

<div data-bbox="303 275 738 350" data-label="Section-Header"> <h2>Chapter 4: Alternatives to the t tools</h2> </div> <div data-bbox="409 365 587 392" data-label="Text"> <p>Class 7 2/23/09 M</p> </div>	<div data-bbox="815 132 1362 207" data-label="Section-Header"> <h3>Slide 1 Chapter 4: Alternatives to the t tools</h3> </div> <div data-bbox="815 294 941 327" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="300 693 699 730" data-label="Section-Header"> <h3>HW 5 due Weds 2/25/09 9:50</h3> </div> <div data-bbox="337 737 670 764" data-label="Text"> <p>Submit as Myname-HW5.doc (or *.rtf)</p> </div> <div data-bbox="251 766 719 987" data-label="List-Group"> <ul style="list-style-type: none"> • Finish Chapter 4 Wilcoxon rank sum, signed rank tests, Fisher's sign test, Welch's unequal variance t test • Comment on Chapter 4 conceptual problems in Blackboard Vista4 • Computation Problem 5 <ul style="list-style-type: none"> ▸ Problem 4.31 Effect of group therapy on breast cancer patients. </div>	<div data-bbox="815 657 1333 695" data-label="Section-Header"> <h3>Slide 2 HW 5 due Weds 2/25/09 9:50</h3> </div> <div data-bbox="815 781 941 816" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="293 1182 711 1222" data-label="Section-Header"> <h3>HW 6 due Monday 3/1/09 9:50</h3> </div> <div data-bbox="337 1224 670 1251" data-label="Text"> <p>Submit as Myname-HW6.doc (or *.rtf)</p> </div> <div data-bbox="251 1255 695 1428" data-label="List-Group"> <ul style="list-style-type: none"> • Read Chapter 5 Comparisons among several samples • Comment on Chapter 5 conceptual problems in Blackboard Vista4 • Computation Problem 6 <ul style="list-style-type: none"> ▸ Problem 4.30 Sunlight protection factor </div>	<div data-bbox="815 1146 1356 1186" data-label="Section-Header"> <h3>Slide 3 HW 6 due Monday 3/1/09 9:50</h3> </div> <div data-bbox="815 1268 941 1304" data-label="Text"> <p>NOTES:</p> </div>

HW 7 due Thursday 3/4/09 Noon

Submit as Myname-HW7.doc (or *.rtf)

- Read Chapter 6 **Comparisons among several samples**
- Comment on Chapter 6 conceptual problems in Blackboard Vista4
- Computation Problem 7
 - Problem 5.25 Duodenal ulcers

Slide 4 HW 7 due Thursday 3/4/09 Noon

NOTES:

Chapter 4: Alternatives to the *t* tools

Note: Sleuth has MANY errors and omissions!

- Permutation tests [Not a solution to unequal variance]
- Wilcoxon's Rank Sum Test (same probability model as Mann-Whitney U test)
 - Ties corrections not in sleuth & exact p values
- Repeated Measures Tests based on ranks: Wilcoxon Sign Rank & Fisher's Sign Test
- Parametric vs. Nonparametrics
 - Power efficiency not an issue, ties not that much of an issue
 - Dealing with covariates & estimating effect sizes can be an issue
 - Hodges-Lehman estimators
- Unequal variance (Welch's) *t* test: some theoretical and practical problems
- Supplemental material
 - Two-sample binomial test (covered in Sleuth Ch 19)
 - The Fligner-Policello test, a rank-based test for samples with unequal variance

Slide 5 Chapter 4: Alternatives to the *t* tools

NOTES:

Case 4.1: Space Shuttle O-Ring Failures

See Case 4.1 Movie, solved in Matlab™ & SPSS

Display 4.1

Numbers of O-ring incidents on 24 space shuttle flights prior to the Challenger disaster

Two problems: unequal variance & ties

Launch Temperature	Number of O-Ring Incidents
Below 65° F	1 1 1 3
Above 65° F	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 2

Summary of Statistical Findings

There is strong evidence that the number of O-ring incidents was associated with launch temperature in these 24 launches ...

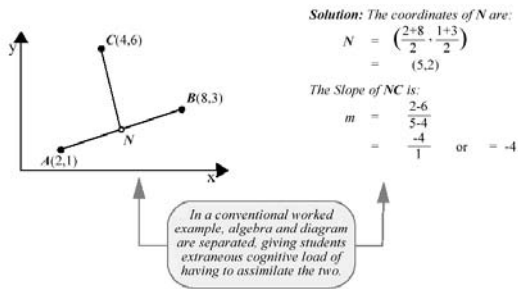
p-value = 0.009 from a **permutation test** on the *t* statistic

Slide 6 Case 4.1: Space Shuttle O-Ring Failures

NOTES:

Display 4.2

Cognitive load experiment: conventional method of instruction (for finding the slope of the line that connects C to the midpoint between A and B)

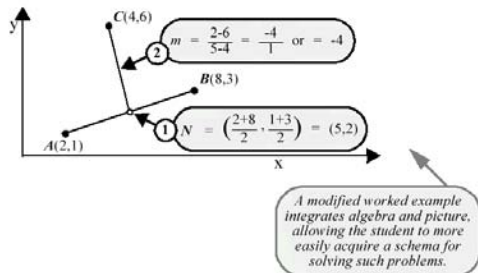


Slide 7 Case 4.2: Cognitive Load

NOTES:

Display 4.3

Cognitive load experiment: modified method of instruction (for finding the slope of the line that connects C to the midpoint between A and B)

