1 – Estudo de Coorte

Avaliar se os indivíduos expostos ao um determinado fator apresentam maior propensão ao desenvolvimento de uma doença do que os indivíduos não expostos a este fator;
Seleção de indivíduos sadios expostos e não expostos ao fator;
Estudo Observacional;
Acompanhamento por um período de tempo;

a) Probabilidade:

Categoria X	Categoria Y		
	$j=1(D)$	$j=2(\bar{D})$	Totais
$i=1(E)$	$p_{(1)1}$	$p_{(1)2}$	n_{1+}
$i=1(\bar{E})$	$p_{(2)1}$	$p_{(2)2}$	n_{2+}
Totais	$p_{(+1)}$	$p_{(+2)}$	n

Totais fixos na categoria de X
A probabilidade de um individuo pertencer a o j de Y, estando na categoria i de X:

$p_{(i)j} = P(Y=j|X=i)$

$p_{(1)1} = incidencia\ entre\ os\ expostos$

$p_{(2)1} = incidencia\ entre\ os\ não-expostos$

$N_{11} \sim Bin(n_{1+}, p_{(1)1})$

$N_{12} \sim Bin(n_{2+}, p_{(2)1})$

Modelo Probabilístico = Produto de duas binomiais

$$P(N_1 = n_1, N_2 = n_2) = \prod_{i=1}^{2} P(N_{i1} = n_{i1}, N_{i2} = n_{i2}) = \prod_{i=1}^{2}\left[(n_{i+})! \prod_{j=1}^{2} \frac{(p_{(i)j})^{n_{ij}}}{(n_{ij})!}\right]$$

Probabilidade $p_{(i)j}$estimada calculo baseado nos totais fixos, doentes sobre total linha;

$$\hat{p}_{(i)j} = \frac{n_{ij}}{n_{i+}}$$

b) Frequências esperadas:

$$e_{ij} = \frac{(n_{i+})(n_{+j})}{n}$$

Calculo baseado no produto das marginais

c) Hipóteses de homogeneidade:

$H_0: p_{(1)1} = p_{(2)1} (= p_{+1})$

$H_a: p_{(1)1} \neq p_{(2)1}$

$Sob\ H_0 => E(N_{i1}) = n_{i+}(p_{+1})\ \ e\ E(N_{i2}) = n_{i+}(p_{+2})$

Sob Ho não existe associação, exposto não difere de não-exposto;

$H_0: p_{(1)1} - p_{(2)1} = 0 => diferença\ entre\ incidencias\ (risco\ atribuivel)$

$H_0: \frac{p_{(1)1}}{p_{(2)1}} = 1 => razão\ de\ incidencia\ (risco\ relativo)$

$H_0: \frac{p_{(1)1}/(1-p_{(1)1})}{p_{(2)1}/(1-p_{(2)1})} = 1 => razão\ produtos\ cruzados\ (razão\ de\ chances/odds\ radio)$

Indica ou não associação entre exposto e não exposto;

d) Estatísticas do teste:

$$Pearson: Q_p = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij}-e_{ij})^2}{e_{ij}} \sim X^2_{(1)}$$

$$RVS: Q_L = -2\sum_{i=1}^{2}\sum_{j=1}^{2} log\left(\frac{e_{ij}}{n_{ij}}\right) \sim X^2_{(1)}$$

$$Neymar: Q_N = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij}-e_{ij})^2}{n_{ij}} \sim X^2_{(1)}$$

e) Estatísticas mensurar grau de associação:

i) Incidência = $\frac{numero\ de\ cados\ novos\ periodo}{numero\ de\ individuos\ experimento}$

ii) Diferença entre incidências: $d_{E/\bar{E}} = p_{(1)1} - p_{(2)1}$

iii) Risco atribuído à exposição: $\hat{d} = p_{(1)1} - p_{(2)1} = \frac{n_{11}}{n_{1+}} - \frac{n_{21}}{n_{2+}}$

iv) Razão de chance odds = $\frac{probabilidade\ do\ evento\ ocorrer}{probabilidade\ do\ evento\ não\ ocorrer}$

$odds_E = \frac{p_{(1)1}}{1-p_{(1)1}} odds_{\bar{E}} = \frac{p_{(2)1}}{1-p_{(2)1}} odds\ radio_{E/\bar{E}} = \frac{odds_E}{odds_{\bar{E}}} = \widehat{OR} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$

v) Risco relativo: $\widehat{RR} = risco\ relativo_{E|\bar{E}} = \frac{p_{(1)1}}{p_{(2)1}}$

Confirmar o se há ou não e qual o grau de associação;

$\hat{d} = variação\ entre\ exposto\ e\ não\ exposto;$

$incidencia = \%$ *de casos novos sob total de individuos experimento*
$\widehat{RR} =$ *probabilidadede expostos é vezes maior que não exposto de ficar doente;*
$\widehat{OR} =$ *dentre o n numeros de expostos apenas n entre não expostos pode ser doente;*

f) Estatísticas Intervalar:

i) Risco relativo: $IC(RR) = \exp\left(\hat{f} \mp z_{\frac{\alpha}{2}}\sqrt{\widehat{V(f)}}\right)$

$$f = \ln(RR) = \ln(p_{(1)1}) - \ln(p_{(2)1})$$
$$V(f) = \frac{(1-p_{(1)1})}{(n_{1+})(p_{(1)1})} + \frac{(1-p_{(2)1})}{(n_{2+})(p_{(2)1})}$$
$$z_{\frac{\alpha}{2}} = 100\left(1 - z_{\frac{\alpha}{2}}\right) \ percentil\ N(0{,}1)$$

ii) Razão de chance: $IC(OR) = \exp\left(\hat{f} \mp z_{\frac{\alpha}{2}}\sqrt{\widehat{V(f)}}\right)$

$$f = \ln(OR)$$
$$V(f) = \frac{1}{(n_{11})} + \frac{1}{(n_{12})} + \frac{1}{(n_{21})} + \frac{1}{(n_{22})}$$
$$z_{\frac{\alpha}{2}} = 100\left(1 - z_{\frac{\alpha}{2}}\right) \ percentil\ N(0{,}1)$$

