MS2 2012

Question 1

Best known equations for HR_{max} = (220 - age) bpm

Experiment to ascertain validity of the (220 - age) equation

Healthy volunteers aged 18 - 80 were recruited and had HR_{max} determined by exercise phys.

N = sample size

A_i = age
H_i = HR_{max}) of ith subject

$N = 40$

$\Sigma A_i = 1880$

$\Sigma A_i^2 = 101,374$

$\Sigma H_i = 7040$

$\Sigma H_i^2 = 1,249,420$

$\Sigma A_i \cdot H_i = 322,284$

$\hat{\sigma}^2 = S^2_{yx} = 123.74 \quad \left(\frac{SSE}{df_E}\right)$

a) Write down a linear model relating HR to age.

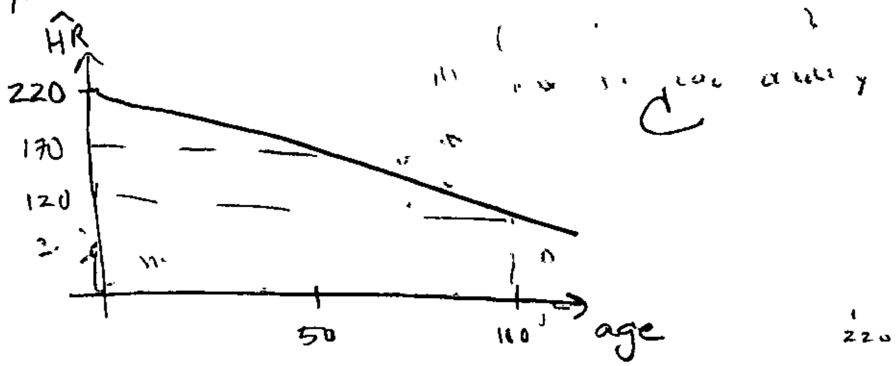
$E(HR_{max})$ = intercept age ← age in years

predicted HR based on age

$= \beta_0 + \beta_1 \text{age}$

when age = 0 (birth) — drop in HR max w/ each add'l year

b) Assuming (220 - age) is correct, state values of the parameters in your linear model.



1.c) fit $^{\varepsilon-1,}$ model to data ⇒ assuming model from A

$\hat{HR} = \beta_0 + \beta_1 \, age$

$X = \begin{pmatrix} 1 & a_1 \\ \vdots & \vdots \\ \cdot & a_{40} \end{pmatrix}$ $(X'X) = \begin{pmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & \end{pmatrix} \begin{pmatrix} 1 & a_1 \\ \vdots & \vdots \\ 1 & a_{40} \end{pmatrix}$ $X'y = \begin{pmatrix} 1 & \cdots & 1 \\ a & \cdots & \end{pmatrix} \begin{pmatrix} hr \\ \vdots \\ hr \end{pmatrix}$

$\hat{\beta} = (X'X)^{-1} X'y$ $= \begin{pmatrix} 40 & \Sigma a_i \\ \Sigma a_i & \Sigma a_i^2 \end{pmatrix}^{-1}$ $= \begin{pmatrix} \Sigma hr \\ \Sigma a \cdot hr \end{pmatrix}$

$\hat{\beta} = \begin{pmatrix} 207.0444 \\ -0.6605 \end{pmatrix} \begin{matrix} = \hat{\beta}_0 \\ = \hat{\beta}_1 \end{matrix}$

$\hat{\beta}_1 = \dfrac{\Sigma A_i H_i - N \bar{A} \bar{H}}{\Sigma A_i^2 - N \bar{A}^2}$

d) Test whether data are consistent w/ $\beta_1 = -1$ (Δ implied w/ formula)

$\hat{\beta}_0 = \bar{H} - \beta_1 \bar{A}$

$H_0: \beta_1 = -1$

$C = (0 \quad 1)$ $\theta_0 = (-1)$

$M^{-1} = C(X'X)^{-1} C'$

$\hat{\theta} = -0.6605$

$$\frac{\{(\hat{\theta} - \theta_0)' M^{-1} (\hat{\theta} - \theta_0)\}/1}{MSE} \sim F_{1,\,38}$$

or $\dfrac{\beta_1 - -1}{\sqrt{Var \beta_1}} =$

either way, test stat is large so reject the null. Different from formula.

e) 50 yr old → $HR_{max} = 160$
different from expectation?

