

Computer Exercise 3. Probability Distributions (Answers)

1. Read in the data set population.dta from the course web site

The data contain observations on the age of the UK adult population, gender and

Summarise the distribution of the age variable (using the summarise, detail command in stata) to obtain the unconditional mean and variance of the distribution of age)

Now draw the pdf of age using the “histogram” command in stata - making sure the bandwidth is equal to one year, (type “help histogram” in stata if you need)

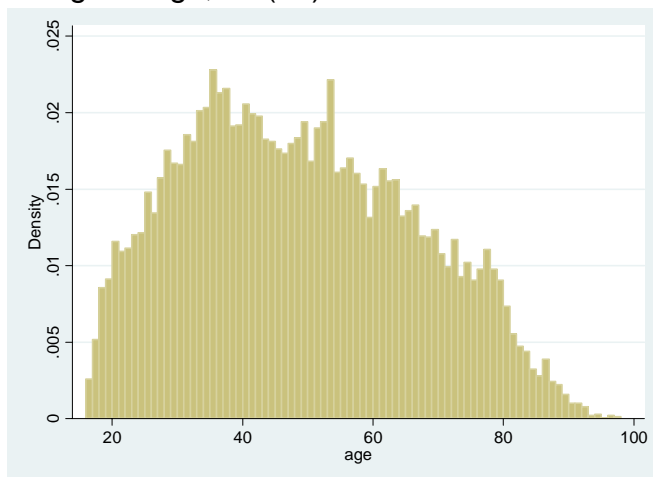
```
su age, det
```

```
-----+-----
```

Percentiles		Smallest		
1%	18	16		
5%	22	16		
10%	26	16	Obs	13909
25%	34	16	Sum of Wgt.	13909
50%		47	Mean	48.28636
		Largest	Std. Dev.	17.58655
75%	62	96		
90%	74	96	Variance	309.2869
95%	79	97	Skewness	.25892
99%	86	98	Kurtosis	2.192978

So $E(\text{Age}) = 48$ and $\text{Var}(\text{Age}) = 309$

```
histogram age, bin(82)
```



(note the number of bins makes the width of the bars equal to one)

Now repeat for each gender to obtain the conditional mean variance and pdf (conditional on gender)

```
. su age if sex==1, det
```

age				

Percentiles		Smallest		
1%	18	16		
5%	21	16		
10%	25	16	Obs	6739
25%	34	16	Sum of Wgt.	6739
50%	47		Mean	48.01469
		Largest	Std. Dev.	17.37704
75%	61	93		
90%	73	95	Variance	301.9613
95%	78	96	Skewness	.2220531
99%	85	97	Kurtosis	2.19032

```
. su age if sex==2, det
```

