#3.

$P(-|C) = \frac{1}{4}$, $P(\bar{C}|+) = \frac{2}{3}$.

$\hookrightarrow$ $\underbrace{P(+|C)}_{\text{Sensitivity}} = 1 - \frac{2}{3} = \frac{1}{3}$. $\underbrace{P(-|\bar{C})}_{\text{Specificity}} = 1 - P(+|\bar{C})$

Not determinable.

#6.

	Homicide	XHomicide
U.S	62.4	
CAN	6.0	
AUS	5.6	
UK	1.3	

per 100,000

relative risk between US And

CAN : $\frac{6.0}{62.4} = 0.0961$

AUS : $\frac{5.6}{62.4} = 0.0897$

UK : $\frac{1.3}{62.4} = 0.0208$

$\leadsto$ RR이 difference of Proportion 보다 낫다!
(∵ 확률이 낮은 사건일수록 그 비율이 중요!).

#9.

SMOKE O : S SMOKE X : $\bar{S}$

Cancer O : C Cancer X : $\bar{C}$

$P(C|\bar{S}) = 0.00023.$

$P(C|S) = 0.01284.$

$D.O.P = 0.01284 - 0.00023 = \underline{0.0000375}.$

NOT informative.

$RR = \frac{0.01284}{0.00023} = 55.8.$

More informative.

#12.

Cohort : Row-Sum fixed.

$P(C|S) = 0.00140.$ $P(C|\bar{S}) = 0.00010.$

$P(H|S) = 0.00669$ $P(H|\bar{S}) = 0.00413.$

$\boxed{a}$ *Lung Cancer - Smoke.

$DOP = 0.00140 - 0.00010 = 0.00130.$

$RR = 0.00140 / 0.00010 = 14.$

$\text{Odds ratio} = \frac{0.00140 \times 0.99990}{0.99860 \times 0.00010} = 14.018.$

RR $\approx$ Odds ratio since DOP $\simeq$ 0.

* Heart disease - Smoke.

$$DOP = 0.00669 - 0.00413 = 0.00256.$$

$$RR = \frac{0.00669}{0.00413} = 1.6198.$$

$$\text{Odds ratio} = \frac{0.00669}{0.00413} \times \frac{0.99331}{0.99587} = 1.6157.$$

By the same reason, RR $\approx$ Odds ratio.

$\boxed{b}$ Heart disease is more related to Cigarette smoking in terms of reduction in # of death

$(\because DOP : H > C)$

#15.

AG marginal odds ratio

$$\theta = 1189 \times 1278 / 1493 \times 557 = 1.827.$$ ~> 여성이 차별받은건 아닌가??

AG Conditional Odds ratio:

$$\theta_A = \frac{512 \times 19}{313 \times 89} = 0.349.$$

$$\theta_B = \frac{207 \times 17}{353 \times 8} = 1.246$$

~> 별 차이 X

⇒ 차별 X.

$\theta_C = \frac{120 \times 391}{205 \times 202} = 1.133$

$\theta_D = 0.9212.$

$\theta_E = 1.2216.$

$\theta_F = 0.8278.$

∴ A, B는 합격률 ↑

⊕ 남자 지원자수 ⇑

⇓

marginal odds ratio

남자가 여자보다 84% ↑

#18.

각 나이별로 A 사망률 > B 사망률. But 전체나이로 보면 A 사망률 < B 사망률 why??

상대적으로 높은 사망률의 나이대의 사람들이 B 지역에 더 많기 때문에.

#21.

	K_1	K_2	Total
Smith	18/40	25/160	43/200
Jones	40/100	15/100	55/200

#24

Sample Conditional Probability.

Black : $(0.14, 0.1, 0.31, 0.45)$.

White : $(0.06, 0.55, 0.29, 0.59)$ $\star$ Stochastically ordered!

$\boxed{\begin{array}{l}\underline{\text{def}}\ \text{Stochastic order } \preceq_{st} \\ A, B : \text{random Variable.} \quad A \leq_{st} B \text{ if} \\ \qquad Pr(A > x) \leq Pr(B > x)\end{array}}$

$$\Delta = P(Y_1 > Y_2) - P(Y_2 > Y_1)$$

$$\hat{\Delta} = \sum_{j>k}\sum \pi_{j|1}\pi_{k|2} - \prod_{j<k} \pi_{j|1}\pi_{k|2}.$$

$$= 0.06 \times (0.1 + 0.31 + 0.45)$$

$$+ 0.55 \times (0.31 + 0.45) + 0.29 \times 0.45.$$

$$= 0.0516 + 0.418 + 0.1305$$

$$= 0.6001.$$

$\Rightarrow P(Y_1 > Y_2) = P(Y_2 > Y_1) + 0.6.$

$\rightsquigarrow Y_1 > Y_2$ tendency $\uparrow$.

