Solutions for Session 2: Summarising Data

07/11/2023

```
. do solution.do

. global datadir http://personalpages.manchester.ac.uk/staff/mark.lunt/stats

. global datadir $datadir/2_summarizing_data/data

. use $datadir/pimax.dta, clear

. sort pimax

. browse
request ignored because of batch mode
```

1.1. Median = 95
1.2 Lower quartile = 75
Upper quartile = 110

```
. sort id

. gen sum = sum(pimax)

. browse
request ignored because of batch mode

. gen n = _n

. gen mean = sum / n

. browse
request ignored because of batch mode
```

1.3 Mean = 92.6

```
. drop mean

. egen mean = mean(pimax)

. gen diff = pimax - mean

. gen diff2 = diff*diff

. gen diff2_sum = sum(diff2)

. gen variance = diff2_sum / n

. gen sd = sqrt(variance)
```

1.4 SD = 24.41

```
. use "$datadir/htwt.dta", clear

. histogram bmi
(bin=20, start=17.344378, width=1.4061145)

. graph export bmi.eps replace
(file bmi.eps written in EPS format)
```

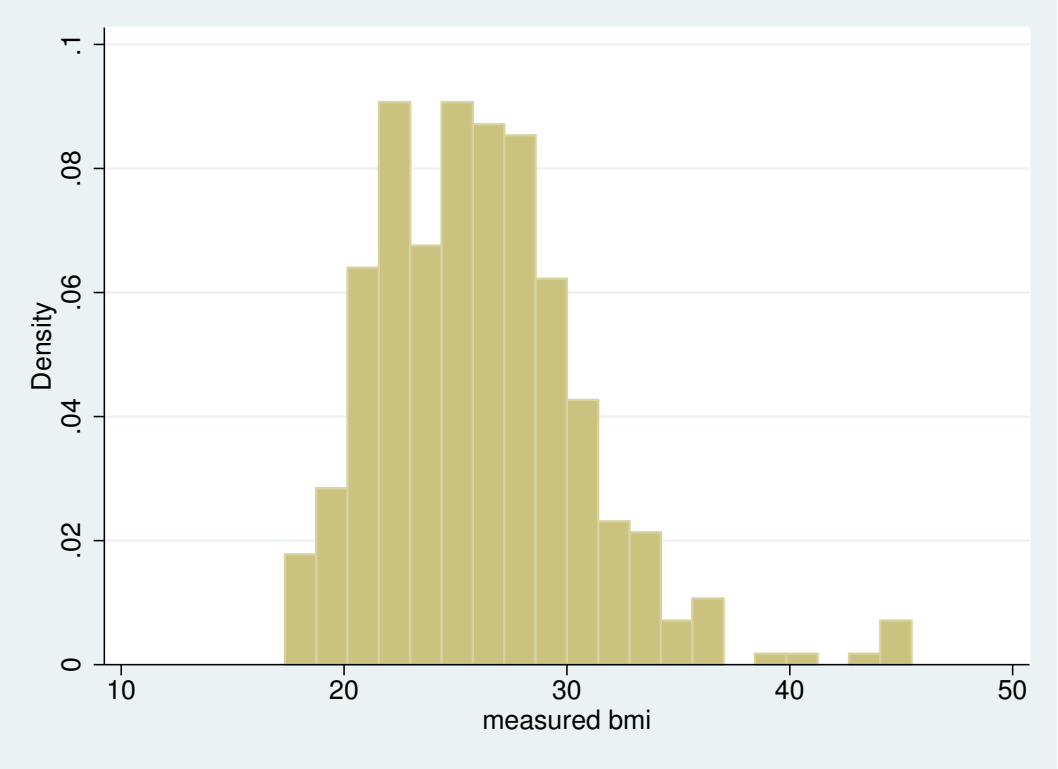

Figure 0.1: . histogram bmi

2.1 There appear to be some very high values of BMI: the distribution is not symmetrical

```
. summarize bmi, det
```

measured bmi

Percentiles		Smallest		
1%	17.9212	17.34438		
5%	19.813	17.36379		
10%	20.77459	17.67052	Obs	400
25%	22.64491	17.88321	Sum of Wgt.	400
50%	25.71006		Mean	26.07662
		Largest	Std. Dev.	4.597912
75%	28.6683	44.11156		
90%	31.45446	44.28855	Variance	21.14079
95%	33.80441	44.92188	Skewness	.9914068
99%	43.88211	45.46667	Kurtosis	5.12302

