

Statistics for Medical and Public Health Researchers

Chapter 8: Tests For Comparing Two Groups of Categorical or Continuous Data

JEFF GILL

Professor of Political Science

Professor of Biostatistics

Professor of Surgery (Public Health Sciences)

Washington University, St. Louis

Background

- This chapter gives additional procedures for comparing two groups:

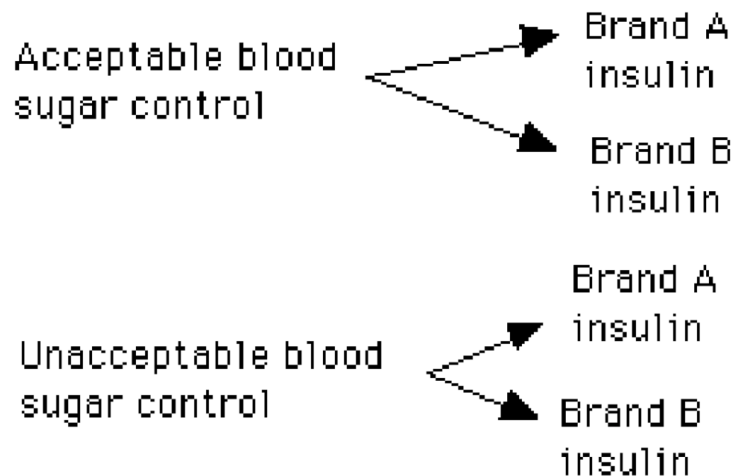
	<i>Level of Measurement</i>		
	Continuous	Ordinal	Nominal
Paired Data			
Independent (Unpaired) Data			

under weaker assumptions:

- ▷ different distributional assumptions due smaller sample sizes providing a weaker justification for normally distributed statistics (CLT),
- ▷ less precise levels of measurement than last chapter (eg. above table),
- ▷ and some experimental settings that impose more heterogeneity on sample comparisons.

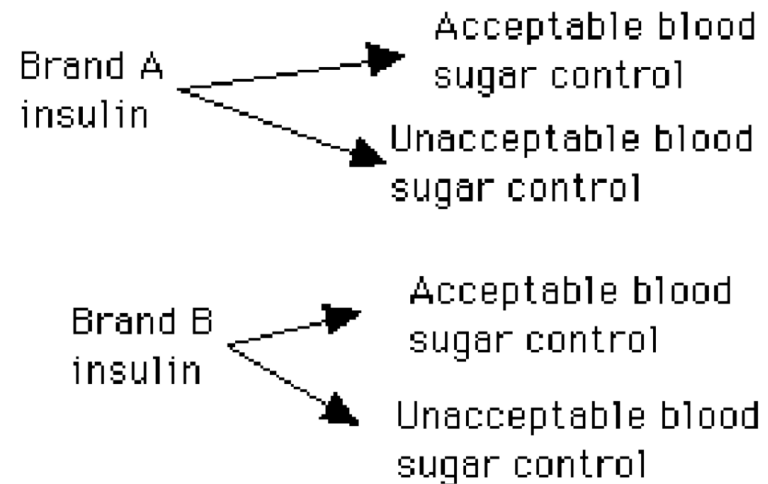
Reminder: Case Control Studies Versus Cohort Studies

CASE-CONTROL STUDY DESIGN



Sampling determined by different levels of *health status*.

COHORT STUDY DESIGN

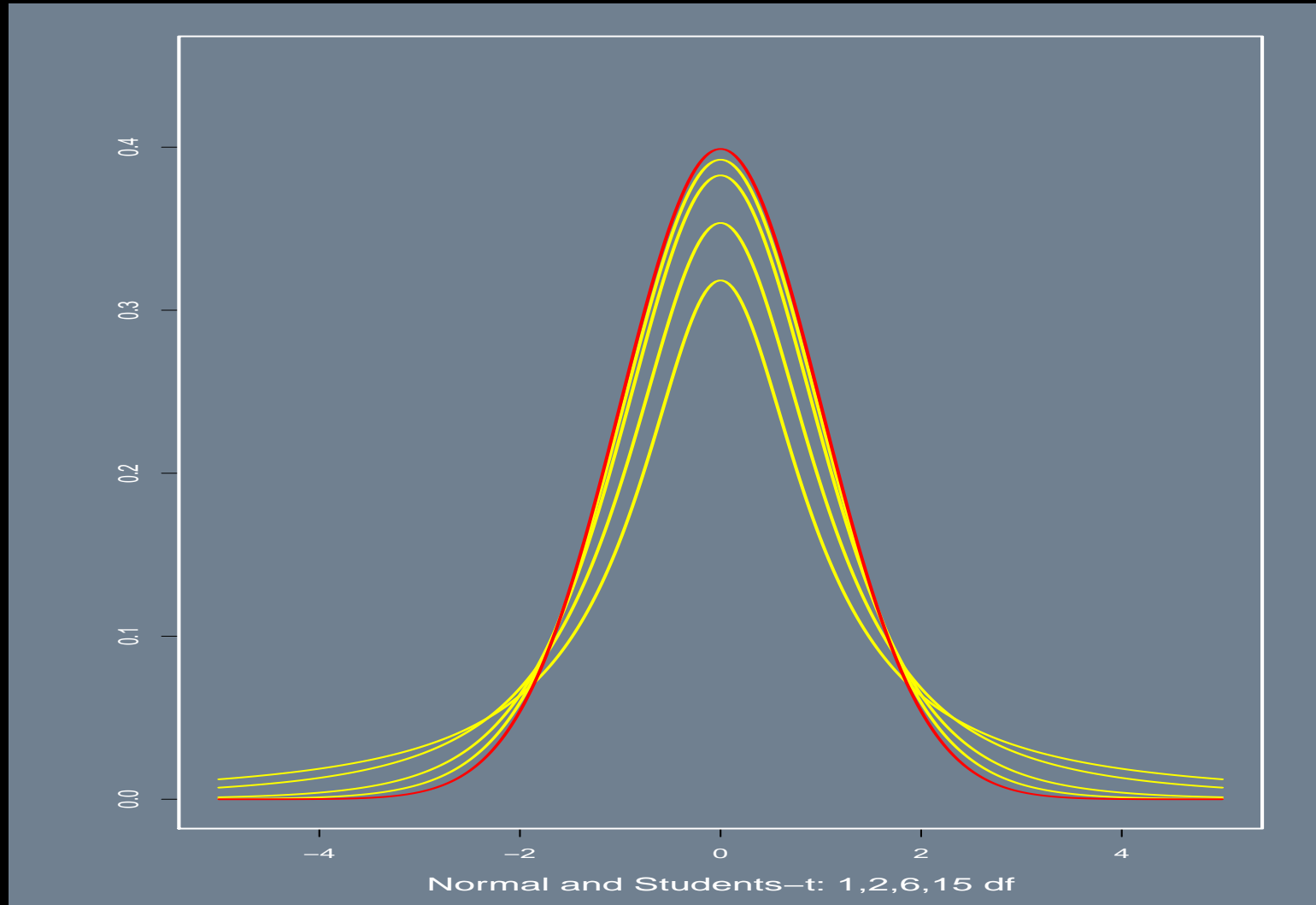


Sampling determined at different levels of *exposure*.

Similarities and Differences Between Chapter 7 and Chapter 8

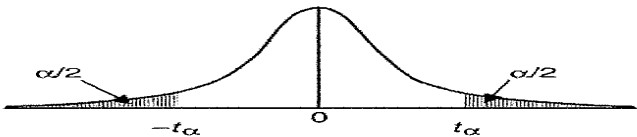
- ▶ We will still be doing hypothesis testing in the same general way.
- ▶ In **Chapter 7** we split cohorts with assumed similarities (*physiotherapy for lung cancer patients, post-natal urinary incontinence, change in blood pressure for the same person, difference in fatigue severity scores*).
- ▶ The perceived similarities before treatment along with the (relatively large sample sizes) allowed us to use the normal distribution assumption for our distributional tests.
- ▶ **Chapter 8** moves on to groups with assumed greater heterogeneity and smaller sample sizes requiring the students- t distribution instead of the normal distribution.
- ▶ The students- t distribution requires a **degrees of freedom** parameter, which is a function of sampling size.
- ▶ There are no hard and fixed “rules,” about model choice here; textbooks vary and you must decide how relatively different your groups are biologically or medically.

Students- t Versus the Normal



Students-*t* Table

Table T3 Student's *t*-distribution. The value tabulated is t_α such that if X is distributed as Student's *t*-distribution with df degrees of freedom, then α is the probability that $X \leq -t_\alpha$ or $X \geq t_\alpha$



df	α							
	0.20	0.10	0.05	0.04	0.03	0.02	0.01	0.001
1	3.078	6.314	12.706	15.895	21.205	31.821	63.657	636.6
2	1.886	2.920	4.303	4.849	5.643	6.965	9.925	31.60
3	1.634	2.353	3.182	3.482	3.896	4.541	5.842	12.92
4	1.530	2.132	2.776	2.999	3.298	3.747	4.604	8.610
5	1.474	2.015	2.571	2.757	3.003	3.365	4.032	6.869
6	1.439	1.943	2.447	2.612	2.829	3.143	3.707	5.959
7	1.414	1.895	2.365	2.517	2.715	2.998	3.499	5.408
8	1.397	1.860	2.306	2.449	2.634	2.896	3.355	5.041
9	1.383	1.833	2.262	2.398	2.574	2.821	3.250	4.781
10	1.372	1.812	2.228	2.359	2.528	2.764	3.169	4.587
11	1.363	1.796	2.201	2.328	2.491	2.718	3.106	4.437
12	1.356	1.782	2.179	2.303	2.461	2.681	3.055	4.318
13	1.350	1.771	2.160	2.282	2.436	2.650	3.012	4.221
14	1.345	1.761	2.145	2.264	2.415	2.624	2.977	4.140
15	1.340	1.753	2.131	2.249	2.397	2.602	2.947	4.073
16	1.337	1.746	2.120	2.235	2.382	2.583	2.921	4.015
17	1.333	1.740	2.110	2.224	2.368	2.567	2.898	3.965
18	1.330	1.734	2.101	2.214	2.356	2.552	2.878	3.922
19	1.328	1.729	2.093	2.205	2.346	2.539	2.861	3.883
20	1.325	1.725	2.086	2.196	2.336	2.528	2.845	3.850
21	1.323	1.721	2.079	2.189	2.327	2.517	2.830	3.819
22	1.321	1.717	2.074	2.183	2.320	2.508	2.818	3.790
23	1.319	1.714	2.069	2.178	2.313	2.499	2.806	3.763
24	1.318	1.711	2.064	2.172	2.307	2.492	2.797	3.744
25	1.316	1.708	2.059	2.166	2.301	2.485	2.787	3.722
26	1.315	1.706	2.056	2.162	2.396	2.479	2.779	3.706
27	1.314	1.703	2.052	2.158	2.291	2.472	2.770	3.687
28	1.313	1.701	2.048	2.154	2.286	2.467	2.763	3.673
29	1.311	1.699	2.045	2.150	2.282	2.462	2.756	3.657
30	1.310	1.697	2.042	2.147	2.278	2.457	2.750	3.646
∞	1.282	1.645	1.960	2.054	2.170	2.326	2.576	3.291

Running Example

- ▶ “Cost effectiveness of community leg ulcer clinics: randomised controlled trial.” C Jane Morrell, Stephen J Walters, Simon Dixon, Karen A Collins, Louise M L Brereton, Jean Peters, Charles G D Brooker (BMJ VOLUME 316 16 MAY 1998).
- ▶ **Objectives:** To establish the relative cost effectiveness of community leg ulcer clinics that use four layer compression bandaging versus usual care provided by district nurses.
- ▶ **Design:** Randomised controlled trial with 1 year of follow up.
- ▶ **Setting:** Eight community based research clinics in four trusts in Trent.
- ▶ **Subjects:** 233 patients with venous leg ulcers allocated at random to intervention (120) or control (113) group.
- ▶ **Interventions:** Weekly treatment with four layer bandaging in a leg ulcer clinic (clinic group) or usual care at home by the district nursing service (control group).

