test
- H_0 : mean biomass ratio is unaffected by reserve protection ($\mu = 1$)
- H_A : mean biomass ratio is affected by reserve protection ($\mu \neq 1$)
- But first you should check if your sample meets test assumptions!

26

Sample of 32 biomass ratios

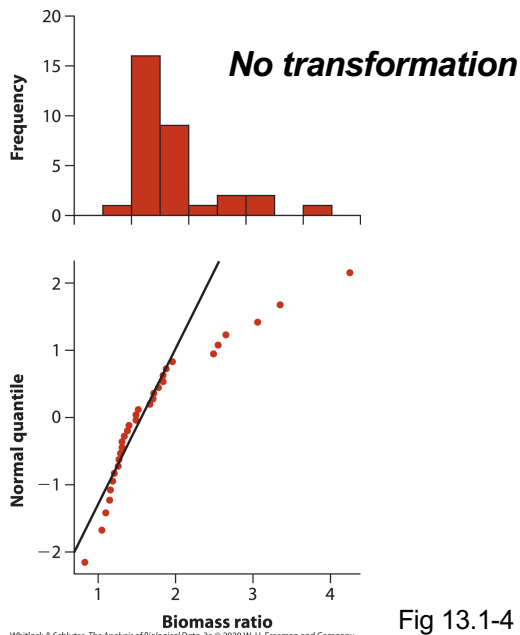


Fig 13.1-4

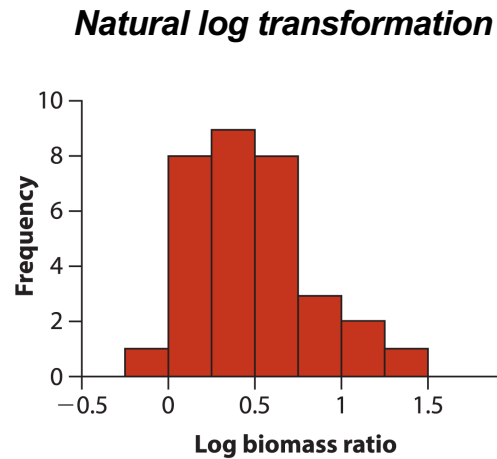


Fig 13.3-2

27

Ex 13.1: The benefits of marine reserves



- H_0 Before: mean biomass ratio is unaffected by reserve protection ($\mu = 1$)
- H_0 After: mean \ln biomass ratio is unaffected by reserve protection ($\mu' = 0$)
- H_A Before: mean biomass ratio is affected by reserve protection ($\mu \neq 1$)
- H_A After: mean \ln biomass ratio is affected by reserve protection ($\mu' \neq 0$)
- $t_{31} = 7.40$
- $P\text{-value} = 2.4 \times 10^{-8}$

28

29

4 potential options

- Ignore the violations
- Transform the data
- Use a nonparametric method

30

Nonparametric alternatives

- A **nonparametric method** makes fewer assumptions than standard parametric methods do about the distributions of the variables
 - Can be used when deviations from normality should not be ignored, and sample remains non-normal even after transformation
- Do not rely on parametric statistics like mean, standard deviation, variance
- Usually based on ranks of the data points rather than the actual values

31

Sign test

- The **sign test** compares the median of a sample to a constant specified in the null hypothesis. It makes no assumptions about the distribution of the measurement of the population
- Each measurement is characterized as above (+) or below (-) the null hypothesis
- If the null is true, then you expect the half the measurements to be + and half to be –
- Uses binomial distribution to test if the proportion of measurements above the null hypothesis is $p = 0.5$

32

Ex 13.4: Sexual conflict and the origin of new species

- Sexual conflict: when males and females have opposing optimal trait values
- Hypothesized to increase rate of evolutionary change, and perhaps speciation rates
- Do insect groups in which females mate multiple times (high sexual conflict) have more species than those in which females mate once (low conflict)?



