

STAT 401A - Statistical Methods for Research Workers

Nonparametric two-sample tests

Jarad Niemi (Dr. J)

Iowa State University

last updated: September 17, 2014

Nonparametric statistics

http://en.wikipedia.org/wiki/Parametric_statistics

Definition

Parametric statistics assumes that the data have come from a certain probability distribution and makes inferences about the parameters of this distribution, e.g. assuming the data come from a normal distribution and estimating the mean μ .

http://en.wikipedia.org/wiki/Nonparametric_statistics

Definition

Nonparametric statistics make no assumptions about the probability distributions of the [data], e.g. randomization and permutation tests.

Central limit theorem

Theorem

Let X_1, X_2, \dots be a sequence of iid random variables with $E[X_i] = \mu$ and $0 < V[X_i] = \sigma^2 < \infty$. Then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{n \rightarrow \infty} N(0, 1)$$

where

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

i.e. the sample mean using the first n variables.

Central limit theorem

Lemma

Let X_1, X_2, \dots be a sequence of iid random variables with $E[X_i] = \mu$ and $0 < V[X_i] = \sigma^2 < \infty$. Then

$$\frac{\bar{X}_n - \mu}{s_n / \sqrt{n}} \xrightarrow{n \rightarrow \infty} N(0, 1)$$

where

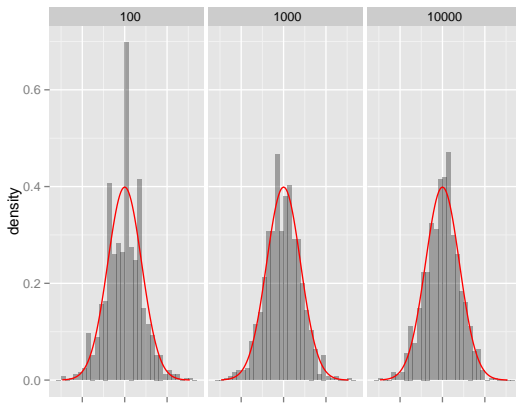
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{and} \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

i.e. the sample mean and variance using the first n variables.

Bernoulli example

Consider $X_i \stackrel{iid}{\sim} \text{Ber}(p)$, i.e. $X_i = 1$ with probability p and $X_i = 0$ with probability $1 - p$. Then $E[X_i] = p$ and $0 < V[X_i] = p(1 - p) < \infty$.

```
stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.  
stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.  
stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



Rusty leaves data

year1	year2	diff	diff>0
38	32	6	1
10	16	-6	0
84	57	27	1
36	28	8	1
50	55	-5	0
35	12	23	1
73	61	12	1
48	29	19	1

If there is no effect, then the “diff>0” column should be a 1 or 0 with probability 0.5, i.e. $X_i \stackrel{iid}{\sim} \text{Ber}(p)$ and $K = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$.

Sign test

The sign test calculates the probability of observing this many ones (or more extreme) if the null hypothesis is true. Here the hypotheses are

$$H_0 : p = 0.5 \quad H_1 : p > 0.5.$$

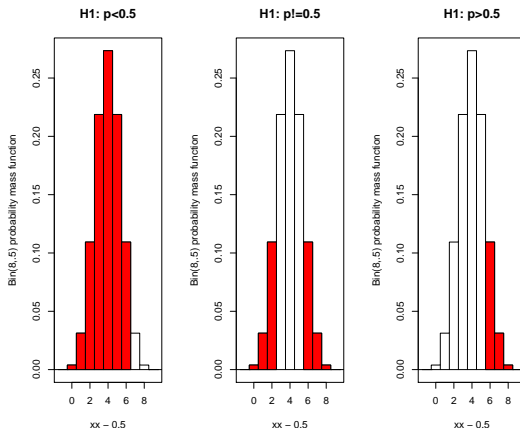
For our one-sided hypothesis (removing leaves will decrease rusty leaves), the pvalue is the probability of observing 6, 7, or 8 ones. This is

$$\binom{8}{6}0.5^8 + \binom{8}{7}0.5^8 + \binom{8}{8}0.5^8 = 0.14$$

```
K = sum(d[,4])
n = nrow(d)
sum(dbinom(K:8,8,.5))
```

```
[1] 0.1445
```

Visualizing pvalues



Sign test using normal approximation

Recall that if $K \sim \text{Bin}(n, p)$, then $E[K] = np$ and $V[K] = np(1 - p)$. Thus, if $p = 0.5$, then

$$Z = \frac{K - (n/2)}{\sqrt{n/4}} \xrightarrow{n \rightarrow \infty} N(0, 1)$$

and we can approximate the pvalue by calculating the area under the normal curve.

