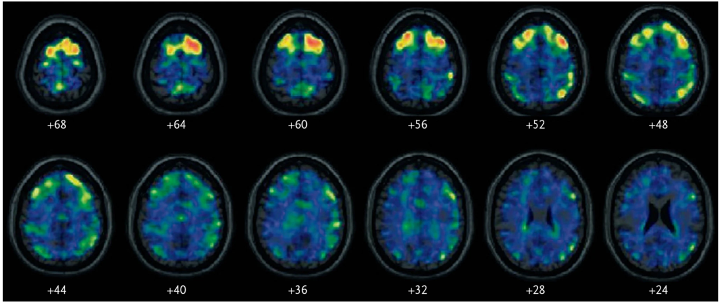
Lecture 36: Non-parametric statistical tests

Nov 27, 2023 (Slides by Corinne Riddell)

Concussions

Concussions: NY times April 10,2019

Abnormal Levels of a Protein Linked to C.T.E. Found in N.F.L Players' Brains, Study Shows



NYtimes article

From the article

ORIGINAL ARTICLE

Tau Positron-Emission Tomography in Former National Football League Players

Robert A. Stern, Ph.D., Charles H. Adler, M.D., Ph.D., Kewei Chen, Ph.D., Michael Navitsky, M.S., Ji Luo, M.S., David W. Dodick, M.D., Michael L. Alosco, Ph.D., Yorghos Tripodis, Ph.D., Dhruman D. Goradia, Ph.D., Brett Martin, M.S., Diego Mastroeni, Ph.D., Nathan G. Fritts, B.A., et al.

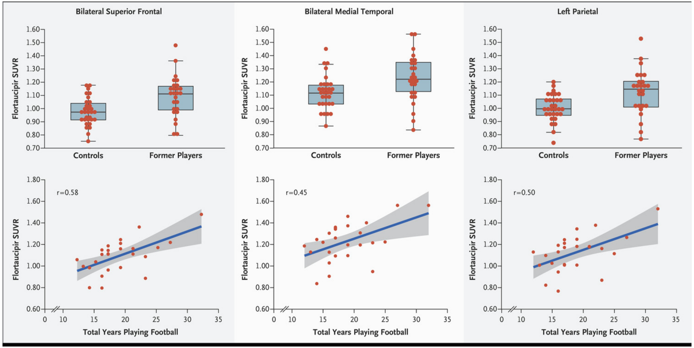
The authors of the study and outside experts stressed that such tau imaging is far from a diagnostic test for C.T.E., which is likely years away and could include other markers, from blood and spinal fluid.

The results of the study, published in The New England Journal of Medicine on Wednesday, are considered preliminary, but constitute a first step toward developing a clinical test to determine the presence of C.T.E. in living players, as well as early signs and potential risk.

From the article

STATISTICAL ANALYSIS Between-group comparisons of age, years of education, and MMSE scores were analyzed with Mann–Whitney U tests. Group differences in race were analyzed with the use of chi-square tests. For between-group comparisons of amyloid-beta plaque burden, chi-square tests were used to compare the proportion of participants with a positive florbetapir PET, and t-tests were used to compare the mean cortical:cerebellar florbetapir standard uptake value ratio (SUVR, the ratio of radioactivity in a cerebral region to that in the cerebellum as a reference) between the groups.

From the article



Roadmap

In part II :

- One sample comparison to a mean (one sample t)
- Two independent samples (two sample t)
- Two non-independent samples (paired t)
- Multiple samples/groups (ANOVA)
  - Bonferroni
  - Tukey's HSD

Roadmap

But all of the methods we have looked at so far depend on some assumptions about the underlying distribution.

What have we assumed?

What do we do if our assumptions are violated?

Non-Parametric testing

Non-Parametric Testing

From http://biostatisticsryangoslingreturns.tumblr.com/

Non-Parametric Testing

PROS: Non-parametric methods make very few assumptions about the variable(s) we samples or their distribution and thus rely less on "parameters".