Running Example

► **Outcome Data:**

- ▷ time to complete ulcer healing in weeks
- ▷ ulcer status (healed or unhealed) at 3 months and 12 months
- ▷ total ulcer free weeks
- ▷ “health-related quality of life (HRQoL) at 0, 3, 12 months, measured from 0 (very poor health) to 100 (very good health)
- ▷ some others not considered in the textbook.

► **Standard Thresholds:** “To have an 80% chance of detecting as significant (at the 5% level) an increase in healing from 50% to 70%, 206 patients were required.”

► **Procedures:** “We used t tests or Mann-Whitney tests for continuous measurements, χ^2 tests for categorical data, the Kaplan-Meier method to calculate the time from recruitment for the initial leg ulcer to heal, and the log rank test to compare the healing times of both groups.”

Morrell, *et al.*

Table 1 Baseline characteristics of patients recruited. Values are numbers (percentages) of patients unless stated otherwise

Characteristic	Clinic group (n=120)*	Home group (n=113)†
Mean (SD) age (years)	73.8 (10.9)	73.2 (11.6)
Mean (SD) area of ulcer (cm ²)	16.2 (28.9)	16.9 (40.8)
Mean (SD) occurrence of first ulcer (years)	14.2 (14.9)	12.9 (15.9)
Mean (SD) maximum duration of current ulcer (months)	27.5 (53.8)	29.7 (82.3)
Mean (SD) body mass index (kg/m ²)	27.0 (6.7)	27.1 (6.0)
Mean (SD) systolic brachial pressure (mm Hg)	154.8 (28.7)	153.9 (24.9)
Mean (SD) ankle brachial pressure index	1.08 (0.17)	1.10 (0.18)
Women	77 (64)	78 (69)
Married	56 (47)	52 (46)
Living alone	30 (25)	33 (29)
Walking with aid	66 (55)	57 (50)
Family history of leg ulcers	82 (28)	25 (22)
History of deep vein thrombosis	28 (23)	25 (22)
Varicose veins	77 (64)	76 (67)
Rheumatoid arthritis	14 (12)	18 (16)
Heart problems	13 (11)	8 (7)
Osteoarthritis	47 (39)	40 (35)
Diabetes mellitus	8 (7)	10 (9)

*Mean values measured in at least 110 patients.

†Mean values measured in at least 100 patients.

Paired Data

- ▶ Paired data come from non-independent groupings.
- ▶ Two main types:
 - ▷ **Matched Cases Data:** a case-control study that “matches” similar cases together for comparison, the same case measured at different points in time.
 - ▷ **Cross-Over Trial Data:** subjects get of different treatments (or exposures) at different points in time, typically in RCT settings.
- ▶ Key statistical issue: *the variability is less than with independent comparisons.*

Standard Tests For Paired Samples

► Continuous Data:

- ▷ Normally distributed data: paired Students- t test.
- ▷ Otherwise: Wilcoxon Signed Rank (matched pairs) test.

► Ordinal Data:

- ▷ sign test or Wilcoxon Signed Rank (matched pairs) test.

► Nominal Data:

- ▷ Two categories: McNemar's Test.
- ▷ More Than Two Categories: tabular analysis, multinomial models, modified Wilcoxon (book: "consult a statistician").

Paired Students- t Test Setup

- ▶ Group 1: $x_{11}, x_{12}, \dots, x_{1n}$.
- ▶ Group 2: $x_{21}, x_{22}, \dots, x_{2n}$.
- ▶ Pair these: $(x_{11}, x_{21}), (x_{12}, x_{22}), \dots, (x_{1n}, x_{2n})$.
- ▶ Define the paired difference: $d_1 = x_{11} - x_{21}, d_2 = x_{12} - x_{22}, \dots, d_n = x_{1n} - x_{2n}$.
- ▶ $H_0 : \delta = 0$.
- ▶ $H_A : \delta \neq 0$ (most common), $H_A : \delta > 0$ or $H_A : \delta < 0$.
- ▶ Assumptions:
 - ▷ $d_i \sim \mathcal{N}(\delta, \sigma_\delta^2)$,
 - ▷ d_i independent of $d_j, \forall i \neq j$.

Paired Students- t Test Steps

- Obtain: d_1, d_2, \dots, d_n from the data.
- Calculate the mean and standard deviation of these values:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i \quad s_d = \frac{1}{n-1} \sqrt{\sum_{i=1}^n (d_i - \bar{d})^2}.$$

- Use the latter to get the standard error of the mean difference:

$$SE(\bar{d}) = \frac{s_d}{\sqrt{n}}.$$

- The test statistic is distributed Students- t with $df = n - 1$:

$$t = \frac{\bar{d} - 0}{SE(\bar{d})}$$

under the null assumption.

Pairing the Same Case At Two Points In Time

► Two Groups HRQoL:

	Mean	<i>n</i>	SD	SE
Zero Months (baseline)	66.3	36	18.8	3.1
Three Months	58.9	36	22.0	3.7

► Students-*t* Test For \bar{d} :

	Mean	SD	SE	95% CI	<i>t</i>	df	<code>2*(1-pt(2.661,df=35))</code>
Baseline-3 Months	7.3	16.5	2.8	[1.7:12.9]	2.661	35	0.011681

► The confidence interval here is calculated by:

$$[\bar{d} \pm t_{df,\alpha} \times SE(\bar{d})] = [7.3 \pm 2.03 \times 2.8] = [1.7 : 12.9]$$

where:

```
qt(0.025,df=35,lower.tail=FALSE)
[1] 2.0301
```

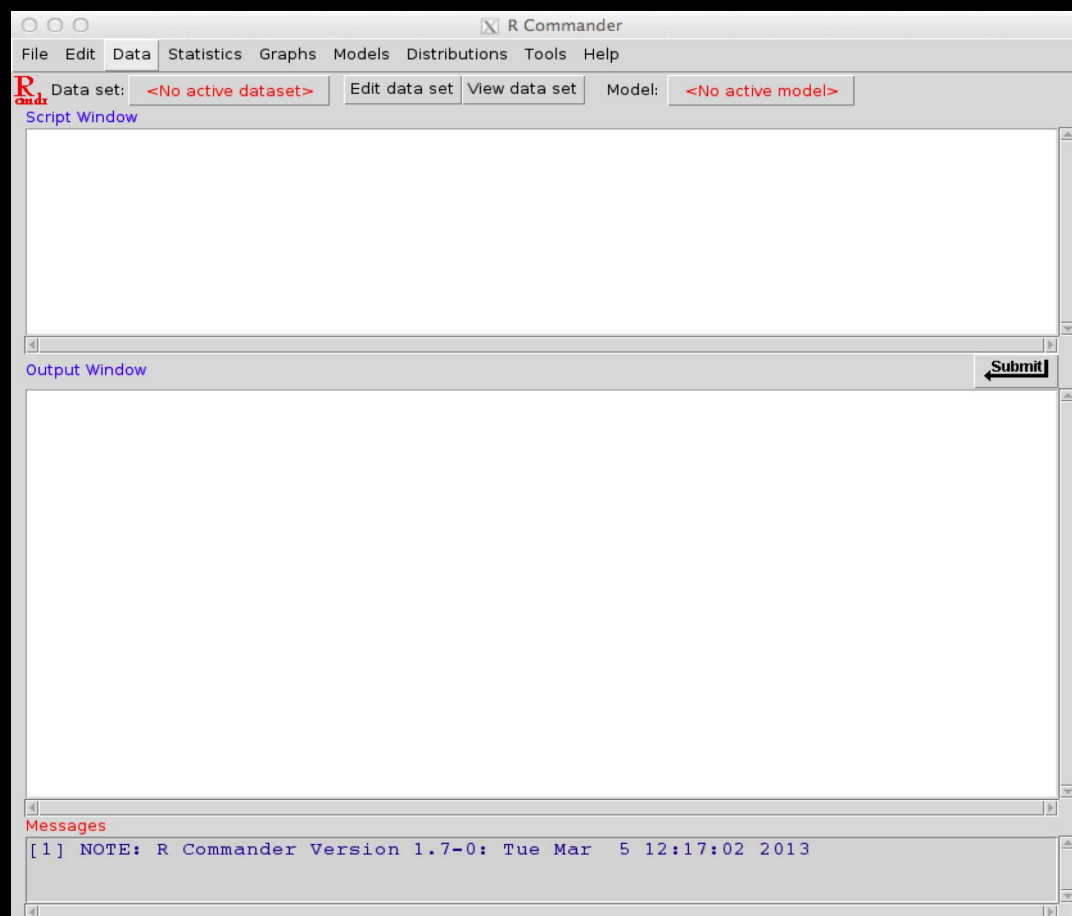
Paired Sample Tests of Center, Example in R

- ▶ Effectiveness of AZT treatment for HIV: 300mg versus 600mg; outcome is p24 antigen level in pg/ml, (a viral protein that is detectable early after infection); only 10 pairs (20 cases).
- ▶ Graphs show rough normality and similarity of variance (page 240).
- ▶ Test:

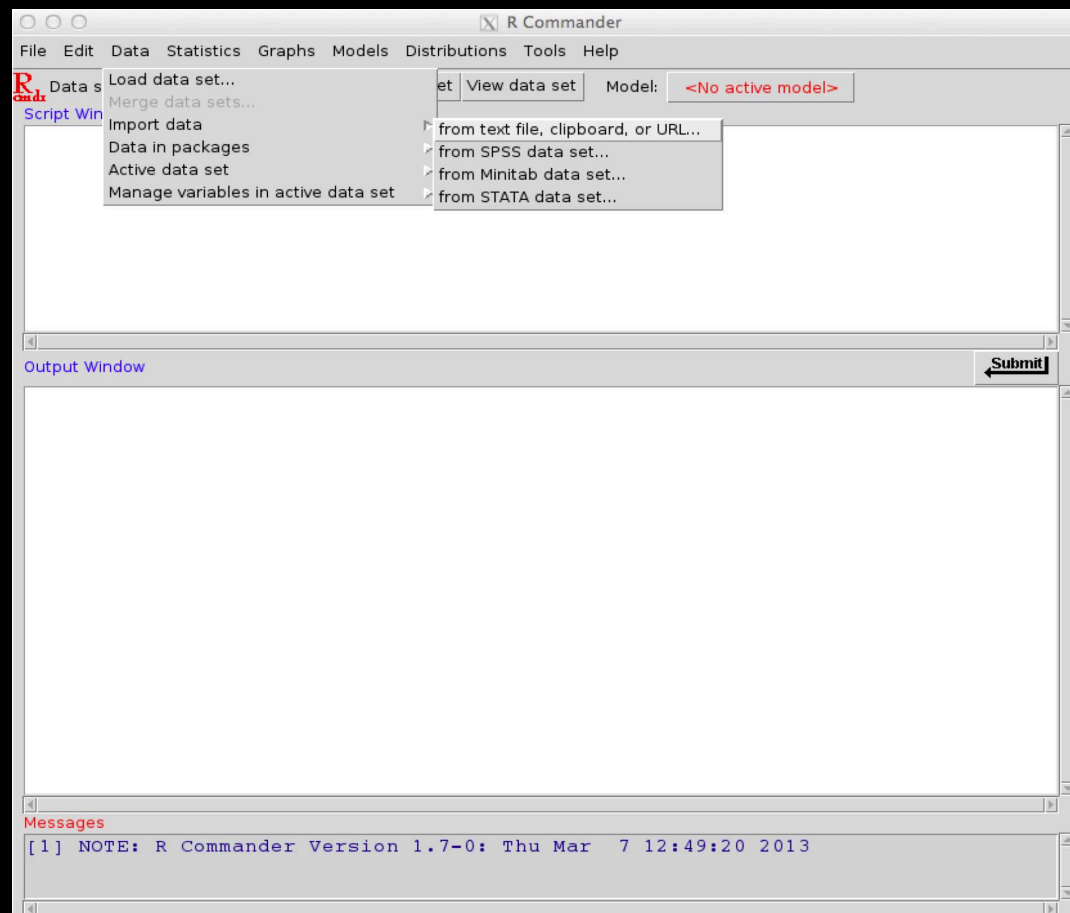
```
x <- c(284, 279, 289, 292, 287, 295, 285, 279, 306, 298)
y <- c(298, 307, 297, 279, 291, 335, 299, 300, 306, 291)
t.test(x, y, var.equal=TRUE, alternative = "two.sided",
       paired = TRUE, conf.level = 0.95)
```

```
t = -2.1437, df = 9, p-value = 0.06066
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.40245    0.60245
sample estimates:
mean of the differences
          -10.9
```

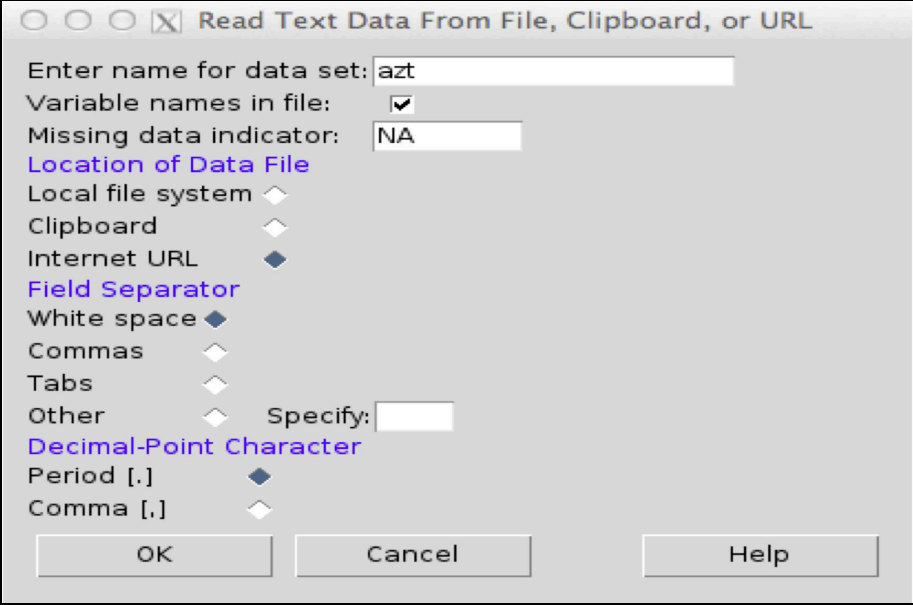
Paired Sample Tests of Center, Example in Rcmdr



Open URL Access In Rcmdr



Using My Data Site



The image shows a dialog box titled "Read Text Data From File, Clipboard, or URL". It contains several input fields and radio buttons for configuring data import. The "Enter name for data set:" field is filled with "azt". The "Variable names in file:" checkbox is checked. The "Missing data indicator:" field is filled with "NA". Under the "Location of Data File" section, the "Internet URL" radio button is selected. Under the "Field Separator" section, the "White space" radio button is selected. Under the "Decimal-Point Character" section, the "Period [.]" radio button is selected. At the bottom are "OK", "Cancel", and "Help" buttons.

Read Text Data From File, Clipboard, or URL

Enter name for data set:

Variable names in file: ☒

Missing data indicator:

Location of Data File

Local file system ☐

Clipboard ☐

Internet URL ☒

Field Separator

White space ☒

Commas ☐

Tabs ☐

Other ☐ Specify:

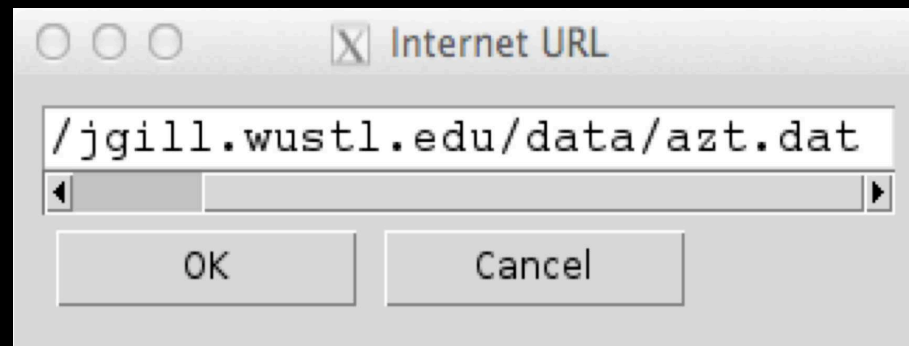
Decimal-Point Character

Period [.] ☒

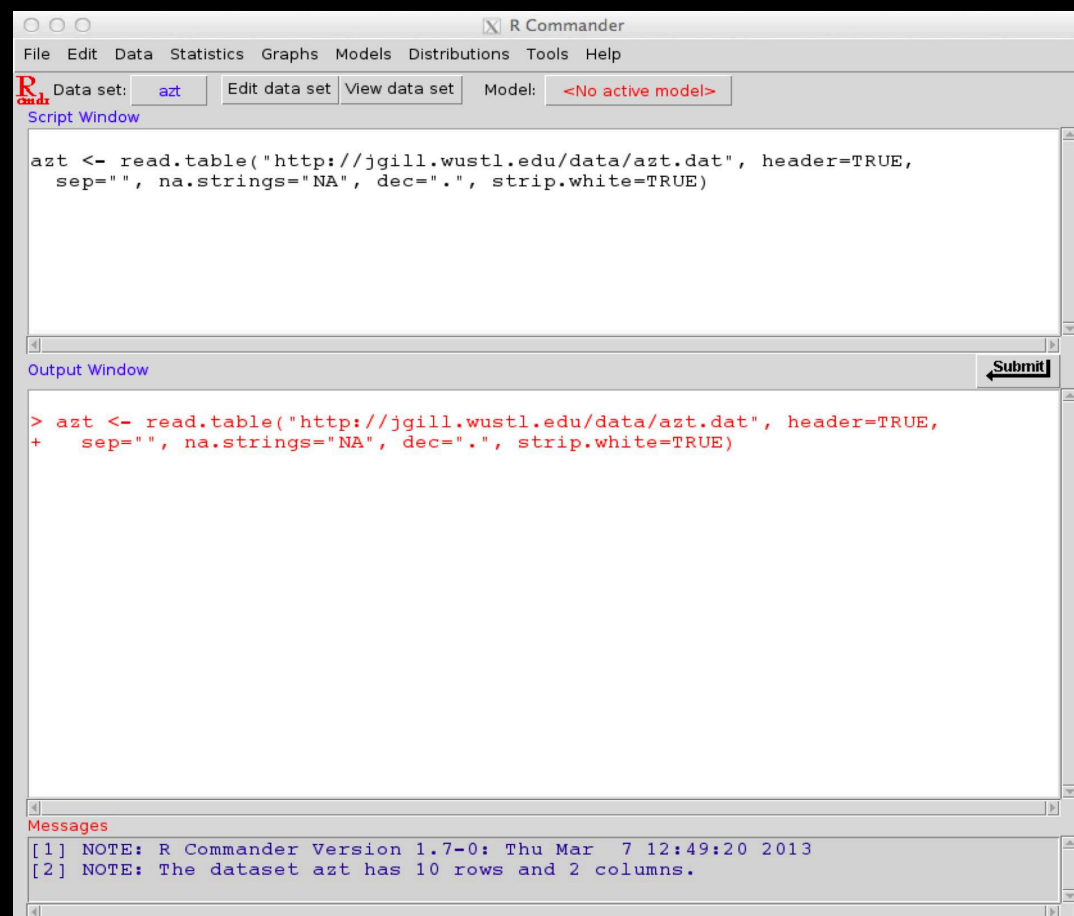
Comma [,] ☐

OK Cancel Help

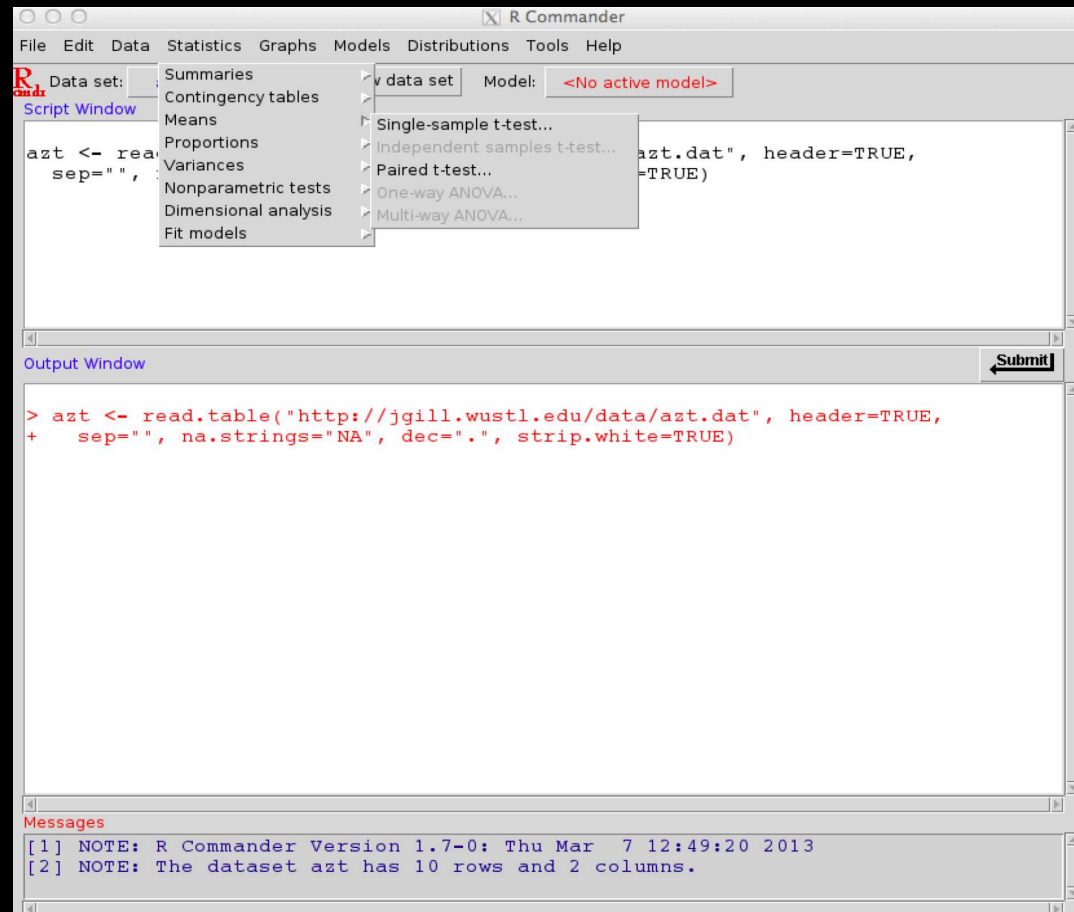
Now We Have the AZT Data



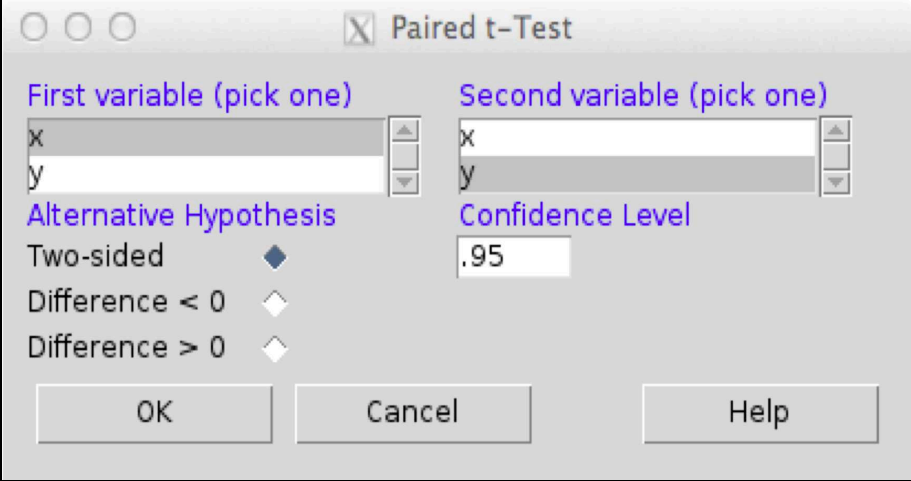
Rcmdr Resulting Window



Statistics → Means → Paired t-test



Rcmdr Window for Students- t Test



A screenshot of the Rcmdr 'Paired t-Test' dialog box. The window has a title bar with standard OS controls and the text 'Paired t-Test'. Inside, there are two columns of options. The first column is headed 'First variable (pick one)' and contains a list box with 'x' and 'y', where 'x' is selected. Below this is the 'Alternative Hypothesis' section with three radio buttons: 'Two-sided' (selected), 'Difference < 0', and 'Difference > 0'. The second column is headed 'Second variable (pick one)' and contains a list box with 'x' and 'y', where 'y' is selected. Below this is a 'Confidence Level' text box containing the value '.95'. At the bottom are three buttons: 'OK', 'Cancel', and 'Help'.

Paired t-Test

First variable (pick one)

x
y

Second variable (pick one)

x
y

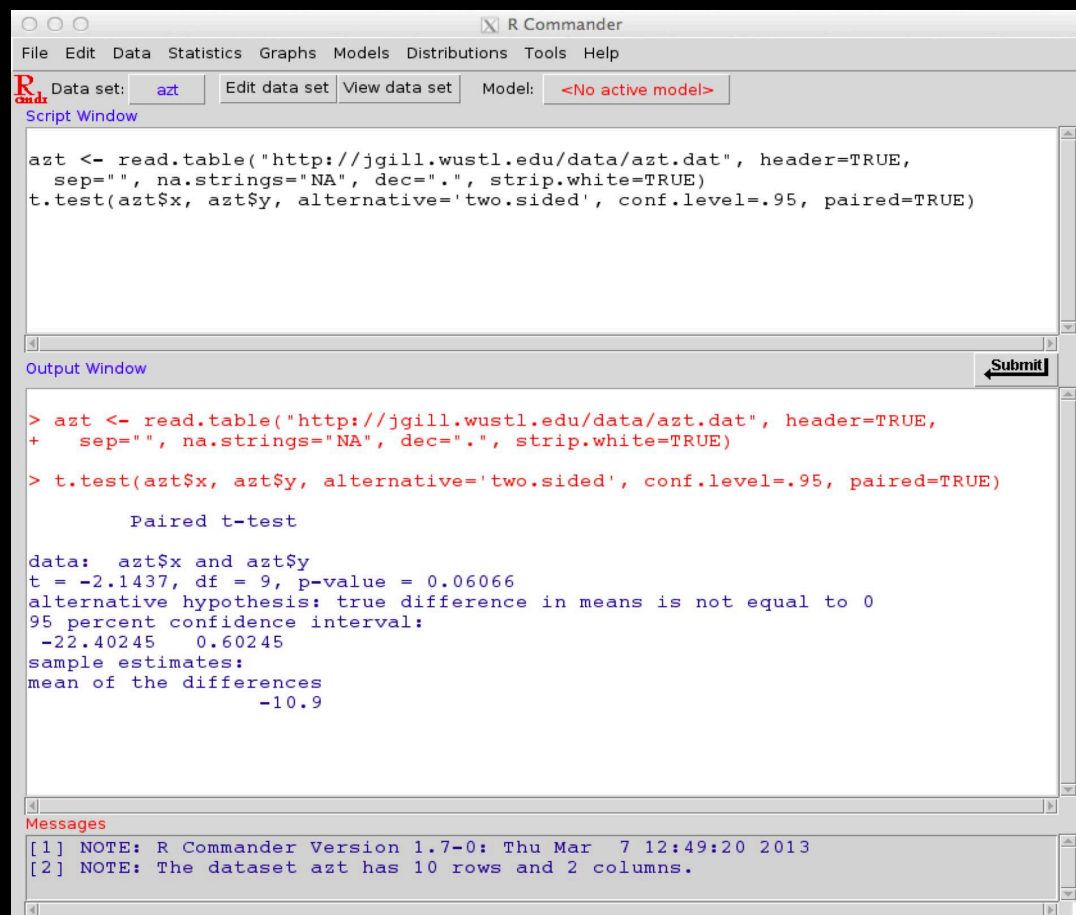
Alternative Hypothesis

Two-sided ☒
Difference < 0 ☐
Difference > 0 ☐

Confidence Level

.95

OK Cancel Help

Students- t Test Results

The screenshot shows the R Commander interface. The top menu bar includes File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, and Help. Below the menu bar, there are buttons for 'Data set: azt', 'Edit data set', 'View data set', and 'Model: <No active model>'. The 'Script Window' contains the following R code:

```
azt <- read.table("http://jgill.wustl.edu/data/azt.dat", header=TRUE,
  sep=" ", na.strings="NA", dec=".", strip.white=TRUE)
t.test(azt$x, azt$y, alternative='two.sided', conf.level=.95, paired=TRUE)
```

The 'Output Window' shows the execution results:

```
> azt <- read.table("http://jgill.wustl.edu/data/azt.dat", header=TRUE,
+   sep=" ", na.strings="NA", dec=".", strip.white=TRUE)
> t.test(azt$x, azt$y, alternative='two.sided', conf.level=.95, paired=TRUE)

    Paired t-test

data:  azt$x and azt$y
t = -2.1437, df = 9, p-value = 0.06066
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.40245   0.60245
sample estimates:
mean of the differences
                -10.9
```

The 'Messages' window at the bottom displays two messages:

```
[1] NOTE: R Commander Version 1.7-0: Thu Mar  7 12:49:20 2013
[2] NOTE: The dataset azt has 10 rows and 2 columns.
```

Wilcoxon Signed Rank Test Setup

- ▶ Consider ordinal data as “ranks.”
- ▶ Ranks are resistant to outliers since an outlying value will only ever be 1 unit away from the following value.
- ▶ Removes information about distributional shape (therefore sometimes called nonparametric).
- ▶ Group 1: $x_{11}, x_{12}, \dots, x_{1n}$.
- ▶ Group 2: $x_{21}, x_{22}, \dots, x_{2n}$.
- ▶ Pair these: $(x_{11}, x_{21}), (x_{12}, x_{22}), \dots, (x_{1n}, x_{2n})$.
- ▶ Define the paired difference: $d_1 = x_{11} - x_{21}, d_2 = x_{12} - x_{22}, \dots, d_n = x_{1n} - x_{2n}$.
- ▶ $H_0 : \text{median}(\delta) = 0$.
- ▶ $H_A : \text{median}(\delta) \neq 0$, (most common), $H_A : \text{median}(\delta) > 0$, or $H_A : \text{median}(\delta) < 0$.
- ▶ Assumptions: $d_i \sim \text{symmetric}$, $\forall i$, d_i independent of $d_j, i \neq j$.

Wilcoxon Signed Rank Test Steps

► From the data obtain:

- ▷ the absolute value of the paired differences: $|d_1|, |d_2|, \dots, |d_n|$ and
- ▷ the sign of the paired differences: $\text{sign}(d_1), \text{sign}(d_2), \dots, \text{sign}(d_n)$.

► Rank the absolute differences, discarding values equal to zero and averaging ties, calling these $R_1, R_2, \dots, R_{n'}$, where the new $n' < n$ from the discards.

► Calculate:

$$T^+ = \left| \sum_{i=1}^{n'} \text{sign}(d_i) \times R_i \right| = \left| \sum_{i=1}^{n'} \text{signed rank of } d_i \right|$$

(also called V and W).

► The test statistic is:

$$z = \frac{T^+ - \frac{n'(n'+1)}{4}}{\sqrt{\frac{n'(n'+1)(2n'+1)}{24}}},$$

which is asymptotically standard normal.

