**QUESTION 1 [7 marks]**

| | | Y | M | Old | |
|---|---|---|---|---|---|
| | | **25-44** | **45-64** | **>64** | |
| H | High.Sch. ( | *0.19* | *0.25* | 0.16 | 0.6 |
| U | University ( | 0.25 | *0.09* | *0.06* | 0.4 |
| | | 0.44 | 0.34 | 0.22 | 1 |

a) [2 marks] Based on this information please complete the contingency table

0.5 marks for each correct answer appearing in the table. Students need not detail logic (though are advised to do so on exam).

b) [1 marks] Is a randomly chosen applicant's educational-attainment independent of age? Defend your answer.

Must demonstrate one of the following numerically

P[Y|H] not = P[Y] (or not = P[Y|U])

0.316667 not = 0.44 Thus the answer is no; the two variables are dependent.

Any variant upon the above is acceptable. Make sure that student has calculated the probabilities in question to make her/his case.

c) [1 marks] Are the event "High-school or less" and the event of being in age category 25-44 disjoint? Defend your answer.

If H disjoint from Y then the probability of their joint event (P[Y and H]) must be zero, but since it is in fact 0.19, we have demonstrated that they are NOT disjoint.

Student must point out that joint probability (probability of AND event) is not zero for full marks (0.5 marks for correct answer with no explanation).

d) [1 marks] Consider a randomly chosen applicant who falls in the "University or more" category. What is the probability that they are younger than 64 years of age?

P["not Old" | U] = P["not Old" & U]/P[U] = 0.85

0.5 mark for logic and 0.5 mark for answer.

e) [1 marks] What is the probability that a randomly chosen applicant is either in "45-64" or the "University or more" category?

P[ M or U ] = P[M] + P[U] - P[M and U]

0.65 0.5 mark for logic and 0.5 mark for answer.

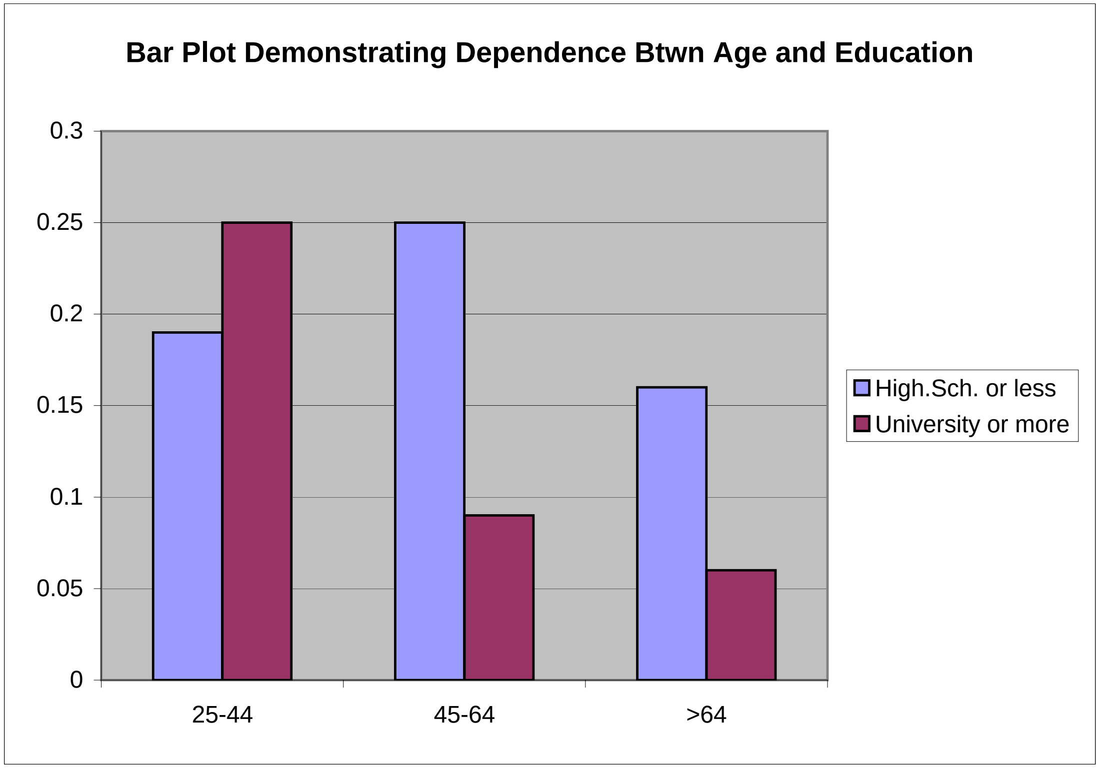
f) [1 marks] You randomly select 5 candidates, what is the probability that at least 1 of these candidates is between the ages of 25 and 64?

P[ single candidate is either Y or M] = 0.78

Let S denote that a single candidate is either Y or M

P[ of at least one S among 5 candidates] = 1 - P[None S's among the 5]
P[None S's among the 5] = P[notS & notS & notS & notS] = (1-P[S])*(1-P[S])*(1-P[S])*(1-P[S])*(1-P[S])
P[notS & notS & notS & notS &no 0.000515
P[ of at least one S among 5 candidates] = 1 0.999485

0.5 mark for logic and 0.5 mark for answer.