3.2 Mean BMI = 26.1
3.3 The mean is slightly higher than the median, as you would expect from the skewness
3.4 p25 = 22.6, p75 = 28.7

```
. sort sex
```

```
. by sex: summ bmi, det
```

```
-> sex = female
                        measured bmi
```

	Percentiles	Smallest		
1%	17.67052	17.34438		
5%	19.66385	17.36379		
10%	20.70313	17.67052	Obs	225
25%	22.25057	17.88321	Sum of Wgt.	225
50%	25.08112		Mean	25.85984
		Largest	Std. Dev.	4.903072
75%	28.65985	43.65266		
90%	31.48201	44.11156	Variance	24.04011
95%	34.53231	44.28855	Skewness	1.161285
99%	44.11156	45.46667	Kurtosis	5.203729

```
-> sex = male
                        measured bmi
```

	Percentiles	Smallest		
1%	18.11375	17.95918		
5%	19.81836	18.11375		
10%	20.89796	18.29623	Obs	175
25%	23.54788	18.89285	Sum of Wgt.	175
50%	26.25182		Mean	26.35534
		Largest	Std. Dev.	4.170245
75%	28.71692	35.64129		
90%	31.42691	35.69304	Variance	17.39094
95%	33.24788	40.69628	Skewness	.6977607
99%	40.69628	44.92188	Kurtosis	4.862646

```
. graph box bmi, by(sex)


. graph export bmi_by_sex.eps replace
(file bmi_by_sex.eps written in EPS format)
```

4.2 The median BMI and the lower quartile are both slightly higher in males
However, the upper quartile and range are very similar in men and women

```
. tabstat nurseht nursewt, by(sex) statistics(mean sd)
Summary statistics: mean, sd
  by categories of: sex (sex)
```

sex	nurseht	nursewt
female	159.774	65.86416
	6.398803	12.7513
male	172.9571	78.8125
	6.911771	12.2367
Total	165.5129	71.53308
	9.307717	14.06939

5.1 Females, height: mean 159.8 cm, SD 6.4cm
weight: mean 65.9 kg, SD 12.8kg
Males, height: mean 173.0 cm, SD 6.9cm
weight: mean 78.8 kg, SD 12.2kg

