52817 – מודלים סטטיסטיים ויישומיהם

סיכום הרצאותיו של פרופ' דוד צוקר (david.zucker@mail.huji.ac.il)
סתו תשע"ז, האוניברסיטה העברית

סובם ע"י גיא עשירי-פרוסנר, guy.ashiri@mail.huji.ac.il, אין לסגל או לי שום אחריות על מה שכתוב פה, אם יש טעויות – זה קורה, אם משהו נראה לא הגיוני אז תחשבו למה, אם יש שקרים אז...

מבנה הציון: 80% מבחן, 20% בחנים (לא מגן). חובת הגשה של 70% מהתרגילים.

תוכן עניינים

הקדמה

סקירה כללית של נושאי הקורס

1. ניתוח שונות חד כיווני – הכללה של מבחן t על שני מדגמים בלתי תלויים. אם מבחן t עוסק בהשוואה בין שתי קבוצות בלתי תלויות, ניתוח שונות חד-כווני הוא השוואה בין i קבוצות. לדוגמא, מדדו תפקוד כליות y_i על $i = 4$ קבוצות שקיבלו מינונים שונים של תרופה.
2. ניתוח שונות דו-כווני – נניח שיודעים שהאוכלוסיה בדוגמא הקודמת היא של חולי סוכרת, חלקם סוג 1 וחלקם סוג 2
3. ניתוח שונות עם אפקט מקרי – יש לנו i בתי ספר, בכל אחד מודדים y ציון על מבחן יכולות הקריאה. אם מניחים $y_{i,j} = \mu + a_i + \epsilon_{i,j}$ אז השונות $var(a_i) = \sigma_a^2, var(\epsilon_{i,j}) = \sigma_\epsilon^2 \to$ $var(y) = \sigma_a^2 + \sigma_\epsilon^2$
4. רגרסיה לוגיסטית – משתנה חזוי y בינארי מול p משתנים חוזים (רציפים או בינאריים)
5. רגרסיה פואסונית – משתנה חזוי y שלם טבעי מול p משתנים חוזים, כלומר $y_i \sim Pois\left(\lambda\left(x_{1_i}, \dots, x_{p_i}\right)\right)$
6. מודלים לוג-לינארים עבור נתונים איכותיים: נניח שעורכים סקר ובו 3 שאלות הבאות:
   א. האם את\ה תומכ\ת בהשקעה ממשלתית גדולה יותר בחינוך? (כן\לא\לא יודע)
   ב. האם את\ה תומכ\ת בהשקעה ממשלתית גדולה יותר בבריאות? (כן\לא\לא יודע)
   ג. סטטוס חברתי-כלכלי? (נמוך\בינוני\גבוה)

   רוצים לבדוק מה הקשר בין סטטוס כלכלי-חברתי ובין התשובות לשתי השאלות.

רגרסיה לינארית (חזרה)

מודל - $y = X\beta + \epsilon$, מגדירים שגיאה לפי $S(b) = \|y - Xb\|^2$, בוחרים את $\hat{\beta}$ כך שימזער את $S(b)$.

פתרון: $\hat{\beta} = (X^TX)^{-1}X^Ty$, $\hat{y} = X\hat{\beta}$, $P_X = X(X^TX)^{-1}X^T$ היא מטריצת ההטלה של ת"מ הנפרש ע"י עמודות X. שאריות: $e = y - \hat{y}$, אומד לשונות $s^2 = \frac{1}{n-p-1}e^Te$. מקבלים כי $\hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1})$ ו- $\frac{(n-p-1)s^2}{\sigma^2} \sim \chi^2_{n-p-1}$.

כלי מרכזי בקורס רגרסיה היה מבחן F – בחירה בין שתי תתי-קבוצות של עמודות X, אחת של r העמודות הראשונות ואחת של p. $F = \frac{\|\hat{y}-\hat{y}^{(0)}\|^2/(p-r)}{\|y-\hat{y}\|^2/(n-p-1)}$, תחת הנחת האפס מתקיים $F \sim F_{p-r,n-p-1}$. כמו כן $SSE = e^Te, SSE^{(0)} = {e^{(0)}}^T e^{(0)}$.

טענה: תהי X מטריצה בגודל n*d כאשר $n \geq d$ ועמודותיה ב"ת. תהי מטריצה H הפיכה מגודל d*d, נגדיר $Z = XH$ אז $P_Z = P_X$, כלומר שינוי בסיס לא משנה את מטריצת ההטלה.

הוכחה:
$$P_Z = Z(Z^TZ)^{-1}Z^T = XH((XH)^TXH)^{-1}(XH)^T = XH(H^TX^TXH)^{-1}H^TX^T =$$
$$XHH^{-1}X^{-1}X^{T^{-1}}H^{T^{-1}}H^TX^T = XX^{-1}X^{T^{-1}}X^T = X(X^TX)^{-1}X^T$$

ניתוח שונות חד-כווני – One Way ANOVA

נתונות I קבוצות ב"ת, עבור קבוצה i יש לנו נתונים $y_{i_1},\dots y_{i_n}\sim_{iid}N(\mu_i,\sigma^2)$, נרצה לעשות הסקה סטטיסטית. שאלה ראשונה – האם בכלל קיימים הבדלים בין הקבוצות? $H_0{:}\,\mu_1=\dots=\mu_I$. נגדיר $\epsilon_{ij}=y_{ij}-\mu_i$, $\epsilon_{i1j},\dots,\epsilon_{inj}\sim_{iid}N(0,\sigma^2)$, כלומר $y_{ij}=\mu_i+\epsilon_{ij}$.

ניסוח נוסף: $\alpha_i=\mu_i-\mu\,,\mu=\frac{1}{I}\sum_{i=1}^{I}\mu_i$ אז $y_{ij}=\mu+\alpha_i+\epsilon_{ij}$. הערה חשובה: נשים לב כי מתקיים $\sum\alpha=\sum(\mu_i-\mu)=0$, לכן ניסוח שקול של השערת האפס הוא $H_0{:}\,\alpha_1=\dots=\alpha_I$. נסמן את הוקטורים $\beta=\begin{bmatrix}\mu_1\\\mu_2\\\vdots\\\mu_I\end{bmatrix},\eta=\begin{bmatrix}\mu\\\alpha_1\\\vdots\\\alpha_{I-1}\end{bmatrix}$ ונקבל $y=W\beta+\epsilon=X\eta+\epsilon$.

דוגמא: עבור $I=3,n_i=4$, נקבל $\epsilon=\begin{bmatrix}\epsilon_{11}\\\epsilon_{12}\\\epsilon_{13}\\\epsilon_{14}\\\epsilon_{21}\\\epsilon_{22}\\\epsilon_{23}\\\epsilon_{24}\\\epsilon_{31}\\\epsilon_{32}\\\epsilon_{33}\\\epsilon_{34}\end{bmatrix}$, $y=\begin{bmatrix}y_{11}\\y_{12}\\y_{13}\\y_{14}\\y_{21}\\y_{22}\\y_{23}\\y_{24}\\y_{31}\\y_{32}\\y_{33}\\y_{34}\end{bmatrix}$ וכן $E[y]=W\beta=X\eta=\begin{bmatrix}\mu_1\\\mu_1\\\mu_1\\\mu_1\\\mu_2\\\mu_2\\\mu_2\\\mu_2\\\mu_3\\\mu_3\\\mu_3\\\mu_3\end{bmatrix}$.

$$E[y]=X\begin{bmatrix}\mu\\\alpha_1\\\alpha_2\end{bmatrix}=\begin{bmatrix}\mu+\alpha_1\\\mu+\alpha_1\\\mu+\alpha_1\\\mu+\alpha_1\\\mu+\alpha_2\\\mu+\alpha_2\\\mu+\alpha_2\\\mu+\alpha_2\\\mu+\alpha_3\\\mu+\alpha_3\\\mu+\alpha_3\\\mu+\alpha_3\end{bmatrix}=\begin{bmatrix}\mu+\alpha_1\\\mu+\alpha_1\\\mu+\alpha_1\\\mu+\alpha_1\\\mu+\alpha_2\\\mu+\alpha_2\\\mu+\alpha_2\\\mu+\alpha_2\\\mu-\alpha_1-\alpha_2\\\mu-\alpha_1-\alpha_2\\\mu-\alpha_1-\alpha_2\\\mu-\alpha_1-\alpha_2\end{bmatrix}=\underbrace{\begin{bmatrix}1&1&0\\1&1&0\\1&1&0\\1&1&0\\1&0&1\\1&0&1\\1&0&1\\1&0&1\\1&-1&-1\\1&-1&-1\\1&-1&-1\\1&-1&-1\end{bmatrix}}_{X}\underbrace{\begin{bmatrix}\mu\\\alpha_1\\\alpha_2\end{bmatrix}}_{\eta}\to X_1=\begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix},X_2=\begin{bmatrix}1&0\\1&0\\1&0\\1&0\\0&1\\0&1\\0&1\\0&1\\-1&-1\\-1&-1\\-1&-1\\-1&-1\end{bmatrix}$$

$$\eta_1=[\mu],\eta_2=\begin{bmatrix}\alpha_1\\\alpha_2\end{bmatrix},H_0{:}\,\eta_2=0\to F=\frac{\frac{\left\|\hat{y}-\hat{y}^{(0)}\right\|^2}{3-1}}{\frac{\|y-\hat{y}\|^2}{12-3}},\hat{y}=P_Xy,\hat{y}_0=P_{X_1}y=X_1(X_1^TX_1)^{-1}X_1^Ty$$

$$X_1^TX_1=[1\dots1]\begin{bmatrix}1\\\vdots\\1\end{bmatrix}=12\to(X_1^TX_1)^{-1}=\frac{1}{12},X_1^Ty=[1\dots1]\begin{bmatrix}y_{11}\\\vdots\\y_{34}\end{bmatrix}=\sum_{i=1}^{3}\sum_{j=1}^{4}y_{ij}$$

$$\to(X_1^TX_1)^{-1}X_1^Ty=\frac{1}{12}\sum_{i=1}^{3}\sum_{j=1}^{4}y_{ij}=\bar{y}\to P_{X_1}y=\begin{bmatrix}1\\\vdots\\1\end{bmatrix}\bar{y}_{..}=\begin{bmatrix}\bar{y}_{..}\\\vdots\\\bar{y}_{..}\end{bmatrix}$$

באותו אופן נחשב עבור $E[y] = W\beta$:

$$E[y] = \overbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}}^{W} \overbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}}^{\beta} = W\beta = WG\eta = W \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} \underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\eta} = \begin{bmatrix} \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu - \alpha_1 - \alpha_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}}_{\beta}$$

$$\to X\eta = E[y] = WG\eta \to \left.\begin{matrix} X = WG \\ W = XH \\ H = G^{-1} \end{matrix}\right\} \to P_W = P_X$$

$$W^T W = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

$$\to (W^T W)^{-1} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}, W^T y = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_{11} \\ \vdots \\ y_{34} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{4} y_{1_i} \\ \sum_{i=1}^{4} y_{2_i} \\ \sum_{i=1}^{4} y_{3_i} \end{bmatrix} \to \hat{\beta} = (W^T W)^{-1} W^T y = \begin{bmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \\ \bar{y}_{3.} \end{bmatrix} \to \hat{y} = \begin{bmatrix} \bar{y}_{1.} \\ \vdots \\ \bar{y}_{1.} \\ \bar{y}_{2.} \\ \vdots \\ \bar{y}_{2.} \\ \bar{y}_{3.} \\ \vdots \\ \bar{y}_{3.} \end{bmatrix}$$

בעת נחשב את המונים עבור מבחן F: $F = \frac{\frac{\left\|\hat{y} - \hat{y}^{(0)}\right\|^2}{2}}{\frac{\|y - \hat{y}\|^2}{9}}$

$$\hat{y}-\hat{y}^{(0)}=\begin{bmatrix}\bar{y}_{1.}-\bar{y}_{..}\\ \vdots\\ \bar{y}_{1.}-\bar{y}_{..}\\ \bar{y}_{2.}-\bar{y}_{..}\\ \vdots\\ \bar{y}_{2.}-\bar{y}_{..}\\ \bar{y}_{3.}-\bar{y}_{..}\\ \vdots\\ \bar{y}_{3.}-\bar{y}_{..}\end{bmatrix}\rightarrow\left\|\hat{y}-\hat{y}^{(0)}\right\|^2=4(\bar{y}_{1.}-\bar{y}_{..})^2+4(\bar{y}_{2.}-\bar{y}_{..})^2+4(\bar{y}_{3.}-\bar{y}_{..})^2$$

$$=4\sum_{i=1}^{3}(\bar{y}_{i.}-\bar{y}_{..})^2,\|y-\hat{y}\|^2=\sum_{i=1}^{3}\sum_{j=1}^{4}\left(\bar{y}_{ij}-\bar{y}_{i.}\right)^2$$

באופן כללי נקבל:

$$(W^TW)^{-1}=\begin{bmatrix}n_1^{-1} & & 0\\ & \ddots & \\ 0 & & n_I^{-1}\end{bmatrix},W^Ty=\begin{bmatrix}\sum_{j=1}^{n_1}y_{1_j}\\ \vdots\\ \sum_{j=1}^{n_I}y_{I_j}\end{bmatrix},\hat{\beta}=\begin{bmatrix}\bar{y}_{1.}\\ \vdots\\ \bar{y}_{I.}\end{bmatrix}$$

$$N=\sum_{i=1}^{I}n_i,\bar{y}_{..}=\frac{1}{N}\sum_{i=1}^{3}\sum_{j=1}^{4}y_{ij},\hat{y}^{(0)}=\begin{bmatrix}\bar{y}_{..}\\ \vdots\\ \bar{y}_{..}\end{bmatrix}$$

$$\left\|\hat{y}-\hat{y}^{(0)}\right\|^2=\sum_{i=1}^{I}n_i(\bar{y}_{i.}-\bar{y}_{..})^2,\|y-\hat{y}\|^2=\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\bar{y}_{ij}-\bar{y}_{i.}\right)^2$$

$$F=\frac{\frac{\left\|\hat{y}-\hat{y}^{(0)}\right\|^2}{I-1}}{\frac{\|y-\hat{y}\|^2}{N-I}}$$

פירוק לריבועים

$$\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\bar{y}_{ij}-\bar{y}_{..}\right)^2}_{SST}=\underbrace{\sum_{i=1}^{I}n_i(\bar{y}_{i.}-\bar{y}_{..})^2}_{SSB\backslash SSR}+\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\bar{y}_{ij}-\bar{y}_{i.}\right)^2}_{SSW\backslash SSE}$$

SSE: Sum of Squares Total, SSB: Between, SSW: Within, SSR: Regression, SSE: Errors

הוכחה: $\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\bar{y}_{ij}-\bar{y}_{..}\right)^2=\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\bar{y}_{ij}-\bar{y}_{i.}+\bar{y}_{i.}-\bar{y}_{..}\right)^2=\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\left(\bar{y}_{ij}-\bar{y}_{i.}\right)^2+\right.$
$\left.(\bar{y}_{i.}-\bar{y}_{..})^2+2\left(\bar{y}_{ij}-\bar{y}_{i.}\right)(\bar{y}_{i.}-\bar{y}_{..})\right)=\underbrace{\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\bar{y}_{ij}-\bar{y}_{i.}\right)^2}_{SSE}+\underbrace{\sum_{i=1}^{I}n_i(\bar{y}_{i.}-\bar{y}_{..})^2}_{SSB}+$
$2\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\bar{y}_{ij}-\bar{y}_{i.}\right)(\bar{y}_{i.}-\bar{y}_{..})$

כאשר מתקיים $\sum_{i=1}^{I}\sum_{j=1}^{n_i}\left(\bar{y}_{ij}-\bar{y}_{i.}\right)(\bar{y}_{i.}-\bar{y}_{..})=\sum_{i=1}^{I}\left((\bar{y}_{i.}-\bar{y}_{..})\underbrace{\sum_{j=1}^{n_i}\left(\bar{y}_{ij}-\bar{y}_{i.}\right)}_{=0}\right)=0$

לוח ANOVA:

	Degrees of Freedom	Sum of Squares	Mean Square	F
Groups	$I-1$	SSB	$MSB=\frac{SSB}{I-1}$	F
Error	$N-I$	SSE	$MSE=\frac{SSE}{N-I}$	
Total	$N-1$	SST		

קונטרסטים

שוב בדוגמת מינוני התרופה, נניח $\mu_1=0,\mu_2=50,\mu_3=100,\mu_4=200$. יהי $c\in\mathbb{R}^I$ המקיים $\sum_{i=1}^I c_i=0$, נסמן קונטרסט (צ"ל של התוחלות) באופן $\psi(c)=\sum_{i=1}^I c_i\mu_i=c^T\beta$, לדוגמא:

$$\mu_4-\mu_1=\begin{bmatrix}-1 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \mu_4\end{bmatrix},\frac{\mu_2+\mu_3+\mu_4}{3}-\mu_1=\begin{bmatrix}-1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\\ \mu_4\end{bmatrix}$$

אם $\psi(c)=c^T\beta$ אז $\hat\psi(c)=c^T\hat\beta=\sum_{i=1}^I c_i\bar y_{i.}$, $E[\hat\psi(c)]=c^TE[\hat\beta]=c^T\beta=\psi(c)$, כלומר אח"ה.

$$Var\left(\hat\psi(c)\right)=Var(c^T\hat\beta)=c^Tcov(\hat\beta)c=c^T\left(\sigma^2(\mathrm{W^TW})^{-1}\right)c=c^T\left(\sigma^2\begin{bmatrix}n_1^{-1} & & 0\\ & \ddots & \\ 0 & & n_I^{-1}\end{bmatrix}\right)c.$$

אם ניקח לדוגמא $I=4$ אז נקבל:

$$Var\left(\hat\psi(c)\right)=c^T\left(\sigma^2\begin{bmatrix}n_1^{-1} & & 0\\ & \ddots & \\ 0 & & n_I^{-1}\end{bmatrix}\right)c=\begin{bmatrix}c_1 & c_2 & c_3 & c_4\end{bmatrix}\sigma^2\begin{bmatrix}n_1^{-1} & & 0\\ & \ddots & \\ 0 & & n_I^{-1}\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ c_3\\ c_4\end{bmatrix}$$
$$=\sigma^2\left(\frac{c_1^2}{n_1}+\cdots+\frac{c_4^2}{n_4}\right)\to Var\left(\hat\psi(c)\right)=\sigma^2V(c)$$

תחת הנחת נורמליות, נקבל $\hat\psi(c)\sim N\left(\psi(c),\sigma^2V(c)\right)$. ידוע כי סטטיסטי Z יהיה $Z(c)=\frac{\hat\psi(c)-\psi(c)}{\sigma\sqrt{V(c)}}\sim N(0,1)$, אם נחליף באומד לשונות נקבל $T(c)=\frac{\hat\psi(c)-\psi(c)}{s\sqrt{V(c)}}=\frac{\frac{\hat\psi(c)-\psi(c)}{\sigma\sqrt{V(c)}}}{\frac{s}{\sigma}}=\frac{Z(c)}{\sqrt{\frac{s^2}{\sigma^2}}}$, השבר מתפלג בצורה $\frac{\sim N(0,1)}{\sqrt{\sim\frac{\chi^2_{N-I}}{N-I}}}$ כלומר בסך הכל $T(c)\overset{H_0}{\sim}t_{N-I}$.

נרצה לבנות רווח סמך עבור $\psi(c)$ בצורה $\hat\psi(c)\pm A$ כך שיתקיים $P\left(\psi(c)\in[\hat\psi(c)\pm A]\right)=1-\alpha$

$$P\left(\psi(c)\in[\hat\psi(c)\pm A]\right)=P\left(\hat\psi(c)-A\le\psi(c)\le\hat\psi(c)+A\right)=P\left(-A\le\psi(c)-\hat\psi(c)\le A\right)$$
$$=P\left(|\psi(c)-\hat\psi(c)|\le A\right)=P\left(|\hat\psi(c)-\psi(c)|\le A\right)$$
$$=P\left(\left|\frac{\hat\psi(c)-\psi(c)}{s\sqrt{V(c)}}\right|\le\frac{A}{s\sqrt{V(c)}}\right)=P\left(|T(c)|\le\frac{A}{s\sqrt{V(c)}}\right)$$
$$\to P\left(|T(c)|\le t_{N-I,1-\frac{\alpha}{2}}\right)=1-\alpha\to\frac{A}{s\sqrt{V(c)}}=t_{N-I,1-\frac{\alpha}{2}}\to A$$
$$=s\sqrt{V(c)}t_{N-I,1-\frac{\alpha}{2}}$$

כך רווח סמך ברמה $1-\alpha$ עבור $\psi(c)$ יהיה $\hat\psi(c)\pm s\sqrt{V(c)}t_{N-I,1-\frac{\alpha}{2}}$.

השוואות מרובות

בכל שמבצעים מספר רב יותר של מבחנים (נניח k), יש הסתברות גדולה יותר לטעות לפחות באחד מהם – אם הניסויים בלתי תלויים אז ההסתברות שלא לטעות היא $(1-\alpha)^k$ וההסתברות לטעות היא $1-(1-\alpha)^k$:

K, $\alpha = 0.05$	$(1-\alpha)^k$	$1-(1-\alpha)^k$
1	0.95	0.05
2	0.9025	0.0975
3	0.857375	0.142625

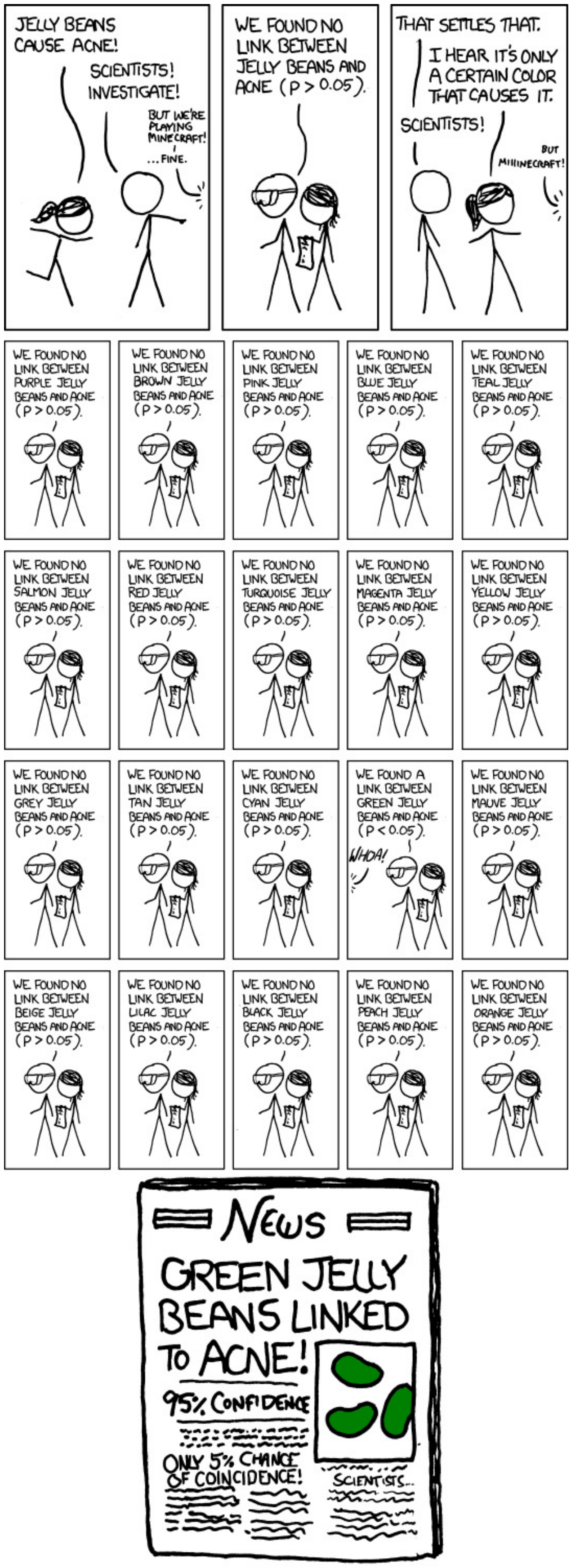
בעמוד הבא: קומיקס של xkcd! שהמרצה הציג!! גם ציין אמרה באנגלית If you torture the data enough, they'll eventually confess to something.

