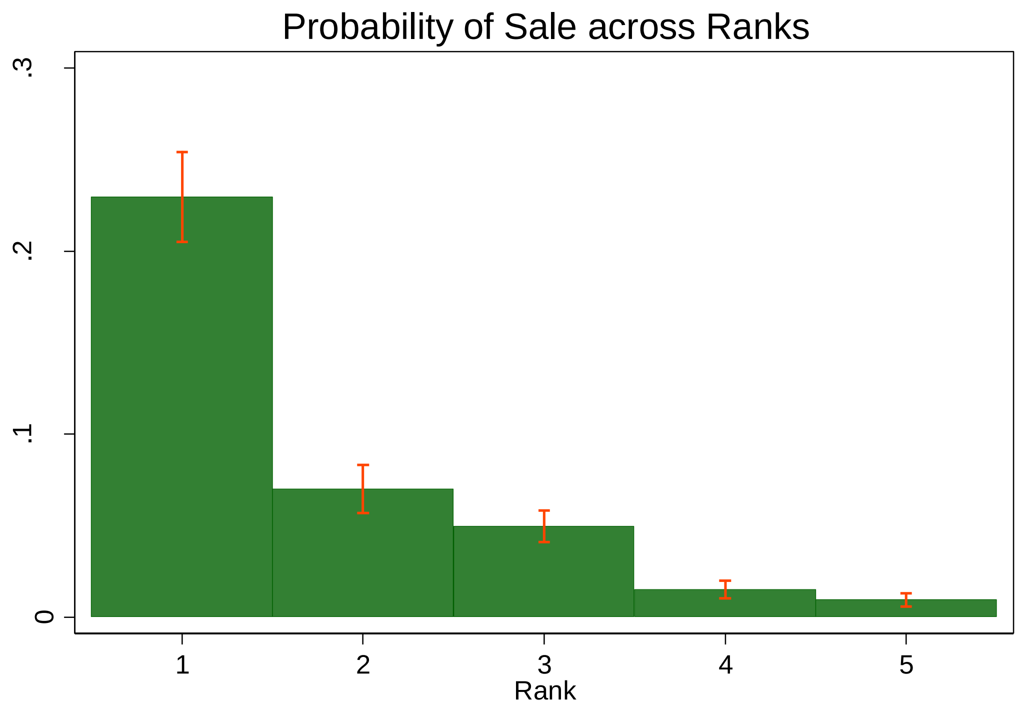
1. Probability of Sale



The difference in the probability of sales is not statistically significant between rank 2 and 3, and between rank 4 and 5. Therefore, a significant increase in the probability of sale happens when rank increases from 4 to 3, and from 2 to 1.