- Use a ranking of the data instead of actual values
- Do not assume a normal distribution of the data
- Less sensitive to outliers and skewed data
- Do not need a large sample size (sometimes)

CONS: Non-parametric methods use less of the information offered in the data

- If the assumptions of for a parametric test are met and a non-parametric test is used, it will have lower power (probability of detecting a false null hypothesis)
- They are less specific in what they test (e.g., independence)

Non-Parametric Testing

We will discuss non-parametric equivalents for:

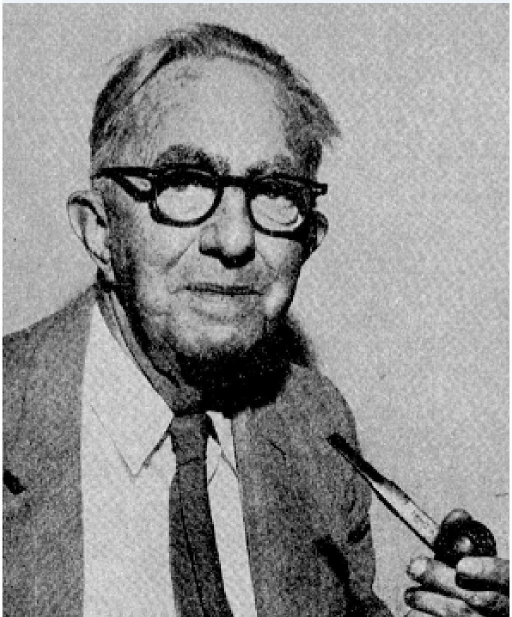
Two sample t : Wilcoxon Rank-Sum

Paired t : Wilcoxon sign-rank

ANOVA: Kruskal Wallis

Wilcoxon two-sample tests

Frank Wilcoxon



Wilcoxon Rank-Sum

- Sometimes also called the Mann-Whitney U test
- Non-parametric test for comparing two independent samples with a continuous outcome
- This is the non-parametric counterpart of the two sample t-test
- Assumes that the distributions have the same general shape but assumes nothing about that shape.
- Evaluates the null hypothesis that the two population distributions are identical.

Wilcoxon Rank-Sum

To calculate a rank sum test

The observations are ordered from lowest to highest and assigned the rank of their order.

If there are "tie" values, these are assigned the average of the ranks, ie if two observations have the same value and the next lower value is rank=3 then the two observations are both given the rank of 4.5 (because they would have been ranks 4 and 5).

Then the sums of ranks belonging to group 1 are compared to the sums of ranks belonging to group 2

Wilcoxon Rank-Sum

Values in group 1: 4,3,5,2,6

Values in group 2: 6,5,7,4,8

Wilcoxon Rank-Sum

Group 1	rank	Group 2	rank
4	3.5	6	7.5
3	2	5	5.5
5	5.5	7	9
2	1	4	3.5
6	7.5	8	10
sum	19.5	sum	35.5

Wilcoxon Rank-Sum

The smaller of the two sums is called W. This is then used in the following equation to generate a Z statistic.

$$Z_w = \frac{W - \mu_w}{\sigma_w}$$

where

$$\mu_w = \frac{n_s(n_s + n_l + 1)}{2}$$

and

$$\sigma_w = \sqrt{\frac{n_s n_l (n_s + n_l + 1)}{12}}$$

Wilcoxon Rank-Sum

So from our example where group 1 had a rank sum of 19.5 and group 2 had a rank sum of 35.5

$$\mu_w = \frac{n_s(n_s + n_l + 1)}{2} = \frac{5(5+5+1)}{2} = 27.5$$

and

$$\sigma_w = \sqrt{\frac{n_s n_l(n_s + n_l + 1)}{12}} = \sqrt{\frac{5 * 5(5+5+1)}{12}} = 4.8$$

$$Z_w = \frac{W - \mu_w}{\sigma_w} = \frac{19.5 - 27.5}{4.8} = -1.67$$

Wilcoxon Rank-Sum

The Z_w we generate follows an approximate standard normal distribution. So we can use our Z score to get a p-value in R

```
2*pnorm(-1.67)
```

```
## [1] 0.09491936
```