נבחין בין מספר מצבים:

1. עבור מספר סופי k של קונטרסטים קבועים מראש, $\psi_k = \psi\left(c^{(k)}\right)$
2. כל ההשוואות הזוגיות $\Delta_{rs} = \mu_r - \mu_s$
3. כל הקונטרסטים האפשריים

נרצה דרכים לבניית רווח סמך עבור בדיקה מרובה, כך שההסתברות כי הרווח S(c) מכיל את הפרמטר $\psi(c)$ לכל $c \in C$) היא לפחות $1-\alpha$. נדון בשיטות המתאימות:

1. שיטת בונפרוני (Bonferroni) והשיטה המדויקת
2. שיטת טוקי (Tukey)
3. שיטת שפה (Scheffé)



JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
...FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

שיטת בונפרוני Bonferroni

נסמן מאורעות באופן $A_k = \{\psi(c^{(k)}) \in S(c^{(k)})\}, B_k = A_k^c = \{\psi(c^{(k)}) \notin S(c^{(k)})\}$, נרצה $P\left(\bigcap_{k=1}^{K} A_k\right) \geq 1-\alpha$, כלומר $P\left(\left(\bigcap_{k=1}^{K} A_k\right)^c\right) \leq \alpha$ ולפי חוקי דה-מורגן $P\left(\bigcup_{k=1}^{K} B_k\right) \leq \alpha$. מתקיים גם $P\left(\bigcup_{k=1}^{K} B_k\right) \leq \sum_{k=1}^{K} P(B_k)$, לכן אם נגדיר את רמת הסמך של רווח בודד ל-$\frac{\alpha}{K}$ נקבל רווח סמך $S(c^{(k)}) = \hat{\psi}(c) \pm s\sqrt{V(c^{(k)})}t_{N-I,1-\frac{\alpha/K}{2}}$, שקול ל- $P\left(\psi(c^{(k)}) \in S(c^{(k)})\right) = 1-\frac{\alpha}{K}$ וכמובן גם $P\left(\psi(c^{(k)}) \notin S(c^{(k)})\right) = \frac{\alpha}{K}$.

השיטה המדויקת

נניח $K=3$ אז $\psi(c^{(1)}) = c_1^{(1)}\mu_1 + \cdots + c_I^{(1)}\mu_I$ ובהתאם $\psi(c^{(2)}) = \sum_{i=1}^{I} c_i^{(2)}\mu_i$ ועבור $c^{(3)}$. נסתכל במטריצה $\begin{bmatrix} c_1^{(1)} & \dots & c_I^{(1)} \\ c_1^{(2)} & \dots & c_I^{(2)} \\ c_1^{(3)} & \dots & c_I^{(3)} \end{bmatrix}\begin{bmatrix}\mu_1 \\ \vdots \\ \mu_I\end{bmatrix} = C\beta = \begin{bmatrix}\psi(c^{(1)}) \\ \psi(c^{(2)}) \\ \psi(c^{(3)})\end{bmatrix} = \Psi$, מתקיים $C\hat{\beta} = \begin{bmatrix}\hat{\psi}(c^{(1)}) \\ \hat{\psi}(c^{(2)}) \\ \hat{\psi}(c^{(3)})\end{bmatrix} = \hat{\Psi}$, כמובן גם $E[\hat{\Psi}] = CE[\hat{\beta}] = C\beta = \Psi$ ועבור השוונויות מתקיים:

$$Cov(\hat{\Psi}) = \sigma^2 C \begin{bmatrix} n_1^{-1} & & 0 \\ & \ddots & \\ 0 & & n_I^{-1}\end{bmatrix} C^T = \sigma^2 \begin{bmatrix} c_1^{(1)} & \dots & c_I^{(1)} \\ c_1^{(2)} & \dots & c_I^{(2)} \\ c_1^{(3)} & \dots & c_I^{(3)} \end{bmatrix}\begin{bmatrix} n_1^{-1} & & 0 \\ & \ddots & \\ 0 & & n_I^{-1}\end{bmatrix}\begin{bmatrix} c_1^{(1)} & c_I^{(2)} & c_I^{(3)} \\ \vdots & \vdots & \vdots \\ c_I^{(1)} & c_I^{(2)} & c_I^{(3)} \end{bmatrix}$$

$$\rightarrow \left(Cov(\hat{\Psi})\right)_{rs} = \sigma^2 \sum_{i=1}^{I} \frac{c_i^{(r)} c_i^{(s)}}{n_i} = \sigma^2 H_{rs}$$

נסתכל על $\hat{\Psi} - \Psi \sim N(0, \sigma^2 H)$ אז $\Omega = \begin{bmatrix}\sqrt{V(C^{(1)})} & & 0 \\ & \ddots & \\ 0 & & \sqrt{V(C^K)}\end{bmatrix} = \begin{bmatrix}\sqrt{H_{11}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{H_{KK}}\end{bmatrix}$, נסמן $Z = \frac{1}{\sigma}\Omega^{-1}(\hat{\Psi} - \Psi) \sim N(0, \Omega^{-1}H\Omega^{-1})$, $z_k = Z(c^{(k)}) = \frac{\hat{\psi}(c^{(k)}) - \psi(c^{(k)})}{\sigma\sqrt{V(c^{(k)})}} = \frac{\hat{\psi}(c^{(k)}) - \psi(c^{(k)})}{\sigma\sqrt{H_{kk}}}$. עבור $K=3$ המטריצה $\Omega^{-1}H\Omega^{-1}$ תקיים:

$$\Omega^{-1}H\Omega^{-1} = \begin{bmatrix} H_{11}^{-\frac{1}{2}} & & 0 \\ & H_{22}^{-\frac{1}{2}} & \\ 0 & & H_{33}^{-\frac{1}{2}}\end{bmatrix}\begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33}\end{bmatrix}\begin{bmatrix} H_{11}^{-\frac{1}{2}} & & 0 \\ & H_{22}^{-\frac{1}{2}} & \\ 0 & & H_{33}^{-\frac{1}{2}}\end{bmatrix} =$$

$$= \begin{bmatrix} H_{11}^{-\frac{1}{2}}H_{11} & H_{11}^{-\frac{1}{2}}H_{12} & H_{11}^{-\frac{1}{2}}H_{13} \\ H_{22}^{-\frac{1}{2}}H_{21} & H_{22}^{-\frac{1}{2}}H_{22} & H_{22}^{-\frac{1}{2}}H_{23} \\ H_{33}^{-\frac{1}{2}}H_{31} & H_{33}^{-\frac{1}{2}}H_{32} & H_{33}^{-\frac{1}{2}}H_{33}\end{bmatrix}\begin{bmatrix} H_{11}^{-\frac{1}{2}} & & 0 \\ & H_{22}^{-\frac{1}{2}} & \\ 0 & & H_{33}^{-\frac{1}{2}}\end{bmatrix} =$$

$$=\begin{bmatrix} H_{11}^{-\frac{1}{2}}H_{11}^{-\frac{1}{2}}H_{11} & H_{22}^{-\frac{1}{2}}H_{11}^{-\frac{1}{2}}H_{12} & H_{33}^{-\frac{1}{2}}H_{11}^{-\frac{1}{2}}H_{13} \\ H_{11}^{-\frac{1}{2}}H_{22}^{-\frac{1}{2}}H_{21} & H_{22}^{-\frac{1}{2}}H_{22}^{-\frac{1}{2}}H_{22} & H_{33}^{-\frac{1}{2}}H_{22}^{-\frac{1}{2}}H_{23} \\ H_{11}^{-\frac{1}{2}}H_{33}^{-\frac{1}{2}}H_{31} & H_{22}^{-\frac{1}{2}}H_{33}^{-\frac{1}{2}}H_{32} & H_{33}^{-\frac{1}{2}}H_{33}^{-\frac{1}{2}}H_{33} \end{bmatrix} = \begin{bmatrix} 1 & \frac{H_{12}}{\sqrt{(H_{11}H_{22})}} & \frac{H_{13}}{\sqrt{(H_{11}H_{33})}} \\ \frac{H_{21}}{\sqrt{(H_{22}H_{11})}} & 1 & \frac{H_{23}}{\sqrt{(H_{22}H_{33})}} \\ \frac{H_{31}}{\sqrt{(H_{33}H_{11})}} & \frac{H_{32}}{\sqrt{(H_{33}H_{22})}} & 1 \end{bmatrix}$$

ובסך הכל $R_{rs}=\frac{H_{rs}}{\sqrt{H_{rr}H_{ss}}}$, $R=\Omega^{-1}H\Omega^{-1}$, זו הקורלציה של z_r, z_s. נגדיר $T=\frac{1}{s}\Omega^{-\frac{1}{2}}(\widehat{\Psi}-\Psi)=\frac{Z}{\sqrt{\frac{s^2}{\sigma^2}}}$, מתפלג רב-t עם מטריצת מתאם R ו-$N-I$ דרגות חופש.

נניח שמצאנו ערך t^* כך ש-$P(|T_k|\le t^*,\forall k)=1-\alpha$ אז $P\left(\left|\frac{\hat\psi(c^{(k)})-\psi(c^{(k)})}{s\sqrt{V(c^{(k)})}}\right|\le t^*,\forall k\right)=1-\alpha$
וזה שקול אלגברית ל- $P\left(\hat\psi(c^{(k)})-t^*s\sqrt{V(c^{(k)})}\le\psi(c^{(k)})\le\hat\psi(c^{(k)})+t^*s\sqrt{V(c^{(k)})}\right)=1-\alpha$
עבור כל k, זה אומר שאם נגדיר $S(c^{(k)})=[S_L(c^{(k)}),S_U(c^{(k)})]$ אז קצות רווח הסמך יהיו $S_L(c^{(k)})=\hat\psi(c^{(k)})-t^*s\sqrt{V(c^{(k)})}, S_U(c^{(k)})=\hat\psi(c^{(k)})+t^*s\sqrt{V(c^{(k)})}$ ואז עבור כל k יתקיים $P\left(\psi(c^{(k)})\in S(c^{(k)})\right)=1-\alpha$, את t^* ניתן (ועדיף) למצוא נומרית (חבילת mvtnorm ב-R).

אלגוריתם פתרון לשיטה המדויקת: מגדירים $R_{rs}=\frac{\sum_{i=1}^{I}\frac{c_i^{(r)}c_i^{(s)}}{n_i}}{\sqrt{V(c^{(r)})V(c^{(s)})}}$, מוצאים את הערך של t^* כך שיתקיים $P(|R|\le t^*)=1-\alpha$, בונים רווחי סמך.

שיטת טוקי Tukey

בדיקה בזוגות: נסמן $\Delta_{ik}=\mu_i-\mu_k$, $\widehat{\Delta}_{ik}=\hat\mu_i-\hat\mu_k=\bar y_{i.}-\bar y_{k.}$, יש $\binom{I}{2}=\frac{I(I-1)}{2}$ כאלו. נגדיר $S_{ik}=[S_{ik}^{(L)},S_{ik}^{(U)}]$ אז $S_{ik}^{(L)}=\widehat\Delta_{ik}-A, S_{ik}^{(U)}=\widehat\Delta_{ik}+A$ ורוצים למצוא את הערך A כך שיתקיים $P(\Delta_{ik}\in S_{ik})=P\left(\widehat\Delta_{ik}-A\le\Delta_{ik}\le\widehat\Delta_{ik}+A\right)=1-\alpha$, נדרוש $P\left(\widehat\Delta_{ik}-A\le\Delta_{ik}\le\widehat\Delta_{ik}+A,\forall i,k\right)=P\left(-A\le\Delta_{ik}-\widehat\Delta_{ik}\le A,\forall i,k\right)=P\left(\left|\Delta_{ik}-\widehat\Delta_{ik}\right|\le A,\forall i,k\right)=P\left(\left|\widehat\Delta_{ik}-\Delta_{ik}\right|\le A,\forall i,k\right)$. באופן כללי $(\bar y_{i.}-\mu_i)\sim N\left(0,\frac{\sigma^2}{n_i}\right)$. נניח כי $\forall i: n_i=n$ אז $\widehat\Delta_{ik}-\Delta_{ik}=(\bar y_{i.}-\bar y_{k.})-(\hat\mu_i-\hat\mu_k)=(\bar y_{i.}-\hat\mu_i)-(\bar y_{k.}-\hat\mu_k)=\sqrt{\frac{\sigma^2}{n}}\left(\frac{(\bar y_{i.}-\hat\mu_i)}{\sqrt{\frac{\sigma^2}{n}}}-\frac{(\bar y_{k.}-\hat\mu_{k.})}{\sqrt{\frac{\sigma^2}{n}}}\right)$, נסמן $\phi=\frac{A}{\frac{s}{\sqrt n}}\to A=\phi\sqrt{\frac{s^2}{n}}$ וכן $z_i=\frac{(\bar y_{i.}-\hat\mu_i)}{\sqrt{\frac{\sigma^2}{n}}}$ אז נקבל עבור ההפרש $\widehat\Delta_{ik}-\Delta_{ik}=\sqrt{\frac{\sigma^2}{n}}(z_i-z_k)$, נציב אותם בנוסחא שדרשנו להסתברות:

$$P\left(\sqrt{\frac{\sigma^2}{n}}|z_i - z_k| \le A, \forall i,k\right) = P\left(\sqrt{\frac{\sigma^2}{n}}|z_i - z_k| \le \phi\sqrt{\frac{s^2}{n}}, \forall i,k\right)$$
$$= P\left(\frac{\sqrt{\frac{\sigma^2}{n}}}{\sqrt{\frac{s^2}{n}}}|z_i - z_k| \le \phi, \forall i,k\right) = P\left(\frac{|z_i - z_k|}{\sqrt{\frac{s^2}{\sigma^2}}} \le \phi, \forall i,k\right)$$
$$= P\left(\frac{\max_{i,k}|z_i - z_k|}{\sqrt{\frac{s^2}{\sigma^2}}} \le \phi\right) = P(Q \le \phi) \to Q = \frac{\max_i z_i - \min_i z_i}{\sqrt{\frac{s^2}{\sigma^2}}}$$

בסך הכל נקבל $P\left(\widehat{\Delta}_{ik} - A \le \Delta_{ik} \le \widehat{\Delta}_{ik} + A, \forall i,k\right) = P(Q \le \phi), \phi = \frac{A}{\frac{s}{\sqrt{n}}}$, המכנה של Q מתפלג $\sim\chi^2_{N-I}$, נרצה למצוא את ההתפלגות של Q כולו.

נניח כי $\xi_1, \dots, \xi_m \sim N(0,1)$ וכן $W \sim \chi^2_p$ ב"ת, נגדיר $Q = \frac{\max_i \xi_i - \min_i \xi_i}{\frac{W}{p}}$ אז $P(Q \le q) = \int_0^\infty \int_{-\infty}^\infty m[\Phi(y) - \Phi(y - qv)]\phi(y) g_p(v) dg\, dv$, זו התפלגות Studentized Range עם פרמטרים m,p (הפקודה ב-R היא ptukey).

בסך הכל נקבל כי Q מפולג Studentized Range עם $I, N-I$. כדי להשיג עבור ההסתברות שדרשנו $P\left(\widehat{\Delta}_{ik} - A \le \Delta_{ik} \le \widehat{\Delta}_{ik} + A, \forall i,k\right) = 1 - \alpha$ צריך לקחת $\phi = q_{I,N-I}^{(1-\alpha)}$ כאשר $q_{I,N-I}^{(1-\alpha)}$ הוא הערך עבורו מתקיים $P\left(Q \le q_{I,N-I}^{(1-\alpha)}\right) = 1 - \alpha$. נקבל כי $A = \phi\frac{s}{\sqrt{n}} = \frac{s}{\sqrt{n}} q_{I,N-I}^{(1-\alpha)}$ ורווח סמך ברמה $1 - \alpha$ עבור Δ_{ik} יהיה $\widehat{\Delta}_{ik} \pm \frac{s}{\sqrt{n}} q_{I,N-I}^{(1-\alpha)}$. נוכל גם לסמן $\Delta_{ik} = c^T\beta$ כאשר $c_n = \begin{cases} 1 & n = i \\ -1 & n = k \\ 0 & else \end{cases}$, נקבל $V(c) = 2$ וכן רווח סמך $\widehat{\Delta}_{ik} \pm \frac{s\sqrt{V(c)}}{\sqrt{2}} q_{I,N-I}^{(1-\alpha)}$.

מה עושים אם n_i שונים?

אפשרות 1 – רווח Tukey-Kramer: $\widehat{\Delta}_{ik} \pm q_{I,N-I}^{(1-\alpha)} s \sqrt{\frac{1}{2}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}$.

אפשרות 2 – לנסח את השאלה במונחים של $K = \binom{I}{2}$ קונטרסטים ולבנות רווחים בשיטה המדויקת.

אומדים ורווחי סמך אי-פרמטריים עבור $\Delta_{ik} = \mu_i - \mu_k$

אומד Hodges-Lehmann: נסתכל על המספרים $D_{rs}^{(i,k)} = y_{ir} - y_{ks}, r = 1{:}n_i, s = 1{:}n_k$, $\widehat{\Delta}_{ik}$ יהיה חציון של $n_i \cdot n_k$ הערכים האלו. נחשב אינדקס $m = \left\lfloor \frac{n_i n_k}{2} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{n_i n_k(n_i + n_k + 1)}{12}} - \frac{1}{2} \right\rfloor$, ר"ס יהיה $\left[D_m, D_{n_1 n_2 + 1 - m}\right]$. אם מחשבים ר"ס בו"ז בשיטת Tukey אז מחליפים את $z_{1-\frac{\alpha}{2}}$ ב-$\frac{q_{I,\infty}(1-\alpha)}{\sqrt{2}}$.

שיטת שפה Scheffé

רוצים כיסוי בו-זמני על כל הקונטרסטים האפשריים, רווח סמך ברמה $1-\alpha$ יהיה נתון בצורה $\hat{\psi}(c) \pm \xi s\sqrt{V(c)}$ כאשר $\xi = \sqrt{(I-1)F_{I-1,N-I}^{(1-\alpha)}}$. הוכחה:

$\left|\hat{\psi}(c)-\psi(c)\right| = \left|\sum_{i=1}^{I} c_i\bar{y}_{i.} - \sum_{i=1}^{I} c_i\mu_i\right| = \left|\underbrace{\sum_{i=1}^{I} c_i\bar{y}_{i.}^*}_{y_{ij}^*=y_{ij}-\mu_i}\right| = \left|\sum_{i=1}^{I} c_i(\bar{y}_{i.}^*-\bar{y}_{..}^*)\right|$, כיוון שמתקיים $\sum_i c_i = 0$ נוכל לסמן $\left|\sum_{i=1}^{I} c_i(\bar{y}_{i.}^*-\bar{y}_{..}^*)\right| = \left|\sum_{i=1}^{I} v_iw_i\right|$ כאשר $v_i = \sqrt{n_i}(\bar{y}_{i.}^*-\bar{y}_{..}^*), w_i = \frac{c_i}{\sqrt{n_i}}$ ומתקיים $\|w\|^2 = V(c)$. אם נסמן $Q = \|v\|^2 = \sum_{i=1}^{I} n_i(\bar{y}_{i.}^*-\bar{y}_{..}^*)^2$ אז נקבל מא"ש קושי-שוורץ:

$$\left|\sum_{i=1}^{I} v_iw_i\right| \le \|v\|\cdot\|w\| \to \frac{\left|\sum_{i=1}^{I} v_iw_i\right|}{\|v\|} \le \|w\| \to \max_{c\in C}\frac{\left|\hat{\psi}(c)-\psi(c)\right|}{\sqrt{V(c)}} \le \max_{c\neq 0}\frac{\left|\hat{\psi}(c)-\psi(c)\right|}{\sqrt{V(c)}}$$

ואם נסמן $H = \frac{Q/(I-1)}{s} = \frac{\sum_{i=1}^{I} n_i\left(\bar{y}_{i.}^*-\bar{y}_{..}^*\right)^2/(I-1)}{s}$ אז $\max_{c\in C}\frac{\left|\hat{\psi}(c)-\psi(c)\right|}{s\sqrt{(I-1)V(c)}} \le \sqrt{H}$ כאשר אומד השאריות מקיים $s^2 = \frac{1}{N-I}\sum_i\sum_j\left(y_{ij}-\bar{y}_{i.}\right)^2 = \frac{1}{N-I}\sum_i\sum_j\left(y_{ij}^*-\bar{y}_{i.}^*\right)^2$. למעשה, H הוא בצורת מבחן F של ניתוח שונות עבור y_{ij}^* במקום y_{ij}. נוכל לרשום $y_{ij}^* = \mu_i^* + \epsilon_{ij}$ כאשר $\forall i: \mu_i^* = 0$ ואז למעשה מתקיימת השערת האפס של מבחן F, כלומר $H \sim F_{I-1,N-I}$ ולכן $P\left(H \le F_{I-1,N-I}^{(1-\alpha)}\right) = 1-\alpha$ ובסך הכל נקבל $P\left(\max_{c\in C}\frac{\left|\hat{\psi}(c)-\psi(c)\right|}{s\sqrt{(I-1)V(c)}} \le \sqrt{F_{I-1,N-I}^{(1-\alpha)}}\right) \ge P\left(\sqrt{H} \le \sqrt{F_{I-1,N-I}^{(1-\alpha)}}\right) = 1-\alpha$.

הנחות מודל – ניתוח שונות

$$y_{ij} = \mu_i + \epsilon_{ij}, \epsilon_{ij} \sim N(0,\sigma^2)$$

שתי הנחות במודל:

1. התפלגות נורמלית
2. שוויון שונויות בין הקבוצות $1,\dots,I$

נרצה לבדוק האם ההנחות מתקיימות ומה עושים כאשר אינן מתקיימות

שאריות:

$$e_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \bar{y}_{i.} = \left(y_{ij}-\mu_i\right) - \left(\bar{y}_{i.}-\mu_i\right) = \epsilon_{ij} - \left(\bar{y}_{i.}-\mu_i\right) \approx \epsilon_{ij}$$

מה נעשה אם השגיאה לא מפולגת נורמלית?

1. טרנספורמציה על הנתונים ($\frac{1}{y_{ij}}, \sqrt{y_{ij}}, \log\left(y_{ij}\right)$ וכן הלאה)
2. שימוש במשפט הגבול המרכזי באופן $\bar{y}_{i.} \sim N\left(\mu_i, \frac{\sigma^2}{n_i}\right)$, דורש מספר דגימות גדול בכל קבוצה.
3. מבחן Kruskal-Wallis, הכללה של מבחן דרגות Wilcoxon למצב של יותר מאשר שתי קבוצות: נסמן $H = (N-1)\frac{\sum_{i=1}^{I} n_i(\bar{R}_{i.}-\bar{R}_{..})^2}{\sum_i\sum_j\left(R_{ij}-\bar{R}_{i.}\right)^2} \sim \chi_{I-1}^2$, אם בכל הקבוצות n_i גדול. ב-R פונקציית qn.test בחבילת kSamples.

שוויון שונויות

מדדי פיזור: $s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{i.}\right)^2$

Mean Absolute Deviation: $MAD = \frac{1}{n_i}\sum_{j=1}^{n_i}\left|y_{ij} - \bar{y}_{i.}\right|$

Mean Absolute Deviation From the Median: $MADM = \frac{1}{n_i}\sum_{j=1}^{n_i}\left|y_{ij} - median\left(y_{i1}, \dots, y_{in_i}\right)\right|$

מבחן Levene לשוויון שונויות:

$$H_0: \sigma_1^2 = \cdots = \sigma_I^2$$

נסמן $x_{ij} = \left|y_{ij} - \bar{y}_{i.}\right| \approx \left|\epsilon_{ij}\right| = \sigma_i\left|\xi_{ij}\right|$, אז $\xi_{ij} = \frac{\epsilon_{ij}}{\sigma_i} \sim N(0,1)$ ומתקיים $E\left[x_{ij}\right] \approx \sigma_i E\left[\left|\xi_{ij}\right|\right] = \sigma_i\sqrt{\frac{2}{\pi}} = \theta_i$, כלומר ניסוח שקול להשערת האפס הוא $H_0^*: \theta_1 = \cdots = \theta_I$ ואז אפשר לבצע מבחן F של ניתוח שונות חד-כווני על תוחלות x_{ij}.

מה עושים אם דחינו את H_0^*?

