ACADGILD ASSIGNMENT 11.2

SESSION 11: Linear Models

1. Use the link given below and locate the bank marketing dataset. https://archive.ics.uci.edu/ml/machine-learning-databases/00222/

a. Is there any association between job and default?

Answer:

with(bankdata,chisq.test(job, default))

with(bankdata, table(job, default))

with(bankdata, prop.table(table(job,default)))

```
> with(bankdata,chisq.test( job, default))

        Pearson's Chi-squared test

data:  job and default
X-squared = 18.401, df = 10, p-value = 0.04857

Warning message:
In chisq.test(job, default) : Chi-squared approximation may be
incorrect
> with(bankdata, table( job, default) )
              default
job               no   yes
  admin.        9129     0
  blue-collar   6088     0
  entrepreneur  1155     0
  housemaid      741     0
  management    2466     0
  retired       1327     0
  self-employed 1139     0
  services      3053     0
  student        771     0
  technician    5765     2
  unemployed     776     1
> with(bankdata, prop.table(table( job,default)))
              default
```

```
job                              no          yes
  admin.           2.816463e-01 0.000000e+00
  blue-collar      1.878259e-01 0.000000e+00
  entrepreneur     3.563385e-02 0.000000e+00
  housemaid        2.286120e-02 0.000000e+00
  management       7.608058e-02 0.000000e+00
  retired          4.094036e-02 0.000000e+00
  self-employed    3.514022e-02 0.000000e+00
  services         9.419060e-02 0.000000e+00
  student          2.378675e-02 0.000000e+00
  technician       1.778607e-01 6.170364e-05
  unemployed       2.394101e-02 3.085182e-05
```

b. Is there any significant difference in duration of last call between? people having housing loan or not?

Answer:

with(bankdata, chisq.test(duration,housing))

with(bankdata, table(duration,housing))

Output from R-Console

```
> with(bankdata, chisq.test(duration,housing))

        Pearson's Chi-squared test

data:  duration and housing
X-squared = 1572.7, df = 1534, p-value = 0.2403

Warning message:
In chisq.test(duration, housing) :
  Chi-squared approximation may be incorrect
> with(bankdata, table( duration,housing) )
        housing
duration no yes
       0   1   3
       1   2   1
       2   1   0
       3   2   1
       4   2  10
       5  16  14
       6  13  24
       7  22  31
       8  27  39
       9  33  42
```

10	36	35
11	34	45
12	24	39
13	44	33
14	25	43
15	34	33
16	35	44
17	34	41
18	43	40
19	24	34
20	29	32
21	30	42
22	35	39
23	21	39
24	30	32
25	25	38
26	23	37
27	31	33
28	25	24
29	31	36
30	17	35
31	28	32
32	20	21
33	19	27
34	31	34
35	34	30
36	42	39
37	32	36
38	26	32
39	32	39
40	26	38
41	33	55
42	35	45
43	31	46
44	33	49
45	28	39
46	37	37
47	25	45
48	43	46
49	49	44
50	41	52
51	41	59
52	48	50
53	49	44
54	46	58
55	48	69
56	51	57
57	41	65
58	44	66
59	53	72
60	47	57
61	49	68
62	49	59

63	55	71
64	63	74
65	57	64
66	57	48
67	65	76
68	68	67
69	63	64
70	59	67
71	76	67
72	63	91
73	74	85
74	64	69
75	65	73
76	67	86
77	56	87
78	56	86
79	71	67
80	69	76
81	69	72
82	57	94
83	59	84
84	66	59
85	85	84
86	55	70
87	70	89
88	64	81
89	68	82
90	73	94
91	75	70
92	72	77
93	76	69
94	66	68
95	63	80
96	86	67
97	79	76
98	70	72
99	55	71
100	61	82
101	62	80
102	64	73
103	68	78
104	70	81
105	62	79
106	67	88
107	60	84
108	62	73
109	66	91
110	71	68
111	72	86
112	61	80
113	62	79
114	74	76
115	58	76