Wilcoxon Rank-Sum in R

The general syntax will be:

wilcox.test(group1, group2, paired=F)

or

wilcox.test(outcome ~ group, paired=F)

remember that you can always type help(wilcox.test) in your console to get the full details

Wilcoxon Rank-Sum example :phenylketonuria

Normalized mental age scores for children with phenylketonuria

Group 1: "low exposure" $<$ 10.0 mg/dl

Group 2: "high exposure" $>=$ 10.0 mg/dl

Wilcoxon Rank-Sum :phenylketonuria

```
##   Group  nMA
## 1   low 34.5
## 2   low 37.5
## 3   low 39.5
## 4   low 40.0
## 5   low 45.5
## 6   low 47.0
```

Wilcoxon Rank-Sum :phenylketonuria

In this example there 18 High and 21 Low exposure individuals.

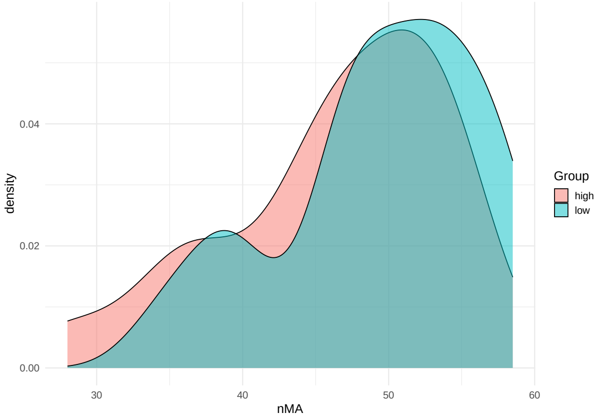
```
group_by(pku,Group) %>%
  summarise(
    count = n(),
    median = median(nMA, na.rm = TRUE),
    IQR = IQR(nMA, na.rm = TRUE)
  )
```

```
## # A tibble: 2 x 4
##   Group count median   IQR
##   <chr> <int>  <dbl> <dbl>
## 1 high     18   48.2  9.12
## 2 low      21   51    7
```

Wilcoxon Rank-Sum: PKU

If we graph the distributions with a density plot what do we notice?

```
ggplot(pku, aes(x = nMA)) +
  geom_density(aes(fill = Group), alpha = 0.5) +
  theme_minimal(base_size = 15)
```



Wilcoxon Rank-Sum: PKU

```r
wilcox.test(nMA ~ Group, data=pku,paired=F)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  nMA by Group
## W = 142, p-value = 0.1896
## alternative hypothesis: true location shift is not equal to 0
```