#27.

	D	$\bar{D}$	
E	90	910	1000
$\bar{E}$	10	590	600
	100	1500	

$$AR = \frac{\frac{100}{1600} - \frac{10}{600}}{\frac{100}{1600}}$$

$$= \frac{100 - 10 \cdot \frac{1600}{600}}{100}$$

1600명이 모두 노출X 경우로 만들어줌.

$$\frac{P(E)\cdot(RR-1)}{1+P(E)\cdot(RR-1)} = \frac{P(E)\cdot\left(\frac{P(D|E)}{P(D|\bar{E})}-1\right)}{1+P(E)\cdot\left(\frac{P(D|E)}{P(D|\bar{E})}-1\right)}$$

$$= \frac{P(E)P(D|E) - P(E)P(D|\bar{E})}{P(D|\bar{E}) + P(E)P(D|E) - P(E)P(D|\bar{E})}$$

$$= \frac{P(E)P(D|E) + P(\bar{E})P(D|\bar{E}) - P(D|\bar{E})}{P(E)P(D|E) + P(\bar{E})P(D|\bar{E})}$$

$$= \frac{P(D) - P(D|\bar{E})}{P(D)}.$$

$$= AR.$$

$\blacksquare$

#30.

$RR = \frac{\pi_1}{\pi_2}$ $\quad \theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$

$\Leftrightarrow \left| (1-\pi_2)\pi_1 - (1-\pi_1)\pi_2 \right| \geq \left| (1-\pi_1)\pi_1 - (1-\pi_1)\pi_2 \right|.$

$\Leftrightarrow |\pi_1 - \pi_2| \geq |\pi_1 - \pi_2||1-\pi_1|$ OK.

$\therefore \quad |\theta - 1| \geq |RR - 1|$

$\rightsquigarrow$ Independence test.

#33.

ⓐ

$\theta = \theta_{XY(1)} = \theta_{XY(2)}$

$$\Leftrightarrow \frac{p(X=1,Y=1|Z=1)\,p(X=2,Y=2|Z=1)}{p(X=1,Y=2|Z=1)\,p(X=2,Y=1|Z=1)} = \frac{p(X=1,Y=1|Z=2)\,p(X=2,Y=2|Z=2)}{p(X=1,Y=2|Z=2)\,p(X=2,Y=1|Z=2)}$$

$$= \frac{p(1.1.2)\,p(2.2.2)}{p(1.2.2)\,p(2.1.2)}$$

$$\theta_C \cdot \pi_1 + (1-\pi_1) = \theta_{(1)YZ}\,\pi_1 + (1-\pi_1)$$

$$= \frac{p(Y=1,Z=1|X=1)\,p(Y=2,Z=2|X=1)}{p(Y=1,Z=2|X=1)\,p(Y=2,Z=1|X=1)} \cdot \frac{\cancel{p(X=1,Y=2,Z=1)}}{\cancel{p(X=1,Y=2)}} + p(Z=2|X=1,Y=2)$$

$$\frac{\cancel{p(Y=1,Y=2)}\,p(Z=2|X=1,Y=2)}{\cancel{p(Y=1)}}$$

$$= p(Z=2|X=1,Y=2)\left(\frac{p(Y=1,Z=1|X=1)}{p(Y=1,Z=2|X=1)} + 1\right)$$

$$= p(Z=2|X=1,Y=2)\,\frac{p(Y=1|X=1)}{p(Y=1,Z=2|X=1)} = \frac{p(X=1,Y=1)}{p(X=1,Y=2)}\,\frac{p(1.2.2)}{p(1.1.2)}$$

$$\theta_C \cdot \pi_2 + (1-\pi_2)$$

$$= \theta_{(2)YZ}\, \pi_2 + (1-\pi_2)$$

$$\cancel{p(Z=2 \mid X=2, Y=2)\, \frac{p(X=2,Y=2)}{p(X=2)}}$$

$$= \frac{p(Y=1,Z=1 \mid X=2)\, \cancel{p(Y=2,Z=2 \mid X=2)}}{p(Y=1,Z=2 \mid X=2)\, \cancel{p(Y=2,Z=1 \mid X=2)}} \cdot \frac{\cancel{p(X=2,Y=2,Z=1)}}{p(X=2, \cancel{Y=2})} + p(Z=2 \mid X=2, Y=2)$$

