

Modelos de respuesta discreta. Máster oficial en Estadística Aplicada
Relación de ejercicios del Tema 2

Para explicar las relaciones existentes entre la ocurrencia de problemas coronarios y algunas de las variables relacionadas se dispone de una muestra de 200 hombres para los que se han observado las siguientes siete variables explicativas:

- Edad en años
- Presión sistólica en milímetros de mercurio
- Presión diastólica en milímetros de mercurio
- Colesterol en miligramos por DL
- Altura en pulgadas
- Peso en libras
- Problemas coronarios (1 si se tiene alguna incidencia en los 10 años previos y 0 en otro caso)

Ejercicio 1. Ajuste un modelo de regresión logística simple para explicar la ocurrencia de problemas coronarios a partir de la edad. Tras el ajuste conteste a las siguientes cuestiones:

1. Indique la estimación puntual del parámetro independiente e interprete su resultado. (Solución: -4.792766)
2. Indique la estimación puntual del parámetro asociado a la edad e interprete su resultado. (Solución: 0.0633289)
3. Indique el error estándar de estimación del parámetro independiente. (Solución: 0.9595327)
4. Indique el error estándar de estimación del parámetro asociado a la edad. (Solución: 0.0192455)
5. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro independiente (Solución: -4.9948958). Concluya sobre el test a una significación $\alpha = 0.1$
6. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro asociado a la edad (Solución: 3.2905749). Concluya sobre el test a una significación $\alpha = 0.1$
7. Indique el p-valor del test de significación (Wald) del parámetro independiente (Solución: 5.8867557×10^{-7}). Concluya sobre el test a una significación $\alpha = 0.05$.
8. Indique el p-valor del test de significación (Wald) del parámetro asociado a la edad (Solución: 9.998288×10^{-4}). Concluya sobre el test a una significación $\alpha = 0.05$.
9. Reflexione cuánto cambiaría multiplicativamente (por cuánto se multiplicaría) la ventaja de padecer enfermedad coronaria (frente a no padecerla) si la edad aumenta 1 año, 5 años y 10 años (Solución: 1.0653772, 1.3725144 y 1.8837957)
10. Indique cuánto debería aumentar la edad para que la ventaja de padecer enfermedad coronaria (frente a no padecerla) se duplique, se triplique y se cuatriplique (Solución: 10.9452003, 17.3477321 y 21.8904007)
11. Indique un intervalo de confianza al 95% para el parámetro independiente (Solución: $\beta_0 \in (-6.6734156, -2.9121164)$)

-
12. Indique un intervalo de confianza al 90% para el parámetro asociado a la edad (Solución: $\beta_1 \in (0.0316728, 0.094985)$)
 13. Indique el valor experimental y el p-valor del test de bondad de ajuste de Hosmer y Lemeshow (tomando los deciles de riesgo para su obtención). Concluya sobre el test.
(Solución: 5.7868017 y 0.6710992)
 14. Indique cuál sería la probabilidad de padecer problemas coronarios de una persona de 50 años según el modelo. Indique también cuál sería la probabilidad de no padecer problemas coronarios, la ventaja de padecer problemas coronarios y la ventaja de no padecer problemas coronarios. Interprete estos dos últimos resultados
(Solución: $\hat{p}(50) = 0.1643348$, $1 - \hat{p}(50) = 0.8356652$, $\hat{p}(50)/(1 - \hat{p}(50)) = 0.1966514$, $(1 - \hat{p}(50))/\hat{p}(50) = 5.0851396$)
 15. Indique un intervalo de confianza al 95% para la probabilidad de padecer problemas coronarios de una persona de 50 años según el modelo (utilice aproximación normal) (Solución: $p(50) \in (0.1053764, 0.2232931)$)
 16. Indique cuál sería la probabilidad de padecer problemas coronarios de una persona de 66 años según el modelo. Indique también cuál sería la probabilidad de no padecer problemas coronarios, la ventaja de padecer problemas coronarios y la ventaja de no padecer problemas coronarios. Interprete estos dos últimos resultados (Solución: $\hat{p}(66) = 0.3513614$, $1 - \hat{p}(66) = 0.6486386$, $\hat{p}(66)/(1 - \hat{p}(66)) = 0.5416905$, $(1 - \hat{p}(66))/\hat{p}(66) = 1.8460727$)
 17. Indique un intervalo de confianza al 90% para la probabilidad de padecer problemas coronarios de una persona de 66 años según el modelo (utilice aproximación normal) (Solución: $p(66) \in (0.201486, 0.5012368)$)
 18. Indique cuántos residuos de Pearson estandarizados son no significativos y el porcentaje de residuos de la deviance no significativos (Solución: 17, 6.5)
 19. Calcule CCR, Sensibilidad y Especificidad (en %) del test diagnóstico asociado a este modelo, utilizando como punto de corte 0.13. Interprete las tres cantidades
(Solución: CC= 64%, Sen=61.5384615%, Espec=64.3678161%)
 20. Calcule el área bajo la curva ROC
(Solución: 0.6984969)

