1 א

$$OR=\frac{odds(D=1|E=1)}{odds(D=1|E=0)}=\frac{P_{11}\cdot P_{00}}{P_{01}\cdot P_{10}}=$$

$$\frac{P(E=1,D=1)\cdot P(E=0,D=0)}{P(E=0,D=1)\cdot P(E=1,D=0)}=\frac{P(E=1|D=1)\cdot \cancel{P(D=1)}\cdot P(E=0|D=0)\cdot \cancel{P(D=0)}}{P(E=0|D=1)\cdot \cancel{P(D=1)}\cdot P(E=1|D=0)\cdot \cancel{P(D=0)}}=$$

הסתברות מותנית

$$\frac{P(E=1|D=1)\cdot P(E=0|D=0)}{P(E=0|D=1)\cdot P(E=1|D=0)}=\frac{P(E=1|D=1)\cdot(1-P(E=1|D=0))}{\underbrace{(1-P(E=1|D=1))}_{ODDS(E=1|D=1)}\underbrace{P(E=1|D=0)}_{\frac{1}{ODDS(E=1|D=0)}}}=$$

$$\frac{ODDS(E=1|D=1)}{ODDS(E=1|D=0)}$$

ב

$$\frac{OR}{RR}=\frac{\cancel{P_{11}}\cdot P_{00}}{\cancel{P_{01}}\cdot P_{10}}\cdot\frac{\cancel{P_{01}}(P_{10}+P_{11})}{\cancel{P_{11}}(P_{00}+P_{01})}=\frac{P_{00}}{P_{10}}\cdot\frac{(P_{10}+\overset{0}{P_{11}})}{(P_{00}+\underset{0}{P_{01}})}=\frac{\cancel{P_{00}}\cdot \cancel{P_{10}}}{\cancel{P_{10}}\cdot \cancel{P_{00}}}=1$$

כלומר נסיק שכאשר יש דעיכה בהסתברות להידבק "במחלה", או ההסתברות לכך קטנה עד אפסית, אזי $\underbrace{\text{יחס הסיכויים}}_{OR}$ הינו קרוב $\underbrace{\text{לסיכון היחסי}}_{RR}$:

2

ההשערות:

$$H_0: P(E=i,D=j)=P(E=i)\cdot P(D=j)\quad \forall i,j\in\{0,1\}$$

המשתנים בלתי תלויים

$$H_1: P(E=i,D=j)\neq P(E=i)\cdot P(D=j)\quad \forall i,j\in\{0,1\}$$

כפי שלמדנו בתרגול: $\hat{P}_{ij}=\frac{X_{ij}}{n}$

תחת H_0: $\hat{P}_{H_{0_{ij}}}=\hat{P}_i\cdot\hat{P}_j=\frac{X_i\cdot X_j}{n^2}$

ת $= 2\log\left(\frac{L(\hat\theta)}{L(\hat\theta_0)}\right)$

סטטיסטי המבחן:

$$\frac{L(\hat\theta)}{L(\hat\theta_0)} = \frac{\cancel{\frac{n!}{x_{11}!\cdot x_{10}!\cdot x_{01}!\, x_{00}!}}\cdot \hat p_{11}^{x_{11}}\cdot \hat p_{10}^{x_{10}}\cdot \hat p_{01}^{x_{01}}\, p_{00}^{x_{00}}}{\left(\cancel{\frac{n!}{x_{11}!\cdot x_{10}!\cdot x_{01}!\cdot x_{00}!}}\cdot \hat p_{11}^{x_{01}}\cdot \hat p_{10}^{x_{10}}\cdot \hat p_{00}^{x_{00}}\, p_{01}^{x_{01}}\right)\Big| H_0} =$$

$$\frac{\left(\frac{x_{11}}{n}\right)^{x_{11}}\cdot\left(\frac{x_{10}}{n}\right)^{x_{10}}\cdot\left(\frac{x_{01}}{n}\right)^{x_{01}}\cdot\left(\frac{x_{00}}{n}\right)^{x_{00}}}{\left(\frac{x_{1.}\cdot x_{.1}}{n^2}\right)^{x_{11}}\cdot\left(\frac{x_{1.}\cdot x_{.0}}{n^2}\right)^{x_{10}}\left(\frac{x_{0.}\cdot x_{.1}}{n^2}\right)^{x_{01}}\left(\frac{x_{0.}\cdot x_{.0}}{n^2}\right)^{x_{00}}} = \frac{(n^2\cdot x_{11})^{x_{11}}\cdot(n^2\cdot x_{10})^{x_{10}}\cdot(n^2\cdot x_{01})^{x_{01}}\cdot(n^2\cdot x_{00})^{x_{00}}}{(x_{1.}\cdot x_{.1})^{x_{11}}\cdot(x_{1.}\cdot x_{.0})^{x_{10}}\cdot(x_{0.}\cdot x_{.1})^{x_{01}}\cdot(x_{0.}\cdot x_{.0})^{x_{00}}}$$