33

TABLE 13.4-1 The number of species in 25 pairs of insect groups. Each pair matches a group of insect species in which females mate only once with a related group of insect species in which females mate multiple times.

| Taxon pair | Number of species | | Difference | Above (+) or below (-) zero |
|------------|-----------------------|---------------------|------------|-----------------------------|
| | Multiple-mating group | Single-mating group | | |
| A | 53 | 10 | 43 | + |
| B | 73 | 120 | -47 | - |
| C | 228 | 74 | 154 | + |
| D | 353 | 289 | 64 | + |
| E | 157 | 30 | 127 | + |
| F | 300 | 4 | 296 | + |
| G | 34 | 18 | 16 | + |
| H | 3400 | 3500 | -100 | - |
| I | 20 | 1000 | -980 | - |
| J | 196 | 486 | -290 | - |
| K | 1750 | 660 | 1090 | + |
| L | 55 | 63 | -8 | - |
| M | 37 | 115 | -78 | - |
| N | 100 | 30 | 70 | + |
| O | 21,000 | 600 | 20,400 | + |
| P | 37 | 40 | -3 | - |
| Q | 7 | 5 | 2 | + |
| R | 15 | 7 | 8 | + |
| S | 18 | 6 | 12 | + |
| T | 240 | 13 | 227 | + |
| U | 15 | 14 | 1 | + |
| V | 77 | 16 | 61 | + |
| W | 15 | 14 | 1 | + |
| X | 85 | 6 | 79 | + |
| Y | 86 | 8 | 78 | + |

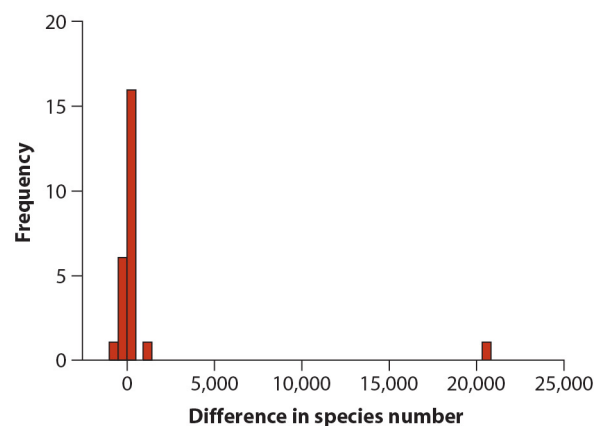


Fig 13.4-1

34

Ex 13.4: Sexual conflict and the origin of new species

- Design suitable for paired t -test, but...
- Sample size small ($n = 25$)
- Data are not normal and transformation doesn't help

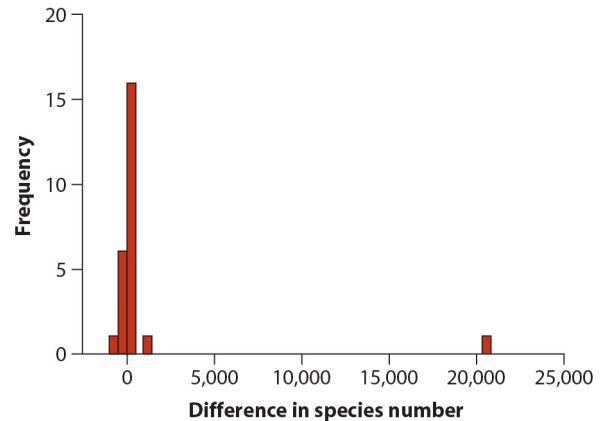


Fig 13.4-1

35

Sign test

- H_0 : median difference in number of species between insect groups is zero
- H_A : median difference in number of species between insect groups is not zero

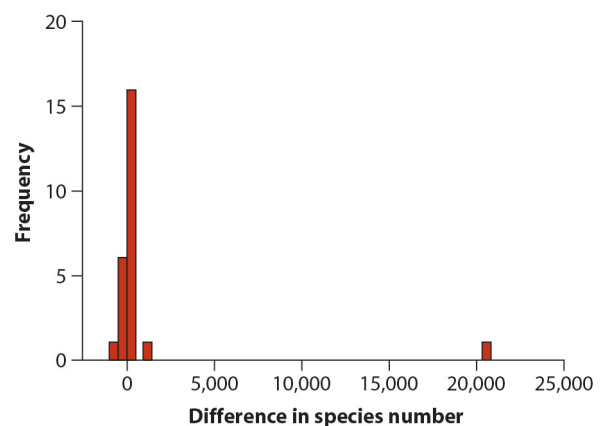


Fig 13.4-1

36

Sign test

- If difference between values is above zero then sign is +
- Dataset has 18 of 25 values above zero (+)
 - Null proportion of + values is 0.5
- Use binomial test:
 - `binom.test(18,25,p=0.5)`
 - $P\text{-value} = 0.04329$
- **Reject** the null hypothesis

37

Sign test

- Can be replacement of one-sample or paired t -test
- Low power compared to t -tests
 - Impossible to reject null when sample size < 5
- What if value is exactly null?
 - Insect example: difference of zero
- Typically you remove these data points (reducing n)

38

Wilcoxon signed-rank test

- More power than standard sign test because information about the magnitude away from the null for each data point
- But test assumes that population is symmetric around the median (i.e., no skew)
- Nearly as restrictive as normality assumption, **thus not recommended**

39

40

Nonparametric test for two samples?

