Analysis of the Accidents Dataset

Information on over 150,000 road accidents with personal injury in 2010 in Great Britain was obtained from the Government of the UK.

We create a variable, the AccidentIndex defined as the product of the number of vehicles and the number of casualties divided by the accident severity. It represent how bad and complicated the accident is. A large value of AccidentIndex implies a very severe accidents with lots of people involved and the complex casualty. Lower values of AccidentIndex means not severe accidents with fewer people involved and simple casualty.

Based on simulated data, we got the mean of the AccidentIndex is around 1 and the 90^{th} percentile is around 1.5 to 2. And then we found that the standard errors of the mean and the 90^{th} percentile is declines. Therefore, we want to check whether the log(SE) is declines proportionally as the Log(n) increase.

In the plot 1, we can see that there is a linear relationship between log(se of mean) and log(n). Based on the table 1, we can say the slope is roughly -0.5 and it means that SE of mean decreases as a factor of$\sqrt{n}$. And then the log(se of 90^{th} percentile) and log(n) seems not has a linear relationship. If we check the table 1, the estimated slop is around -0.395 but -0.5 is still within its 95% confidence interval. Overall, SE roughly decreases as a factor of$\sqrt{n}$. And for the improvement in precision, we get half of the standard error when n quadrupled.

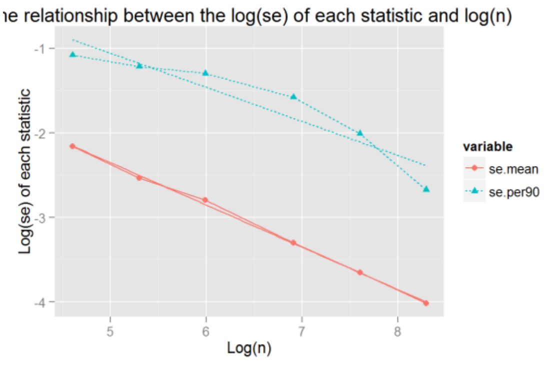

Table 1: The estimated slop and their 95% confidence interval of the relationship between log(se) and log(n)

Log(SE)against Log(n)	Estimated slop	95% LCL	95% UCL
Log(SE.mean)	-0.469	-0.517	-0.422
Log(SE.pre90)	-0.395	-0.602	-0.188

Plot 1: the relationship between log(se) and log(n)