2 – Estudo Ensaio Clinico Aleatorizado

População de interesse doente;Comparação entre tratamentos;Tratamentos alocados aleatoriamente;Estudos experimentais;

X Grupos	Y Desfecho		
	$j = 1(Com)$	$j = 2(Sem)$	Totais
$i = 1(Tratado)$	$p_{(1)1}$	$p_{(1)2}$	n_{1+}
$i = 1(Controle)$	$p_{(2)1}$	$p_{(2)2}$	n_{2+}
Totais	$p_{(+1)}$	$p_{(+2)}$	n

Delineamento similar ao estudo de Coorte;
$\hat{d} =$ *variação entre com medicação e sem medicação;*
$incidencia = \%$ *de casos tratados sob total de individuos experimento*
$\widehat{RR} =$ *probabilidadede tratados é vezes maior que caso controle;*
$\widehat{OR} =$ *dentre o n numeros de tratados apenas n entre controles teve melhora;*

3 – Estudo Caso-Controle

Avaliar se exposição a um fator esta associado ao desenvolvimento da doença;
Seleção de dois grupos de indivíduos casos e controles;
Seleção de localidade que assegure comparabilidade;
Emparelhamento de casos e controles;
Seleção de mais de grupo de controle;
Após seleção verifica-se exposto e não-exposto;
Estudo retrospectivo;

a) Probabilidade:

Categoria X	Categoria Y		
	$j = 1(C)$	$j = 2(\bar{C})$	Totais
$i = 1(E)$	$p_{(1)1}$	$p_{(1)2}$	p_{1+}
$i = 1(\bar{E})$	$p_{(2)1}$	$p_{(2)2}$	p_{2+}
Totais	$n_{(+1)}$	$n_{(+2)}$	n

Totais fixos na categoria de Y
$N_{11} \sim Bin(n_{+1}, p_{1(1)})$
$N_{12} \sim Bin(n_{+2}, p_{1(2)})$
Modelo Probabilístico = Produto de duas binomiais

$$P(N_1 = n_1, N_2 = n_2) = \prod_{i=1}^{2} P\left(N_{1j} = n_{1j}, N_{2j} = n_{2j}\right) = \prod_{i=1}^{2}\left[(n_{+j})! \prod_{j=1}^{2} \frac{(p_{i(j)})^{n_{ij}}}{(n_{ij})!}\right]$$

Probabilidade $p_{i(j)}$ estimada calculo baseado nos totais fixos, doentes sobre total coluna;

$$\hat{p}_{i(j)} = \frac{n_{ij}}{n_{i+}}$$

b) Frequências esperadas:

$$e_{ij} = \frac{(n_{i+})(n_{+j})}{n}$$

Calculo baseado no produto das marginais;

c) Hipóteses de homogeneidade:

$H_0: p_{1(1)} = p_{1(2)}$

$H_a: p_{1(1)} \neq p_{1(2)}$

$Sob\ H_0 => E(N_{1j}) = n_{+j}(p_{1+})\ e\ E(N_{2j}) = n_{+j}(p_{2+})$

Sob Ho não existe associação, exposição não difere entre casos e controles;

$$H_0: \frac{p_{1(1)}/(1-p_{2(1)})}{p_{1(2)}/(1-p_{2(2)})} = \frac{p_{1(1)}p_{2(2)}}{p_{1(2)}p_{2(1)}} = 1$$

$razão\ produtos\ cruzados\ (razão\ de\ chances/odds\ radio)$

Indica ou não associação entre casos e controles;

d) Estatísticas do teste:

$$Pearson: Q_p = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij}-e_{ij})^2}{e_{ij}} \sim X^2_{(1)}$$

$$RVS: Q_L = -2\sum_{i=1}^{2}\sum_{j=1}^{2} log\left(\frac{e_{ij}}{n_{ij}}\right) \sim X^2_{(1)}$$

$$Neymar: Q_N = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij}-e_{ij})^2}{n_{ij}} \sim X^2_{(1)}$$

e) Estatísticas mensurar grau de associação:

i) Razão de chance odds = $\frac{probabilidade\ de\ exposição\ ao\ fator}{probabilidade\ de\ não\ exposição\ ao\ fator}$

$odds_C = \frac{p_{(1)1}}{1-p_{(1)1}} odds_{\bar{C}} = \frac{p_{(2)1}}{1-p_{(2)1}} odds\ radio_{C/\bar{C}} = \frac{odds_C}{odds_{\bar{C}}} = \widehat{OR} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$

Chance de exposição ao fator sob estudo é maior ou menor aos casos do que de controles;

f) Estatísticas Intervalar:

i) Razão de chance: $IC(OR) = \exp\left(\hat{f} \mp z_{\frac{\alpha}{2}}\sqrt{\widehat{V(f)}}\right)$

$f = \ln(OR)$

$$V(f) = \frac{1}{(n_{11})} + \frac{1}{(n_{12})} + \frac{1}{(n_{21})} + \frac{1}{(n_{22})}$$

$z_{\frac{\alpha}{2}} = 100\left(1 - z_{\frac{\alpha}{2}}\right)\ percentil\ N(0{,}1)$