1. טרנספורמציה מייצבת שונות, אם $\sigma_i \approx c\mu_i^\alpha$ אז ניתן לבצע על y טרנספורמציה בצורה $\tilde{y}_{ij} = \begin{cases} \log y_{ij} & \alpha = 1 \\ y_{ij}^{1-\alpha} & \alpha \neq 1 \end{cases}$
2. מבחן Welch לניתוח שונות חד-כווני עם שונויות שונות: נניח זמנית כי σ_i^2 ידוע לכל i, נוכל לסמן $\bar{y}_{i.} \sim N\left(\mu_i, \frac{\sigma_i^2}{n_i}\right) \to w_i = Var(\bar{y}_{i.})^{-1} = \frac{n_i}{\sigma_i^2} \to W = \sum_{i=1}^I w_i \to \bar{y}_{..w} = \frac{1}{W}\sum_{i=1}^I w_i\bar{y}_{i.} \to$ $SSB_w = \sum_{i=1}^I w_i(\bar{y}_{i.} - \bar{y}_{..w})^2$, טענה: תחת H_0 מתקיים $SSB_w \sim \chi_{I-1}^2$. נוכל לבצע מבחן F בצורה $F = \frac{SSB_w}{(I-1)C}$ כאשר תחת השערת האפס מתקיים בקירוב $\frac{SSB_w}{(I-1)C} \sim F_{I-1,d}$ כאשר $\pi_i = \frac{w_i}{W}, C = 1 + \frac{2(I+2)}{I^2-1}\sum_{i=1}^I \frac{(1-\pi_i)^2}{n_i-1}, d = \left(\frac{3}{I^2-1}\sum_{i=1}^I \frac{(1-\pi_i)^2}{n_i-1}\right)^{-1}$

ניתוח שונות דו-כווני – Two Way ANOVA

יש לנו מינונים שונים של תרופה (0, 50, 100, 200 מ"ג) ושני סוגים של חולים (סוכרת סוג 1, סוג 2). נאמר כי יש לנו שני גורמים: A מינון התרופה, עם I אפשרויות ו-B סוג המחלה עם J אפשרויות. נבנה טבלא של $I\cdot J$ תאים, בכל אחד מהם יש תצפיות $y_{ij_1},\dots,y_{ij_{n_{ij}}}\sim_{iid}N(\mu_{ij},\sigma^2)$, נניח אי תלות של התצפיות בין התאים השונים. בשלב זה נניח כי $n_{ij}=n$ לכל i,j (המקרה המאוזן):

B

	i \ j	1	2	…	J
A	1				
	2				
	⋮				
	I				

המודל הוא $y_{ijk}=\mu_{ij}+\epsilon_{ijk}$ כאשר $i=1,\dots,I$ רמה של A ו-$j=1,\dots,J$ רמה של B, $k=1,\dots n_{ij}$ תצפית ברמה הנתונה ומתקיים $\epsilon_{ijk}\overset{iid}{\sim}N(0,\sigma^2)$. ניסוח חלופי למודל: $\bar{\mu}_{i.}=\frac{1}{J}\sum_{j=1}^{J}\mu_{ij}$ (מיצוע על פני השורות) ובדומה $\bar{\mu}_{.j}=\frac{1}{I}\sum_{i=1}^{I}\mu_{ij}$ (מיצוע על העמודות), מיצוע כללי $\mu=\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\mu_{ij}$ וכן נגדיר $\alpha_i=\bar{\mu}_{i.}-\mu,\beta_j=\bar{\mu}_{.j}-\mu,\gamma_{ij}=\mu_{ij}-\left(\mu+\alpha_i+\beta_j\right)$. אם נקבע ערך של j ונסתכל על הביטוי $\Delta_{A,ir}^{(j)}=\mu_{ij}-\mu_{ir}=\mu+\alpha_i+\beta_j+\gamma_{ij}-\left(\mu+\alpha_r+\beta_j+\gamma_{rj}\right)=(\alpha_i-\alpha_r)+(\gamma_{ij}-\gamma_{ir})$, מתברר כי תחת המצב של $\gamma_{ij}=0,\forall i,j$ (מצב של חוסר אינטראקציה) נקבל $\Delta_{A,ir}^{(j)}=\alpha_i-\alpha_r$ לכל j. בהתאם נקבל במצב של חוסר אינטראקציה כי $\Delta_{B,tn}^{(i)}=\beta_t-\beta_n$ לכל i. נוכל להסיק כי שלושת התנאים הבאים שקולים זה לזה:

1. $\forall i,j:\gamma_{ij}=0$
2. $\forall j:\Delta_{A,ik}^{(j)}=\delta_{A,ik}$ לכל i,k
3. $\forall i:\Delta_{B,tn}^{(i)}=\delta_{B,tn}$ לכל t,n

במודל זה נוכל להתייחס לשלוש השערות:

1. $H_0^A:\alpha_i=0,\forall i$
2. $H_0^B:\beta_j=0,\forall j$
3. $H_0^{AB}:\gamma_{ij}=0,\forall i,j$

הערה: הפרמטרים α,β,γ מקיימים את התנאים הבאים:

$$\sum_{i=1}^{I}\alpha_i=0,\sum_{j=1}^{J}\beta_j=0,\forall rt:\sum_{i=1}^{I}\gamma_{it}=0,\forall r:\sum_{j=1}^{J}\gamma_{rj}=0$$

בדי לבנות פרוצדורות לבדיקת ההשערות הנ"ל נרצה לנסח את המודל בצורה של רגרסיה. נניח כי הרמות מקיימות $I=3, J=2, n=2$ אז $y = \begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \\ y_{311} \\ y_{312} \\ y_{321} \\ y_{322} \end{bmatrix}$ ובדומה $\epsilon = \begin{bmatrix} \epsilon_{111} \\ \vdots \\ \epsilon_{IJn} \end{bmatrix}$ ובן $\eta = \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_{I-1} \\ \beta_i \\ \vdots \\ \beta_{J-1} \\ \gamma_{11} \\ \vdots \\ \gamma_{(I-1),(J-1)} \end{bmatrix}$, כלומר עבור I,J שהנחנו, $\eta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \gamma_{11} \\ \gamma_{21} \end{bmatrix}$ ומתקיים $y = X\eta + \epsilon$. צריך לשנות את מבנה וקטור התוחלות $E[y] = X\eta$ כך שיתקיימו כל התנאים על הסכום האפסי של הפרמטרים:

$$\alpha_3 = -(\alpha_1+\alpha_2), \beta_2 = -\beta_1, \gamma_{12} = -\gamma_{11}, \gamma_{22} = -\gamma_{21}, \gamma_{31} = -(\gamma_{11}+\gamma_{21}), \gamma_{32} = -\gamma_{31}$$

ובסך הכל נקבל:

$$E[y] = \begin{bmatrix} \mu+\alpha_1+\beta_1+\gamma_{11} \\ \mu+\alpha_1+\beta_1+\gamma_{11} \\ \mu+\alpha_1+\beta_2+\gamma_{12} \\ \mu+\alpha_1+\beta_2+\gamma_{12} \\ \mu+\alpha_2+\beta_1+\gamma_{21} \\ \mu+\alpha_2+\beta_1+\gamma_{21} \\ \mu+\alpha_2+\beta_2+\gamma_{22} \\ \mu+\alpha_2+\beta_2+\gamma_{22} \\ \mu+\alpha_3+\beta_1+\gamma_{31} \\ \mu+\alpha_3+\beta_1+\gamma_{31} \\ \mu+\alpha_3+\beta_2+\gamma_{32} \\ \mu+\alpha_3+\beta_2+\gamma_{32} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \gamma_{11} \\ \gamma_{21} \end{bmatrix}}_{\eta}$$

עבור $I=3, J=2, n_{ij}=1 \forall i,j$ נקבל את המטריצה H באופן הבא:

$$\eta=\begin{bmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \beta_1\\ \gamma_{11}\\ \gamma_{21}\end{bmatrix}=\underbrace{\begin{bmatrix}\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6}\\ \frac{1}{2}-\frac{1}{6} & \frac{1}{2}-\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6}\\ -\frac{1}{6} & -\frac{1}{6} & \frac{1}{2}-\frac{1}{6} & \frac{1}{2}-\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6}\\ \frac{1}{3}-\frac{1}{6} & -\frac{1}{6} & \frac{1}{3}-\frac{1}{6} & -\frac{1}{6} & \frac{1}{3}-\frac{1}{6} & -\frac{1}{6}\\ 1-\frac{2}{3} & 0-\frac{1}{3} & 0-\frac{1}{6} & 0--\frac{1}{6} & 0-\frac{1}{6} & 0--\frac{1}{6}\\ 0-\frac{1}{6} & 0--\frac{1}{6} & 1-\frac{2}{3} & 0-\frac{1}{3} & 0-\frac{1}{6} & 0--\frac{1}{6}\end{bmatrix}}_{H}\underbrace{\begin{bmatrix}\mu_{11}\\ \mu_{12}\\ \mu_{21}\\ \mu_{22}\\ \mu_{31}\\ \mu_{32}\end{bmatrix}}_{\beta}$$

$$=\underbrace{\begin{bmatrix}\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6}\\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6}\\ -\frac{1}{6} & -\frac{1}{6} & \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6}\\ \frac{1}{6} & -\frac{1}{6} & \frac{1}{6} & -\frac{1}{6} & \frac{1}{6} & -\frac{1}{6}\\ \frac{1}{3} & -\frac{1}{3} & -\frac{1}{6} & \frac{1}{6} & -\frac{1}{6} & \frac{1}{6}\\ -\frac{1}{6} & \frac{1}{6} & \frac{1}{3} & -\frac{1}{3} & -\frac{1}{6} & \frac{1}{6}\end{bmatrix}}_{H}\begin{bmatrix}\mu_{11}\\ \mu_{12}\\ \mu_{21}\\ \mu_{22}\\ \mu_{31}\\ \mu_{32}\end{bmatrix}$$

הסיפור ההפוך עם $\beta=G\eta$:

$$\beta=\begin{bmatrix}\mu_{11}\\ \mu_{12}\\ \mu_{21}\\ \mu_{22}\\ \mu_{31}\\ \mu_{32}\end{bmatrix}=\underbrace{\begin{bmatrix}1 & 1 & 0 & 1 & 1 & 0\\ 1 & 1 & 0 & -1 & -1 & 0\\ 1 & 0 & 1 & 1 & 0 & 1\\ 1 & 0 & 1 & -1 & 0 & -1\\ 1 & -1 & -1 & 1 & -1 & -1\\ 1 & -1 & -1 & -1 & 1 & 1\end{bmatrix}}_{G}\begin{bmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \beta_1\\ \gamma_{11}\\ \gamma_{21}\end{bmatrix}$$

ובדיוק כמו במקרה החד-כווני מתקיים $\eta=H\beta,\beta=G\eta\rightarrow G=H^{-1}$.

לפי המודל $y=W\beta+\epsilon=X\eta+\epsilon\rightarrow\eta=H\beta$, מתקיים $\hat{\beta}=(W^TW)^{-1}W^Ty,\hat{\eta}=(X^TX)^{-1}X^Ty$; מצד אחד אנחנו יודעים $\hat{y}=W\hat{\beta}=P_Wy$, מצד שני $\hat{y}=X\hat{\eta}=P_Xy$ וראינו $P_X=P_W$. באותה מידה הזהויות הוקטוריות נכונות גם עבור האומדים של הפרמטרים, כלומר $\hat{\eta}=H\hat{\beta}$:

$$E[y]=W\beta=WG\eta=X\eta\rightarrow X=WG$$

$$\begin{aligned}\hat{\eta}&=(X^TX)^{-1}X^Ty=((WG)^TWG)^{-1}(WG)^Ty=(G^TW^TWG)^{-1}G^TW^Ty\\ &=G^{-1}(W^TW)^{-1}G^{T^{-1}}G^TW^Ty=\underbrace{G^{-1}}_{H}\underbrace{(W^TW)^{-1}W^Ty}_{\hat{\beta}}\end{aligned}$$

המטריצה W המתקבלת עבור $I=3, J=3, n=2$:

$$W=\begin{bmatrix}1&0&0&0&0&0&0&0&0\\1&0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0\\0&0&0&1&0&0&0&0&0\\0&0&0&0&1&0&0&0&0\\0&0&0&0&1&0&0&0&0\\0&0&0&0&0&1&0&0&0\\0&0&0&0&0&1&0&0&0\\0&0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&0&1\\0&0&0&0&0&0&0&0&1\end{bmatrix}$$

אז $\hat{\mu}_{ij}=\bar{y}_{ij.}$ כאשר $\bar{y}_{ij.}=\frac{1}{n}\sum_k y_{ijk}$ ולכן $\hat{y}_{ijk}=\hat{\mu}_{ij}=\bar{y}_{ij.}$. בגלל השוויון $\hat{\eta}=H\hat{\beta}$ נוכל להסיק את האומד עבור μ:

$$\hat{\mu}=\frac{1}{IJ}\sum_i\sum_j\hat{\mu}_{ij}=\frac{1}{IJ}\sum_i\sum_j\bar{y}_{ij.}=\frac{1}{IJ}\sum_i\sum_j\left[\frac{1}{n}\sum_k y_{ijk}\right]=\frac{1}{IJn}\sum_i\sum_j\sum_k y_{ijk}=\bar{y}_{...}$$

אנחנו יודעים גם $\alpha_i=\frac{1}{J}\sum_{j=1}^J\mu_{ij}-\frac{1}{IJ}\sum_{r=1}^I\sum_{s=1}^J\mu_{rs}$ לכן האומד מקיים (בגלל השוויון $\hat{\eta}=H\hat{\beta}$):

$$\hat{\alpha}_i=\frac{1}{J}\sum_{j=1}^J\hat{\mu}_{ij}-\frac{1}{IJ}\sum_{r=1}^I\sum_{s=1}^J\hat{\mu}_{rs}=\frac{1}{J}\sum_{j=1}^J\bar{y}_{ij.}-\frac{1}{IJ}\sum_{r=1}^I\sum_{s=1}^J\bar{y}_{rs.}=\bar{y}_{i..}$$

ובסך הכל נקבל את האומדים:

$$\hat{\mu}=\bar{y}_{...}\quad\hat{\alpha}_i=\bar{y}_{i..}\quad\hat{\beta}_j=\bar{y}_{.j.}\quad\gamma_{ij}=\hat{\mu}_{ij}-\left(\hat{\mu}+\hat{\alpha}_i+\hat{\beta}_j\right)=\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}$$

השערת חוסר אינטראקציה (בשעור הקודם): $H_0^{AB}:\gamma_{ij}=0,\forall i,j$. לוקחים את $X\eta$ ומוחקים ממנו את כל הפרמטרים ששייכים ל-γ ומוחקים מ-X את העמודות הרלוונטיות. למשל, עבור מקרה לא מאוזן עם $I=J=3, n=2$ יש לנו את הוקטור $\eta=\begin{bmatrix}\mu\\\alpha_1\\\alpha_2\\\beta_1\\\beta_2\\\gamma_{11}\\\gamma_{12}\\\gamma_{21}\\\gamma_{22}\end{bmatrix}$ עם 9 פרמטרים ומטריצה X הבאה:

$$X=\left[\begin{array}{ccccccccc}
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0\\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0\\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0\\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0\\
1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0\\
1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0\\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0\\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0\\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1\\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1\\
1 & 0 & 1 & -1 & -1 & 0 & 0 & -1 & -1\\
1 & 0 & 1 & -1 & -1 & 0 & 0 & -1 & -1\\
1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0\\
1 & -1 & -1 & 1 & 0 & -1 & 0 & -1 & 0\\
1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1\\
1 & -1 & -1 & 0 & 1 & 0 & -1 & 0 & -1\\
1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1\\
1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1\\
\underbrace{\quad}_{Q_0} & \multicolumn{2}{c}{\underbrace{\qquad\qquad}_{Q_1}} & \multicolumn{2}{c}{\underbrace{\qquad\qquad}_{Q_2}} & \multicolumn{4}{c}{\underbrace{\qquad\qquad\qquad\qquad}_{Q_3}}
\end{array}\right]$$

כאשר Q_0 מתאים ל-μ, Q_1 ל-α, Q_2 ל-β ו-Q_3 ל-γ. לאחר הסרת העמודות מתקבל למעשה מודל רגרסיה חדש, לשם כך נעזר במשפט הבא:

נניח מודל רגרסיה $y=X\beta+\epsilon$ ונניח כי נפרק את X,β באופן $X=\begin{bmatrix}X^{(1)} & X^{(2)}\end{bmatrix}$, $\beta=\begin{bmatrix}\beta^{(1)}\\ \beta^{(2)}\end{bmatrix}$. המודל המלא מקיים $\hat{\beta}=(X^TX)^{-1}X^Ty$ ובהתאם ניתן להסיק את הפירוק $\begin{bmatrix}\hat{\beta}^{(1)}\\ \hat{\beta}^{(2)}\end{bmatrix}$ במקביל לפירוק של β עצמו.

אם נשמיט מהמודל את האיבר $X^{(2)}\beta^{(2)}$ נקבל את $\tilde{\beta}^{(1)}=\left({X^{(1)}}^TX^{(1)}\right)^{-1}{X^{(1)}}^Ty$.

משפט: אם ${X^{(1)}}^TX^{(2)}=0$ אז $\tilde{\beta}^{(1)}=\hat{\beta}^{(1)}$, כלומר, אם מתקיים תנאי הניצבות אז האומדים לא משתנים גם כאשר חלקים מוסרים.

הוכחה:

$$X=\begin{bmatrix}X^{(1)} & X^{(2)}\end{bmatrix},\ \ X^T=\begin{bmatrix}{X^{(1)}}^T\\ {X^{(2)}}^T\end{bmatrix}\to X^TX=\begin{bmatrix}{X^{(1)}}^TX^{(1)} & {X^{(1)}}^TX^{(2)}\\ {X^{(2)}}^TX^{(1)} & {X^{(2)}}^TX^{(2)}\end{bmatrix},X^Ty=\begin{bmatrix}{X^{(1)}}^Ty\\ {X^{(2)}}^Ty\end{bmatrix}$$

אם מתקיים ${X^{(1)}}^TX^{(2)}=0$ אז מתקבל $X^TX=\begin{bmatrix}{X^{(1)}}^TX^{(1)} & 0\\ 0 & {X^{(2)}}^TX^{(2)}\end{bmatrix}$ ולכן $(X^TX)^{-1}=\begin{bmatrix}\left({X^{(1)}}^TX^{(1)}\right)^{-1} & 0\\ 0 & \left({X^{(2)}}^TX^{(2)}\right)^{-1}\end{bmatrix}$ ונקבל:

$$\hat{\beta}=(X^TX)^{-1}X^Ty=\begin{bmatrix}\left({X^{(1)}}^TX^{(1)}\right)^{-1} & 0\\ 0 & \left({X^{(2)}}^TX^{(2)}\right)^{-1}\end{bmatrix}\begin{bmatrix}{X^{(1)}}^Ty\\ {X^{(2)}}^Ty\end{bmatrix}=\begin{bmatrix}\left({X^{(1)}}^TX^{(1)}\right)^{-1}{X^{(1)}}^Ty\\ \left({X^{(2)}}^TX^{(2)}\right)^{-1}{X^{(2)}}^Ty\end{bmatrix}$$

כלומר $\begin{bmatrix} \left(X^{(1)^T}X^{(1)}\right)^{-1} X^{(1)^T} y \\ \left(X^{(2)^T}X^{(2)}\right)^{-1} X^{(2)^T} y \end{bmatrix} = \begin{bmatrix} \hat{\beta}^{(1)} \\ \hat{\beta}^{(2)} \end{bmatrix}$ ובדיוק $\hat{\beta}^{(1)} = \tilde{\beta}^{(1)}$.

נחזור לניתוח שונות דו-כווני:

נוכל לרשום $C = \begin{bmatrix} \gamma_{11} \\ \vdots \\ \gamma_{(I-1),(J-1)} \end{bmatrix}, B = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{J-1} \end{bmatrix}, A = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{I-1} \end{bmatrix}, \eta = \begin{bmatrix} \mu \\ A \\ B \\ C \end{bmatrix}$, החלוקה התואמת של X היא כאמור $X = [Q_0 \quad Q_1 \quad Q_2 \quad Q_3]$, במצב המאוזן מתקיים:

$$Q_0^T Q_1 = Q_0^T Q_2 = Q_0^T Q_3 = Q_1^T Q_2 = Q_1^T Q_3 = Q_2^T Q_3 = 0$$

נוותר על ההוכחה המסורבלת, בבל מקרה נקבל כי המטריצה $X^T X$ היא בצורה:

$$X^T X = \begin{bmatrix} Q_0^T Q_0 & & & 0 \\ & Q_1^T Q_1 & & \\ & & Q_2^T Q_2 & \\ 0 & & & Q_3^T Q_3 \end{bmatrix}$$

כך למשל עבור המטריצה X שראינו קודם מתקיים:

$$X^T X = \begin{bmatrix} 18 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 12 & 6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6 & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12 & 6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & 12 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 & 4 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 4 & 8 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 & 4 & 2 & 8 & 4 \\ 0 & 0 & 0 & 0 & 0 & 2 & 4 & 4 & 8 \end{bmatrix}$$

במקרה הכללי, אם נשמיט את הפרמטרים בוקטור C (כלומר $H_0^{AB}: \gamma_{ij} = 0, \forall i, j$) יתקבל מהמשפט ומיחסי הניצבות כי: $\hat{\mu}^{(0)} = \hat{\mu}, \hat{A}^{(0)} = \hat{A}, \hat{B}^{(0)} = \hat{B}$. מבחן F: $F_{AB} = \frac{\frac{\|\hat{y} - \hat{y}^{(0)}\|^2}{[(I-1)(J-1)]}}{\frac{\|y - \hat{y}\|^2}{N - IJ}}$.

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij} = \bar{y}_{ij.}, \hat{y}_{ijk}^{(0)} = \hat{\mu}^{(0)} + \hat{\alpha}_i^{(0)} + \hat{\beta}_j^{(0)} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j \rightarrow \hat{y}_{ijk} - \hat{y}_{ijk}^{(0)} = \hat{\gamma}_{ij}$$
$$= \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

$$F_{AB} = \frac{\sum_i \sum_j \sum_k \hat{\gamma}_{ij}^2 / [(I-1)(J-1)]}{\sum_i \sum_j \sum_k \left(y_{ijk} - \hat{y}_{ijk}\right)^2 / (N - IJ)} = \frac{n \sum_i \sum_j \hat{\gamma}_{ij}^2 / [(I-1)(J-1)] / [(I-1)(J-1)]}{\sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij.}\right)^2 / (N - IJ)}$$
$$= \frac{n \sum_i \sum_j \left(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}\right)^2 / [(I-1)(J-1)]}{\sum_i \sum_j \sum_k \left(y_{ijk} - \bar{y}_{ij.}\right)^2 / (N - IJ)}$$

באופן דומה:

$$F_A = \frac{\sum_i \sum_j \sum_k \left(\hat{y}_{ijk} - \hat{y}_{ijk}^{(0)}\right)^2 / (I-1)}{\sum_i \sum_j \sum_k \left(y_{ijk} - \hat{y}_{ijk}\right)^2 / (N - IJ)} = \cdots = \frac{Jn \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 / (I-1)}{\sum_i \sum_j \sum_k \left(y_{ijk} - \hat{y}_{ijk}\right)^2 / (N - IJ)}$$

$$F_B = \frac{\sum_i \sum_j \sum_k \left(\hat{y}_{ijk} - \hat{y}_{ijk}^{(0)}\right)^2 / (J-1)}{\sum_i \sum_j \sum_k \left(y_{ijk} - \hat{y}_{ijk}\right)^2 / (N-IJ)} = \cdots = \frac{In \sum_i \left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2 / (J-1)}{\sum_i \sum_j \sum_k \left(y_{ijk} - \hat{y}_{ijk}\right)^2 / (N-IJ)}$$

המקרה הלא-מאוזן

$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$ כאשר i,j כרגיל ו-$k = 1, \dots, n_{ij}$. נבחר משתנים באופן הבא:

$\pi_1, \pi_2, \dots, \pi_I : \pi_i > 0, \sum_{i=1}^{I} \pi_i = 1$ ובדומה $\tau_1, \tau_2, \dots \tau_J : \tau_j > 0, \sum_{j=1}^{J} \tau_j = 1$. בברירת מחדל ניקח $n_{i.} =$ $\sum_j n_{ij}$ $n_{.j} = \sum_i n_{ij}$ $N = \sum_i \sum_j n_{ij} \rightarrow \pi_i = \frac{n_{i.}}{N}, \tau_j = \frac{n_{.j}}{N}$ ונגדיר:

$$\mu = \sum_i \sum_j \pi_i \tau_j \mu_{ij} \qquad \bar{\mu}_{i.} = \sum_j \tau_j \mu_{ij} \qquad \bar{\mu}_{.j} = \sum_i \pi_i \mu_{ij}$$
$$\alpha_i = \bar{\mu}_{i.} - \mu \qquad \beta_j = \bar{\mu}_{.j} - \mu \qquad \gamma_{ij} = \mu_{ij} - \left(\mu + \alpha_i + \beta_j\right)$$

במצב של חוסר אינטראקציה $(\forall i,j : \gamma_{ij} = 0)$ נקבל את אותו פירוש של אי-תלות בין גורמים – כלומר ההפרש בין שני תאים בעמודה זהה למקביליהם בעמודה אחרת (למשל $\mu_{11} - \mu_{31} = \mu_{14} - \mu_{34}$) וכן להפרשים בין שני תאים מקבילים בשורות (כלומר $\mu_{11} - \mu_{14} = \mu_{31} - \mu_{34}$)

יהיו שני סטטיסטיקאים, ראובן ושמעון. נניח כי ראובן בוחר משקלות π_i, τ_j כנ"ל ואילו שמעון בוחר משקלות π_i^*, τ_j^*. מחד יש לנו את המודל $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$ ומאידך $\mu_{ij} = \mu^* + \alpha_i^* + \beta_j^* + \gamma_{ij}^*$, התנאי $\forall i,j : \gamma_{ij} = 0$ שקול לתנאי $\forall i,j : \gamma_{ij}^* = 0$, כלומר חוסר אינטראקציה אינו תלוי בבחירת המשקולות.

