

# Stats 506, F18, Problem Set 1

*Xun Wang, xunwang@umich.edu*

*September 30, 2018*

## Question 1

Using bash to execute the shell script, the answers are below:

```
## Response to Part A part one:
## 2010
## Response to Part A part two:
## Created a .csv.gz file containing only DOEID, NWEIGHT, and BRRWT1-BRRWT96 for part two.
## Response to part B part one:
## observations within region "1"
## 794
## observations within region "2"
## 1327
## observations within region "3"
## 2010
## observations within region "4"
## 1555
## Response to Part B part two:
## set up a .txt file and contents of the file are:
## Regionc Division
## "1" "1"
## "1" "2"
## "2" "3"
## "2" "4"
## "3" "5"
## "3" "6"
## "3" "7"
## "4" "8"
## "4" "9"
## "4" "10"
```

## Question 2: Including Plots

### Question 2 Part a

Table 1: **Table 1.** *Airlines names and percentage:* airlines responsible for at least 1% flights from Jan. 1 to Oct. 31, 2013.

Airlines Names	Percentage(%)
Endeavor Air Inc.	5.41
American Airlines Inc.	9.75
JetBlue Airways	16.21
Delta Air Lines Inc.	14.28
ExpressJet Airlines Inc.	16.13
AirTran Airways Corporation	1.01
Envoy Air	7.89
United Air Lines Inc.	17.37
US Airways Inc.	6.12
Virgin America	1.51
Southwest Airlines Co.	3.60

According to table 1, 11 listed airlines were responsible for at least 1% of the flights departing any of the three NYC airports between January 1 and October 31, 2013. And the right column is the percent of thier flights.

### Question 2 Part b

Because the table is too long, we seperate it into three parts: the table for 2013 year, 2014 year and the difference between them. These tables include airline names, flight numbers, percents and limits of confidence intervals.

Table 2: **Table 2.** *Number, percent and confidence interval:* flights in the first 10 months of 2013.

Airlines	Number 2013	Percent 2013(%)	CI Lower limit 2013(%)	CI Upper limit 2013(%)
Endeavor Air Inc.	15232	5.41	5.33	5.50
American Airlines Inc.	27447	9.75	9.65	9.86
JetBlue Airways	45605	16.21	16.07	16.34
Delta Air Lines Inc.	40168	14.28	14.15	14.40
ExpressJet Airlines Inc.	45395	16.13	16.00	16.27
AirTran Airways Corporation	2845	1.01	0.97	1.05
Envoy Air	22202	7.89	7.79	7.99
United Air Lines Inc.	48880	17.37	17.23	17.51
US Airways Inc.	17232	6.12	6.04	6.21
Virgin America	4235	1.51	1.46	1.55
Southwest Airlines Co.	10143	3.60	3.54	3.67

Table 3: **Table 3.** *Number, percent and confidence interval:* flights in the first 10 months of 2014.

Airlines	Numbers 2014	Percent 2014(%)	CI Lower limit 2014(%)	CI Upper limit 2014(%)
American Airlines Inc.	26302	10.38	10.26	10.50
JetBlue Airways	44479	17.56	17.41	17.71
Delta Air Lines Inc.	41683	16.45	16.31	16.60

Airlines	Numbers 2014	Percent 2014(%)	CI Lower limit 2014(%)	CI Upper limit 2014(%)
ExpressJet Airlines Inc.	39819	15.72	15.58	15.86
AirTran Airways Corporation	1251	0.49	0.47	0.52
Envoy Air	18559	7.33	7.22	7.43
United Air Lines Inc.	46267	18.26	18.11	18.42
US Airways Inc.	16750	6.61	6.52	6.71
Virgin America	4797	1.89	1.84	1.95
Southwest Airlines Co.	11902	4.70	4.62	4.78

Table 4: **Table 4.** *number, percent and confidence interval: flights difference in the first 10 months of 2013 and the first 10 months of 2014. Sorted by percent of difference.*

