1. Conditional probability

C=cancer
¬C=not-cancer
S=smoke
¬S=not-smoke
P(A,B) = 'probability of A *and* B'; e.g., P(¬S,C) = 'probability of not-smoke *and* cancer'
P(A|B) = 'probability of A, *given* B'; e.g., P(¬S|C) = 'probability of not-smoke, *given* cancer'

Table 1a. Contingency table for smoking and cancer with n=100 persons.

	C	¬C	
S	20	40	60
¬S	10	30	40
	30	70	100

Table 1b. Probability distribution associated with Table 1a.

	C	¬C	
S	2/10	4/10	6/10
¬S	1/10	3/10	4/10
	3/10	7/10	1

Table 1c. Table legend for probability distribution in Table 1b.

	C	¬C	
S	P(S,C)	P(S,¬C)	P(S)
¬S	P(¬S,C)	P(¬S,¬C)	P(¬S)
	P(C)	P(¬C)	1=P(S)+P(¬S) =P(C)+P(¬C) = P(S,C)+ P(S,¬C)+ P(¬S,C)+ P(¬S,¬C)

Formula for conditional probabilities: P(A|B)=P(A,B)/P(B)

For instance: P(C|S)=P(S,C)/P(S)=0.2/0.6=1/3

2. Independence

Table 2a. Contingency table for smoking and cancer with n=100 persons.

	C	¬C	
S	20	40	60
¬S	10	30	40
	30	70	100

P(C|S)=20/60=1/3
P(C|¬S)=10/40=1/4
P(C)=30/100=3/10

S carries information about C: Learning that S should increase your confidence that C; S and C are not independent in this table.

Table 2b. Contingency table for smoking and cancer with n=100 persons.

	C	¬C	
S	18	42	60
¬S	12	28	40
	30	70	100

P(C|S)=18/60=3/10
P(C|¬S)=12/40=3/10
P(C)=30/100=3/10

S carries no information about C: Learning that S should not increase your confidence that C; S and C are independent in this table.

Formula for independence: A and B are independent iff P(A|B)=P(A), or, which is the same, iff P(A,B)=P(A)P(B)

3. Conditional independence

C=cancer
¬C=not-cancer
S=smoke
¬S=not-smoke
F=stained fingers
¬F=not-stained fingers
P(A,B) = ‘probability of A *and* B’; e.g., P(¬S,C) = ‘probability of not-smoke *and* cancer’
P(A|B) = ‘probability of A, *given* B’; e.g., P(¬S|C) = ‘probability of not-smoke, *given* cancer’

Table 3a. Contingency table for stained fingers and cancer

	C	¬C	
F	16	74	90
¬F	14	96	110
	30	170	200

In this table:
P(C|F)=16/90=0,18
P(C|¬F)=14/110=0,13
P(C)=30/200=0,15
Hence, P(C|F)≠P(C); therefore C and F are not independent

Now suppose that among the people in 2a there are 100 smokers and 100 non-smokers. Make a separate contingency table for each group.

Table 3b. Contingency table for stained fingers and cancer, given S (only smokers here)

	C	¬C	
F	14	56	70
¬F	6	24	30
	20	80	100

In this table:
P(C|F)=14/70=0,2
P(C|¬F)=6/30=0,2
P(C)=20/100=0,2
Hence, P(C|F)=P(C); therefore C and F are independent in this table

Table 3c. Contingency table for stained fingers and cancer, given ¬S (only non-smokers here)

	C	¬C	
F	2	18	20
¬F	8	72	80
	10	90	100

In this table:
P(C|F)=2/20=0,1
P(C|¬F)=8/80=0,1
P(C)=10/100=0,1
Hence, P(C|F)=P(C); therefore C and F are independent in this table

Conclusion: Conditioning on smoking renders F and C probabilistically independent; we say that 'C and F are independent given S'.

Formula for conditional independence: A and B are conditionally independent given C iff P(A|B,C)=P(A|C).

Here:

P(C|F,S)=P(C,F,S)/P(F,S)=(14/200)/(70/200)=0,2
P(C|S)=P(S,C)/P(S)=(20/200)/(100/200)=0,2