מהאילוצים על הפרמטרים $(y = W\beta + \epsilon = X\eta + \epsilon)$ נקבל:

$$\sum_i \pi_i \alpha_i = 0 \rightarrow \alpha_I = -\sum_{i=1}^{I-1} \frac{\pi_i \alpha_i}{\pi_I}, \qquad \sum_j \tau_j \beta_j = 0 \rightarrow \beta_J = -\sum_{j=1}^{J-1} \frac{\tau_j \beta_j}{\tau_J}$$
$$\forall j : \sum_i \pi_i \gamma_{ij} = 0 \rightarrow \gamma_{Ij} = -\sum_{i=1}^{I-1} \frac{\pi_i \gamma_{iJ}}{\pi_I}, \qquad \forall i : \sum_j \tau_j \gamma_{ij} = 0 \rightarrow \gamma_{iJ} = -\sum_{j=1}^{J-1} \frac{\tau_j \gamma_{ij}}{\tau_J}$$

ונוכל לבנות את המטריצה X בדומה למקרה הקודם, כלומר כך שיתקיים $E[y] = X\eta$.

לדוגמא, במקרה בו $I = J = 3$ עם ערכי n_{ij} הבאים:

i\j	1	2	3	$n_{i.}$
1	3	3	3	9
2	2	2	2	6
3	2	2	2	6
$n_{.j}$	7	7	7	21

אם נשתמש במשקלות π_i, τ_j נקבל $\tau_1 = \tau_2 = \tau_3 = \frac{n_{ij}}{N} = \frac{7}{21} = \frac{1}{3}$ ובדומה $\pi_1 = \frac{3}{7}, \pi_2 = \pi_3 = \frac{2}{7}$, כך מהאילוצים $\alpha_3 = -\frac{\pi_1}{\pi_3}\alpha_1 - \frac{\pi_2}{\pi_3}\alpha_2 = -\frac{3}{2}\alpha_1 - \alpha_2$ וכן $\beta_3 = -\frac{\tau_1}{\tau_3}\beta_1 - \frac{\tau_2}{\tau_3}\beta_2 = -\beta_1 - \beta_2$.

המטריצה X תהיה:

$$X=\begin{bmatrix}
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0\\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0\\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0\\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0\\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0\\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0\\
1 & 1 & 0 & -1 & -1 & -1 & -1 & 0 & 0\\
1 & 0 & 1 & -1 & -1 & -1 & -1 & 0 & 0\\
1 & 0 & 1 & -1 & -1 & -1 & -1 & 0 & 0\\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0\\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0\\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1\\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1\\
1 & 0 & 1 & -1 & -1 & 0 & 0 & -1 & -1\\
1 & -\frac{3}{2} & -1 & -1 & -1 & 0 & 0 & -1 & -1\\
1 & -\frac{3}{2} & -1 & 1 & 0 & -\frac{3}{2} & 0 & -\frac{3}{2} & 0\\
1 & -\frac{3}{2} & -1 & 1 & 0 & -\frac{3}{2} & 0 & -\frac{3}{2} & 0\\
1 & -\frac{3}{2} & -1 & 0 & 1 & 0 & -\frac{3}{2} & 0 & -\frac{3}{2}\\
1 & -\frac{3}{2} & -1 & 0 & 1 & 0 & -\frac{3}{2} & 0 & -\frac{3}{2}\\
1 & -\frac{3}{2} & -1 & -1 & -1 & \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & \frac{3}{2}\\
1 & -\frac{3}{2} & -1 & -1 & -1 & \frac{3}{2} & \frac{3}{2} & \frac{3}{2} & \frac{3}{2}
\end{bmatrix}$$

עבור המודל המלא נוכל להשתמש בניסוח $y = W\beta + \epsilon$, ממנו נובע $\hat{y}_{ijk} = \bar{y}_{ij.}$, במקודם נוכל לפרק את הוקטור η באופן $C = \begin{bmatrix} \gamma_{11} \\ \vdots \\ \gamma_{(I-1),(J-1)} \end{bmatrix}, B = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{J-1} \end{bmatrix}, A = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{I-1} \end{bmatrix}, \eta = \begin{bmatrix} \mu \\ A \\ B \\ C \end{bmatrix}$ ובמקביל את X באופן $X = [Q_0 \quad Q_1 \quad Q_2 \quad Q_3]$, אך בניגוד למצב המאוזן כאן אין לנו ניצבות, לכן נקבל שינוי בנוסחאות.

$SSE(\mu,\alpha,\beta,\gamma)$ יהיה SSE במודל המלא, $SSE(\mu,\alpha,\beta)$ יהיה SSE במודל בלי γ, $SSE(\mu,\beta,\gamma)$ יהיה SSE במודל בלי α וכן הלאה. ההשערה $H_0^{AB}: \gamma_{ij} = 0, \forall i,j$ תוליד את הסטטיסטי F_{AB} באופן:

$$F_{AB} = \frac{\frac{\left\|\hat{y} - \hat{y}^{(-\gamma)}\right\|^2}{[(I-1)(J-1)]}}{\frac{\|y - \hat{y}\|^2}{N - IJ}} = \frac{\frac{SSE(\mu,\alpha,\beta) - SSE(\mu,\alpha,\beta,\gamma)}{[(I-1)(J-1)]}}{\frac{SSE(\mu,\alpha,\beta,\gamma)}{N - IJ}}$$

ראינו כי מתקיים במודל המלא

$$SSE(\mu,\alpha,\beta,\gamma) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}} \left(y_{ijk} - \hat{y}_{ijk}\right)^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{n_{ij}} \left(y_{ijk} - \bar{y}_{ij.}\right)^2$$

אבל אין לנו ביטוי מפורש עבור $SSE(\mu,\alpha,\beta)$ וצריך להריץ שתי רגרסיות: אחת עם γ ואחת בלעדיהן ואז לחשב את ההפרש בין ערכי SSE. באופן דומה אם נרצה לבדוק את ההשערה $H_0^A: \alpha_i = 0, \forall i$ אז הסטטיסטי F_A יהיה

$$F_A = \frac{\sum_i \sum_j \sum_k \left(\hat{y}_{ijk} - \hat{y}_{ijk}^{(-\alpha)}\right)^2 /(I-1)}{\sum_i \sum_j \sum_k \left(y_{ijk} - \hat{y}_{ijk}\right)^2 /(N-IJ)} = \frac{\left(SSE(\mu,\beta,\gamma) - SSE(\mu,\alpha,\beta,\gamma)\right)/(I-1)}{SSE(\mu,\alpha,\beta,\gamma)/(N-IJ)}$$

ובדומה עבור $H_0^B: \beta_j = 0, \forall j$ הסטטיסטי F_B יהיה

$$F_B = \frac{\sum_i \sum_j \sum_k \left(\hat{y}_{ijk} - \hat{y}_{ijk}^{(0)}\right)^2 /(I-1)}{\sum_i \sum_j \sum_k \left(y_{ijk} - \hat{y}_{ijk}\right)^2 /(N-IJ)} = \frac{\left(SSE(\mu,\alpha,\gamma) - SSE(\mu,\alpha,\beta,\gamma)\right)/(I-1)}{SSE(\mu,\alpha,\beta,\gamma)/(N-IJ)}$$

כאשר בכל המקרים המכנה $\frac{SSE(\mu,\alpha,\beta,\gamma)}{N-IJ}$ הוא למעשה ה-MSE, כלומר s^2.

בדיקת השערות

מבחנים של H_A, H_B במודל בלי אינטראקציה:

$$SSE^* = SSE(\mu,\alpha,\beta) \rightarrow \begin{cases} SSA^* = SSE(\mu,\beta) - SSE(\mu,\alpha,\beta) \\ SSB^* = SSE(\mu,\alpha) - SSE(\mu,\alpha,\beta) \end{cases}$$

$$d = N - I - J + 1 \rightarrow \begin{cases} F_A = \dfrac{SSA^*/(I-1)}{SSE^*/d} \overset{H_A}{\sim} F_{I-1,d} \\ F_B = \dfrac{SSB^*/(J-1)}{SSE^*/d} \overset{H_B}{\sim} F_{J-1,d} \end{cases}$$

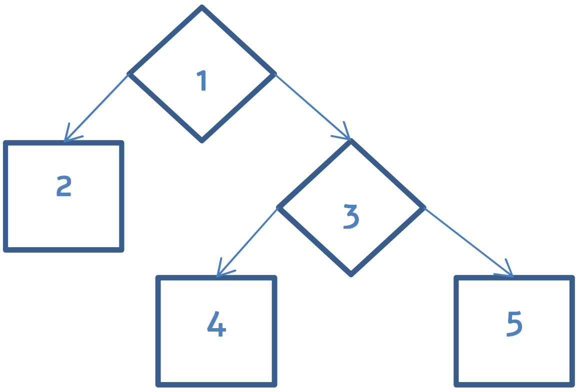
במודל המאוזן נקבל:

$$SSA^* = SSA = Jn \sum_{i=1}^{I} (\bar{y}_{i..} - \bar{y}_{...})^2 \quad SSB^* = SSB = In \sum_{j=1}^{J} \left(\bar{y}_{.j.} - \bar{y}_{...}\right)^2 \quad SSE^* = SSE = \sum_i \sum_j \sum_k \left(y_{ijk} - \hat{y}_{ijk}\right)^2$$

כאשר $\hat{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + \left(\bar{y}_{.j.} - \bar{y}_{...}\right) = \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...}$. במקרה של חוסר אינטראקציה הערך של $\alpha_r - \alpha_s$ אינו תלוי במשקלות:

$$\left.\begin{aligned} \alpha_r &= \bar{\mu}_{r.} - \mu = \sum_{j=1}^{J} \tau_j \mu_{rj} - \mu \\ \alpha_s &= \bar{\mu}_{s.} - \mu = \sum_{j=1}^{J} \tau_j \mu_{sj} - \mu \end{aligned}\right\} \rightarrow \alpha_r - \alpha_s = \sum_{j=1}^{J} \tau_j \left(\mu_{rj} - \mu_{sj}\right) = \sum_{j=1}^{J} \tau_j \Delta_{rs} = \Delta_{rs} \sum_{j=1}^{J} \tau_j = \Delta_{rs}$$

באופן כללי, אם נרצה לבדוק השערות על A,B, נעבוד לפי התרשים הבא:



1. בודקים את השערת חוסר האינטראקציה, H_{AB}. אם $P_{AB} > \begin{matrix}0.10\\0.15\end{matrix}$ אז עוברים לשלב 2, אחרת לשלב 3.
2. משמיטים את משתני האינטראקציה γ_{ij} ואת העמודות הרלוונטיות ב-X ועובדים עם SSA^*, SSB^*, SSE^*.
3. האם יש משקלות לרמות השונות של B (τ_j) שהן בעלות משמעות? אם כן, עוברים לשלב 4, אחרת לשלב 5.
4. משתמשים ב-τ_j הרלוונטיים ומבצעים הסקה עבור ההשפעה של A לפיהן.
5. חוקרים את ההשפעה של A עבור כל רמה של B בנפרד.

קונטרסטים בניתוח שונות דו-כווני

הערה: נכון למצב של חוסר אינטראקציה, אינטראקציה ברת הזנחה או משקלות בעלות משמעות.

$\psi(d) = \sum_{i=1}^{I} d_i\alpha_i$ ובשל האילוץ על הפרמטרים $\sum_{i=1}^{I}\pi_i\alpha_i = 0 \to \alpha_I = -\sum_{i=1}^{I-1}\frac{\pi_i\alpha_i}{\pi_I}$ נקבל $\psi(d) = \sum_{i=1}^{I} d_i\alpha_i - d_I\sum_{i=1}^{I-1}\frac{\pi_i\alpha_i}{\pi_I} \to c_i = d_i - d_I\frac{\pi_i}{\pi_I} \to \psi(d) = \sum_{i=1}^{I-1} c_i\alpha_i = c^T\alpha$. נרצה לראות את ההתפלגות:

$$\hat{\eta} = (X^TX)^{-1}X^Ty \to Cov(\hat{\eta}) = \sigma^2\Lambda = \sigma^2(X^TX)^{-1}$$

$$\hat{\psi}(d) = c^T\hat{\alpha} \to Var\left(\hat{\psi}(d)\right) = c^TG_\alpha c = V(c) \to G_\alpha = Cov(\hat{\alpha}) = [\Lambda_{rs}]_{r,s=2,\dots,I}$$

כלומר השונות של $\hat{\alpha}$ מגיעה מהעמודות של α במטריצה X.

$$c^T\hat{\alpha} \sim N\left(c^T\alpha, \sigma^2(c^TG_\alpha c)\right) \to \frac{c^T\hat{\alpha} - c^T\alpha}{s\sqrt{V(c)}} \sim t_m, m = \begin{cases} N - IJ \\ N - I - J + 1 \end{cases}$$

ועבור רמת סמך נדרשת $1-\omega$ רווח הסמך הבודד עבור $c^T\alpha$ יהיה $c^T\hat{\alpha} \pm t_m(1-\omega)s\sqrt{V(c)}$. עבור רווחי סמך מרובים:

1. שיטת בונפרוני:

$$c^{(k)^T}\hat{\alpha} \pm t_m\left(1 - \frac{\omega}{k}\right)s\sqrt{V(c)}$$

2. השיטה המדויקת:

אם $\psi^{(k)} = {c^{(k)}}^T \alpha \rightarrow \hat{\psi}^{(k)} = {c^{(k)}}^T \hat{\alpha}$ אז $\Psi = \begin{bmatrix} \psi^{(1)} \\ \vdots \\ \psi^{(k)} \end{bmatrix} \rightarrow \widehat{\Psi} = \begin{bmatrix} \hat{\psi}^{(1)} \\ \vdots \\ \hat{\psi}^{(k)} \end{bmatrix} \sim N(\Psi, \sigma^2 C^T G_\alpha C)$

כאשר $C = \begin{bmatrix} c_1^{(1)} & \cdots & c_{I-1}^{(1)} \\ \vdots & \ddots & \vdots \\ c_{I-1}^{(1)} & \cdots & c_{I-1}^{(I-1)} \end{bmatrix}$ מטריצת רכיבי הקונטרסטים. נסמן $z_k = \frac{\hat{\psi}^{(k)} - \psi^{(k)}}{\sigma\sqrt{V(c)}}$ אז מתקיים

$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix} \sim N(0, R)$ כאשר $R_{uv} = \frac{{c^{(u)}}^T G_\alpha c^{(v)}}{\sqrt{V(c^{(u)}) V(c^{(v)})}} = Corr(z_u, z_v)$

3. כל ההפרשים Δ_{rs}:

$$\Delta_{rs} = \alpha_r - \alpha_s = {c^{(rs)}}^T \alpha, \qquad r, s = 1, \dots, I-1, \qquad c_i^{(rs)} = \begin{cases} 1 & i = r \\ -1 & i = s \\ 0 & else \end{cases}$$

ונוכל להתייחס כמו בשיטה המדויקת.

במקרה המאוזן מתקיים $\hat{\alpha}_r - \hat{\alpha}_s = \bar{y}_{r..} - \bar{y}_{s..}$ כאשר ידוע לנו $\bar{y}_{i..} \sim N\left(\mu + \alpha_i, \frac{\sigma^2}{Jn}\right)$ לכן נקבל רווח סמך בצורה $\bar{y}_{r..} - \bar{y}_{s..} \pm \frac{s}{\sqrt{Jn}} q_{I-1}^{(1-\omega)}$ (q ערך קריטי עבור התפלגות Studentized range)

4. כל הקונטרסטים (מקביל לשיטת שפה):

$$c^T \hat{\alpha} \pm s\sqrt{(I-1)V(c)F_{I-1,m}^{(1-\omega)}}$$

ניתוח שונות רב-כווני

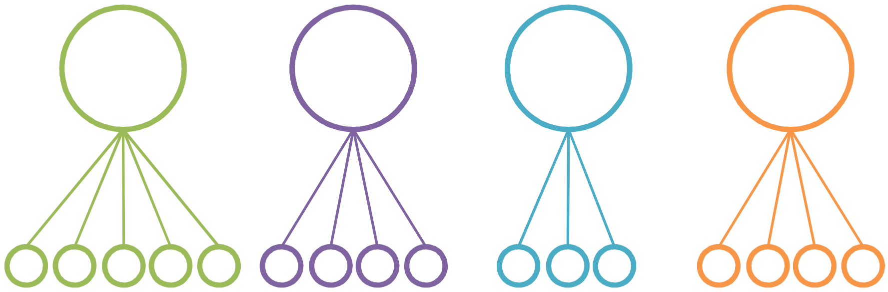
המקרה התלת כווני: $y_{ijkl} = \mu_{ijk} + \epsilon_{ijkl} \quad i = 1, \dots, I \quad j = 1, \dots, J \quad k = 1, \dots, K \quad l = 1, \dots, n_{ijk}$, אפשר להכליל לצורה $y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \psi_{ij} + \phi_{ik} + \lambda_{jk} + \omega_{ijk} + \epsilon_{ijkl}$ כאשר ψ, ϕ, λ משתני אינטראקציה הדדיים בין הגורמים ו-ω ההשפעה של כולם על כולם. אם ניקח למשל $I = J = K = 3$ נקבל כי יש 27 קומבינציות שונות, זה מודל לא אינטואיטיבי ומסובך למימוש. מודל אחר נקרא Fractional Design, מוותרים על משתני האינטראקציה לטובת נוחות חישוב. מסמנים $\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$ ואז עבור $I = J = K = 2$ יש רק 4 קומבינציות במקום 8. לדוגמא:

	A	B	C
n_1	1	1	2
n_2	1	2	1
n_3	2	1	1
n_4	2	2	2

ואז הפרמטרים יהיו $\eta = \begin{bmatrix} \mu \\ \alpha \\ \beta \\ \gamma \end{bmatrix}$ ולפי $y = X\eta + \epsilon$ נקבל את המטריצה $X = \begin{bmatrix} 1 & 1 & 1 & -1 \\ & \vdots \times n_1 & & \\ 1 & 1 & -1 & 1 \\ & \vdots \times n_2 & & \\ 1 & -1 & 1 & 1 \\ & \vdots \times n_3 & & \\ 1 & -1 & -1 & -1 \\ & \vdots \times n_4 & & \end{bmatrix}$

ניתוח שונות חד-כווני עם אפקט מקרי – One Way ANOVA With Random Effect

יש לנו מספר בתי ספר, בכ"א יש מספר תלמידים:



התצפיות הן y_{ij} כאשר i הוא בית הספר ו-j התלמיד, אלו דגימות מסוימות מתוך אוכלוסיה גדולה יותר של בתי ספר ורוצים לאפיין את התצפית y (למשל יכולת קריאה) ולחקור כיצד זהות בית הספר משפיעה על y כאשר יש מספר רב של בתי ספר. בונים מודל מהצורה $y_{ij} = \mu + a_i + \epsilon_{ij}$, דומה למודל החד-כווני אך הפעם a_i הוא משתנה מקרי, $a_i \sim N(0, \sigma^2 a)$ וכן $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, מניחים כי $a_1, \dots, a_I$ ב"ת וכן כי כל ה-ϵ_{ij} בלתי תלויים. נסמן $Y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{bmatrix}$, נניח $n_i = 4$ אז $Y_i = \begin{bmatrix} \mu + a_i + \epsilon_{i1} \\ \mu + a_i + \epsilon_{i2} \\ \mu + a_i + \epsilon_{i3} \\ \mu + a_i + \epsilon_{i4} \end{bmatrix} = \mu \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}}_{Q} \begin{bmatrix} a_i \\ \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \end{bmatrix}$, זה וקטור שמפולג רב-נורמלית. מתקבל $E[Y_i] = \mu \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + Q \cdot E \begin{bmatrix} a_i \\ \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \end{bmatrix} = \mu \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + Q \cdot 0 = \mu \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. מתקיים $Var(y_{ij}) = Var(\mu + a_i + \epsilon_{ij}) = \sigma_a^2 + \sigma_\epsilon^2$ ולכן עבור $r \neq s$ נקבל $Cov(y_{ir}, y_{is}) = Cov(\mu + a_i + \epsilon_{ir}, \mu + a_i + \epsilon_{is}) = Cov(a_i, a_i) + \underbrace{Cov(a_i, \epsilon_{ir}) + Cov(\epsilon_{ir}, a_i) + Cov(\epsilon_{ir}, \epsilon_{is})}_{=0} = \sigma_a^2$. בסך הכל נקבל כי מתקיים

$$Y_i \sim N\left(\mu \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 + \sigma_\epsilon^2 & \sigma_a^2 & \sigma_a^2 & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 + \sigma_\epsilon^2 & \sigma_a^2 & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2 + \sigma_\epsilon^2 & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2 & \sigma_a^2 + \sigma_\epsilon^2 \end{bmatrix}\right)$$

ובמקרה הכללי $Y_i \sim N\left(\mu 1_{n_i}, \sigma_a^2 1_{n_i} 1_{n_i}^T + \sigma_\epsilon^2 I_{n_i}\right) = Y_i \sim N\left(\mu 1_{n_i}, \sigma_\epsilon^2 \left(I_{n_i} + \underbrace{\theta}_{=\sigma_a^2/\sigma_\epsilon^2} 1_{n_i} 1_{n_i}^T\right)\right)$. פונקצית ההתפלגות תהיה :

$$F_{Y_i}(y) = (2\pi)^{-\frac{n_i}{2}} \left(\det\left(\sigma_\epsilon^2 \left(I_{n_i} + \theta 1_{n_i} 1_{n_i}^T\right)\right)\right)^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2} \left(y_i - \mu 1_{n_i}\right)^T \left[\sigma_\epsilon^2 \left(I_{n_i} + \theta 1_{n_i} 1_{n_i}^T\right)\right]^{-1} \left(y_i - \mu 1_{n_i}\right)\right)$$

ופונקצית הנראות:

$$L(\mu,\theta,\sigma_\epsilon^2) = \prod_{i=1}^{I} f_{Y_i}(y_i)$$
$$= (2\pi)^{\frac{-N}{2}} \prod_{i=1}^{I} \left[\det\left(\sigma_\epsilon^2\left(I_{n_i} + \theta 1_{n_i}1_{n_i}^T\right)\right)\right]^{-\frac{1}{2}}$$
$$\cdot \exp\left(-\frac{1}{2}\sum_{i=1}^{I}\left[\left(y_i - \mu 1_{n_i}\right)^T\left[\sigma_\epsilon^2\left(I_{n_i} + \theta 1_{n_i}1_{n_i}^T\right)\right]^{-1}\left(y_i - \mu 1_{n_i}\right)\right]\right)$$

נסמן $V_i = \sigma_\epsilon^2\left(I_{n_i} + \theta 1_{n_i}1_{n_i}^T\right)$, לוג הנראות הוא

$$l(\mu,\theta,\sigma_\epsilon^2) = \log L(\mu,\theta,\sigma_\epsilon^2)$$
$$= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{I}\log(\det(V_i)) - \frac{1}{2}\sum_{i=1}^{I}\left(y_i - \mu 1_{n_i}\right)^T V_i^{-1}\left(y_i - \mu 1_{n_i}\right)$$

נגדיר $H = I_{n_i} + \theta 1_{n_i}1_{n_i}^T$, נבחר מטריצה U המקיימת $U^TU = UU^T = I$ אז ניתן לרשום:

$$H = I_{n_i} + \theta \underbrace{1_{n_i}1_{n_i}^T}_{G} = I_{n_i} + \theta G = UI_{n_i}U^T + \theta U \underbrace{\begin{bmatrix} n_i & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}}_{P} U^T$$
$$= U \underbrace{\begin{bmatrix} 1 + n_i\theta & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}}_{Q} U^T$$