$$= \left(\frac{n\cdot x_{11}}{x_{1.}\cdot x_{.1}}\right)^{x_{11}}\cdot\left(\frac{n\cdot x_{10}}{x_{1.}\cdot x_{.0}}\right)^{x_{10}}\cdot\left(\frac{n\cdot x_{01}}{x_{0.}\cdot x_{.1}}\right)^{x_{01}}\cdot\left(\frac{n\cdot x_{00}}{x_{0.}\cdot x_{.0}}\right)^{x_{00}}$$

$$\log\left(\frac{L(\hat\theta)}{L(\hat\theta_0)}\right) = \log\left(\left(\frac{n\cdot x_{11}}{x_{1.}\cdot x_{.1}}\right)^{x_{11}}\cdot\left(\frac{n\cdot x_{10}}{x_{1.}\cdot x_{.0}}\right)^{x_{10}}\cdot\left(\frac{n\cdot x_{01}}{x_{0.}\cdot x_{.1}}\right)^{x_{01}}\cdot\left(\frac{n\cdot x_{00}}{x_{0.}\cdot x_{.0}}\right)^{x_{00}}\right) =$$

$$= x_{11}\cdot\log\left(\frac{n\cdot x_{11}}{x_{1.}\cdot x_{.1}}\right) + x_{10}\cdot\log\left(\frac{n\cdot x_{10}}{x_{1.}\cdot x_{.0}}\right) + x_{01}\log\left(\frac{n\cdot x_{01}}{x_{0.}\cdot x_{.1}}\right) + x_{00}\log\left(\frac{n\cdot x_{00}}{x_{0.}\cdot x_{.0}}\right)$$

$$= \sum_{i=0}^{1}\sum_{j=0}^{1} x_{ij}\log\left(\frac{n\cdot x_{ij}}{x_{i.}\cdot x_{.j}}\right)$$

$$\Rightarrow T = 2\log\left(\frac{L(\hat\theta)}{L(\hat\theta_0)}\right) = 2\sum_{i=0}^{1}\sum_{j=0}^{1} x_{ij}\cdot\log\left(\frac{n\cdot x_{ij}}{x_{i.}\cdot x_{.j}}\right)$$

לסיכום סה"כ המבחן הוא:

$$H_0: P(E=i, D=j) = P(E=i)\cdot P(D=j) \quad \forall i,j\in\{0,1\}$$
$$H_1: P(E=i, D=j) \neq P(E=i)\cdot P(D=j) \quad \forall i,j\in\{0,1\}$$

$$T: 2\sum_{i=0}^{1}\sum_{j=0}^{1} x_{ij}\cdot\log\left(\frac{n\cdot x_{ij}}{x_{i.}\cdot x_{.j}}\right), \quad R.R: T > \chi^2_{1,\alpha}$$

3

1. X_1 עולה ביחידה:

$$\pi_2=\pi(\beta_0+\beta_1(X_1+1)+\beta_2X_2+\beta_{1,2}(X_1+1)\cdot X_2)$$

$$\log\left(\frac{\frac{\tilde{\pi}_i}{1-\tilde{\pi}_i}}{\frac{\pi_i}{1-\pi_i}}\right)=\log\left(\frac{\tilde{\pi}_1}{1-\tilde{\pi}_i}\right)-\log\left(\frac{\pi_i}{1-\pi_i}\right)=$$

$$\beta_0+\beta_1(X_1+1)+\beta_2X_2+\beta_{1,2}(X_1+1)\cdot X_2-(\beta_0+\beta_1\cdot X_1+\beta_2X_2+\beta_{1,2}\cdot X_1\cdot X_2)=\beta_1+\beta_{1,2}\cdot X_2$$

