Answers

1.(a) The sensitivity of the test is given as 95%.

(b) The specificity of the test is given as 98%.

(c) The false positive rate is given by

$$P(\text{HIV} - \text{negative}|\text{ Positive test}) = \frac{P(\text{Positive test}|\text{ HIV} - \text{negative})P(\text{HIV} - \text{negative})}{P\text{Positive test})}$$

$$= \frac{P(\text{Positive test}|\text{ HIV} - \text{negative})P(\text{HIV} - \text{negative})}{P(\text{Positive test}|\text{ HIV} - \text{pos})P(\text{HIV} - \text{pos}) + P(\text{Positive test}|\text{ HIV} - \text{neg})P(\text{HIV} - \text{neg})}$$

$$= \frac{(.02)(.998)}{(.95)(.002) + (.02)(.998)}$$

$$= .91;$$

that is, 91% of all positive test results using this blood tests for a sample from the population of U.S. college students would be false positives.

2. Here are the marginal distributions for the variables, obtained using `Stat` → `Tables` → `Tally`.

`Summary Statistics for Discrete Variables`

Employme	Count	Percent	CumPct
1	11	3.04	3.04
2	11	3.04	6.08
3	30	8.29	14.36
4	107	29.56	43.92
5	203	56.08	100.00
N=	362		
*=	2		

Size of	Count	Percent	CumPct
1	8	2.20	2.20
2	8	2.20	4.41
3	45	12.40	16.80
4	142	39.12	55.92
5	160	44.08	100.00
N=	363		
*=	1		

Quality	Count	Percent	CumPct
1	8	2.20	2.20
2	8	2.20	4.41
3	48	13.22	17.63
4	137	37.74	55.37
5	162	44.63	100.00
N=	363		
*=	1		

Career p	Count	Percent	CumPct
1	10	2.75	2.75
2	20	5.49	8.24
3	91	25.00	33.24
4	137	37.64	70.88
5	106	29.12	100.00
N=	364		

Career g	Count	Percent	CumPct
1	19	5.22	5.22
2	49	13.46	18.68
3	119	32.69	51.37
4	108	29.67	81.04
5	69	18.96	100.00
N=	364		

Need for	Count	Percent	CumPct
1	11	3.04	3.04
2	43	11.88	14.92
3	106	29.28	44.20
4	108	29.83	74.03
5	94	25.97	100.00
N=	362		
*=	2		

Entrepre	Count	Percent	CumPct
1	9	2.48	2.48
2	46	12.67	15.15
3	107	29.48	44.63
4	128	35.26	79.89
5	73	20.11	100.00
N=	363		
*=	1		

Quantita	Count	Percent	CumPct
1	25	6.87	6.87
2	57	15.66	22.53
3	119	32.69	55.22
4	112	30.77	85.99
5	51	14.01	100.00
N=	364		

Qualitat	Count	Percent	CumPct
1	11	3.02	3.02
2	22	6.04	9.07
3	86	23.63	32.69
4	154	42.31	75.00
5	91	25.00	100.00
N=	364		

Peer pre	Count	Percent	CumPct
1	146	40.44	40.44
2	85	23.55	63.99
3	87	24.10	88.09
4	35	9.70	97.78
5	8	2.22	100.00
N=	361		
*=	3		

Coursewo	Count	Percent	CumPct
1	80	22.04	22.04
2	100	27.55	49.59
3	118	32.51	82.09
4	47	12.95	95.04
5	18	4.96	100.00
N=	363		
*=	1		

Career e	Count	Percent	CumPct
1	4	1.10	1.10
2	8	2.20	3.30
3	33	9.07	12.36
4	103	28.30	40.66
5	216	59.34	100.00
N=	364		

Parental	Count	Percent	CumPct
1	129	35.54	35.54
2	81	22.31	57.85
3	81	22.31	80.17
4	49	13.50	93.66
5	23	6.34	100.00
N=	363		
*=	1		

The patterns for the first four variables are similar, with opinions being dominated by high levels of importance (4 or 5) attached to job opportunities, salary, quality of life, prestige, and enjoyment. This is all very sensible, of course. Glamour is noticeably less important (perhaps the results would be different at USC or UCLA?), being centered at 3's and 4's. Other measures viewed as moderately important include the need for communications skills, entrepreneurial aspects, and the quantitative nature of the required coursework. Somewhat surprising to me is that the qualitative nature of required coursework was viewed as more important than the quantitative nature; I'm not sure how to interpret that. The students consider themselves self–reliant, with both peer pressure and parental influence being relatively unimportant. The ease of coursework isn't too important, either, so apparently the students don't consider themselves lazy.