$\widehat{OR} = dentre\ o\ n\ numeros\ de\ casos\ apenas\ n\ entre\ controles\ melhora\ com\ medicamento;$

Doenças raras => RR ~ OR

4 – Estudo Transversal

Investigar as possíveis relações de causa;Seleção da amostra aleatória, coleta simultânea de uma variedade de características, realizada em único ponto do tempo, avaliação fotográfica; Total n é fixo e demais aleatórios;

a) Probabilidade:

Categoria X	Categoria Y		
	$j = 1$	$j = 2$	Totais
$i = 1$	p_{11}	p_{12}	p_{1+}
$i = 2$	p_{21}	p_{22}	p_{2+}
Totais	p_{+1}	p_{+2}	n

Totais fixo n;

Modelo Probabilístico = Multinomial;

$$P(N_{11} = n_{11}, N_{12} = n_{12}, N_{21} = n_{21}, N_{22} = n_{22}) = n! \prod_{i=1}^{2}\prod_{j=1}^{2} \frac{(p_{ij})^{n_{ij}}}{(n_{ij})!}$$

$$n_{ij} \geq 0, \sum_{i,j=1}^{2} n_{ij} = n\ (unidades)\ e\ \sum_{i,j=1}^{2} p_{ij} = 1\ (probabilidades)$$

Probabilidade p_{ij}estimadacalculo baseado nos totais fixos, doentes sobre total coluna;

$$\hat{p}_{ij} = \frac{n_{ij}}{n}$$

b) Frequências esperadas:

$$e_{ij} = \frac{(n_{i+})(n_{+j})}{n}$$

Calculo baseado no produto das marginais;

c) Hipóteses de Independência:

$H_0: p_{ij} = (p_{i+})(p_{+j})$

$H_a: p_{ij} \neq (p_{i+})(p_{+j})$ - pelo menos um par difere entre si;

d) Estatísticas do teste:

$$Pearson: Q_p = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij}-e_{ij})^2}{e_{ij}} \sim X^2_{(1)}$$

$$RVS: Q_L = -2\sum_{i=1}^{2}\sum_{j=1}^{2} log\left(\frac{e_{ij}}{n_{ij}}\right) \sim X^2_{(1)}\, Neymar: Q_N = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-e_{ij})^2}{n_{ij}} \sim X^2_{(1)}$$

e) Estatísticas mensurar grau de associação (intensidade):

i) Prevalência = $\frac{numero\ de\ individuos\ com\ a\ resposta\ na\ data\ da\ coleta}{numero\ de\ individuos\ pesqiosados\ no\ estudo}$

$$\widehat{RP} = \frac{n_{11}/n_{1+}}{n_{21}/n_{2+}}$$

ii) Incidência = $\frac{numero\ de\ indeividuos\ que\ tornaram\ doentes\ no\ \ perido\ \ da\ coleta}{numero\ de\ individuos\ pesqiosados\ no\ estudo}$

$$\widehat{OR} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

f) Estatísticas Intervalar:

i) Prevalência: $IC(RP) = \exp\left(\hat{f} \mp z_{\frac{\alpha}{2}}\sqrt{\widehat{V(f)}}\right)$

$$f = \ln(RR) = \ln(p_{11}) - \ln(p_{21})$$

$$V(f) = \frac{(1-p_{11})}{(n_{1+})(p_{11})} + \frac{(1-p_{21})}{(n_{2+})(p_{21})}$$

$$z_{\frac{\alpha}{2}} = 100\left(1 - z_{\frac{\alpha}{2}}\right)\ percentil\ N(0{,}1)$$

ii) Incidência: $IC(OR) = \exp\left(\hat{f} \mp z_{\frac{\alpha}{2}}\sqrt{\widehat{V(f)}}\right)$

$$f = \ln(OR)$$

$$V(f) = \frac{1}{(n_{11})} + \frac{1}{(n_{12})} + \frac{1}{(n_{21})} + \frac{1}{(n_{22})}$$

$$z_{\frac{\alpha}{2}} = 100\left(1 - z_{\frac{\alpha}{2}}\right)\ percentil\ N(0{,}1)$$

$\widehat{RP} = proporção\ de\ individuos\ que\ estão\ doentes\ em\ dum\ determinado\ tempo\ especifico;$

$\widehat{OR} = proporção\ de\ individuos\ que\ tornaram - se\ doentes\ no\ decorrer\ de\ um\ periodo\ especifico\ de\ acompanhamento;$

5 – Estudo Tempo de Duração Fixo

Contagem de indivíduos aleatória;

a) Probabilidade:

Categoria X	Categoria Y		
	$j = 1$	$j = 2$	Totais
$i = 1$	μ_{11}	μ_{12}	μ_{1+}
$i = 2$	μ_{21}	μ_{22}	μ_{2+}
Totais	μ_{+1}	μ_{+2}	μ