$H_0: \hat{\beta}_0 + \hat{\beta}_1(50) = 160$

expected = $207.044 - 0.6605(60)$

$c = (1 \quad 50)$

$= \cancel{} = 167.414$

craigs is wrong

$H_0: HR_{50} = 160 \qquad \hat{H} = 167.414$

$t = \cancel{\frac{160 - 167.414}{}}$ ← don't like this way

$c = (1 \quad 60) \qquad \theta_0 = 160$

$c\hat{\beta} \doteq \hat{\theta} \doteq 167.414$

$F = 11.69 \quad {}_{1,50}$ reject, it is diff

f) add interaction

g) new $\Sigma A_i = 1880 + 47$

$\Sigma A_i^2 = 101374 + 47^2$

$\Sigma A \cdot H = 322284 + (47 \cdot 196)$

$\Sigma H = 7040 + 196$

then recalc as in c)

for the others, do it and then explain

Question 2 | MS-Q 2012

(a) Income, age, educ.

$y = \beta_0 + \beta_1 \, age + \beta_1 \, educ + \varepsilon_i \sim N(0, \sigma^2)$ $\quad n = 20$

$y_i \perp y_j$

$\vec{\varepsilon} \sim N_{20}(0, I_{20}(\sigma^2))$

$\vec{y} \sim N_{20}(X\beta, I_{20}\sigma^2)$

or $\alpha_1 = \dfrac{\Sigma x_i y_i - N\bar{x}\bar{y}}{\Sigma x_i^2 - N\bar{x}^2} = \dfrac{88328.1 - 20\left(\frac{757}{20}\right)\left(\frac{2301.6}{20}\right)}{29199 - 20\left(\frac{757}{20}\right)^2}$

$\alpha_0 = \bar{y} - \beta_1 \bar{x} =$

b) $Ey = \alpha_0 + \alpha_1 \, age$

$\Sigma y_i = 2301.6$

$\Sigma x_i y_i = 88328.1$

$\Sigma x_i = 757$

$\Sigma x_i^2 = 29199$

$$\hat{\alpha} = (X'X)^{-1}(X'y)$$

$$= \left\{\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_{20} \end{pmatrix}\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{20} \end{pmatrix}\right\}^{-1}\left\{\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_{20} \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_{20} \end{pmatrix}\right\}$$

$$= \begin{pmatrix} 20 & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{pmatrix}^{-1}\begin{pmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{pmatrix}$$

$$= \begin{pmatrix} 20 & 757 \\ 757 & 29199 \end{pmatrix}^{-1}\begin{pmatrix} 2301.6 \\ 88328.1 \end{pmatrix} = \begin{pmatrix} 31.108471 \\ 2.218534 \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{pmatrix}$$

$(2\times1)^T \quad 1\times2 \quad = 2\times2$

this way

c) consider two simple regs $\quad$ inc = age, $\quad$ inc = educ

$\qquad 0.67 \qquad 0.71 \quad = R^2$

NO ⇒ individually, they explain that much. however, age and education may / are not independent. Thus we cannot assume their effect together would be additive.

d) $\hat{E}Y = \beta_0 + \beta_1 \text{age} + \beta_2 \text{educ}$ $\hat{\sigma} = MSE = \frac{SSS}{df_E} = .5159$

$(X'X)^{-1} = \Big(\;\Big)_{3\times3}$

$H_0: \beta_0 = \beta_1 = \beta_2$ using GLH approach $\rightarrow \theta = C\beta = \theta_0$

$C = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$ $\theta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\beta_0 - \beta_1 = 0$, $\beta_1 - \beta_2 = 0$

don't have $\vec{\hat{\beta}}$

$\hat{\theta} = C\hat{\beta} =$

e) $\text{corr}(\hat{\beta}, \hat{\beta}_2) = \dfrac{\text{cov}(\hat{\beta}_1, \hat{\beta}_2)}{\sqrt{\text{var}\hat{\beta}_1 \, \text{var}\hat{\beta}_2}}$ — cell (1,2) of $((X'X)^{-1}\sigma^2)$

cell 11 cell 22

← same approach diff. #'s

f) the x's don't need to be normal.
The normality assumption of HILE Gauss (only applies to the conditional distributions and error

Question 3 y = treatment efficacy

$= \beta_0 + trt + age + gender + trt \times age + trt \times gender$

n=100, 25 per gender / trt combo (2 genders, 2 treatments) → med or counsel

a) write the relation btwn age and trt efficacy for each combo

TRT	Gender	Relation
med	female	$\beta_0 + \beta_{age}$
med	male	$\beta_0 + \beta_{age} + \beta_{gender}$
counsel	female	$\beta_0 + \beta_{trt} + \beta_{age} + \beta_{trt \times age}$
counsel	male	$\beta_0 + \beta_{trt} + \beta_{age} + \beta_{gender} + \beta_{trt \times age} + \beta_{trt \times gender}$

b) Reg. coeffs obtained.

