PHP 2500 Introduction to Biostatistics

Problem Set Two Solutions

---

1. (a) A and B is the event "The individual is exposed to high levels of both carbon monoxide and nitrogen dioxide."
   (b) A or B is the event "The individual is exposed to high levels of either carbon monoxide or nitrogen dioxide (or both)."
   (c) The complement of A is the event "The individual is not exposed to high levels of carbon monoxide."
   (d) The events A and B are not mutually exclusive. An individual can be exposed to high levels of both gases.

2. (a) There are ten sample points:
   S={ abc, abd, abe, acd, ace, ade, bcd, bce, bde, cde}

   (b) (i) A = {abc, abd, abe, acd, ace, ade}

   $$P(A) = P(abc) + P(abd) + \ldots + P(ade) = 6(1/10) = 0.6$$

The events B and C also contain 6 points, so their probabilities are also 0.6.

(ii) AB = {abc, abd, abe}, so its probability is
$P(AB) = P(AC) = P(BC) = 3/10$

(iii) Using (i) and (ii) we have

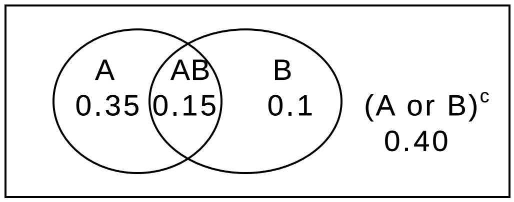
$$P(A \text{ or } B) = P(A) + P(B) - P(AB) = 0.6 + 0.6 - 0.3 = 0.9$$

(Alternatively, you can use brute force: A or B = {abc, abd, abe, acd, ace, ade, bcd, bce, bde} so P(A or B) = 9/10)
Same reasoning applies for P(B or C) and P(A or C)

(iv) $P(ABC) = P(\{abc\}) = 1/10$, $P(A \text{ or } B \text{ or } C) = P(S) = 1$

3. Given is 50% have parasite A, 25% have parasite B, and 40% have neither parasite A nor parasite B. Mathematically, $P(A) = 0.50$, $P(B) = 0.25$, and $P((A \text{ or } B)^c) = 1 - P(A \text{ or } B) = 0.40$.

(a)



(b) $P(A \text{ or } B) = 1 - 0.40 = 0.60$

(c) Since $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, we have $P(AB) = P(A) + P(B) - P(A \text{ or } B) = 0.50 + 0.25 - 0.60 = 0.15$

(c) No, because the intersection of A and B is not empty.

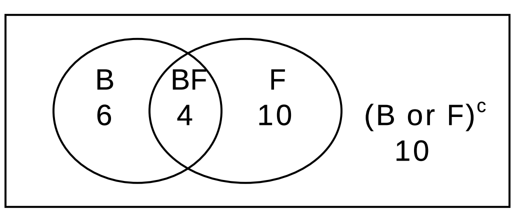
(e) No, because $P(A \text{ and } B) = 0.15$, while $P(A)P(B) = (0.50)(0.25) = 0.125$, so $P(A \text{ and } B) \neq P(A)P(B)$, which means that A and B are not independent.

(f) Yes. Any set and its complement are mutually exclusive, for they can have no points in common.

(g) No. An event and its complement are never independent. In this example $P(A|A^c) = 0$, which is not equal to $P(A) = 0.50$. (Alternatively, $P(A \text{ and } A^c) = 0 \neq P(A)P(A^c) = (0.50)(0.50) = 0.25$)

4. Let B stand for Burns and F stand for Fractures.

(a)



(b) We are told that $P(B) = 10/30$, $P(F) = 14/30$, and $P((B \text{ or } F)^c) = 1 - P(B \text{ or } F) = 10/30$. Since $P(B \text{ or } F) = P(B) + P(F) - P(B \text{ and } F)$ and $P(B \text{ or } F) = 20/30$, we find $P(B \text{ and } F) = P(B) + P(F) - P(B \text{ or } F) = 10/30 + 14/30 - 20/30 = 4/30$

4. (c) $P(B \text{ and } F^c) = P(B) - P(B \text{ and } F) = 10/30 - 4/30 = 6/30 = 0.20$

(d) $P(F \mid B) = P(B \text{ and } F) / P(B) = (4/30) / (10/30) = 0.4$

(e) Not independent, because $P(B \text{ and } F) \neq P(B)P(F)$:
$P(B \text{ and } F) = 4/30$, while $P(B)P(F) = (14/30) / (10/30)$