116	62	59
117	64	68
118	49	81
119	72	79
120	52	68
121	61	76
122	69	84
123	69	74
124	65	94
125	67	80
126	62	88
127	68	75
128	70	73
129	64	70
130	62	76
131	69	60
132	56	53
133	57	82
134	74	58
135	78	69
136	77	83
137	61	56
138	51	64
139	87	63
140	64	79
141	52	62
142	57	58
143	63	65
144	64	71
145	59	68
146	54	45
147	56	68
148	51	66
149	51	68
150	49	82
151	56	70
152	49	64
153	47	67
154	62	66
155	63	68
156	47	76
157	71	70
158	43	82
159	65	71
160	48	85
161	57	69
162	46	57
163	46	64
164	55	71
165	66	67
166	73	68
167	44	57
168	59	71

169	50	53
170	37	60
171	59	53
172	53	68
173	51	71
174	57	63
175	48	65
176	39	58
177	47	56
178	59	63
179	46	55
180	58	56
181	49	63
182	39	57
183	39	50
184	56	54
185	56	55
186	37	48
187	54	52
188	58	55
189	63	49
190	54	40
191	54	53
192	49	51
193	45	58
194	40	54
195	31	54
196	40	48
197	50	57
198	56	49
199	49	44
200	60	47
201	41	67
202	45	47
203	50	40
204	47	42
205	41	46
206	50	41
207	54	58
208	46	41
209	45	47
210	42	56
211	50	57
212	44	46
213	47	40
214	39	50
215	38	49
216	33	44
217	41	49
218	36	39
219	40	45
220	37	48
221	45	43

222	52	39
223	38	31
224	30	42
225	42	45
226	29	61
227	39	39
228	28	44
229	26	40
230	36	40
231	34	45
232	38	36
233	26	34
234	37	37
235	29	30
236	27	44
237	31	34
238	32	43
239	35	55
240	31	52
241	36	37
242	23	32
243	34	45
244	34	41
245	49	40
246	39	37
247	31	52
248	24	43
249	33	41
250	40	39
251	34	48
252	36	43
253	31	33
254	27	41
255	38	33
256	25	39
257	36	32
258	29	38
259	33	36
260	23	38
261	29	32
262	35	34
263	37	36
264	31	30
265	23	43
266	26	32
267	33	38
268	35	32
269	23	31
270	38	18
271	27	25
272	26	36
273	23	37
274	23	32

275	21	34
276	23	24
277	33	29
278	27	25
279	19	22
280	27	30
281	30	31
282	21	31
283	31	30
284	27	26
285	29	30
286	19	29
287	25	21
288	24	31
289	23	28
290	29	26
291	29	28
292	18	35
293	28	34
294	24	35
295	35	28
296	33	23
297	28	32
298	29	25
299	22	24
300	18	28
301	35	25
302	20	25
303	16	23
304	25	29
305	26	35
306	23	35
307	18	19
308	16	21
309	16	27
310	10	23
311	24	16
312	25	23
313	20	27
314	27	27
315	18	19
316	23	20
317	21	33
318	18	37
319	28	16
320	16	30
321	14	21
322	31	25
323	20	27
324	20	33
325	17	25
326	19	31
327	24	16

328	19	29
329	25	34
330	10	18
331	18	16
332	27	24
333	20	34
334	13	29
335	17	25
336	14	24
337	21	26
338	15	14
339	28	25
340	21	28
341	16	20
342	21	25
343	15	24
344	17	23
345	26	30
346	19	22
347	15	25
348	17	13
349	26	14
350	18	19
351	14	17
352	16	22
353	19	23
354	12	29
355	16	28
356	15	10
357	18	11
358	17	28
359	17	25
360	24	22
361	28	15
362	24	18
363	25	12
364	20	22
365	12	19
366	19	14
367	14	19
368	14	14
369	14	25
370	21	20
371	20	12
372	11	16
373	12	10
374	13	21
375	12	15
376	11	21
377	19	20
378	18	21
379	9	22
380	14	13