Tempo fixo

Modelo Probabilístico = Poisson Independentes;

$$P(N = n) = \prod_{i=1}^{2}\prod_{j=1}^{2} P(N_{ij}) = \prod_{i=1}^{2}\prod_{j=1}^{2}\frac{e^{-\mu_{ij}}(\mu_{ij})^{n_{ij}}}{(n_{ij})!}, \mu > 0$$

$$(N = n) = (N_{11} = n_{11}, N_{12} = n_{12}, N_{21} = n_{21}, N_{22} = n_{22})$$

$$\hat{\mu}_{ij} = n_{ij}$$

b) Frequências esperadas:

$$e_{ij} = \frac{(n_{i+})(n_{+j})}{n} = \frac{(\mu_{i+})(\mu_{+j})}{\mu}$$

c) Hipóteses de Multiplicatividade:

$$H_0: \frac{\mu_{1j}}{\mu_{1+}} = \frac{\mu_{2j}}{\mu_{2+}}\left(= \frac{\mu_{+j}}{\mu}\right), para\ j = 1{,}2, \ldots\ H_a: \frac{\mu_{1j}}{\mu_{1+}} \neq \frac{\mu_{2j}}{\mu_{2+}}$$

$$H_0: \mu_{ij} = \frac{(\mu_{i+})(\mu_{+j})}{\mu}, para\ j = 1{,}2, \ldots$$

$$H_a: \mu_{ij} \neq \frac{(\mu_{i+})(\mu_{+j})}{\mu}, para\ ao\ menos\ um\ par\ ij$$

d) Estatísticas do teste:

$$Pearson: Q_p = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} \sim X^2_{(1)}$$

$$RVS: Q_L = -2\sum_{i=1}^{2}\sum_{j=1}^{2} log\left(\frac{e_{ij}}{n_{ij}}\right) \sim X^2_{(1)}\, Neymar: Q_N = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(n_{ij}-e_{ij})^2}{n_{ij}} \sim X^2_{(1)}$$

e) Estatísticas mensurar grau de associação (intensidade):

i) Incidência = $\frac{numero\ de\ indeividuos\ que\ tornaram\ doentes\ no\ perido\ da\ coleta}{numero\ de\ individuos\ pesqiosados\ no\ estudo}$

$$\widehat{OR} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

f) Estatísticas Intervalar:

i) Incidência: $IC(OR) = \exp\left(\hat{f} \mp z_{\frac{\alpha}{2}}\sqrt{\widehat{V(f)}}\right)$

$f = \ln(OR)$

$V(f) = \frac{1}{(n_{11})} + \frac{1}{(n_{12})} + \frac{1}{(n_{21})} + \frac{1}{(n_{22})}$

$z_{\frac{\alpha}{2}} = 100\left(1 - z_{\frac{\alpha}{2}}\right)\ percentil\ N(0{,}1)$

$\widehat{OR} = proporção\ de\ individuos\ que\ tornaram - se\ doentes\ no\ decorrer\ de\ um\ periodo\ especifico\ de\ acompanhamento;$

6 – Tabelas de Contingência r x s;

6.1 – Variáveis X e Y Nominais:

a) Totais marginais linhas-fixos $(n_{i+}\ fixos - categorias\ de\ X)$;

Hipóteses Nulas Homogeneidade:

$H_0: p_{(1)1} = p_{(2)1} (= p_{+1})$

$H_a: p_{(1)1} \neq p_{(2)1}$

Estatísticas do Teste:

$$Pearson: Q_p = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \sim X^2_{(r-1)(s-1)}$$

$$RVS: Q_L = -2\sum_{i=1}^{2}\sum_{j=1}^{2} log\left(\frac{e_{ij}}{n_{ij}}\right) \sim X^2_{(r-1)(s-1)}$$

$$Neymar: Q_N = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij} - e_{ij})^2}{n_{ij}} \sim X^2_{(r-1)(s-1)}$$

$$e_{ij} = \frac{(n_{i+})(n_{+j})}{n}$$

b) Totais n fixo $(n\ fixo - categorias\ X\ e\ Y\ aleatórios)$;

Hipóteses Nulas Independência:

$H_0: p_{ij} = (p_{i+})(p_{+j})$

$H_a: p_{ij} \neq (p_{i+})(p_{+j})$ - pelo menos um par difere entre si;

Estatísticas do Teste:

$$Pearson: Q_p = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \sim X^2_{(r-1)(s-1)}$$

$$RVS: Q_L = -2\sum_{i=1}^{2}\sum_{j=1}^{2} log\left(\frac{e_{ij}}{n_{ij}}\right) \sim X^2_{(r-1)(s-1)}$$

$$Neymar: Q_N = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij} - e_{ij})^2}{n_{ij}} \sim X^2_{(r-1)(s-1)}$$

$$e_{ij} = \frac{(n_{i+})(n_{+j})}{n}$$

c) Totais aleatórios $(n, n_{i+}, n_{+j}, categorias\ X\ e\ Y\ aleatórios)$;

Hipóteses Nulas Multiplicatividade:

$H_0: \frac{\mu_{1j}}{\mu_{1+}} = \frac{\mu_{2j}}{\mu_{2+}}\left(= \frac{\mu_{+j}}{\mu}\right), para\ j = 1,2,\ldots$

$H_a: \frac{\mu_{1j}}{\mu_{1+}} \neq \frac{\mu_{2j}}{\mu_{2+}}$

$H_0: \mu_{ij} = \frac{(\mu_{i+})(\mu_{+j})}{\mu}, para\ j = 1,2,\ldots$

$H_a: \mu_{ij} \neq \frac{(\mu_{i+})(\mu_{+j})}{\mu}, para\ ao\ menos\ um\ par\ ij$

Estatísticas do Teste:

$$Pearson: Q_p = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \sim X^2_{(r-1)(s-1)}$$

$$RVS: Q_L = -2\sum_{i=1}^{2}\sum_{j=1}^{2} log\left(\frac{e_{ij}}{n_{ij}}\right) \sim X^2_{(r-1)(s-1)}$$

$$Neymar: Q_N = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij}-e_{ij})^2}{n_{ij}} \sim X^2_{(r-1)(s-1)}\, e_{ij} = \frac{(n_{i+})(n_{+j})}{n}$$

6.2 – Variáveis: X Nominal e Y Ordinal:

a) Totais marginais linhas-fixos (n_{i+} *fixos – categorias de X*);