```
Z = (K-n/2)/(sqrt(n/4))
1-pnorm(Z)
```

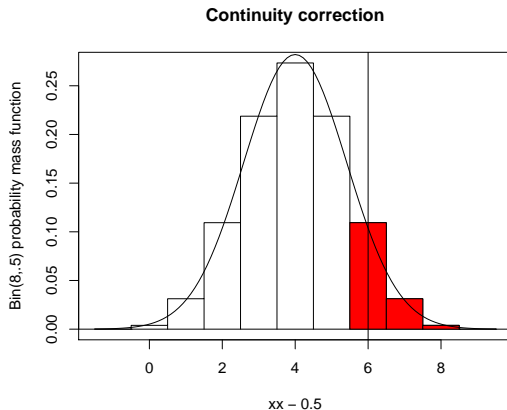
```
[1] 0.07865
```

The continuity correction accounts for the fact that K is discrete:

```
Z = (K-n/2-1/2)/(sqrt(n/4))
1-pnorm(Z)
```

```
[1] 0.1444
```

Continuity correction



Wilcoxon signed-rank test

Also known as the Wilcoxon signed-rank test:

- 1 Compute the difference in each pair.
- 2 Drop zeros from the list.
- 3 Order the absolute differences from smallest to largest and assign them their ranks.
- 4 Calculate S : the sum of the ranks from the pairs for which the difference is positive.
- 5 Calculate $E[S] = n(n+1)/4$ where n is the number of pairs.
- 6 Calculate $SD[S] = [n(n+1)(2n+1)/24]^{1/2}$.
- 7 Calculate $Z = (S - E[S] + c)/SD[S]$ where c is the appropriate continuity correction.
- 8 Calculate the pvalue comparing Z to a standard normal.

Signed rank test

year1	year2	diff	diff>0	absdiff	rank
50	55	-5	0	5	1.0
38	32	6	1	6	2.5
10	16	-6	0	6	2.5
36	28	8	1	8	4.0
73	61	12	1	12	5.0
48	29	19	1	19	6.0
35	12	23	1	23	7.0
84	57	27	1	27	8.0

- $S = 32.5$
- $E[S] = 18$
- $SD[S] = 7.14$
- $Z = 1.96$ (with continuity correction of -0.5)
- $p = 0.02$

Signed-rank test in R

```
# By hand
S = sum(d$rank[d$"diff">0"]==1)
n = nrow(d)
ES = n*(n+1)/4
SDS = sqrt(n*(n+1)*(2*n+1)/24)
z = (S-0.5-ES)/SDS
1-pnorm(z)

[1] 0.02497

# Using a function
wilcox.test(d$year1, d$year2, paired=T)

Warning: cannot compute exact p-value with ties

Wilcoxon signed rank test with continuity correction

data: d$year1 and d$year2
V = 32.5, p-value = 0.04967
alternative hypothesis: true location shift is not equal to 0
```

Divide this two-sided pvalue by 2 since the data are in agreement with the alternative hypothesis (fewer rusty leaves after removal).

SAS code for paired nonparametric test

```
DATA leaves;
  INPUT tree year1 year2;
  diff = year1-year2;
  DATALINES;
1 38 32
2 10 16
3 84 57
4 36 28
5 50 55
6 35 12
7 73 61
8 48 29
;

PROC UNIVARIATE DATA=leaves;
  VAR diff;
RUN;
```

SAS code for paired nonparametric tests

The UNIVARIATE Procedure
Variable: diff

Moments

N	8	Sum Weights	8
Mean	10.5	Sum Observations	84
Std Deviation	12.2007026	Variance	148.857143
Skewness	-0.1321468	Kurtosis	-1.2476273
Uncorrected SS	1924	Corrected SS	1042
Coeff Variation	116.197167	Std Error Mean	4.31359976

Basic Statistical Measures

Location		Variability	
Mean	10.50000	Std Deviation	12.20070
Median	10.00000	Variance	148.85714
Mode	.	Range	33.00000
		Interquartile Range	20.50000

Tests for Location: Mu0=0

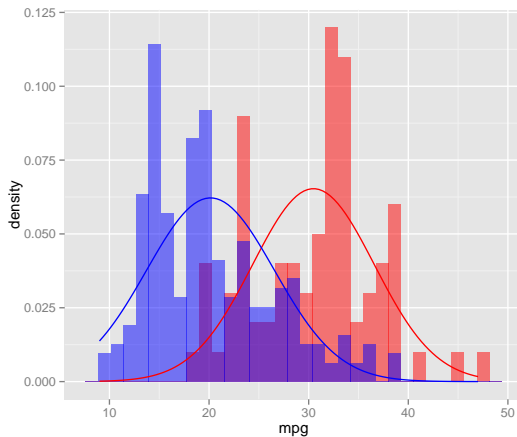
Test	-Statistic-	-----p Value-----		
Student's t	t 2.434162	Pr > t	0.0451	
Sign	M 2	Pr >= M	0.2891	
Signed Rank	S 14.5	Pr >= S	0.0469	

Conclusion

Removal of red cedar trees within 100 yards is associated with a significant reduction in rusty apple leaves (Wilcoxon signed rank test, $p=0.023$).

Do these data look normal?

