

Gus Lipkin

Final Exam Take Home

1. Summary statistics for continuous and categorical variables

Continuous

Variable	Obs	Mean	Std. Dev.	Min	Max
time	151	11.66225	3.257581	6	21
flow	151	493.5762	186.7685	120	985
signals	151	8.900662	3.526575	2	23
distance	151	48.09934	11.19866	29	90
passengers	151	0.1986755	0.5035854	0	3
children	151	0.2980132	0.7098798	0	4
year	151	81.90728	3.988903	65	86

Categorical

route	Freq.	Percent	Cum.		seatbelts	Freq.	Percent	Cum.
1	136	90.07	90.07		0	44	29.14	29.14
2	15	9.93	100		1	107	70.86	100
Total	151	100			Total	151	100	
origin	Freq.	Percent	Cum.		marital	Freq.	Percent	Cum.
0	75	49.67	49.67		0	64	42.38	42.38
1	76	50.33	100		1	87	57.62	100
Total	151	100			Total	151	100	
age	Freq.	Percent	Cum.		income	Freq.	Percent	Cum.
1	11	7.28	7.28		1	55	36.42	36.42
2	63	41.72	49.01		2	52	34.44	70.86
3	54	35.76	84.77		3	28	18.54	89.4
4	10	6.62	91.39		4	8	5.3	94.7
5	13	8.61	100		5	8	5.3	100
Total	151	100			Total	151	100	
gender	Freq.	Percent	Cum.					
0	66	43.71	43.71					
1	85	56.29	100					
Total	151	100						

2. Your best logistic regression model output that includes both the coefficient estimates and the odds ratio

Best

Logistic regression	Number of obs	151				
	LR chi2(8)	89.31				
	Prob > chi2	0				
Log likelihood = -4.2149005	Pseudo R2	0.9137				
route	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
time	-0.5583822	0.3811426	-1.47	0.143	-1.305408	0.1886436
flow	0.0610483	0.0300643	2.03	0.042	0.0021234	0.1199732
signals	3.05003	1.694326	1.8	0.072	-0.2707881	6.370847
passengers	-1.608229	6.441543	-0.25	0.803	-14.23342	11.01696
1.gender	1.308856	2.710426	0.48	0.629	-4.003481	6.621194
1.marital	-2.817294	3.630845	-0.78	0.438	-9.933619	4.299032
children	1.241401	19.3798	0.06	0.949	-36.74232	39.22512
income	0.3664817	1.684526	0.22	0.828	-2.935129	3.668092
_cons	-31.60328	21.22399	-1.49	0.136	-73.20155	9.994982
route	Odds Ratio	Std. Err.			[95% Conf. Interval]	
time	0.5721339	0.2180646			0.2710619	1.20761
flow	1.06295	0.0319568			1.002126	1.127467
signals	21.11597	35.77733			0.7627781	584.5528
passengers	0.200242	1.289867			6.58E-07	60898.43
1.gender	3.701938	10.03383			0.018252	750.8408
1.marital	0.0597675	0.2170064			0.0000485	73.62847
children	3.460458	67.06299			1.10E-16	1.08E+17
income	1.44265	2.430182			0.0531239	39.17709
_cons	1.88E-14	4.00E-13			1.62E-32	21916.21

2nd Best

Logistic regression	Number of obs	151				
	LR chi2(2)	85.06				
	Prob > chi2	0				

Log likelihood = -6.3387136	Pseudo R2	0.8703				
route	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
flow	0.0463908	0.0176995	2.62	0.009	0.0117005	0.0810811
signals	2.440057	0.9797151	2.49	0.013	0.5198512	4.360264
_cons	-30.36408	11.70953	-2.59	0.01	-53.31434	-7.41382
route	Odds Ratio	Std. Err.			[95% Conf.	Interval]
flow	1.047484	0.0185399			1.011769	1.084459
signals	11.4737	11.24096			1.681777	78.27777
_cons	6.50E-14	7.61E-13			7.01E-24	0.0006029

3. Interpretation of the coefficients and odds ratio

- For the second best model, the rural route was 1.04 times more likely to have been chosen if there was higher traffic flow and 11.47 times more likely to have been chosen if there were more traffic signals.

4. Stata code

(R for some processing)

```
```{r}
library(readxl)
library(tidyverse)

df <- read_xlsx("../TakeHome Final/Data_FinalTakeHome.xlsx")
df

df$route[df$route == 2] <- 0
df$income <- df$income - 1

write.csv(df, "../TakeHome Final/cleaned.csv")
```
```

```
capture log close
log using takehome, replace
import excel "/Users/guslipkin/Documents/Fall2020/QMB 3200 ~ Advanced
Quantitative Methods/TakeHome Final/Data_FinalTakeHome.xlsx",
sheet("Sheet1") firstrow clear
```

```
summ time flow signals distance passengers children year
```

```
tabulate route
```

```
tabulate seatbelts
```

```
tabulate age
```

```
tabulate gender
```

```
tabulate marital
```

```
tabulate income
```

```
tabulate origin
```

```
logit route time flow signals passengers i.gender i.marital children  
income
```

```
logit , or
```

```
logit route flow signals
```

```
logit , or
```

```
log close
```