Rank-Sum Test

STAT 401 - Statistical Methods for Research Workers

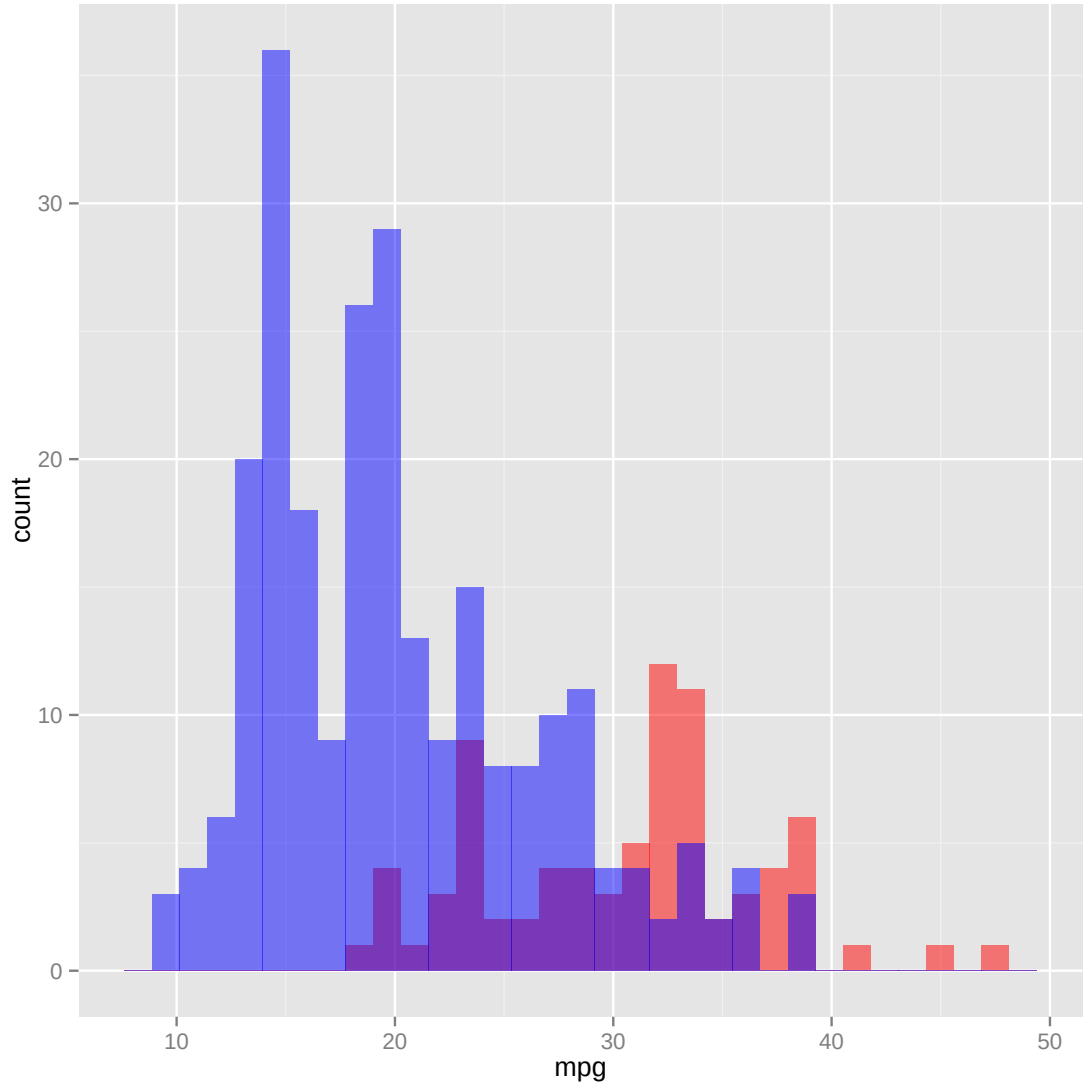
Jarad Niemi

Iowa State University

6 September 2013