Interpret coeffs:

- ME = Δ in efficacy for (non-ref or 1 unit increase) in x
- INT = add'l effect of an increase in age / being male for those on trt = 1 (add effect of BOTH)

To test whether there are gender diff in trt eff given they are the same age, we should test whether

$H_0: \beta_{trt \times gender} = 0$ $\quad C = (0\ 0\ 0\ 0\ 0\ 1)$ $\quad \theta_0 = 0$ $\quad \hat{\theta} = C\hat{\beta}$

$F = \frac{(\hat{\theta} - \theta_0)}{MSE}$ or $\frac{\beta_6}{\sqrt{Var \beta_6}} \sim t_{94}$

$F \sim 1, 100 - 6 = 94$

c) added last. (Wald test of coeff = 0)

non-sign terms are trt x age, age (close), and trt.

probably want to keep trt given trt x gender is sign.

d) Fill in ANCOVA table

Type I / MSE

Source	df	Type I	MS	F	p
trt	1		— 45.72234	48.584	< 0...
age	1		— 0.5817	0.6181	0.43
gender	1		— 107.163	113.8701	< 0..
trt x age	1		— 0.3369	$0.3580 = \frac{MS}{MSE}$	0.55
trt x gender	1	13.649	13.649	$\boxed{12.845056}$	
error	94	88.46328	$MSE = \frac{SSE}{94} = 0.9410987$		
total	99				

↑ mse*df

↑ t^2 from b)

last added in order = added last

F-stat, and df, for testing

$E(y) = \beta_0 + \beta_1 trt + \beta_2 age + \beta_3 gender + \beta_4 trt \times age + \beta_5 trt \times g$

craig $SS = (SSH) = (SS\beta_4 + SS\beta_5)/2$

vs.

$E(y) = \beta_0 + \beta_1 trt + \beta_2 age + \beta_3 gender$

$\Rightarrow H_0: \beta_4 = \beta_5 = 0$

$\hat{\theta} = C\hat{\beta} = \begin{pmatrix} -0.004731 \\ 1.529518 \end{pmatrix}$

$C = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \theta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

or use SSH of β_4 and β_5 w/ MSE etc. testing.

$F = \frac{(\hat{\theta} - \theta_0)' M^{-1} (\hat{\theta} - \theta_0)/df\,H}{MSE} = $ etc.

$\sim 2, 9$

where

$M^{-1} = C(X'X)^{-1}C'$

e) μ_m = mean med when $g=0$ age $=33$

μ_c = mean counsel when $g=0$ age $=33$

$H_0: \mu_m = \mu_c$

$$\cancel{\beta_0} + \cancel{\beta_2(33)} = \cancel{\beta_0} + \beta_1 + \cancel{\beta_2}(33) + \beta_4(33)$$

$$\Rightarrow \beta_1 + \beta_4(33) = 0$$

$C = (0\ 1\ 0\ 0\ 33\ 0)$

$\theta_0 = 0$

$F \sim 1, 94$

$\perp$ age when age $=0$ (?)

~~2012~~ ~~Applied~~ Question 4 ([illegible]) | MS-2 2012

④ current smoking status, Y/n cough that day

a) Two separate logistic reg. models are specified. Provide the algebraic expressions for each as well as interpretations.

i. logit(cough Y/n) = int + smoke [illegible] [illegible]
β_0 = log odds for having a cough (if non-smoker))
β_{smoke} = 0.5415 (for every 1 unit increase in smoking freq, the log odds of having a cough increase by 0.5415 → treating as linear

ii logit(cough) = β_0 + $\beta_{occas.}$ + $\beta_{regular}$
β_0 = log. odds for coughing when $X_1 = X_2 = 0$ (non-smoker)
β_{occ} : log OR for having a cough if an occas smoker compared w/ a non-smoker
β_{reg} = " " regular smoker --

b) Interpret the model w/ smoke as a covariate but without an intercept term. (this model wasn't run... correct?)

$$\text{logit}\left(\frac{Y}{n}\right) = \beta_1 (\text{smoke})$$

increase in prob. of smoking assuming no coughing among non-smokers.

c) adv interval → easier to understand (simpler)

disadv interval → assumes the jump from each level of smoking is =

adv cat → allows the effect of each successive 'level' of coughing to differ

disadv cat → might be harder to interpret /use.

ha, almost exactly the same as Craig

d) estimate P(coughing) for occassional smokers

i. $P(\text{cough}) = \dfrac{\exp(\beta_0 + \beta_1(1))}{1 + \exp(\beta_0 + \beta_1(1))} = 0.298615$

ii. $P(\text{cough}) = \dfrac{\exp(\beta_0 + \beta_1)}{1 + (\exp(\beta_0 + \beta_1))} = 0.287901$

e) Pick preferred model, write 2-3+ sentences summarizing