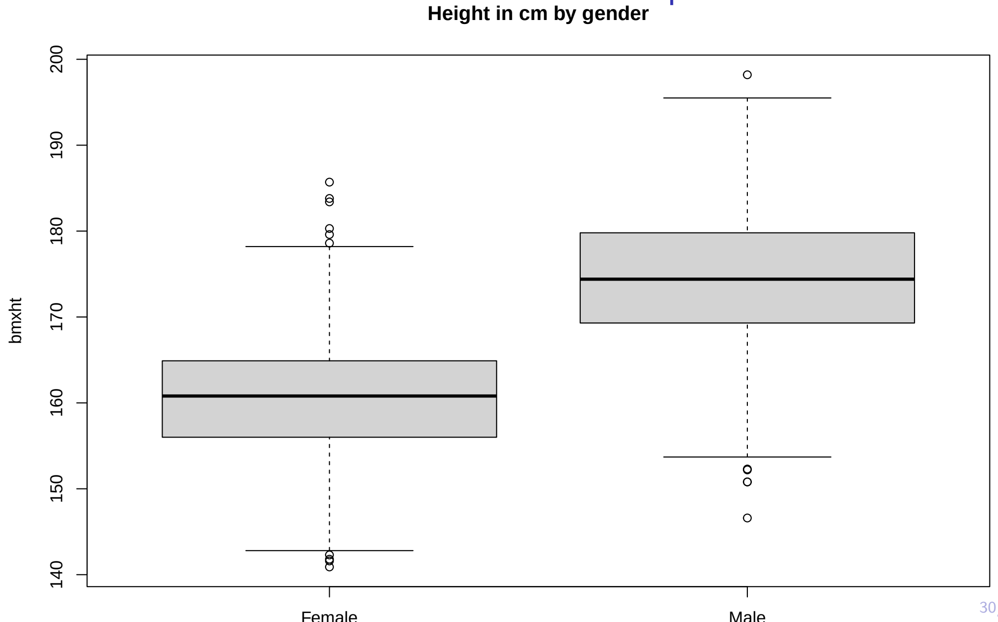
Wilcoxon Rank-Sum vs T : NHANES example

Here I will again use the NHANES data as an example, looking at height by gender

```r
# Read CSV into R
nhanes <- read.csv(file="./data/nhanes.csv", header=TRUE, sep=",")
names(nhanes)
```

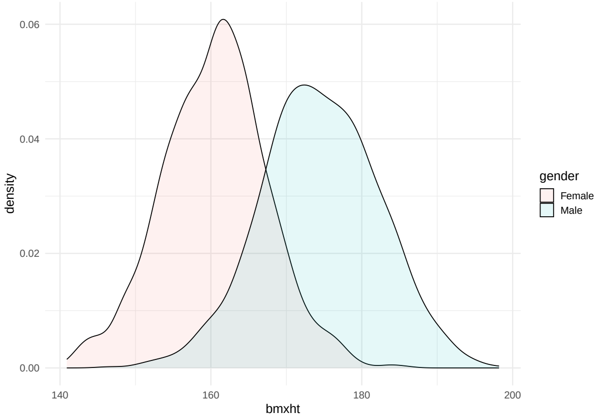
```
##  [1] "ridageyr"   "agegroup"   "gender"     "military"   "born"       "citizen"
##  [7] "drinks"     "drinkscat"  "bmxwt"      "bmxht"      "bmxbmi"     "bmicat"
## [13] "bpxpls"     "bpxsy1"     "bpxsy2"     "sys1d"      "sys2d"      "bpxdi1"
## [19] "bpxdi2"     "dias1d"     "dias2d"     "bpcat"      "chest"      "fs1"
## [25] "fs2"        "fs3"        "lbdhdd"     "hdlcat"     "highhdl"    "hi"
## [31] "asthma"     "vwa"        "vra"        "va"         "aspirin"    "sleep"
## [37] "is"         "hs"         "lbdldl"     "highldl"
```

Wilcoxon Rank-Sum vs T : NHANES example



Wilcoxon Rank-Sum vs T

```
ggplot(nhanes, aes(x = bmxht)) +
  geom_density(aes(fill=gender), alpha=0.1) +
  theme_minimal(base_size = 15)
```



Wilcoxon Rank-Sum vs T

```r
t.test(malesht, femalesht, paired=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  malesht and femalesht
## t = 47.285, df = 2384, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  13.37441 14.53172
## sample estimates:
## mean of x mean of y
##  174.4717  160.5186
```

Wilcoxon Rank-Sum vs T

```r
wilcox.test(malesht,femalesht,paired=F)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  malesht and femalesht
## W = 1402065, p-value < 2.2e-16
## alternative hypothesis: true location shift is not equal to 0
```