```
stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.  
stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```



Rank-sum test

Also referred to as the Wilcoxon rank-sum test and the Mann-Whitney U test:

- ➊ Transform the data to ranks
- ➋ Calculate U , the sum of ranks of the group with a smaller sample size
- ➌ Calculate $E[U] = n_1 \bar{R}$
 - ➊ n_1 : sample size of the smaller group
 - ➋ \bar{R} : average rank
- ➍ Calculate $SD(U) = s_R \sqrt{\frac{n_1 n_2}{(n_1 + n_2)}}$
 - ➊ n_2 : sample size of the larger group
 - ➋ s_R : standard deviation of the ranks
- ➎ Calculate $Z = (U + c - E[U])/SD(U)$ where c , the continuity correction, is either 0.5 or -0.5.
- ➏ Determine the pvalue using a standard normal distribution.

Example on a small dataset

mpg	country	rank
13	US	1.0
15	US	2.0
17	US	3.0
22	US	4.0
26	Japan	5.5
26	US	5.5
28	US	7.0
32	Japan	8.0
33	Japan	9.0

- $U = 22.5$
- $E[U] = 15$
- $SD[U] = 3.86$
- $z = 1.81$ (appropriate continuity correction is -0.5)
- $p = 0.07$

Example on a small dataset

```
n1 = sum(sm$country=="Japan")
n2 = sum(sm$country=="US")
U = sum(sm$rank[sm$country=="Japan"])
EU = n1*mean(sm$rank)
SDU = sd(sm$rank) * sqrt(n1*n2/(n1+n2))
Z = (U-.5-EU)/SDU
2*pnorm(-Z)
```

```
[1] 0.06953
```

```
wilcox.test(mpg~country, sm)
```

```
Warning: cannot compute exact p-value with ties
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: mpg by country
```

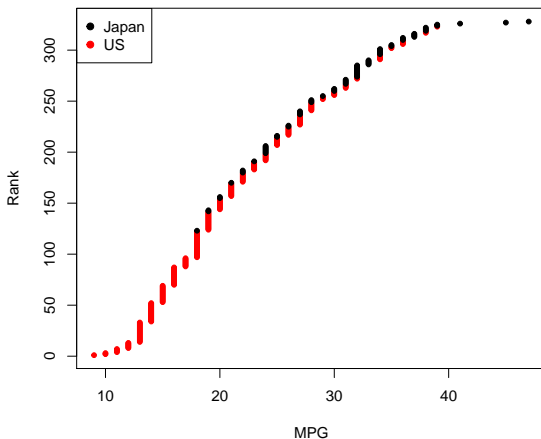
```
W = 16.5, p-value = 0.06953
```

```
alternative hypothesis: true location shift is not equal to 0
```

Visual representation of Rank Sum Test

```
ordr = order(mpg$mpg)
mpg.ordered = mpg[ordr,]

par(mar=c(5,4,0,0)+.1)
plot(mpg.ordered$mpg, 1:nrow(mpg), col=mpg.ordered$country, pch=19, xlab="MPG", cex=0.7, ylab="Rank")
legend("topleft", c("Japan", "US"), col=1:2, pch=19)
```



R code and output for Rank Sum Test

```
wilcox.test(mpg~country,mpg)
```

Wilcoxon rank sum test with continuity correction

data: mpg by country

W = 17150, p-value < 2.2e-16

alternative hypothesis: true location shift is not equal to 0

SAS code for Wilcoxon rank sum test

```
DATA mpg;  
    INFILE 'mpg.csv' DELIMITER=', ' FIRSTOBS=2;  
    INPUT mpg country $;  
  
PROC NPAR1WAY DATA=mpg WILCOXON;  
    CLASS country;  
    VAR mpg;  
    RUN;
```

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable mpg
Classified by Variable country

country	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
US	249	33646.50	40960.50	733.579091	135.126506
Japan	79	20309.50	12995.50	733.579091	257.082278

Average scores were used for ties.

Wilcoxon Two-Sample Test

Statistic 20309.5000

Normal Approximation

Z 9.9696

One-Sided Pr > Z <.0001

Two-Sided Pr > |Z| <.0001

t Approximation

One-Sided Pr > Z <.0001

Two-Sided Pr > |Z| <.0001

Z includes a continuity correction of 0.5.

Kruskal-Wallis Test

Chi-Square 99.4068

DF 1

Conclusion

Average miles per gallon of Japanese cars are significantly different than average miles per gallon of American cars (Wilcoxon rank sum test, $p < 0.0001$).