Airlines	Number_diff	Percent_diff(%)	CI Lowerlimit_diff(%)	CI Upperlimit_diff(%)
Delta Air Lines Inc.	1515	2.18	1.99	2.37
JetBlue Airways	-1126	1.35	1.15	1.55
Southwest Airlines Co.	1759	1.09	0.99	1.20
United Air Lines Inc.	-2613	0.89	0.69	1.10
American Airlines Inc.	-1145	0.63	0.47	0.79
US Airways Inc.	-482	0.49	0.36	0.62
Virgin America	562	0.39	0.32	0.46
ExpressJet Airlines Inc.	-5576	-0.41	-0.61	-0.22
AirTran Airways Corporation	-1594	-0.52	-0.56	-0.47
Envoy Air	-3643	-0.56	-0.71	-0.42
Endeavor Air Inc.	NA	NA	NA	NA

Limiting tables to the airlines identified in part a, and sorted by percent of difference, it could be observed that the largest increase appeared in Delta Air Lines Inc.. And Envoy Air had the largest decrease from year 2013 to year 2014 except the missing value of company Endeavor Air Inc.. We could get the same result by sorting by R functions.

```
## [1] "Delta Air Lines Inc."
```

```
## [1] "Envoy Air"
```

Then filtering in R, the below table are airlines showed increase in the percent of flights but a decrease in the number of flights. Calculating in R, up to 2013 October there were 279384 times flight in total but 2014 only had 251809 times. So decrease in the number of total flights made this phenomenon.

Table 5: **Table 5.** *Number and percent: flights in the first 10 months of 2013 and the first 10 months of 2014. Sorted by increase in percent and a decrease in number.*

Airlines	Numbers13	Percent13(%)	numbers14	Percent14(%)	number_diff	Percent_diff(%)
JetBlue Airways	45605	16.21	44479	17.56	-1126	1.35
United Air Lines Inc.	48880	17.37	46267	18.26	-2613	0.89
American Airlines Inc.	27447	9.75	26302	10.38	-1145	0.63
US Airways Inc.	17232	6.12	16750	6.61	-482	0.49

## Question 2 Part c

Among of the three NYC airports, we produce a table showing the percent of flights each airline was responsible for in 2013, limiting the airlines in part a, including confidence intervals for estimates.

Table 6: **Table 6.** *Number,airports,percent and confidence interval:* flights in 2013.

year	Airports	Airlines	Percent2013(%)	CI Lowerlimit 2013(%)	CI Upperlimit 2013(%)
2013	EWR	Endeavor Air Inc.	1.05	0.99	1.11
2013	EWR	American Airlines Inc.	2.89	2.79	2.98
2013	EWR	JetBlue Airways	5.43	5.30	5.55
2013	EWR	Delta Air Lines Inc.	3.59	3.49	3.70
2013	EWR	ExpressJet Airlines Inc.	36.36	36.09	36.63
2013	EWR	Envoy Air	1.88	1.81	1.96
2013	EWR	United Air Lines Inc.	38.14	37.87	38.41
2013	EWR	US Airways Inc.	3.65	3.54	3.75
2013	EWR	Virgin America	1.30	1.23	1.36
2013	EWR	Southwest Airlines Co.	5.12	5.00	5.25
2013	JFK	Endeavor Air Inc.	13.17	12.97	13.36
2013	JFK	American Airlines Inc.	12.39	12.19	12.58
2013	JFK	JetBlue Airways	37.81	37.53	38.10
2013	JFK	Delta Air Lines Inc.	18.60	18.37	18.83
2013	JFK	ExpressJet Airlines Inc.	1.27	1.20	1.33
2013	JFK	Envoy Air	6.46	6.32	6.61
2013	JFK	United Air Lines Inc.	4.07	3.96	4.19
2013	JFK	US Airways Inc.	2.69	2.60	2.79
2013	JFK	Virgin America	3.23	3.13	3.34
2013	LGA	Endeavor Air Inc.	2.43	2.33	2.52
2013	LGA	American Airlines Inc.	14.77	14.56	14.99
2013	LGA	JetBlue Airways	5.73	5.59	5.88
2013	LGA	Delta Air Lines Inc.	22.04	21.79	22.29
2013	LGA	ExpressJet Airlines Inc.	8.43	8.26	8.60
2013	LGA	AirTran Airways Corporation	3.11	3.01	3.22
2013	LGA	Envoy Air	16.17	15.95	16.40
2013	LGA	United Air Lines Inc.	7.69	7.52	7.85
2013	LGA	US Airways Inc.	12.55	12.35	12.75
2013	LGA	Southwest Airlines Co.	5.82	5.67	5.96

Calculating in R, the following results are largest carrier in airport EWR, JFK,LGA, respectively.

```
## [1] "United Air Lines Inc."
## [1] "JetBlue Airways"
## [1] "Delta Air Lines Inc."
```

And getting the table for 2014 using the same ways.