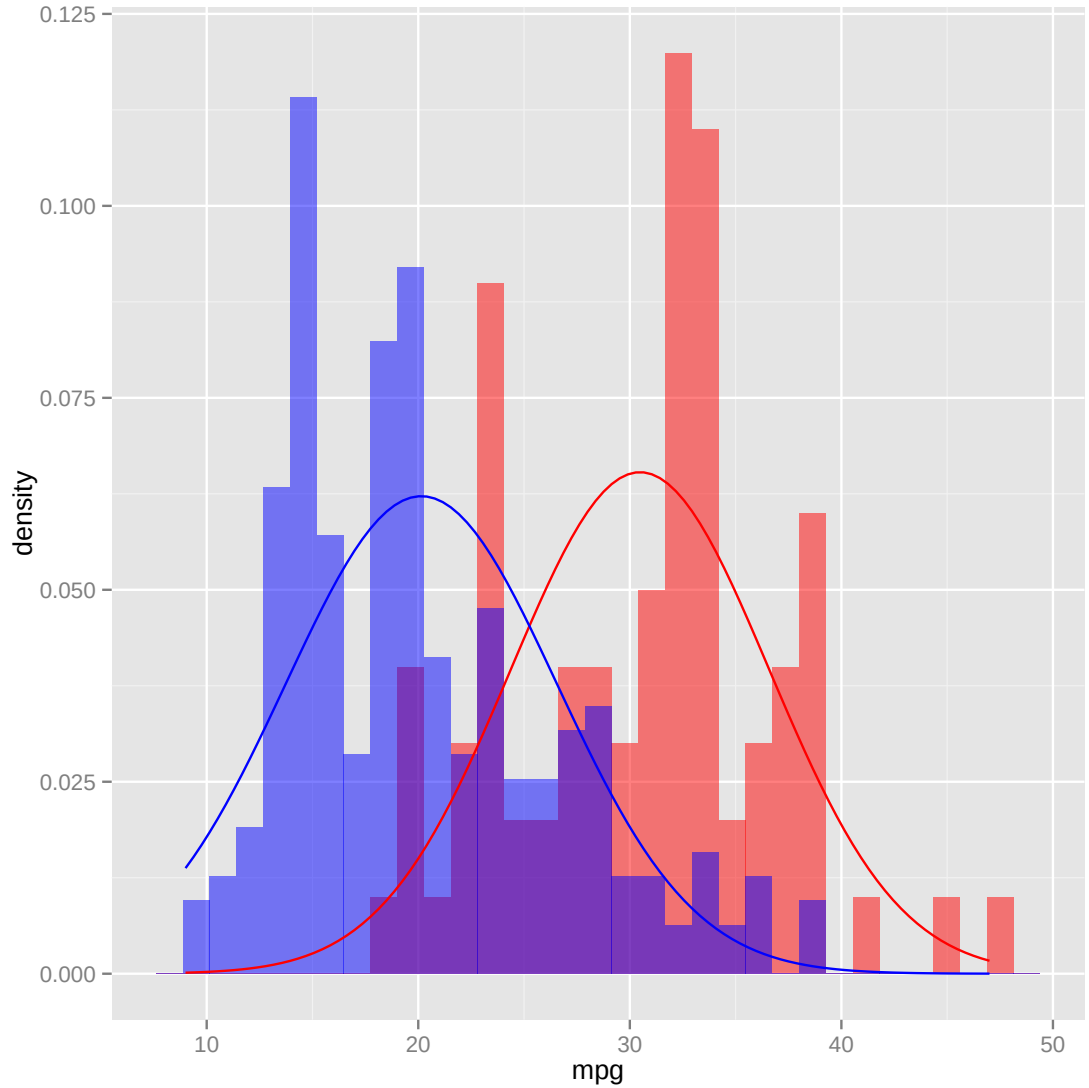
Do these data look normal?

Raw histogram of mpg for US and Japanese cars.



Do these data look normal?

Compared to best fitting normal curves.