381	11	18
382	9	14
383	14	23
384	14	19
385	15	17
386	17	22
387	12	17
388	21	13
389	14	15
390	13	18
391	14	13
392	16	14
393	14	14
394	24	18
395	12	20
396	13	26
397	15	11
398	13	21
399	19	19
400	12	23
401	14	11
402	14	13
403	7	10
404	17	14
405	12	14
406	15	17
407	18	9
408	13	19
409	15	17
410	11	17
411	14	15
412	8	17
413	10	17
414	13	15
415	11	15
416	8	9
417	9	15
418	14	17
419	11	13
420	12	18
421	13	11
422	16	18
423	11	9
424	11	8
425	10	13
426	10	12
427	12	17
428	12	11
429	7	14
430	13	10
431	13	15
432	12	16
433	9	14

434	8	11
435	8	17
436	7	16
437	10	12
438	13	11
439	11	12
440	5	9
441	9	9
442	12	16
443	10	9
444	10	10
445	13	18
446	13	13
447	7	12
448	5	13
449	8	13
450	7	5
451	10	10
452	9	19
453	7	10
454	9	9
455	14	10
456	11	13
457	6	10
458	6	11
459	11	7
460	12	17
461	8	11
462	12	7
463	12	8
464	14	17
465	13	9
466	10	16
467	9	13
468	7	7
469	6	6
470	12	9
471	13	12
472	16	10
473	14	12
474	9	8
475	7	12
476	5	13
477	10	9
478	5	4
479	14	12
480	13	6
481	5	11
482	10	10
483	11	16
484	12	9
485	10	10
486	10	9

```
487  10  11
488   6   8
489  11   8
490   9   8
491   7  13
492   7  10
493  10  10
494   7   8
495   3   8
496   6   5
497   5   5
498  12   9
499   4   4
[ reached getOption("max.print") -- omitted 1044 rows ]
```

c. Is there any association between consumer price index and consumer?

Answer:

with(bankdata, chisq.test(cons.price.idx,cons.conf.idx))

with(bankdata, table(cons.price.idx,cons.conf.idx))

Output from R-console:

```
> with(bankdata, chisq.test(cons.price.idx,cons.conf.idx))

        Pearson's Chi-squared test

data:  cons.price.idx and cons.conf.idx
X-squared = 1029700, df = 625, p-value < 2.2e-16

Warning message:
In chisq.test(cons.price.idx, cons.conf.idx) :
  Chi-squared approximation may be incorrect
> with(bankdata, table(cons.price.idx,cons.conf.idx))
              cons.conf.idx
cons.price.idx -50.8   -50 -49.5 -47.1 -46.2 -45.9 -42.7   -42 -41.8 -40.8
-40.4 -40.3
        92.201     0     0     0     0     0     0     0     0     0     0
0      0
        92.379     0     0     0     0     0     0     0     0     0     0
0      0
        92.431     0     0     0     0     0     0     0     0     0     0
```

cons.price.idx	-40	-39.8	-38.3	-37.5	-36.4	-36.1	-34.8	-34.6	-33.6	-33	-31.4	-30.1
92.469	0	0	0	0	0	0	0	0	0	0	0	0
92.649	0	0	0	0	0	0	0	0	0	0	0	0
92.713	0	0	0	0	0	0	0	0	0	0	0	0
92.756	0	0	0	0	0	10	0	0	0	0	0	0
92.843	0	282	0	0	0	0	0	0	0	0	0	0
92.893	0	0	0	0	5794	0	0	0	0	0	0	0
92.963	0	0	0	0	0	0	0	0	0	715	0	0
93.075	0	0	0	2458	0	0	0	0	0	0	0	0
93.2	0	0	0	0	0	0	0	3616	0	0	0	0
93.369	0	0	0	0	0	0	0	0	0	0	0	0
93.444	0	0	0	0	0	0	0	0	0	0	0	0
93.749	0	0	0	0	0	0	0	0	0	0	0	0
93.798	0	0	0	0	0	0	0	0	0	0	67	0
93.876	0	0	0	0	0	0	0	0	0	0	0	0
93.918	0	0	0	0	0	0	6685	0	0	0	0	0
93.994	0	0	0	0	0	0	0	0	0	0	0	0
94.027	0	0	0	0	0	0	0	0	0	0	0	0
94.055	0	0	0	0	0	0	0	0	0	0	0	0
94.199	0	0	0	0	0	0	0	0	0	0	0	0
94.215	0	0	0	0	0	0	0	0	0	0	0	311
94.465	0	0	0	0	0	0	0	0	4374	0	0	0
94.601	0	0	204	0	0	0	0	0	0	0	0	0
94.767	128	0	0	0	0	0	0	0	0	0	0	0