Ejercicio 2. Ajuste un modelo de regresión logística simple para explicar la ocurrencia de problemas coronarios a partir de la presión sistólica. Tras el ajuste conteste a las siguientes cuestiones:

1. Indique la estimación puntual del parámetro independiente e interprete su resultado. (Solución: -5.324702)
2. Indique la estimación puntual del parámetro asociado a la presión sistólica e interprete su resultado. (Solución: 0.0274914)
3. Indique el error estándar de estimación del parámetro independiente. (Solución: 1.3980421)
4. Indique el error estándar de estimación del parámetro asociado a la presión sistólica. (Solución: 0.0108544)
5. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro independiente (Solución: -3.808685). Concluya sobre el test a una significación $\alpha = 0.1$
6. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro asociado a la presión sistólica (Solución: 2.5327378). Concluya sobre el test a una significación $\alpha = 0.1$

-
7. Indique el p-valor del test de significación (Wald) del parámetro independiente (Solución: 1.3970778×10^{-4}). Concluya sobre el test a una significación $\alpha = 0.05$.
 8. Indique el p-valor del test de significación (Wald) del parámetro asociado a la presión sistólica (Solución: 0.0113176). Concluya sobre el test a una significación $\alpha = 0.05$.
 9. Reflexione cuánto cambiaría multiplicativamente (por cuánto se multiplicaría) la ventaja de padecer enfermedad coronaria (frente a no padecerla) si la presión sistólica aumenta 1 unidad, 10 unidades y 20 unidades (Solución: 1.0278727, 1.3164171 y 1.7329541)
 10. Indique cuánto debería aumentar la presión sistólica para que la ventaja de padecer enfermedad coronaria (frente a no padecerla) se duplique, se triplique y se cuatriplique (Solución: 25.2132599, 39.9620714 y 50.4265198)
 11. Indique un intervalo de confianza al 95% para el parámetro independiente (Solución: $\beta_0 \in (-8.0648142, -2.5845899)$)
 12. Indique un intervalo de confianza al 90% para el parámetro asociado a la presión sistólica (Solución: $\beta_1 \in (0.0096375, 0.0453453)$)
 13. Indique el valor experimental y el p-valor del test de bondad de ajuste de Hosmer y Lemeshow (tomando los deciles de riesgo para su obtención). Concluya sobre el test. (Solución: 17.5468682 y 0.0248924)
 14. Indique cuál sería la probabilidad de padecer problemas coronarios de una persona con 150 de presión sistólica según el modelo. Indique también cuál sería la probabilidad de no padecer problemas coronarios, la ventaja de padecer problemas coronarios y la ventaja de no padecer problemas coronarios. Interprete estos dos últimos resultados
(Solución: $\hat{p}(150) = 0.2312981$, $1 - \hat{p}(150) = 0.7687019$, $\hat{p}(150)/(1 - \hat{p}(150)) = 0.3008944$, $(1 - \hat{p}(150))/\hat{p}(150) = 3.3234247$)
 15. Indique un intervalo de confianza al 95% para la probabilidad de padecer problemas coronarios de una persona de 150 de presión sistólica según el modelo (utilice aproximación normal) (Solución: $p(150) \in (0.1174378, 0.3451585)$)
 16. Indique cuál sería la probabilidad de padecer problemas coronarios de una persona de 145 de presión sistólica según el modelo. Indique también cuál sería la probabilidad de no padecer problemas coronarios, la ventaja de padecer problemas coronarios y la ventaja de no padecer problemas coronarios. Interprete estos dos últimos resultados (Solución: $\hat{p}(145) = 0.2077646$, $1 - \hat{p}(145) = 0.7922354$, $\hat{p}(145)/(1 - \hat{p}(145)) = 0.2622511$, $(1 - \hat{p}(145))/\hat{p}(145) = 3.8131387$)
 17. Indique un intervalo de confianza al 90% para la probabilidad de padecer problemas coronarios de una persona de 145 de presión sistólica según el modelo (utilice aproximación normal) (Solución: $p(145) \in (0.1297984, 0.2857309)$)
 18. Indique cuántos residuos de Pearson estandarizados son no significativos y el porcentaje de residuos de la deviance no significativos (Solución: 22, 8.5)
 19. Calcule CCR, Sensibilidad y Especificidad (en %) del test diagnóstico asociado a este modelo, utilizando como punto de corte 0.13. Interprete las tres cantidades
(Solución: CC= 70.5%, Sen=57.6923077%, Espec=72.4137931%)
 20. Calcule el área bajo la curva ROC (Solución: 0.6337312)

