① elderly (65-90) - diabetic or not; obese or not

a) cross-sectional cohort design

b) p obese = $\frac{250+140}{\text{total}}$ z b/c n>>)

95% CI = $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$

c) RR = $\frac{R \text{ obese}}{R \text{ non}} = \frac{250/390}{260/530}$ ± CI formula $\frac{\text{risk diab | ob}}{\text{risk diab | not}}$

d) yes/no ⇒ RR doesn't/does include 1 in the interval

e) case-control (unmatched)

f)

	O	O^c	T
D	250	260	510
D^c	140	370	510
T			1020

can only do OR

OR = $\frac{ad}{bc} = \frac{250(370)}{(140)(260)}$ ± CI formula

NP (control)

g) matched case-control

pairs that are both obese, both not, etc.

total pairs = 510 (1 diabetic / 1 not)

D:	ob	not	
ob			250
not			260
	140	370	510

h) concordant pairs etc.

i) yes → interaction is sig (p = 0.084). Therefore there is a significant added effect on the odds of obesity if the subject is male.

j) OR male = $\exp(\beta_0 + \beta_{obese} + \beta_{male} + \beta_{int} + \beta_{age})$

$= \exp(\beta_0 + \beta_{obese} + \beta_{age})$

k) $\beta_{age}(5)$ = inc/dec in log odds of obesity for someone 5 yrs older.

estimate two OR as above.

② dose - factor w/ 3 levels

full 2-way model w/ interaction

a)

$$y = \alpha_0 + \underbrace{\alpha_1 X_1}_{=1 \text{ if batch}=2} + \beta_2 X_2 + \beta_3 X_3 + \gamma_2 X_1 X_2 + \gamma_3 X_1 X_3 + e$$

X_2: dose=100; X_3: dose=200; X_1X_2: batch=2/dose=100; X_1X_3: batch=2/dose=200

batch	dose	mean
ref 1	0	α_0
1	100	$\alpha_0 + \beta_2$
1	200	$\alpha_0 + \beta_3$
2	0	$\alpha_0 + \alpha_1$
2	100	$\alpha_0 + \alpha_1 + \beta_2 + \gamma_2$
2	200	$\alpha_0 + \alpha_1 + \beta_3 + \gamma_3$

γ_2 = addle expected change in the cholestrol levels of an individual if they are on 100mg of the second batch of drugs, above and beyond the main effects.

b) MS — F

894.4/1 — model / MSE

9060.2/2 — for each 3

4992.2/2

29745.1/84 = MSE

added in order ⇒ if it was added last, the p-val for batch should be the same in the Wald and ANOVA table tests ($t^2 = F$)

c) now, dose is continuous

$y = X\eta + e$

$y = 90 \times 1$

$X = 90 \times 4$ — into batch dose interaction

$\eta = 4 \times 1$

$e = 90 \times 1$

$a = \text{rank}(c)$

$$F = \frac{((\hat\theta - \theta_0)\, M^{-1}\, (\hat\theta - \theta_0))/a}{MSE}$$

$F \sim 1, 86$

$H_0: \beta_{dose} = \beta_{dose*batch} \Rightarrow \beta_D - \beta_{DB} = 0$

where $\hat\theta = c\hat\beta$

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

d) can think of dose as categorical nested w/in numerical.
- ~~If numeric is correct,~~

then $2\beta_{100} = \beta_{200}$ (the effect of #1 dose should be ½ that of #2 dose