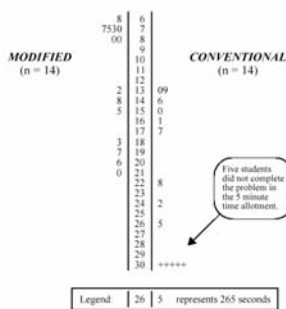


Slide 8

NOTES:

Display 4.4

Numbers of seconds to solution of a problem in coordinate geometry, for students instructed with conventional and modified materials



Problem: censored (truncated) data

Statistical Summary
There was convincing evidence that a student could solve the problem more quickly if taught with the modified method (1-sided p-value = 0.003 from the rank-sum test). The modified method shortened solution times by an estimated 152 s (95% CI 58 to 159 s)

Slide 9

NOTES:

Wilcoxon's rank-sum test

Analogous to 2-sample t test
Same test as the Mann-Whitney U test; different method for calculating test statistic, but identical results



Frank Wilcoxon of American Cyanamide [See Salsburg, 2001, The Lady Tasting Tea, for a brief history of his discovery of rank based statistics]

Slide 10 Wilcoxon's rank-sum test

NOTES:

Display 4.5

Rank-sum test statistic, T , for the cognitive load experiment

Y	Group	Order	Rank
68	M	1	1
70	M	2	2
73	M	3	3
75	M	4	4
77	M	5	5
80	M	6	6.5
80	M	7	6.5
130	C	8	8
132	M	9	9
139	C	10	10
146	C	11	11
148	M	12	12
150	C	13	13
155	M	14	14
161	C	15	15
177	C	16	16
183	M	17	17
197	M	18	18
206	M	19	19
210	M	20	20
228	C	21	21
242	C	22	22
265	C	23	23
300	C	24	26
300	C	25	26
300	C	26	26
300	C	27	26
300	C	28	26

⑥ Record the tied group membership patterns [2, 5]

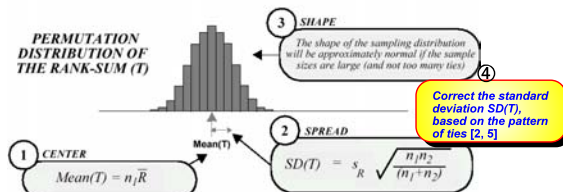
P-values identical to Mann-Whitney U test, different equations to obtain the same mathematically equivalent result

Slide 11 Wilcoxon Rank Sum Statistic

NOTES:

Display 4.6

Facts about the randomization (or sampling) distribution of the rank-sum statistic—the sum of ranks in group 1—when there is no group difference



where \bar{R} and s_R are the average and the sample standard deviation for the combined set of ranks (e.g. the 4th column of Display)

Note error: Sleuth doesn't include the ties correction for the variance of Wilcoxon T

Slide 12

NOTES:

Display 4.7

Finding the p-value with the normal approximation to the permutation distribution of the rank-sum statistic; calculations for the cognitive load data continued from .

① Calculate the average and sample standard deviation of the ranks from the combined sample (column 4 of Display)

$$\bar{R} = 14.5 \quad s_R = 8.202$$

② Compute the theoretical "null hypothesis" mean and standard deviation of T , using the formulas in

$$\text{Mean}(T) = (14)(14.5) = 203; \quad \text{SD}(T) = 8.202 \sqrt{\frac{14 \times 14}{(14+14)}} = 21.70$$

③ Determine the Z-statistic: $Z = \frac{(137 - 203)}{21.70} = -3.04$

④ Find the p-value from a standard normal table \rightarrow one-sided p-value = .00118

③ Correct the standard deviation $\text{SD}(T)$, based on the pattern of ties [2, 5]

Slide 13

NOTES:

Ties Correction for Rank Sum

From Hollander & Wolfe (1999); ties reduce $\text{Var}(W)$ and rate test

Ties

If there are ties, give tied observations the average of the ranks for which those observations are competing. After computing W using average ranks, use procedure (4.4), (4.5) or (4.6) and refer the value of W to Table A.6. Now, however, the test is approximate rather than exact. (To get an exact test, even in the tied case, see Comment 5.)

When applying the large-sample approximation, the following modification should be made. When there are ties, the null mean of W is unaffected, but the null variance is reduced to

$$\text{var}_0(W) = \frac{mn}{12} \left[m + n + 1 - \frac{\sum_{j=1}^g t_j(t_j - 1)(t_j + 1)}{(m + n)(m + n - 1)} \right], \quad (4.13)$$

or, equivalently,

$$\text{var}_0(W) = \frac{mn(N+1)}{12} - \left\{ \frac{mn}{12N(N-1)} \cdot \sum_{j=1}^g t_j(t_j - 1)(t_j + 1) \right\}, \quad (4.14)$$

To apply the large-sample approximation when ties are present, compute W using average ranks, and compute

$$W^* = \frac{W - [n(m + n + 1)/2]}{[\text{var}_0(W)]^{1/2}}$$

where $\text{var}_0(W)$ is given by display (4.13). With this modified value of W^* , approximations (4.10), (4.11) and (4.12) can be applied.

g = tied groups
 t_j = items in each tied group
Note, ties reduce the variance

EEOS611

Slide 14 Ties Correction for Rank Sum

NOTES:

Mann-Whitney U test

Wilcoxon (1945), Mann & Whitney (1947)