Ejercicio 3. Tras agrupar a los individuos de la muestra según su presión sistólica se tiene la siguiente información

Presión	Cor=0	Cor=1	Presión	Cor=0	Cor=1
90	2	0	130	15	6
95	1	0	132	1	0
100	13	0	134	3	0
105	2	0	135	1	0
106	1	0	136	1	0
108	4	0	138	2	2
110	35	9	140	5	2
112	4	0	141	0	1
114	4	0	148	1	1
115	10	0	150	2	0
118	2	0	154	1	0
120	40	2	155	1	0
122	4	0	156	1	0
124	4	0	160	4	1
125	5	0	164	1	0
126	1	0	168	0	1
128	1	0	190	2	1

Ajuste un modelo de regresión logística simple para explicar la ocurrencia de problemas coronarios a partir de la presión sistólica. Tras el ajuste conteste a las siguientes cuestiones:

1. Indique la estimación puntual del parámetro independiente e interprete su resultado. (Solución: -5.324702)
2. Indique la estimación puntual del parámetro asociado a la presión sistólica e interprete su resultado. (Solución: 0.0274914)
3. Indique el error estándar de estimación del parámetro independiente. (Solución: 1.3980548)
4. Indique el error estándar de estimación del parámetro asociado a la presión sistólica. (Solución: 0.0108545)
5. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro independiente (Solución: -3.8086505). Concluya sobre el test a una significación $\alpha = 0.1$
6. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro asociado a la presión sistólica (Solución: 2.5327204). Concluya sobre el test a una significación $\alpha = 0.1$
7. Indique el p-valor del test de significación (Wald) del parámetro independiente (Solución: 1.3972731×10^{-4}). Concluya sobre el test a una significación $\alpha = 0.05$.
8. Indique el p-valor del test de significación (Wald) del parámetro asociado a la presión sistólica (Solución: 0.0113181). Concluya sobre el test a una significación $\alpha = 0.05$.
9. Reflexione cuánto cambiaría multiplicativamente (por cuánto se multiplicaría) la ventaja de padecer enfermedad coronaria (frente a no padecerla) si la presión sistólica aumenta 1 unidad, 10 unidades y 20 unidades (Solución: 1.0278727, 1.3164171 y 1.7329541)
10. Indique cuánto debería aumentar la presión sistólica para que la ventaja de padecer enfermedad coronaria (frente a no padecerla) se duplique, se triplique y se cuatriplique (Solución: 25.2132598, 39.9620712 y 50.4265195)