Rank-sum test

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

Rank-sum test

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

- Transform the data to ranks

Rank-sum test

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

- Transform the data to ranks
- Calculate U, the sum of ranks of the group with a smaller sample size

Rank-sum test

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

- Transform the data to ranks
- Calculate U, the sum of ranks of the group with a smaller sample size
- Calculate $E[U] = n_1\overline{R}$

Rank-sum test

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

- Transform the data to ranks
- Calculate U, the sum of ranks of the group with a smaller sample size
- Calculate $E[U] = n_1\overline{R}$
  - n_1: sample size of the smaller group

Rank-sum test

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

- Transform the data to ranks
- Calculate U, the sum of ranks of the group with a smaller sample size
- Calculate $E[U] = n_1 \overline{R}$
  - n_1: sample size of the smaller group
  - $\overline{R}$: average rank

Rank-sum test

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

- Transform the data to ranks
- Calculate U, the sum of ranks of the group with a smaller sample size
- Calculate $E[U] = n_1\overline{R}$
  - n_1: sample size of the smaller group
  - $\overline{R}$: average rank
- Calculate $SD(U) = s_R\sqrt{\frac{n_1 n_2}{(n_1+n_2)}}$

Rank-sum test

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

- Transform the data to ranks
- Calculate U, the sum of ranks of the group with a smaller sample size
- Calculate $E[U] = n_1\overline{R}$
  - n_1: sample size of the smaller group
  - $\overline{R}$: average rank
- Calculate $SD(U) = s_R\sqrt{\frac{n_1 n_2}{(n_1+n_2)}}$
  - n_2: sample size of the larger group

Rank-sum test

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

- Transform the data to ranks
- Calculate U, the sum of ranks of the group with a smaller sample size
- Calculate $E[U] = n_1\overline{R}$
  - n_1: sample size of the smaller group
  - $\overline{R}$: average rank
- Calculate $SD(U) = s_R\sqrt{\frac{n_1 n_2}{(n_1+n_2)}}$
  - n_2: sample size of the larger group
  - s_R: standard deviation of the ranks

Rank-sum test

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

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

Rank-sum test

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

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

Example on a small dataset

	mpg	country	rank
1	13	US	1
2	15	US	2
3	17	US	3
4	22	US	4
5	26	Japan	5.5
6	26	US	5.5
7	28	US	7
8	32	Japan	8
9	33	Japan	9

Example on a small dataset

	mpg	country	rank
1	13	US	1
2	15	US	2
3	17	US	3
4	22	US	4
5	26	Japan	5.5
6	26	US	5.5
7	28	US	7
8	32	Japan	8
9	33	Japan	9

- $U = 22.5$

Example on a small dataset

	mpg	country	rank
1	13	US	1
2	15	US	2
3	17	US	3
4	22	US	4
5	26	Japan	5.5
6	26	US	5.5
7	28	US	7
8	32	Japan	8
9	33	Japan	9

- $U = 22.5$
- $E[U] = 15$

Example on a small dataset

	mpg	country	rank
1	13	US	1
2	15	US	2
3	17	US	3
4	22	US	4
5	26	Japan	5.5
6	26	US	5.5
7	28	US	7
8	32	Japan	8
9	33	Japan	9

- $U = 22.5$
- $E[U] = 15$
- $SD[U] = 3.86$

Example on a small dataset

	mpg	country	rank
1	13	US	1
2	15	US	2
3	17	US	3
4	22	US	4
5	26	Japan	5.5
6	26	US	5.5
7	28	US	7
8	32	Japan	8
9	33	Japan	9

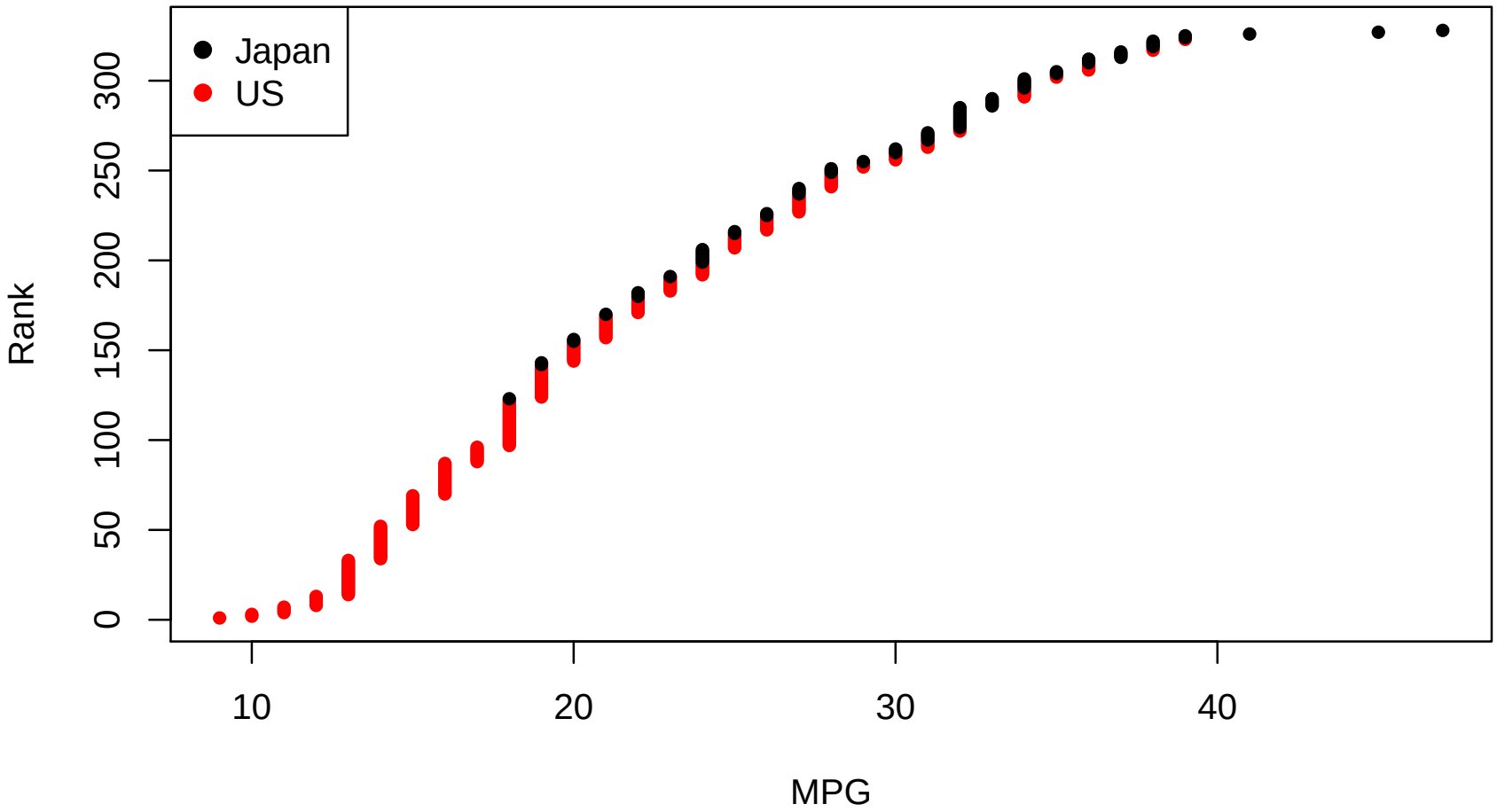
- $U = 22.5$
- $E[U] = 15$
- $SD[U] = 3.86$
- $z = 1.81$ (appropriate continuity correction is -0.5)

Example on a small dataset

	mpg	country	rank
1	13	US	1
2	15	US	2
3	17	US	3
4	22	US	4
5	26	Japan	5.5
6	26	US	5.5
7	28	US	7
8	32	Japan	8
9	33	Japan	9

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

Transform data to ranks



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
Pr > Chi-Square	<.0001

Conclusion

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