Assignment 1, Part II
QUESTION 2

| | Background Check | | | TR = Theft risk |
|---|---|---|---|---|
| | TR | nTR | P[TR \| row-group] | |
| Innocent (I) | 50 | 950 | 0.05 | Note these are separate studies |
| Guilty (G) | 20 | 5 | 0.8 | Note these are separate studies |
| | 70 | 955 | | |
| | | | 1025 | |

Based on the results in Tables 1 and 2

(a)**[1 marks]** What is the probability of being labeled "theft risk" among the innocent?

P[ TR | I ] = 0.0500

(b) **[1 marks] What is the probability of being labeled "theft risk" among the guilty?**

P[ TR | G ] = 0.8000

(c) [1 mark] In order to determine the probability that an applicant would end up committing theft, one might be tempted to observe that a total of 1025 employees were surveyed in the two separate studies outlined above, and to note that 70 out of those 1025 employees were "guilty" thus estimating the probability as 70/1025. This would be incorrect. Explain why

The two studies described above are separate studies and thus can not be combined. While the conditional probabilities that they provide can be reasonably assumed to apply to the general population, the marginal probabilities that one would obtain by naively combining the two studies are meaningless.

(d) [3 marks] Construct probability tree for this problem being sure to annotate it.

See below for Tree

Students are free to use there own notation to define the events but the students definitions must be clear to TA. Please give 1 mark for tree, 1 mark for labelling parts, 1 mark for entering all probabilities. Su

A national database reveals that among the general population of retail job applicants only X out of every 225 will end up committing a theft in their next job ("guilty")

**(e) [1 mark] What is the probability that a randomly chosen applicant would be labeled "theft risk" by a BC ?**

Students may choose to pick the AND event probabilities from their probability tree, without replicating the mathematical expressions. This approach is acceptable provided that they make clear which values they chose from the

| Guilty | Applicants |
|---|---|
| 1 | 250 |

p[ G ] = 0.004 — 0.25 marks

P[TR]=P[ I ]P[TR| I ]+P[G]P[TR|G]

P[TR]=(1-P[ G ])P[TR| I ]+P[G]P[TR|G] — 0.75 marks for logic

where P[TR| I ] and P[TR| G ] were calculated in part (b)

P[TR] = 0.0530 — 0.25 mark for answer

(f) [3 marks] Construct a contingency table for this problem, being sure to specify all the joint and marginal probabilities.

| | TR | nTR | total |
|---|---|---|---|
| G | 0.0032 | 0.0008 | **0.004** |
| I | 0.0498 | 0.9462 | **0.996** |
| **Total** | 0.053 | 0.947 | 1 |

P[G| nTR ] 0.000845

(g) [1.5 marks] Given that a person is labeled "theft risk" what is the probability that they are actually innocent?

P[I|TR] = P[I&TR]/P[TR]  1 mark for logic
P[I|TR] = (P[TR|I]*P[I])/P[TR]

P[I|TR]= **0.9396** 0.5 mark for answer

(h) [0.5 marks] Given that a person is labeled "theft risk" what is the probability that they are actually guilty?

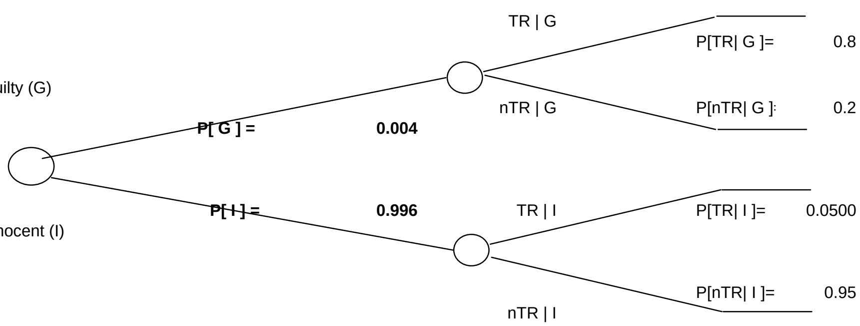
P[G|TR] = 1-P[I|TR]  0.25 mark for logic
P[G|TR] = **0.0604** 0.25 mark for answer

(i) [2 marks] Your retail company is considering a BC policy, which involves submitting all applicants to a BCA, and screening out all those who the BCA labels as "theft risk." Comment on the advise-ability of this policy (two sentence maximum).

Explanation must mention the high false implication (postive) rate. 94% of those flagged as TR are innocent. Adopting this test would screen these erroneously flagged TR candidates out, preventing them from having a job.

Please give partial credit if students show their work

G = guilty (i.e., perpetrator of theft)
I = innocent, upstanding, employee
"TR" -> flagged as theft risk
"nTR" -> flagged as NON theft risk




**ANDS COLUMNs**

| **and Events** | **Pr[ANDs]** | | Out of 2000 |
|---|---|---|---|
| G & TR | 0.0032 | = P[G]*P[TR \| G] | 6.4 |
| G & nTR | 0.0008 | | 1.6 |
| I & TR | 0.0498 | | 99.6 |
| I & nTR | 0.9462 | | 1892.4 |
| | | | 2000 |