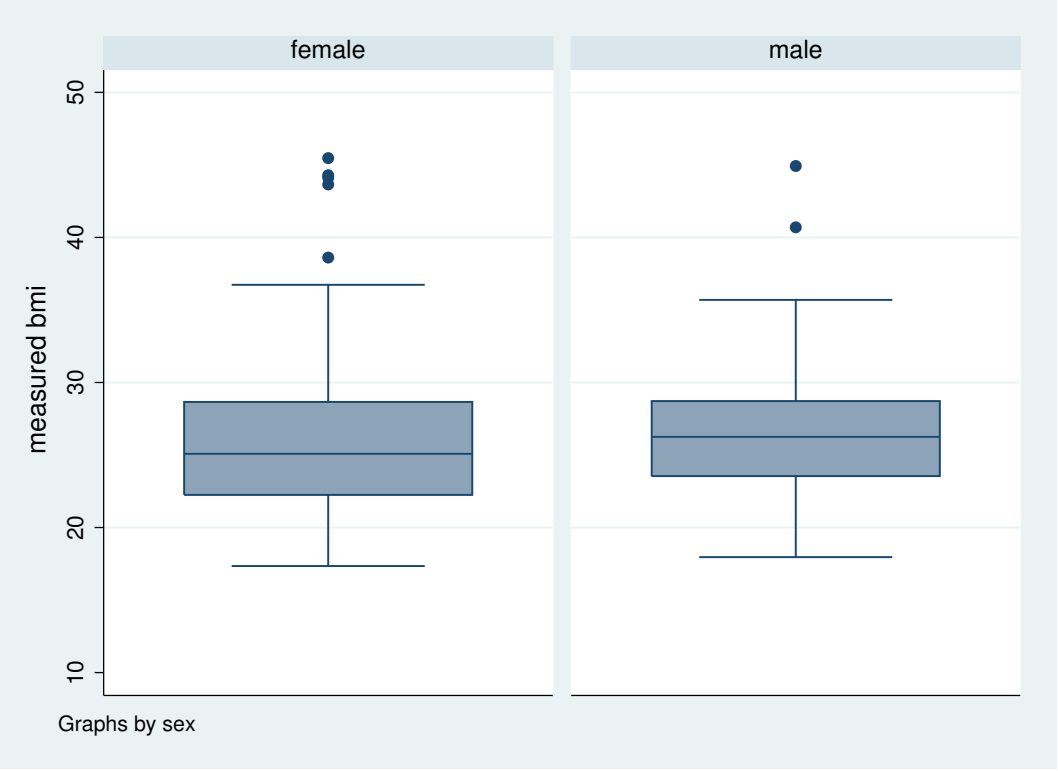

Figure 0.2: . graph box bmi, by(sex)

```
. table sex, c(mean nurseht sd nurseht mean nursewt sd nursewt)
```

sex	mean(nurseht)	sd(nurseht)	mean(nursewt)	sd(nursewt)
female	159.774	6.398803	65.86416	12.7513
male	172.9571	6.911771	78.8125	12.2367

6.1 Should be the same as 6.1

```
. summarize age
```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	412	48.41262	15.23696	19	76

7.1 48.4

```
. histogram age
(bin=20, start=19, width=2.85)

. graph export age.eps replace
(file age.eps written in EPS format)
```