Wilcoxon Rank-Sum vs T

- When the sample size is quite large (as with these NHANES data) the assumption of approximate normality (due to CLT) is reasonable one and the results of the hypothesis testing will generally not be different using a parametric or non-parametric approach.
- In smaller sample sizes, with potential outliers, can get more reliable results using Wilcoxon (exact version) than equivalent t-test

Wilcoxon sign rank

Wilcoxon Sign rank

- Non-parametric test for comparing two non-independent (paired) sample means
- This is the non-parametric counterpart of the paired t-test
- Assumes that the distributions have the same general shape but assumes nothing about that shape.
- Evaluates the null hypothesis that the difference between the first and second measures is 0.

Wilcoxon Sign rank

Steps:

1) Calculate the difference between each pair of observations
2) Rank the difference by absolute value from smallest to largest (again, tie values get the average of the ranks). Any pair where difference $= 0$ is thrown out.
3) Assign a "sign" for whether the difference was positive or negative
4) Take the sum of the positive ranks and the sum of the negative ranks (the smaller sum is denoted with a T).

Wilcoxon Sign rank

Under the null hypothesis that the difference is 0, we would expect the sample to have equal numbers of positive and negative ranks with equivalent sums. This expectation is tested against the statistic

$$Z_T = \frac{T - \mu_T}{\sigma_T}$$

Where

$$\mu_T = \frac{n(n+1)}{4}$$

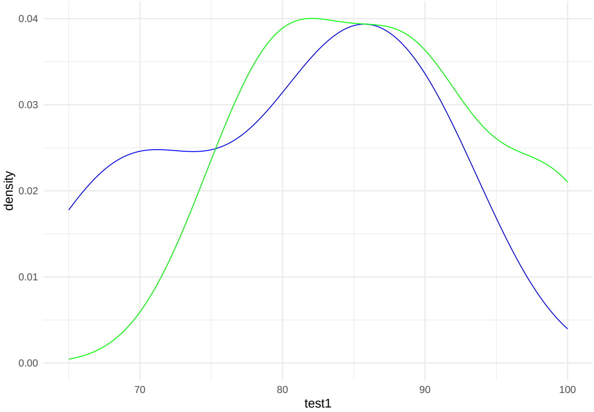
and

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Wilcoxon Sign rank: Example Pre and post test

Time 1	Time 2
65	77
87	100
77	75
90	89
70	80
84	81
92	91
83	96
85	84
91	89
68	88
72	100
81	81

Sign rank example



Sign Rank example: calculate difference and sign

Time 1	Time 2	Difference	sign
65	77	12	+
87	100	13	+
77	75	-2	-
90	89	-1	-
70	80	10	+
84	81	-3	-
92	91	-1	-
83	96	13	+
85	84	-1	-
91	89	-2	-
68	88	20	+
72	100	18	+
81	81	0	?

Sign Rank example: sort by absolute value and assign rank

Time 1	Time 2	Difference	sign	rank
90	89	-1	-	2
92	91	-1	-	2
85	84	-1	-	2
77	75	-2	-	4.5
91	89	-2	-	4.5
84	81	-3	-	6
70	80	10	+	7
65	77	12	+	8
87	100	13	+	9.5
83	96	13	+	9.5
72	100	18	+	11
68	88	20	+	12
81	81	0	?	drop

Sign Rank example: sum the positive and negative ranks

Negative signs

Time 1	Time 2	Difference	sign	rank
90	89	-1	-	2
92	91	-1	-	2
85	84	-1	-	2
77	75	-2	-	4.5
91	89	-2	-	4.5
84	81	-3	-	6

Sum of Negative sign ranks is 21

Sign Rank example: sum the positive and negative ranks

Time 1	Time 2	Difference	sign	rank
70	80	10	+	7
65	77	12	+	8
87	100	13	+	9.5
83	96	13	+	9.5
72	100	18	+	11
68	88	20	+	12