- The statistic U can be computed as follows (Hollander & Wolfe 1999)
 - For the two groups X_i and Y_j with m & n cases, consider each of the $m \times n$ pairs
 - For each pair of values X_i and Y_j , observe which is smaller.
 - If the X_i value is smaller, score a 1 for that pair. If the Y_j value is smaller, score a 0 for that pair.
 - Mann & Whitney showed that in the case of no ties:
 - $T = U + [n(n+1)/2]$, where T is the sum of ranks from the Wilcoxon rank sum test
 - Thus, the Wilcoxon & MW-U tests are exactly equivalent
 - When, X_i and Y_j are tied, score $\frac{1}{2}$

EEOS611

Slide 15 Mann-Whitney U test

NOTES:

SPSS's Mann-Whitney U test

C:\Program Files\SPSS\Help\Algorithms\mnpwr_tests.pdf

Mann-Whitney U Test

Calculation of Sums of Ranks

The combined data from both groups are sorted and ranks assigned to all cases, with average rank being used in the case of ties. The sum of ranks for each of the groups (S_1 and S_2) is calculated, as well as, for tied observations, $T_i = \frac{t^3 - t}{12}$, where t is the number of observations tied for rank i .

The average rank for each group is

$$\bar{S}_i = S_i / n_i$$

where n_i is the sample size in group i .

EEUS011

Slide 16 SPSS's Mann-Whitney U test

NOTES:

Test Statistic and Significance Level

The U statistic for group 1 is

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - S_1$$

• If $U > n_1 n_2 / 2$, the statistic used is

$$U' = n_1 n_2 - U$$

• If $n_1 n_2 \leq 400$ and $n_1 n_2 / 2 + \min(n_1, n_2) \leq 220$ the exact significance level based on an algorithm of Dineen and Blakesley (1973).

• The test statistic corrected for ties is

$$Z = \frac{(U' - n_1 n_2 / 2)}{\sqrt{\frac{n_1 n_2}{N(N-1)} \left(\frac{N^3 - N}{12} - \sum_i T_i \right)}}$$

which is distributed approximately as a standard normal. A two-tailed significance level is printed.

T_i
number of
terms in
each tied
group

Slide 17

NOTES:

Exact p values tabulated (no ties)

If no ties, exact P values tabulated

- CRC Handbook of Tables for Probability and Statistics
 - Tabulated values of Mann-Whitney U statistic
 - Can readily convert from sum of ranks of smaller group to U statistics
- Or, Hollander & Wolf's (1999) tabulated values of the Wilcoxon T statistic
- Note: the p values for tabulated exact tests are not appropriate if there are any tied ranks; but an exact p value can be calculated using all combinations of data (Gallagher provides a Matlab m.file implementing Hollander & Wolfe algorithm)

Slide 18 Exact p values tabulated (no ties)

NOTES:

Other alternatives for two independent samples

4.3.1 Permutation tests

Slide 19 Other alternatives for two independent samples

NOTES:

Ties and Case 4.1

Sleuth argues that rank tests inappropriate because of ties. The real problem is unequal variance (Behrens

Display 4.1

Numbers of O-ring incidents on 24 space shuttle flights prior to the Challenger disaster

Launch Temperature	Number of O-Ring Incidents
Below 65° F	1 1 1 3
Above 65° F	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 2

Sleuth argues that tied values pose problems for the Wilcoxon rank sum test, but Siegel (1956) states that the Wilcoxon rank sum test is robust in the presence of ties. But the Wilcoxon test p values are only approximate with ties, and the normal approximation is conservative. There is an exact test with ties (computer intensive but Gallagher has programmed).

Slide 20 Ties and Case 4.1

NOTES:

Display 4.1

Numbers of O-ring incidents on 24 space shuttle flights prior to the Challenger disaster

Launch Temperature	Number of O-Ring Incidents
Below 65° F	1 1 1 3
Above 65° F	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 2

Independent Samples Test				95% Confidence Interval of the Difference				Test Statistics ^b	
	t	df	Sig. (2-tailed)	Lower	Upper				Incident
Equal variances assumed	3.9	22	.00079	.61	2.0	Mann-Whitney U			6.000
						Wilcoxon W			216.000
						Z			-3.301
Equal variances not assumed	2.5	3.34	.07690	-.25	2.8	Asymp. Sig. (2-tailed)			.000963
						Exact Sig. [2*(1-tailed Sig.)]			.005082 ^a

Note the 100-fold difference in p values

^a. Not corrected for ties.
^b. Grouping Variable: Temperature

Slide 21

NOTES:

Wilcoxon rank sum test assumptions

Nonparametric tests distribution-free, NOT assumption free

- The observations of X_1, \dots, X_m are a random sample from population 1, independent and identically distributed. The observations of Y_1, \dots, Y_m are a random sample from population 2, independent and identically distributed
- The X's and Y's are mutually independent
- Populations 1 and 2 are continuous [i.e., no ties]
- **"Robustness of level: The significance level of the rank sum test is not preserved if the two populations differ in dispersion or shape. This is also the case for the normal theory 2-sample t test."** Hollander & Wolfe, p. 120

Slide 22 Wilcoxon rank sum test assumptions

NOTES:

Display 4.10

A summary of the t-statistics calculated from all 10,626 rearrangements of the O-ring data into a "Low" group of size 4 and a "High" group of size 20

Number of rearrangements with identical t-statistics	t-statistic	Total number of rearrangements into two groups of size 4 and 20: 10,626
2,380	-1.188	
3,400	-0.463	
2,040	0.231	
1,530	0.939	
855	1.716	
316	2.643	
95	3.888	
10	5.952	
		Number of rearrangements with t-statistics greater than or equal to 3.888: 105
		<u>1-sided p-value from a permutation test of the t-statistic:</u> $105/10626 = .00988$