$$Prob(Sale|rank = 2) \approx Prob(Sale|rank = 3)$$

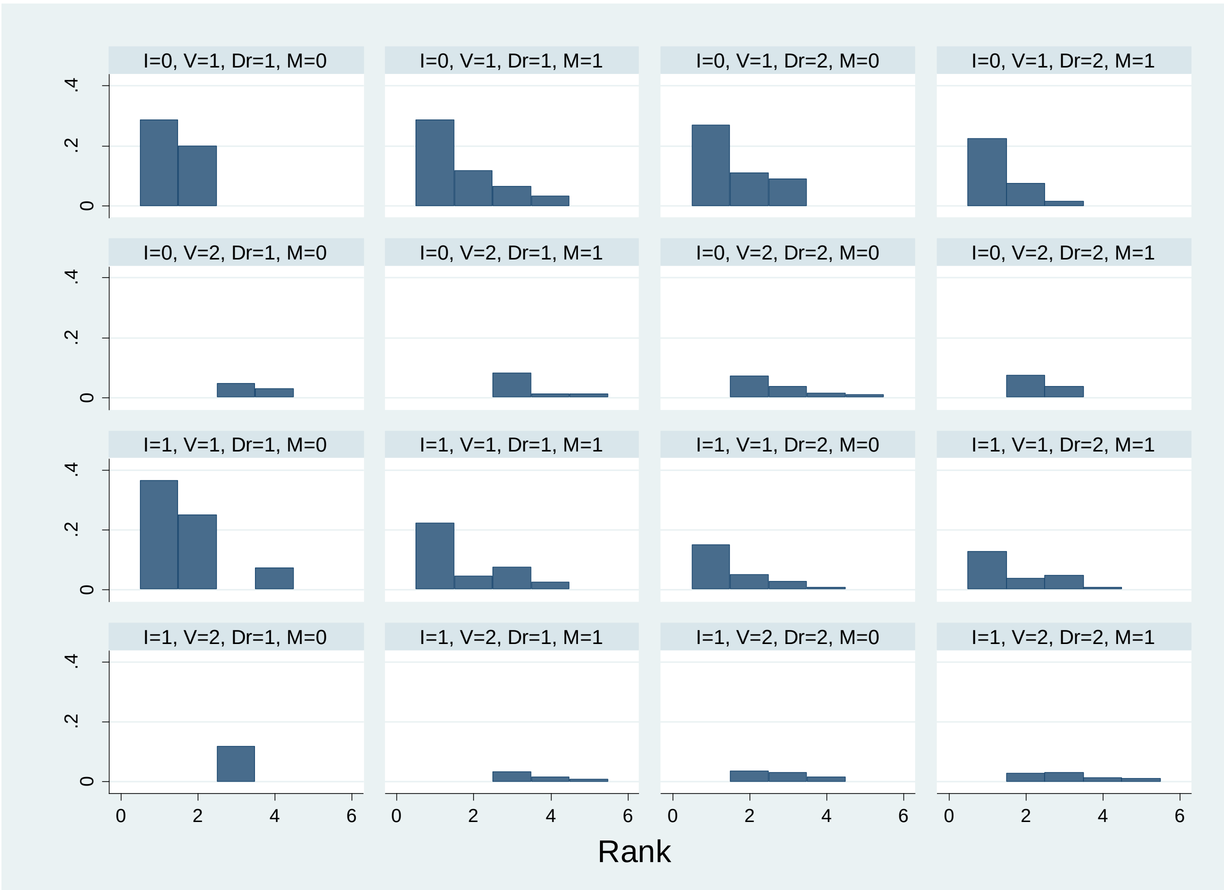
$$Prob(Sale|rank = 4) \approx Prob(Sale|rank = 5)$$

Here are the numbers illustrated in the bar graph above.

	mean	sd	n	Confidence Interval
Prob(Sale|rank=1)	0.229682	0.420814	1132	(0.205142, .2542223)
Prob(Sale|rank=2)	0.070139	0.25547	1440	(0.056933, .0833449)
Prob(Sale|rank=3)	0.049798	0.217571	2470	(0.041213, .05838203)
Prob(Sale|rank=4)	0.015245	0.122552	2427	(0.010367, .02012325)
Prob(Sale|rank=5)	0.009482	0.096934	2531	(0.005704, .01326063)