5. A table or Venn diagram is useful here:

	Bachelors	**Masters**	Total
Less than 30	9	6	15
Over 30	8	2	10
Total	17	8	25

(a) $P(>30 \text{ or } M) = P(>30) + P(M) - P(>30 \text{ and } M) = 16/25$

(b) $P(>30|M)=P(>30 \text{ and } M)/P(M)=(2/25)/(8/25)=1/4$

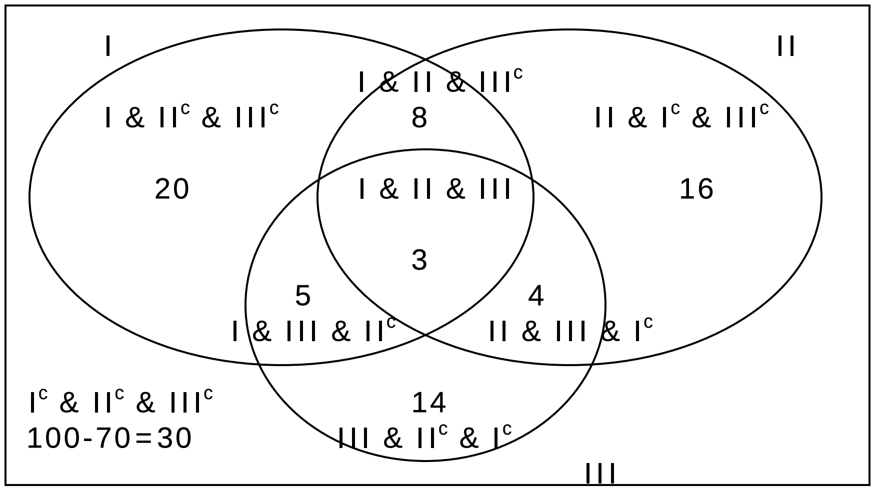
6. We are given that P(Course)=0.8, P(No Accident|C)=.85, and P(NA|No Course)=.60. A table or Venn diagram is useful here as well, but we don't have a denominator. **USEFUL TRICK:** when the probabilities are expressed are percentages we actually don't need a denominator because we can just pretend that it is 100, from there we can fill out the table by multiplication of various conditional probabilities.

(Pretend)	**Course**	**No course**	Total
Accident	80-68=12	8	12+8=20
No Accident	.85(80)=68	.60(20)=12	68+12=80
Total	.80(100)=80	.20(100)=20	100

(a) From the table we have, P(No Accident) = 80/100.
For the hard way (intended for this problem) use the Law of Total Probability:
$P(NA) = P(NA|C)P(C) + P(NA|NC)P(NC)$
$= (68/80)(80/100) + (12/20)(20/100)=(68+12)/100$

(b) From the table we have, P(C|A)=12/20=60%
For the hard way use Bayes' Theorem:
$P(C|A) = P(C \text{ and } A)/P(A)$
$= P(A|C)P(C)/\{ P(A|C)P(C)+ P(A|NC)P(NC)\}$
$= (1-.85)(.8)/\{(1-.85)(.8)+(1-.6)(.2)\}$
$= 60\%$

7. Probably the only way to make sense of this problem is to use a Venn Diagram. (A table approach might work here but I don't see any easy way to construct it.)



(a) P(yes to at least one question) = 70/100

This probability can also be obtained from the following formula:

P(I or II or II) = P(I) + P(II) + P(III) – P(I and II) – P(I and III) – P(II and II) + P(I and II and III)

= 0.36 + 0.31 + 0.26 - 0.11 - 0.08 - 0.07 + 0.03

= .70

※ Do you know why this formula works?

(b) P(no to all thee questions) = 30/100

Note that: P(I^c and II^c and III^c) = 1 - P(I or II or III)

= 1 - 0.7 = 0.3 (from (a))

(c) P(yes to 1 & II| yes to at least one)=

= P((yes to I & II) and (At least one))/P(at least one)

= P(yes to I & II)/P(at least one)

= (11/100)/(70/100) = 11/70

8. (a) (i) X=number of heads on three tosses of a coin
(ii) Let a 1 represent a heads, S={(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1),}}
(iii) The probability of each triplet is 1/8.

(b) P(no heads) = 1/8, P(one heads)=3/8, P(two heads)=3/8, P(three heads)=1/8.