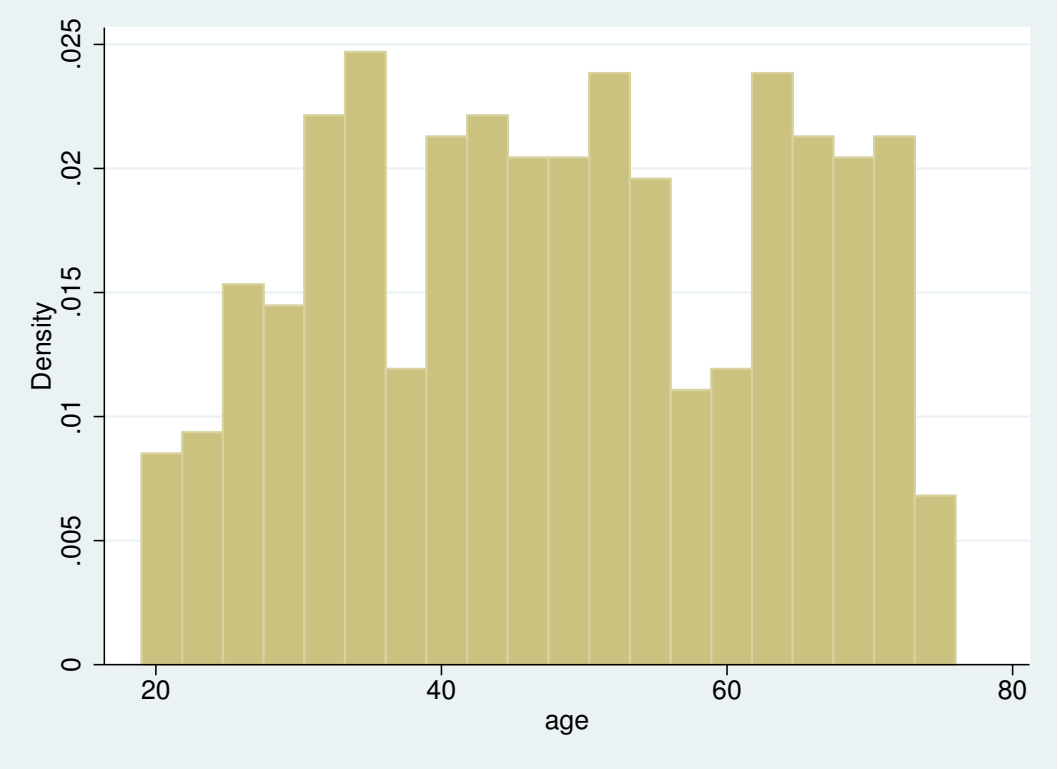

Figure 0.3: . histogram age

7.2 No, the distribution does not decrease at the ends as a normal distribution would

```
. sort sex
```

```
. by sex: summ age, det
```

```
-> sex = female
                            age
-------------------------------------------------------------
      Percentiles      Smallest
 1%           19             19
 5%           24             19
10%           26             19       Obs                 234
25%           35             20       Sum of Wgt.         234
50%           46                      Mean            46.7906
                        Largest       Std. Dev.       15.18791
75%           61             72
90%           68             73       Variance        230.6727
95%           71             73       Skewness        .074776
99%           73             74       Kurtosis        1.841795
```

```
-> sex = male
                            age
-------------------------------------------------------------
      Percentiles      Smallest
 1%           19             19
 5%           25             19
10%           31             20       Obs                 178
25%           39             20       Sum of Wgt.         178
50%           51                      Mean            50.54494
                        Largest       Std. Dev.       15.07948
75%           64             75
90%           71             75       Variance        227.3906
95%           73             75       Skewness       -.1555397
99%           75             76       Kurtosis        2.015873
```

7.3 Males 19 - 76, Females 19-74

```
. summ bmi bmirep
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bmi	400	26.07662	4.597912	17.34438	45.46667
bmirep	406	24.90835	4.084013	15.66737	45.28271

7.4 The mean of the reported BMI is less than the mean of the measured BMI

```
. gen bmidiff = bmi - bmirep
(18 missing values generated)

. summ bmidiff
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bmidiff	394	1.091431	1.845408	-9.373114	7.106318

7.5 Mean = 1.1, SD = 1.8

```
. histogram nurseht, by(sex)
```

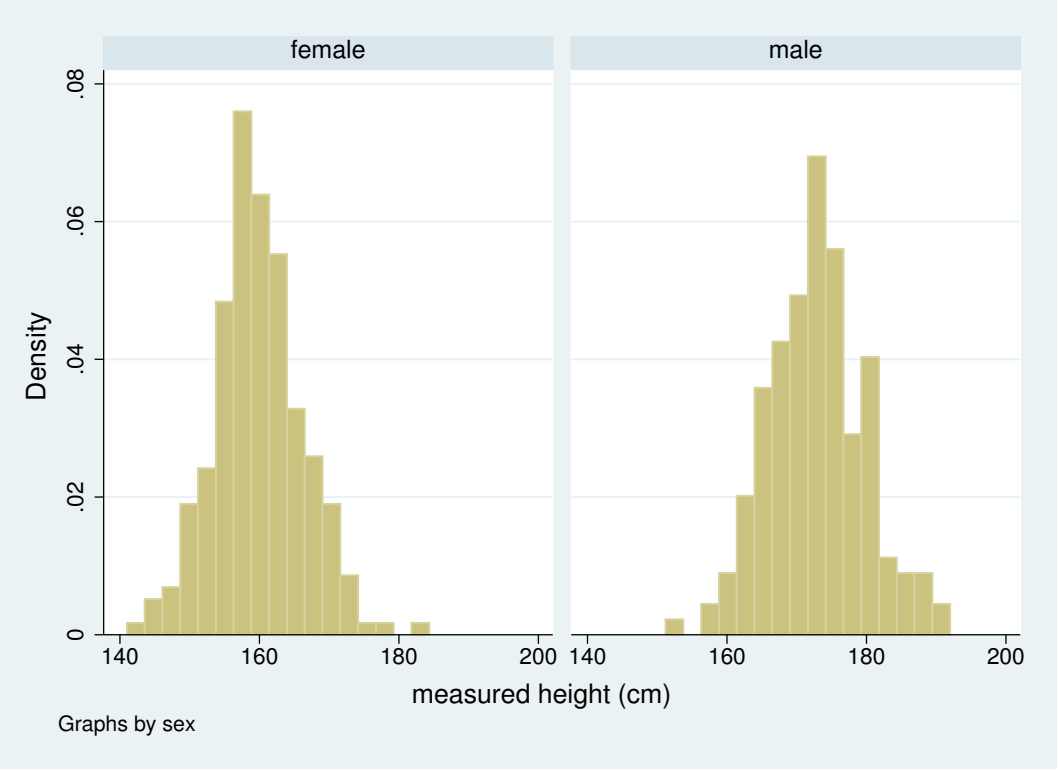

Figure 0.4: . histogram nurseht, by(sex)

```
. graph export nurseht_by_sex.eps replace
(file nurseht_by_sex.eps written in EPS format)

. histogram nurseht, by(sex) normal


. graph export nurseht_by_sexn.eps replace
(file nurseht_by_sexn.eps written in EPS format)

. histogram nursewt, by(sex)


. graph export nursewt_by_sex.eps replace
(file nursewt_by_sex.eps written in EPS format)

. histogram nursewt, by(sex) normal


. graph export nursewt_by_sexn.eps replace
(file nursewt_by_sexn.eps written in EPS format)
end of do-file
```