-
11. Indique un intervalo de confianza al 95% para el parámetro independiente (Solución: $\beta_0 \in (-8.0648391, -2.584565)$)
 12. Indique un intervalo de confianza al 90% para el parámetro asociado a la presión sistólica (Solución: $\beta_1 \in (0.0096373, 0.0453454)$)
 13. Indique el valor experimental y el p-valor del test de bondad de ajuste mediante el estadístico de Wilks. (Solución: 37.1642724 y 0.2431694)
 14. Indique cuál sería la probabilidad de padecer problemas coronarios de una persona con 150 de presión sistólica según el modelo. Indique también cuál sería la probabilidad de no padecer problemas coronarios, la ventaja de padecer problemas coronarios y la ventaja de no padecer problemas coronarios. Interprete estos dos últimos resultados
(Solución: $\hat{p}(150) = 0.2312981$, $1 - \hat{p}(150) = 0.7687019$, $\hat{p}(150)/(1 - \hat{p}(150)) = 0.3008944$, $(1 - \hat{p}(150))/\hat{p}(150) = 3.3234247$)
 15. Indique un intervalo de confianza al 95% para la probabilidad de padecer problemas coronarios de una persona de 150 de presión sistólica según el modelo (utilice aproximación normal) (Solución: $p(150) \in (0.1174375, 0.3451588)$)
 16. Indique cuál sería la probabilidad de padecer problemas coronarios de una persona de 145 de presión sistólica según el modelo. Indique también cuál sería la probabilidad de no padecer problemas coronarios, la ventaja de padecer problemas coronarios y la ventaja de no padecer problemas coronarios. Interprete estos dos últimos resultados (Solución: $\hat{p}(145) = 0.2077646$, $1 - \hat{p}(145) = 0.7922354$, $\hat{p}(145)/(1 - \hat{p}(145)) = 0.2622511$, $(1 - \hat{p}(145))/\hat{p}(145) = 3.8131387$)
 17. Indique un intervalo de confianza al 90% para la probabilidad de padecer problemas coronarios de una persona de 145 de presión sistólica según el modelo (utilice aproximación normal) (Solución: $p(145) \in (0.1297981, 0.2857312)$)
 18. Indique cuántos residuos de Pearson estandarizados son no significativos y el porcentaje de residuos de la deviance no significativos (Solución: 45, 22.5)
 19. Calcule CCR, Sensibilidad y Especificidad (en %) del test diagnóstico asociado a este modelo, utilizando como punto de corte 0.13. Interprete las tres cantidades
(Solución: CC= 70.5%, Sen=57.6923077%, Espec=72.4137931%)
 20. Calcule el área bajo la curva ROC (Solución: 0.6337312)

Ejercicio 4. Ajuste un modelo de regresión logística múltiple para explicar la ocurrencia de problemas coronarios a partir del peso y el nivel de colesterol con interacción. Tras el ajuste conteste a las siguientes cuestiones:

1. Indique la estimación puntual del parámetro independiente e interprete su resultado. (Solución: 5.5057806)
2. Indique la estimación puntual del parámetro asociado al peso e interprete su resultado. (Solución: -0.0606286)
3. Indique la estimación puntual del parámetro asociado al nivel de colesterol e interprete su resultado. (Solución: -0.0370668)
4. Indique la estimación puntual del parámetro asociado a la interacción e interprete su resultado. (Solución: 2.7505793×10^{-4})