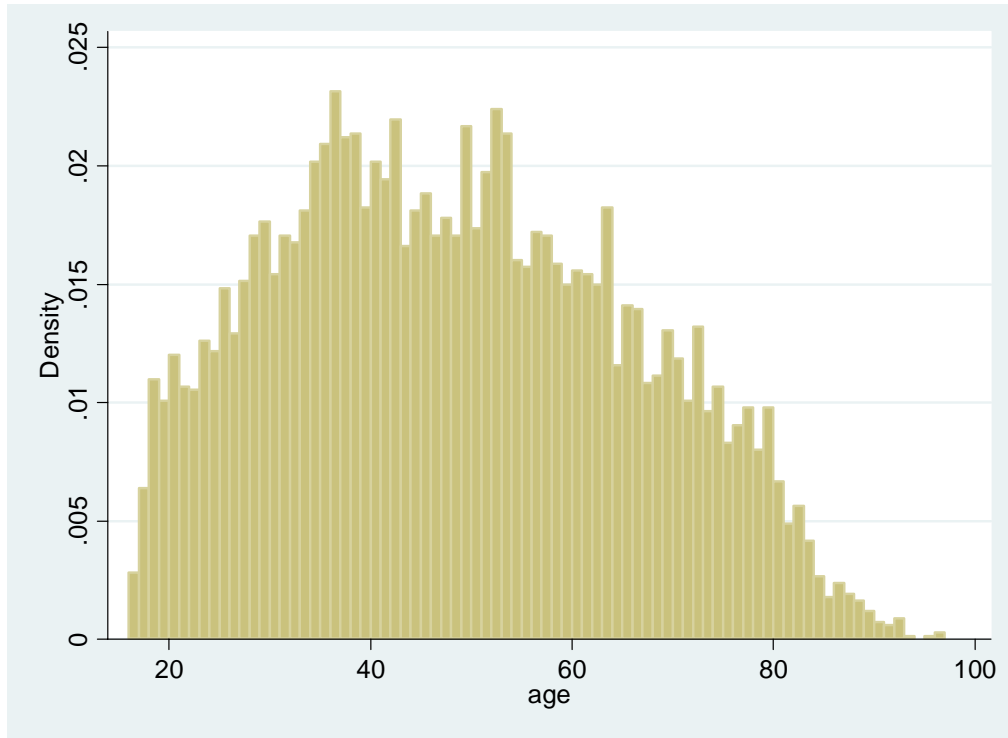
age				

Percentiles		Smallest		
1%	18	16		
5%	22	16		
10%	26	16	Obs	7170
25%	34	16	Sum of Wgt.	7170
50%	47		Mean	48.5417
		Largest	Std. Dev.	17.77866
75%	62	94		
90%	74.5	96	Variance	316.0806
95%	79	96	Skewness	.2894185
99%	87	98	Kurtosis	2.187752

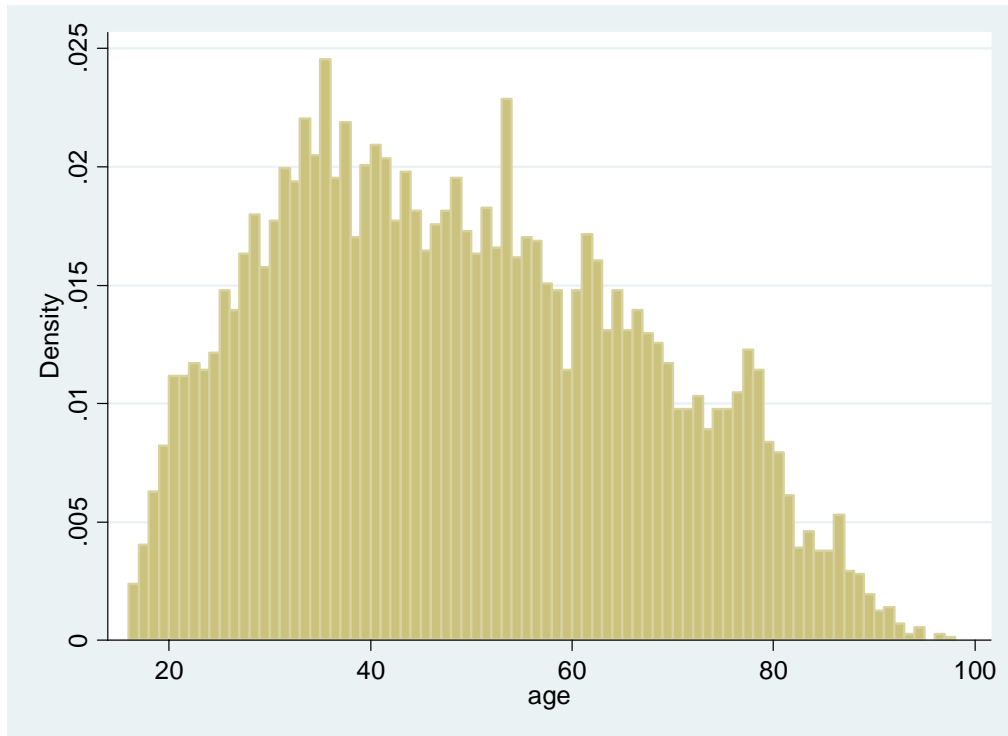
Can see $E(\text{Age}/\text{Female}) > E(\text{Age}/\text{Male})$

and $\text{Var}(\text{Age}/\text{Female}) > \text{Var}(\text{Age}/\text{Male})$

Men
histogram age if sex==1, bin(81)



