

## Male and Female Drivers

The following data are from the National Highway and Traffic Safety Institute, and concern the gender, age and percentage of automobile crashes experienced by U.S. drivers recently. We will use R to analyze this data to answer a few interesting questions.

Age Group	Number of Male licensed Drivers (1000's)	Number of crashes by Males (1000's)	Number of Female licensed Drivers (1000s)	Number of crashes by Females (1000's)
16	816	244	764	178
17	1198	233	1115	175
18	1342	243	1212	164
19	1454	229	1333	145
20-24	7866	951	7394	618
25-29	9356	899	8946	595
30-34	10121	875	9871	571
35-39	10521	901	10439	566
40-44	9776	692	9752	455
45-49	8754	667	8710	390
50-54	6840	390	6763	247
55-59	5341	290	5258	165
60-64	4565	218	4486	133
65-69	4234	191	4231	121
70-74	3604	167	3749	104
75-7	2563	118	2716	77
80-84	1400	61	1516	45
>=85	767	34	767	20
<b>TOTAL</b>	<b>90518</b>	<b>7403</b>	<b>89022</b>	<b>4769</b>

**1** Using proportions, determine which age group(s) are involved in the most crashes, by using relative frequency histograms, box plots, or other graphs, as well as determining relevant statistics (tables of crash %'s by age group, etc.)

**2** Using proportions, determine which gender is involved in the most crashes, using the same graphical procedures and statistics and tables as used above.

**3** Using proportions, determine which age group(s) are involved in the most crashes, by gender.

**4** Find a useful, appropriate linear model (either using transformations or not), relating CRASHES (the response variable) to AGE GROUP (the explanatory variable). Use the median age for the age of each age group interval. Be sure

to include scatter plots (properly labeled), residual plots, statistical values of regression parameters,  $r$  and  $r^2$  values, etc. to justify the appropriateness of your model.

**[5]** Find a useful, appropriate linear model relating CRASHES (the response variable) to AGE GROUP (the explanatory variable), but make separate models for Males and Females. Again, include proper graphs, labeled correctly, and statistics appropriate to a regression study.

**[6]** Compare the two models for Males and Females. Specifically, compare slopes of each model and determine who has more crashes (by proportion), and by how much. Show appropriate graphs if and where necessary.

**[7]** Conjecture why your results show what they show. You may have some theories yourself, or use information from the internet, as you desire. Answer questions like, why do Males/Females have the most crashes?, why do young/old people have more crashes?, what environmental elements contribute, you conjecture, to the high incidence of crashes for \_\_\_\_ age group?, etc.