

Solutions for Session 1

Mark Lunt

04/10/2022


```

. do solution.do

. clear

2.1

. global mydir P:/Documents/statacourse

. capture mkdir $mydir

. cd $mydir
P:
Documents
statacourse

. !del /Q *.*

2.2

. sysuse auto
(1978 Automobile Data)

. save $mydir/auto
file P:/Documents/statacourse/auto.dta saved

. dir
<dir> 10/04/22  9:48  .
<dir>  9/30/22 13:01 ..
11.9k 10/04/22  9:48 auto.dta

. clear

. use auto
(1978 Automobile Data)

. save myauto
file myauto.dta saved

2.3.1

. generate wtkg = weight/2.2046

. label variable wtkg "Weight (kg)"

2.3.2

. generate short = 0

. replace short = 1 if length < 190
(36 real changes made)

. generate short2 = (length < 190)

```

```
. tab short short2
```

| short | short2 | | Total |
|-------|--------|----|-------|
| | 0 | 1 | |
| 0 | 38 | 0 | 38 |
| 1 | 0 | 36 | 36 |
| Total | 38 | 36 | 74 |

2.3.3

```
. egen wtt = cut(weight), group(3)
```

```
. label variable wtt "Tertiles of weight"
```

```
. tab wtt
```

| Tertiles of weight | Freq. | Percent | Cum. |
|--------------------|-------|---------|--------|
| 0 | 24 | 32.43 | 32.43 |
| 1 | 25 | 33.78 | 66.22 |
| 2 | 25 | 33.78 | 100.00 |
| Total | 74 | 100.00 | |

```
. label def tertiles 0 "Lowest tertile" 1 "Middle tertile" 2 "Highest tertile"
```

```
. label values wtt tertiles
```

```
. tab wtt
```

| Tertiles of weight | Freq. | Percent | Cum. |
|--------------------|-------|---------|--------|
| Lowest tertile | 24 | 32.43 | 32.43 |
| Middle tertile | 25 | 33.78 | 66.22 |
| Highest tertile | 25 | 33.78 | 100.00 |
| Total | 74 | 100.00 | |

. tab make

| Make and Model | Freq. | Percent | Cum. |
|-------------------|-------|---------|--------|
| AMC Concord | 1 | 1.35 | 1.35 |
| AMC Pacer | 1 | 1.35 | 2.70 |
| AMC Spirit | 1 | 1.35 | 4.05 |
| Audi 5000 | 1 | 1.35 | 5.41 |
| Audi Fox | 1 | 1.35 | 6.76 |
| BMW 320i | 1 | 1.35 | 8.11 |
| Buick Century | 1 | 1.35 | 9.46 |
| Buick Electra | 1 | 1.35 | 10.81 |
| Buick LeSabre | 1 | 1.35 | 12.16 |
| Buick Opel | 1 | 1.35 | 13.51 |
| Buick Regal | 1 | 1.35 | 14.86 |
| Buick Riviera | 1 | 1.35 | 16.22 |
| Buick Skylark | 1 | 1.35 | 17.57 |
| Cad. Deville | 1 | 1.35 | 18.92 |
| Cad. Eldorado | 1 | 1.35 | 20.27 |
| Cad. Seville | 1 | 1.35 | 21.62 |
| Chev. Chevette | 1 | 1.35 | 22.97 |
| Chev. Impala | 1 | 1.35 | 24.32 |
| Chev. Malibu | 1 | 1.35 | 25.68 |
| Chev. Monte Carlo | 1 | 1.35 | 27.03 |
| Chev. Monza | 1 | 1.35 | 28.38 |
| Chev. Nova | 1 | 1.35 | 29.73 |
| Datsun 200 | 1 | 1.35 | 31.08 |
| Datsun 210 | 1 | 1.35 | 32.43 |
| Datsun 510 | 1 | 1.35 | 33.78 |
| Datsun 810 | 1 | 1.35 | 35.14 |
| Dodge Colt | 1 | 1.35 | 36.49 |
| Dodge Diplomat | 1 | 1.35 | 37.84 |
| Dodge Magnum | 1 | 1.35 | 39.19 |
| Dodge St. Regis | 1 | 1.35 | 40.54 |
| Fiat Strada | 1 | 1.35 | 41.89 |
| Ford Fiesta | 1 | 1.35 | 43.24 |
| Ford Mustang | 1 | 1.35 | 44.59 |
| Honda Accord | 1 | 1.35 | 45.95 |
| Honda Civic | 1 | 1.35 | 47.30 |
| Linc. Continental | 1 | 1.35 | 48.65 |
| Linc. Mark V | 1 | 1.35 | 50.00 |
| Linc. Versailles | 1 | 1.35 | 51.35 |
| Mazda GLC | 1 | 1.35 | 52.70 |
| Merc. Bobcat | 1 | 1.35 | 54.05 |
| Merc. Cougar | 1 | 1.35 | 55.41 |
| Merc. Marquis | 1 | 1.35 | 56.76 |
| Merc. Monarch | 1 | 1.35 | 58.11 |
| Merc. XR-7 | 1 | 1.35 | 59.46 |
| Merc. Zephyr | 1 | 1.35 | 60.81 |
| Olds 98 | 1 | 1.35 | 62.16 |
| Olds Cutl Supr | 1 | 1.35 | 63.51 |
| Olds Cutlass | 1 | 1.35 | 64.86 |
| Olds Delta 88 | 1 | 1.35 | 66.22 |
| Olds Omega | 1 | 1.35 | 67.57 |
| Olds Starfire | 1 | 1.35 | 68.92 |
| Olds Toronado | 1 | 1.35 | 70.27 |
| Peugeot 604 | 1 | 1.35 | 71.62 |
| Plym. Arrow | 1 | 1.35 | 72.97 |
| Plym. Champ | 1 | 1.35 | 74.32 |
| Plym. Horizon | 1 | 1.35 | 75.68 |
| Plym. Sapporo | 1 | 1.35 | 77.03 |
| Plym. Volare | 1 | 1.35 | 78.38 |
| Pont. Catalina | 1 | 1.35 | 79.73 |
| Pont. Firebird | 1 | 1.35 | 81.08 |
| Pont. Grand Prix | 1 | 1.35 | 82.43 |
| Pont. Le Mans | 1 | 1.35 | 83.78 |
| Pont. Phoenix | 1 | 1.35 | 85.14 |
| Pont. Sunbird | 1 | 1.35 | 86.49 |
| Renault Le Car | 1 | 1.35 | 87.84 |
| Subaru | 1 | 1.35 | 89.19 |
| Toyota Celica | 1 | 1.35 | 90.54 |
| Toyota Corolla | 1 | 1.35 | 91.89 |
| Toyota Corona | 1 | 1.35 | 93.24 |
| VW Dasher | 1 | 1.35 | 94.59 |
| VW Diesel | 1 | 1.35 | 95.95 |
| VW Rabbit | 1 | 1.35 | 97.30 |
| VW Scirocco | 1 | 1.35 | 98.65 |
| Volvo 260 | 1 | 1.35 | 100.00 |
| Total | 74 | 100.00 | |