Atribuir scores $a = (a_1, \ldots, a_n)$ para categoria de Y, e atribuir um score médio para cada sub população:

$$\bar{F}_i = \sum_{j=1}^{r} a_j(p_{(i)j}) \text{ , } i = 1, \ldots, s;$$

Se s = 2

Hipóteses Nulas Scores Médios Não Diferem:

H_0: $\bar{F}_1 = \bar{F}_2$

H_a: $\bar{F}_1 \neq \bar{F}_2$

Estatísticas do Teste score médio:

$$Q_s = \frac{(\hat{f}_1 - \mu_a)^2}{\frac{(n-n_{1+})}{(n_{1+})(n-1)}v_a} = \frac{(n-n_{1+})}{(n_{1+})(n-1)}\frac{(n_{1+})(\hat{f}_1 - \mu_a)^2}{v_a} \sim X^2{}_1$$

Se s > 2

H_0: $\bar{F}_1 = \bar{F}_2 = \cdots = \bar{F}_s$

H_a: $\bar{F}_1 \neq \bar{F}_2 \neq \cdots \neq \bar{F}_s$

Estatísticas do Teste score médio:

$$Q_s = \frac{(n-1)}{n}\frac{\sum_{i=1}^{s}(n_{i+})(\hat{f}_1 - \mu_a)^2}{v_a} \sim X^2{}_{(s-1)}$$

$$\bar{f}_i = \sum_{j=1}^{r} a_j\,(\hat{p}_{(i)j}) = \sum_{j=1}^{r} a_j\left(\frac{n_{ij}}{n_{i+}}\right), \quad i = 1,2$$

$$E(\bar{f}_1|H_0) = \sum_{j=1}^{r} a_j\left(\frac{E(N_{1j})}{n_{1+}}\right) = \sum_{j=1}^{r} a_j\left(\frac{n_{+j}}{n}\right) = \mu_a$$

$$V(\bar{f}_1|H_0) = \frac{(n-n_{1+})}{(n_{1+})(n-1)}\underbrace{\sum_{j=1}^{r}(a_j-\mu_a)^2\left(\frac{n_{+j}}{n}\right)}_{v_a} = \frac{(n-n_{1+})}{(n_{1+})(n-1)}v_a$$

Para totais fixos de colunas substituir os valores de n_{i+} por n_{+j} e seguir o próximo item;

6.3 – Variáveis: X Ordinal e Y Nominal:

a) Totais marginais colunas-fixos (n_{+j} *fixos – categorias de Y*);

Atribuir scores $a = (a_1, \ldots, a_n)$ para categoria de Y, e atribuir um score médio para cada sub população:

$$\bar{F}_j = \sum_{i=1}^{r} a_j(p_{i(j)}) \text{ , } j = 1, \ldots, s;$$

Se s = 2

Hipóteses Nulas Scores Médios Não Diferem:

H_0: $\bar{F}_1 = \bar{F}_2$

H_a: $\bar{F}_1 \neq \bar{F}_2$

Estatísticas do Teste score médio:

$$Q_s = \frac{(\hat{f}_1 - \mu_a)^2}{\frac{(n-n_{+1})}{(n_{+1})(n-1)}v_a} = \frac{(n-n_{+1})}{(n_{+1})(n-1)}\frac{(n_{+1})(\hat{f}_1 - \mu_a)^2}{v_a} \sim X^2{}_1$$

Se s > 2

H_0: $\bar{F}_1 = \bar{F}_2 = \cdots = \bar{F}_s$

H_a: $\bar{F}_1 \neq \bar{F}_2 \neq \cdots \neq \bar{F}_s$

Estatísticas do Teste score médio:

$$Q_s = \frac{(n-1)}{n}\frac{\sum_{j=1}^{s}(n_{+j})(\bar{f}_1 - \mu_a)^2}{v_a} \sim X^2{}_{(r-1)}$$

6.4 – Variáveis: X ou Y Ordinal:

a)Totais n fixo (n *fixo – categorias X e Y aleatórios*);

Hipóteses Nulas Independência:

$H_0: p_{ij} = (p_{i+})(p_{+j})$

$H_a: p_{ij} \neq (p_{i+})(p_{+j})$ - pelo menos um par difere entre si;

Estatísticas do Teste:

$$Pearson: Q_p = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij}-e_{ij})^2}{e_{ij}} \sim X^2_{(r-1)(s-1)} \quad RVS: Q_L = -2\sum_{i=1}^{2}\sum_{j=1}^{2} log\left(\frac{e_{ij}}{n_{ij}}\right) \sim X^2_{(r-1)(s-1)}$$

$$Neyman: Q_N = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij}-e_{ij})^2}{n_{ij}} \sim X^2_{(r-1)(s-1)} \quad e_{ij} = \frac{(n_{i+})(n_{+j})}{n}$$

6.5 – Variáveis: X e Y Ordinal:

a) Totais geral fixos $(n\ fixos - categorias\ de\ X\ e\ Y\ aleatórios)$**;**

Hipóteses Nulas Ausência de Tendência Linear Multinominal:

$H_0: p_{ij} = (p_{i+})(p_{+j})$

$H_a: p_{ij} \neq (p_{i+})(p_{+j})$ - pelo menos um par difere entre si;

Definir scores médio;

Estatísticas do Teste Correlação:

$$Q_{CS} = \frac{(\bar{f}-\mu_c\mu_a)^2}{V(\bar{f})} = \cdots = (n-1)(r_{ac})^2 \sim X_1^2$$

$$\bar{F} = \sum_{i=1}^{r}\sum_{j=1}^{s} c_i a_j p_{ij}$$

Estimativas

$$\bar{f} = \sum_{i=1}^{r}\sum_{j=1}^{s} c_i a_j \hat{p}_{ij} = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{c_i a_j n_{ij}}{n}$$

