Modelos Log–Lineales

Hasta los años 60 las tablas de contingencia de 2×2 eran analizadas calculando estadísticos tipo χ^2 para testear independencia. Cuando las tablas involucraban más variables se solía repetir este análisis para las subtablas para determinar las interacciones o asociaciones entre las variables. A partir de los 70 con los trabajos de Goodman y la difusión de estos en libros como el de Bishop, Finberg y Holland (1975) y Haberman (1975) hubo un cambio sustancial en el tratamiento de estos problemas, en particular con la inclusión de los modelos log–lineales.

Un modelo loglineal puede ser visto como un caso particular del GLM.

Los modelos log–lineales se usan con frecuencia para analizar la relación de tres o más variables categóricas en una tabla de contingencia.

El objetivo de estos modelos es estudiar asociación, es por ello que no se hace distinción entre variables de respuesta y covariables. Cuando interesa estudiar algunas variables como independientes y otras como dependientes es más adecuado un modelo lineal generalizado. Lo mismo es cierto si las variables estudiadas son continuas y no se pueden discretizar apropiadamente.

La estrategia básica consiste en ajustar un modelo a las frecuencias observadas en la tabla cruzada. Los modelos contienen componentes que reflejan las distintas asociaciones entre las variables. Los modelos son representados por las frecuencias esperadas. Los patrones de asociación entre las variables pueden describirse en términos de los odds y los odds ratios.

Por ahora supondremos que observamos $Y = (Y_{11}, \ldots., Y_{IJ})$ un vector multinomial con $n = y_{++}$ y $\mu = n\boldsymbol{\pi}$.

Modelos para tablas bidimensionales

Comenzamos por considerar el caso más sencillo de tablas de contingencia, es decir bidimensional. Luego, los conceptos que aquí veremos se extienden a tablas más complejas.

Podríamos tener un tabla de 2×2 como en el ejemplo que vimos en nuestras primeras clases:

	C: Cree en la vida postmortem		Total
S:Sexo	Si	No	
Mujer	435	147	582
Hombre	375	134	509
Total	810	281	1091

En general, tendremos:

A	**B** Si	No	Total
Si	n_{11}	n_{12}	n_{1+}
No	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	$n_{++} = n$

En general,podríamos tener una tabla bidimensional de $I \times J$.

Modelo de Independencia

Bajo el modelo de independencia tenemos que $\pi_{ij} = \pi_{i+}\pi_{+j} \quad \forall i\,\forall j$, por lo tanto si μ_{ij} denota la esperanza de la casilla (i, j), entonces

$$\begin{aligned} \log \mu_{ij} &= \log n + \log \pi_{i+} + \log \pi_{+j} \\ \log \mu_{ij} &= \lambda + \lambda_i^A + \lambda_j^B \end{aligned}$$

donde λ_i^A y λ_j^B representan el efecto de la fila i y de la columna j, respectivamente.

La interpretación de los parámetros es más sencilla para respuestas binarias. Por ejemplo, en una tabla de $I \times 2$, donde las columnas corresponden a la respuesta dicotómica Y, para cada fila i el logit para la probabilidad π_i de que $Y = 1$ es

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \log\left(\frac{\mu_{i1}}{\mu_{i2}}\right)$$

Bajo independencia

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \lambda_1^B - \lambda_2^B$$

$\Longrightarrow$ no depende de i, es decir no depende de la fila.

Esto corresponde al caso en que

$$\text{logit}(\pi_i) = \alpha$$

por lo tanto la chance de clasificar en una columna particular es constante a lo largo de las filas.

Identificabilidad y Restricciones sobre los parámetros

En una tabla de 2×2, por ejemplo, el modelo independiente especifica 5 parámetros, por lo tanto está sobreespecificado.

La siguiente tabla muestra tres conjuntos de parámetros diferentes para los datos de *creencia* que dan los mismos valores estimados para las frecuencias esperadas.

Como en el caso lineal, podemos imponer restricciones a los parámetros de manera de obtener unicidad, por ejemplo, pidiendo que para el primer nivel de cada factor el parámetro sea 0 o bien pidiendo que la suma de los parámetros dentro de un factor sea 0.