This permutation test is invalid! The underlying t test assumes equal variances, and that problem is not corrected by using permutations. See Manly

Slide 23

NOTES:

$nCr=24$ Choose 4= $24!/((24-4)!*4!)$

=10.626 combinations

Display 4.10

A summary of the t-statistics calculated from all 10,626 rearrangements of the O-ring data into a "Low" group of size 4 and a "High" group of size 20

Number of rearrangements with identical t-statistics	t-statistic	Total number of rearrangements into two groups of size 4 and 20: 10,626
2,380	-1.188	
3,400	-0.463	
2,040	0.231	
1,530	0.939	
855	1.716	
316	2.643	
95	3.888	
10	5.952	
		Number of rearrangements with t-statistics greater than or equal to 3.888: 105
		<u>1-sided p-value from a permutation test of the t-statistic:</u> $105/10626 = .00988$

Which test is appropriate?

Independent samples t test: exceptionally strong evidence against null ($p=0.00038$); Wilcoxon's rank sum test [with ties correction] (0.000963, very strong evidence); ~~Permutation test~~ (strong evidence 0.0098); ~~Unequal variance t test~~ (some evidence, 0.038)

Slide 24 $nCr=24$ Choose 4= $24!/((24-4)!*4!)$

NOTES:

Gallagher's Matlab solutions, normal approximation with ties correction & exact test

Randomization test: p-value = 0.00988
 >>[pvalue,W,U]=Wilcoxranksun(X,Y,0)
 pvalue = **9.6335e-004** [With ties correction;
 identical to SPSS approximation]
 W = 84 U = 74

Gallagher's exact Wilcoxon rank sum test in Matlab
 [algorithms from Hollander & Wolfe (1999)]
 >> [pvalue,W,U]=Wilcoxranksun(X,Y,1)
 2-tailed pvalue = 0.0038 or **9.4 times Wilcoxon rank sum
 with ties correction; but 75% of SPSS exact value without
 ties correction (0.00508)**

Launch Temperature	Number of O-Ring Incidents
Below 65° F	1 1 1 3
Above 65° F	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 2

EEOS611

Slide 25 Gallagher's Matlab solutions, normal approximation with ties correction & exact test

NOTES:

Test Statistics^b

	Incident
Mann-Whitney U	6.000
Wilcoxon W	216.000
Z	-3.301
Asymp. Sig. (2-tailed)	.00096
Exact Sig. [2*(1-tailed Sig.)]	.00508 ^a

a. Not corrected for ties.

b. Grouping Variable: Temperature

**All versions of SPSS, including Version 14:
 The exact tests in SPSS are not corrected for ties!**

Slide 26 SPSS solution with Wilcoxon Rank sum test, not conservative

NOTES:

Non-parametric tests are distribution-free, not assumption free

- No specific distributional assumptions, like normally distributed errors, but all nonparametric tests have assumptions
- Mann-Whitney U, Underwood 1997, p. 131, "MW/Wilcoxon has nearly identical assumptions to Student's t test"
- Zar (1999, p. 49) the test is not particularly sensitive to differences in dispersion
 - Gallagher: Not true in my experience
 - Matlab simulation program available

EEOS611

Slide 27 Non-parametric tests are distribution-free, not assumption free

NOTES:

Randomization doesn't solve problems with unequal variance

- Randomization is often superior to the t -distribution for 2-sample problems. It does not remedy the common problems with the t distributions though.
- The most common problem with Student's t test is the so-called Fisher-Behrens problem: testing for differences in the average if the distributions have different variances
 - This is an open question
 - **Neither** Wilcoxon Rank sum tests **nor** randomization provide a clear solution



EEOS611

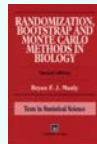
Slide 28 Randomization doesn't solve problems with unequal variance

NOTES:

Neither randomization nor permutation tests solve the unequal variance problem

Manly (1997, p. 141)

- "The randomization test for the difference in two means can be upset if the samples come from sources that have the same mean, but different variances. This is apparent because the null hypothesis for the randomization test is that the samples come from exactly the same source, which is not true if the variances are not constant"
 - A variety of modifications have been proposed, but all require further study.
- O-ring data may not be a test between mean failure rates!



EEOS611

Slide 29 Neither randomization nor permutation tests solve the unequal variance problem

NOTES:

Gallagher's Matlab Case0401b.m

Exact tests based on Student's t test
Why not just use a 2-sample binomial test?

```
>> X = 1 1 1 3; Y = 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 2
```

The observed mean difference in O-ring failures was 1.3 incidents per launch with 95% CI of [0.6 2.0]; The t statistic was 3.9 with 22 df and 2-tailed p of 0.00079;

Exact tests: The total number of ways of selecting 4 items from 24 items is 10626; The 1-tailed probability of observing a t statistic greater than observed (3.9) is **0.009881**; The exact 1-tailed p value is: 5/506; The 1-tailed probability of observing a difference greater than observed (1.3) is **0.009881**; The 2-tailed probability of observing a t statistic with abs value greater than observed (3.9) is **0.009881**... The exact two-tailed probability for the Wilcoxon rank sum test is **0.003764** The normal approximation for the 2-tailed probability for the Wilcoxon rank sum test is 0.000963;

The probability of an O ring incident if cold ($<65^\circ\text{F}$) was $4/4=1.00$; The probability of an O ring incident in warm ($\geq 65^\circ\text{F}$) was $3/20=0.15$; **The two sample binomial test for equal proportions (0.29) of failure has a 2-sided p value of 0.000640**