מכאן כי הע"ע של H הם $\lambda_1 = 1 + n_i\theta, \lambda_2 = \dots = \lambda_n = 1$ ולכן הדטרמיננטה $= \prod_{j=1}^{n}\lambda_j = |H|$ $1 + n_i\theta = (1 + n_i\theta)\cdot 1 \cdot \ldots \cdot 1$. בנוסף, מההצגה $H = UQU^T$ נקבל כי $H^{-1} = U^{-1}(QU^T)^{-1} =$ $U^{-1}(U^T)^{-1}Q^{-1} = (U^TU)^{-1}Q^{-1} = I^{-1}Q^{-1} = Q^{-1}$, בעת נוכל שוב לכפול ב- U, U^T כלומר

$$H^{-1} = UQ^{-1}U^T = U\begin{bmatrix} \frac{1}{1 + n_i\theta} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}U^T = UI_{n_i}U^T + U\begin{bmatrix} \frac{1}{1 + n_i\theta} - 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}U^T$$
$$= I_{n_i} + \left(\frac{-n_i\theta}{1 + n_i\theta}\right)U\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}U^T$$
$$= I_{n_i} + \frac{1}{n_i}\left(\frac{-n_i\theta}{1 + n_i\theta}\right)\underbrace{U\begin{bmatrix} n_i & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}U^T}_{G} = I_{n_i} + \left(\frac{-\theta}{1 + n_i\theta}\right)G$$
$$= I_{n_i} - \frac{\theta}{(1 + n_i\theta)}1_{n_i}1_{n_i}^T$$

ובסך הכל $1_{n_i}1_{n_i}^T \frac{\theta}{(1+n_i\theta)} - \left(I + \theta 1_{n_i}1_{n_i}^T\right)^{-1} = I_{n_i}$, $\det\left(I_{n_i} + \theta 1_{n_i}1_{n_i}^T\right) = (1 + n_i\theta)$. מכאן נקבל כי $\det(V_i) = \det\left(\sigma_\epsilon^2\left(I_{n_i} + \theta 1_{n_i}1_{n_i}^T\right)\right) = (\sigma_\epsilon^2)^{n_i}\det\left(I_{n_i} + \theta 1_{n_i}1_{n_i}^T\right) = (\sigma_\epsilon^2)^{n_i}(1 + n_i\theta)$ ואם נסמן $\omega = \sigma_\epsilon^2$ אז $\det(V_i) = \omega^{n_i}(1 + n_i\theta)$ ומתקבל $\log\det(V_i) = n_i\log(\omega) + \log(1 + n_i\theta)$. בנוסף יש לנו $\left(y_i - \mu 1_{n_i}\right)^T V_i^{-1}\left(y_i - \mu 1_{n_i}\right) = \frac{1}{\omega}\left(y_i - \mu 1_{n_i}\right)^T\left(I_{n_i} - \frac{\theta}{(1+n_i\theta)}1_{n_i}1_{n_i}^T\right)\left(y_i - \mu 1_{n_i}\right)$. מכאן כי

$$
\begin{aligned}
l = \log L(\mu, \theta, \omega)& \\
&= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\omega - \frac{1}{2}\sum_{i=1}^{I}\log(1 + n_i\theta) \\
&\quad - \frac{1}{2}\sum_{i=1}^{I}\left(y_i - \mu 1_{n_i}\right)^T V_i^{-1}\left(y_i - \mu 1_{n_i}\right)
\end{aligned}
$$

נסמן $e_i = y_i - \bar{y}_{i.}1_{n_i}$, $\bar{y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$ אז $e_{ij} = y_{ij} - \bar{y}_{i.} \to \sum_{j=1}^{n_i} e_{ij} = 0$ כלומר $1_{n_i}^T e_i = 0$ וגם $1_{n_i}^T 1_{n_i} = n_i$, כעת: $y_i - \mu 1_{n_i} = y_i - \mu 1_{n_i} - 1_{n_i}\bar{y}_{i.} + 1_{n_i}\bar{y}_{i.} = y_i - 1_{n_i}\bar{y}_{i.} + e_i$ אם נציב בביטוי ל- $\left(y_i - \mu 1_{n_i}\right)^T V_i^{-1}\left(y_i - \mu 1_{n_i}\right)$ נקבל $\frac{1}{\omega}\left(e_i^T e_i + \frac{n_i}{1+n_i\theta}(\bar{y}_{i.} - \mu)^2\right)$. נסמן גם $SSE = \sum_{i=1}^{I}\sum_{j=1}^{n_i} e_{ij}$ ואז פונקצית לוג הנראות תהיה

$$
\begin{aligned}
l &= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\omega - \frac{1}{2}\sum_{i=1}^{I}\log(1 + n_i\theta) - \frac{1}{2\omega}\sum_{i=1}^{I}\left(e_i^T e_i + \frac{n_i}{1 + n_i\theta}(\bar{y}_{i.} - \mu)^2\right) \\
&= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log\omega - \frac{1}{2}\sum_{i=1}^{I}\log(1 + n_i\theta) \\
&\quad - \frac{1}{2\omega}\left(SSE + \sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}(\bar{y}_{i.} - \mu)^2\right)\right)
\end{aligned}
$$

$$
\frac{dl}{d\mu} = -\frac{1}{2\omega}\sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}(\bar{y}_{i.} - \mu)\right)\cdot(-2) = \frac{1}{\omega}\sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}(\bar{y}_{i.} - \mu)\right)
$$

$$
\frac{dl}{d\theta} = -\frac{1}{2}\sum_{i=1}^{I}\frac{n_i}{(1 + n_i\theta)} + \frac{1}{2\omega}\sum_{i=1}^{I}\left(\left(\frac{n_i}{1 + n_i\theta}\right)^2(\bar{y}_{i.} - \mu)^2\right)
$$

$$
\frac{dl}{d\omega} = -\frac{N}{2\omega} + \frac{1}{2\omega^2}\left(SSE + \sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}(\bar{y}_{i.} - \mu)^2\right)\right)
$$

נרצה למצוא מתי הנגזרות מתאפסות: $\frac{dl}{d\mu} = 0, \frac{dl}{d\theta} = 0, \frac{dl}{d\omega} = 0$

$$
\frac{dl}{d\mu} = 0 = \frac{1}{\omega}\sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}(\bar{y}_{i.} - \mu)\right) \to \hat{\mu}(\theta) = \frac{\sum_{i=1}^{I}\frac{n_i}{1 + n_i\theta}\bar{y}_{i.}}{\sum_{i=1}^{I}\frac{n_i}{1 + n_i\theta}}
$$

$$\frac{dl}{d\omega} = 0 = -\frac{N}{2\omega} + \frac{1}{2\omega^2}\left(SSE + \sum_{i=1}^{I}\left(\frac{n_i}{1+n_i\theta}(\bar{y}_{i.}-\mu)^2\right)\right) \rightarrow \widehat{\omega}(\theta)$$
$$= \frac{1}{N}\left(SSE + \sum_{i=1}^{I}\left(\frac{n_i}{1+n_i\theta}\left(\bar{y}_{i.}-\hat{\mu}(\theta)\right)^2\right)\right)$$

$$\frac{dl}{d\theta} = 0 = -\frac{1}{2}\sum_{i=1}^{I}\frac{n_i}{(1+n_i\theta)} + \frac{1}{2\widehat{\omega}(\theta)}\sum_{i=1}^{I}\left(\left(\frac{n_i}{1+n_i\theta}\right)^2\left(\bar{y}_{i.}-\hat{\mu}(\theta)\right)^2\right)$$

זו משוואה לא לינארית ואין לנו דרך אנליטית נוחה לפתור, לכן נשתמש בשיטה נומרית.

שיטת ניוטון רפסון

תהי $g:\mathbb{R}\rightarrow\mathbb{R}$, נרצה לפתור $g(\theta)=0$ עבור $\theta\in\mathbb{R}$. נבחר θ^* וטור טיילור מסדר ראשון יתן לנו $g(\theta) = g(\theta^*) + (\theta-\theta^*)g'(\theta^*)$, אז $g(\theta^*) + (\theta-\theta^*)g'(\theta^*) = 0$ ונקבל $\theta = \theta^* - \frac{g(\theta^*)}{g'(\theta^*)}$, נבצע שוב ושוב עד שנגיע לפתרון. באופן כללי נרצה לראות אם $\theta_m = \theta_{m-1} - \frac{g(\theta_{m-1})}{g'(\theta_{m-1})}$ מתכנס.

למשל $g(x) = x^2-3$:

$$x_0 = 1, g'(x) = 2x \rightarrow x_{m+1} = x_m - \frac{x_m^2-3}{2x_m}$$

$$x_1 = 1 - \frac{1^2-3}{2} = 1+1 = 2$$

$$x_2 = 2 - \frac{2^2-3}{2\cdot 2} = 2 - \frac{1}{4} = 1.75$$

$$x_3 = 1.75 - \frac{1.75^2-3}{2\cdot 1.75} \approx 1.73214$$

$$x_4 = 1.732 - \frac{1.732^2-3}{2\cdot 1.732} \approx 1.73205$$

כלומר מהר מאד מתכנסים לערך $\sqrt{3}$, אך יש בעיה אפשרית אם בוחרים נקודה שנמצאת בין שני פתרונות כי אז הסדרה לא תתכנס (למשל לו היינו בוחרים $x_0=0$ לא היינו מגיעים להתכנסות כי הוא נמצא באמצע בין $\sqrt{3}$ ו-$-\sqrt{3}$).

(בחזרה לניתוח שונות)

ראינו $\hat{\mu}(\theta) = \frac{\sum_{i=1}^{I}\frac{n_i}{1+n_i\theta}\bar{y}_{i.}}{\sum_{i=1}^{I}\frac{n_i}{1+n_i\theta}}$ ומתקיים $Var(\bar{y}_{i.}) = Var(\mu+\alpha_i+\bar{\epsilon}_{i.}) = 0 + Var(\alpha_i) + Var(\bar{\epsilon}_{i.}) = \sigma_a^2 + \frac{\sigma_\epsilon^2}{n_i} = \frac{1}{n_i}(n_i\sigma_a^2+\sigma_\epsilon^2) = \sigma_\epsilon^2\left(\frac{1}{n_i}(1+n_i\theta)\right) \rightarrow \left(Var(\bar{y}_{i.})\right)^{-1} = \frac{1}{\sigma_\epsilon^2}\frac{n_i}{1+n_i\theta}$, נציב ונקבל כי $\hat{\mu}(\theta) = \frac{\sum_{i=1}^{I}\left(Var(\bar{y}_{i.})\right)^{-1}\bar{y}_{i.}}{\sum_{i=1}^{I}\left(Var(\bar{y}_{i.})\right)^{-1}} = \frac{\sum_{i=1}^{I}\frac{n_i}{1+n_i\theta}\bar{y}_{i.}}{\sum_{i=1}^{I}\frac{n_i}{1+n_i\theta}}$, כלומר $\hat{\mu}$ הוא שקלול של אומדי השונות של הממוצעים – ככל שהשונות גדולה יותר משקל הממוצע יהיה קטן יותר.

ראינו בקורס הסקה כי אם $\hat{\phi}_{MLE}$ הוא אנ"מ של ϕ אז אסימפטוטית $\hat{\phi}_{MLE} \sim N(\phi, V)$ וכן $\hat{V} = \left(E\left[\frac{-\partial^2 l}{\partial\phi_r \partial\phi_s}\right]\right)^{-1}\Big|_{\phi=\hat{\phi}_{MLE}} = \left(\left[\frac{-\partial^2 l}{\partial\phi_r \partial\phi_s}\right]\right)^{-1}\Big|_{\phi=\hat{\phi}_{MLE}}$. ר"ס עבור θ: $\hat{\theta} \pm z_{1-\frac{\alpha}{2}}\sqrt{\hat{V}_{22}}$, נסמן $[\theta_L, \theta_U]$. נגדיר $R = \frac{\sigma_a^2}{\sigma_a^2+\sigma_\epsilon^2}$ אז $1 - R = \frac{\sigma_\epsilon^2}{\sigma_a^2+\sigma_\epsilon^2} \to (1-R)^{-1} = \frac{\sigma_a^2+\sigma_\epsilon^2}{\sigma_\epsilon^2} = 1 + \frac{\sigma_a^2}{\sigma_\epsilon^2} = 1 + \theta$, ר"ס עבור $(1-R)^{-1}$ יהיה $[1+\theta_L, 1+\theta_U]$; ר"ס עבור $(1-R)$ יהיה $\left[\frac{1}{1+\theta_L}, \frac{1}{1+\theta_U}\right]$; ר"ס עבור R יהיה $\left[1 - \frac{1}{1+\theta_L}, 1 - \frac{1}{1+\theta_U}\right] = \left[\frac{\theta_L}{1+\theta_L}, \frac{\theta_U}{1+\theta_U}\right]$. ר"ס עבור σ_a^2: $\sigma_a^2 = \omega\theta = \phi_2\phi_3 \to \hat{\sigma}_a^2 = \hat{\phi}_2\hat{\phi}_3$.

תזכורת – שיטת דלתא: עבור מ"מ כלשהו ψ מתקיים $\hat{\psi} \sim N(\psi, \tau^2)$, נסמן $\eta = g(\psi)$ אז $\hat{\eta} = g(\hat{\psi})$ מקיים $\hat{\eta} \sim N\left(\eta, \left[g'(\hat{\psi})\right]^2 \tau^2\right)$.

בעת נגדיר $\eta = g(\phi), g: \mathbb{R}^p \to \mathbb{R}$ אז $\hat{\eta} = g(\hat{\phi})$, אם $\nabla g = \begin{bmatrix} \frac{\partial g}{\partial \phi_1} \\ \vdots \\ \frac{\partial g}{\partial \phi_p} \end{bmatrix}$ וקטור הגרדיאנט אז $\hat{\eta} \sim N\left(\eta, \nabla g(\hat{\phi})^T V, \nabla g(\hat{\phi})\right)$. עבור $\eta = \begin{bmatrix} g_1(\phi) \\ \vdots \\ g_m(\phi) \end{bmatrix}$ מטריצת היעקוביאן היא $J = \begin{bmatrix} \frac{\partial g_1}{\partial \phi_1} & \cdots & \frac{\partial g_m}{\partial \phi_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial \phi_p} & \cdots & \frac{\partial g_m}{\partial \phi_p} \end{bmatrix}$ ומתקיים $\hat{\eta} \sim N\left(\eta, J(\hat{\phi})^T V, J(\hat{\phi})\right)$. בעת נסמן $\tilde{\eta} = \log\eta = \log\phi_2 + \log\phi_3 = g(\phi)$ אז $\nabla g = \begin{bmatrix} 0 \\ \frac{1}{\phi_2} \\ \frac{1}{\phi_3} \end{bmatrix}$ ונקבל $\hat{\tilde{\eta}} = g(\hat{\phi}) \sim N\left(\tilde{\eta}, \widehat{Var}(\hat{\tilde{\eta}})\right)$. מתקיים $\widehat{Var}(\hat{\tilde{\eta}}) = \begin{bmatrix} 0 & \frac{1}{\hat{\phi}_2} & \frac{1}{\hat{\phi}_3} \end{bmatrix} \hat{V} \begin{bmatrix} 0 \\ \frac{1}{\hat{\phi}_2} \\ \frac{1}{\hat{\phi}_3} \end{bmatrix}$ ולכן ר"ס עבור $\tilde{\eta}$ יהיה $g(\hat{\phi}) \pm z_{1-\frac{\alpha}{2}}\sqrt{\widehat{Var}(\hat{\tilde{\eta}})} = [\tilde{\eta}_L, \tilde{\eta}_U]$. מכאן כי ר"ס עבור η יהיה $[\eta_L, \eta_U] = \left[e^{\tilde{\eta}_L}, e^{\tilde{\eta}_U}\right]$ וניתן להסיק את רווח הסמך עבור σ_a^2.

דרכים אפשריות לחישוב $\hat{V} = \left(E\left[\frac{-\partial^2 l}{\partial\phi_r \partial\phi_s}\right]\right)^{-1}$ באמצעות הנגזרות השניות:

$$E[(\bar{y}_{i.} - \mu)^2)] = Var(\bar{y}_{i.}) = \sigma_a^2 + \frac{\sigma_\epsilon^2}{n_i} = \frac{\omega(1 + n_i\theta)}{n_i}, \qquad E[SSE] = (N - I)\omega$$

$$\frac{\partial^2 l}{\partial \theta^2} = \frac{1}{2}\sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}\right)^2 - \frac{1}{\omega}\sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}\right)^3 (\bar{y}_{i.} - \mu)^2 \to E\left[-\frac{\partial^2 l}{\partial \theta^2}\right]$$
$$= -\frac{1}{2}\sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}\right)^2 + \frac{1}{\omega}\sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}\right)^3 \frac{\omega(1 + n_i\theta)}{n_i} = \frac{1}{2}\sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}\right)^2 \to \hat{V}$$
$$= \frac{2}{\sum_{i=1}^{I}\left(\frac{n_i}{1 + n_i\theta}\right)^2}$$

$$\frac{\partial^2 l}{\partial\omega^2} = \frac{N}{2\omega^2} - \frac{1}{\omega^3}\left[SSE + \sum_{i=1}^{I}\left(\frac{n_i}{1+n_i\theta}(\bar{y}_{i.}-\mu)^2\right)\right] \to E\left[-\frac{\partial^2 l}{\partial\omega^2}\right]$$

$$= -\frac{N}{2\omega^2} + \frac{1}{\omega^3}\left[(N-I)\omega + \sum_{i=1}^{I}\left(\frac{n_i}{1+n_i\theta}\frac{\omega(1+n_i\theta)}{n_i}\right)\right]$$

$$= -\frac{N}{2\omega^2} + \frac{1}{\omega^3}[N\omega - I\omega + I\omega] = -\frac{N}{2\omega^2} + \frac{N}{\omega^2} = \frac{N}{2\omega^2} \to \hat{V} = \frac{2\omega^2}{N}$$

$$\frac{\partial^2 l}{\partial\omega\partial\theta} = -\frac{1}{2\omega^2}\sum_{i=1}^{I}\left(\frac{n_i}{1+n_i\theta}\right)^2(\bar{y}_{i.}-\mu)^2 \to E\left[-\frac{\partial^2 l}{\partial\omega\partial\theta}\right] = \frac{1}{2\omega^2}\sum_{i=1}^{I}\left(\frac{n_i}{1+n_i\theta}\right)^2\frac{\omega(1+n_i\theta)}{n_i}$$

$$= \frac{1}{2\omega}\sum_{i=1}^{I}\frac{n_i}{1+n_i\theta} \to \hat{V} = \frac{2\omega}{\sum_{i=1}^{I}\frac{n_i}{1+n_i\theta}}$$

רגרסיה לוגיסטית – Logistic Regression

יש לנו משתנים חוזים $x_{1i}, x_{2i}, \dots, x_{pi}$ וגם $x_{0i} = 1$. המשתנה החזוי $y_i = \begin{cases} 1 \\ 0 \end{cases}$. מודל רגרסיה לינארית שהכרנו $y_i = x_i^T \beta + \epsilon_i$ לא מתאים כאן, נחפש מודל אחר.

נרשום $E[y_i|x_i] = P(y_i = 1|x_i) = g\left(\sum_{j=0}^{p} \beta_j x_{ij}\right)$, כאשר $g: (-\infty, \infty) \to (0,1)$ פונקציה עולה. אם נבחר $g(n) = \Phi(n)$ נקבל את מודל Probit Regression, אם ניקח $g(n) = \frac{e^n}{1+e^n}$ נקבל את מודל Logistic Regression.

זהו מקרה פרטי של מודל לינארי מוכלל (GLM – Generalized Linear Model), במודלים מסוג זה מניחים $E[y_i] = g(\beta^T x_i)$, את הפונקציה h המקיימת $h(E[y_i]) = \beta^T x_i$ מכנים Link Function. עבור Probit מתקיים $h(w) = \Phi^{-1}(w)$, עבור לוגיסטית $h(w) = \log\left(\frac{w}{1-w}\right) = logit(w)$.

נאמוד את β ע"י ML:

$$f_{y_i} = \begin{cases} g(\beta^T x_i) & y = 1 \\ 1 - g(\beta^T x_i) & y = 0 \end{cases} \to L(\beta) = \prod_{i=1}^{n} f_{y_i}(y_i)$$

$$f_{y_i} = \begin{cases} p & y = 1 \\ 1 - p & y = 0 \end{cases} = p^y (1-p)^{1-y} \to L(\beta) = \prod_{i=1}^{n} \left(g(\beta^T x_i)\right)^{y_i} \left(1 - g(\beta^T x_i)\right)^{1-y_i}$$

$$l = \log L(\beta) = \sum_{i=1}^{n} \log\left(\left(g(\beta^T x_i)\right)^{y_i} \left(1 - g(\beta^T x_i)\right)^{1-y_i}\right)$$
$$= \sum_{i=1}^{n} y_i \log(\beta^T x_i) + \sum_{i=1}^{n} (1 - y_i) \log\left(1 - g(\beta^T x_i)\right)$$

$$\frac{\partial l}{\partial \beta_r} = 0, r = 0, \dots, p: \quad \frac{\partial}{\partial \beta_r} \log\left(g(\beta^T x_i)\right) = \frac{1}{g(\beta^T x_i)} g'(\beta^T x_i) \left[\frac{\partial}{\partial \beta_r} \beta^T x_i\right]$$

$$\frac{\partial}{\partial \beta_r} \beta^T x_i = \frac{\partial}{\partial \beta_r} \sum_{j=0}^{p} \beta_j x_{ij} = \sum_{j=0}^{p} \frac{\partial \beta_j}{\partial \beta_r} x_{ij}, \frac{\partial \beta_j}{\partial \beta_r} = \begin{cases} 1 & j = r \\ 0 & j \neq r \end{cases} \to \frac{\partial}{\partial \beta_r} \beta^T x_i = \begin{cases} \sum_{j=0}^{p} x_{ir} & j = r \\ 0 & j \neq r \end{cases}$$

$$\to \frac{\partial l}{\partial \beta_r} = \sum_{i=1}^{n} y_i x_{ir} \frac{g'(\beta^T x_i)}{g(\beta^T x_i)} - \sum_{i=1}^{n} (1 - y_i) x_{ir} \frac{g'(\beta^T x_i)}{1 - g(\beta^T x_i)}$$

מקבלים מערכת משוואות לא לינארית, כדי לפתור צריך שיטה נומרית – ניוטון רפסון (❤) רב-ממדי:

יש לנו m משוואות בצורה $\begin{matrix} Q_1(u_1, \dots, u_m) = 0 \\ \vdots \\ Q_m(u_1, \dots, u_m) = 0 \end{matrix}$ וידועה נקודה $u^{(0)}$ (או כמה נקודות כאלו). נסמן $Q = \begin{bmatrix} Q_1 \\ \vdots \\ Q_m \end{bmatrix}, u = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}$ ומטריצת היעקוביאן $J = \begin{bmatrix} \frac{\partial Q_1}{\partial u_1} & \cdots & \frac{\partial Q_m}{\partial u_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial Q_1}{\partial u_m} & \cdots & \frac{\partial Q_m}{\partial u_m} \end{bmatrix}$ אז הפתרון מתקבל באמצעות $u^{(k+1)} = u^{(k)} - J^{-1}\left(u^{(k)}\right) Q\left(u^{(k)}\right)$. בהקשר של MLE מתקיים $Q_r = \frac{\partial l}{\partial \beta_r}$ לכן מטריצת היעקוביאן של Q היא

למעשה מטריצת ההסיאן (Hessian, מטריצת נגזרות שניות) של l (כי $\frac{\partial Q_1}{\partial u_1}=\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}$) ומתקיים $\beta^{(k+1)}=\beta^{(k)}-([H(l)]^{-1}Q)|_{\beta=\beta^{(k)}}$. נחשב את מטריצת ההסיאן:

$$H(u)=\frac{g'(u)}{g(u)},Q(u)=\frac{g'(u)}{1-g(u)}\rightarrow\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}$$
$$=\sum_{i=1}^n y_ix_{ir}x_{is}H'(\beta^Tx_i)-\sum_{i=1}^n(1-y_i)x_{ir}x_{is}Q'(\beta^Tx_i)$$
$$=\sum_{i=1}^n x_{ir}x_{is}[y_iH'(\beta^Tx_i)-(1-y_i)Q'(\beta^Tx_i)]=-\sum_{i=1}^n x_{ir}x_{is}W_i(\beta)$$