(c) P(you get two heads and I get two heads) = ?
Let A = {two heads in three flips}, then P(A) = 3/8.
By independence we have
$P(A_{you} \text{ and } A_{me}) = P(A_{you})P(A_{me}) = P(A)P(A) = 9/64$

(d) P(you and I get the same number of heads) = ?
P(we both get no heads or we both get one heads or … or we both get three heads) =
P(both none) + P(both one) +P(both two) + P(both three) =
(1/8)(1/8) + (3/8)(3/8) + (3/8)(3/8) +(1/8)(1/8) = 20/64 =.31
(the second step because the events are disjoint and the third step because you and I are independent.)

(e) This is the same as (e) but with 5 people instead of two.
P(all five see the same number of heads)=
P(all five none) + … +P(all five see three)=
$(1/8)^5 + (3/8)^5 + (3/8)^5 + (1/8)^5 = 488/32768 = 1.5\%$
(the second step because the events are disjoint and the third step because all five coin flippers are independent.)

9. (a) P(live to 99| live to 85)= P(live to 99 and live to 85)/P(live to 85)
= P(live to 99)/P(live to 85)
= 0.708/0.853 = 0.83

(b) Let A be the event "Man who lives to age 85 lives to 99," and let B be the event "Woman who lives to age 85 lives to 99". We found in (a) that P(A) = P(B) = 0.83. And because A and B are assumed to be independent events,
P(A andB) = P(A)P(B) = $(0.83)^2$ = 0.689.

(c) P(A or B but not both) = P((A and not B) or (B and not A))
$= P(AB^c \text{ or } A^cB)$
$= P(AB^c) + P(A^cB)$ (because AB^c and A^cB are disjoint)
= P(A) - P(AB) + P(B) - P(AB)
= P(A) + P(B) - 2P(AB)
= 0.83 + 0.83 - 2(0.689) (from parts (a) and (b))
= 0.282 (Sad but true)

10. (a) False negative result = P(Test neg | have cancer) = 1-0.85 =0.15
(b) False positive result = P(Test pos| no cancer) = 1-0.8=0.2
(c) Given: P(T+ | C) = 0.85, P(T- | C^c) = 0.80, and P(C) = 0.0025

$$P(C\,|\,T+) = \frac{P(T+|\,C)\,P(C)}{P(T+|\,C)\,P(C) + P(T+|\,C^c)\,P(C^c)}$$

$$= \frac{(0.85)(0.0025)}{(0.85)(0.0025) + (1-0.80)(1-0.0025)}$$

$$= \frac{0.002125}{0.201625} = 0.01$$

11. (a) The <u>estimated</u> sensitivity is 302/481 = 0.628
The <u>estimated</u> Specificity is 372/452 = 0.823

(b) The predictive value of a positive test:

$$P(D|\,T+) = \frac{P(T+|\,D)\,P(D)}{P(T+|\,D)\,P(D) + P(T+|\,D^c)\,P(D^c)}$$

$$= \frac{(0.628)(0.10)}{(0.628)(0.10) + (1-0.823)(1-0.10)}$$

$$= 0.28$$

(b) The predictive value of a negative test:

$$P(D^c\,|T+) = \frac{P(T+|\,D^c)\,P(D^c)}{P(T+|\,D^c)\,P(D^c) + P(T+|\,D)\,P(D)}$$

$$= \frac{(0.823)(1-0.10)}{(0.823)(1-0.10) + (1-0.628)(0.10)}$$

$$= 0.95$$

12. (a) The false positive rate, P(T+|D-)=1-secificity, goes <u>down</u> as the cutoff point is raised.
The false negative rate, P(T-|D+)=1-sensitivity, goes <u>up</u> as the cutoff point is raised.

13. (a) Initially I noticed the wide variability among the PPV lines. Also note that the PPV ranges from zero to one depending on the prevalence, for any sensitivity and specificity.

(b) Prevalence will play an important role in determining the required sensitivity and specificity for a particular diagnostic test because these tests are often aimed at achieving some PPV or NPV for the overall population. Cost benefit analyses can also be based on PPV and NPV statistics.

(c) Somewhat. In most cases the PPV varied little (5%-10%) when the sensitivity was decreased. But again it depended on the prevalence.