2.3.4

```
. gen str20 company = substr(make, 1, index(make, " "))  
(1 missing value generated)
```

```
. tab company
```

| company | Freq. | Percent | Cum. |
|---------|-------|---------|--------|
| AMC | 3 | 4.11 | 4.11 |
| Audi | 2 | 2.74 | 6.85 |
| BMW | 1 | 1.37 | 8.22 |
| Buick | 7 | 9.59 | 17.81 |
| Cad. | 3 | 4.11 | 21.92 |
| Chev. | 6 | 8.22 | 30.14 |
| Datsun | 4 | 5.48 | 35.62 |
| Dodge | 4 | 5.48 | 41.10 |
| Fiat | 1 | 1.37 | 42.47 |
| Ford | 2 | 2.74 | 45.21 |
| Honda | 2 | 2.74 | 47.95 |
| Linc. | 3 | 4.11 | 52.05 |
| Mazda | 1 | 1.37 | 53.42 |
| Merc. | 6 | 8.22 | 61.64 |
| Olds | 7 | 9.59 | 71.23 |
| Peugeot | 1 | 1.37 | 72.60 |
| Plym. | 5 | 6.85 | 79.45 |
| Pont. | 6 | 8.22 | 87.67 |
| Renault | 1 | 1.37 | 89.04 |
| Toyota | 3 | 4.11 | 93.15 |
| VW | 4 | 5.48 | 98.63 |
| Volvo | 1 | 1.37 | 100.00 |
| Total | 73 | 100.00 | |

```
. replace company = make if company == ""  
(1 real change made)
```

```
. save, replace  
file myauto.dta saved
```

2.4

```
. sysuse bplong  
(fictional blood-pressure data)
```

```
. save $mydir/bplong  
file P:/Documents/statacourse/bplong.dta saved
```

```
. preserve
```

```
. keep if when == 1  
(120 observations deleted)
```

```
. save $mydir/bpbefore  
file P:/Documents/statacourse/bpbefore.dta saved
```

```
. restore
```

```
. keep if when == 2  
(120 observations deleted)
```

```
. save "$mydir/bpafter"
file P:/Documents/statacourse/bpafter.dta saved
```