cons.price.idx	cons.conf.idx -40	-39.8	-38.3	-37.5	-36.4	-36.1	-34.8	-34.6	-33.6	-33	-31.4	-30.1
92.201	0	0	0	0	0	0	0	0	0	0	770	0
92.379	0	0	0	0	0	0	0	0	0	0		

	0	0										
92.431	0	0	0	0	0	0	0	0	0	0	0	0
92.469	0	0	0	0	0	0	0	0	178	0	0	0
92.649	0	0	0	0	0	0	0	0	0	0	0	357
92.713	0	0	0	0	0	0	0	0	0	172	0	0
92.756	0	0	0	0	0	0	0	0	0	0	0	0
92.843	0	0	0	0	0	0	0	0	0	0	0	0
92.893	0	0	0	0	0	0	0	0	0	0	0	0
92.963	0	0	0	0	0	0	0	0	0	0	0	0
93.075	0	0	0	0	0	0	0	0	0	0	0	0
93.2	0	0	0	0	0	0	0	0	0	0	0	0
93.369	0	0	0	0	0	0	264	0	0	0	0	0
93.444	0	0	0	0	0	5175	0	0	0	0	0	0
93.749	0	0	0	0	0	0	0	174	0	0	0	0
93.798	0	0	0	0	0	0	0	0	0	0	0	0
93.876	212	0	0	0	0	0	0	0	0	0	0	0
93.918	0	0	0	0	0	0	0	0	0	0	0	0
93.994	0	0	0	0	7763	0	0	0	0	0	0	0
94.027	0	0	233	0	0	0	0	0	0	0	0	0
94.055	0	229	0	0	0	0	0	0	0	0	0	0
94.199	0	0	0	303	0	0	0	0	0	0	0	0
94.215	0	0	0	0	0	0	0	0	0	0	0	0
94.465	0	0	0	0	0	0	0	0	0	172	0	0
94.601	0	0	0	0	0	0	0	0	0	0	0	0
94.767	0	0	0	0	0	0	0	0	0	0	0	0

	cons.conf.idx	
cons.price.idx	-29.8	-26.9
92.201	0	0
92.379	267	0

```
92.431      0    447
92.469      0     0
92.649      0     0
92.713      0     0
92.756      0     0
92.843      0     0
92.893      0     0
92.963      0     0
93.075      0     0
93.2        0     0
93.369      0     0
93.444      0     0
93.749      0     0
93.798      0     0
93.876      0     0
93.918      0     0
93.994      0     0
94.027      0     0
94.055      0     0
94.199      0     0
94.215      0     0
94.465      0     0
94.601      0     0
94.767      0     0
```

>

d. Is the employment variation rate consistent across Job types?

Answer:

with(bankdata, chisq.test(job,emp.var.rate))

with(bankdata, table(job,emp.var.rate))

Output from R-Console:

```
> with(bankdata, chisq.test( job,emp.var.rate))

        Pearson's Chi-squared test

data:  job and emp.var.rate
X-squared = 4600.8, df = 90, p-value < 2.2e-16

Warning message:
In chisq.test(job, emp.var.rate) :
  Chi-squared approximation may be incorrect
> with(bankdata, table( job,emp.var.rate) )
```

```
             emp.var.rate
job             -3.4   -3 -2.9 -1.8 -1.7 -1.1 -0.2 -0.1  1.1  1.4
  admin.         321   47  562 2231  246  187    3  940 1601 4284
  blue-collar     64    9   99 2519   58   33    3  575 2295 3599
  entrepreneur    24    1   38  306   14    7    0  265  289  512
  housemaid       32    9   41  120   18   16    1   70  229  524
  management      98   12  121  593   47   38    0  522  553  940
  retired        193   33  181  338   96   83    0   72  215  509
  self-employed   40    6   60  287   24   12    0  187  253  552
  services        32    2   88 1040   47   40    0  311  932 1477
  student         62   20  144  311   72   73    0   21   66  106
  technician     145   22  234 1243  110  115    2  575 1060 3237
  unemployed      44    9   76  164   31   28    1  141  171  349

>
```

e. Is the employment variation rate same across Education?