$H_0: 2\beta_{100} - \beta_{200} = 0$

$c = (0 \quad 0 \quad 2 \quad -1 \quad 0)$

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

$\theta_0 = 0$

$\hat{\theta} = (2 * 14.919 - 12.471)$

→ compare models using F-test

$$\frac{SSE(small) - SSE(large)/dfE(small) - dfE(large)}{SSE(large)/dfE(large)}$$

$$\frac{(\hat{\theta}(M^{-1})\hat{\theta})/1}{MSE}$$

$MSE \leftarrow SSE/dfE$

$(y - x\beta)^T(y - x\beta)$

$= y^Ty - (x\beta)^T y - y^T(x\beta) + (x\beta)^T(x\beta)$

$= y^Ty - \beta^T x^T y - y^T x\beta + \beta^T x^T x \beta$

e) $e \sim N(0, 2\sigma^2)$ BATCH 1
$e \sim N(0, \sigma^2)$ BATCH 2

use weighted least squares →

$\sum_{i=1}^{n} w_i(y_i - X_i\eta)^2$ $\quad w=1$ for BATCH 1 $\quad w = 2$ for BATCH 2 $\quad W^T = W$

$y = n \times 1$
$x = n \times p$
$\eta = p \times 1$
$W = n \times n$

$(y - x\eta)^T W(y - x\eta)$

show $\hat{\eta} = (x^TWx)^{-1}(x^TWy)$

$(yW - x\eta W)^T(y - x\eta)$

$(x^TWx)\eta = x^TWy$

$= (yW)^Ty - (x\eta W)^Ty$

$d\eta \int \; \underbrace{(x^TWx)\eta}_{vector} - \underbrace{x^TWy}_{scalar} = 0$

$-(yw)^Txn + (x\eta W)^T(x\eta)$

$= y^Ty + \eta^T(x^TWx)\eta - \eta x^TWy$

$\frac{d}{d\eta}$ $2W^Ty^Tx\eta + \eta^T(xW)^Tx\eta =$

$2W^Ty^Tx + 2(x^T)Wx\eta \neq 0$

$\eta = (x^TWx)^{-1}\overbrace{(Wy^Tx)}^{\text{dimensions don't work}}$

... but close!

$(y-xn)^T W(y-xn)$ norm equn: $(X^TWX)n = X^TWy$

$(y-xn)^T(Wy - Wxn)$ e) cont

$= y^TWy - y^TWxn - (xn)^TWy + (xn)^TWxn$

$- y^TWxn - n^Tx^TWy$

$= y^TWy - 2nx^TWy + n^Tx^TWxn$

$\frac{d}{dn} = -\cancel{2}x^TWy + \cancel{2}x^TWxn = 0$

$x^TWxn = x^TWy$

$\hat{n} = (x^TWx)^{-1}(x^TWy)$ ✓

f) $E(\hat{n}) = (x^TWx)^{-1}x^TW\,E(y) = n$ (y is the R.V.)

$E(y) \to xn$

$Cov(\hat{n}) = Cov((x^TWx)^{-1}x^TWy)$

$= ((x^TWx)^{-1}x^TW)\,Cov(y)\,((x^TWx)^{-1}x^TW)'$

still $\sigma^2 I_n$

$= \sigma^2\{(x^TWx)^{-1}x^TWW\,x\,((x^TWx)^{-1})'\}$ ☆

W?

2014 II, #3

(3) a), i. False. β's are parameters.

ii. False. ε's are random error for each person (based on pred values ↳ those are based off stats ↳ not parameter)

iii. False. Mean $X\beta$

iv. True. A statistic is a R.V.

v. True.

b) SE → $\sqrt{\text{diagonals}}$ of $\hat{\sigma}^2 (X^T X)^{-1}$

int $\sqrt{361.33 \cdot 30.32} =$

age $\sqrt{\text{''} \cdot 0.0134} =$

wt $\sqrt{\text{''} \cdot 0.0013} =$

then t-values are $\beta / SE(\beta)$

c) $\text{corr}(\beta_0, \beta_1) = \dfrac{\text{cov}(\beta_0, \beta_1)}{\sqrt{\text{var}\beta_0 \text{ var}\beta_1}} = \dfrac{\sigma^2 \cdot (X'X)^{-1}_{12}}{SE\beta_0 \cdot SE\beta_1}$

d) $H_0: \beta_0 = \beta_1 = \beta_2$ using GLH approach

$C\beta = \begin{bmatrix} \beta_0 - \beta_1 \\ \beta_0 - \beta_2 \end{bmatrix}$ $C = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$ $\theta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ $\hat{\theta} = \begin{pmatrix} 119.1577 \\ 118.4314 \end{pmatrix}$