- The **Mann-Whitney *U*-test** compares the distributions of two groups. It does not require as many assumptions as the two-sample *t*-test

41

Ex 13.5: Sexual cannibalism in sagebrush crickets



- In sage crickets, female eats fleshy hind wings of male during mating
- Are females more likely to mate if they are hungry?
- Two treatment groups:
 - Females starved ($n = 11$)
 - Females fed ($n = 13$)
- Measured time to mating

42

Ex 13.5: Sexual cannibalism in sagebrush crickets

TABLE 13.5-1 Times to mating (in hours) for female sagebrush crickets that were recently starved or fed. The measurements of fed females are in red to facilitate comparison after ranking (see Table 13.5-2).

| Starved | Fed |
|---------|------|
| 1.9 | 1.5 |
| 2.1 | 1.7 |
| 3.8 | 2.4 |
| 9.0 | 3.6 |
| 9.6 | 5.7 |
| 13.0 | 22.6 |
| 14.7 | 22.8 |
| 17.9 | 39.0 |
| 21.7 | 54.4 |
| 29.0 | 72.1 |
| 72.3 | 73.6 |
| | 79.5 |
| | 88.9 |

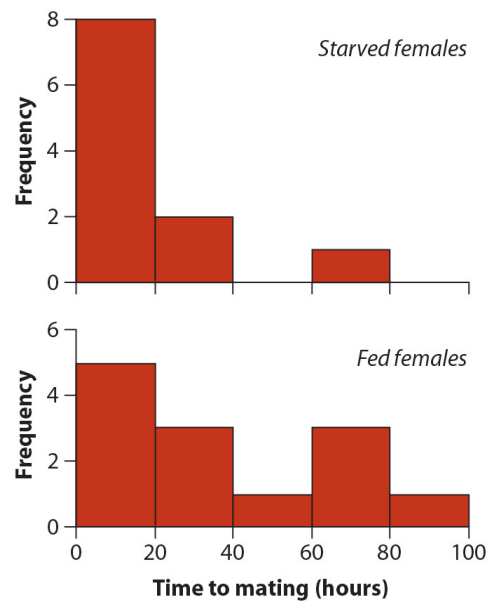


Fig 13.5-1

43

Deviations from normality?

- Both skewed right
- Transforming distributions don't make them ~normal
- Ignore?
 - Low sample sizes means we can't rely on central limit theorem

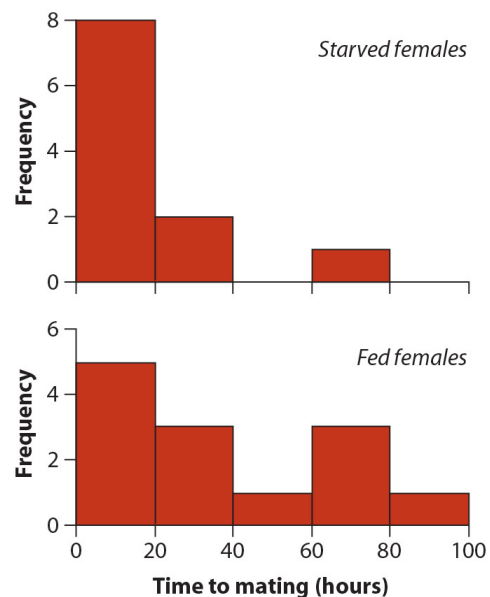


Fig 13.5-1

44

Mann-Whitney U-test

- H_0 : time to mating is the same for female crickets that are starved and those that are fed
- H_A : time to mating differs between the two groups

45

TABLE 13.5-1 Times to mating (in hours) for female sagebrush crickets that were recently starved or fed. The measurements of fed females are in red to facilitate comparison after ranking (see [Table 13.5-2](#)).