2.4.1

```
. use bpbefore, clear
(fictional blood-pressure data)

. gen fromfile = 1

. append using bpafter
(label when already defined)
(label sex already defined)
(label agegrp already defined)

. replace fromfile = 2 if fromfile == .
(120 real changes made)

. tab fromfile when
```

| fromfile | Status | | Total |
|----------|--------|-------|-------|
| | Before | After | |
| 1 | 120 | 0 | 120 |
| 2 | 0 | 120 | 120 |
| Total | 120 | 120 | 240 |

```
. label variable fromfile "Whether the measurement is a before or after"
. label define fromfile 1 "Before" 2 "After"
. label values fromfile fromfile
```

2.4.2

```
. save mybplong
file mybplong.dta saved

. use bpbefore, clear
(fictional blood-pressure data)

. rename bp bp_before

. save, replace
file bpbefore.dta saved

. use bpafter
(fictional blood-pressure data)

. rename bp bp_after

. save, replace
file bpafter.dta saved

. use bpbefore
(fictional blood-pressure data)

. sort patient

. save, replace
file bpbefore.dta saved
```

```
. use bpafter
(fictional blood-pressure data)
```

```
. sort patient
```

```
. save, replace
file bpafter.dta saved
```

```
. merge 1:1 patient using bpbefore
(label agegrp already defined)
(label sex already defined)
(label when already defined)
```

| Result | # of obs. |
|-------------|-----------------|
| not matched | 0 |
| matched | 120 (_merge==3) |

```
. tab _merge
```

| _merge | Freq. | Percent | Cum. |
|-------------|-------|---------|--------|
| matched (3) | 120 | 100.00 | 100.00 |
| Total | 120 | 100.00 | |

```
. save bpwide
file bpwide.dta saved
```

2.5.1

```
. use myauto
(1978 Automobile Data)
```

```
. gen lengthm = length*0.0254
```

2.5.2

```
. gen heavy = weight > 3000 if weight < .
```

```
. tab heavy
```

| heavy | Freq. | Percent | Cum. |
|-------|-------|---------|--------|
| 0 | 35 | 47.30 | 47.30 |
| 1 | 39 | 52.70 | 100.00 |
| Total | 74 | 100.00 | |

2.5.3

```
. egen wtpt = cut(weight), group(3)
```



```
. tab wtt wtpt
```

| Tertiles of weight | wtpt | | | Total |
|--------------------|------|----|----|-------|
| | 0 | 1 | 2 | |
| Lowest tertile | 24 | 0 | 0 | 24 |
| Middle tertile | 0 | 25 | 0 | 25 |
| Highest tertile | 0 | 0 | 25 | 25 |
| Total | 24 | 25 | 25 | 74 |

2.5.4

```
. egen comp2 = ends(make), head
```

```
. tab comp2
```

| comp2 | Freq. | Percent | Cum. |
|---------|-------|---------|--------|
| AMC | 3 | 4.05 | 4.05 |
| Audi | 2 | 2.70 | 6.76 |
| BMW | 1 | 1.35 | 8.11 |
| Buick | 7 | 9.46 | 17.57 |
| Cad. | 3 | 4.05 | 21.62 |
| Chev. | 6 | 8.11 | 29.73 |
| Datsun | 4 | 5.41 | 35.14 |
| Dodge | 4 | 5.41 | 40.54 |
| Fiat | 1 | 1.35 | 41.89 |
| Ford | 2 | 2.70 | 44.59 |
| Honda | 2 | 2.70 | 47.30 |
| Linc. | 3 | 4.05 | 51.35 |
| Mazda | 1 | 1.35 | 52.70 |
| Merc. | 6 | 8.11 | 60.81 |
| Olds | 7 | 9.46 | 70.27 |
| Peugeot | 1 | 1.35 | 71.62 |
| Plym. | 5 | 6.76 | 78.38 |
| Pont. | 6 | 8.11 | 86.49 |
| Renault | 1 | 1.35 | 87.84 |
| Subaru | 1 | 1.35 | 89.19 |
| Toyota | 3 | 4.05 | 93.24 |
| VW | 4 | 5.41 | 98.65 |
| Volvo | 1 | 1.35 | 100.00 |
| Total | 74 | 100.00 | |

```
. save myauto2
file myauto2.dta saved
```

2.5.5

```
. use bpwide
(fictional blood-pressure data)

. gen bpdiff = bp_after - bp_before
```

2.5.6

```
. egen gsex = group(sex agegrp)
```

```

. tab gsex
  group(sex
  agegrp) |
  Freq.   Percent   Cum.
-----+-----+-----
      1      20     16.67    16.67
      2      20     16.67    33.33
      3      20     16.67    50.00
      4      20     16.67    66.67
      5      20     16.67    83.33
      6      20     16.67   100.00
-----+-----+-----
    Total      120    100.00
end of do-file

```

```

. cd $basedir/stats/1_Intro_to_Stata
P:
Documents
home
teaching
stats
1_Intro_to_Stata

```