Table 7: **Table 7.** *Number,airports,percent and confidence interval: flights in 2014.*

year	Airports	Airlines	Percent2014(%)	CI Lowerlimit 2014(%)	CI Upperlimit 2014(%)
2014	EWR	American Airlines Inc.	3.03	2.92	3.14
2014	EWR	JetBlue Airways	6.26	6.10	6.42
2014	EWR	Delta Air Lines Inc.	4.75	4.61	4.89
2014	EWR	ExpressJet Airlines Inc.	32.41	32.10	32.72
2014	EWR	Envoy Air	0.19	0.16	0.22
2014	EWR	United Air Lines Inc.	41.33	41.01	41.66
2014	EWR	US Airways Inc.	3.98	3.85	4.11
2014	EWR	Virgin America	1.88	1.79	1.97
2014	EWR	Southwest Airlines Co.	5.50	5.35	5.65
2014	JFK	American Airlines Inc.	14.63	14.39	14.88
2014	JFK	JetBlue Airways	42.00	41.66	42.34
2014	JFK	Delta Air Lines Inc.	23.15	22.86	23.44
2014	JFK	ExpressJet Airlines Inc.	1.31	1.23	1.39
2014	JFK	Envoy Air	6.68	6.51	6.85
2014	JFK	United Air Lines Inc.	4.82	4.67	4.96
2014	JFK	US Airways Inc.	3.25	3.12	3.37
2014	JFK	Virgin America	3.85	3.72	3.98
2014	LGA	American Airlines Inc.	13.89	13.66	14.13
2014	LGA	JetBlue Airways	5.67	5.51	5.82
2014	LGA	Delta Air Lines Inc.	22.11	21.83	22.39
2014	LGA	ExpressJet Airlines Inc.	12.34	12.12	12.57
2014	LGA	AirTran Airways Corporation	1.48	1.40	1.56
2014	LGA	Envoy Air	15.34	15.09	15.58
2014	LGA	United Air Lines Inc.	7.37	7.19	7.54
2014	LGA	US Airways Inc.	12.58	12.36	12.81
2014	LGA	Virgin America	0.02	0.01	0.03
2014	LGA	Southwest Airlines Co.	8.40	8.22	8.59

Similarly, we get the largest carrier in airports EWR, JFK,LGA in 2014, respectively.

```
## [1] "United Air Lines Inc."
```

```
## [1] "JetBlue Airways"
```

```
## [1] "Delta Air Lines Inc."
```

### Question 3: Including Plots

#### Question 3 Part a

Table 8: **Table 8.** *Divisions, percentage and BRR standard error: stucco as the major outside wall material within each division. Sorted by percent of stucco.*

Divisions	Percentage of stucco(%)	standard error	relative standard error
Mountain South	64.25	4.49	6.99
Pacific	44.59	1.69	3.78
Mountain North	16.61	3.24	19.53
South Atlantic	10.62	1.42	13.38
West North Central	4.87	2.03	41.67
West South Central	2.99	0.69	23.23
Middle Atlantic	2.06	0.73	35.62
New England	1.23	0.80	65.31
East North Central	0.66	0.28	43.17
East South Central	0.42	0.41	96.69

From the table above, we could see percent of homes having stucco construction as their major outside wall material within each division. Obviously, Mountain South had the largest proportion of stucco homes and East South Central had the lowest percent. It is the same with extracting their names in R.

```
## [1] "Mountain South"
```

```
## [1] "East South Central"
```

#### Question 3 Part b

The table 9 below shows the average total electricity usage in kilowatt hours in each division and also show the result stratified by urban and rural status in table 10.