Wilcoxon Signed Rank Test Steps Example

1	2	3	4	5	6	7
Subj.	X_A	X_B	original $X_A - X_B$	absolute $X_A - X_B$	rank of absolute $X_A - X_B$	signed rank
1	78	78	0	0	---	---
2	24	24	0	0	---	---
3	64	62	+2	2	1	+1
4	45	48	-3	3	2	-2
5	64	68	-4	4	3.5	-3.5
6	52	56	-4	4	3.5	-3.5
7	30	25	+5	5	5	+5
8	50	44	+6	6	6	+6
9	64	56	+8	8	7	+7
10	50	40	+10	10	8.5	+8.5
11	78	68	+10	10	8.5	+8.5
12	22	36	-14	14	10	-10
13	84	68	+16	16	11	+11
14	40	20	+20	20	12	+12
15	90	58	+32	32	13	+13
16	72	32	+40	40	14	+14
W = 67.0						
N = 14						

Two-Sample Tests of Center, Example in R

- Consider comparing two groups measuring tumour size (grams) in female Buffalo rats:

```
A <- c(5.8, 1.0, 1.1, 2.1, 2.5, 1.1, 1.0, 1.2, 3.2, 2.7)
B <- c(1.5, 2.7, 6.6, 4.6, 1.1, 1.2, 5.7, 3.2, 1.2, 1.3)
wilcox.test(A,B,paired = TRUE,exact=FALSE)
```

Wilcoxon signed rank test with continuity correction

data: A and B

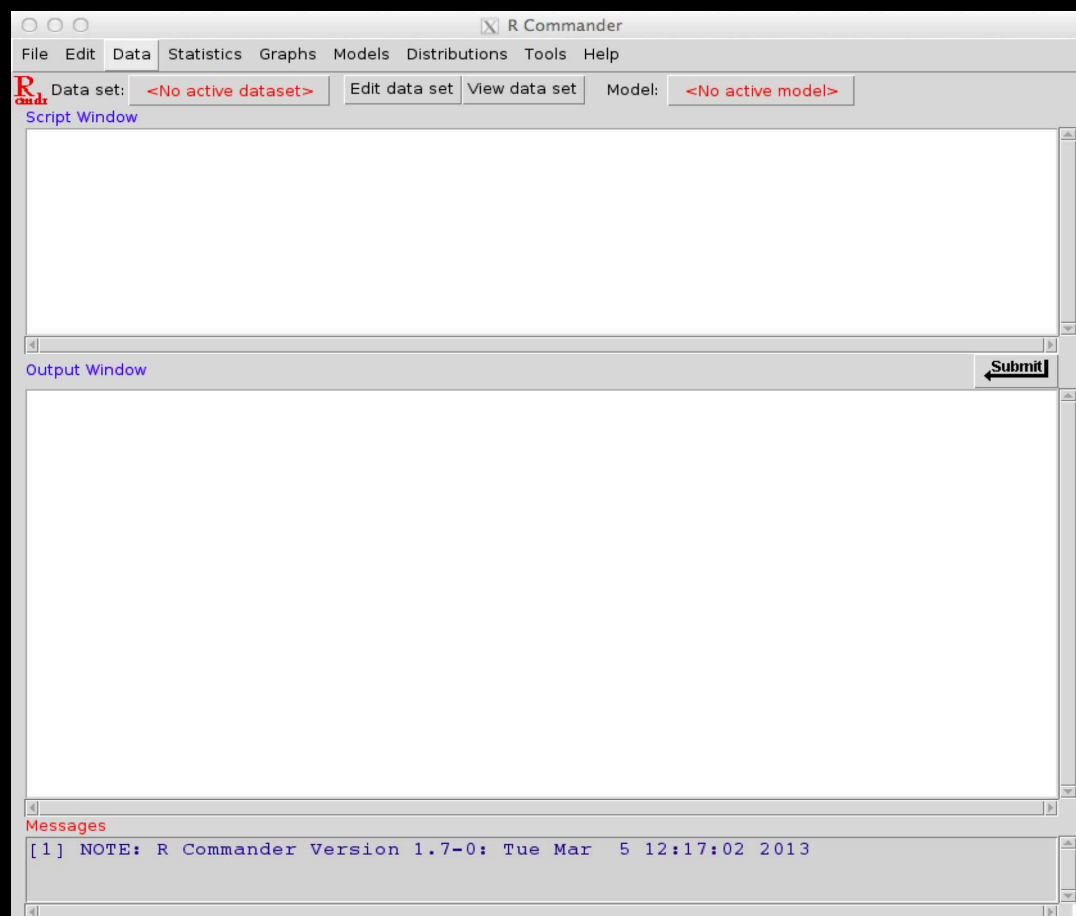
V = 18.5, p-value = 0.386

alternative hypothesis: true location shift is not equal to 0

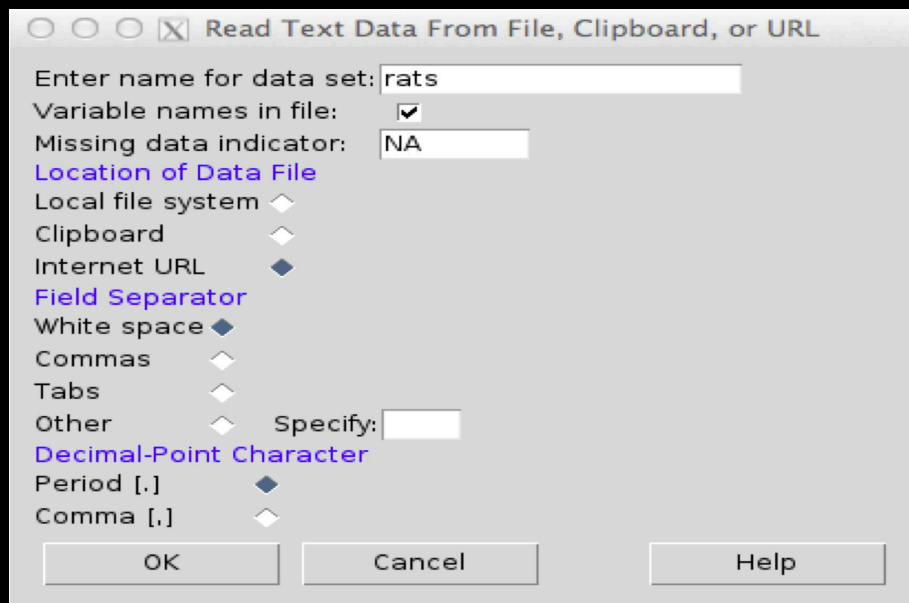
- The “continuity correction” adds 0.5 in the direction of the mean to each value before the sum, to help with small sample size.
- Otherwise stipulate:

```
wilcox.test(A,B,paired = TRUE,exact=FALSE, correct = FALSE)
```

Two-Sample Tests of Center, Example in Rcmdr



Download File in Rcmdr



The screenshot shows a dialog box titled "Read Text Data From File, Clipboard, or URL". It contains several input fields and radio buttons for configuring data import. The "Enter name for data set:" field is filled with "rats". The "Variable names in file:" checkbox is checked. The "Missing data indicator:" field is filled with "NA". Under "Location of Data File", the "Internet URL" radio button is selected. Under "Field Separator", the "White space" radio button is selected. Under "Decimal-Point Character", the "Period [.] " radio button is selected. The "OK", "Cancel", and "Help" buttons are at the bottom.

☐ ☐ ☒ Read Text Data From File, Clipboard, or URL

Enter name for data set: rats

Variable names in file: ☒

Missing data indicator: NA

Location of Data File

Local file system ☐

Clipboard ☐

Internet URL ☒

Field Separator

White space ☒

Commas ☐

Tabs ☐

Other ☐ Specify:

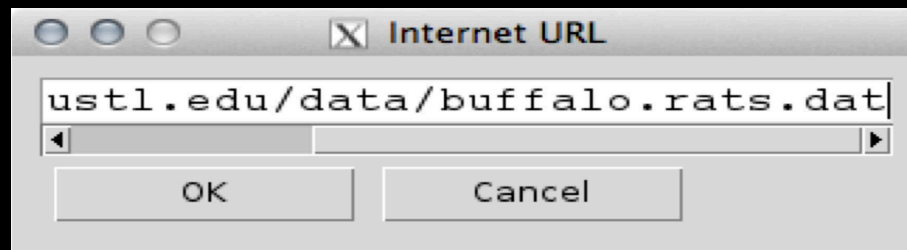
Decimal-Point Character

Period [.] ☒

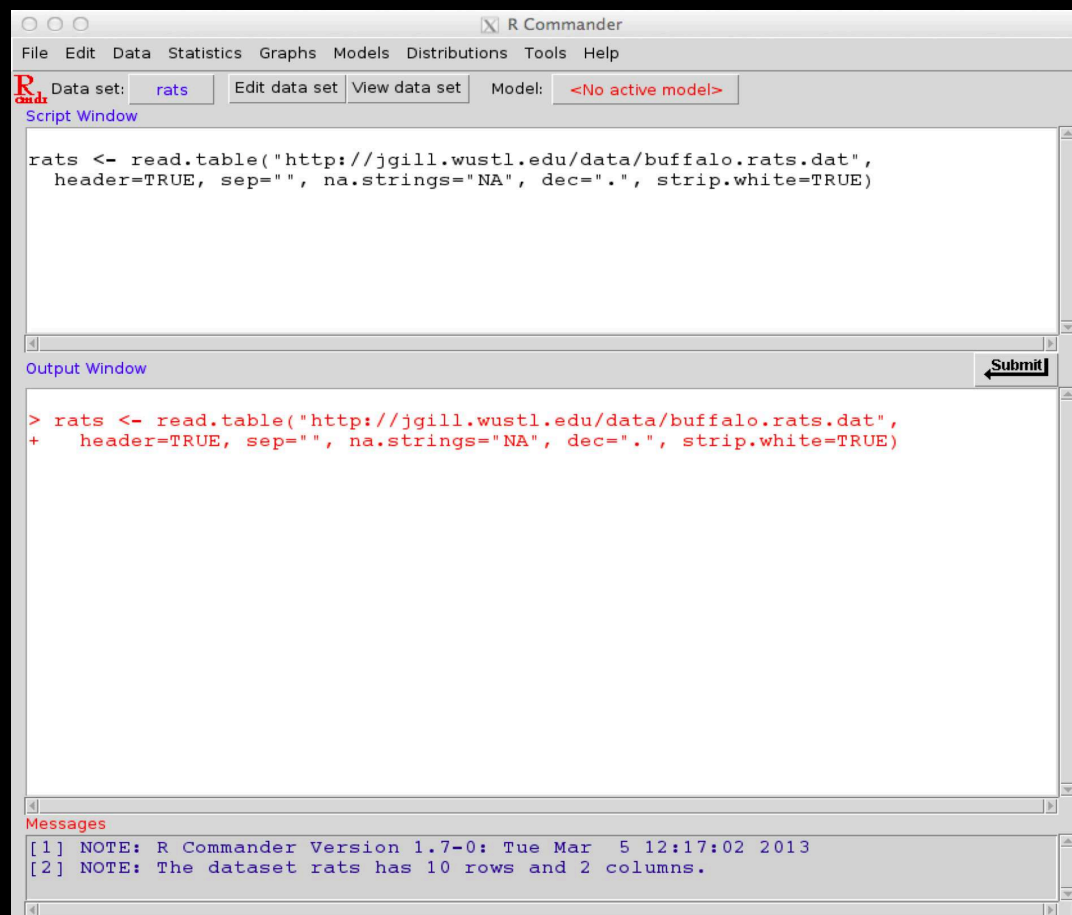
Comma [,] ☐

OK Cancel Help

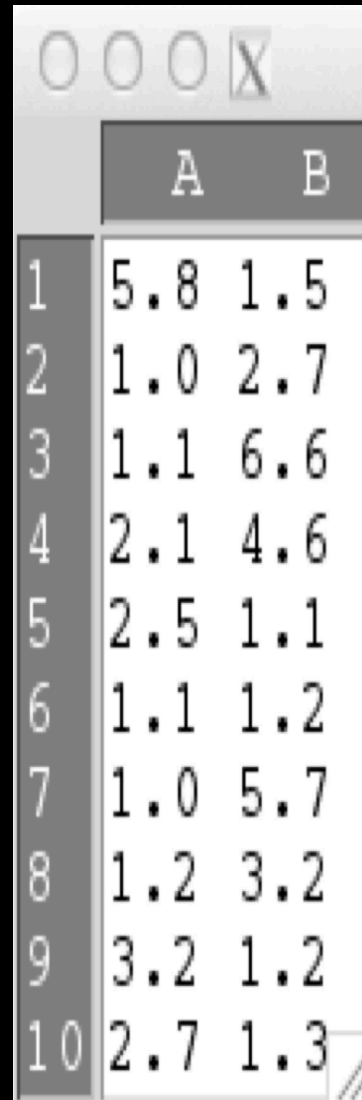
Using My Data Site



Now We Have Our Buffalo Rats

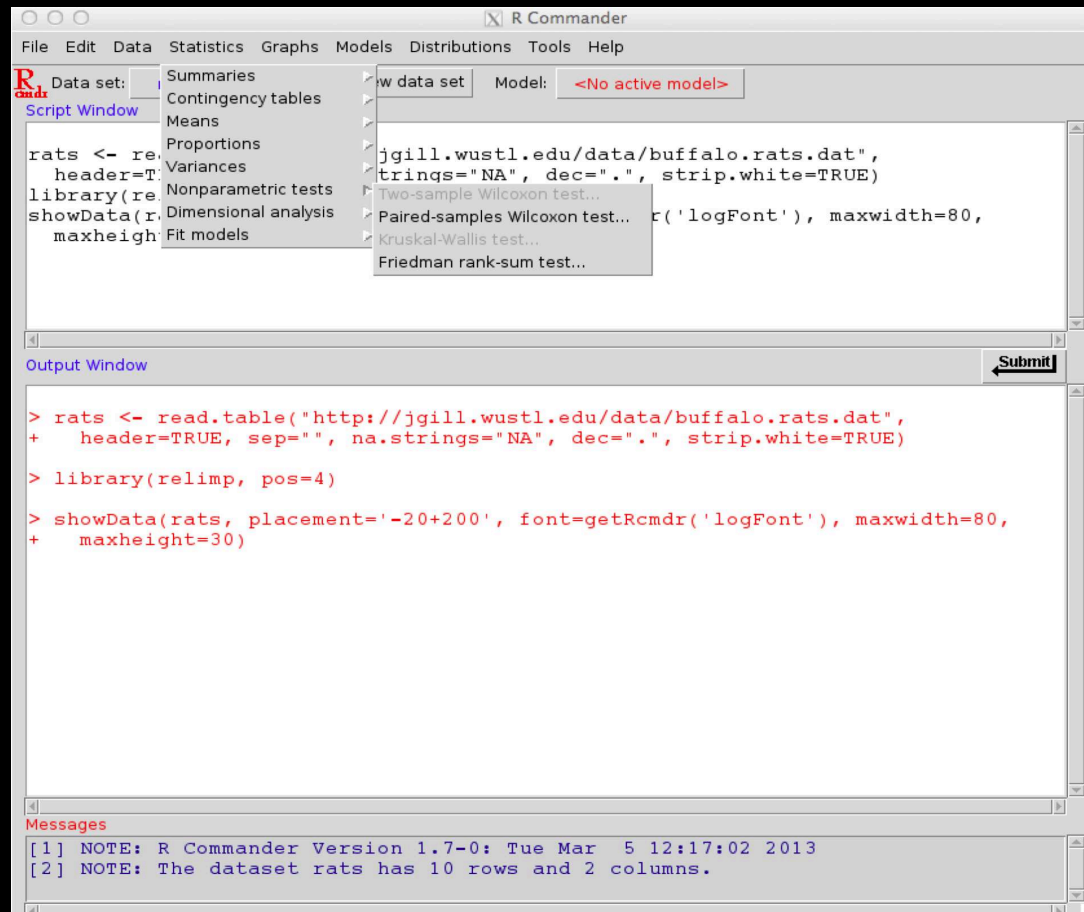


Clicking “View data set”

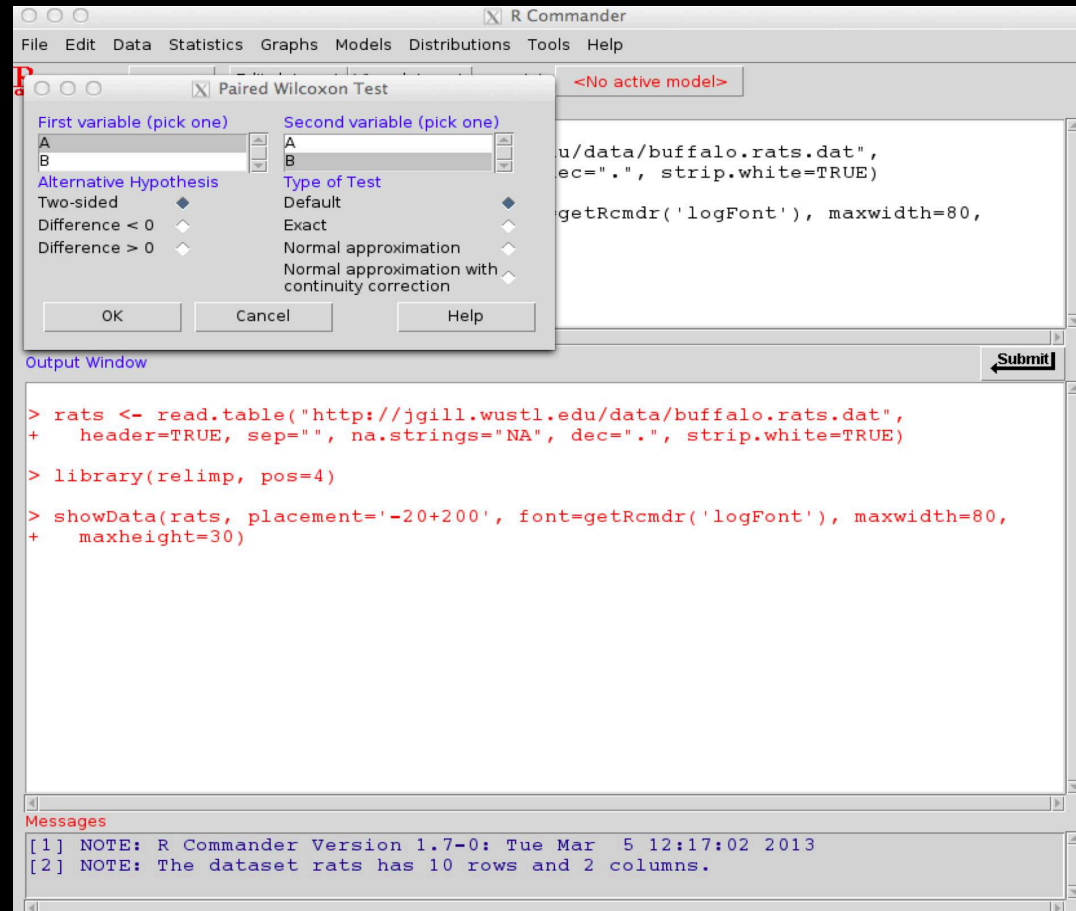


	A	B
1	5.8	1.5
2	1.0	2.7
3	1.1	6.6
4	2.1	4.6
5	2.5	1.1
6	1.1	1.2
7	1.0	5.7
8	1.2	3.2
9	3.2	1.2
10	2.7	1.3

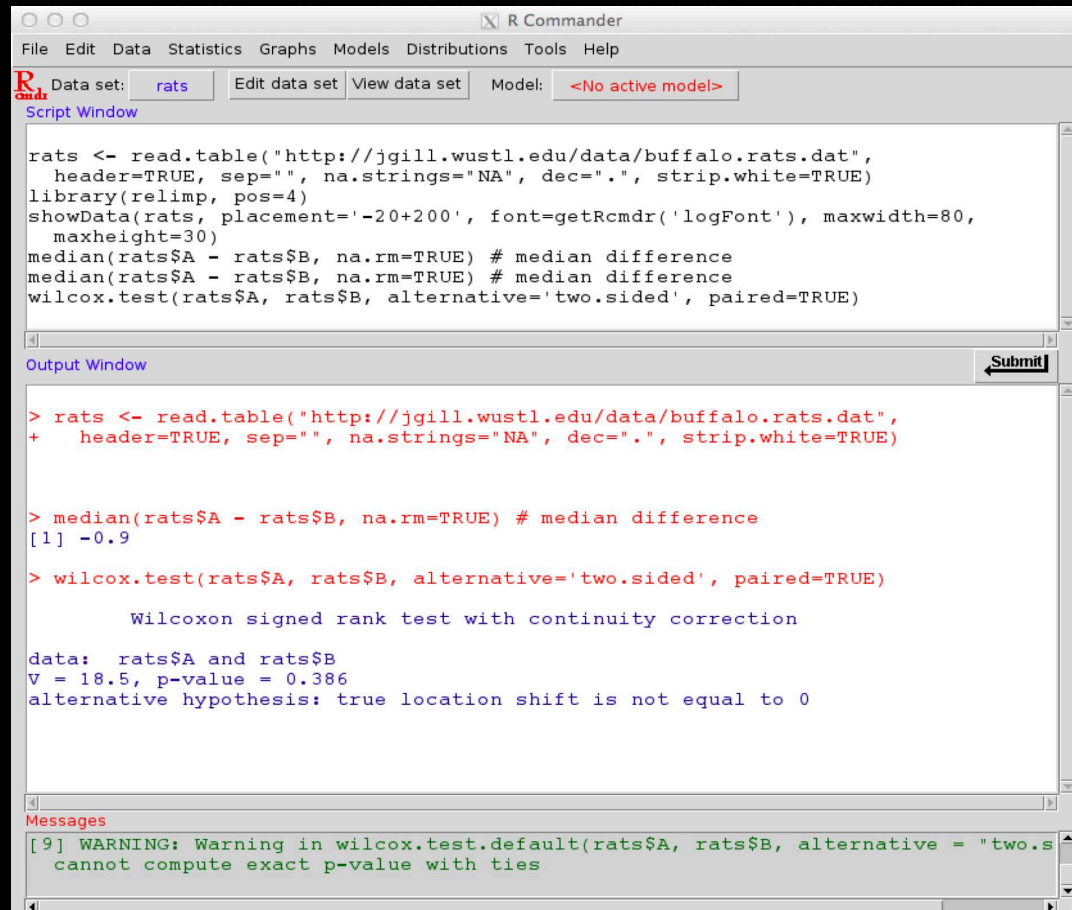
Statistics → Nonparametric tests → Paired-samples Wilcoxon test



Select Two Variables, Rcmdr forces “Default” since there are ties



Rcmdr Only Let Me Run As Default (warning appears about ties)



The screenshot shows the R Commander window with the following components:

- Menu Bar:** File, Edit, Data, Statistics, Graphs, Models, Distributions, Tools, Help.
- Buttons:** Data set: **rats**, Edit data set, View data set, Model: **<No active model>**.
- Script Window:**

```
rats <- read.table("http://jgill.wustl.edu/data/buffalo.rats.dat",
  header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)
library(relimp, pos=4)
showData(rats, placement='-20+200', font=getRcmdr('logFont'), maxwidth=80,
  maxheight=30)
median(rats$A - rats$B, na.rm=TRUE) # median difference
median(rats$A - rats$B, na.rm=TRUE) # median difference
wilcox.test(rats$A, rats$B, alternative='two.sided', paired=TRUE)
```
- Output Window:** (Contains a **Submit** button)

```
> rats <- read.table("http://jgill.wustl.edu/data/buffalo.rats.dat",
+   header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)

> median(rats$A - rats$B, na.rm=TRUE) # median difference
[1] -0.9

> wilcox.test(rats$A, rats$B, alternative='two.sided', paired=TRUE)

      Wilcoxon signed rank test with continuity correction

data:  rats$A and rats$B
V = 18.5, p-value = 0.386
alternative hypothesis: true location shift is not equal to 0
```
- Messages:**

```
[9] WARNING: Warning in wilcox.test.default(rats$A, rats$B, alternative = "two.s
cannot compute exact p-value with ties
```

Comparing Two Paired Groups, Nominal Outcome

- ▶ Brown *et al.* (1987) performed a **matched case-control study** on *all* cases of testicular cancer of young men in the DC area from 1/1/76 to 6/30/86 (page 137 in the book).
- ▶ Control: men in the same hospital as the cases diagnosed with malignancy other than cancer of the genital tract, similar in age (± 2 years), race, vital status, and year of diagnosis.
- ▶ Key explanatory variable of interest: undescended testes at birth.
- ▶ The analysis focuses on the discordant pairs: no undescended testes and cancer, undescended testes and no cancer:

		<u>Undescended</u>		Total
		Yes	No	
Testicular Cancer	Yes	4	11	15
	No	3	241	244
	Total	7	252	259

- ▶ The book takes the ***b*** and ***c*** values from the 2×2 table and relabels them as ***f*** and ***g***, “discordant pairs.”
- ▶ The null hypothesis is that the discordant pairs are equal, $H_0 : f = g$.