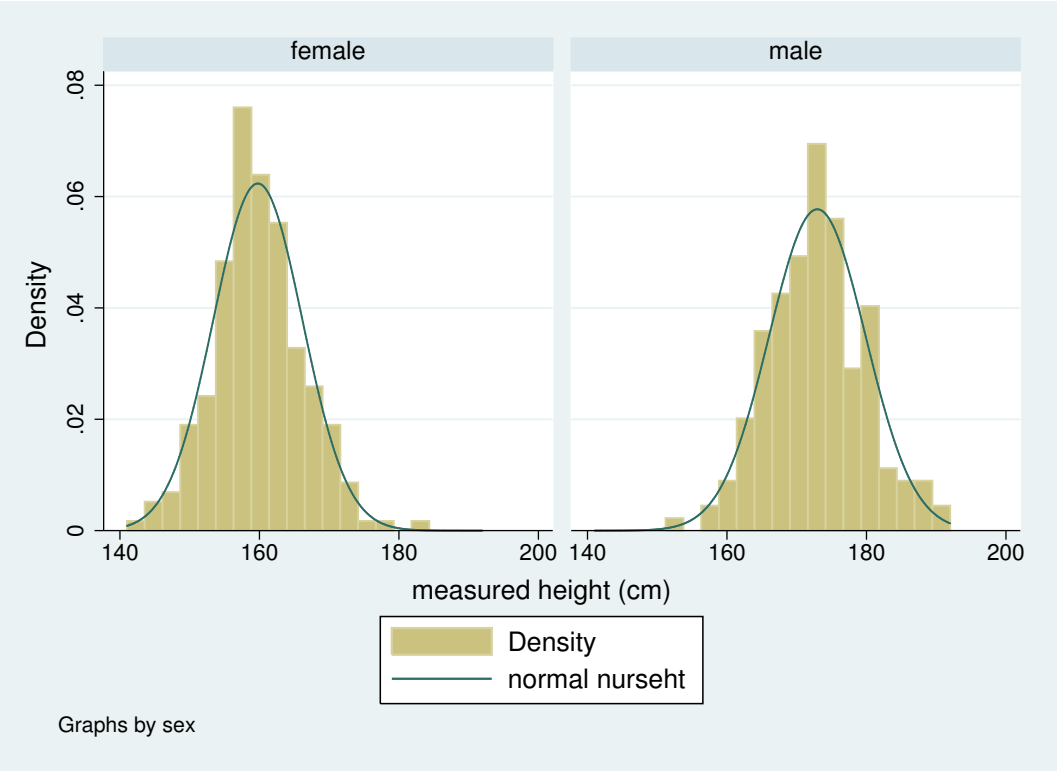

Figure 0.5: . histogram nurseht, by(sex) normal

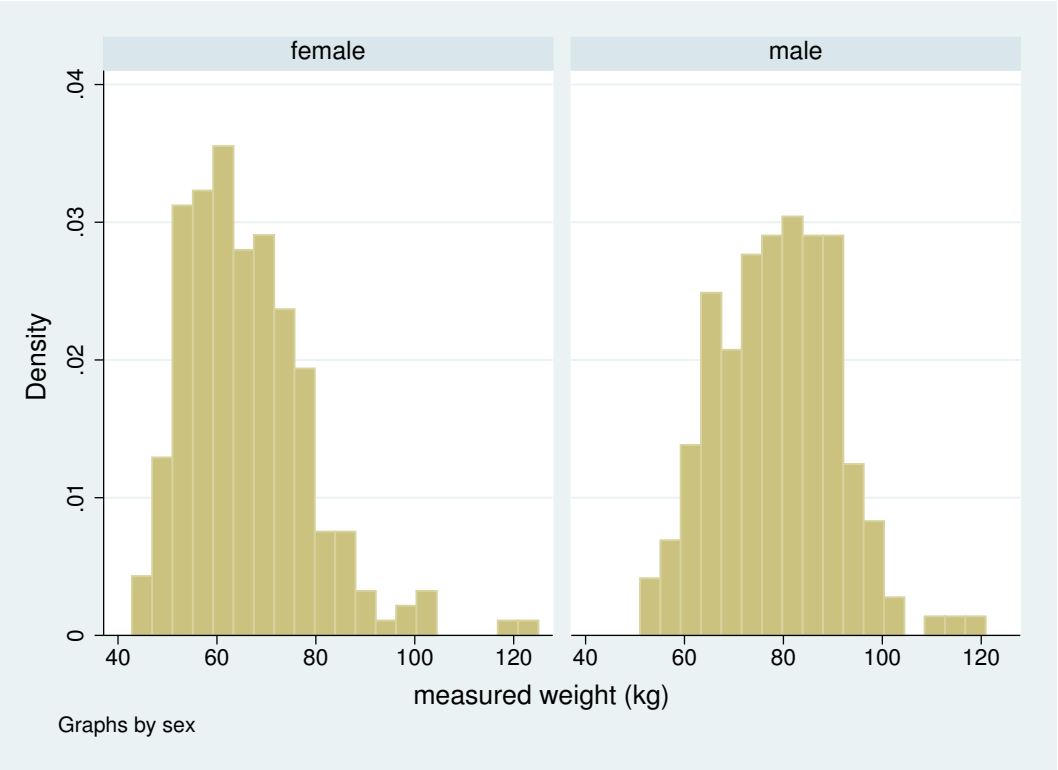