14. The formulae are $odds_i = p_i/(1-p_i)$; $OR = odds_1/odds_2$; $RR = p_1/p_2$

(a) - (e): See following table

P_1	P_2	$odds_1$	$odds_2$	OR	RR
0.5	0.5	1	1	1	1
0.3	0.4	0.429	0.667	0.643	0.75
0.05	0.1	0.053	0.111	0.474	0.500
0.95	0.9	19.000	9.000	2.111	1.056
0.1	0.9	0.111	9.000	0.012	0.111
0.9	0.1	9.000	0.111	81.000	9.000
0.01	0.04	0.010	0.042	0.242	0.25
0.99	0.96	99.000	24.000	4.125	1.031
0.96	0.99	24.000	99.000	0.242	0.969

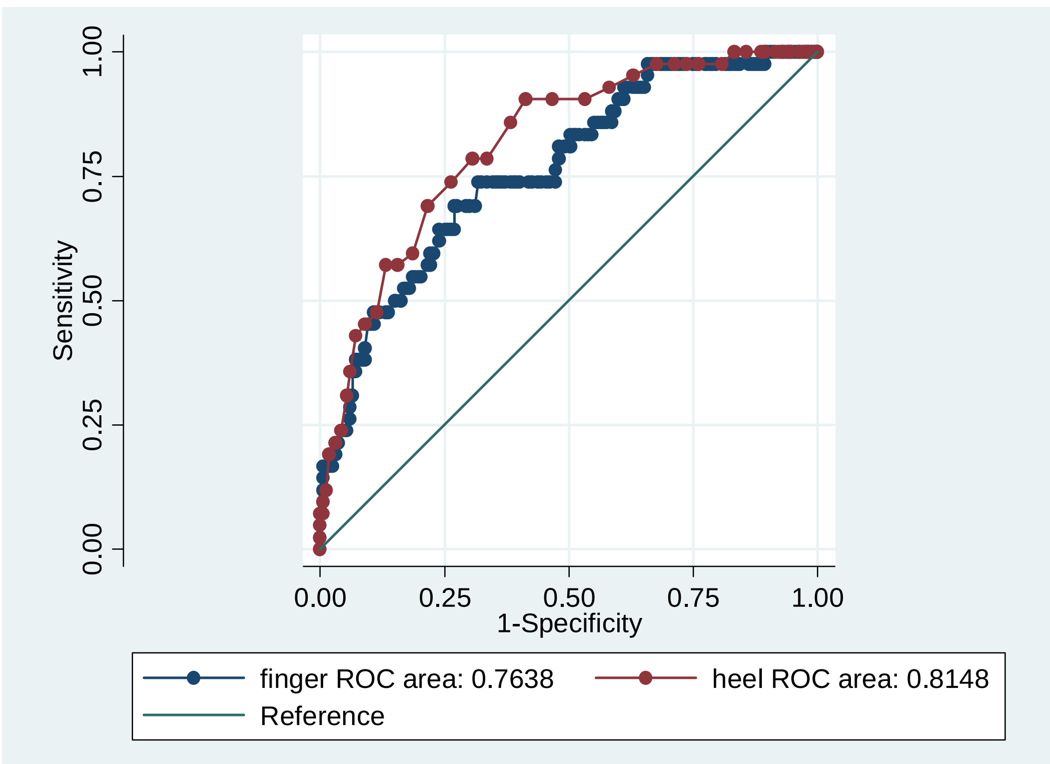
(f) The odds ratio is a good approximation of the relative risk when the event (success rate) is rare in both groups (i.e., when P_1 and P_2 are both small).

(g) (a) true and (b) false.

15. (a) Here is the graph and output generated from STATA:

```
. roccomp ostio finger heel, graph aspectratio(1) summary

                              ROC                    -Asymptotic Normal--
                   Obs        Area     Std. Err.      [95% Conf. Interval]
-------------------------------------------------------------------------
finger             209      0.7638       0.0408        0.68373     0.84379
heel               209      0.8148       0.0351        0.74596     0.88364
-------------------------------------------------------------------------
Ho: area(finger) = area(heel)
    chi2(1) =     1.28       Prob>chi2 =   0.2584
```



(b) The AUC is larger for Heel BMD measurements, which means that overall the predictive ability for Heel BMD is better.

To answer the next questions I used the commands (remember to take the negative of the BMD measurement):

```
roctab ostio finger, detail
roctab ostio heel, detail
```

(c) (>=.6) yields a sensitivity=92.86% and specificity=34.73%

(d) (>=.6) yields a sensitivity=85.71% and specificity=61.68%

(e) A standardized finger BMD of -2.15 or less would be considered positive for Ostio.

(f) A standardized heel BMD of -1.4 or less would be considered positive for Ostio.

(g) The results above show that the ROC curve hides information about the cutoff. Two test can have the same sensitivity and specificity, but operate at difference cutoffs (like above). So, if you wanted to use both tests in practice and you wanted to be equally accurate with each, you would need to use a cutoff of -2.15 or less for finger BMD and -1.4 or less for heel BMD as a surrogate for Osteoporosis.