Veamos cómo se obtienen estas restricciones.

Frecuencias Observadas		Frecuencias Ajustadas		Log Frecuencias Observadas	
435	147	432.1	149.9	6.069	5.010
375	134	377.9	131.1	5.935	4.876
Parámetro			Conjunto 1	Conjunto 2	Conjunto 3
λ			4.876	6.6069	5.472
λ_1^A			0.134	0	0.067
λ_2^A			0	-0.134	-0.067
λ_1^B			1.059	0	0.529
λ_2^B			0	-1.059	-0.529

En nuestro ejemplo de 2×2 resultaría

$$\lambda_1^A + \lambda_2^A = 0 \quad \lambda_1^B + \lambda_2^B = 0$$

En general, trabajaremos con esta restricción. De todos modos, lo que todos cumplirán es que la diferencia entre dos efectos principales es la misma. En nuestro ejemplo, tenemos que $\lambda_1^B - \lambda_2^B = 1{,}059$ para los tres conjuntos de parámetros.

Modelo Saturado

Cuando las variables son dependientes satisfacen un modelo más complejo

$$\log \mu_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

donde los parámetros λ_{ij}^{AB} reflejan la asociación entre A y B. Este modelo describe perfectamente cualquier conjunto de frecuencias y es el modelo más general para una tabla de contingencia bivariada. El caso de independencia corresponde a $\lambda_{ij}^{AB} = 0$

Existe una relación directa entre los *log odds ratios* y los parámetros de asociación λ_{ij}^{AB}. Consideremos una tabla de 2×2 y las restricciones

$$\sum_{i=1}^{I} \lambda_{ij}^{AB} = 0 \; \forall j, \quad \sum_{j=1}^{J} \lambda_{ij}^{AB} = 0 \; \forall i \, .$$

$$\begin{aligned}
\log\theta &= \log\left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right) = \log\mu_{11} + \log\mu_{22} - \log\mu_{12} - \log\mu_{21} \\
&= (\lambda + \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB}) + (\lambda + \lambda_2^A + \lambda_2^B + \lambda_{22}^{AB}) \\
&\quad -(\lambda + \lambda_1^A + \lambda_2^B + \lambda_{12}^{AB}) + (\lambda + \lambda_2^A + \lambda_1^B + \lambda_{21}^{AB}) \\
&= \lambda_{11}^{AB} + \lambda_{22}^{AB} - \lambda_{12}^{AB} - \lambda_{21}^{AB} \\
&= 4\lambda_{11}^{AB}\,,
\end{aligned}$$

donde la última igualdad resulta de las condiciones impuestas.

Los λ_{ij}^{AB} determinan los log odds ratios. Cuando $\lambda_{ij}^{AB} = 0$ los odds ratios valen 1 y A e B son independientes.

En la tabla de *creencia* el odd ratio es

$$\theta = \frac{435 \times 134}{147 \times 375} = 1.057$$

y $\log\theta = 0.056$, por lo tanto

$$\lambda_{11}^{AB} + \lambda_{22}^{AB} - \lambda_{12}^{AB} - \lambda_{21}^{AB} = 0.056 \qquad (*)$$

Los parámetros de asociación se pueden ajustar de manera que el primero de cada fila y el primero de cada columna sea 0 o que la suma sobre cada fila y la suma sobre cada columna sea 0.

Cualquiera de estas combinaciones satisfará $(*)$.

El modelo saturado tiene $IJ = 1+(I-1)+(J-1)+(I-1)(J-1)$ parámetros no redundantes, es decir tiene tantos parámetros como observaciones, dando un ajuste perfecto.

En la práctica se trata de usar modelos no saturados en tanto su ajuste *suaviza* a los datos y dan origen a interpretaciones más simples.

Los modelos log–lineales que hemos visto son **modelos jerárquicos**. Decimos que un modelo es **jerárquico** cuando incluye todos los términos de orden menor que están presentes en un término de orden mayor. Así, si el modelo contiene λ_{ij}^{AB}, entonces también están presentes en el modelo λ_i^A y λ_j^B. Estos son los modelo más frecuentes.

Como en ANOVA cuando hay interacciones, debemos ser cuidadosos en interpretar los efectos principales cuando hay términos de orden mayor. En general, para cada variable la atención se restringe a los términos de orden mayor.