| Starved | R_1 | Fed | R_2 |
|---------|-------|------|-------|
| 1.9 | 3 | 1.5 | 1 |
| 2.1 | 4 | 1.7 | 2 |
| 3.8 | | 2.4 | 5 |
| 9.0 | | 3.6 | |
| 9.6 | | 5.7 | |
| 13.0 | | 22.6 | |
| 14.7 | | 22.8 | |
| 17.9 | | 39.0 | |
| 21.7 | | 54.4 | |
| 29.0 | | 72.1 | |
| 72.3 | | 73.6 | |
| | | 79.5 | |
| | | 88.9 | |

Rank sums

46

TABLE 13.5-1 Times to mating (in hours) for female sagebrush crickets that were recently starved or fed. The measurements of fed females are in red to facilitate comparison after ranking (see [Table 13.5-2](#)).

| Starved | R_1 | Fed | R_2 |
|---------|-------|------|-------|
| 1.9 | 3 | 1.5 | 1 |
| 2.1 | 4 | 1.7 | 2 |
| 3.8 | 7 | 2.4 | 5 |
| 9.0 | 9 | 3.6 | 6 |
| 9.6 | 10 | 5.7 | 8 |
| 13.0 | 11 | 22.6 | 15 |
| 14.7 | 12 | 22.8 | 16 |
| 17.9 | 13 | 39.0 | 18 |
| 21.7 | 14 | 54.4 | 19 |
| 29.0 | 17 | 72.1 | 20 |
| 72.3 | 21 | 73.6 | 22 |
| | | 79.5 | 23 |
| | | 88.9 | 24 |

Rank sums

47

TABLE 13.5-2 Times to mating of female crickets from both groups, ordered from smallest to largest and then ranked. Data from group 2 (fed crickets) are highlighted in red to facilitate comparison.

| Group | Time to mating | Rank |
|-------|----------------|------|
| 2 | 1.5 | 1 |
| 2 | 1.7 | 2 |
| 1 | 1.9 | 3 |
| 1 | 2.1 | 4 |
| 2 | 2.4 | 5 |
| 2 | 3.6 | 6 |
| 1 | 3.8 | 7 |
| 2 | 5.7 | 8 |
| 1 | 9.0 | 9 |
| 1 | 9.6 | 10 |
| 1 | 13.0 | 11 |
| 1 | 14.7 | 12 |
| 1 | 17.9 | 13 |
| 1 | 21.7 | 14 |
| 2 | 22.6 | 15 |
| 2 | 22.8 | 16 |
| 1 | 29.0 | 17 |
| 2 | 39.0 | 18 |
| 2 | 54.4 | 19 |
| 2 | 72.1 | 20 |
| 1 | 72.3 | 21 |
| 2 | 73.6 | 22 |
| 2 | 79.5 | 23 |
| 2 | 88.9 | 24 |

Rank sums

Sum ranks of group 1:

$$R_1 = 121$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 88$$

$$U_2 = n_1 n_2 - U_1 = 55$$

Whichever value is bigger is U test statistic

$$U = U_1 = 88$$

48

Mann-Whitney U-test

- $U = 88$, with sample sizes of 11 and 13
- Use Mann-Whitney U -distribution to calculate P -value
 - Use stats table or computer command
- $P > 0.05$, **fail to reject** the null hypothesis that starved and fed females have equal times to mating

49

Mann-Whitney U-test

- What if you have tied ranks?
- Assign all instances of the same measurement the average of ranks that the tied points would have received

TABLE 13.5-3 Measurements from two hypothetical groups, illustrating how ties are ranked.

| Group | Y | Rank |
|-------|----|------|
| 2 | 12 | 1 |
| 2 | 14 | 2 |
| 1 | 17 | 3 |
| 1 | 19 | 4.5 |
| 2 | 19 | 4.5 |
| 1 | 24 | 6 |
| 2 | 27 | 7 |
| 1 | 28 | 8 |

50

Assumptions of nonparametric tests

- Still assume that both samples are random samples from their populations
- Wilcoxon signed-rank test assumes distributions are symmetrical (big limitation – not recommended)
- Rejecting null hypothesis of Mann-Whitney U -test means two groups have different distributions of ranks, but does not necessary imply that means or medians of groups differ
- To make this inference there is an assumption that the shapes of the distributions are similar

51

Notes

- Skipping sections 13.7 and 13.8

52