Sob Ho:

$$E(\bar{f}) = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{c_i a_j}{n} E(N_{ij}) = \sum_{i=1}^{r} c_i \frac{n_{i+}}{n} \sum_{j=1}^{s} a_j \frac{n_{+j}}{n} = \mu_c\mu_a$$

$$V(\bar{f}) = \sum_{i=1}^{r} (c_i-\mu_c)^2 \frac{n_{i+}}{n} \sum_{j=1}^{s} \frac{(a_j-\mu_a)^2 / \frac{n_{+j}}{n}}{n(-1)}$$

b) Totais n_{i+} **linha-fixos** $(n_{i+}\ fixos - categorias\ de\ X)$**;**

Assim, se as variáveis Y e X forem ordinais e escores puderem ser assumidos para as categorias de ambas

$\Downarrow$

- **Se total n fixo** $\Rrightarrow$ estatística da correlação

$$Q_{CS} = (n-1)(r_{ac})^2 \sim \chi_1^2$$

$r_{ac} =$ coeficiente de correlação de Pearson.

- **Se n_{i+} fixos** $\Rrightarrow$ estatísticas escore e/ou da correlação

Q_S e/ou Q_{CS}.

7 – Análise Estratificada

Confundimento: distorcem a associação entre X e Y alterando força ou sentido;

Modificadora de efeito: mostram o efeito de X sobre Y variando de acordo com a categoria sob uma variável Z, o que indica interação entre X e Y;

Comparação das OR ou RR sem e com estratificação

Tabelas 2 x 2 x 2

$OR \approx OR_1 \approx OR_2 \Rightarrow$ ausência de ambos

$OR \neq OR_1$ e OR_2, mas $OR_1 \approx OR_2 \Rightarrow$ confundimento

$OR \neq OR_1$ e OR_2, mas $OR_1 \approx OR_2 \Rightarrow$ confundimento

$OR \neq OR_1$ e OR_2, mas $OR_1 < OR_2 \Rightarrow$ interação

$OR \neq OR_1$ e OR_2, mas $OR_1 > OR_2 \Rightarrow$ interação

Se interação significativa (variável modificadora de efeito) $\Rightarrow$ apresentar RR ou OR associadas a cada estrato

Se confundimento $\Rightarrow$ apresentar medida global ajustada pela variável de confundimendo $\Rightarrow$ proposta Mantel-Haenszel.

Estatística Mantel-Haenszel (1959).

Neste estudo multicentros tem-se:

- conjunto de $q = 2$ tabelas de cont. 2 $\times$ 2 $(h = 1, \ldots, q)$

Tratamentos	Resposta j = 1	j = 2	Totais
i = 1	n_{h11}	n_{h12}	n_{h1+}
i = 2	n_{h21}	n_{h22}	n_{h2+}
Totais	n_{h+1}	n_{h+2}	n_h

- totais marginais-linha n_{hi+} fixos nas $q = 2$ tabelas.
- interesse em testar H_0: $p_{h(1)1} = p_{h(2)1}$, $h = 1, \ldots, q$
- condicional a H_0, $N_{h11} \sim$ Hipergeométrica tal que

$$e_{h11} = E(N_{h11} \mid n_h, n_{h1+}, n_{h+1}) = \frac{(n_{h1+})(n_{h+1})}{n_h}$$

$$v_{h11} = V(N_{h11} \mid n_h, n_{h1+}, n_{h+1}) = \frac{(n_{h1+})(n_{h2+})(n_{h+1})(n_{h+2})}{(n_h)^2(n_h - 1)}$$

Sob H_0 e para $\sum_{h=1}^{q} n_h$ suficientemente grande

$$Q_{MH} = \frac{\left(\sum_{h=1}^{q} n_{h11} - \sum_{h=1}^{q} e_{h11}\right)^2}{\sum_{h=1}^{q} v_{h11}} \sim \chi^2_{(1)}$$

- Q_{MH} é eficaz para avaliar associações se a maioria das diferenças $(p_{h(1)1} - p_{h(2)1})$ apresentar o mesmo sinal.
- Havendo homogeneidade das OR nas q tabelas 2 $\times$ 2

$$\widehat{OR}_{MH} = \frac{\sum_{h=1}^{q} \frac{n_{h11}\, n_{h22}}{n_h}}{\sum_{h=1}^{q} \frac{n_{h12}\, n_{h21}}{n_h}} \quad \Rightarrow \text{odds ratio comum}$$

Intervalo de confiança aproximado para OR_{MH}

$$IC_{95\%}(OR_{MH}) = \left(\widehat{OR}_{MH} \exp(z_{\alpha/2}\, \widehat{\sigma}), \widehat{OR}_{MH} \exp(-z_{\alpha/2}\, \widehat{\sigma})\right)$$

$$\widehat{\sigma}^2 = \frac{\sum_{h=1}^{q} \frac{(n_{h11}+n_{h22})(n_{h11}n_{h22})}{(n_h)^2}}{2\left(\sum_{h=1}^{q} \frac{(n_{h11}n_{h22})}{n_h}\right)^2} + \frac{\sum_{h=1}^{q} \frac{(n_{h12}+n_{h21})(n_{h12}n_{h21})}{(n_h)^2}}{2\left(\sum_{h=1}^{q} \frac{(n_{h12}n_{h21})}{n_h}\right)^2} + \frac{\sum_{h=1}^{q} \frac{(n_{h11}+n_{h22})(n_{h12}n_{h21})+(n_{h12}+n_{h21})(n_{h11}n_{h22})}{(n_h)^2}}{2\left(\sum_{h=1}^{q} \frac{(n_{h11}n_{h22})}{n_h}\right)\left(\sum_{h=1}^{q} \frac{(n_{h12}n_{h21})}{n_h}\right)}.$$

$$IC_{95\%}(OR_{MH}) = (2,1; 7,7)$$

- Y ordinal e X nominal com totais n_{i+} fixos $\Rightarrow Q_{SMH}$
- Y e X ordinais com totais n_{i+} fixos $\Rightarrow Q_{SMH}$ ou Q_{CSMH}
- Y e X ordinais com total n fixo $\Rightarrow Q_{CSMH}$