נסמן גם $R_i(\beta)=y_iH(\beta^Tx_i)-(1-y_i)Q(\beta^Tx_i)$ ונקבל:

$$\frac{\partial l}{\partial\beta_r}=\sum_{i=1}^n y_ix_{ir}H(\beta^Tx_i)-\sum_{i=1}^n(1-y_i)x_{ir}Q(\beta^Tx_i)=\sum_{i=1}^n x_{ir}R_i(\beta)$$

יהיו $W(\beta)=\begin{bmatrix}W_1(\beta) & & 0\\ & \ddots & \\ 0 & & W_n(\beta)\end{bmatrix},R(\beta)=\begin{bmatrix}R_1\\ \vdots\\ R_n\end{bmatrix},X=\begin{bmatrix}x_{10} & \cdots & x_{1p}\\ \vdots & \ddots & \vdots\\ x_{n0} & \cdots & x_{np}\end{bmatrix}$, לכן מתקיים

$$\nabla l(\beta)=X^TR,\nabla^2 l(\beta)=\left[\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}\right]=-X^TWX$$

נקבל בניוטון-רפסון:

$$\beta^{(k+1)}=\beta^{(k)}-([\nabla^2 l]^{-1}\nabla l)|_{\beta=\beta^{(k)}}=\beta^{(k)}-((X^TWX)^{-1}X^TR)|_{\beta=\beta^{(k)}}$$

$$R_i(\beta)=y_iH(\beta^Tx_i)-(1-y_i)Q(\beta^Tx_i)=y_i[H(\beta^Tx_i)+Q(\beta^Tx_i)]-Q(\beta^Tx_i)$$
$$=y_i\Omega(\beta^Tx_i)-Q(\beta^Tx_i)=\Omega(\beta^Tx_i)\left[y_i-\frac{Q(\beta^Tx_i)}{\Omega(\beta^Tx_i)}\right]=\Omega(\beta^Tx_i)[y_i-g(\beta^Tx_i)]$$

עבור רגרסיה לוגיסטית מתקיים $g(u)=\frac{e^u}{1+e^u}\rightarrow 1-g(u)=\frac{1}{1+e^u}\rightarrow g'(u)=g(u)\big(1-g(u)\big)$ אז $H(u)=\frac{g'(u)}{g(u)}=\frac{g(u)(1-g(u))}{g(u)}=\big(1-g(u)\big)=\frac{1}{1+e^u}$ ובן $Q(u)=\frac{g'(u)}{1-g(u)}=\frac{g(u)(1-g(u))}{1-g(u)}=g(u)=\frac{e^u}{1+e^u}$. סימנו $\Omega(u)=H(u)+Q(u)$ לכן $\Omega(u)=\big(1-g(u)\big)+g(u)=1$, מכאן נקבל:

$$R_i(\beta)=y_i-g(\beta^Tx_i)=y_i-\frac{e^{\beta^Tx_i}}{1+e^{\beta^Tx_i}}$$

$$W_i(\beta)=y_iH'(\beta^Tx_i)-(1-y_i)Q'(\beta^Tx_i)=\Omega(\beta^Tx_i)g'(\beta^Tx_i)-\Omega'(\beta^Tx_i)\big(y_i-g(\beta^Tx_i)\big)$$

ראינו כי עבור רגרסיה לוגיסטית $\Omega(u)=1$, לכן $\Omega'=0$ ונקבל בסך הכל:

$$W_i(\beta)=1\cdot g'(\beta^Tx_i)-0\cdot\big(y_i-g(\beta^Tx_i)\big)=g'(\beta^Tx_i)=g(\beta^Tx_i)\big(1-g(\beta^Tx_i)\big)$$
$$=\frac{e^{\beta^Tx_i}}{\big(1+e^{\beta^Tx_i}\big)^2}$$

מתקיים $E\left[\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}\right] = -\sum_{i=1}^{n} x_{ir}x_{is}\Omega(\beta^T x_i)g'(\beta^T x_i)$, נסמן $W_i^*(\beta) = \Omega(\beta^T x_i)g'(\beta^T x_i)$ אז $E\left[\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}\right] = -\sum_{i=1}^{n} x_{ir}x_{is}W_i^*(\beta)$ ובמקרה של רגרסיה לוגיסטית נקבל $W_i(\beta) = W_i^*(\beta) =$ $g'(\beta^T x_i) = g(\beta^T x_i)\left(1-g(\beta^T x_i)\right)$. ראינו כי בשיטת ניוטון-רפסון

$$\beta^{(m+1)} = \beta^{(m)} - \left(\nabla^2 l\left(\beta^{(m)}\right)\right)^{-1} \nabla l\left(\beta^{(m)}\right) = \beta^{(m)} + \left(X^T W\left(\beta^{(m)}\right)X\right)^{-1} X^T R(\beta)$$

ובסך הכל מתקבל עבור אומד MLE:

$$\hat{\beta} \sim N\left(\beta, (-E[\nabla^2 l])^{-1}\right) = N\left(\beta, (X^T W^*(\beta)X)^{-1}\right) \to \hat{\beta} \sim N(\beta, V) \to V = (X^T W^*(\beta)X)^{-1}$$
$$\to \hat{V} = \left(X^T W^*(\hat{\beta})X\right)^{-1}$$

עבור ההסתברות $\hat{p}(x) = g(\hat{\beta}^T x)$ נבנה רווח סמך באופן $[g(\theta_L), g(\theta_U)]$:

$$\theta = \beta^T x = x^T\beta \to \hat{\theta} = x^T\hat{\beta} \to \hat{\theta} \sim N(\theta, x^T V x) \to \theta \in \left[\hat{\theta} \pm z_{1-\frac{\alpha}{2}}\sqrt{x^T V x}\right]$$
$$\to \begin{cases} \theta_L = \hat{\theta} - z_{1-\frac{\alpha}{2}}\sqrt{x^T V x} \\ \theta_U = \hat{\theta} + z_{1-\frac{\alpha}{2}}\sqrt{x^T V x} \end{cases}$$
$$\to p(x) \in \left[g\left(\hat{\theta} - z_{1-\frac{\alpha}{2}}\sqrt{x^T V x}\right), g\left(\hat{\theta} + z_{1-\frac{\alpha}{2}}\sqrt{x^T V x}\right)\right]$$

מנת יחס סיכויים (Odds Ratio)

נניח כי A הוא מאורע כלשהו אז **יחס הסיכויים** (odds) מוגדר $\frac{P(A)}{1-P(A)} = \frac{P(A)}{P(A^c)}$. במודל הלוגיסטי $P(y=1|x=x) = \frac{e^{\beta^T x}}{1+e^{\beta^T x}}$ לכן $P(y=0|x=x) = 1 - \frac{e^{\beta^T x}}{1+e^{\beta^T x}} = \frac{1}{1+e^{\beta^T x}}$ ויחס הסיכויים הוא $e^{\beta^T x}$.

אם יש לנו וקטורים x_1, x_2 אז **מנת יחס הסיכויים** (odds ratio) מוגדרת באופן

$$\frac{\frac{P(y=1|x=x_2)}{P(y=0|x=x_2)}}{\frac{P(y=1|x=x_1)}{P(y=0|x=x_1)}}$$

היחס $p = \frac{3}{4}$ מתאר סיכוי הצלחה של 3:1. מכאן $\frac{p}{1-p} = \frac{3/4}{1-3/4} = \frac{3/4}{1/4} = \frac{3}{1}$. ברגרסיה לוגיסטית נקבל $\frac{e^{\beta^T x_2}}{e^{\beta^T x_1}} = e^{\beta^T(x_2-x_1)}$. אם ניקח וקטור x_2 שנבדל מ-x_1 רק בגודל c ברכיב ה-k אז מנת יחס הסיכויים היא $e^{\beta_k c}$.

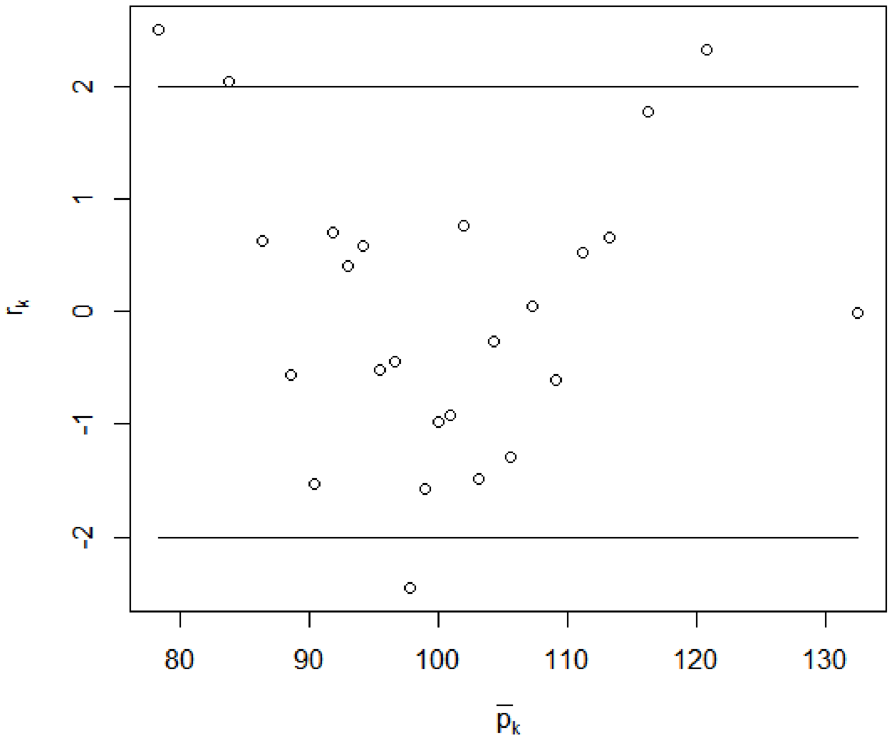
נסמן $\Delta = x_2 - x_1$, ר"ס עבור $e^{\beta^T\Delta}$: $\left[e^{\psi_L}, e^{\psi_U}\right]$

$$e^{\beta^T\Delta} = e^{\Delta^T\beta} \to \psi = \Delta^T\beta \to \hat{\psi} = \Delta^T\hat{\beta} \sim N(\psi, \Delta^T V\Delta) \to \begin{cases} \psi_L = \hat{\psi} - z_{1-\frac{\alpha}{2}}\sqrt{\Delta^T V\Delta} \\ \psi_L = \hat{\psi} + z_{1-\frac{\alpha}{2}}\sqrt{\Delta^T V\Delta} \end{cases}$$
$$\to e^{\beta^T\Delta} \in \left[e^{\hat{\psi} - z_{1-\frac{\alpha}{2}}\sqrt{\Delta^T V\Delta}}, e^{\hat{\psi} + z_{1-\frac{\alpha}{2}}\sqrt{\Delta^T V\Delta}}\right]$$

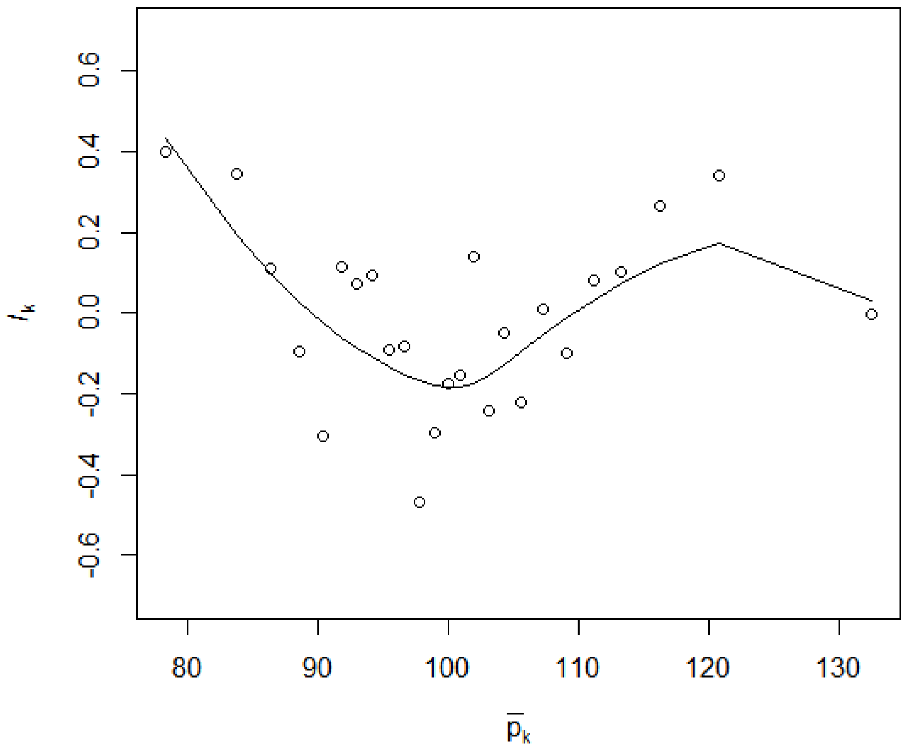
ממשים רגרסיה לוגיסטית ב-R ע"י הפקודה `glm(y~x_1+...+x_p,family=binomial(link=logit))`.

בחינת הנחות מודל

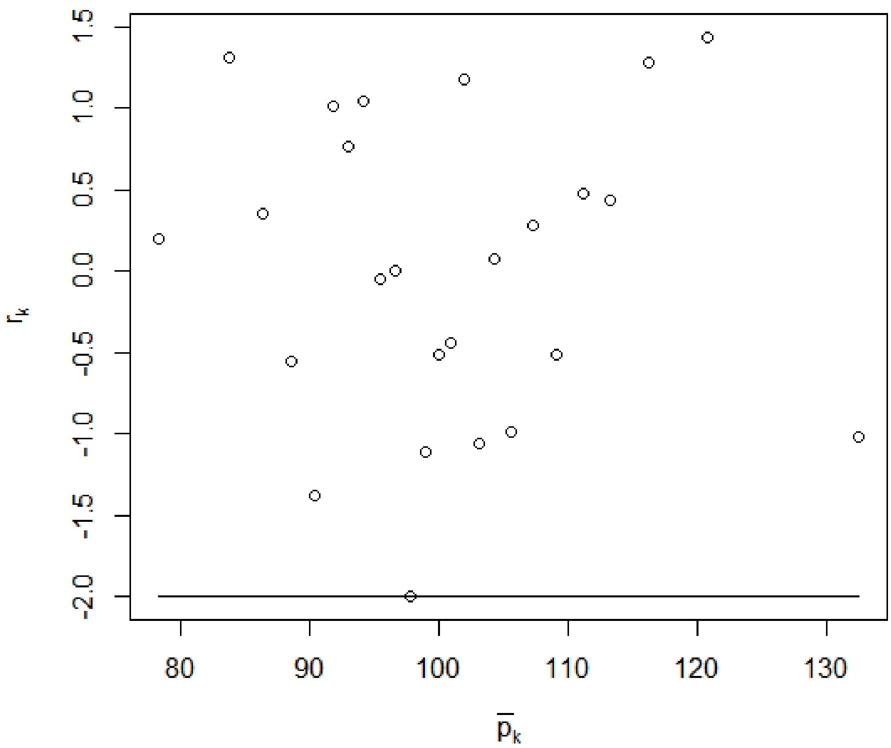
נסמן שאריות $e_i = y_i - \hat{y}_i$, בניגוד למודל הלינארי, הגרף של e_i מול x_{ij} לא ילמד אותנו הרבה. נסמן I_k קטעים במדגם, את האוסף $C_k = \{i: i \in I_k\}$, $n_k = \#C_k$ (מספר התצפיות). נסמן גם $\bar{y}_k = \frac{1}{n_k}\sum_{i\in C_k} y_i$, $\bar{p}_k = \frac{1}{n_k}\sum_{i\in C_k} \hat{p}_i \approx p_k$, $\bar{x}_{kj} = \frac{1}{n_k}\sum_{i\in C_k} x_{ij}$ אז $n_k\bar{y}_k = \sum_{i\in C_k} y_i \sim Bin(n_k, p_k)$ ומתקיים $\frac{\bar{y}_k - p_k}{\sqrt{\frac{p_k(1-p_k)}{n_k}}} \sim N(0,1)$, לכן $r_k = \frac{\bar{y}_k - \bar{p}_k}{\sqrt{\frac{\bar{p}_k(1-\bar{p}_k)}{n_k}}} \sim N(0,1)$ ונוכל לשרטט גרף של r_k מול $\bar{p}_k$ כדי למצוא חריגות:

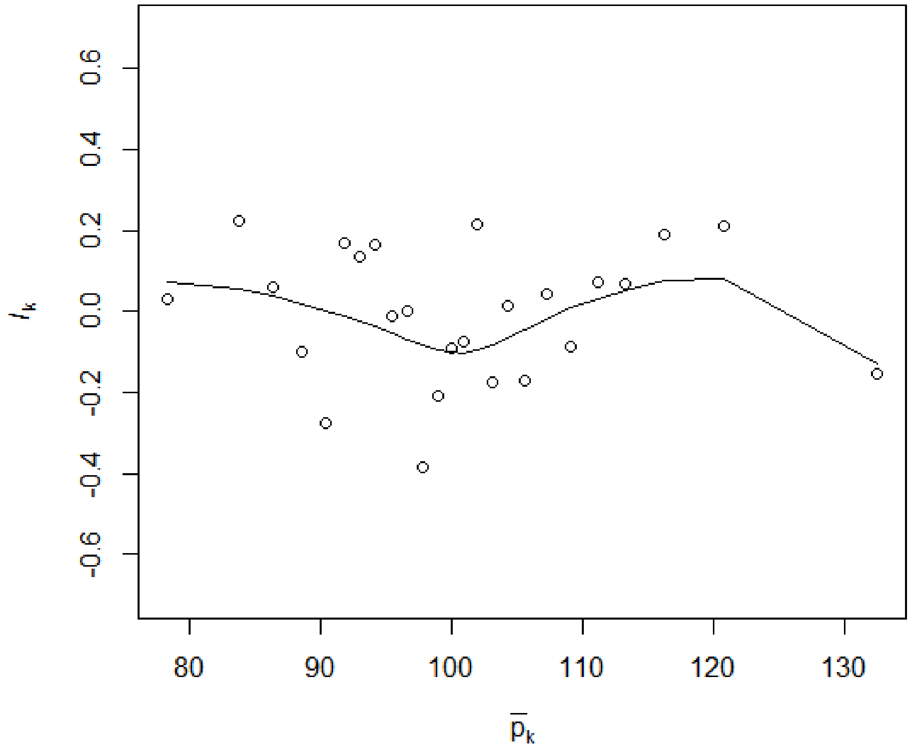


אם נסמן $l_k = logit(\bar{y}_k) - logit(\bar{p}_k)$, גרף של l_k מול $\bar{p}_k$ עם התאמה לא-פרמטרית של פולינום (למשל ע"י LOESS) ילמד אותנו על נקודות משפיעות שדורשות הוספת גורמים למודל:



בגרף הנ"ל ניתן לראות כי יש לפחות שתי נקודות משפיעות; הוספה של שני גורמים נוספים (במקרה הזה חזקות שניה ושלישית של המשתנה map) תמתן את ההשפעה ואף תסדיר חריגות:





אין כלל מדויק לבחירת K, בדרך כלל מנסים כמה ערכים שונים כדי לבחור איזה מתאים יותר.

ניתן לבחון טיב התאמה כללי ע"י הסטטיסטי $\chi^2_{res} = \sum_{k=1}^{K} r_k^2$ שתחת השערת האפס מתפלג χ^2- מתואר בספרים הבאים:

Cox, D. R., & Snell, E. J. (1989). Analysis of binary data (Vol. 32). CRC Press.

Hosmer Jr, D. W., & Lemeshow, S. (2000). Applied logistic regression. John Wiley & Sons.

Cox-Snell מציעים (עמ' 72) להניח $K-(p+1)$ דרגות חופש בהתפלגות χ^2, זה יעבוד רק אם יש מספיק דגימות כך שגם K יהיה גדול דיו. Hosmer-Lemeshow מציעים מבחן אחר, הידוע בשם Hosmer-Lemeshow Test (מפתיע!).

כמו הגרפים שתוארו כאן, ניתן לשרטט גם גרפים של e_i מול $\bar{x}_{ij}$ עבור משתנה חוזה מסוים, כך ניתן לזהות משתנים 'בעייתיים'. בציר ה-y משתמשים ב-r_k, l_k המחושבים כמו קודם, בציר ה-x משתמשים בערכי $\bar{x}_{jk} = \frac{1}{n_k}\sum_{i \in C_k} x_{ij}$.

רגרסיה פואסונית – Poisson Regression

יש לנו x_i כרגיל, הפעם $y_i \in \{0,1,2,3,\ldots\}$. מניחים $y_i \sim Poi(\lambda_i)$ כאשר $\lambda_i = g(\beta^T x_i)$. בגלל המודל הפואסוני מתקיים $\lambda_i = E[y_i]$, הפונקציה g צריכה להיות פונקציה עולה ולקיים $g: (-\infty, \infty) \to [0, \infty)$ – הבחירה הפופולרית היא $g(u) = e^u$. נאמוד MLE עבור β:

$$L(\beta) = \prod_{i=1}^{n} f_{y_i}(y_i) = \prod_{i=1}^{n} \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \to l = \log\big(L(\beta)\big) = \sum_{i=1}^{n} y_i \log \lambda_i - \sum_{i=1}^{n} \lambda_i - \sum_{i=1}^{n} \log(y_i!)$$
$$= \sum_{i=1}^{n} y_i \log\big(g(\beta^T x_i)\big) - \sum_{i=1}^{n} g(\beta^T x_i) - \sum_{i=1}^{n} \log(y_i!)$$

כמו ברגרסיה לוגיסטית, נסמן $\phi(u) = \log\big(g(u)\big)$:

$$\frac{\partial l}{\partial \beta_r} = \sum_{i=1}^{n} y_i \phi'(\beta^T x_i) x_{ir} - \sum_{i=1}^{n} g'(\beta^T x_i) x_{ir} = \sum_{i=1}^{n} x_{ir} R_i \to R_i = y_i \phi'(\beta^T x_i) - g'(\beta^T x_i)$$
$$= y_i \phi'(\beta^T x_i) \left[y_i - \frac{g'(\beta^T x_i)}{\phi'(\beta^T x_i)} \right]$$

מתקיים $\phi'(u) = \frac{g'(u)}{g(u)}$, לכן $\frac{g'(u)}{\phi'(u)} = g(u)$. נסמן $\Omega(u) = \phi'(u)$ ונקבל:

$$y_i \phi'(\beta^T x_i) \left[y_i - \frac{g'(\beta^T x_i)}{\phi'(\beta^T x_i)} \right] = \Omega(\beta^T x_i)\big(y_i - g(\beta^T x_i)\big) = R_i(\beta)$$

$$\to \frac{\partial l}{\partial \beta_r} = \sum_{i=1}^{n} x_{ir} R_i = \sum_{i=1}^{n} x_{ir} \cdot \Omega(\beta^T x_i)\big(y_i - g(\beta^T x_i)\big)$$

הנגזרת השנייה:

$$\frac{\partial^2 l}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^{n} x_{ir} \frac{\partial l}{\partial \beta_s} R_i \to \frac{\partial l}{\partial \beta_s} R_i = \Omega'(\beta^T x_i)\big(y_i - g(\beta^T x_i)\big) x_{is} - \Omega(\beta^T x_i) g'(\beta^T x_i) x_{is}$$
$$= -W_i(\beta) x_{is} \to W_i(\beta) = \Omega(\beta^T x_i) g'(\beta^T x_i) - \Omega'(\beta^T x_i)\big(y_i - g(\beta^T x_i)\big)$$
$$\to \frac{\partial^2 l}{\partial \beta_r \partial \beta_s} = -\sum_{i=1}^{n} W_i(\beta) x_{ir} x_{is} \to E\left[\frac{\partial^2 l}{\partial \beta_r \partial \beta_s}\right] = -\sum_{i=1}^{n} x_{ir} x_{is} E[W_i(\beta)]$$
$$= -\sum_{i=1}^{n} x_{ir} x_{is} E\big[\Omega(\beta^T x_i) g'(\beta^T x_i) - \Omega'(\beta^T x_i)\big(y_i - g(\beta^T x_i)\big)\big]$$
$$= -\sum_{i=1}^{n} x_{ir} x_{is} \Big(E[\Omega(\beta^T x_i) g'(\beta^T x_i)] - E\big[\Omega'(\beta^T x_i)\big(y_i - g(\beta^T x_i)\big)\big]\Big)$$
$$= -\sum_{i=1}^{n} x_{ir} x_{is} \big(\Omega(\beta^T x_i) g'(\beta^T x_i) - E[\Omega'(\beta^T x_i)(y_i - E[y_i])]\big)$$
$$= -\sum_{i=1}^{n} x_{ir} x_{is} \Omega(\beta^T x_i) g'(\beta^T x_i) = -\sum_{i=1}^{n} x_{ir} x_{is} W_i^*(\beta)$$

כאשר $W_i^*(\beta) = \Omega(\beta^T x_i) g'(\beta^T x_i)$. במקודם, מטריצות הנגזרות יקיימו

$$\nabla l = X^T R(\beta), \qquad \nabla^2 l = -X^T W(\beta) X, \qquad E[\nabla^2 l] = -X^T W^*(\beta) X$$

ושוב נקבל את נוסחאות ניוטון-רפסון בצורה $\beta^{(m+1)}=\beta^{(m)}+\left(X^TW\left(\beta^{(m)}\right)X\right)^{-1}X^TR\left(\beta^{(m)}\right)$, מכאן כי $\hat{V}=\left(X^TW^*\left(\hat{\beta}\right)X\right)^{-1}$, $V=(X^TW^*(\beta_{true})X)^{-1}$, $\hat{\beta}\sim N(\beta,V)$.