Modelos para tablas tridimensionales

Los diferentes modelos que veremos representan distintos patrones de independencia y asociación.

Supongamos que tenemos tres variables categóricas A, B y C que tienen valores posibles:

A: 1, 2, . . . , I

B: 1, 2, . . . , J

C: 1, 2, . . . , K

Para desplegar los casos observados deberemos combinar tablas bidimensionales, como la que sigue.

Asumiremos que en una tabla como ésta, un individuo puede clasificar con una probabilidad π_{ijk} en la casilla ijk. Si las n unidades experimentales son independientes, entonces el vector $(n_{111}, \ldots, n_{ijk}, \ldots, n_{IJK})$ tiene distribución

Víctima	Defendido	Pena de Muerte Si	Pena de Muerte No	Porcentaje de Si	
Blanca	Blanco	53	414	11.3	Tabla Parcial
	Negro	11	37	22.9	
Negra	Blanco	0	16	0.0	Tabla Parcial
	Negro	4	139	2.8	
Total	Blanco	53	430	11.0	Tabla Marginal
	Negro	15	176	7.9	

Cuadro 1: Pena de Muerte por color del defendido y de la víctima

multinomial de parámetros n y $\boldsymbol{\pi} = (\pi_{111}, \ldots, \pi_{ijk}, \ldots, \pi_{IJK})'$. La única restricción que se impone al vector de probabilidades $\boldsymbol{\pi}$ es que sume 1 y el EMV será el vector de componentes

$$\widehat{\pi}_{ijk} = \frac{n_{ijk}}{n}$$

Calculemos el odds ratio en las tablas parciales. Cuando la víctima es blanca tenemos

$$\theta_B = \frac{53 \times 37}{11 \times 414} = 0.4306$$

por otro lado si la víctima es negra

$$\theta_N = \frac{0 \times 139}{4 \times 16} = 0$$

Sin embargo, si consideramos la tabla colapsada obtenemos

$$\theta_{Total} = \frac{53 \times 176}{15 \times 430} = 1.44$$

es decir que la conclusión a partir de este valor sería la opuesta que a partir de los odds ratios de las parciales. Este cambio de dirección en la asociación de dos variables al considerar una tercera se conoce como **paradoja de Simpson**. Por esta razón debemos tener mucho cuidado antes de colapsar una tabla, tratando de entender cual es la asociación entre las variables en primera instancia y a partir de ella decidir si es razonable colapsar o no.

Otro Ejemplo

Supongamos que tenemos los siguientes datos de sobrevida de pacientes sometidos a cirugía en dos hospitales, A y B donde vivos significa que el paciente sobrevivió 6 semanas a la cirugía.

Hospital	Muertos	Vivos
A	63	2037
B	16	784

Cuadro 2: Sobrevida a una cirugía según hospital

A patir de esta tabla obtenemos que

$$\theta = \frac{63 \times 784}{16 \times 2037} = 1.515464$$

con lo que parecemás conveniente el hospital B. Sin embargo, si tenemos en cuenta una tercera variable **C: Estado inicial del paciente** la información sería

	Malas condiciones		Buenas Condiciones	
Hospital	Muertos	Vivos	Muertos	Vivos
A	6	594	57	1443
B	8	692	8	92

Cuadro 3: Sobrevida a una cirugía según hospital y estado inicial

Si analizamos la información teniendo en cuenta el estado del paciente al ser intervenido vemos que

$$\theta_{MC} = \frac{6 \times 692}{8 \times 594} = 0.8737374 \qquad \theta_{BC} = \frac{57 \times 92}{8 \times 1443} = \ 0.454262$$

Es decir, el hospital A es siempre preferible, pero es sensiblemente mejor en caso en que el paciente esté en malas condiciones iniciales. El hospital A tiene mayor porcentaje de muertos en general, pero menor porcentaje de muertos al considerar los grupos de buenas/malas condiciones. Estamos otra vez ante la paradoja de Simpson.