Slide 30 Gallagher's Matlab Case0401b.m

NOTES:

<div data-bbox="352 170 647 203" data-label="Section-Header"> <h3>Matlab Case0401b.m</h3> </div> <div data-bbox="393 214 617 237" data-label="Section-Header"> <h4>Summary of conclusions</h4> </div> <div data-bbox="233 235 725 277" data-label="Text"> <p>The exact two-tailed probability for the Wilcoxon rank sum test is 0.003764 (Can't use: not robust to unequal variances)</p> </div> <div data-bbox="233 277 730 340" data-label="Text"> <p>The normal approximation for the 2-tailed probability for the Wilcoxon rank sum test is 0.000963 (but this result should NOT be used — it is a large sample approximation)</p> </div> <div data-bbox="233 340 347 363" data-label="Text"> <p>Binomial test:</p> </div> <div data-bbox="233 365 734 390" data-label="Text"> <p>The probability of an O ring incident if cold (<65F) was $4/4=1.00$</p> </div> <div data-bbox="233 390 682 432" data-label="Text"> <p>The probability of an O ring incident in warm ($\geq 65F$) was $3/20=0.15$</p> </div> <div data-bbox="233 434 725 476" data-label="Text"> <p>The two-sample binomial test for equal proportions ($p=0.29$) of failure has a 2-sided p value of 0.000640</p> </div>	<div data-bbox="815 132 1240 168" data-label="Section-Header"> <h3>Slide 31 Matlab Case0401b.m</h3> </div> <div data-bbox="815 256 940 289" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="358 657 656 693" data-label="Section-Header"> <h3>Fligner-Policello test</h3> </div> <div data-bbox="284 699 714 745" data-label="Text"> <p>Wilcoxon-rank sum test for unequal variances Matlab m.file available, $p < 3 \times 10^{-14}$ for O-ring data!</p> </div> <div data-bbox="238 749 742 819" data-label="Text"> <p>% Trujillo-Ortiz, A., F. A. Trujillo-Rodriguez, R. Hernandez-Walls, M. A. Fligner and S. Perez-Osuna (2003), Fptest: Non-parametric Fligner-Policello test of two combined random variables with continuous cumulative distribution. A MATLAB file.</p> </div> <div data-bbox="238 819 602 856" data-label="Text"> <p>% [WWW document]. URL http://www.mathworks.com/matlabcentral/fileexchange/</p> </div> <div data-bbox="238 854 568 877" data-label="Text"> <p>% loadFile.do?objectId=4226&objectType=FILE</p> </div> <div data-bbox="238 877 347 898" data-label="Text"> <p>% References:</p> </div> <div data-bbox="238 898 737 980" data-label="Text"> <p>Fligner, M. A. and Policello, G. E. (1981), Robust rank procedure for the Behrens-Fisher Problem. Journal of the American Statistical Association, 76(373): 162-168. Hollander, M. and Wolfe, D. (1999), Nonparametric Statistical Methods (2nd ed.), New York: John Wiley & Sons, Inc. p. 135-139.</p> </div>	<div data-bbox="815 623 1232 659" data-label="Section-Header"> <h3>Slide 32 Fligner-Policello test</h3> </div> <div data-bbox="815 743 940 777" data-label="Text"> <p>NOTES:</p> </div>
<div data-bbox="305 1148 716 1182" data-label="Section-Header"> <h3>Asymptotic Power Efficiency</h3> </div> <div data-bbox="284 1188 737 1232" data-label="Text"> <p>Ratio of sample sizes required to obtain the same p values</p> </div> <div data-bbox="246 1228 737 1472" data-label="List-Group"> <ul style="list-style-type: none"> For normally distributed data, the power efficiency of the Wilcoxon rank sum test is 95.5% of the Student's t test. <ul style="list-style-type: none"> For other distributions (e.g., exponential distributions), the power efficiency can be $\gg 100\%$ (300% for exponential) <ul style="list-style-type: none"> Hollander & Wolfe p. 140 Strengths of Wilcoxon's Rank-sum test <ul style="list-style-type: none"> Resistant to outliers Can handle censored data Weakness: generality, determining effect sizes & confidence limits </div> <div data-bbox="652 1482 781 1514" data-label="Text"> <p>EEOS611</p> </div>	<div data-bbox="815 1113 1347 1148" data-label="Section-Header"> <h3>Slide 33 Asymptotic Power Efficiency</h3> </div> <div data-bbox="815 1232 940 1266" data-label="Text"> <p>NOTES:</p> </div>

Measuring effects sizes

Difficult with nonparametric procedures

Display 4.8

Using a rank-sum test to construct a confidence interval for an additive treatment effect; cognitive load study

Hypothesized Effect (seconds)	2-sided p-value	Confidence Interval Inclusion?
-50	.0286	No
-60	.0800	Yes
-55	.0403	No
-58	.0502	Yes
-150	.1227	Yes
-160	.0476	No
-155	.0589	Yes
-158	.0530	Yes
-159	.0502	Yes

Try several hypothesized values for δ to identify those that have 2-sided p-values $\geq .05$

A 95% confidence interval is -159 seconds to -58 seconds.

EEOS611

Slide 34 Measuring effects sizes

NOTES:

Hodges-Lehman estimator for 95% CI

Add a fixed amount to one of the groups: Sleuth's

Display 4.8

Using a rank-sum test to construct a confidence interval for an additive treatment effect; cognitive load study

Hypothesized Effect (seconds)	2-sided p-value	Confidence Interval Inclusion?
-50	.0286	No
-60	.0800	Yes
-55	.0403	No
-58	.0502	Yes
-150	.1227	Yes
-160	.0476	No
-155	.0589	Yes
-158	.0530	Yes
-159	.0502	Yes

Try several hypothesized values for δ to identify those that have 2-sided p-values $\geq .05$

A 95% confidence interval is -159 seconds to -58 seconds.