$$= p(Z=2 \mid X=2, Y=2) \left(\frac{p(Y=1,Z=1 \mid X=2)}{p(Y=1,Z=2 \mid X=2)} + 1 \right) = p(Z=2 \mid X=2, Y=2)\, \frac{p(Y=1 \mid X=2)}{p(Y=1, Z=2 \mid X=2)}$$

$$= \frac{p(X=2,Y=1)}{p(X=2,Y=2)} \cdot \frac{p(2.2.2)}{p(2.1.2)}.$$

$$\Rightarrow \theta\, \frac{\theta_C \pi_1 + (1-\pi_1)}{\theta_C \pi_2 + (1-\pi_2)} = \theta \cdot \frac{p(X=1,Y=1)\, p(X=2,Y=2)}{p(X=1,Y=2)\, p(X=2,Y=1)} \cdot \frac{p(2.1.2)\, p(1.2.2)}{p(2.2.2)\, p(1.1.2)}$$

$$= \theta_{XY}\, \frac{p(1.1.2)\, \cancel{p(2.2.2)}}{\cancel{p(1.2.2)}\, p(2.1.2)} \cdot \frac{p(2.1.2)\, \cancel{p(1.2.2)}}{\cancel{p(2.2.2)}\, p(1.1.2)}$$

$$= \theta_{XY}.$$

(b)

Collapsibility of Odds ratio:

When $\theta_{XY(k)}$ is identical at every level k of Z, that value equals

if either Z and X are conditionally independent or

if Z and Y are conditionally independent.

By Symmetry, it suffices to show for the first case. $(Z \perp\!\!\!\perp X \mid Y)$

$$\pi_1 = p(Z=1 \mid X=1, Y=2)$$
$$= p(Z=1 \mid Y=2)$$
$$= p(Z=1 \mid X=2, Y=2) = \pi_2.$$

$$\therefore \frac{\theta_{XY}}{\theta} = 1. \Rightarrow \theta_{XY(1)} = \theta_{XY(2)} = \theta = \theta_{XY}.$$

ⓒ θ_{XY}/θ far from 1.0. ?

$$\frac{\theta_{XY}}{\theta} = \frac{\theta_c \pi_1 + (1-\pi_1)}{\theta_c \pi_2 + (1-\pi_2)}$$
$$= \frac{(\theta_c - 1)\pi_1 + 1}{(\theta_c - 1)\pi_2 + 1}.$$

① $\theta_c > 1$, $\pi_1 > \pi_2$.
Z is positively associated with X and Y.

② $\theta_c < 1$, $\pi_1 < \pi_2$.
Z is negatively associated with X and Y.

#36.

Variables are independent iff

$$\pi_{j|1} = \pi_{j|2} = \cdots = \pi_{j|I} \quad \forall \; j=1,\cdots,J-1.$$

(pf)

$\Rightarrow$: obvious

$\Leftarrow$: $\pi_{+j} = \sum_{i=1}^{I} \pi_{ij} = \sum_{i=1}^{I} \pi_{j|i}\pi_{i+}$

$$= \pi_{j|i} \sum_{i=1}^{I} \pi_{i+}$$

$$= \pi_{j|i} \qquad \forall_{j} = 1,\cdots,J-1.$$

#39.

$$V(Y) = \sum_{j=1}^{J} \pi_{+j}(1-\pi_{+j}) = \sum P(\text{1st observation} \in j \;\&\; \text{2nd observation} \notin j)$$

independent

$$= \sum_{j=1}^{J} \pi_{+j} - \pi_{+j}^2 = 1 - \sum \pi_{+j}^2.$$

a. If $\pi_{+j} = 1$ for some j, then

$$\sum_{j=1}^{J} \pi_{+j}(1-\pi_{+j}) \geq \pi_{+j}(1-\pi_{+j}) = 1\cdot 0 = 0.$$

$$1 - \sum_{j=1}^{J} \pi_{+j}^2 \leq 1 - \pi_{+j}^2 = 1 - 1^2 = 0.$$

$\therefore$ $V(Y) = 0$.

maximize $V(Y) = 1 - \sum \pi_{+j}^2$

subject to $\sum \pi_{+j} = 1$.

$\Rightarrow$ By Lagrange multiplier, $-2\pi_{+j} + \lambda = 0 \quad \forall_j$

$$\Rightarrow \pi_{+j} = \frac{\lambda}{2} \quad \forall_j \Rightarrow \frac{\lambda}{2} = \frac{1}{J}.$$