$Q_{SMH} \Rightarrow$ estatística escore médio de MH

$Q_{CSMH} \Rightarrow$ estatística da correlação de MH

Supondo $\mathbf{a}_h = (a_{h1}, a_{h2}, \cdots, a_{hr})$, Mantel definiu para o tratamento Ativo a seguinte soma de escores-estratos:

$$f_{+1+} = \sum_{h=1}^{2} (n_{h1+})(\bar{f}_{h1}),$$

$\bar{f}_{h1} = \sum_{j=1}^{3} a_{hj}\left(\frac{n_{h1j}}{n_{h1+}}\right)$: escore médio do trat. Ativo na tabela h.

Hipótese nula de ausência de associação entre tratamento e grau de melhora fica expressa por:

$$H_0\text{: } \bar{F}_{h1} = \bar{F}_{h2}\ (= \bar{F}),\ h = 1, 2.$$

Sob a hipótese nula H_0: $\bar{F}_{h1} = \bar{F}_{h2}$ $(= \bar{F})$, h = 1, 2, tem-se:

$$\mu = E(f_{+1+}) = \sum_{h=1}^{2} (n_{h1+})(\mu_h)$$

$$v = Var(f_{+1+}) = \sum_{h=1}^{2} \frac{(n_{h1+})\,(n_h - n_{h1+})}{(n_h - 1)}\, v_h$$

com $\mu_h = \sum_{j=1}^{3} a_{hj}\left(\frac{n_{h+j}}{n_h}\right)$ e $v_h = \sum_{j=1}^{3} (a_{hj} - \mu_h)^2 \left(\frac{n_{h+j}}{n_h}\right)$

- Quando os tamanhos amostrais $n_{+i+} = \sum_{h=1}^{2}\sum_{j=1}^{3} n_{hij}$ forem suficientemente grandes, f_{+1+} terá distribuição aproximada normal, de modo que:

$$Q_{SMH} = \frac{(f_{+1+} - \mu)^2}{v} \sim \chi^2_{(1)}$$

Estatística escore médio estendida de Mantel-Haenszel

- Q_{SMH} será eficiente para detectar padrões de diferenças quando $(\bar{f}_{h1} - \bar{f}_{h2})$ apresentarem predominantemente o mesmo sinal.

Ainda,

$$E(\bar{f}_h) = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{c_i a_j}{n_h} E(N_{hij}) = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{c_i a_j}{n_h} \frac{(n_{hi+})(n_{h+j})}{n_h}$$
$$= \sum_{i=1}^{3} c_i \left(\frac{n_{hi+}}{n_h}\right) \sum_{j=1}^{2} a_j \left(\frac{n_{h+j}}{n_h}\right) = \mu_{hc}\mu_{ha}$$

$$V(\bar{f}_h) = \left\{ \sum_{i=1}^{3} (c_i - \mu_{hc})^2 \left(\frac{n_{hi+}}{n_h}\right) \sum_{j=1}^{2} \frac{(a_j - \mu_{ha})^2 (n_{h+j}/n_h)}{(n_h - 1)} \right\} = \frac{v_{hc} v_{ha}}{(n_h - 1)}$$

Quando o tamanho amostral combinado das q tabelas $s \times 2$ for grande (na prática $\sum_{h=1}^{q} n_h \geq 40$), tem-se:

$$Q_{CSMH} \sim \chi^2_{(1)}$$

- Sob H_0: $p_{hij} = (p_{hi+})(p_{h+j})$, $h = 1, \ldots, q$, e considerando os escores **a** = (a_{h1}, a_{h2}) e **c** = (c_{h1}, c_{h2}, c_{h3}), segue que:

$$Q_{CSMH} = \frac{\left[\sum_{h=1}^{q} n_h \left(\bar{f}_h - E(\bar{f}_h)\right)\right]^2}{\sum_{h=1}^{q} n_h^2 \, V(\bar{f}_h)} = \frac{\left[\sum_{h=1}^{q} n_h (v_{hc} v_{ha})^{1/2} r_{ac.h}\right]^2}{\sum_{h=1}^{q} n_h^2 \frac{v_{hc} v_{ha}}{(n_h - 1)}},$$

Estatística da correlação estendida de Mantel-Haenszel

$$\bar{f}_h = \sum_{i=1}^{3}\sum_{j=1}^{2} c_i a_j \widehat{p}_{hij} = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{c_i a_j n_{hij}}{n_h},$$

$r_{ac.h}$ = coeficiente de correlação de Pearson associado à tabela h.

Extensões em tabelas de contingência $s \times r$

- Para um conjunto de tabelas $s \times r$, com $s, r > 2$, é possível testar as associações de interesse por meio de extensões das estatísticas apresentadas. Assim,

- Se as categorias de X e Y forem nominais $\Rrightarrow Q_{MH}$
- Se as categorias de Y forem ordinais e n_{i+} fixos $\Rrightarrow Q_{SMH}$
- Se as categorias de X e Y ordinais e n ou n_{i+} fixos $\Rrightarrow Q_{CSMH}$

Dados relacionados em tabelas de contingência 2×2

McNemar (1947) propôs um teste o qual se baseia na argumentação de que somente os elementos fora da diagonal são importantes para determinar a existência de diferenças entre essas proporções.

$$Q_{Mc} = \frac{(n_{12} - n_{21})^2}{(n_{12} + n_{21})} \sim \chi^2_{(1)}$$

Embora o teste de McNemar possa ser utilizado, duas medidas adicionais auxiliam a mensurar a acurácia do resultado de um exame clínico.