EEOS611

Slide 35 Hodges-Lehman estimator for 95% CI

NOTES:

Unequal variance t test

Welch's t test with Satterthwaite approximation for d.f.

Slide 36 Unequal variance t test

NOTES:

Recall the equal variance t test

Pooled estimate of the standard deviation

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1 + n_2 - 2)}}, \quad \text{d.f.} = n_1 + n_2 - 2.$$

The degrees of freedom going into this estimate are the combined degrees of freedom from the individual estimates: $(n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$.

Standard Error for the Difference

The standard deviation for the difference between averages from two independent samples is given in Display 2.6. If the two populations have equal spread, the common value of the variance factors from the two terms under the radical can be removed from under the radical as the common standard deviation. Using the pooled standard deviation as its estimate, the standard error for the difference in sample averages is:

$$SE(\bar{Y}_2 - \bar{Y}_1) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

EEOS611

Slide 37 Recall the equal variance t test

NOTES:

Recall Display 2.8, page 40

Equal variance t test standard error

Calculation of the pooled estimate of SD and the standard error for the difference between two sample averages: Bumpus' data

1 SUMMARY STATISTICS			
Group	n	Average (in.)	Sample SD (in.)
1: Died	24	0.72792	0.02354
2: Survived	35	0.73800	0.01984

2 THE POOLED SD			
s_p	=	$\sqrt{\frac{(24-1)(0.02354)^2 + (35-1)(0.01984)^2}{(24+35-2)}}$	
	=	$\sqrt{\frac{0.00128}{57}}$	These are the degrees of freedom associated with the pooled SD.
	=	$\sqrt{0.004584}$	This is the pooled variance.
Answer $\Rightarrow s_p$	=	0.02141 inches	

3 THE STANDARD ERROR			
$SE(\bar{Y}_2 - \bar{Y}_1)$	=	$0.02141 \sqrt{\frac{1}{24} + \frac{1}{35}}$	
	=	0.00567 inches	Answer

SE of difference used for 95% CIs

EEOS611

Slide 38 Recall Display 2.8, page 40

NOTES:

The Mechanics of Confidence Interval Construction

A confidence interval with confidence level $100(1 - \alpha)\%$ is the following:

Equal variance t test

$100(1 - \alpha)\%$ Confidence Limits for the Difference Between Means

$$(\bar{Y}_2 - \bar{Y}_1) \pm t_{df}(1 - \alpha/2)SE(\bar{Y}_2 - \bar{Y}_1).$$

Display 2.8 Construct...n of a 95% confidence interval for the difference between the mean humerus lengths of sparrows that died and those that survived

Group	n	Average (in.)	SD (in.)
1: Died	24	0.72792	0.02354
2: Survived	35	0.73800	0.01984

$\bar{Y}_2 - \bar{Y}_1 = 0.73800 - 0.72792 = 0.01008$

$SE(\bar{Y}_2 - \bar{Y}_1) = 0.00567$ inches ← From Display 2.6

Degrees of freedom = $24 + 35 - 2 = 57$ ← From tables of the t-distribution with 57 degrees of freedom

$t_{57}(.975) = 2.002$

Half-width = $(2.002)(0.00567) = 0.01136$

Lower 95% confidence limit = $0.01008 - 0.01136 = -0.00128$ inches

Upper 95% confidence limit = $0.01008 + 0.01136 = 0.02144$ inches

Slide 39

NOTES:

4.3.2 The Welch t-Test for Comparing Two Normal Populations with Unequal Spreads

Welch's *t*-test employs the individual sample standard deviations as separate estimates of their respective population standard deviations, rather than pooling to obtain a single estimate of a population standard deviation. The result is a different formula for the standard error of the difference in averages:

$$SE_W(\bar{Y}_2 - \bar{Y}_1) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

This becomes the denominator in the *t*-statistic for comparing the means of populations with different spreads. Even when the populations are normal, however, the exact sampling distribution of the Welch *t*-ratio is unknown. It can be approximated by a *t*-distribution with *d.f.*_W degrees of freedom, known as Satterthwaite's approximation:

$$d.f._W = \frac{[SE_W(\bar{Y}_2 - \bar{Y}_1)]^4}{\frac{[SE(\bar{Y}_2)]^4}{(n_2 - 1)} + \frac{[SE(\bar{Y}_1)]^4}{(n_1 - 1)}}$$

where

$$SE(\bar{Y}_1) = \frac{s_1}{\sqrt{n_1}} \quad \text{and} \quad SE(\bar{Y}_2) = \frac{s_2}{\sqrt{n_2}}$$

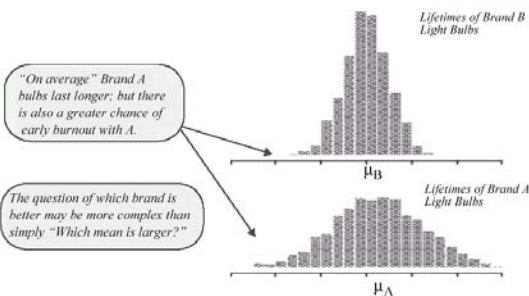
Note the reduction in degrees of freedom. This is the Satterthwaite approximation

Slide 40 Welch's t test, p. 97 in Sleuth

NOTES:

Display 4.11

The conceptual difficulty with comparing population means when population spreads are not the same



Slide 41

NOTES:

Problems with unequal variance *t* tests (Welch *t* test)

- Conceptual difficulties interpreting difference in central tendency if variances unequal
- Exact distribution of test statistic unknown, the Satterthwaite *d.f.* is an approximation, Usually non-integer *d.f.*
- Test isn't necessarily conservative
- Doesn't generalize easily to more than 2 groups
- There are alternate procedures to Welch
 - Variance-stabilizing transformations
 - Nonparametric tests based on ranks: Fligner test
 - Note that this uses the large-sample normal approximation
 - Probably not appropriate for small sample sizes

Slide 42 Problems with unequal variance *t* tests (Welch *t* test)

NOTES:

Example: Stream temperatures**Unequal variance t test not necessarily conservative**

- Temperatures taken from different portions of a stream:
 - Portion 1: 15.8, 16.9, 17, 17.1, 18, 18.7
 - mean = 17.25, variance = 0.995
 - Portion 2: 18.3, 18.5
 - mean = 18.4, variance = 0.02
- Obviously the variances are unequal and an equal variance 2-sample t test may be inappropriate
 - Welch [unequal variance]: $t = 2.74$ w/ 5 df $p = 0.037$.
 - Pooled [equal variance]: $t = 1.54$ w/ 6 df $p = 0.17$.
- Why is the equal-variance t-test giving a lower t-value and a higher p value?

Slide 43 Example: Stream temperatures

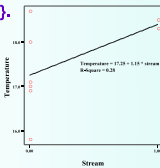
NOTES:

The avg temp different: Exact test**P=3/28, 1-tailed**

There are only 28 ways the 8 temperatures can be arranged into groups of 6 and 2 [8 Choose 2], and in only 3 of these arrangements would the difference in means be equal or greater than the 1.15 °C difference observed. These 3 arrangements include the observed data and two others: {18.3, 18.5}, {18.3, 18.7}, {18.5, 18.7}.

P=3/28=0.107

This is the appropriate p value, unless you argue that the variances are different between the 2 portions of stream, but there are too few data to provide strong evidence for this

**Slide 44 The avg temp different: Exact test**

NOTES:

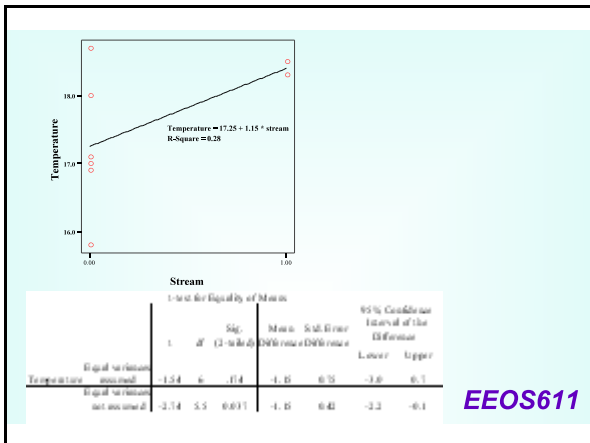
Levene's test, Section 4.5.3**Three different Levene's tests in the literature**

- Sleuth's Levene's test on page 102-103 is not the same as the Levene's test used by SPSS in Student's t test.
 - Sleuth: squared deviations from the mean used as the variables in a t test
 - SPSS: absolute values used
 - SPSS: |observations-mean| used in an F test
 - Other Leven tests |observations-median| used in t or F test
- The results can often be quite different
- Levene's tests have largely replaced the F_{\max} and Bartlett's tests for equal variance

Slide 45 Levene's test, Section 4.5.3

NOTES:

Slide 46



NOTES:

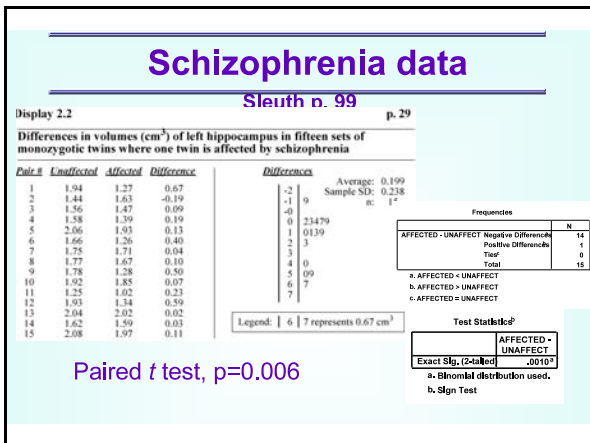
Slide 47 4.4 Alternatives to the paired t test

4.4 Alternatives to the paired t test

Wilcoxon sign-rank and Fisher Sign tests

NOTES:

Slide 48 Schizophrenia data



NOTES:

Slide 49 Wilcoxon signed rank test

NOTES:

Wilcoxon signed rank test

Display 4.12

Signed-rank test statistic computations: schizophrenia study

Pair	Unaffected	Affected	Difference	Ordered Magnitude	Order	Rank	+Ranks	-Ranks
1	1.94	1.27	.67	.02 (+)	1	1	1	
2	1.44	1.63	-.19	.03 (+)	2	2		2
3	1.56	1.47	.09	.04 (+)	3	3	3	
4	1.58	1.39	.19	.07 (+)	4	4	4	
5	2.06	1.93	.13	.09 (+)	5	5	5	
6	1.66	1.26	.40	.10 (+)	6	6	6	
7	1.75	1.71	.04	.11 (+)	7	7	7	
8	1.77	1.67	.10	.13 (+)	8	8	8	
9	1.78	1.28	.50	.19 (+)	9	9.5	9.5	
10	1.92	1.85	.07	.19 (-)	10			9.5
11	1.25	1.02	.23	.23 (+)	11	11	11	
12	1.93	1.34	.59	.40 (+)	12	12	12	
13	2.04	2.02	.02	.50 (+)	13	13	13	
14	1.62	1.59	.03	.59 (+)	14	14	14	
15	2.08	1.97	.11	.67 (+)	15	15	15	

