Nonparametrics

Hypothesis Tests

OUTCOME VARIABLE

EXPOSURE VARIABLE

	Continuous	Binary
1 group	One-group t-test	Normal approximation test [or] exact binomial test
2 groups	Two-group t-test	χ^2 test [or] Fisher's exact test
>2 groups	ANOVA	χ^2 test [or] Fisher's exact test

Hypothesis Tests

Red = has additional distributional assumptions

OUTCOME VARIABLE

EXPOSURE VARIABLE

	Continuous	Binary
1 group	One-group t-test	Normal approximation test [or] exact binomial test Check: $np_0(1-p_0) \geq 5$
2 groups	Two-group t-test	χ^2 test [or] Fisher's exact test Check: all expected counts ≥ 5
>2 groups	ANOVA	χ^2 test [or] Fisher's exact test Check: no more than 20% of expected counts < 5 and none < 1

Hypothesis Tests

Red = has additional distributional assumptions

OUTCOME VARIABLE

EXPOSURE VARIABLE

	Continuous	Binary
1 group	One-group t-test [or] Wilcoxon signed-rank test	Normal approximation test [or] exact binomial test
2 groups	Two-group t-test [or] Wilcoxon rank sum test	χ^2 test [or] Fisher's exact test
>2 groups	ANOVA [or] Kruskal-Wallis test	χ^2 test [or] Fisher's exact test

Parametric & Nonparametric Methods

- One-group t-test, two-group t-test, and ANOVA all use an underlying assumption of normally-distributed data
- These are called **parametric methods**
  - Parametric methods rely on an underlying statistical distribution in the data
  - Inferences about population parameters are not valid if assumptions are not met
- Each of these methods has a corresponding **nonparametric method**
  - Nonparametric methods do not assume that the data or population have any specific structure

Nonparametric Methods for Continuous Data

Parametric Method	Nonparametric Equivalent
One-group t-test/paired t-test	Wilcoxon signed-rank test
Two-group t-test	Wilcoxon rank sum test (also called Mann Whitney test)
ANOVA	Kruskal-Wallis test

General Idea of Rank-Based Methods

1. Ignore the groups
2. Assign ranks to the observations from smallest to largest (1, 2, …)
3. Separate subjects back into their groups
4. Add up the ranks in each group
5. Bigger difference in sum of the ranks in each group → more evidence to reject H_0

Original data

ID	Weight	Group
001	174	1
002	152	1
003	145	1
004	166	1
005	171	1
006	155	2
007	137	2
008	199	2
009	204	2
010	146	2

Sorted by weight

ID	Weight	Rank	Group
007	137	1	2
003	145	2	1
010	146	3	2
002	152	4	1
006	155	5	2
004	166	6	1
005	171	7	1
001	174	8	1
008	199	9	2
009	204	10	2

Sum of ranks in Group 1:
2 + 4 + 6 + 7 + 8 = 27

Sum of ranks in Group 2:
1 + 3 + 5 + 9 + 10 = 28

Wilcoxon Signed-Rank Test

- Nonparametric equivalent of one-group t-test and paired t-test
  - For one group:

    Let m be the median in the population. We want to test if m is equal to a certain value, m_0.

    $H_0\colon m = m_0 \qquad H_A\colon m \neq m_0$

  - For two paired groups:

    Let Δ_m be the median of the differences in the population.

    $H_0\colon \Delta_m = 0 \qquad H_A\colon \Delta_m \neq 0$

Low Birth Weight Data

- Information on 100 low birth weight infants born in two teaching hospitals in Boston, Massachusetts

Variable	Description
sex	Sex of the baby (Male, Female)
gestage	Gestational age at time of birth (weeks)
length	Length of the baby (cm)
birthwt	Birth weight of the baby (g)
headcirc	Baby’s head circumference (cm)
apgar	Apgar score (integers, min=0, max=10). This is a scoring system used for assessing the clinical status of a newborn. 7 or higher is generally considered normal, 4-6 is low, and 3 or below is critically low.

