



### Inference Methods So Far

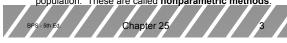
- Variables have had Normal distributions.
- In practice, few variables have true Normal distributions, but our methods have been robust (not sensitive to moderate lack of Normality).
- If the data is clearly not Normal, then using the methods of the previous chapters will yield inaccurate results. Other approaches must be investigated.



## Options for Non-Normal Data

Suppose plots suggest the data is non-Normal.

- If there are extreme outliers, investigate their cause before deciding how to proceed; do not remove outliers indiscriminately.
- In some instances, data can be transformed to be nearly Normal.
- Perhaps the data follow **another standard distribution**. If so, use methods based on that distribution.
- Finally, there are inference procedures that do not assume any specific form of the distribution of the population. These are called nonparametric methods



## **Ranking Procedures**

- The procedures we will study will replace *t* tests and ANOVA when the Normality condition is not met.
- Analyses will be based on the ranks (place in order) of the data and not their actual values.
- Rank tests concern the center of the population(s).
  for Normal distributions, center is described by the mean.
  for skewed distributions, center is described by the median.
- Ranking the data is essential to finding the median as you need to order the values before you can determine which one is in the center.



# Comparison of Testing Procedures

Setting	Normal test	Rank test
One sample	One-sample t test Chapter 17	Wilcoxon signed rank test
Matched pairs	Apply one-sample test to o	differences within pairs
Two independent samples	Two-sample <i>t</i> test Chapter 18	Wilcoxon rank sum test
Several independent samples	One-way ANOVA F test Chapter 24	Kruskal-Wallis test

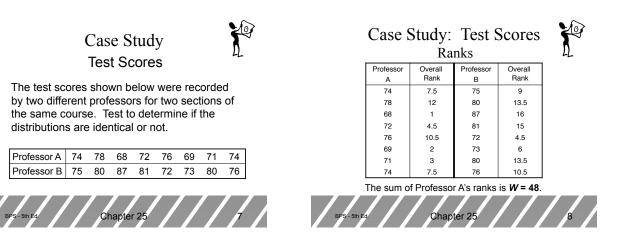


# Ranks

- To rank observations, first arrange them in order from smallest to largest.
- The rank of each observation is its position in this ordered list, starting with rank 1 for the smallest value.
- The **sum of the ranks** for *N* total observations will be:

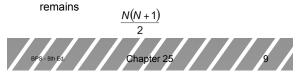
$$1+2+\cdots+N=\frac{N(N+1)}{2}$$





### Ties in Ranks

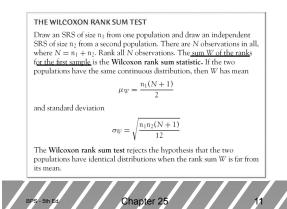
- If two or more values are tied, assign each the average of the corresponding ranks. Then continue the ranking with the next value.
  - if the 3<sup>rd</sup> and 4<sup>th</sup> values are tied, assign each a rank of 3.5; the 5th value will get a rank of 5. (*Ranks*: 1, 2, 3.5, 3.5, 5).
    if the 4<sup>th</sup>, 5<sup>th</sup>, and 6<sup>th</sup> values are tied, assign each a rank of 5; the 7th value will get a rank of 7. (*Ranks*: 1, 2, 3, 5, 5, 7).
- The sum of the ranks for N total observations

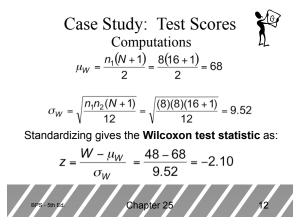


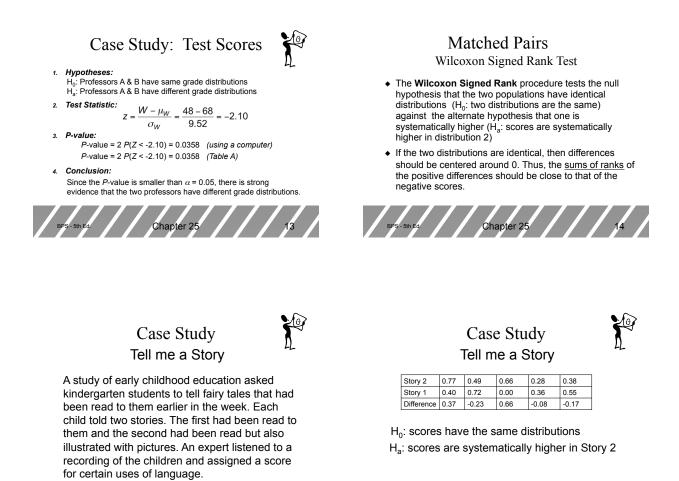
#### Comparing Two Samples Wilcoxon Rank Sum Test

- The Wilcoxon Rank Sum procedure tests the null hypothesis that the two populations have identical distributions. (H<sub>0</sub>: two distributions are the same)
- If the two distributions are identical, then samples of the same size should have roughly the same number of small values, the same number of medium values, and the same number of large values. Thus, the <u>sums of</u> <u>ranks</u> for each sample should be roughly the same.
- If instead, the sample sizes differ, then they should have proportionally similar sums of ranks.













THE WILCOXON SIGNED RANK TEST FOR MATCHED PAIRS

Case Study:	Tell me a Story	. 20
	Ranks	Π_

Absolute difference	0.08	0.17	0.23	0.37	0.66
Rank	1	2	3	4	5

#### The sum of ranks of positive differences is W+ = 9.

3.708

$$u_{w^*} = \frac{(5)(6)}{4} = 7.5$$
  $\sigma_{w^*} = \sqrt{\frac{(5)(6)(11)}{24}} = z = \frac{W^* - \mu_{w^*}}{\sigma_{w^*}} = \frac{9 - 7.5}{3.708} = 0.27$ 

μ