Comparing Two Paired Groups, Nominal Outcome

- The χ^2 McNemar's test statistic is given by:

$$\chi_{\text{McNemar}}^2 = \frac{(f - g)^2}{f + g} = \frac{(11 - 3)^2}{11 + 3} = 4.5714,$$

which has $df = 1$ since it comes from a 2×2 table (number of rows minus one times number of columns minus one).

- So from `pchisq(4.5714,df=1,lower.tail=FALSE)` we get 0.03251.

- A Yate's-style correction for small samples uses: $\chi_{\text{McNemar}}^2 = \frac{(|f - g| - 1)^2}{f + g},$

but this is not necessary here.

- Directly in R:

```
x <- matrix(c(4,3,11,241),2,2)
mcnemar.test(x,correct=FALSE)
```

McNemar's Chi-squared test

```
data:  x
```

```
McNemar's chi-squared = 4.5714, df = 1, p-value = 0.03251
```

Paired Samples With Nominal Outcome

- ▶ Now let's ask about the *exact* probability of getting this table or one more “extreme.”
- ▶ This is a form of Fisher's Exact Test, which we will see in more detail with independent data.
- ▶ Returning to the Brown *et al.* data, we start with the table:

$$\begin{array}{r|l} \text{(i)} & \\ \hline 4 & 11 \\ 3 & 241 \end{array}$$

- ▶ We then identify 3 more tables that make the off-diagonal values more discordant:

(ii)	(iii)	(iv)
$\begin{array}{r l} 4 & 12 \\ 2 & 241 \end{array}$	$\begin{array}{r l} 4 & 13 \\ 1 & 241 \end{array}$	$\begin{array}{r l} 4 & 14 \\ 0 & 241 \end{array}$

- ▶ For the original table and the three more discordant tables, calculate:

$$p = \frac{(f+g)!}{f!g!} \left(\frac{1}{2}\right)^{f+g}.$$

Paired Samples With Nominal Outcome

► For the 4 tables, the probabilities are:

$$p(i) = \frac{14!}{11!3!} \left(\frac{1}{2}\right)^{11+3} = 0.022217$$

$$p(i) = \frac{14!}{12!2!} \left(\frac{1}{2}\right)^{11+3} = 0.005554$$

$$p(i) = \frac{14!}{13!1!} \left(\frac{1}{2}\right)^{11+3} = 0.000854$$

$$p(i) = \frac{14!}{14!0!} \left(\frac{1}{2}\right)^{11+3} = 0.000061$$

► The total is 0.028686, which is 0.057372 when doubled to put all of the rejection region in one tail (the book rounds this to 0.06).

Standard Tests For Independent (Unpaired) Samples

► Continuous Data:

- ▷ Normally distributed data, modest samples: independent samples Students- t test.
- ▷ Non-normally distributed data, modest samples: Wilcoxon (as before) or Mann-Whitney U test (generally preferred).

► Ordinal Data:

- ▷ Mann-Whitney U test for difference.
- ▷ Chi-Squared test for trend.

► Nominal Data:

- ▷ Two Categories: proportions Chi-Square test, add Yates correction if all expected frequencies are not greater than five.
- ▷ More Than Two Categories: Chi-Square cross-tabs test, collapse categories if all expected frequencies are not greater than five.

Independent Samples With Continuous Outcome, Students- t Test Setup

- ▶ Group 1: $x_{11}, x_{12}, \dots, x_{1n_1}$.
- ▶ Group 2: $x_{21}, x_{22}, \dots, x_{2n_2}$.
- ▶ $H_0 : \delta = 0$.
- ▶ $H_A : \delta \neq 0$ (most common), $H_A : \delta > 0$ or $H_A : \delta < 0$.
- ▶ Assumptions:
 - ▷ variables are continuous and independent,
 - ▷ the group standard deviations are similar (one no more than about twice the other),
 - ▷ the **population** data for each group is distributed *normal*.
- ▶ Restriction: sample sizes are modest.
- ▶ We are now switching from a normal distribution to a students- t distribution for the test statistic.

Independent Samples With Continuous Outcome, Students- t Test Steps

► From the two samples get $\bar{x}_1, \bar{x}_2, s_1, s_2$, based on the sample sizes n_1 and n_2 .

► Calculate the mean difference:

$$\bar{d} = \bar{x}_1 - \bar{x}_2.$$

► Calculate the *pooled* standard deviation:

$$SD_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

► Use this to get the standard error of the mean difference:

$$SE(\bar{d}) = SD_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

► The test statistic is distributed Students- t with $df = n_1 + n_2 - 2$ under the null:

$$t = \frac{\bar{d} - 0}{SE(\bar{d})}$$

Independent Samples With Continuous Outcome, Students- t Test Example

► Leg Ulcer-Free Time in Weeks:

	Mean	n	SD	SE
Intervention Group	20.1	120	18.5	1.7
Control Group	14.2	113	17.6	1.7

► Students- t Test:

	t	df	$2*(1-pt(2.485, df=35))$	Mean Diff	SE Diff	95% CI
Difference in Weeks	2.485	231		5.9	2.4	[1.2:10.5]

► The confidence interval here is calculated by:

$$[\bar{d} \pm t_{df,\alpha} \times SE(\bar{d})] = [5.9 \pm 1.97 \times 2.4] = [1.2 : 10.5]$$

using:

```
> qt(0.025,df=231,lower.tail=FALSE)
[1] 1.9703
```

Independent Samples With Continuous Outcome, No Normal Data Assumption

- ▶ For this case we prefer the **Mann-Whitney Test** (also confusingly called a two-sample Wilcoxon test).
- ▶ Start: combine the two groups and rank the full sample in decreasing order (again tied values get the mean of the two adjacent ranks), to produce a statistic **W** (to be described).
- ▶ The test statistic is distributed standard normal:

$$z = \frac{W - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}}.$$

- ▶ Some software uses slight variations on this with better small sample properties.
- ▶ $H_0 : \delta = 0$.
- ▶ $H_A : \delta \neq 0$ (most common), $H_A : \delta > 0$ or $H_A : \delta < 0$.
- ▶ Assumptions: groups are independent, cases can be ordered.

Independent Samples With Continuous Outcome, No Normal Data Assumption

► Steps for a small sample Mann-Whitney test:

- ▷ select the group with smaller total rank (for convenience),
- ▷ order both groups together,
- ▷ for each observation in this group count the number of times it is lower than the other group,
- ▷ W is the sum.

► Buffalo Rats Example:

A:	5.8	1.0	1.1	2.1	2.5	1.1	1.0	1.2	3.2	2.7
B:	1.5	2.7	6.6	4.6	1.1	1.2	5.7	3.2	1.2	1.3

Independent Samples With Continuous Outcome, No Normal Data Assumption

```

apply(rats,2,sum)
  A      B
21.7 29.1
( o.rats <- order(rats,decreasing=TRUE) )
[1] 13  1 17 14  9 18 10 12  5  4 11 20  8 16 19  3  6 15  2  7
sum.rats <- rep(0,20)
  for (i in 1:length(o.rats)) {
    if (o.rats[i] < 11) {
      rats.greater <- 0
      for (j in 1:i) {
        if ((o.rats[j] > o.rats[i]) & (o.rats[j] > 10))
          rats.greater <- rats.greater + 1
      }
      sum.rats[i] <- rats.greater
    }
  }
sum.rats
[1]  0  1  0  0  3  0  4  0  5  5  0  0  7  0  0  9  9  0 10 10
sum(sum.rats)
[1] 63

```

Independent Samples With Continuous Outcome, No Normal Data Assumption

► Steps for a *large sample* Mann-Whitney test:

▷ for group 1 calculate the sum of the ranks out of the total (combined groups), R_1 , and calculate:

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

where n_1 is the size of group 1.

▷ for group 2 calculate the sum of the ranks out of the total (combined groups), R_2 , and calculate:

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

where n_2 is the size of group 2.

▷ for the test use $U = \min(U_1, U_2)$ in:

$$z = \frac{U - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}.$$

► There are many other versions used in the literature (see Agresti [2012], *Categorical Data Analysis*).

Independent Samples With Continuous Outcome, No Normal Data Assumption

► Leg Ulcer-Free Time in Weeks:

	n	Mean Rank	Sum of Ranks
Intervention Group	120	126.87	15224.0
Control Group	113	106.52	12037.0
Total	233		

► U and W Test:

	test statistic	z	p-value
Mann-Whitney Test	5596.0	-16.42	0
Wilcoxon Test	12037.0	25.53	0

```

n1 <- 120; n2 <- 113
m <- (n1*(n1+n2+1)/2); s <- sqrt(n1*n2*(n1+n2+1)/12)
( z <- (5596 - m)/s )    # [1] -16.42106
2*(pnorm(z))           # [1] 9.810308e-05
m <- n2*(n2+1)/4; s <- sqrt(n2*(n2+1)*(2*n2+1)/24)
( z <- (12037 - m)/s )   # [1] 25.53199
2*(1-pnorm(25.53))     # [1] 0

```


Independent Samples With Nominal Outcome

► Approach 1: Difference In Proportions:

▷ $H_0 : \pi_{\text{Intervention}} - \pi_{\text{Control}} = 0$ versus $H_A : \pi_{\text{Intervention}} - \pi_{\text{Control}} \neq 0$.

▷ The test assumes that there is a common underlying population proportion, π , estimated by:

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

with standard error estimated by:

$$SE(p_1 - p_2) = \sqrt{p(1 - p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}.$$

▷ The test statistic is therefore:

$$z = \frac{p_1 - p_2}{SE(p_1 - p_2)}$$

where tail values imply a difference in the two proportions.

