regression

Code ▾

Hide

```
head(data)
```

	economic_stability <int>	education <chr>	education_id <int>	facebook_user_rank <int>	gen... <chr>	ha <c
1	27	Completed High School	1	10	M	NA
2	8	Completed College	2	18	M	NA
3	13	null	0	19	M	NA
4	23	Completed College	2	16	M	NA
5	23	null	0	16	F	NA
6	25	null	0	15	F	NA

6 rows | 1-7 of 20 columns

Hide

```
#multivariate regression
regression_1 <- lm (insurance_segment_id~ income + economic_stability + education_id + f
acebook_user_rank + race_code + youtube_user_rank + gender_dum,  data=data)
summary(regression_1)
```

```
Call:
lm(formula = insurance_segment_id ~ income + economic_stability +
    education_id + facebook_user_rank + race_code + youtube_user_rank +
    gender_dum, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-7.9773 -1.8173  0.4073  1.7124  5.1956

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)         6.448911   0.031061 207.617  < 2e-16 ***
income              0.003212   0.000106  30.315  < 2e-16 ***
economic_stability -0.073079   0.001076 -67.912  < 2e-16 ***
education_id        0.058476   0.007382   7.921 2.37e-15 ***
facebook_user_rank -0.054007   0.003367 -16.042  < 2e-16 ***
race_code           0.033800   0.004399   7.684 1.56e-14 ***
youtube_user_rank  -0.028307   0.003360  -8.425  < 2e-16 ***
gender_dum         -0.078893   0.014068  -5.608 2.05e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.208 on 99992 degrees of freedom
Multiple R-squared:  0.112,  Adjusted R-squared:  0.112
F-statistic:  1803 on 7 and 99992 DF,  p-value: < 2.2e-16
```

Strongest correlation with insurance segment

1. Gender
2. Economic stability
3. Education
4. Facebook User Rank
5. Race
6. Youtube User

Hide

```
# chi square for race by segment
options(scipen = 999)
library(gmodels)
CrossTable(data$race, data$insurance_segment_id)
```

```
   Cell Contents
|-----------------------|
|                     N |
| Chi-square contribution |
|         N / Row Total |
|         N / Col Total |
|       N / Table Total |
|-----------------------|


Total Observations in Table:  100000
```

data$race	data$insurance_segment_id 1	2	3	4	5	6	7	8	Row Total
American Indian	41	28	38	27	73	47	58	30	342
	3.979	0.643	6.728	0.010	56.121	1.535	19.847	5.306	
	0.120	0.082	0.111	0.079	0.213	0.137	0.170	0.088	0.003
	0.003	0.004	0.005	0.003	0.008	0.003	0.002	0.005	
	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	
Asian	130	113	126	120	73	319	533	65	1479
	51.704	0.765	2.922	0.259	28.340	23.440	16.709	4.897	
	0.088	0.076	0.085	0.081	0.049	0.216	0.360	0.044	0.015
	0.008	0.016	0.017	0.015	0.008	0.019	0.018	0.011	
	0.001	0.001	0.001	0.001	0.001	0.003	0.005	0.001	
Black	1654	1865	1240	1383	1323	1296	3585	1122	13468
	136.386	887.751	65.750	110.761	7.412	382.586	57.109	151.907	
	0.123	0.138	0.092	0.103	0.098	0.096	0.266	0.083	0.135
	0.101	0.265	0.169	0.179	0.145	0.079	0.119	0.194	
	0.017	0.019	0.012	0.014	0.013	0.013			

| | 0.036 | 0.011 | |

Chinese	98	45	35	37	13	263	361	24	876
	14.281	4.494	13.209	14.024	55.964	97.878	35.185	13.995	
	0.112	0.051	0.040	0.042	0.015	0.300	0.412	0.027	0.009
	0.006	0.006	0.005	0.005	0.001	0.016	0.012	0.004	
	0.001	0.000	0.000	0.000	0.000	0.003	0.004	0.000	

Hispanic	1452	1101	705	993	550	1558	3584	733	10676
	49.359	162.806	7.426	33.384	183.974	22.619	40.238	21.867	
	0.136	0.103	0.066	0.093	0.052	0.146	0.336	0.069	0.107
	0.089	0.156	0.096	0.128	0.060	0.095	0.119	0.127	
	0.015	0.011	0.007	0.010	0.005	0.016	0.036	0.007	

Japanese	31	18	8	10	2	43	47	7	166
	0.549	3.418	1.415	0.635	11.395	8.992	0.195	0.700	
	0.187	0.108	0.048	0.060	0.012	0.259	0.283	0.042	0.002
	0.002	0.003	0.001	0.001	0.000	0.003	0.002	0.001	
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	

Other	6104	1412	1409	1474	1471	5147	8468	1427	26912
	659.809	122.573	159.342	178.718	393.139	116.053	14.341	10.532	
	0.227	0.052	0.052	0.055	0.055	0.191	0.315	0.053	0.269
	0.373	0.201	0.193	0.190	0.161	0.313	0.280	0.247	
	0.061	0.014	0.014	0.015	0.015	0.051	0.085	0.014	

Portuguese	0	0	0	1	0	0	3	0	4
	0.654	0.281	0.293	1.538	0.365	0.658			

```
|     2.659 |     0.231 |           |
|               |     0.000 |     0.000 |     0.000 |     0.250 |     0.000 |     0.000
|     0.750 |     0.000 |     0.000 |
|               |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
|     0.000 |     0.000 |           |
|               |     0.000 |     0.000 |     0.000 |     0.000 |     0.000 |     0.000
|     0.000 |     0.000 |           |
---------------|-----------|-----------|-----------|-----------|-----------|-----------
|-----------|-----------|-----------|
         White |     6840 |     2455 |     3756 |     3700 |     5610 |     7788
|    13558 |     2370 |    46077 |
|          63.856 |   191.232 |    43.861 |     4.834 |   473.421 |     5.447
|     9.102 |    32.098 |           |
|           0.148 |     0.053 |     0.082 |     0.080 |     0.122 |     0.169
|     0.294 |     0.051 |     0.461 |
|           0.418 |     0.349 |     0.513 |     0.478 |     0.615 |     0.473
|     0.449 |     0.410 |           |
|           0.068 |     0.025 |     0.038 |     0.037 |     0.056 |     0.078
|     0.136 |     0.024 |           |
---------------|-----------|-----------|-----------|-----------|-----------|-----------
|-----------|-----------|-----------|
  Column Total |    16350 |     7037 |     7317 |     7745 |     9115 |    16461
|    30197 |     5778 |   100000 |
|           0.164 |     0.070 |     0.073 |     0.077 |     0.091 |     0.165
|     0.302 |     0.058 |           |
---------------|-----------|-----------|-----------|-----------|-----------|-----------
|-----------|-----------|-----------|
```