

Stats 506, F18, Problem Set 1

Xinye Jiang (xinyej@umich.edu)

September 30, 2018

Question 1

Part A

i.

Solution: `cut -d ',' -f 2 recs2015_public_v3.csv | grep 3 | wc -l`

Result: 2010

There are 2010 rows for region 3 in the RECS 2015 data set. (Region has column number 2.)

ii.

Solution: `cut -d ',' -f 1,475-571 recs2015_public_v3.csv > Aii.csv | gzip Aii.csv`

Result: Get a compressed data set containing only the variables: DOEID, NWEIGHT, and BRRWT1-BRRWT96 (Column numbers are 1,475-571.) in the working directory.

Part B

i.

Solution: `awk -F "," 'NR>1 {col[$2]++} END {for(i in col) print i,col[i]}' recs2015_public_v3.csv`

Result:

```
"1" 794
"2" 1327
"3" 2010
"4" 1555
```

The result shows the number of observations within each region.

ii.

Solution: `awk -F ',' '{print $2,$3}' recs2015_public_v3.csv | (read -r; printf "%s\n" "$REPLY"; sort -V) | uniq > region_division.txt`

Result: Get a file “region_division.txt” in the working directory and its contents is shown below.

```
"REGIONC" "DIVISION"
"1" "1"
"1" "2"
"2" "3"
"2" "4"
"3" "5"
"3" "6"
"3" "7"
```

"4" "8"
"4" "9"
"4" "10"

Question 2

a.

Table 1: **Table 1.** *Proportion of flights which was at least 1% between January 1 and October 31, 2013 by airlines.* Each row shows the flights distribution for one airline. Rows are sorted by the proportion of flights.

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

The above table shows the names of those airlines which were responsible for at least 1% of the flights departing any of the three NYC airports between January 1 and October 31, 2013.

b.

Divide the required table into several tables to take a deep look at the data.

Table 2: **Table 2.** *Proportion with 95% CI and number of flights by airlines in the first 10 months of 2013.* Each row shows percent with 95% CI and number for one airline.

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

Table 3: **Table 3.** *Proportion with 95% CI and number of flights by airlines in the first 10 months of 2014.* Each row shows percent with 95% CI and number for one airline.

Airline	2014 number (%)	2014 proportion (%)	2014 lwr (%)	2014 upr (%)
Envoy Air	18559	7.33	7.22	7.43
AirTran Airways Corporation	1251	0.49	0.47	0.52
ExpressJet Airlines Inc.	39819	15.72	15.58	15.86
Virgin America	4797	1.89	1.84	1.95
US Airways Inc.	16750	6.61	6.52	6.71
American Airlines Inc.	26302	10.38	10.26	10.50
United Air Lines Inc.	46267	18.26	18.11	18.42
Southwest Airlines Co.	11902	4.70	4.62	4.78
JetBlue Airways	44479	17.56	17.41	17.71
Delta Air Lines Inc.	41683	16.45	16.31	16.60
Endeavor Air Inc.	NA	NA	NA	NA

Table 4: **Table 4.** *Change in proportion with 95% CI and in number of flights by airlines between the first 10 months of 2013 and 2014.* Each row shows percent with 95% CI and number of change for one airline. Rows are sorted by the change in percent.

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

Among the airlines from part “a”, Delta Air Lines Inc. showed the largest increase in percent and Envoy Air showed the largest decrease in percent. Southwest Airlines Co. showed the largest increase in the number of flights and ExpressJet Airlines Inc. showed the largest decrease in number. We can see that some airlines like JetBlue Airways, showed an increase in the percent of flights but a decrease in the number of flights. This is because the number of total flights in the first ten months of 2014 is smaller than the one of 2013.

c.

Table 5: **Table 5.** *Proportion of flights by airlines in each airport in 2013.* Each row shows the proportion and 95% CI of flights for one combination of the airport and the airline in 2013. Rows are sorted by the airport and the proportion of flights in 2013.

Airport	Airline	2013 proportion (%)	2013 lwr (%)	2013 upr (%)
John F Kennedy Intl	JetBlue Airways	37.81	37.53	38.10

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

Based on the table above, JetBlue Airways was the largest carrier at John F Kennedy Intl airport in 2013. Delta Air Lines Inc. was the largest carrier at La Guardia airport in 2013. United Air Lines Inc. was the largest carrier at Newark Liberty Intl airport in 2013.

Question 3

a.

Table 6: **Table 6.** *Proportion of homes having stucco construction as the major outside wall material by division.* Each row shows the proportion and standard error for one division. Rows are sorted by the proportion.

Division	Proportion (%)	Standard Error (%)
Mountain South	64.25	4.49
Pacific	44.59	1.69
Mountain North	16.61	3.24
South Atlantic	10.62	1.42
West North Central	4.87	2.03
West South Central	2.99	0.69
Middle Atlantic	2.06	0.73
New England	1.23	0.80
East North Central	0.66	0.28
East South Central	0.42	0.41

The percent of homes having stucco construction as the major outside wall material within each division with its standard error is shown above. Mountain South division had the highest proportion, and East South Central division had the lowest division.

b.

Table 7: **Table 7.** *Average total electricity usage in kilowatt hours by division.* Each row shows the average total electricity usage and its standard error for one division. Rows are sorted by the average total electricity usage.

Division	Average Total Electricity Usage	Standard Error
East South Central	14535.97	620.63
West South Central	14324.26	423.04
South Atlantic	13446.62	276.88
West North Central	10523.83	453.62
Mountain South	10442.02	1271.40
East North Central	9128.66	203.63
Middle Atlantic	8465.44	201.10
Mountain North	8384.47	644.43
Pacific	8100.41	178.60
New England	7514.56	532.02

The average total electricity usage in kilowatt hours in each division with its standard error is shown above. East South Central had the largest average total electricity usage, while New England had the smallest.

Table 8: **Table 8.** *Average total electricity usage in kilowatt hours stratified by urban and rural status.* Each row shows the average total electricity usage and its standard error for one status. Rows are sorted by the average total electricity usage.

Status	Average Total Electricity Usage	Standard Error
Rural	14164.87	390.05
Urban	9867.52	130.23

The average total electricity usage in kilowatt hours stratified by urban and rural status with its standard error is shown above. Rural place had larger average total electricity usage.

c.

Table 9: **Table 9.** *Disparity between urban and rural areas in terms of the proportion of homes with internet access by division.* Each row shows the proportion disparity and its standard error by division. Rows are sorted by the proportion disparity.

Divison	Disparity (%)	Standard Error (%)
Mountain South	18.52	5.78
East South Central	9.33	5.47
West North Central	7.68	5.17
Mountain North	5.50	5.96
West South Central	5.10	3.77
Pacific	3.43	4.05
South Atlantic	3.26	3.47
New England	1.78	2.18
East North Central	0.04	2.73
Middle Atlantic	-1.95	3.64

From the output above, we can see that Mountain South division has the largest disparity between urban and rural areas in terms of the proportion of homes with internet access.