לא במקרה קיבלנו נוסחאות זהות לאלו שבמקרה הלוגיסטי – זה נובע מכך ששני המודלים שייכים למשפחת המודלים GLM – Generalized Linear Models.

במקרה שבו $g(u)=e^u$ (הנפוץ ביותר ברגרסיה פואסונית):

$$g(u)=e^u\to g'(u)=e^u,\phi(u)=\log\left(g(u)\right)=u\to\phi'(u)=1\to\Omega(u)=\phi'(u)=1$$

והנגזרות:

$$\frac{\partial l}{\partial\beta_r}=\sum_{i=1}^{n}x_{ir}\left(y_i-e^{\beta^Tx_i}\right),\qquad\frac{\partial^2 l}{\partial\beta_r\partial\beta_s}=-\sum_{i=1}^{n}x_{ir}x_{is}e^{\beta^Tx_i}$$

אם תחת $\hat{\beta}\sim N(\beta,V)$ נרצה לבדוק את ההשערה $H_0:\beta_j=0$ אז נוכל להשתמש בבך ש- $\frac{\hat{\beta}_j}{\sqrt{\hat{V}_{jj}}}\overset{H_0}{\sim}N(0,1)$,

$\mu(x)=E[Y|X=x]=g(\beta^Tx)=g(x^T\beta)$, במקודם נבנה רווח סמך בשלבים:

$$\theta=x^T\beta\to\hat{\theta}=x^T\hat{\beta}\sim N(\theta,x^TVx)\to\begin{cases}\theta_L=\hat{\theta}-z_{1-\frac{\alpha}{2}}\sqrt{x^TVx}\\\theta_U=\hat{\theta}+z_{1-\frac{\alpha}{2}}\sqrt{x^TVx}\end{cases}\to\theta\in[g(\theta_L),g(\theta_U)]$$

ב-R: glm(y~x_1+...+x_p,family=poisson(link=log)).

מודלים לוג-ליניאריים – Log-Linear Models

תגובות בינאריות בלוחות שכיחות

(פרופ' צוקר היה חולה ולכן הנושא נלמד ע"י פרופ' שמואל אומן)

בעיתון רפואי הופיע מחקר לפיו נטילת תרופה לדלקת שרירים מובילה לירידה בסיכון לחלות באלצהיימר.
במחקר היו 6989 נבדקים מעל גיל 55, התוצאות סוכמו בטבלא:

אלצהיימר / תרופה	לא חלה	כן חלה	
לא נטל	2343	210	2553
כן נטל	4353	83	4436
	6696	293	6989

את ננתח לפי התרופה כגורם, באופן יחסי נקבל את הטבלא הבאה:

אלצהיימר / תרופה	לא	כן
לא	0.92	0.08
כן	0.98	0.02

נסמן באופן הבא:

		B		
		1	2	
A	1	f_{11}	f_{12}	$f_{1.}$
	2	f_{21}	f_{22}	$f_{2.}$
		$f_{.1}$	$f_{.2}$	$f_{..} = n$

נרצה לבדוק את ההשערה H_0 לפיה A,B ב"ת: נחשב שכיחויות צפויות תחת H_0: $e_i = \frac{f_{i.}f_{.j}}{f_{..}}$

$$e_{ij} = \begin{pmatrix} 2446 & 107 \\ 4250 & 186 \end{pmatrix} \rightarrow \chi^2_{obs} = \sum_{i,j} \frac{\left(f_{ij} - e_{ij}\right)^2}{e_{ij}} = 162.91$$

תחת H_0 מתקיים כי $\chi^2_{obs} \sim \chi^2$ ולכן נדחה את H_0.

יחד עם זאת, הטענה במחקר היתה גם על כיוון הקשר, נרצה למצוא פרמטר לכיוון – נניח כי A,B מ"מ באשר:

	B		
	1	2	
A 1	π_{11}	π_{12}	$\pi_{1.}$
2	π_{21}	π_{22}	$\pi_{2.}$
	$\pi_{.1}$	$\pi_{.2}$	$\pi_{..}=1$

הגדרנו בעבר את יחס הסיכויים $\frac{p}{1-p}$, כאשר $0<p<1$. יחס הסיכויים מקיים $0<\frac{p}{1-p}<\infty$, לכן הלוג שלו יקיים $-\infty<\frac{p}{1-p}<\infty$. אם נסמן $\theta=\frac{p}{1-p}$ אז נוכל "לפרש" לפיו את התלות בין A ו-B: אם $\theta=1$ אין תלות; אם $\theta>1$ יש תלות חיובית, כלומר $A=1$ מגדיל את הסיכוי לקבל $B=1$; אם $\theta<1$ יש תלות שלילית, כלומר $A=1$ מקטין את הסיכוי לקבל $B=1$. נרצה לבדוק את מנת יחס הסיכויים:

$$\alpha=\frac{P(B=1|A=1)}{P(B=1|A=2)}=\frac{\pi_{11}/\pi_{1.}}{\pi_{21}/\pi_{2.}}$$

ובמונחים של π_{ij}:

$$\alpha=\frac{odds(B=1|A=1)}{odds(B=1|A=2)}=\frac{P(B=1|A=1)/P(B=2|A=1)}{P(B=1|A=2)/P(B=2|A=2)}=\frac{\left(\frac{\pi_{11}}{\pi_{1.}}\right)/\left(\frac{\pi_{12}}{\pi_{1.}}\right)}{\left(\frac{\pi_{21}}{\pi_{2.}}\right)/\left(\frac{\pi_{22}}{\pi_{2.}}\right)}=\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}$$

הפרמטר α נקרא cross-product ratio. אם $\alpha=1$ נסיק כי אין קשר בין A,B; אם $A>1$ אז יש קשר חיובי; אם $\alpha<1$ יש קשר שלילי. גם לגודל של α יש משמעות.

$$\hat{\alpha}=\frac{f_{11}f_{22}}{f_{12}f_{21}}=0.2127$$

קיבלנו כי יש קשר שלילי בין 'לא לחלות $(B=1)$ ו'לא לקחת תרופה' $(A=1)$, כלומר קשר חיובי בין לא לחלות $(B=1)$ ו'כן לקחת תרופה' $(B=2)$. השאלה הבאה שנרצה לבדוק היא האם התוצאה הזו מובהקת, לשם כך נצטרך להבין מה ההתפלגות של $\hat{\alpha}$: נניח כי f_{ij} תוצאה של n תצפיות ב"ת מההתפלגות של π_{ij} של A,B:

	x_1	x_2	...	x_n	Sum
$\pi_{11}(A=1,B=1)$	1	0	...	0	f_{11}
$\pi_{12}(A=1,B=2)$	0	0	...	1	f_{12}
$\pi_{21}(A=2,B=1)$	0	1	...	0	f_{21}
$\pi_{22}(A=2,B=2)$	0	0	...	0	f_{22}

כאשר וקטור הסכומים מתפלג באופן $\begin{bmatrix} f_{11} \\ f_{12} \\ f_{21} \\ f_{22} \end{bmatrix} \sim Mult\left(n, \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{21} \\ \pi_{22} \end{bmatrix}\right)$. באמצעות פיתוח (חסר) של טור טיילור ו-CLT מקבלים כי $\log(\hat{\alpha}) \underset{n\to\infty}{\sim} N\left(\log(\alpha), \frac{1}{n}\sum_{i,j}\frac{1}{\pi_{ij}}\right)$. נסמן $\theta = \log(\alpha)$, הפרמטר הזה נקרא log-odds ratio. $\hat{\theta} = \log(\hat{\alpha})$ הוא ה-log-odds ratio במדגם ומתקיים עבורו $\hat{\theta} \sim N\left(\theta, \sum_{i,j}\frac{1}{f_{ij}}\right)$. בדוגמא קיבלנו $\hat{\alpha} = 0.2127$, רווח סמך ברמה של 95% עבור θ נתון לפי $\log(\hat{\alpha}) \pm z_{1-\frac{\alpha}{2}}\sqrt{\sum_{i,j}\frac{1}{f_{ij}}}$ ונקבל $\log(0.2127) \pm 1.96 \cdot \sqrt{\frac{1}{2343} + \cdots} = -1.5477 \pm 0.259$ כלומר $\theta \in [-1.8069, -1.2888]$ וכן $\alpha \in [0.1642, 0.2756]$.

הגדרת המודל הלוג-ליניארי

מה קורה כאשר יש לנו יותר מאשר שני גורמים, או כשלגורם יש מספר ערכים אפשריים?

נניח כי נתונים לנו משתנים איכותיים באופן $A \in \{1, \dots, I\},\ B \in \{1, \dots, J\},\ C \in \{1, \dots, K\}$ אז נוכל לכתוב $\pi_{ijk} = P(A = i, B = j, C = k)$ וכן $\pi_{i.k} = P(A = i, C = k)$; ... ; $\pi_{i..} = P(A = i)$ כאשר $\pi_{ijk} > 0$ וכן $\sum_{i,j,k}\pi_{ijk} = 1$.

יש לנו $IJK - 1$ פרמטרים חופשיים, נחפש פרמטריזציה של π_{ijk} (השערות מעניינות יתקבלו כאשר חלק מהפרמטרים מתאפסים): מודל $\pi_{ijk} = \mu + \alpha_i + \cdots + \gamma_{ijk}$ (בדומה לANOVA) לא יתאים לנו כי הוא מבריח אילוצים על הפרמטרים, לכן נגדיר את המודל הלוג-ליניארי:

$$\theta_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

כאשר $\theta_{ijk} = \log(\pi_{ijk})$, $-\infty < \theta_{ijk} < 0$, $\lambda \in \mathbb{R}$ ומתקיים $\lambda_{\cdot}^{A} = \cdots = \lambda_{.k}^{AC} = \cdots = \lambda_{\dots}^{ABC} = 0$.

$$\pi_{ijk} = e^{\theta_{ijk}} \to \forall i,j,k: 0 < \pi_{ijk} < 1$$

$$1 = \sum_{i,j,k}\pi_{ijk} = e^{\mu}\sum_{i,j,k} e^{\lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}}$$

הפירוש יגיע מתוך סדרה מקוננת של מודלים היררכיים:

המודל (A, B, C) מניח כי $\theta_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C$, כלומר $H_0: \lambda_{ij}^{AB} = \lambda_{ik}^{AC} = \lambda_{jk}^{BC} = \lambda_{ijk}^{ABC} = 0$; המודל (AB, C) מניח $\theta_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB}$ (AB אינטראקציה ולכן גם הגורמים עצמם נכללים), $H_0: \lambda_{ik}^{AC} = \lambda_{jk}^{BC} = \lambda_{ijk}^{ABC} = 0$; (AB, AC) מניח $\theta_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC}$, כלומר $H_0: \lambda_{jk}^{BC} = \lambda_{ijk}^{ABC} = 0$; (AB, AC, BC) מניח $\theta_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$, כלומר $H_0: \lambda_{ijk}^{ABC} = 0$; במודל המלא (ABC) $\theta_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$.

לדוגמא, מחקר עמדות בנושא הפלות שנערך בארה"ב שנות ה-70 – יש לנו את A כמשתנה קטגורי המציין שנה, B משתנה קטגורי המציין שייכות לזרם נוצרי, C משתנה אורדינלי (ordinal, הגודל תואם את הסדר בין הערכים) המציין שנות לימוד ו-D משתנה קטגורי של גישה להפלות:

שנה	זרם דתי	מספר שנות לימוד	גישה			
			חיובית	מעורבת	שלילית	סה"כ
1972	פרוטסטנטי צפוני	≤ 8	9	16	41	66
		$9-12$	85	52	105	242
		≥ 13	77	30	38	145
	פרוטסטנטי דרומי	≤ 8	8	8	46	62
		$9-12$	35	29	54	118
		≥ 13	37	15	22	74
	קתולי	≤ 8	11	14	38	63
		$9-12$	47	35	115	197
		≥ 13	25	21	42	88
1973	פרוטסטנטי צפוני	≤ 8	17	17	42	76
		$9-12$	102	38	84	224
		≥ 13	88	15	31	134
	פרוטסטנטי דרומי	≤ 8	14	11	34	59
		$9-12$	61	30	59	150
		≥ 13	49	11	19	79
	קתולי	≤ 8	6	16	26	48
		$9-12$	60	29	108	197
		≥ 13	31	18	50	99
1974	פרוטסטנטי צפוני	≤ 8	23	13	32	68
		$9-12$	106	50	88	244
		≥ 13	79	21	31	131
	פרוטסטנטי דרומי	≤ 8	5	15	37	57
		$9-12$	38	39	54	131
		≥ 13	52	12	32	96
	קתולי	≤ 8	8	10	24	42
		$9-12$	65	39	89	193
		≥ 13	37	18	43	98

טענות (משאיר את ההוכחה לפרופ' צוקר):

$(A,B,C) \Leftrightarrow$ המשתנים A,B,C ב"ת בזוגות.

$(AB,C) \Leftrightarrow$ המשתנים A,B ב"ת במשתנה C.

$(AB,AC) \Leftrightarrow$ המשתנים B,C ב"ת באופן מותנה במשתנה A.

$(AB,AC,BC) \Leftrightarrow$ לכל i,j,i',j' מנת יחס הסיכויים $\frac{P(i|j,k)/P(i'|j,k)}{P(i|j',k)/P(i'|j',k)}$ אינה תלויה ב-k.

אמידת פרמטרים

(מכאן והלאה שוב צוקר)

נסמן F_{ijk} את מספר הפריטים במדגם עם $A=i,B=j,C=k$. $\sum_{i,j,k} F_{ijk}=n$.

$$F=\left(F_{111},F_{112},F_{113},\dots,F_{11K},F_{121},F_{122},\dots,F_{12K},\dots,F_{211},\dots,F_{IJK}\right)\sim Mult\left(n,\pi_{111},\dots,\pi_{ijk}\right)$$

$$P(F=f)=\frac{n!}{f_{111}!\cdot\ldots\cdot f_{IJK}!}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K}\pi_{ijk}^{f_{ijk}}$$

וקטור הפרמטרים:

$$\omega=\begin{pmatrix}\lambda_1^A,\dots,\lambda_{I-1}^A,\lambda_1^B,\dots,\lambda_{J-1}^B,\lambda_1^C,\dots,\lambda_{K-1}^C,\lambda_1^{AB},\dots,\lambda_{(I-1)(J-1)}^{AB},\\ \lambda_1^{AC},\dots,\lambda_{(I-1)(K-1)}^{AC},\lambda_1^{BC},\dots,\lambda_{(J-1)(K-1)}^{BC},\lambda_1^{ABC},\dots,\lambda_{(I-1)(J-1)(K-1)}^{ABC}\end{pmatrix}$$

נגדיר $\xi_{ijk}=\theta_{ijk}-\bar{\theta}_{\dots}=\theta_{ijk}-\mu\rightarrow\log\left(\pi_{ijk}\right)=\theta_{ijk}=\bar{\theta}_{\dots}+\xi_{ijk}$ ולכן נוכל לרשום את

$$\pi_{ijk}=e^{\bar{\theta}_{\dots}+\xi_{ijk}}=e^{\bar{\theta}_{\dots}}e^{\xi_{ijk}}\rightarrow\underbrace{\sum_i\sum_j\sum_k\pi_{ijk}}_{=1}=e^{\bar{\theta}_{\dots}}\sum_i\sum_j\sum_k e^{\xi_{ijk}}\rightarrow e^{\bar{\theta}_{\dots}}$$

$$=\left(\sum_i\sum_j\sum_k e^{\xi_{ijk}}\right)^{-1}\rightarrow\pi_{ijk}=\frac{e^{\xi_{ijk}}}{\sum_{i'}\sum_{j'}\sum_{k'}e^{\xi_{i'j'k'}}}=\pi_{ijk}(\omega)$$

פונקצית הנראות:

$$L(\omega)=\frac{n!}{f_{111}!\cdot\ldots\cdot f_{IJK}!}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K}\left(\pi_{ijk}(\omega)\right)^{F_{ijk}}=\frac{n!}{f_{111}!\cdot\ldots\cdot f_{IJK}!}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K}\left(e^{\theta_{ijk}(\omega)}\right)^{F_{ijk}}$$

$$=\frac{n!}{f_{111}!\cdot\ldots\cdot f_{IJK}!}\prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K}e^{\theta_{ijk}(\omega)F_{ijk}}$$

$$l(\omega)=\log L(\omega)=\sum_i\sum_j\sum_k\theta_{ijk}(\omega)F_{ijk}+const.$$

בתת-מודל, וקטור הפרמטרים מסומן τ (ללא פרמטרי λ המיותרים), כך למשל עבור (A,B,C) נקבל $\tau=\left(\lambda_1^A,\dots,\lambda_{I-1}^A,\lambda_1^B,\dots,\lambda_{J-1}^B,\lambda_1^C,\dots,\lambda_{K-1}^C\right)$, עבור (AB,C) נקבל $\tau=\left(\lambda_1^A,\dots,\lambda_{I-1}^A,\lambda_1^B,\dots,\lambda_{J-1}^B,\lambda_1^C,\dots,\lambda_{K-1}^C,\lambda_1^{AB},\dots,\lambda_{(I-1)(J-1)}^{AB}\right)$. נרצה לעבור למספור נח יותר של התאים, בצורה $m=1,2,3,\dots,K,K+1,\dots,IJK$ ואז לוג הנראות $l(\tau)=\sum_{m=1}^{IJK}\theta_m(\tau)F_m+const.$ וכן $\pi_{ijk}(\omega)=\pi_m(\omega)=\frac{e^{\xi_m}}{\sum_{m'=1}^{IJK}e^{\xi_{m'}}}\rightarrow\theta_m=\log(\pi_m)=\xi_m-\log\left(\sum_{m'=1}^{IJK}e^{\xi_{m'}}\right)$, והנגזרת תהיה:

$$\frac{\partial l}{\partial\tau_r}=\sum_{m=1}^{IJK}F_m\frac{\partial\theta_m(\tau)}{\partial\tau_r},\frac{\partial\theta_m(\tau)}{\partial\tau_r}=\frac{\partial}{\partial\tau_r}\left(\xi_m-\log\left(\sum_{m'=1}^{IJK}e^{\xi_{m'}}\right)\right)=\frac{\partial\xi_m}{\partial\tau_r}-\frac{\sum_{m'=1}^{IJK}\frac{\partial\xi_{m'}}{\partial\tau_r}e^{\xi_{m'}}}{\sum_{m'=1}^{IJK}e^{\xi_{m'}}}$$

נסמן $Z_{mr}=\frac{\partial\xi_m}{\partial\tau_r}$ אז $\frac{\partial\theta_m(\tau)}{\partial\tau_r}=Z_{mr}-\frac{\sum_{m'=1}^{IJK}Z_{m'r}e^{\xi_{m'}}}{\sum_{m'=1}^{IJK}e^{\xi_{m'}}}$ ונקבל:

$$\frac{\partial l}{\partial \tau_r} = \sum_{m=1}^{IJK} F_m \frac{\partial \theta_m(\tau)}{\partial \tau_r} = \sum_{m=1}^{IJK} F_m \left(Z_{mr} - \frac{\sum_{m'=1}^{IJK} Z_{m'r} e^{\xi_{m'}}}{\sum_{m'=1}^{IJK} e^{\xi_{m'}}} \right)$$
$$= \sum_{m=1}^{IJK} F_m Z_{mr} - \sum_{m=1}^{IJK} F_m \frac{\sum_{m'=1}^{IJK} Z_{m'r} e^{\xi_{m'}}}{\sum_{m'=1}^{IJK} e^{\xi_{m'}}}$$
$$= \sum_{m=1}^{IJK} F_m Z_{mr} - \frac{\sum_{m'=1}^{IJK} Z_{m'r} e^{\xi_{m'}}}{\sum_{m'=1}^{IJK} e^{\xi_{m'}}} \underbrace{\sum_{m=1}^{IJK} F_m}_{=n} = \sum_{m=1}^{IJK} F_m Z_{mr} - n \cdot \frac{\sum_{m'=1}^{IJK} Z_{m'r} e^{\xi_{m'}}}{\sum_{m'=1}^{IJK} e^{\xi_{m'}}}$$
$$= \sum_{m=1}^{IJK} F_m Z_{mr} - n \cdot \frac{\sum_{m=1}^{IJK} Z_{mr} e^{\xi_m}}{\sum_{m'=1}^{IJK} e^{\xi_{m'}}} = \sum_{m=1}^{IJK} F_m Z_{mr} - n \cdot \sum_{m=1}^{IJK} Z_{mr} \underbrace{\frac{e^{\xi_m}}{\sum_{m'=1}^{IJK} e^{\xi_{m'}}}}_{=\pi_m}$$
$$= \sum_{m=1}^{IJK} Z_{mr}\left(F_m - n\pi_m(\tau)\right)$$

רשמנו $Z_{mr} = \frac{\partial \xi_m}{\partial \tau_r}$ כאשר ξ_m הוא צ"ל של λ שונים, למשל $\xi_1 = \lambda_1^A + \lambda_1^B + \lambda_1^C + \lambda_{11}^{AB} + \lambda_{11}^{AC} + \lambda_{11}^{BC} +$ λ_{111}^{ABC}, סה"כ נוכל לרשום במודל המלא $\xi = X\omega$ ובמודל חלקי $\xi = Z\tau$ כאשר Z היא המטריצה X ללא העמודות המיותרות, כמו כן מתקיים $Z_{mr} = \frac{\partial}{\partial \tau_r} \sum_{s=1}^{p} Z_{ms}\tau_s = \frac{\partial}{\partial \tau_r}(Z\tau)_m = \frac{\partial \xi_m}{\partial \tau_r}$. כדי למצוא אומדי ML נרצה לפתור $\frac{\partial l}{\partial \tau_r} = 0$ עבור $r = 1, \dots, p$:

$$\frac{\partial^2 l}{\partial \tau_r \partial \tau_s} = \frac{\partial}{\partial \tau_s}\frac{\partial l}{\partial \tau_r} = -n\left(\frac{\sum_{m'} Z_{m'r} Z_{m's} e^{\xi_{m'}}}{\sum_{m'} e^{\xi_{m'}}} - \frac{\sum_{m'} Z_{m'r} e^{\xi_{m'}}}{\sum_{m'} e^{\xi_{m'}}} \frac{\sum_{m'} Z_{m's} e^{\xi_{m'}}}{\sum_{m'} e^{\xi_{m'}}} \right)$$

נסמן $\nabla l = Z^T(F - n\pi), \nabla^2 l = -nZ^TWZ$ כאשר $W = \begin{bmatrix} \pi_1 & & 0 \\ & \ddots & \\ 0 & & \pi_{IJK} \end{bmatrix} - \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_{IJK} \end{bmatrix} [\pi_1 \quad \cdots \quad \pi_{IJK}]$

ואז ניוטון-רפסון:

$$\tau^{(u+1)} = \tau^{(u)} + (nZ^TWZ)^{-1} Z^T \left(F - n\pi\left(\tau^{(u)}\right) \right) = \tau^{(u)} + (nZ^TWZ)^{-1} Z^T \left(\underbrace{p}_{=\frac{1}{n}F} - \pi\left(\tau^{(u)}\right) \right)$$
$$= (Z^TWZ)^{-1} \left((nZ^TWZ)\tau^{(u)} + Z^T \left(p - \pi\left(\tau^{(u)}\right) \right) \right)$$
$$= (Z^TWZ)^{-1} Z^T W \left(Z\tau^{(u)} + W^{-1} \left(p - \pi\left(\tau^{(u)}\right) \right) \right)$$
$$= (Z^TWZ)^{-1} Z^T Y\left(\tau^{(u)}\right), Y\left(\tau^{(u)}\right) = W\left(\tau^{(u)}\right) Z\tau^{(u)} + \left(p - \pi\left(\tau^{(u)}\right) \right)$$

ובסך הכל נקבל $\hat{\tau} \sim N\left(\tau, \frac{1}{n}(Z^TWZ)^{-1}\right)$.