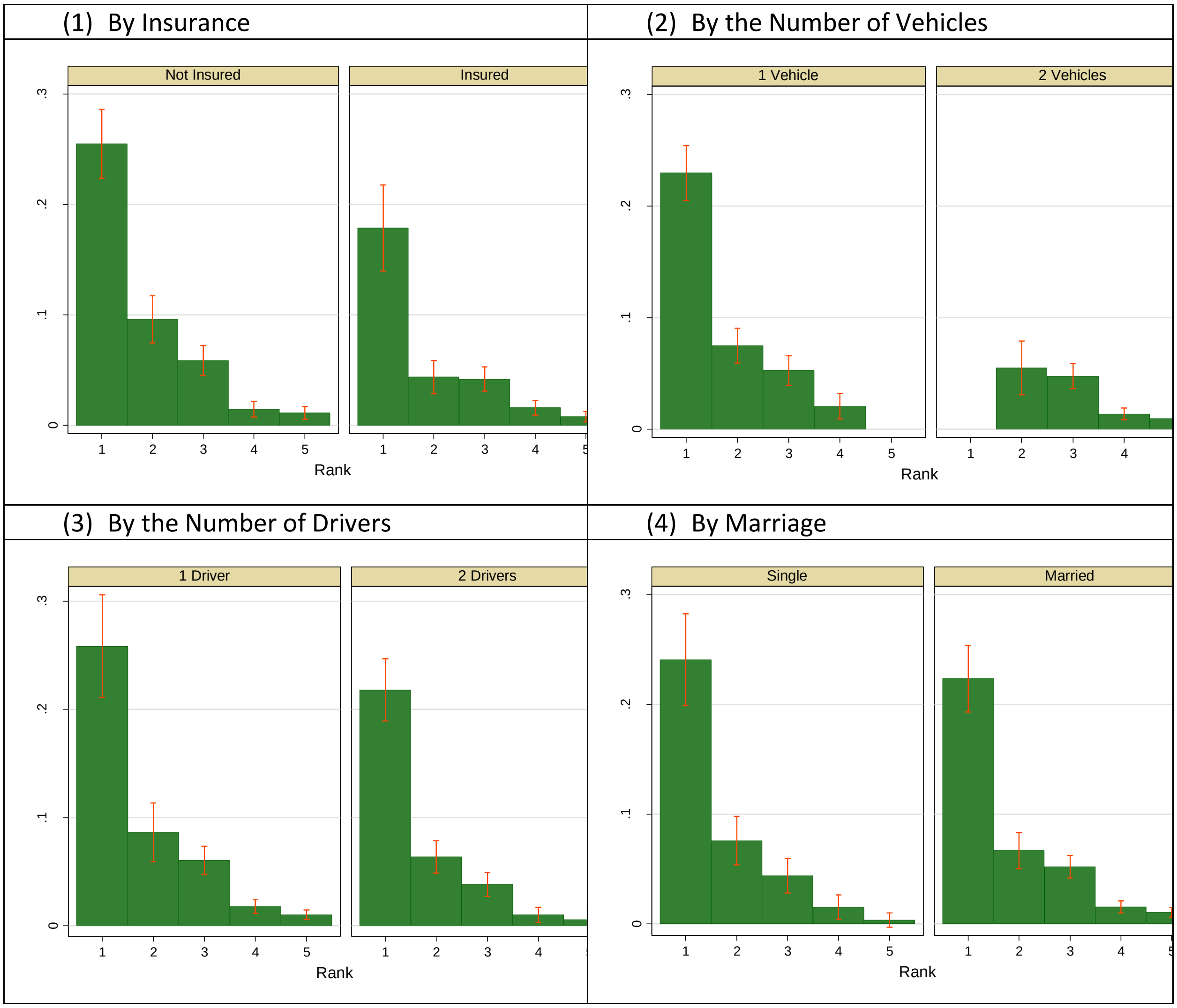
2. Probability of Sale by Group

I visualized the sample distribution of sale by the 16 groups. I think this is not a good way to estimate the distribution because (1) there are a lot of missing ranks, and (2) I cannot see any statistical difference given the small sample in each group.



\

Instead of the 16 groups above, I divided the sample into 2 groups by each characteristic. **My impression is that the probability distribution of sale does NOT vary by the characteristics**. The probability is significantly higher for rank1, there is not significant difference between rank 2 and 3, or rank 4 and 5. Therefore, we should estimate only 1 distribution using the total sample rather than estimate multiple distributions by subgroups.


(1) By Insurance
Not Insured
Insured
(2) By the Number of Vehicles
1 Vehicle
2 Vehicles
(3) By the Number of Drivers
1 Driver
2 Drivers
(4) By Marriage
Single
Married
0
.1
.2
.3
1
2
3
4
5
Rank