Table 9: **Table 9.** *Divisions, average electricity usage and BRR standard error: in total.*

Divisions	Average usage of electricity	standard error	relative standard error
East North Central	9128.66	203.63	2.23
East South Central	14535.97	620.63	4.27
Middle Atlantic	8465.44	201.10	2.38
Mountain North	8384.47	644.43	7.69
Mountain South	10442.02	1271.40	12.18
New England	7514.56	532.02	7.08
Pacific	8100.41	178.60	2.20
South Atlantic	13446.62	276.88	2.06
West North Central	10523.83	453.62	4.31
West South Central	14324.26	423.04	2.95

Table 10: **Table 10.** *Divisions, average electricity usage and BRR standard error: stratified by urban and rural status.*

Divisions	status	Average usage of electricity	standard error	relative standard error
East North Central	Rural	13500.02	754.27	5.59

Divisions	status	Average usage of electricity	standard error	relative standard error
East North Central	Urban	7980.07	218.20	2.73
East South Central	Rural	16332.68	1145.51	7.01
East South Central	Urban	13747.31	791.21	5.76
Middle Atlantic	Rural	12223.39	811.45	6.64
Middle Atlantic	Urban	7987.29	167.61	2.10
Mountain North	Rural	9356.00	1866.28	19.95
Mountain North	Urban	8099.38	358.81	4.43
Mountain South	Rural	8610.43	1058.29	12.29
Mountain South	Urban	10742.91	1308.63	12.18
New England	Rural	9001.05	1140.32	12.67
New England	Urban	6963.86	533.80	7.67
Pacific	Rural	14114.76	1078.51	7.64
Pacific	Urban	7348.88	226.60	3.08
South Atlantic	Rural	15941.99	562.96	3.53
South Atlantic	Urban	12724.60	301.49	2.37
West North Central	Rural	14173.93	798.83	5.64
West North Central	Urban	9466.52	379.83	4.01
West South Central	Rural	16317.16	1148.01	7.04
West South Central	Urban	13628.52	396.21	2.91

### Question 3 Part c

In order to get the largest disparity between these proportions, we calculate the proportion of homes with internet access of urban and rural areas first, and the BBR standard error to estimate these proportions respectively.

Table 11: **Table 11.** *Divisions, proportion of internet access and BRR standard error.*

Divisions	status	Proportion of internet access(%)	standard error	relative standard error
East North Central	Rural	86.21	2.33	2.71
East North Central	Urban	86.25	1.27	1.47
East South Central	Rural	69.03	2.82	4.09
East South Central	Urban	78.36	3.99	5.09
Middle Atlantic	Rural	91.29	3.05	3.34
Middle Atlantic	Urban	89.34	2.80	3.13
Mountain North	Rural	81.93	4.14	5.05
Mountain North	Urban	87.42	2.77	3.17
Mountain South	Rural	66.75	4.33	6.49
Mountain South	Urban	85.27	2.01	2.36
New England	Rural	85.79	1.75	2.04
New England	Urban	87.57	2.59	2.95
Pacific	Rural	85.28	4.00	4.69
Pacific	Urban	88.71	1.29	1.46
South Atlantic	Rural	82.04	2.94	3.58
South Atlantic	Urban	85.30	1.36	1.59
West North Central	Rural	80.33	4.51	5.61
West North Central	Urban	88.00	1.72	1.95
West South Central	Rural	76.50	2.23	2.92
West South Central	Urban	81.61	2.65	3.25

And then we get their disparity through spreading the tibble.

Table 12: **Table 12.** *Divisions, disparity of proportion of internet access and BRR standard error: Sorted by disparity.*

Divisions	Proportion of rural status(%)	Proportion of urban status(%)	Disparity of proportion(%)
Mountain South	66.75	85.27	18.52
East South Central	69.03	78.36	9.33
West North Central	80.33	88.00	7.68
Mountain North	81.93	87.42	5.50
West South Central	76.50	81.61	5.10
Pacific	85.28	88.71	3.43
South Atlantic	82.04	85.30	3.26
Middle Atlantic	91.29	89.34	1.95
New England	85.79	87.57	1.78
East North Central	86.21	86.25	0.04

The largest disparity happened in:

```
## [1] "Mountain South"
```