SPSS:
Discard
sample
pairs with
equal
values*

Correct the standard
deviation SD(T), based
on the pattern of ties [2,
5]

1 Order the absolute differences and assign ranks to them

2 Signed rank statistics = sum of ranks for positive differences:

= 110.5

Slide 50 Dealing with tied pairs

NOTES:

Dealing with tied pairs

Two sorts of ties in the signed rank test

There are two sorts of ties with the signed rank test. If you have identical values in both pairs, Wilcoxon recommended that those paired observations be dropped from the analysis. That is still the standard recommendation. There is another sort of tie resulting after the absolute values of the differences between paired observations are ranked. You could have two or more differences with the same absolute values. Those ties are not discarded, and the variance formula is adjusted to take into account the number of tied groups [See next slide]

Hollander and Wolfe's Nonparametric statistics, 2nd ed (p. 46) covers the problem of dropping ties of the first sort. If there are many ties, H & W recommend using another test. They also state that you could leave the tied samples in, and use a random number generator to randomly assign positive or negative differences. This apparently is discussed in Pratt (1959). If you want a more conservative 1-sided test, assign all of the tied differences to the group that would make it less likely to reject the null. For example, if you are testing lipitor's effects on cholesterol and a patient had identical cholesterol levels before and after, then assign that difference as if the lipitor blood sample had the higher cholesterol or the placebo value had the lower cholesterol. If you still reject the null, your conservative test would be less likely to result in a Type I error, but of course the probability of Type II error (accepting a false null would be increased). Pratt (1959), cited in both Lehmann and Hollander & Wolfe, provides a more thorough review. Lehmann cites a couple of more recent papers on dealing with the 1st sort of ties in paired rank tests.

Slide 51 SPSS algorithms, signed rank test

NOTES:

SPSS algorithms, signed rank test

There are exact tests, if no tied ranks

$$Z = \frac{\min(S_p, S_n) - (n(n+1)/4)}{\sqrt{n(n+1)(2n+1)/24 - \sum_{j=1}^l (t_j^3 - t_j)/48}}$$

L = tied groups
t_j = items in each tied group

where

Asymptotic relative efficiency > 0.864,
95.5% for normally distributed data

n Number of cases with non-zero differences

l Number of ties

t_j Number of elements in the j-th tie, j = 1, ..., l

Fisher's sign test

Straightforward application of the 2-sample binomial test

- Given that the probability of a + sign = probability of a minus sign = 0.5,
- What is the probability of observing exactly k positive signs in n Bernoulli (binomial) trials
- $P(X=k) = n \text{ Choose } k \cdot p^k (1-p)^{n-k}$
 - X has a binomial distribution
 - Must sum probability for observed value of k , and all more extreme values of k .
- Statistical sleuth provides only the normal approximation to the binomial, but SPSS will provide the exact test for $n < 30$.

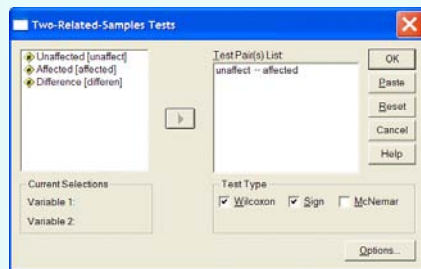
Frequencies			N
AFFECTED - UNAFFECT	Negative Differences		14
	Positive Differences		1
	Ties ^a		0
	Total		15
a. AFFECTED < UNAFFECT			
b. AFFECTED > UNAFFECT			
c. AFFECTED = UNAFFECT			
Test Statistics ^b			
	AFFECTED -		
	UNAFFECT		
Exact Sig. (2-tailed)			.0010 ^a
a. Binomial distribution used.			
b. Sign Test			

EEOS611

Slide 52 Fisher's sign test

NOTES:

Sign test in SPSS



EEOS611

Slide 53 Sign test in SPSS

NOTES:

Conclusions (1 of 2)

Chapter 4 Alternatives to the t tools

- Consider using alternatives to the t tools if
 - The assumptions are grossly violated or
 - The sample sizes are too small to test distributional assumptions
- Wilcoxon rank sum test
 - Appropriate for small sample sizes
 - Appropriate in the presence of outliers
 - Ties are not a problem if the ties-correction used
 - Not appropriate for samples with unequal variances (try Fligner-Policello if the sample sizes are large)

EEOS611

Slide 54 Conclusions (1 of 2)

NOTES:

Slide 55 Conclusions (2 of 2)	
<div> <div>Conclusions (2 of 2)</div> <div>Chapter 4 Alternatives to the t tools</div> <ul style="list-style-type: none"> Permutation test <ul style="list-style-type: none"> Appropriate for small sample sizes, when the Student's t distribution might not be appropriate Does not protect against the problem of unequal variances (the Fisher-Behrens problem) Paired data: tests based on ranks <ul style="list-style-type: none"> Wilcoxon signed rank test: high power efficiency Sign test, simple application of the 1-sample binomial <div>EEOS611</div> </div>	
	NOTES: