Question 11

So ΣX for IE deaths $= 441 - 80$

antib. $= 26$

IE cases $= 91$

---

a) Has there been a change in avg. monthly Rx?

$H_0: \bar{X}_{Before} = \bar{X}_{after}$

$\bar{X}_{Before} = \frac{5568}{51}$ $\quad Var = \frac{\Sigma(X_i - \bar{X})^2}{n-1} = \Sigma \bar{X}_i^2 - n\bar{X} = \frac{612890 - 51\left(\frac{5568}{51}\right)}{51-1}$

$\bar{X}_{After} = \frac{667}{19}$ $\quad Var = \quad = \frac{27783 - 19\left(\frac{667}{19}\right)}{~~1~~ 19-1}$

$\Rightarrow$ compute t test of 2 means

---

b) monthly IE cases $\nearrow X$ and deaths $\nearrow Y$ correlated?

$$\rho = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2 \, \Sigma(Y_i - \bar{Y})^2}} = \frac{\Sigma(X_iY_i - X_i\bar{Y} - Y_i\bar{X} + \bar{X}\bar{Y})}{\sqrt{(\Sigma X_i^2 - n\bar{X})(\Sigma Y_i^2 - n\bar{Y})}}$$

$$= \frac{\Sigma X_iY_i - n\bar{X}\bar{Y} - \cancel{n\bar{Y}\bar{X}} + \cancel{n\bar{X}\bar{Y}}}{\sqrt{(\Sigma X_i^2 - n\bar{X})(\Sigma Y_i^2 - n\bar{Y})}}$$

$\leadsto$ calculate based on sums of squares

* adjust ΣX_i = IE deaths by subtracting 80
* adjust $\Sigma X_iY_i = 53470$ by sub - 7280

$97(91) = 8827 \quad 17(91) = 1547$

c) Through Mar 2008

y = cases

x = month

$(\cancel{X'X})^{-1}$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}\bar{x}$$

662 notes
Lin Reg Pt 7 p. 13

SE's by similar logic $= \sigma^2 / (\sum x_i^2 - n\bar{x})$ p. 24

$\hat{\beta}_1$ = change in 1Σ cases w/ each addl month.

test $\dfrac{\hat{\beta}_1}{\sqrt{\operatorname{Var}\hat{\beta}_1}} \sim t_{n-2}$ $n = 57$

---

d) prediction interval

$X_f = 10$

$$\text{var} = \sigma^2 \underbrace{\left(1 + \frac{1}{N} + \frac{(\overset{10}{x} - \bar{x})^2}{\sum (x_i - \bar{x})^2}\right)}_{\text{only holds for simple lin. reg. otherwise use matrix formulation}}$$

e) does it look diff?

---

f) discuss avgs / trends etc.

Question 2

Body weight = y (lbs)

exercise time = x_1 (hrs)

calorie intake = x_2 (cal/1000)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

a)

source	df	S.Sq	MS ($= SS/df$)	F-val	p-val
model	$p-1=2$	147538	73769	2862.85	small.
error	$n-p=297$	7653	25.76768		
total	$n-1=299$	155191			

b) compute LS estimates and their SE's.

$$\hat\beta = (X'X)^{-1}X'y = \begin{pmatrix} -65.37564 \\ -12.19020 \\ 22.89048 \end{pmatrix}$$

$$\mathrm{cov}(\hat\beta) = \sigma^2 (X'X)^{-1} = \begin{pmatrix} 53.08 & 14.61 & -49.22 \\ 14.61 & 4.25 & -13.76 \\ -49.22 & -13.76 & 45.81 \end{pmatrix}$$

$\to$ MSE = 25.76768

$H_0: \beta_1 = 0$

$$t = \frac{\beta_1}{\sqrt{Var\beta_1}} = \frac{-12.19020}{\sqrt{4.25}} \sim t_{297} = -5.911956 \qquad t_{\frac{\alpha}{2}=0.05,\ 297=df} = -1.967$$

$\Rightarrow \beta_1$ is significantly diff from 0.

c) 95% CI for body wt when $x_1 = 2$ $x_2 = 1200$

conditional means: $\hat y \sim N(\underbrace{x_f\beta}_{\text{calculate}}, \sigma^2 H) \to \underbrace{\sigma^2\left(X_f (X'X)^{-1} X_f' + I\right)}_{\text{calculate} = A}$

$X_f = (1 \quad 2 \quad \frac{1200}{1000})$

95% CI =

$$X_f\beta \pm 1.96\cdot\sqrt{A}$$

d) $H_0: 20\beta_1 + \beta_2 = 0$

$C = (0 \quad 20 \quad 1) \qquad \theta_0 = 0$

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

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

MSE = in table.

$\sim F(1, 297)$

$\Rightarrow$ matrix calcs etc.

reject if $> F_{1,297}$

$X_{11} - 1.25$
$X_{12} - 1.25$

$\Sigma (X_{i1} - 1.25)$
$= \Sigma X_{i1} - \overset{300}{n}(1.25)$

$= \Sigma (X_{i1} - 1.25) \cdot (X_{i1} - 1.25)$
$= \Sigma (X_i^2 - 2(1.25)X_i + 1.5625)$
$= \underbrace{\Sigma X_i^2}_{687.5} - 2.5 \underbrace{\Sigma X_i}_{375} + 300(1.5625)$

e) center ex. on 1.25 hours and cal on 1500

$1500/1000 = 1.5$

$$\text{new } X = \begin{bmatrix} 1 & X_{i1} - 1.25 & X_{i2} - 1.5 \\ \vdots & & \\ 1 & X_{i300} - 1.25 & X_{i300} - 1.5 \end{bmatrix}$$

$$(X'X) = \begin{bmatrix} 1 & \cdots & 1 \\ X_{i1} - 1.25 & \cdots & X_{i300} - 1.25 \\ X_{i2} - 1.5 & \cdots & X_{300} - 1.5 \end{bmatrix} \begin{bmatrix} 1 & X - 1.25 & X - 1.5 \\ \vdots & \vdots & \vdots \\ 1 & X - 1.25 & X - 1.5 \end{bmatrix}$$

$$= \begin{bmatrix} 300 & \Sigma(X_{i1} - 1.25) & \Sigma(X_2 - 1.5) \\ \Sigma(X_1 - 1.25) & \Sigma(X_{i1} - 1.25)(X_{i1} - 1.25) & \Sigma(X_{i1} - 1.25)(X_{i2} - 1.5) \\ \Sigma(X_2 - 1.5) & \Sigma(X_{i1} - 1.25)(X_{i2} - 1.5) & \Sigma(X_{i2} - 1.5)(X_{i2} - 1.5 \end{bmatrix} = \begin{bmatrix} 300 & \Sigma X_{i1} - 300(1.25) & \Sigma X_{i2} - 300(1.5) \\ '' & \text{above} & \\ '' & & \text{etc.} \end{bmatrix}$$

e) cont.

$(X'X)^{-1}$ via R etc.

$$\underbrace{\begin{bmatrix} 1 & \cdots & 1 \\ X_{i1}-1.25 & \cdots & X_{i1}-1.25 \\ X_{i2}-1.5 & & X_{i2}-1.5 \end{bmatrix}}_{3\times 300} \underbrace{\begin{bmatrix} Y_1 \\ \vdots \\ Y_{300} \end{bmatrix}}_{300\times 1} = \underset{3\times 1}{} = \begin{pmatrix} \Sigma Y_i \\ \Sigma\left(y_i\,(X_i-1.25)\right) \\ \Sigma\,(y_i)(X_i-1.5) \end{pmatrix} = \begin{pmatrix} \Sigma Y_i \\ \Sigma X_{i1}Y_i - \Sigma Y_i\,(1.25) \\ \Sigma X_{i2}Y_i - \Sigma Y_i\,(1.5) \end{pmatrix}$$

then do est as $(X'X)^{-1}(X'Y)$

se's by $\sigma^2(X'X)^{-1} \Rightarrow \sigma^2 = MSE$ (shouldn't Δ w/ centering)