Answer:

with(bankdata, chisq.test(education,emp.var.rate))

with(bankdata, table(education, emp.var.rate))

Output from R-console:

```
> with(bankdata, chisq.test( education,emp.var.rate))

        Pearson's Chi-squared test

data:  education and emp.var.rate
X-squared = 1364.7, df = 54, p-value < 2.2e-16

Warning message:
In chisq.test(education, emp.var.rate) :
  Chi-squared approximation may be incorrect
> with(bankdata, table( education, emp.var.rate) )
                     emp.var.rate
education             -3.4   -3 -2.9 -1.8 -1.7 -1.1 -0.2 -0.1  1.1  1.4
  basic.4y             141   17  106  843   75   59    3  238  993 1701
  basic.6y              36    0   35  584   18    9    0  154  592  864
  basic.9y              69   16  110 1628   53   27    0  504 1428 2210
  high.school          216   36  358 2366  183  143    4  809 1857 3543
  illiterate             0    0    3    3    0    0    0    3    2    7
  professional.course  131   19  196 1041   93  113    3  470  887 2290
  university.degree    411   70  758 2403  301  242    0 1414 1627 4942
```

f. Which group is more confident?

Answer:

library(psych)

pairs.panels(bankdata[,1:6])

pairs.panels(bankdata[,1:9])

summary(bankdata)

Output from R-console

```
> summary(bankdata)
      age                 job
 Min.   :17.00    Length:41188
 1st Qu.:32.00    Class :character
 Median :38.00    Mode  :character
 Mean   :40.02
 3rd Qu.:47.00
 Max.   :98.00
   marital              education
 Length:41188        Length:41188
 Class :character    Class :character
 Mode  :character    Mode  :character



    default              housing
 Length:41188        Length:41188
 Class :character    Class :character
 Mode  :character    Mode  :character



     loan                contact
 Length:41188        Length:41188
 Class :character    Class :character
 Mode  :character    Mode  :character



     month               day_of_week
```

```
                                                      Length:41188        Length:41188
                                                      Class :character    Class :character
                                                      Mode  :character    Mode  :character



    duration          campaign           pdays
 Min.   :   0.0   Min.   : 1.000   Min.   :
0.0
 1st Qu.: 102.0   1st Qu.: 1.000   1st
Qu.:999.0
 Median : 180.0   Median : 2.000   Median
:999.0
 Mean   : 258.3   Mean   : 2.568   Mean
:962.5
 3rd Qu.: 319.0   3rd Qu.: 3.000   3rd
Qu.:999.0
 Max.   :4918.0   Max.   :56.000   Max.
:999.0
    previous        poutcome
 Min.   :0.000   Length:41188
 1st Qu.:0.000   Class :character
 Median :0.000   Mode  :character
 Mean   :0.173
 3rd Qu.:0.000
 Max.   :7.000
  emp.var.rate       cons.price.idx
cons.conf.idx
 Min.   :-3.40000   Min.   :92.20   Min.   :-
50.8
 1st Qu.:-1.80000   1st Qu.:93.08   1st Qu.:-
42.7
 Median : 1.10000   Median :93.75   Median :-
41.8
 Mean   : 0.08189   Mean   :93.58   Mean   :-
40.5
 3rd Qu.: 1.40000   3rd Qu.:93.99   3rd Qu.:-
36.4
 Max.   : 1.40000   Max.   :94.77   Max.   :-
26.9
    euribor3m        nr.employed           y
 Min.   :0.634   Min.   :4964    Length:41188
 1st Qu.:1.344   1st Qu.:5099    Class
:character
 Median :4.857   Median :5191    Mode
:character
 Mean   :3.621   Mean   :5167
 3rd Qu.:4.961   3rd Qu.:5228
 Max.   :5.045   Max.   :5228

>
```