Homework 06

Spencer Pease

2/28/2020

(Q1) Association between age and CHD events

(Q1.a)

Fitted logistic regression model with *having a CHD event* as the binary response and *age* and the predictor:

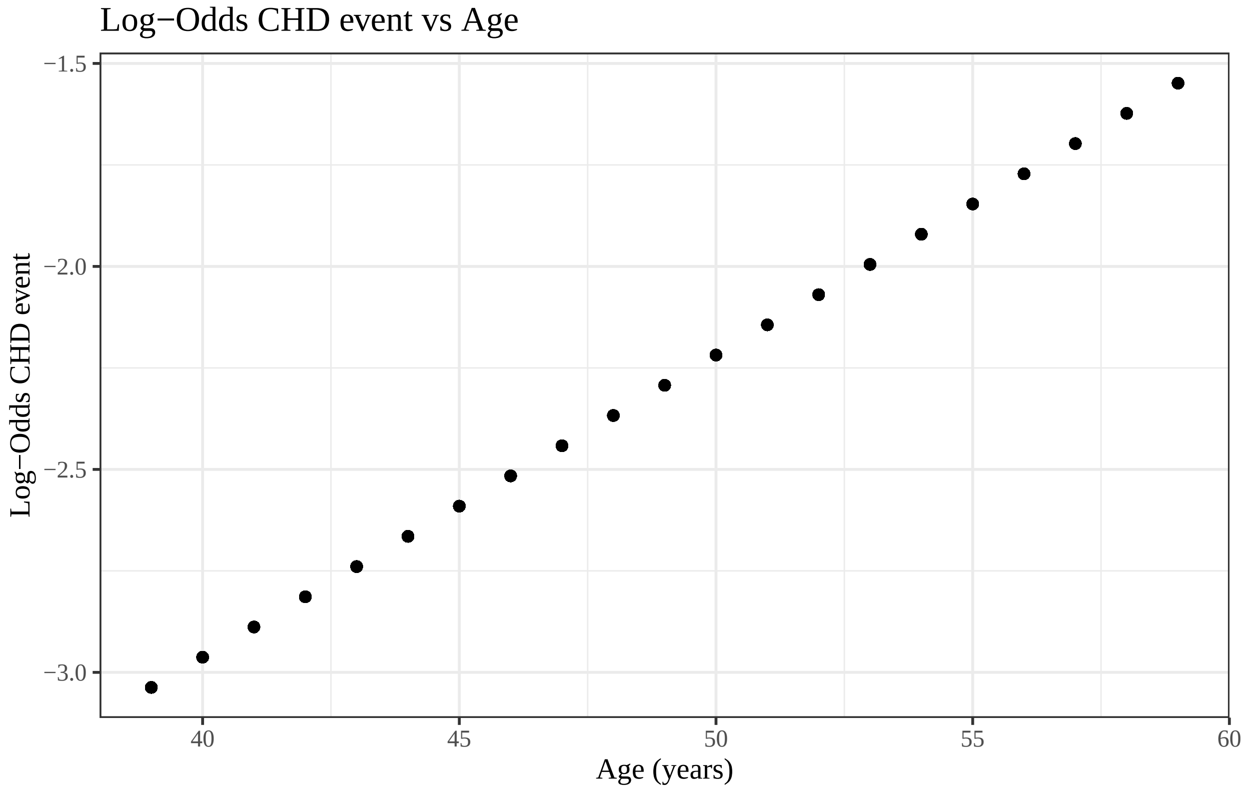
$$logit(P(chd = 1 \mid age)) = \beta_0 + \beta_1 \cdot age$$