כלומר השינוי יהיה: $e^{\beta_1+\beta_{1,2}\cdot X_2}$

X_2 עולה ביחידה:

$$\pi_2=\pi(\beta_0+\beta_1X_1+\beta_2(X_2+1)+\beta_{1,2}\cdot X_1(X_2+1))$$

$$\log\left(\frac{\frac{\tilde{\pi}_i}{1-\tilde{\pi}_i}}{\frac{\pi_i}{1-\pi_i}}\right)=\beta_0+\beta_1X_1+\beta_2(X_2+1)+\beta_{1,2}\cdot X_1(X_2+1)-(\beta_0+\beta_1X_1+\beta_2X_2+\beta_{1,2}X_1X_2)=\beta_2+\beta_{12}X_1$$

כלומר השינוי יהיה $e^{\beta_2+\beta_{12}X_1}$

2. X_1 עולה ביחידה:

$$\pi_2=\pi(\beta_0+\beta_1(X_1+1)+\beta_2X_2+\beta_3(X_1+1)^2)$$

$$\log\left(\frac{\frac{\tilde{\pi}_i}{1-\tilde{\pi}_i}}{\frac{\pi_i}{1-\pi_i}}\right)=\beta_0+\beta_1(X_1+1)+\beta_2X_2+\beta_3(X_1+1)^2-(\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_1^2)$$

$$=\beta_1+\beta_2(X_1+1)^2-\beta_3X_1^2=\beta_1+\beta_3X^2+\beta_3 2X_1+\beta_3-\beta_3X_1^2=\beta_1+(2X_1+1)\cdot\beta_3$$

כלומר השינוי יהיה $e^{\beta_1+(2X_1+1)\beta_3}$

X_2 עולה ביחידה:

$$\pi_2=\pi(\beta_0+\beta_1X_1+\beta_2(X_2+1)+\beta_3X_1^2)$$

$$\log\left(\frac{\frac{\tilde{\pi}_i}{1-\tilde{\pi}_i}}{\frac{\pi_i}{1-\pi_i}}\right)=\beta_0+\beta_1X_1+\beta_2(X_2+1)+\beta_3\cdot X_1^2-(\beta_0+\beta_1X_1+\beta_2X_2-\beta_3X^2)=\beta_2$$

כלומר השינוי יהיה e^{β_2}

(4) בעצם נרצה להבין האם המשתנה $\beta_{1,2}$

עוזר לנו להסביר את התוצאות.

$$H_0: \beta_{1,2} = 0$$
$$H_1: \beta_{1,2} \neq 0$$

נפתח את הסטטיסטי:

$$\lambda = 2\log\left(\frac{L(\hat{\theta})}{L(\hat{\theta}_0)}\right) = 2\left(\log(L(\hat{\theta})) - \log(L(\hat{\theta}_0))\right) =$$

$$\log(L(\hat{\theta})) = \log\left(\prod_{i=1}^{n} \pi^{Y_i}(1-\pi_i)^{1-Y_i}\right) = \sum_{i=1}^{n} \log\left(\pi_i^{Y_i}(1-\pi_i)^{1-Y_i}\right)$$

$$= \sum_{i=1}^{n} Y_i \log(\pi_i) + \sum_{i=1}^{n}(1-Y_i)\log(1-\pi_i) = \sum_{i=1}^{n} Y_i \log(\pi_i) + \sum_{i=1}^{n}\log(1-\pi_i) - \sum_{i=1}^{n} Y_i \log(1-\pi_i) =$$

$$\sum_{i=1}^{n}\log\left(\frac{\pi_i^{Y_i}}{(1-\pi_i)^{Y_i}}\right) + \sum_{i=1}^{n}\log(1-\pi_i) = \sum_{i=1}^{n} Y_i\left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_{12} X_{1i} X_{2i}\right) -$$

$$\log\left(\frac{1}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_{12} X_{1i} \cdot X_{2i}}}\right) =$$

$$\sum_{i=1}^{n} Y_i\left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_{12} X_{1i} X_{2i}\right) - \log\left(1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_{12} X_{1i} X_{2i}}\right)$$

$$\log(L(\hat{\theta}_0)) = \sum_{i=1}^{n} Y_i\left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}\right) - \log\left(1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}}\right)$$