Many of the variables are related to each other. Here is a cross–tabulations of opinions on employment opportunities and salaries:

Rows: Employme   Columns: Size of

	1	2	3	4	5	All
1	7	0	2	0	2	11
	63.64	--	18.18	--	18.18	100.00
	87.50	--	4.44	--	1.25	3.04
2	0	3	4	2	2	11
	--	27.27	36.36	18.18	18.18	100.00
	--	37.50	8.89	1.42	1.25	3.04
3	0	2	12	8	8	30
	--	6.67	40.00	26.67	26.67	100.00
	--	25.00	26.67	5.67	5.00	8.29
4	0	2	16	59	30	107
	--	1.87	14.95	55.14	28.04	100.00
	--	25.00	35.56	41.84	18.75	29.56
5	1	1	11	72	118	203
	0.49	0.49	5.42	35.47	58.13	100.00
	12.50	12.50	24.44	51.06	73.75	56.08
All	8	8	45	141	160	362
	2.21	2.21	12.43	38.95	44.20	100.00
	100.00	100.00	100.00	100.00	100.00	100.00

Cell Contents --
Count
% of Row
% of Col

It is apparent that the two variables are related, with higher values of one occurring with higher values of the other. This is also true for prestige and glamour, even though their marginal distributions were somewhat dissimilar:

Rows: Career p   Columns: Career g

	1	2	3	4	5	All
1	8	1	0	0	1	10
	80.00	10.00	--	--	10.00	100.00
	42.11	2.04	--	--	1.45	2.75
2	2	14	2	2	0	20
	10.00	70.00	10.00	10.00	--	100.00
	10.53	28.57	1.68	1.85	--	5.49
3	6	20	53	12	0	91
	6.59	21.98	58.24	13.19	--	100.00
	31.58	40.82	44.54	11.11	--	25.00

```
4          1      14      46      72       4     137
        0.73   10.22   33.58   52.55    2.92  100.00
        5.26   28.57   38.66   66.67    5.80   37.64

5          2       0      18      22      64     106
        1.89      --   16.98   20.75   60.38  100.00
       10.53      --   15.13   20.37   92.75   29.12

All       19      49     119     108      69     364
        5.22   13.46   32.69   29.67   18.96  100.00
      100.00  100.00  100.00  100.00  100.00  100.00

 Cell Contents --
                 Count
                 % of Row
                 % of Col
```

On the other hand, some variables are much less related; for example, entrepreneurial aspects and job opportunities done' have much connection, which might make sense if you consider that entrepreneurs try to make their own jobs anyway.

```
Rows: Employme     Columns: Entrepre

           1       2       3       4       5     All

1          2       2       0       1       5      10
       20.00   20.00      --   10.00   50.00  100.00
       22.22    4.35      --    0.79    6.85    2.77

2          0       2       5       2       2      11
          --   18.18   45.45   18.18   18.18  100.00
          --    4.35    4.67    1.59    2.74    3.05

3          0       3      13       8       6      30
          --   10.00   43.33   26.67   20.00  100.00
          --    6.52   12.15    6.35    8.22    8.31

4          3      11      38      40      15     107
        2.80   10.28   35.51   37.38   14.02  100.00
       33.33   23.91   35.51   31.75   20.55   29.64

5          4      28      51      75      45     203
        1.97   13.79   25.12   36.95   22.17  100.00
       44.44   60.87   47.66   59.52   61.64   56.23

All        9      46     107     126      73     361
        2.49   12.74   29.64   34.90   20.22  100.00
      100.00  100.00  100.00  100.00  100.00  100.00

 Cell Contents --
                 Count
                 % of Row
                 % of Col
```