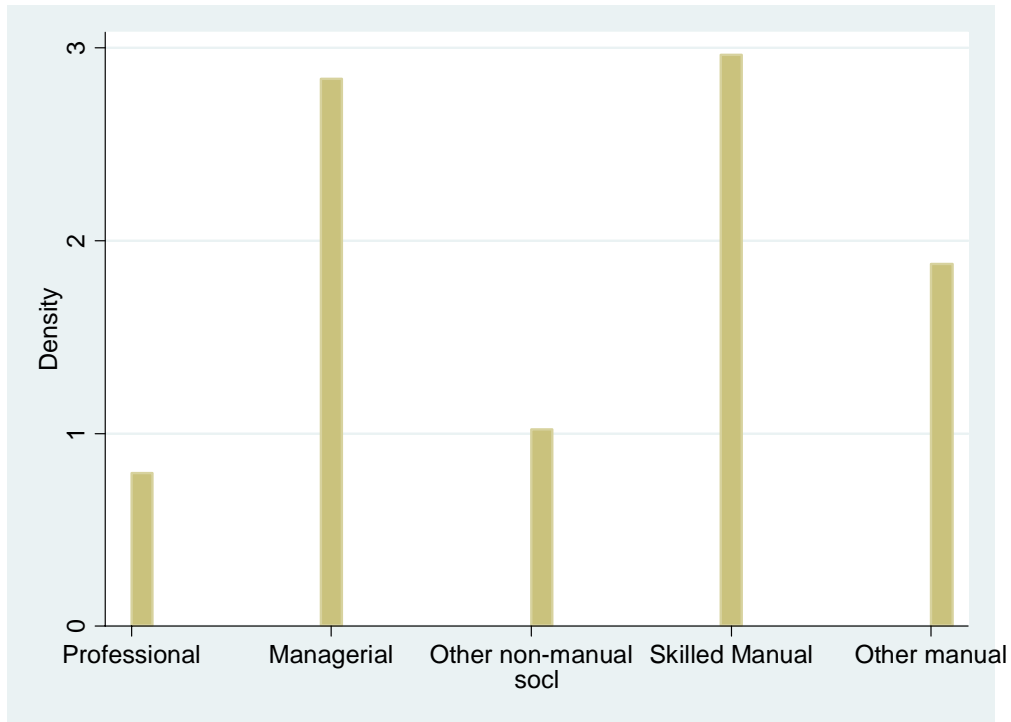
Women
histogram age if sex==2, bin(85)