-
5. Indique el error estándar de estimación del parámetro independiente. (Solución: 9.0257328)
 6. Indique el error estándar de estimación del parámetro asociado al peso. (Solución: 0.0538084)
 7. Indique el error estándar de estimación del parámetro asociado al nivel de colesterol. (Solución: 0.0313064)
 8. Indique el error estándar de estimación del parámetro asociado a la interacción. (Solución: 1.8603658×10^{-4})
 9. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro independiente (Solución: 0.6100093). Concluya sobre el test a una significación $\alpha = 0.1$
 10. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro asociado al peso (Solución: -1.1267484). Concluya sobre el test a una significación $\alpha = 0.1$
 11. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro asociado al nivel de colesterol (Solución: -1.184001). Concluya sobre el test a una significación $\alpha = 0.1$
 12. Indique el valor experimental (valor estimado del estadístico) del test de significación (Wald) del parámetro asociado a la interacción (Solución: 1.4785153). Concluya sobre el test a una significación $\alpha = 0.1$
 13. Indique el p-valor del test de significación (Wald) del parámetro independiente (Solución: 0.5418557). Concluya sobre el test a una significación $\alpha = 0.05$.
 14. Indique el p-valor del test de significación (Wald) del parámetro asociado al peso (Solución: 0.2598489). Concluya sobre el test a una significación $\alpha = 0.05$.
 15. Indique el p-valor del test de significación (Wald) del parámetro asociado al nivel de colesterol (Solución: 0.2364127). Concluya sobre el test a una significación $\alpha = 0.05$.
 16. Indique el p-valor del test de significación (Wald) del parámetro asociado a la interacción (Solución: 0.1392699). Concluya sobre el test a una significación $\alpha = 0.05$.
 17. Indique un intervalo de confianza al 95% para el parámetro independiente (Solución: $\beta_0 \in (-12.1843307, 23.1958918)$)
 18. Indique un intervalo de confianza al 90% para el parámetro asociado al peso (Solución: $\beta_1 \in (-0.1491355, 0.0278784)$)
 19. Indique un intervalo de confianza al 92% para el parámetro asociado al nivel de colesterol (Solución: $\beta_2 \in (-0.0918744, 0.0177409)$)
 20. Indique un intervalo de confianza al 99% para el parámetro asociado a la interacción (Solución: $\beta_3 \in (-2.0414054 \times 10^{-4}, 7.5425639 \times 10^{-4})$)
 21. Indique el valor experimental y el p-valor del test de bondad de ajuste de Hosmer y Lemeshow (tomando los deciles de riesgo para su obtención). Concluya sobre el test. (Solución: 7.855347 y 0.4477265)
 22. Indique cuál sería la probabilidad de padecer problemas coronarios de una persona que pesa 170 libras y tiene un nivel de colesterol de 280 según el modelo. Indique también cuál sería la probabilidad de no padecer problemas coronarios, la ventaja de padecer problemas coronarios y la ventaja de no padecer problemas coronarios. Interprete estos dos últimos resultados
(Solución: $\hat{p}(170, 280) = 0.1103651$, $1 - \hat{p}(170, 280) = 0.8896349$, $\hat{p}(170, 280)/(1 - \hat{p}(170, 280)) = 0.1240567$, $(1 - \hat{p}(170, 280))/\hat{p}(170, 280) = 8.0608317$)
 23. Indique un intervalo de confianza al 95% para la probabilidad de padecer problemas coronarios de una persona que pesa 170 libras y tiene un nivel de colesterol de 280 según el modelo (utilice aproximación normal) (Solución: $p(170, 280) \in (0.0629236, 0.1578067)$)

-
24. Indique cuál sería la probabilidad de padecer problemas coronarios de una persona que tiene un nivel de colesterol de 315 y pesa 191 libras según el modelo. Indique también cuál sería la probabilidad de no padecer problemas coronarios, la ventaja de padecer problemas coronarios y la ventaja de no padecer problemas coronarios. Interprete estos dos últimos resultados (Solución: $\hat{p}(191, 315) = 0.2312172$, $1 - \hat{p}(191, 315) = 0.7687828$, $\hat{p}(191, 315)/(1 - \hat{p}(191, 315)) = 0.3007575$, $(1 - \hat{p}(191, 315))/\hat{p}(191, 315) = 3.3249382$)
25. Indique un intervalo de confianza al 90% para la probabilidad de padecer problemas coronarios de una persona que tiene un nivel de colesterol de 315 y pesa 191 libras según el modelo (utilice aproximación normal) (Solución: $p(191, 315) \in (0.1477944, 0.31464)$)
26. Indique cuántos residuos de Pearson estandarizados son no significativos y el porcentaje de residuos de la deviance no significativos (Solución: 17, 6.5)
27. Calcule CCR, Sensibilidad y Especificidad (en %) del test diagnóstico asociado a este modelo, utilizando como punto de corte 0.13. Interprete las tres cantidades (Solución: CC= 71%, Sen=57.6923077%, Espec=72.9885057%)
28. Calcule el área bajo la curva ROC (Solución: 0.6978338)

Ejercicio 5. Repita el ejercicio anterior sin incluir la interacción e interprete los parámetros. Observe las diferencias.