קשר לרגרסיה פואסונית

נניח כי Q_m מ"מ ב"ת עם התפלגות פואסונית $Q_m \sim Pois(\psi_m)$ וכן $\sum_m \theta_m = 1$, $\psi_m = e^{\alpha+\theta_m}$, פונקצית הנראות תהיה:

$$L(\tau) = \prod_{m=1}^{M} \frac{\psi_m^{Q_m} e^{-\psi_m}}{Q_m!} = \prod_{m=1}^{M} \frac{\left(e^{\alpha+\theta_m(\tau)}\right)^{Q_m} e^{-e^{\alpha+\theta_m(\tau)}}}{Q_m!} \to l(\tau) = \log L(\tau)$$
$$= -\sum_m \log(Q_m!) + \sum_m Q_m\left(\alpha + \theta_m(\tau)\right) - \sum_m e^{\alpha+\theta_m(\tau)}$$
$$= const. + N\alpha - e^{\alpha} \underbrace{\sum_m e^{\theta_m(\tau)}}_{=1} + \sum_m Q_m \theta_m(\tau)$$
$$= N\alpha - e^{\alpha} + \sum_m Q_m \theta_m(\tau) + const.$$

$$\bar{\theta} = \frac{1}{M}\sum_m \theta_m \to \theta_m = \bar{\theta} + \xi_m \to e^{\theta_m} = \frac{e^{\xi_m}}{\sum_{m'} e^{\xi_{m'}}} \to \theta_m = \xi_m - \log\left(\sum_{m'} e^{\xi_{m'}}\right), \sum_m e^{\theta_m} = 1$$
$$\to l(\tau) = N\alpha - e^{\alpha} + \sum_m Q_m \theta_m(\tau) + const.$$
$$= N\alpha - e^{\alpha} + \sum_m Q_m \xi_m + \sum_m Q_m \log\left(\sum_{m'} e^{\xi_{m'}}\right) + const.$$
$$= (N\alpha - e^{\alpha}) + \left(\sum_m Q_m \xi_m - N\log\left(\sum_{m'} e^{\xi_{m'}}\right)\right) + const.$$

$$\frac{\partial l}{\partial \tau_r} = \sum_m Q_m \frac{\partial \xi_m}{\partial \tau_r} - N\frac{\sum_{m'} \frac{\partial \xi_{m'}}{\partial \tau_r} e^{\xi_{m'}}}{\sum_{m'} e^{\xi_{m'}}}, \qquad \frac{\partial l}{\partial \alpha} = N - e^{\alpha}, \qquad \frac{\partial^2 l}{\partial \tau_r \partial \alpha} = 0$$

אם נסמן $\gamma = \begin{bmatrix} \alpha \\ \tau_1 \\ \vdots \\ \tau_p \end{bmatrix}$ אז נקבל $\nabla^2 l = -\begin{bmatrix} e^{\alpha} & 0 \\ 0 & \boxed{n(Z^TWZ)} \end{bmatrix}$ וכן $(\nabla^2 l)^{-1} = -\begin{bmatrix} e^{-\alpha} & 0 \\ 0 & \boxed{\frac{1}{n}(Z^TWZ)^{-1}} \end{bmatrix}$.

נחזור למודל הלוג-לינארי:

$$\frac{\partial l}{\partial \tau_r} = \sum_{m=1}^{IJK} Z_{mr}\left(F_m - n\pi_m(\tau)\right) = \sum_{i,j,k} Z_{[ijk],r}\left(F_{ijk} - n\pi_{ijk}(\tau)\right)$$

רשמנו $\sum_{i,j,k} Z_{[ijk],r}\left(F_{ijk} - n\pi_{ijk}(\tau)\right) = \frac{\partial \xi_{ijk}}{\partial \tau_r}$ כאשר $\xi_{ijk} = \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$, מתקיים $\sum_i \lambda_i^A = 0 \to \lambda_I^A = -\sum_{i=1}^{I-1} \lambda_i^A$. נסמן:

$$\phi_{ia}^A = \begin{cases} 1 & i = a \\ -1 & i = I \\ 0 & ow \end{cases}, \qquad \phi_{jb}^B = \begin{cases} 1 & j = b \\ -1 & j = J \\ 0 & ow \end{cases}, \qquad \phi_{kc}^C = \begin{cases} 1 & k = c \\ -1 & k = K \\ 0 & ow \end{cases}$$

ונוכל לרשום:

$$\xi_{ijk} = \sum_{a=1}^{I-1} \phi_{ia}^A \lambda_a^A + \sum_{b=1}^{J-1} \phi_{jb}^B \lambda_b^B + \sum_{c=1}^{K-1} \phi_{kc}^C \lambda_c^C + \sum_{a=1}^{I-1}\sum_{b=1}^{J-1} \phi_{ia}^A \phi_{jb}^B \lambda_{ab}^{AB} + \sum_{a=1}^{I-1}\sum_{c=1}^{K-1} \phi_{ia}^A \phi_{kc}^C \lambda_{ac}^{AC}$$
$$+ \sum_{b=1}^{J-1}\sum_{c=1}^{K-1} \phi_{jb}^B \phi_{kc}^C \lambda_{bc}^{BC} + \sum_{a=1}^{I-1}\sum_{b=1}^{J-1}\sum_{c=1}^{K-1} \phi_{ia}^A \phi_{jb}^B \phi_{kc}^C \lambda_{abc}^{ABC}$$

$$\frac{\partial \xi_{ijk}}{\partial \lambda_a^A} = \phi_{im}^A \rightarrow \frac{\partial l}{\partial \lambda_a^A} = \sum_{i,j,k} \phi_{ia}^A \left(F_{ijk} - n\pi_{ijk}(\omega)\right)$$
$$= \sum_{j,k} 1 \cdot \left(F_{ajk} - n\pi_{ajk}(\omega)\right) + \sum_{j,k} -1 \cdot \left(F_{Ijk} - n\pi_{Ijk}(\omega)\right)$$
$$= (F_{a..} - n\pi_{a..}) - (F_{I..} - n\pi_{I..})$$

בדי למצוא אומדי ML משווים את הנגזרות לאפס:

$$\forall r = 1, \dots, (I-1): \left(F_{r..} - n\pi_{r..}(\omega)\right) - \left(F_{I..} - n\pi_{I..}(\omega)\right) = 0$$

$$\frac{\partial l}{\partial \lambda_a^A} = (F_{a..} - n\pi_{a..}) - (F_{I..} - n\pi_{I..}) = 0 \rightarrow \begin{cases} \sum_{i=1}^{I-1} (F_{i..} - n\pi_{i..}) - (I-1)(F_{I..} - n\pi_{I..}) = 0 \\ \sum_{i=1}^{I-1} (F_{i..} - n\pi_{i..}) + (F_{I..} - n\pi_{I..}) = 0 \end{cases}$$

אנחנו יודעים גם $\sum_{a,b,c} \pi_{abc} = 1$, $\sum_{a,b,c} F_{abc} = n$, $F_{...} = n\pi_{...}$:

$$\rightarrow I(F_{I..} - n\pi_{I..}) = 0 \rightarrow \hat{\pi}_{I..} = \frac{1}{n} F_{I..} \rightarrow F_{i..} - n\pi_{i..} = 0 \rightarrow \hat{\pi}_{i..} = \frac{1}{n} F_{i..}$$

עבור $(J-1)$, $\dots$, $s = 1$, $(I-1)$, $\dots$, $r = 1$, λ_{rs}^{AB} נקבל באופן דומה $\hat{\pi}_{rs.} = p_{rs.}$. בסך הכל אם האברים מסוג λ^A נמצאים במודל אז $\hat{\pi}_{r..} = p_{r..} = \frac{F_{r..}}{n}$ כאשר $p_{ijk} = \frac{F_{ijk}}{n}$, בדומה אם λ^{AB} במודל אז $\hat{\pi}_{rs.} = p_{rs.}$ וכן הלאה:

$$(1)\ \exists \lambda^A \rightarrow \forall r: \hat{\pi}_{r..} = p_{r..}$$
$$(2)\ \exists \lambda^B \rightarrow \forall s: \hat{\pi}_{.s.} = p_{.s.}$$
$$(3)\ \exists \lambda^C \rightarrow \forall t: \hat{\pi}_{..t} = p_{..t}$$
$$(4)\ \exists \lambda^{AB} \rightarrow \forall r,s: \hat{\pi}_{rs.} = p_{rs.}$$
$$(5)\ \exists \lambda^{AC} \rightarrow \forall r,t: \hat{\pi}_{r.t} = p_{r.t}$$
$$(6)\ \exists \lambda^{BC} \rightarrow \forall s,t: \hat{\pi}_{.st} = p_{.st}$$
$$(7)\ \exists \lambda^{ABC} \rightarrow \forall r,s,t: \hat{\pi}_{rst} = p_{rst}$$

עבור המודל (A,B,C) מתקיימות המשוואות 3-1, נזכור כי במודל זה A,B,C ב"ת, כלומר מתקיים בו $\underbrace{P(A=i, B=j, C=k)}_{\pi_{ijk}} = \underbrace{P(A=i)}_{\pi_{i..}} \underbrace{P(B=j)}_{\pi_{.j.}} \underbrace{P(C=k)}_{\pi_{..k}}$ כלומר $\hat{\pi}_{ijk} = \hat{\pi}_{i..}\hat{\pi}_{.j.}\hat{\pi}_{..k}$ ומהמשוואות נקבל $\hat{\pi}_{ijk} = p_{i..}p_{.j.}p_{..k}$ שהם ביטויים אותם אנחנו יכולים לחשב ודרך כך להגיע לביטויים של θ, λ.

במודל (AB,C) יש לנו את הפרמטרים $\lambda^A,\lambda^B,\lambda^C,\lambda^{AB}$, ממשוואה 4 מקבלים $\forall r,s:\hat{\pi}_{rs.}=p_{rs.}$ וממשוואה 3 $\forall t:\hat{\pi}_{..t}=p_{..t}$. במודל הזה AB,C ב"ת ולכן $\underbrace{P(A=i,B=j,C=k)}_{\pi_{ijk}}=\underbrace{P(A=i,B=j)}_{\pi_{ij.}}\underbrace{P(C=k)}_{\pi_{..k}}$

ומכאן $\hat{\pi}_{ijk}=\hat{\pi}_{ij.}\hat{\pi}_{..k}=p_{ij.}p_{..k}$ ושוב יש פתרון מלא מפורש.

במודל (AB,AC) יש לנו את הפרמטרים $\lambda^A,\lambda^B,\lambda^C,\lambda^{AB},\lambda^{AC}$, ממשוואה 5 מקבלים $\forall r,t:\hat{\pi}_{r.t}=p_{r.t}$ ומאי-תלות של B,C בהנתן A מקבלים $\underbrace{P(B=j,C=k|A=i)}_{\frac{\pi_{ijk}}{\pi_{i..}}}=\underbrace{P(B=j|A=i)}_{\frac{\pi_{ij.}}{\pi_{i..}}}\underbrace{P(C=k|A=i)}_{\frac{\pi_{i.k}}{\pi_{i..}}}$

כלומר $\hat{\pi}_{ijk}=\frac{\hat{\pi}_{ij.}\hat{\pi}_{i.k}}{\hat{\pi}_{i..}}=\frac{p_{ij.}p_{i.k}}{p_{i..}}$. עבור המודל (AB,AC,BC) אין פתרון סגור וצריך להשתמש בניוטון-רפסון לפתרון.

השוואת מודלים

נרצה להשוות בין תת-מודל מסוים M' ובין המודל המלא M – יש שני מבחנים אפשריים: יחס נראות ו-χ^2 של Pearson:

פונקציית לוג הנראות היא $l=const.+\sum_{i,j,k}F_{ijk}\log\pi_{ijk}$. נסמן ב-d את ההפרש בין מספר הפרמטרים במודל המלא $(\dim M=IJK-1)$ ומספר הפרמטרים בתת המודל M^0, לפי משפט Wilks מתקיים $G^2=2\left(l(M)-l(M^0)\right)\overset{H_0}{\sim}\chi^2_d$, כמו כן

$$G^2=2\left(\left(const.+\sum_{i,j,k}F_{ijk}\log\hat{\pi}^{M}_{ijk}\right)-\left(const.+\sum_{i,j,k}F_{ijk}\log\hat{\pi}^{M^0}_{ijk}\right)\right)=2\sum_{i,j,k}F_{ijk}\log\frac{\hat{\pi}^{M}_{ijk}}{\hat{\pi}^{M^0}_{ijk}}$$
$$=2\sum_{i,j,k}F_{ijk}\log\frac{p_{ijk}}{\hat{\pi}^{M^0}_{ijk}}=2\sum_{i,j,k}F_{ijk}\log\frac{n}{n}\cdot\frac{p_{ijk}}{\hat{\pi}^{M^0}_{ijk}}=2\sum_{i,j,k}F_{ijk}\log\frac{F_{ijk}}{\hat{F}^{(0)}_{ijk}}$$

מבחן Pearson יהיה בצורה

$$\mathrm{X}^2=\sum_{i,j,k}\frac{\left(F_{ijk}-\hat{F}^{(0)}_{ijk}\right)^2}{\hat{F}^{(0)}_{ijk}}\overset{H_0}{\sim}\chi^2_d$$

סטטיסטי יחס הנראות בין מודלים $M_1\subset M_2$ מסומן $G^2(M_1|M_2)=2\sum_{i,j,k}F_{ijk}\log\frac{\hat{F}^{(2)}_{ijk}}{\hat{F}^{(1)}_{ijk}}$, כאשר יש לנו שלושה מודלים מקוננים בצורה $M_1\subset M_2\subset M_3$, יחסי הנראות עבורם מקיימים $G^2(M_1|M_3)=G^2(M_1|M_2)+G^2(M_2|M_3)$, כאשר $G^2(M_1|M_2)\overset{M_1}{\sim}\chi^2_{d_{21}},G^2(M_1|M_3)\overset{M_1}{\sim}\chi^2_{d_{31}}$ ודרגות החופש הן $d_{21}=\dim M_2-\dim M_1\,,d_{31}=\dim M_3-\dim M_1$. ייתכן כי נדחה את $G^2(M_1|M_2)$ אך לא נדחה את $G^2(M_2|M_3)$.

שאריות

שאריות Pearson: $r_{ijk}^P = \frac{F_{ijk}-\hat{F}_{ijk}}{\sqrt{\hat{F}_{ijk}}}$ כאשר $\hat{F}_{ijk} = n\hat{\pi}_{ijk}$, מתקיים $X^2 = \sum_{i,j,k}\left(r_{ijk}^P\right)^2$.

שאריות Deviance: $r_{ijk}^D = sign\left(F_{ijk}-\hat{F}_{ijk}\right)\cdot\sqrt{2\left(F_{ijk}\log\frac{F_{ijk}}{\hat{F}_{ijk}} - \left(F_{ijk}-\hat{F}_{ijk}\right)\right)}$, מתקיים $G^2 = \sum_{i,j,k}\left(r_{ijk}^D\right)^2$.

r_{ijk}^P, r_{ijk}^D הם שני מדדים למרחק בין המודל שבנינו לבין הנתונים, כמו במודלים קודמים גם כאן אם המודל תקין אז השאריות אמורות להתפלג נורמלית $N(0,1)$.

דוגמא - הערכה עצמית (Self-Esteem)

נתונים ממחקר על הערכה עצמית (Self-Esteem, בקצרה SE) של סטודנטים באוניברסיטה בארה"ב כפונקציה של ציון ממוצע (GPA), מין (Gender) וגזע (Race), כל אחד מהמשתנים הוא קטגוריאלי עם 2 רמות (גבוה\נמוך, זכר\נקבה, שחור\לבן):

		Black		White	
Gender	GPA	High SE	Low SE	High SE	Low SE
Male	High	15	9	17	10
	Low	26	17	22	26
Female	High	13	22	22	32
	Low	24	23	3	17

(מחקר של Demo and Parker, 1987). ב-R אפשר להשתמש בפקודה loglm שהיא ספציפית למודלים לוג-לינאריים (ומקבלת את נתוני הקלט שלה כוקטור רב ממדי) או ב-glm עם מודל פואסוני.

נסמן $A = Sex, B = GPA, C = Race, D = Esteem$. עם loglm מקבלים כי המודל הבלתי-תלוי (A,B,C,D) אינו מובהק וכך גם המודל עם גורמים מסדר שני (AB,AC,AD,BC,BD,CD). המודל עם גורמים מסדר שלישי (ABC,ABD,ACD,BCD) מובהק ומקבלים מהרצה של glm כי הגורמים ABD,ACD אינם מובהקים (כלומר נוכל לוותר על הגורמים $\lambda^{Sex,GPA,Esteem},\lambda^{Sex,Race,Esteem}$ במודל – זהו המודל (AD,ABC,BCD)). נסמן את תתי-המודלים באופן הבא: $M_3 = (ABC,ABD,ACD,BCD), M_2 = (AD,ABC,BCD), M_1 = (ABC,D)$. מתקיים $M_1 \subset M_2 \subset M_3$ אך $G^2(M_1|M_2)$ לא מובהק בעוד $G^2(M_2|M_3)$ כן מובהק.

לאור התוצאות שהתקבלו בוחרים במודל (AD,ABC,BCD), ממנו ניתן להסיק את המשמעויות הבאות:

- הקשר בין A ו-D לא תלוי ב-B או C.
- הקשר בין A ו-B תלוי ב-C.
- הקשר בין B ו-C תלוי ב-A.
- הקשר בין B ו-D תלוי ב-C אבל לא ב-A.

נרצה להעמיק בקשר האחרון:

$$\psi_{ik}^{BD}(j,j',l,l') = \frac{\frac{P(D=l'|B=j',A=i,C=k)}{P(D=l|B=j',A=i,C=k)}}{\frac{P(D=l'|B=j,A=i,C=k)}{P(D=l|B=j,A=i,C=k)}}$$

$$= \frac{\frac{\exp\left(\bar{\theta}_{\dots}+\lambda_i^A+\lambda_{j'}^B+\lambda_k^C+\lambda_{l'}^D+\lambda_{ij'}^{AB}+\lambda_{ik}^{AC}+\lambda_{il'}^{AD}+\lambda_{j'k}^{BC}+\lambda_{j'l'}^{BD}+\lambda_{kl'}^{CD}+\lambda_{ij'k}^{ABC}+\lambda_{j'kl'}^{BCD}\right)}{\exp\left(\bar{\theta}_{\dots}+\lambda_i^A+\lambda_{j'}^B+\lambda_k^C+\lambda_l^D+\lambda_{ij'}^{AB}+\lambda_{ik}^{AC}+\lambda_{il}^{AD}+\lambda_{j'k}^{BC}+\lambda_{j'l}^{BD}+\lambda_{kl}^{CD}+\lambda_{ij'k}^{ABC}+\lambda_{j'kl}^{BCD}\right)}}{\frac{\exp\left(\bar{\theta}_{\dots}+\lambda_i^A+\lambda_j^B+\lambda_k^C+\lambda_{l'}^D+\lambda_{ij}^{AB}+\lambda_{ik}^{AC}+\lambda_{il'}^{AD}+\lambda_{jk}^{BC}+\lambda_{jl'}^{BD}+\lambda_{kl'}^{CD}+\lambda_{ijk}^{ABC}+\lambda_{jkl'}^{BCD}\right)}{\exp\left(\bar{\theta}_{\dots}+\lambda_i^A+\lambda_j^B+\lambda_k^C+\lambda_l^D+\lambda_{ij}^{AB}+\lambda_{ik}^{AC}+\lambda_{il}^{AD}+\lambda_{jk}^{BC}+\lambda_{jl}^{BD}+\lambda_{kl}^{CD}+\lambda_{ijk}^{ABC}+\lambda_{jkl}^{BCD}\right)}}$$

$$= \frac{\frac{\exp\left(\lambda_{l'}^D+\lambda_{il'}^{AD}+\lambda_{j'l'}^{BD}+\lambda_{kl'}^{CD}+\lambda_{j'kl'}^{BCD}\right)}{\exp\left(\lambda_l^D+\lambda_{il}^{AD}+\lambda_{j'l}^{BD}+\lambda_{kl}^{CD}+\lambda_{j'kl}^{BCD}\right)}}{\frac{\exp\left(\lambda_{l'}^D+\lambda_{il'}^{AD}+\lambda_{jl'}^{BD}+\lambda_{kl'}^{CD}+\lambda_{jkl'}^{BCD}\right)}{\exp\left(\lambda_l^D+\lambda_{il}^{AD}+\lambda_{jl}^{BD}+\lambda_{kl}^{CD}+\lambda_{jkl}^{BCD}\right)}}$$

$$= \frac{\exp\left(\left(\lambda_{l'}^D-\lambda_l^D\right)+\left(\lambda_{il'}^{AD}-\lambda_{il}^{AD}\right)+\left(\lambda_{j'l'}^{BD}-\lambda_{j'l}^{BD}\right)+\left(\lambda_{kl'}^{CD}-\lambda_{kl}^{CD}\right)+\left(\lambda_{j'kl'}^{BCD}-\lambda_{j'kl}^{BCD}\right)\right)}{\exp\left(\left(\lambda_{l'}^D-\lambda_l^D\right)+\left(\lambda_{il'}^{AD}-\lambda_{il}^{AD}\right)+\left(\lambda_{jl'}^{BD}-\lambda_{jl}^{BD}\right)+\left(\lambda_{kl'}^{CD}-\lambda_{kl}^{CD}\right)+\left(\lambda_{jkl'}^{BCD}-\lambda_{jkl}^{BCD}\right)\right)}$$

$$= \frac{\exp\left(\left(\lambda_{j'l'}^{BD}-\lambda_{j'l}^{BD}\right)+\left(\lambda_{j'kl'}^{BCD}-\lambda_{j'kl}^{BCD}\right)\right)}{\exp\left(\left(\lambda_{jl'}^{BD}-\lambda_{jl}^{BD}\right)+\left(\lambda_{jkl'}^{BCD}-\lambda_{jkl}^{BCD}\right)\right)}$$

$$= \exp\left(\left(\lambda_{j'l'}^{BD}-\lambda_{j'l}^{BD}-\lambda_{jl'}^{BD}+\lambda_{jl}^{BD}\right)+\left(\lambda_{j'kl'}^{BCD}-\lambda_{j'kl}^{BCD}-\lambda_{jkl'}^{BCD}+\lambda_{jkl}^{BCD}\right)\right)$$

קיבלנו שאין אינדקסים של i אבל יש של k לכן כפי שטענו קודם המודל תלוי ב-C אבל לא תלוי ב-A.
ניתן לחשב גם את הביטוי של λ_{ru}^{BD} עבור כל רמה של C, מתקבל:

$$\lambda_{ru}^{BD} = \begin{cases} \exp\left(4\lambda_{11}^{BD}+4\lambda_{111}^{BCD}\right) & k=1 \\ \exp\left(4\lambda_{11}^{BD}-4\lambda_{111}^{BCD}\right) & k=2 \end{cases}$$

יש לנו את וקטור הפרמטרים $\tau = [\lambda_1^A \quad \dots \quad \lambda_{11}^{AB} \quad \dots \quad \lambda_{111}^{ABC} \quad \dots]^T$, נרצה לבנות וקטור q שיקיים $4\lambda_{11}^{BD}+4\lambda_{111}^{BCD} = q_1\tau$, כלומר $q_1 = [0 \quad \dots \quad 0 \quad 4 \quad 0 \quad \dots \quad 0 \quad 4]^T$ ובאותו אופן עבור $4\lambda_{11}^{BD} - 4\lambda_{111}^{BCD} = q_2\tau$, $q_2 = [0 \quad \dots \quad 0 \quad 4 \quad 0 \quad \dots \quad 0 \quad -4]^T$. כעת:

$$\widehat{q_1^T\tau} = q_1^T\hat{\tau}, \qquad \widehat{Var}(q_1^T\hat{\tau}) = q_1^T\widehat{Var}(\hat{\tau})q_1$$

$$\widehat{Var}(\hat{\tau}) = \frac{1}{n}(Z^TWZ)^{-1} \rightarrow \widehat{Var}(q_1^T\hat{\tau}) = \frac{1}{n}q_1^T(Z^TWZ)^{-1}q_1$$

מכאן כי ר"ס עבור $q_1^T\tau$ נתון לפי $q_1^T\hat{\tau} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{n}q_1^T(Z^TWZ)^{-1}q_1}$.