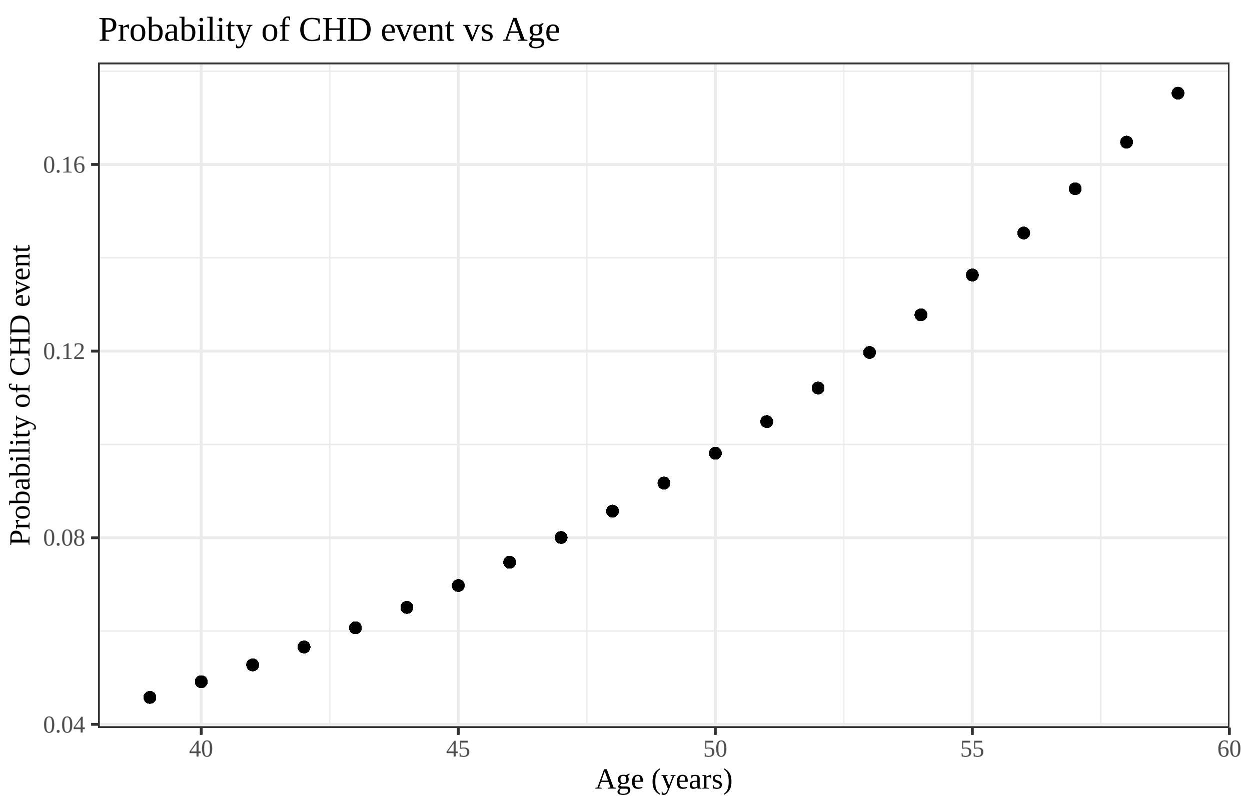
Table 1: Inferential statistics of log-odds CHD event vs age

Parameter	Estimate	Robust SE	e(Est)	e(95%L)	e(95%H)	Pr(>|z|)
(Intercept)	-5.940	0.547	0.003	0.001	0.008	5.65e-27
age	0.074	0.011	1.077	1.054	1.101	4.52e-11

From the fitted model, we estimate that on average the odds of having a CHD event within five years of the study start date for the subset of the population zero years old (e^{β_0}) is 0 (95% CI: 0, 0.01). This value is nonsensical, since it is far outside the range of observed data and we have no reason to think the relationship between middle-aged men and CHD events would apply to children. Comparing two groups differing by one year in age, on average the odds ratio e^{β_1} of experiencing a CHD event within five years of the study start date is 1.08 (95% CI: 1.05, 1.1). This is of scientific interest for testing for an association between age and CHD events in the population within the range of observed data.

(Q1.b,c)


Log−Odds CHD event vs Age
−1.5
−2.0
−2.5
−3.0
Log−Odds CHD event
40
45
50
55
60
Age (years)


Probability of CHD event vs Age
0.16
0.12
0.08
0.04
Probability of CHD event
40
45
50
55
60
Age (years)

(Q1.d)

Table 2: Calculated risk ratio (RR) and odds ratio (OR)

Age	Prob.	Odds	RR	OR
40	0.05	0.05	-	-
45	0.07	0.07	1.42	1.45
50	0.10	0.11	1.41	1.45
55	0.14	0.16	1.39	1.45
60	0.19	0.23	1.37	1.45

From the table, we see that both the probability and odds of having a CHD event are estimated to be higher on average for older ages. When looking at the risk ratio (RR) and odds ratio (OR) between ages, we see that RR is lower between older ages when the difference in age is constant, while the OR is the same between any two ages of equal separation.