Sum of the positive sign ranks is 57

Wilcoxon Sign rank: Example

Our expectation would be

$$\mu_T = \frac{n(n+1)}{4} = \frac{12(12+1)}{4} = 39$$

remember that we had 13 observations, but we dropped one because the scores at times 1 and 2 were the same and

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{12(12+1)(2*12+1)}{24}} = 12.75$$

Wilcoxon Sign rank: Example

And we compare our expectation to the smaller rank value (Sum of negative ranks was 21, sum of positive ranks was 57)

$$Z_T = \frac{T - \mu_T}{\sigma_T} = \frac{21 - 39}{12.75} = -1.412$$

```
2*pnorm(-1.412)
```

```
## [1] 0.15795
```

Wilcoxon Rank-Sum in R

The general syntax will be:

```
wilcox.test(group1, group2, paired=T)
```

or

```
wilcox.test(outcome ~ group, paired=T)
```

Wilcoxon Sign rank: Example

```r
wilcox.test(test1,test2,paired=T, correct=FALSE)
```

```
## Warning in wilcox.test.default(test1, test2, paired = T, correct = FALSE):
## cannot compute exact p-value with ties

## Warning in wilcox.test.default(test1, test2, paired = T, correct = FALSE):
## cannot compute exact p-value with zeroes

##
##  Wilcoxon signed rank test
##
## data:  test1 and test2
## V = 21, p-value = 0.157
## alternative hypothesis: true location shift is not equal to 0
```

Wilcox Sign rank: compare to T

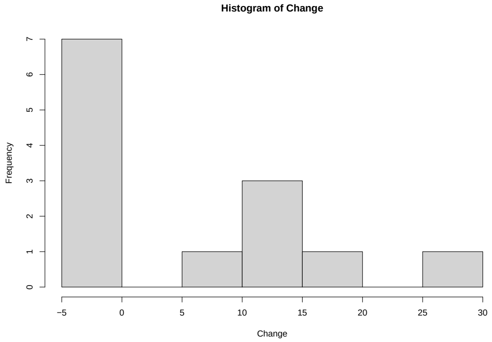
```
t.test(test1,test2,paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  test1 and test2
## t = -2.3684, df = 12, p-value = 0.0355
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -12.7011701  -0.5295991
## sample estimates:
## mean difference
##       -6.615385
```

Wilcox Sign rank: Compare to T

With this study, our sample is 13 and the distribution of changes looks like this - remember that the 0 difference value gets thrown out of sign rank test:

```
hist(Change)
```



Non-parametric test for three or more samples

Kruskal Wallis

The Kruskal Wallis test is a non-parametric alternative to the ANOVA test Kruskal-Wallis looks at the medians of the groups, not the means and tests if at least one is significantly different from another (but not which one) - H_0: There is no difference between the group medians - H_1: There is a statically significant difference in the group medians

Kruskal Wallis

This test can be thought of as an extension of the rank sum test as it is based on the Rank-sum test. We will not do this one by hand.
In R the syntax is generally:

kruskal.test(outcome ~ group, dataset)

Kruskal Wallis


```
##
##  Kruskal-Wallis rank sum test
##
## data:  outcome by treatment
## Kruskal-Wallis chi-squared = 13.096, df = 3, p-value = 0.004434
```

Non parametric summary

Most parametric tests have an analogous non-parametric test We have covered the following:

Samples	Parametric	Non Parametric
Two independent samples	two sample ttest	Wilcoxon rank sum
Two paired samples	paired ttest	Wilcoxon sign rank
Three or more samples	ANOVA	Kruskal Wallis

Non parametrics in R

Samples	test name	R function
Two independent samples	Wilcoxon rank sum	wilcox.test(group1,group2,paired=F)
Two paired samples	Wilcoxon sign rank	wilcox.test(group1,group2,paired=T)
Three or more samples	Kruskal Wallis	kruskal.test(outcome $\sim$ group)

Parting humor