Figure 0.6: . histogram nursewt, by(sex)

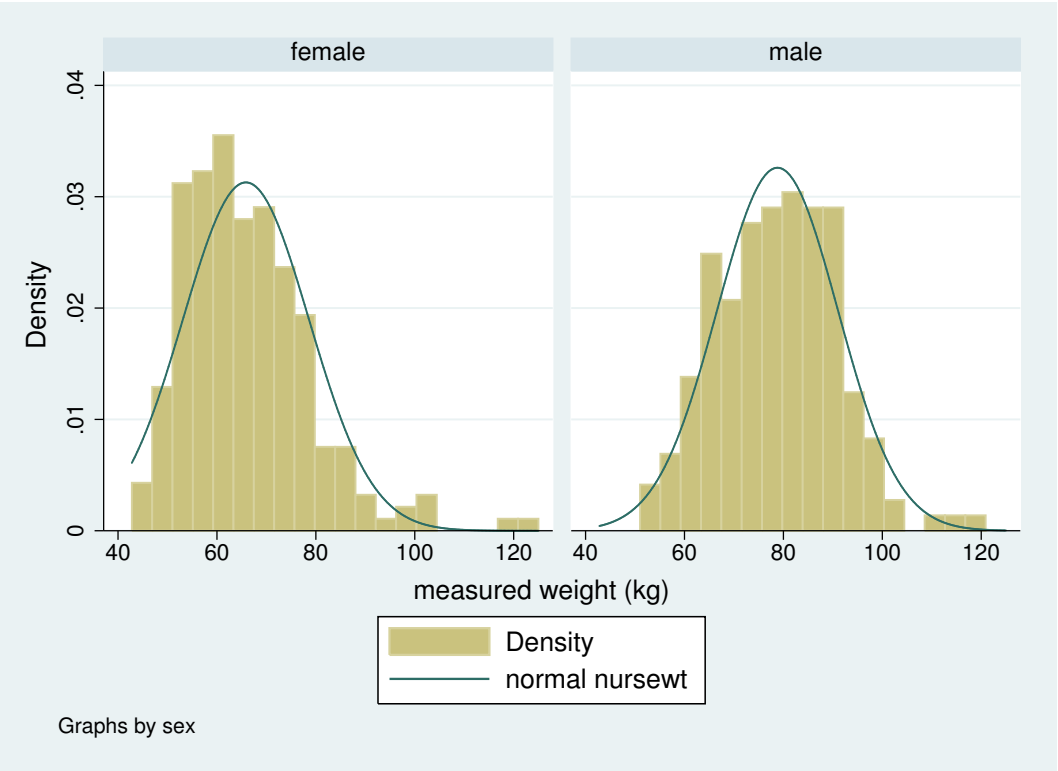

Figure 0.7: . histogram nursewt, by(sex) normal