(Q1.e)

$$OR_5 = \frac{Odds_{x+5}}{Odds_x} = \frac{e^{\beta_0}e^{\beta_1(x+5)}}{e^{\beta_0}e^{\beta_1(x)}} = e^{\beta_1(x+5)-\beta_1(x)} = e^{\beta_1(5)}$$

Table 3: Odds ratio comparing men 5 years apart in age

Estimate	95%L	95%H
1.45	1.3	1.62

The estimated average odds ratio between two subsets of the population five years apart is 1.45 (95% CI: 1.3, 1.62). This odds ratio matches the calculated odds ratios from (Q1.d), confirming that the odds ratio is constant for a given interval between two groups, regardless of the absolute value of the groups.

(Q1.f)

Table 4: Wald test of age not being associated with CHD events

Estimate	Robust SE	Pr(>|z|)
0.074	0.011	4.52e-11

Performing a Wald test on our logistic regression model fit with robust standard error estimates, we get a P-value of 4.52e-11. Combined with an estimated $\beta_1 \neq 0$, we have sufficient evidence to reject the null hypothesis that age is not associated with CHD events in the population.

(Q1.g)

Table 5: likelihood ratio test of age not being associated with CHD events

model	Resid. Df	Resid. Dev	df	Deviance	Pr(>Chi)
chd ~ 1	3153	236.1	-	-	-
chd ~ age	3152	232.7	1	3.4	1.54e-11

Performing a likelihood ratio test also provides sufficient evidence to reject the null hypothesis that age is not associated with CHD events. Comparing the P-value from this test to the Wald test, we see that they are similar (on the same order of magnitude), which shows that while both tests are not equivalent, they become more similar when working with a large number of observations.

(Q2) Association between behavior and CHD events

(Q2.a)

Table 6: Contingency table of CHD events and behavior types

chd	A	B	Sum
0	1411	1486	2897
1	178	79	257
Sum	1589	1565	3154

Table 7: Chi-square test of CHD events and behavior types

X-squared	P-value	df
39.08	4.07e-10	1

With a χ^2-test statistic not equal to zero and a P-value much less than 0.05, we have sufficient evidence to reject the null hypothesis that there is no association between CHD events and behavior type in the population.

Estimating Probability:

- $P(chd \mid type = A) : 178 \div 1589 = 0.11$
- $P(chd \mid type = B) : 79 \div 1565 = 0.05$

Estimating Odds:

- $odds(chd \mid type = A) : 178 \div 1411 = 0.13$
- $odds(chd \mid type = B) : 79 \div 1486 = 0.05$

Estimating Odds Ratio:

- $OR : (1411 \cdot 79) \div (178 \cdot 1486) = 0.42$

(Q2.b)

Table 8: Inferential statistics of LRM with response CHD and predictor Type

Parameter	Estimate	Robust SE	e(Est)	e(95%L)	e(95%H)	Pr(>|z|)
(Intercept)	-2.070	0.08	0.126	0.108	0.147	2.26e-135
typeB	-0.864	0.14	0.421	0.320	0.555	8.13e-10

From the logistic regression model fit with robust standard error estimates, we estimate that the odds of having a CHD event given type B behavior are on average 0.42 (95% CI: 0.32, 0.55) times higher than the odds of having a CHD event given type A behavior in the population. A P-value of 8.13e-10 suggests we can reject the null hypothesis that there is no association between CHD events and behavior type.

(Q2.c)

Table 9: Inferential statistics of LRM with response Type and predictor CHD

Parameter	Estimate	Robust SE	e(Est)	e(95%L)	e(95%H)	Pr(>|z|)
(Intercept)	0.052	0.037	1.053	0.979	1.133	1.64e-01
chd	-0.864	0.140	0.421	0.320	0.555	8.13e-10

(Q2.d)

The odds ratios (e^{β_1}) in *Q2.a* and *Q2.b* are the same, since they both capture the ratio of odds of a CHD event given each behavior type. The P-values are different, because each question is performing a different kind of test of association.

(Q2.e)

The odds ratios (e^{β_1}) in *Q2.b* and *Q2.c* are the same because odds ratios are symmetric when swapping binary predictor and binary response (as seen in *homework 5 Q4*).