נחשב את מבחן הנראות:

$$\log(L(\hat{\theta})) - \log(L(\hat{\theta}_n)) =$$

$$\sum_{i=1}^{n} Y_i\left(\hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_{12} X_{1i} X_{2i}\right) + \log\left(\frac{1}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_{12} X_{1i} \cdot X_{2i}}}\right) -$$

$$\left(\sum_{i=1}^{n} Y_i(\hat\beta_0+\hat\beta_1 X_{1i}+\hat\beta_2 X_{2i})-\log\left(1+e^{\hat\beta_0+\hat\beta_1 X_{1i}+\hat\beta_2 X_{2i}}\right)\right)=$$

$$\sum_{i=1}^{n}\hat\beta_{12}X_{1i}X_{2i}+\log\left(\frac{1+e^{\hat\beta_0+\hat\beta_1 X_{1i}+\hat\beta_2 X_{2i}}}{1+e^{\hat\beta_0+\hat\beta_1 X_1+\hat\beta_2 X_2+\hat\beta_{12}X_{1j}X_{2i}}}\right)$$

כלומר סה"כ המבחן הוא:

$H_0: \beta_{12}=0$

$H_1: \beta_{12}=0$

$$\underline{\lambda}=2\sum_{i=1}^{n}Y_i\hat\beta_{12}X_{1i}X_{2i}+\log\left(\frac{1+e^{\hat\beta_0+\hat\beta_1 X_{1i}+\hat\beta_2 X_{2i}}}{1+e^{\hat\beta_0+\hat\beta_1 X_1+\hat\beta_2 X_2+\hat\beta_{12}X_{1j}X_{2i}}}\right)$$

$\underline{R_0R}$: $\lambda>\chi^2_{1,\alpha}$

5 ג) המימד של S עבור תצפית אחת: $p\times1$

מימד של S עבור n תצפיות: $n\times p$

מימד האינפורמציה של פישר: $p\times p$

$$s(x;\beta)=\frac{\partial}{\partial\beta}\log f(x_i\beta)=\frac{\partial}{\partial\beta}\left[Y_i\log\left(\frac{\pi_i}{1-\pi_i}\right)+\log(1-\pi_i)\right]=$$ ב

$$\frac{\partial}{\partial\beta}\left[Y_i\beta^TX-\log\left(1+e^{\beta^TX}\right)\right]=Y_iX_i-\frac{X_ie^{\beta^TX_i}}{1+e^{\beta^TX_i}}=$$

$$\frac{\partial f(\beta)}{\partial\beta}=\left(\frac{e^{\beta^TX_i}\cdot X_i^T\cdot(1+e^{\beta^TX_i})-e^{\beta^TX_i}e^{\beta^TX_i}X_i^T}{(1+e^{\beta^TX})^2}\right)X_i=\frac{e^{\beta^TX_i}\cdot X_i^T}{(1+e^{\beta^TX})^2}\cdot X_i$$

$$=\frac{e^{\beta^TX_i}}{1+e^{\beta^TX_i}}\cdot\frac{X_i}{1+e^{\beta^TX_i}}\cdot X_i^T=\pi_1\cdot(1-\pi_1)\cdot X_1\cdot X^T$$

$$\frac{\partial S(X;\beta)}{\partial \beta} = -\pi_1 \cdot (1-\pi_1) \cdot X_1 \cdot X_1^T$$

$$\Rightarrow I(\beta) = \pi_1(1-\pi_1) \cdot X_1 \cdot X_1^T$$

$$I_n(\beta) = \sum_i \pi_i(1-\pi_i) X_i X_i^T = X^T \begin{pmatrix} \pi_1(1-\pi_1) & & \\ & \ddots & \\ & & \pi_n(1-\pi_n) \end{pmatrix} X \quad 6$$

$$Var(\hat{\beta}) = I_n(\hat{\beta})^{-1} = (X^T \hat{V} X)^{-1}$$

שאלה 6

	Result
Logistic Regression Model Equation	y = e ^ (b0 + x1 * b_age + x2 * b_obesity + x3 * b_alcohol) / (1 + e ^ (b0 + x1 * b_age + x2 * b_obesity + x3 * b_alcohol))
Beta Estimator	[-9.90606, -0.02169, 0.01961, 0.20531]
Log Likelihood Value for Beta Estimator	-6.827
Variance of Beta Estimator	[[58.75593, -1.17187, -0.026, -0.52854], [-1.17187, 0.05356, 0.0007, -0.0055], [-0.026, 0.0007, 0.00074, -0.00018], [-0.52854, -0.0055, -0.00018, 0.0136]]
Forecast for Expected Value of new observation	0.483
CI for Expected Value Forecast of new observation	[0.26496705546448235, 0.8792969465547116]