Modelos de dependencia

Mutua o Completa Independencia

El modelo más simple es aquel en que

$$P(A = i, B = j, C = k) = P(A = i).P(B = j).P(C = k) \qquad \forall i, j, k$$

de manera que si

$$\begin{aligned} \alpha_i &= P(A = i) \quad i = 1, \ldots, I \\ \beta_j &= P(B = j) \quad j = 1, \ldots, J \\ \delta_k &= P(C = k) \quad k = 1, \ldots, K \end{aligned}$$

$$\pi_{ijk} = \alpha_i \beta_j \delta_k$$

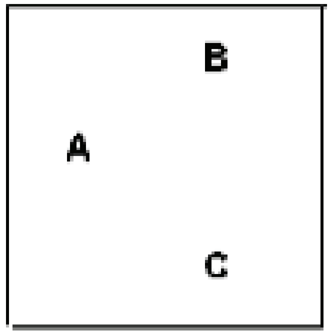
Como la suma de los α's, de los β's y de los δ's es 1 tenemos en total $(I-1)+(J-1)+(K-1)$ parámetros a estimar.

Además, bajo este modelo los vectores marginales tienen distribución:

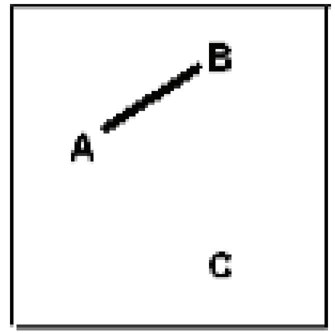
$$\begin{aligned}
(n_{1++},\dots,n_{I++}) &\sim M(n,\alpha_1,\alpha_2,\dots,\alpha_I) \\
(n_{+1+},\dots,n_{+J+}) &\sim M(n,\beta_1,\beta_2,\dots,\beta_J) \\
(n_{++1},\dots,n_{++K}) &\sim M(n,\delta_1,\delta_2,\dots,\delta_K)
\end{aligned}$$

por lo tanto, cada vector de parámetros podría estimarse por separado. Más aún, el EMV será

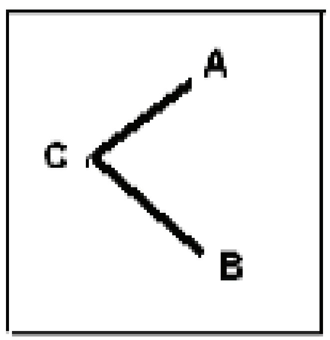
$$\widehat{\alpha}_i = \frac{n_{i++}}{n}$$
$$\widehat{\beta}_j = \frac{n_{+j+}}{n}$$
$$\widehat{\delta}_k = \frac{n_{++k}}{n}$$



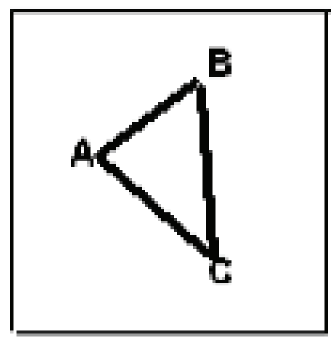
(1)
Mutua Independencia



(2)
Dos independientes de una tercera



(3)
Independencia Condicional



(4)
Asociacion Homogenea

Gráficamente, este modelo se representa como en el plot (1) del gráfico que se presenta más arriba.

En este gráfico no hay conexión entre los tres nodos, lo que indica que no hay relación entre las tres variables. En la notación de modelos log–lineales jerárquicos este modelo se representa como (A, B, C).

En términos de los odds ratios en este modelo significa las tablas marginales $A \times B$, $A \times C$ y $B \times C$ tienen odds ratios iguales a 1.

El modelo log–lineal correspondiente es

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C \quad (1)$$

Independencia conjunta

En el gráfico tenemos a A y B conectadas entre sí, pero no conectadas con C. Esto indica que C es conjuntamente independientes de A y B. Que los nodos A y B estén conectados indica que están posiblemente relacionados. De manera que el modelo de **mutua independencia** es un caso particular de este modelo que indicaremos como (AB, C).

Bajo este modelo tenemos que

$$\pi_{ijk} = \pi_{ij+}\,\pi_{++k} \quad \forall i, j, k$$

Si este modelo se cumple, A y C son independientes en la tabla marginal y B y C también son independientes en la tabla marginal.