CHS: "Amniocentesis, blood tests, and Down's syndrome"

The proportion of women who test positive where Down's syndrome is present can be calculated as follows:

$$\begin{aligned}
&P(\text{Positive test result and presence of Down's syndrome}) \\
&\qquad = P(\text{Positive test result} \mid \text{Presence of Down's syndrome}) \\
&\qquad\quad \times P(\text{Presence of Down's syndrome}) \\
&\qquad = (.89)(.003704) \\
&\qquad = .003296.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&P(\text{Positive test result and absence of Down's syndrome}) \\
&\qquad = P(\text{Positive test result} \mid \text{Absence of Down's syndrome}) \\
&\qquad\quad \times P(\text{Absence of Down's syndrome}) \\
&\qquad = (1 - .75)(.996296) \\
&\qquad = .24907.
\end{aligned}$$

Thus, the probability of a positive test result is the sum of .00396 and .24907, or .25237.

The proportion of positive test results that are false positives is determined as in the "Random drug and disease testing" case. Since

$$\begin{aligned}
&P(\text{Presence of Down's syndrome} \mid \text{Positive test result}) \\
&\qquad = \frac{P(\text{Presence of Down's syndrome and positive test result})}{P(\text{Positive test result})} \\
&\qquad = \frac{.003296}{.25237} \\
&\qquad = .01306,
\end{aligned}$$

the required probability is $1 - .01306$, or .98694.

The false negative rate is calculated the same way:

$$\begin{aligned}
&P(\text{Presence of Down's syndrome} \mid \text{Negative test result}) \\
&\qquad = \frac{P(\text{Presence of Down's syndrome and negative test result})}{P(\text{Negative test result})} \\
&\qquad = \frac{.003704 - .003296}{.74763} \\
&\qquad = .000545.
\end{aligned}$$

(3) To calculate this value, we first need to know what we're given (let W represent being a woman, $Y98$ represent being a member of the class of 1998, and $Y99$ represent being a member of the class of 1999):

$$\begin{aligned}
P(W|Y98) &= .37 \\
P(W|Y99) &= .44 \\
P(Y98) &= 269/(269 + 312) = .463 \\
P(Y99) &= 1 - .463 = .537
\end{aligned}$$

We want $P(Y99|W)$, which equals $P(W \text{ and } Y99)/P(W)$. From the information above, we know that

$$P(W \text{ and } Y99) = P(W|Y99)P(Y99) = (.44)(.537) = .236,$$

so all we need is $P(W)$. But

$$\begin{aligned} P(W) &= P(W \text{ and } Y98) + P(W \text{ and } Y99) \\ &= P(W|Y98)P(Y98) + .236 \\ &= (.37)(.463) + .236 \\ &= .171 + .236 \\ &= .407 \end{aligned}$$

Thus, $P(Y99|W) = P(W \text{ and } Y99)/P(W) = .236/.407 = .58$.

(4) To calculate this value, we first need to know what we're given (let W represent the winner of the Best Picture Oscar having the most Oscar wins for that year and N represent the winner of the Best Picture Oscar having the most Oscar nominations):

$$\begin{aligned} P(W) &= .843 \\ P(N) &= .7 \\ P(N|W) &= .78 \end{aligned}$$

We want $P(W|N)$, which equals $P(W \text{ and } N)/P(N)$. From the information above, we know that

$$P(W \text{ and } N) = P(N|W)P(W) = (.78)(.843) = .658,$$

so $P(W|N) = P(W \text{ and } N)/P(N) = .658/.7 = .939$. In fact, *Shakespeare in Love* did win the most Oscars in 1998 (7), as this probability would have suggested. Note, by the way, that these calculations are only relevant if the movie that has the most Oscar nominations wins the Best Picture Oscar; every Best Picture winner since 1982 either led or tied for most nominations except for the 1991 winner, *The Silence of the Lambs* (which, interestingly enough, ended up with the most wins, including wins for Best Director, Best Actor, and Best Actress).