$t = \dfrac{\text{param}}{SE}$

~~Question~~ Question 3

logit(p disease) = $\beta_1 + \beta_1 I(aA) + \beta_2 (AA)$

a) prob of having disease for indiv. w/ aA genotype $X = (1 \ 1 \ 0)$

$$p = \frac{\exp(X\beta)}{1+\exp(X\beta)} = \frac{\exp(0.1001 - 1.1656)}{1 + \exp(0.1001 - 1.1656)} = 0.2563$$

---

b) OR of genotype aa vs. aA?

$\exp(\beta_1)$ = OR for disease if aA, compared to aa

$= \exp(-1.1656) = 0.3117$

95% CI: $\exp(-1.1656 \pm 1.96(0.3415))$

$= (0.1596, 0.6088)$

---

c) IS: H_0: $\beta_1 + \beta_2 = 0$

$\beta_0 + 2\beta_2 = 3$

$\beta_0 + \beta_1 + 3\beta_2 = 3$

testable? $\rightarrow$ to be testable, must be:

C full rank

& estimable

$$C = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 3 \end{pmatrix} \quad \theta_0 = \begin{pmatrix} 0 \\ 3 \\ 3 \end{pmatrix}$$

* assuming X is full rank b/c logistic reg. ran in SAS

C $\Rightarrow$ not full rank! $\Rightarrow$ the hypothesis is not testable.

col1(2)+col2 = col3

d) Same question for:

H_0: $\beta_1 + \beta_2 = 0$

$\beta_0 + \beta_2 = 2$

$\beta_0 + \beta_1 + 2\beta_2 = 5$

$$C = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix} \quad \theta_0 = \begin{pmatrix} 0 \\ 2 \\ 5 \end{pmatrix}$$

C not full rank → not testable.

d) ~~cultivated~~ area of 340 in 1983

predict the level (prediction interval)

$X_f = (1 \quad 1983 \quad 340)$

$\hat{Y}_f = 906.2465$

$\hat{Y}_f = SN(X_f\beta, \underbrace{\overset{MSE}{\sigma^2}(X_f(X'X)^{-1}X_f' + I_{17})}_{var})$

so 95% CI is $X_f\beta \pm 1.96\sqrt{var}$

$$\frac{(x_f \beta_i)}{\sqrt{var}} \sim N(0,1)$$

---

e) H : there is a pattern in the R-P plot: they are not random

I : independence may be violated through data collection / plots may be dependent

L : again, R-P plot is not great

E : OK (our data exist.)

Gauss : pretty linear QQ plot

f) centering data
interaction terms
non-linear effects
categorical predictors

Question 4|

2011 Nov?

data on manioc production in various areas over the years

a) using matrix notation, write an expression for estimating the least sq. reg. coeffs →
estimate production based on year + area

$$\vec{Y}_{17\times1} = \underset{(17\times3)(3\times1)}{\vec{X}\vec{\beta}} + \vec{\varepsilon}_{17\times1}$$

$\vec{y}$ = 17×1 matrix of manioc production values

$\vec{X}$ = 17×3 matrix of an intercept column, year, and cultivated area.
{should center year so 1961 = 0 (yr - 1961)
could/should center area as well. }

$\vec{\beta}$ = 3×1 matrix of coefficients = $\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix}$

$\vec{\varepsilon}$ = 17×1 matrix of random errors

---

b) Yes, the overall regression is significant ($F = 4.617$, $p=0.02885$)
($H_0: \beta_0=\beta_1=\beta_2$ is rejected) * not sure about this *
Neither coefficient is sign. diff from 0 based on their Wald tests.

---

c) yr: $4.2142 \pm t_{1-\frac{0.05}{2}, df=14}(8.3969)$
$= 4.2142 \pm 2.144787(8.3969)$
$= (-13.788, 22.217)$

age $= 0.5647 \pm 2.144(1.2913)$
$\boxed{= (-2.204, 3.333)}$