También podemos escribirlo como

$$\pi_{ijk} = \theta_{ij}\,\delta_k \quad \forall i, j, k$$

donde

$$\sum_i \sum_j \theta_{ij} = 1 \qquad \sum_k \delta_k = 1$$

El número de parámetros es: $(IJ-1)+(K-1)$.

El EMV de estas probabilidades son: $\widehat{\theta}_{ij} = \dfrac{n_{ij+}}{n}$ y $\widehat{\delta}_k = \dfrac{n_{++k}}{n}$ y por lo tanto los valores esperados son

$$\widehat{\mu}_{ijk} = \frac{n_{ij+}n_{++k}}{n}$$

Esto corresponde al concepto de independencia habitual entre la variable C y una nueva variable formada por la IJ combinaciones de A y B.

El modelo log–lineal jerárquico correspondiente resulta

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} \qquad (2)$$

Independencia Condicional

Ahora consideremos la relación entre A y B controlando por C. Si A y B son independientes en la tabla *parcial* correspondiente al nivel k de C, decimos que

A y B son condicionalmente independientes en el nivel k de C.

Notemos $\pi_{ij|k} = \frac{\pi_{ijk}}{\pi_{++k}}$ a la distribución conjunta de A y B en el nivel k de C. Luego, la independencia condicional de A y B al nivel k de C equivale a

$$\pi_{ij|k} = \pi_{i+|k}\pi_{+j|k} \quad \forall i, j$$

Por lo tanto, diremos que A y B son condicionalmente independientes dado C si la condición anterior vale para todo k.

Entonces, tenemos que

$$\pi_{ijk} = \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}} \quad \forall i, j, k$$

Este modelo de dependencia correspone al gráfico en el que A y C están conectados y también lo están B y C.

El modelo de **mutua independencia** es un caso particular de este modelo.

Independencia condicional de A y B corresponde al modelo log–lineal:

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} \quad (3)$$

En la nomenclatura de los modelos log–lineales este modelo se llama (AC, BC).

Independencia marginal vs. Independencia Condicional

Consideremos el siguiente ejemplo en que se considera la distribución conjunta de las variables Sexo, Ingreso y Carrera

		Ingreso	
Carrera	Sexo	Bajo	Alto
Social	Mujer	18/100	12/100
	Hombre	12/100	8/100
Ciencias	Mujer	2/100	8/100
	Hombre	8/100	32/100
Total	Mujer	20/100	20/100
	Hombre	20/100	40/100

Tenemos que $\theta_{Social} = \frac{18\times 8}{12\times 12} = 1$ y $\theta_{Ciencias} = \frac{2\times 32}{8\times 8} = 1$, es decir que hay independencia en cada nivel de carrera, sin embargo en la tabla marginal $\theta = \frac{20\times 40}{20\times 20} = 2$ y por lo tanto no hay independencia marginal.

Por otro lado, en los odds de Ciencias son 6 veces más grandes en Hombres que en Mujeres dado Ingreso y los odds condicionales de Ingreso Alto son 6 veces más altos en Ciencias que en Sociales dado Sexo.

Ciencias tiene relativamente más hombres y Ciencias tiene relativamente ingresos más altos.

La *independencia condicional* y la *independencia marginal* se verifican simultáneamente cuando alguna independencia má s fuerte es válida.

Tenemos la siguiente relación:

Mutua Independencia entre A, B y C

$$\downarrow$$

B independiente conjuntamente de A y C

$$\swarrow \quad \searrow$$

A y B condicionalmente independientes $\qquad$ A y B marginalmente independientes

Cuando tenemos tres factores podemos tener tres, dos o un par de variables condicionalmente independientes de acuerdo a que tengamos el modelo (1), (2) o (3).

Asociación Homogénea

En efecto, los térmimos de la forma λ_{ts}^{XY} identifican a las variables condicionalmente dependientes.

Para permitir que las tres pares de variables sean condicionalmente dependientes debemos agregar al modelo anterior (AC, BC) una conexión entre A y B:

$$\log \mu_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{ij}^{AB} + \lambda_{jk}^{BC} \quad (4)$$

que corresponde al modelo (AB, AC, BC), conocido como modelo de **asociación homogénea**, pues en este modelo los odds ratios condicionales entre dos variables son idénticos para cada nivel de la tercera variable.