A **sensibilidade** $=$ proporção de resultados positivos que um teste apresenta, quando realizado em sujeitos com a doença $\Rrightarrow$ proporção de verdadeiros positivos.

A **especificidade** $=$ a proporção de resultados negativos que um teste apresenta, quando realizado em sujeitos livres da doença $\Rrightarrow$ proporção de verdadeiros negativos.

O desejado de um exame clínico é que ele tenha alta sensibilidade e alta especificidade simultaneamente.

Coeficiente kappa (Cohen, 1960)

Ausência de concordância entre os observadores $\Rrightarrow H_0$: $p_{ij} = (p_{i+})(p_{+j})$

$$\kappa = \frac{\Pi_0 - \Pi_e}{1 - \Pi_e}$$

- $\Pi_0 = \sum_{i=1}^{s} p_{ii} \Rrightarrow$ probabilidade de concordância
- p_{ii} = probabilidade de um indivíduo ser classificado na categoria i por ambos os observadores
- $\Pi_e = \sum_{i=1}^{s} (p_{i+})(p_{+i}) \Rrightarrow$ probabilidade de concordância sob H_0.

$$\widehat{\kappa} = \frac{\widehat{\Pi}_0 - \widehat{\Pi}_e}{1 - \widehat{\Pi}_e}$$

$$\widehat{\Pi}_0 = \sum_{i=1}^{s} \frac{n_{ii}}{n} \quad \text{e} \quad \widehat{\Pi}_e = \sum_{i=1}^{s} \frac{n_{i+}}{n}\frac{n_{+i}}{n}$$

Se todos os elementos fora da diagonal $= 0 \Rrightarrow \Pi_0 = 1$.
$\Rrightarrow$ $\kappa = 1$ quando existir concordância perfeita.
$\Rrightarrow$ $\kappa = 0$ quando a concordância for aquela sob H_0.
Assim quanto mais próximo de 1 for o valor de κ, maior a concordância entre os observadores.

- Embora não usual, pode-se obter valores negativos para κ.
- Considera-se, em geral,
  $\kappa <$ 0,4 $\Rrightarrow$ concordância fraca,
  $\kappa \in$ [0,4; 0,8) $\Rrightarrow$ concordância moderada
  $\kappa \geq$ 0,8 $\Rrightarrow$ concordância forte.

Um I.C.(κ) pode ser obtido por:

$$\widehat{\kappa} \pm z_{\alpha/2}\sqrt{\widehat{var}(\widehat{\kappa})}$$

$z_{\alpha/2} =$ 100(1 $-$ $\alpha/2$) percentil da N(0,1)

$$var(\widehat{\kappa}) = \frac{(A+B-C)}{n(1-\Pi_e)^2}$$

$$A = \sum_i p_{ii}\left[1-\left((p_{i+})+(p_{+i})\right)(1-\kappa)\right]^2$$

$$B = (1-\kappa)^2\sum_{i\neq j} p_{ij}\left((p_{i+})(p_{+j})\right)^2$$

$$C = \left[\kappa - \Pi_e(1-\kappa)\right]^2$$

$$\widehat{\kappa}_w = \frac{\widehat{\Pi}_0(w)-\widehat{\Pi}_e(w)}{1-\widehat{\Pi}_e(w)}$$

A variância assintótica do coeficiente kappa ponderado:

$$var(\widehat{\kappa}_w) = \frac{\sum_i\sum_j p_{ij}\left[w_{ij}-(\bar{w}_{i+}+\bar{w}_{+j})(1-\kappa_w)\right]^2 - \left[\kappa_w - \Pi_e(w)(1-\kappa_w)\right]^2}{n[1-\Pi_e(w)]^2},$$

sendo $\bar{w}_{i+} = \sum_j (p_{+j})(w_{ij})$ e $\bar{w}_{+j} = \sum_j (p_{i+})(w_{ij})$.

Assim, I.C.(κ_w) pode ser obtido por:

$$\widehat{\kappa}(w) \pm z_{\alpha/2}\sqrt{\widehat{var}(\widehat{\kappa}_w)},$$

em que $z_{\alpha/2}$ é o 100(1 $-$ $\alpha/2$) percentil da N(0,1)

Quando as categorias da resposta forem ordinais:

$$\kappa_w = \frac{\Pi_0(w)-\Pi_e(w)}{1-\Pi_e(w)} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{s} w_{ij}\,p_{ij} - \sum_{i=1}^{s}\sum_{j=1}^{s} w_{ij}(p_{i+})(p_{+j})}{1-\sum_{i=1}^{s}\sum_{j=1}^{s} w_{ij}(p_{i+})(p_{+j})},$$

em que w_{ij} são pesos com valores entre 0 e 1

Um possível conjunto de pesos é dado por:

$$w_{ij} = 1 - \frac{|\,\text{escore}_{(i)} - \text{escore}_{(j)}\,|}{\text{escore}_{(s)} - \text{escore}_{(1)}}$$

sendo escore$_{(1)}$, escore$_{(i)}$, escore$_{(j)}$ e escore$_{(s)}$ os escores associados, respectivamente, à primeira, i-ésima, j-ésima e s-ésima linhas da tabela $s \times s$.