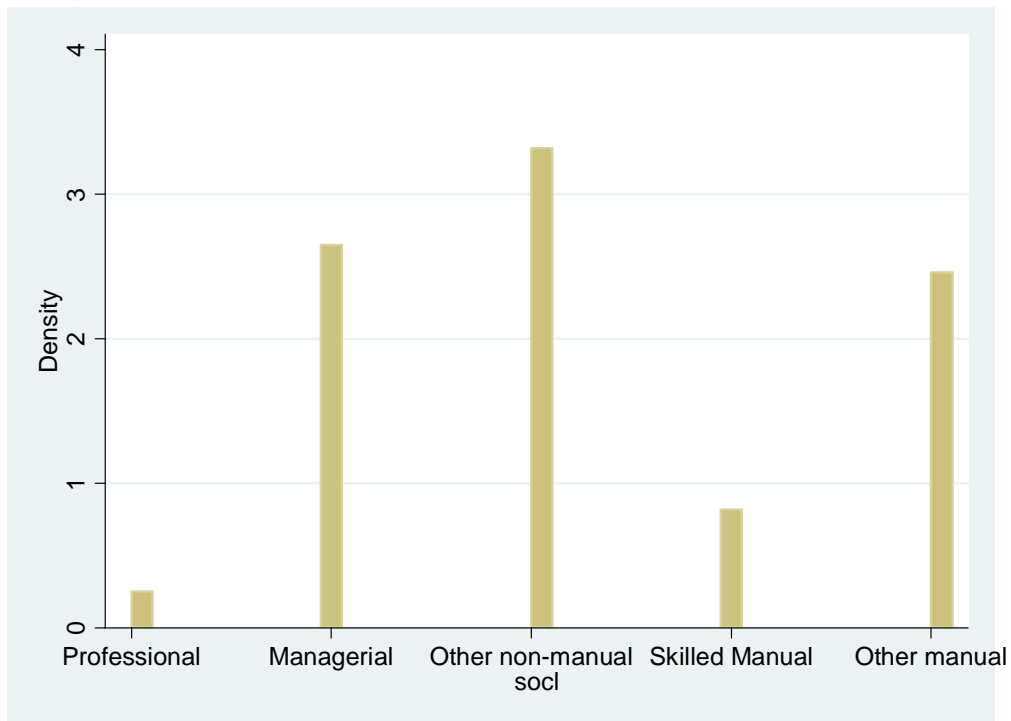
Can see fatter right tail for women and greater concentration of women around 30

Now look at the discrete distribution of the variable social class. In particular, find the mean, variance and conditional mean and variances by gender

histogram soclasad if sex==1, xlabel(, valuelabel)



histogram soclasad if sex==1, xlabel(, valuelabel)



Can see more women in other-non-manual (and correspondingly less in professional, managerial and skilled manual)

. su soclas, det

		socl			
Percentiles		Smallest			
1%	1	1			
5%	1	1			
10%	2	1	Obs	13909	
25%	2	1	Sum of Wgt.	13909	
50%		3	Mean	3.257387	
		Largest	Std. Dev.	1.245947	
75%	4	5			
90%	5	5	Variance	1.552384	
95%	5	5	Skewness	.0478603	
99%	5	5	Kurtosis	1.786439	

. su soclas if sex==1, det

		socl			
Percentiles		Smallest			
1%	1	1			
5%	1	1			
10%	2	1	Obs	6739	
25%	2	1	Sum of Wgt.	6739	
50%		4	Mean	3.24143	
		Largest	Std. Dev.	1.295939	
75%	4	5			
90%	5	5	Variance	1.679459	
95%	5	5	Skewness	-.1393928	
99%	5	5	Kurtosis	1.709189	

. su soclas if sex==2, det

		socl			
Percentiles		Smallest			
1%	1	1			
5%	2	1			
10%	2	1	Obs	7170	
25%	2	1	Sum of Wgt.	7170	
50%		3	Mean	3.272385	
		Largest	Std. Dev.	1.196955	
75%	5	5			
90%	5	5	Variance	1.432701	
95%	5	5	Skewness	.2774789	
99%	5	5	Kurtosis	1.836898	

.

Need to think whether conditional mean of a discrete variable is a meaningful number.

Now write a simple “do-loop” to calculate the mean age of each group and hence the overall mean for the sample

```

local i= 1
while `i'<=2 {

su age if sex==`i'
g mean`i'=r(mean)
g n`i'=r(N)

local i=`i'+1

}

g meanage=(( n1*mean1)+(n2*mean2) )/ (n1+n2)

su meanage age

```

Variable	Obs	Mean	Std. Dev.	Min	Max
age	6739	48.01469	17.37704	16	97
Variable	Obs	Mean	Std. Dev.	Min	Max
age	7170	48.5417	17.77866	16	98

```

. g meanage=(( n1*mean1)+(n2*mean2) )/ (n1+n2)

```

```

. su meanage age

```

Variable	Obs	Mean	Std. Dev.	Min	Max
meanage	13909	48.28636	0	48.28636	48.28636
age	13909	48.28636	17.58655	16	98