Dada la tabla parcial $A-B$ para cada nivel k de C podemos decribir la asociación parcial mediante los odds ratios condicionales como

$$\theta_{ij(k)} = \frac{\pi_{ijk}\pi_{i+1,j+1,k}}{\pi_{i,j+1,k}\pi_{i+1,j,k}} \quad 1 \leq i \leq I-1 \quad 1 \leq j \leq J-1$$

Probaremos que si el modelo (4) vale, entonces

$$\log \theta_{ij(k)} = \lambda^{AB}_{ij} + \lambda^{AB}_{i+1,j+1} - \lambda^{AB}_{i+1,j} - \lambda^{AB}_{i,j+1}$$

es decir que

$$\theta_{ij(1)} = \theta_{ij(2)} = \ldots = \theta_{ij(K)} \quad \forall i, j$$

Lo mismo es cierto para $\theta^{AC}_{ik(j)}$ y para $\theta^{BC}_{jk(i)}$.
Luego, la asociación entre dos variables es idéntica para cada nivel de la tercera variable.

Modelo saturado o Con Interacción Triple

El modelo más general para tres variables es

$$\log \mu_{ijk} = \lambda + \lambda^{A}_{i} + \lambda^{B}_{j} + \lambda^{C}_{k} + \lambda^{AC}_{ik} + +\lambda^{AB}_{ij} + \lambda^{BC}_{jk} + \lambda^{ABC}_{ijk} \quad (5)$$

En este caso las tres variables son condicionalmente dependientes, pero además los odds ratios de cualquier par de variables puede variar a lo largo de los niveles de la tercera. Identificamos este modelo como (ABC).

Condiciones para asociación marginal y parcial idénticas

El siguiente teorema establece condiciones suficientes para que los odds ratios entre A y B sean los mismos en la tabla marginal que en las tablas parciales.

Cuando esto es así podemos estudiar la asociación entre A y B de manera más sencilla colapsando la tabla sobre C.

Teorema: En una tabla tridimensional una variable es colapsable con respecto a la interacción entre las otras dos variables si es al menos condicionalmente independiente de otra dada la tercera.

Bishop, Fienberg y Holland (1975)

En otra palabras, A y B tienen la misma asociación marginal que parcial si A y C son condicionalmente independientes (vale el modelo (AB, BC)) o si B y C son condicionalmente independientes (vale el modelo (AB, AC))

Cuando colapsar en una tabla

Dicho de otra manera el resultado que hemos visto nos dice que si tenemos tres variables, A, B y C podemos colapsar en C si se cumplen las dos condiciones siguientes:

1. No hay interacción ABC, es decir $\lambda_{ijk}^{ABC} = 0$ para todo i, j, k.
2. La interacción AC o BC es nula, es decir $\lambda_{ik}^{AC} = 0$ para todo i, k o $\lambda_{jk}^{BC} = 0$ para todo j, k.

Sin embargo, del Teorema enunciado más arriba fue probado por Bishop, Fienberg y Holland (1975, pp 39, 47)) como un si y sólo si. Whittemore (1978) mostró contraejemplos de la necesidad. De hecho mostró los siguientes ejemplos:

TABLE 4

F_3	1		2		3		Sum over F_3	
F_1 \ F_2	1	2	1	2	1	2	1	2
1	4	2	2	1	1	4	7	7
2	2	1	4	2	1	4	7	7

$u^{(3)}_{123} = u^{(3)}_{12} = u^{(2)}_{12,3} = 0;\ u^{(3)}_{23} \neq 0,\ u^{(3)}_{13} \neq 0.$

TABLE 5

F_3	1		2		3		Sum over F_3	
F_1 \ F_2	1	2	1	2	1	2	1	2
1	75	24	25	8	20	16	120	48
2	20	16	60	48	16	32	96	96

$u^{(3)}_{123} = 0;\ u^{(3)}_{12} = u^{(2)}_{12,3} \neq 0;\ u^{(3)}_{13} \neq 0,\ u^{(3)}_{23} \neq 0.$