Independent Samples With Nominal Outcome

► Example of *leg ulcers healed at 3 months*:

$$p_1 = \frac{\text{number healed under intervention}}{\text{number in intervention}} = \frac{22}{120} = 0.183$$

$$p_2 = \frac{\text{number healed under control}}{\text{number in control}} = \frac{17}{113} = 0.150$$

$$p = \frac{(120 \cdot 0.183) + (113 \cdot 0.150)}{(120 + 113)} = 0.167$$

$$SE(p_1 - p_2) = \sqrt{0.167(1 - 0.167) \left(\frac{1}{120} + \frac{1}{113} \right)} = 0.049$$

$$z = \frac{0.183 - 0.150}{0.049} = 0.673$$

► Using R:

```
2*(1-pnorm(0.673))  
[1] 0.5009473}
```

Independent Samples With Nominal Outcome

- A **confidence interval** without the (restrictive) assumption of a single underlying population means that we can use the smaller:

$$SE(p_1 - p_2) = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}} = \sqrt{\frac{0.183 \cdot 0.817}{120} + \frac{0.150 \cdot 0.85}{113}} = 0.0487$$

- The 95% CI for the difference of proportions is now:

$$(p_1 - p_2) \pm [z_\alpha \times SE(p_1 - p_2)] = (0.183 - 0.150) \pm [1.95 \times 0.0487] = [-0.062452 : 0.128452].$$

Independent Samples With Nominal Outcome

► Approach 2: $r \times c$ Contingency Tables:

- ▷ H_0 : no association/relationship between row and column variables versus H_A : there is a relationship between levels of row variables and levels of column variables.
- ▷ The test assumes that (1) two independent unordered variables, (2) at least 80% of expected cell counts have more than five values, and (3) no zero expected cell counts.
- ▷ Expected cell counts come from the product of marginals divided by N :

$$E_{ij} = \frac{\text{Row Total}(R_i) \times \text{Column Total}(C_j)}{N}.$$

- ▷ Calculate the difference between observed and expected for each cell: $(O_{ij} - E_{ij})$.
- ▷ Square these differences and divide by the expectation: $(O_{ij} - E_{ij})^2 / E_{ij}$.
- ▷ The test statistic is:

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

which has degrees of freedom equal to $df = (r - 1)(c - 1)$.

Independent Samples With Nominal Outcome

- ▶ With small samples the distributional assumption is suspect, and **Yate's Correction** is designed to improve this: $\chi_Y^2 = \sum_{ij} \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}$.
- ▶ As a reminder, the ulcer table is given by:

	Healed	Not Healed	Total
Intervention	$a = 22$	$b = 98$	$m = 120$
Control	$c = 17$	$d = 96$	$n = 113$
Total	$r = 39$	$s = 194$	$N = 233$

Independent Samples With Nominal Outcome

► Table of *leg ulcers healed at 3 months*:

	O	E	$O - E$	$(O - E)^2 / E$
Intervention/healed	22	20.086	1.914	0.1824
Intervention/not healed	98	99.914	-1.914	0.0367
Control/healed	17	18.914	-1.914	0.1937
Control/not healed	96	94.086	-1.914	0.0389
Total	233	233	0	0.4518

► Low-tech R code:

```

exd <- obs <- c(22,17,98,96)
exd[1] <- (obs[1]+obs[2])*(obs[1]+obs[3])/sum(obs)
exd[2] <- (obs[1]+obs[2])*(obs[2]+obs[4])/sum(obs)
exd[3] <- (obs[3]+obs[4])*(obs[1]+obs[3])/sum(obs)
exd[4] <- (obs[3]+obs[4])*(obs[2]+obs[4])/sum(obs)
diff <- obs - exd
chi <- diff^2/exd
( chi.sq <- sum(chi) )
[1] 0.45175

```

Independent Samples With Nominal Outcome, Fisher's Exact Test

- Approach 3: **Fisher's Exact Test**, used for (mostly 2×2) tables with small samples where Yate's correction is not sufficient.
- Consider the standard table setup:

		<i>Factor A</i>		
		Present	Absent	Total
<i>Factor B</i>	Present	<i>a</i>	<i>b</i>	<i>m</i>
	Absent	<i>c</i>	<i>d</i>	<i>n</i>
Total		<i>r</i>	<i>s</i>	<i>N</i>

- The hypergeometric analysis asks, given all marginals fixed, what is the *exact* probability of observing these cell values:

$$p(\text{table}) = \frac{m!n!r!s!}{N!a!b!c!d!}$$

- Next we do this calculation for *every* table with the same marginal values but giving the same or higher levels of association, add the probabilities, and double this for a two-sided test.

Independent Samples With Nominal Outcome, Fisher's Exact Test

- ▶ Consider data on deaths 6 months after fractured neck of femur in a specialized orthopaedic ward (Factor A) and a general ward (Factor B).
- ▶ So: $a = 2, b = 2, c = 18, d = 14, m = 8, n = 8, r = 20, s = 20, N = 40$ (page 142).



Independent Samples With Nominal Outcome, Fisher's Exact Test

► This gives the table:

	Orthopaedic Ward	General Ward	Total
Present	$a = 2$	$b = 6$	$m = 8$
Absent	$c = 18$	$d = 14$	$n = 32$
Total	$r = 20$	$s = 20$	$N = 40$

► Applying the formula:

$$p(\text{table}) = \frac{m!n!r!s!}{N!a!b!c!d!} = \frac{8!32!20!20!}{40!2!6!18!14!} = 0.0957601$$

calculated in R by:

```
(gamma(9)*gamma(33)*gamma(21)*gamma(21))/
  (gamma(41)*gamma(3)*gamma(7)*gamma(19)*gamma(15))
[1] 0.0957601
```

Independent Samples With Nominal Outcome, Fisher's Exact Test

- There are two more tables (page 142) with the same marginal values that give stronger associations, and these have $p(\text{table})$ values of 0.020160 and 0.001638:

2	6	1	7	0	8
18	14	19	13	20	12

with odds ratios of $bc/ad = 10.231$, and $bc/ad = \infty$.

- For some reason the book sets up this table backwards, hence bc/ad instead of ad/bc .
- So the summed probability is: $0.0957601 + 0.020160 + 0.001638 = 0.1175581$.
- The probability of interest is double this for a two-sided test: $2 \times 0.1175581 = 0.2351162$.

Independent Samples With Nominal Outcome, Fisher's Exact Test

- Using R based on a tabular approach:

```
ortho.df <- data.frame(c(2,18,6,14),expand.grid(1:2,1:2))
names(ortho.df) <- c("counts","death","ward")
( ortho.x <- xtabs(counts ~ death + ward, data=ortho.df) )
```

```
      ward
death 1  2
      1  2  6
      2 18 14
```

```
fisher.test(ortho.x)
```

```
p-value = 0.2351
```

```
alternative hypothesis: true odds ratio is not equal to 1
```

```
95 percent confidence interval for the odds ratio:
```

```
0.02308176 1.80076397
```

```
sample estimates:
```

```
odds ratio
```

```
0.2680041
```

Independent Samples With Nominal Outcome, Fisher's Exact Test

- Alternatively, using R with a matrix approach:

```
( ortho.mat <- matrix(c(2,18,6,14), nrow=2,  
                      dimnames = list(c("Death", "No Death"), c("Ward A", "Ward B"))) )  
      Ward A Ward B  
Death      2      6  
No Death   18     14  
  
fisher.test(ortho.x)  
  
p-value = 0.2351  
alternative hypothesis: true odds ratio is not equal to 1  
95 percent confidence interval for the odds ratio:  
 0.02308176 1.80076397  
sample estimates:  
odds ratio  
 0.2680041
```

Crosstabs For a Trend in $2 \times c$ Tables

- Suppose we have a nominal/dichotomous variable crossed with an ordinal variable (such as pain index or cancer stage).
- The regular chi-square test is not efficient since it does not account for the ordering of the second variable.
- The book gives the following contrived example on page 143:

	Outcome of Trial					Total
	Worse	Same	Slightly better	Moderately better	Much better	
Original Score (x_i)	-2	-1	0	1	2	
Treatment A (a_i)	11	53	42	27	11	144
Treatment B	1	13	16	15	7	52
Total (n_i)	12	66	58	42	18	196 (N)
$p_i = a_i/n_i$	0.9167	0.8030	0.7241	0.6429	0.6111	0.7347 (\bar{q})
$q_i = 1 - a_i/n_i$	0.0833	0.1970	0.2759	0.3571	0.3889	0.2653 (\bar{p})

Crosstabs for a Trend in $2 \times c$ Tables

- First calculate the treatment score variability:

$$T_{xp} = \sum_{i=1}^c n_i(p_i - \bar{p})(x_i - \bar{x}) = \sum_{i=1}^c a_i x_i - \frac{1}{N} \left(\sum_{i=1}^c a_i \sum_{i=1}^c n_i x_i \right).$$

where c is the number of columns.

- Then calculate the score variability:

$$T_{xx} = \sum_{i=1}^c n_i x_i^2 - \frac{1}{N} \left(\sum_{i=1}^c (n_i x_i)^2 \right).$$

- This gives the test statistic for the trend:

$$X_{\text{trend}}^2 = \frac{T_{xp}^2}{T_{xx}} \times \frac{1}{\bar{p}} \times \frac{1}{\bar{q}},$$

which has $df = 1$.

Crosstabs for a Trend in $2 \times c$ Tables

```
a <- c(11, 53, 42, 27, 11)
n <- c(12, 66, 58, 42, 18)
x <- -2:2
p <- a/n; q <- 1-p
Txp <- sum( n*(p-mean(p))*(x-mean(x)) )
Txx <- sum(n*(x^2)) - (sum(n*x))^2/196
( X.trend <- Txp^2/(Txx*mean(p)*mean(q)) )
[1] 6.699793
pchisq(X.trend,df=1,lower.tail=FALSE)
[1] 0.15263
```

Related Rate Models

- ▶ Accounts for occurrences, maximum possible events, time.
- ▶ Note that the binomial does not account for repeat events on the same unit.
- ▶ A key problem is that units may differ in size: disease counts are higher in bigger cities, etc.
- ▶ Focus on rate:

$$\text{Rate} = \frac{\text{\#events}}{\text{unit}} = \frac{\text{occurrences}}{\text{possibilities}}$$

- ▶ Cellular abnormalities from gamma radiation:
 - ▷ **ca** is the count of chromosomal abnormalities,
 - ▷ **cells** is the number (in hundreds) of exposed cells,
 - ▷ **doseamt** = dose amount,
 - ▷ **doserate** = rate of application.

Chi-Square Testing of Rate Models

- Generally we use the more developed tools in R:

```
library(faraway)
data(dicentric)
round(xtabs(ca/cells ~ doseamt + doserate, dicentric),2)
```

	doserate									
doseamt	0.1	0.25	0.5	1	1.5	2	2.5	3	4	
1	0.05	0.05	0.07	0.07	0.06	0.07	0.07	0.07	0.07	
2.5	0.16	0.28	0.29	0.32	0.38	0.41	0.41	0.37	0.44	
5	0.48	0.82	0.90	0.88	1.23	1.32	1.34	1.24	1.43	

```
summary(xtabs(ca/cells ~ doseamt + doserate, dicentric))
```

Number of cases in table: 13.27542

Number of factors: 2

Test for independence of all factors:

Chisq = 0.04019, df = 16, p-value = 1

Homework Assignment Using the Hemodialysis Data

- ▶ Split the sample into the two clinical groups, as we have been doing.
- ▶ Pick an explanatory variable where the Wilcoxon Signed Rank Test is appropriate and run the test (unpaired). Explain your results.
- ▶ Pick a categorical explanatory variable and test a clinical comparison using two-by-two table with chi-square inference. Explain your results.