Find the dataset (lowbwt.xlsx) and the full data dictionary (lowbwt Data Dictionary.pdf) in the Data Module on the Canvas site

Example: Length

- Use a nonparametric method to test whether the median length of low birth weight infants equals 39 cm.

Let m be the median length of low birth weight infants

H0: m = 39 cm

HA: m != 39 cm

Example: Length

Wilcoxon signed-rank test:

```
        Wilcoxon signed rank test with continuity correction

data:  length
V = 668, p-value = 0.00000006825
alternative hypothesis: true location is not equal to 39
```

$H_0: m = 39$

$H_A: m \neq 39$

p-value = 0.00000007

Since the p-value is less than 0.05, we reject the null hypothesis and conclude that there is sufficient evidence to suggest that the median length of low birth weight infants is not equal to 39 cm.

Wilcoxon Rank Sum Test

- Nonparametric equivalent of two-group t-test
  - For two independent groups:

    Let m_1 be the median in the population in Group 1. Let m_2 be the median in the population in Group 2.

    $H_0: m_1 = m_2 \qquad H_A: m_1 \neq m_2$

Example: Apgar Score/Hemorrhage

- Use a nonparametric method to test whether the median Apgar score is different in infants with and without germinal matrix hemorrhage.

Let m1 be the median apgar score in infants with GMH
Let m2 be the median apgar score in infants without GMH

H0: m1 = m2
HA: m1 != m2

Example: Apgar Score/Hemorrhage

Wilcoxon rank sum test:

```
        Wilcoxon rank sum test with continuity correction

data:  apgar by hemorrhage
W = 928.5, p-value = 0.004392
alternative hypothesis: true location shift is not equal to 0
```

$H_0: m_1 = m_2$

$H_A: m_1 \neq m_2$

p-value = 0.0044

Since the p-value is less than 0.05, we reject the null hypothesis and conclude that there is sufficient evidence to suggest that the median Apgar score is different among low birth weight infants with and without germinal matrix hemorrhage.

Kruskal-Wallis Test

- Nonparametric equivalent of ANOVA
  - For >2 groups:

    Let m_1 be the median in the population in Group 1. Let m_2 be the median in the population in Group 2. Let m_3 be the median in the population in Group 3.

    $H_0: m_1 = m_2 = m_3$
    H_A: at least one of the group medians is different from the others

Example: Apgar Score/Gestational Age

- Recall that *gestage3* is a categorical variable with 3 categories denoting whether the infant was born Preterm (32-36 weeks), Very Preterm (28-31 weeks), or Extremely Preterm (<28 weeks).
- Use a nonparametric method to test whether the median Apgar score is different in the three gestational age groups.

Let m1 be the median apgar score among infants that are pre-term
Let m2 be the median apgar score among infants that are very pre-term
Let m3 be the median apgar score among infants that are extremely pre-term

H0: m1=m2=m3
HA: at least one median is different

Example: Apgar Score/Gestational Age

Kruskal-Wallis test:

```
        Kruskal-Wallis rank sum test

data:  apgar by gestage3
Kruskal-Wallis chi-squared = 4.2664, df = 2, p-value = 0.1185
```

$H_0: m_1 = m_2 = m_3$

H_A: median Apgar score is different in at least one of the gestational age groups

p-value = 0.1185

Since the p-value is greater than 0.05, we fail to reject the null hypothesis and conclude that there is not sufficient evidence to suggest that the median Apgar score is different among the three gestational age groups.

Parametric vs. Nonparametric

- When to use nonparametric methods:
  - Small sample size
  - Strongly skewed data
  - Outliers in the data

  normality cannot be assumed
- Why can't we always just use nonparametric methods?
  - When both parametric and nonparametric methods are appropriate, the parametric test will have higher power

Regression and Correlation

- There are also nonparametric methods for correlation and regression
  - Correlation: Spearman's rank correlation
  - Regression: Splines, tree-based methods, kernel smoothing, k-nearest neighbors, etc.

This would be its own semester-long course!

Important Points

- General idea of how rank-based methods work
- Set up and interpretation of Wilcoxon signed-rank test, Wilcoxon rank sum test, and Kruskal-Wallis test
- When you should use a parametric or a nonparametric test