$F = \underbrace{(\hat{\theta} - \theta_0)'}_{1\times 2} \underbrace{(M^{-1})}_{2\times 2} \underbrace{(\hat{\theta} - \theta_0)}_{2\times 1} / \hat{\sigma}^2 \; / \; a$ $= 1\times 1$, rank c $a = 2$

$M = \underset{2\times 3}{C} \; \underset{3\times 3}{(X'X)^{-1}} \; \underset{3\times 2}{C'} = 2\times 2$

$= 0.6386$

$F \sim 2,$ ~~5~~ .

e) β_0 = expected BP when age = wt = 0 (not meaningful. center age/wt so the intercept is the value @ both centering values

β_1 / β_2 = inc in y w/ a 1 unit increase in age/wt

f) Dichotomize y

$\tilde{y}_i = 1$ if $y_i > 120$

$p_i = P(y_i > 120)$

$\tilde{y}_i \sim \text{Bern}(p_i)$

$E\tilde{y}_i = p_i$

$\text{Var}\,\tilde{y}_i = p_i(1-p_i)$

g) Estimate OR relating BP > 120 and wt > 132, give 95% CI

$\Rightarrow$ make a 2x2 table then OR formula

	BP>120	BP≤120
wt>132	||||	|
wt≤132	|	|||

or

$$\frac{\left\{\dfrac{P(BP>120)}{1-P(BP>120)}\right\}}{\left\{\dfrac{P(wt>132)}{1-P(wt>132)}\right\}}$$

③ $n = 8$

1a) i: wrong ⇒ $\hat{\beta}$ are stats; β parameters

ii:

iii: wrong ⇒ $y \sim N(X\beta, \sigma^2 I_{8\times 8})$

iv: false, its a statistic

v: true

b) → same as practice exam

do on 6/5/2016 morning

---

④ $n=100$. random assignment to A/B (50 each), 2 cycles

A

	② +	② −
① +	48	2
① −	7	3

→ #'s are wrong → prob 38. would ask

a) test whether toxicity in cycle 1 ⊥ cycle 2 toxicity

chi-square test of indep.

b) assump. → expected counts >5 (yes/no based on that)

c) $p_1 = 10/60$
$p_2 = 5/60$ test $p_1 = p_2$ w/ appropriate test. (formula)

d) odds toxic in cycle 1 $= \frac{(15/50)}{(35/50)} = \frac{15}{35}$

cycle 2 $= \frac{(10/50)}{(40/50)} = \frac{10}{40}$

$$OR_{1:2} \frac{\left(\frac{15}{35}\right)}{\left(\frac{10}{40}\right)} = \frac{15}{35} \cdot \frac{40}{10}$$

e) Data only from cycle 1 used in logistic reg.

$$\text{logit } P(\text{toxicity}) = \beta_1 + \beta_2 \, \text{drug}(B)_{100\times 1}$$

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

10 toxic on drug A

n = 60 " "

$$y = 100\times 1 = \begin{pmatrix} 1_{10} \\ 0_{50} \\ 1_{15} \\ 0_{35} \end{pmatrix} \qquad X = 100\times 2 = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ \vdots & 0_{50} \\ & 1 \\ & \vdots \\ 1 & 1_{50} \end{pmatrix} \Big\} \text{ 50 on drug B}$$

fine w/ appropriate n (100 or 110?)

f) This approach is worthwhile, but needs more work.

how is toxicity being measured?

either cycle toxicity, or BOTH?

think about an interaction term, the cycle could be related to the drugs toxicity

g) p toxicity should be equal

p toxicity drug A = p Toxicity drug B

Question 2

Batch → 2 options (1/2) Ref = 1ST

Dose → 3 options (0/1/2) Ref = 0

$$Y = \alpha_0 + \alpha_1 I(\text{batch } 2) + \beta_2 I(\text{dose}=100) + \beta_3 I(\text{dose}=200)$$

$$+ \gamma_2 I(\text{batch}2) I(\text{dose}=100) + \gamma_3 I(\text{batch } 3) I(\text{dose}=200) + e$$

a)

Batch	Dose	Mean
1	0	α_0
1	100	$\alpha_0 + \beta_2$
1	200	$\alpha_0 + \beta_3$
2	0	$\alpha_0 + \alpha_1$
2	100	$\alpha_0 + \alpha_1 + \beta_2 + \gamma_2$
2	200	$\alpha_0 + \alpha_1 + \beta_3 + \gamma_3$

γ_2 = additional effect on cholestrol level when an individual receives 200mg from batch 2

b)

source	df	SS	MS	F	p
batch	1	894.4	894.4	2.52	0..115
dose	2	9060.2	4530.1	12.79	$1.41e^{-05}$
batch*dose	2	4992.2	2496.1	7.048	0.0014
error	84	29745.1	354.1083		
total	89				

added-in-order. If they were added-last, the F-value for batch should equal t^2 for batch, and they are not equal.

c) Fit dose as numeric

$y = X\eta + e \qquad y = 90\times 1 \qquad X = 90\times 4 \quad \eta = 4\times 1 \quad e = 90\times 1$

$\eta =$ (int batch dose dose*batch)

$H_0: \beta_{dose} = \beta_{dose*batch}$

$c = (0 \quad 0 \quad 1 \; -1) \qquad \theta_0 = 0$

$F \sim 1, 86$ using the appropriate F-calcs.

---

d) Compare dose as numeric vs. dose categorical

$\Rightarrow H_0: 2\beta_{100} = \beta_{200}$

$c = (0 \; 0 \; 2 \; -1 \; 0 \; 0) \Rightarrow$ don't have $(X'X)^{-1}$ to do this

$\Rightarrow$ view numeric/interval as 'nested' w/in categorical model

use F to compare nested models

---

e) $(y - X\eta)^T W (y - X\eta)$ show $\hat{\eta} = (X'WX)^{-1}(X'Wy)$

$(y - X\eta)'(Wy - WX\eta)$

$W = \begin{pmatrix} \ddots & 0 \\ 0 & \end{pmatrix}_{n\times n}$

$= y'Wy - (X\eta)'Wy - y'WX\eta - (X\eta)'WX\eta$

$= y'Wy - \underbrace{\eta'X'Wy) + y'WX\eta}_{\text{both are scalars so} =} - \eta'X'WX\eta$

$(1\times 4)(4\times 90)(90\times 90)(90\times 1)$

$(1\times 90)(90\times 90)(90\times 4)(4\times 1)$

$= y'Wy - 2\eta'X'Wy - \eta'X'WX\eta$

$\frac{d}{d\eta} = -2X'Wy - 2X'WX\eta = 0$

$X'WX\eta = X'Wy$

$\hat{\eta} = (X'WX)^{-1}X'Wy$ ☺

$E\hat{\eta} = E((X'WX)^{-1}X'Wy)$

$= (X'WX)^{-1}X'WEy$

$= (X'WX)^{-1}X'WX\eta = \eta$

$(W'W) = W$

$\text{cov}\,\hat{\eta} = \text{cov}((X'WX)^{-1}X'Wy)$

$= (X'WX)^{-1}X'W\,\text{cov}(y)((X'WX)^{-1}X'W)'$

$= (\sigma I_n)(X'WX)^{-1}X'WW'X(X'WX)^{-1}$

$= \sigma I_n (X'WX)^{-1}$