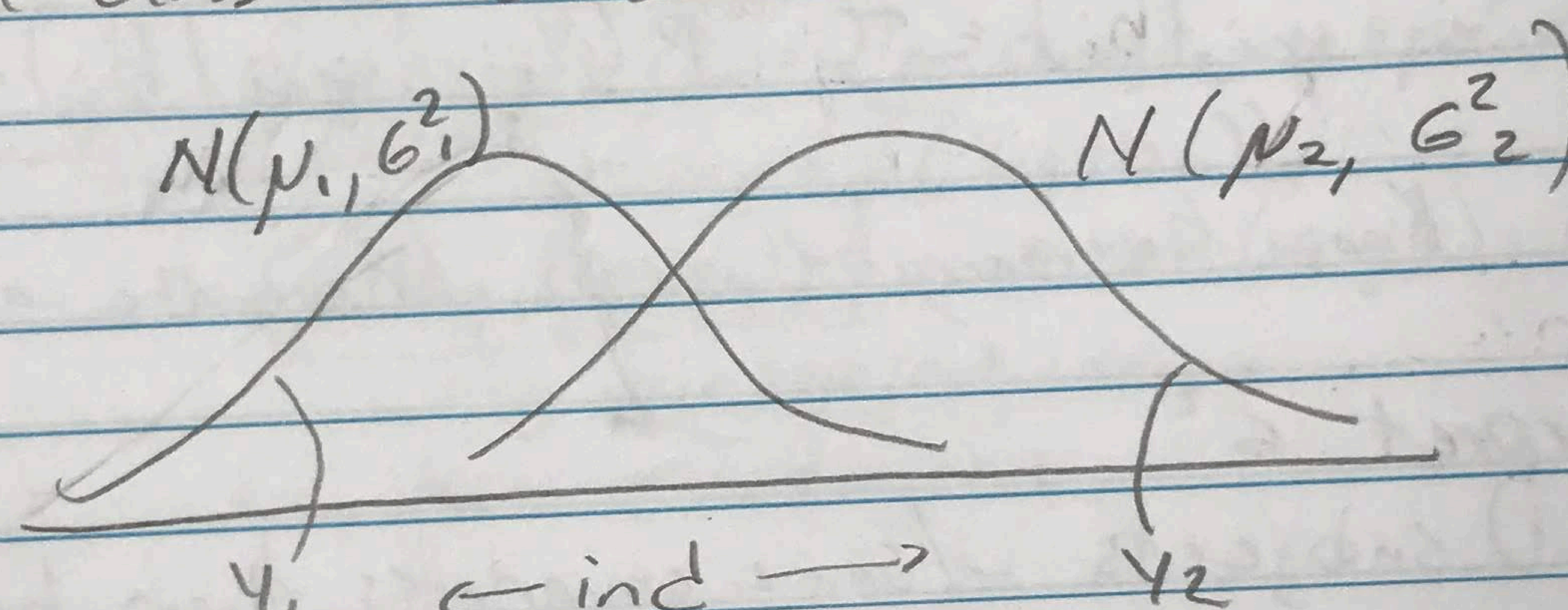
Dec 2nd Class

- Don't just copy SAS output
  - point est and std errors when fit model (don't present other stuff)
  - 3 or 4 digits is enough
- Don't say "I used SAS"
- Plot, properly labelled

- Final Exam
  - short homework 50% ↗ Wednesday - Friday final exam → 2-3 hrs - 80-90% is fine
  - in class 50% of final exam



$$Y_1 - Y_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$E[e^{Y_1 - Y_2}] = e^{\mu_1 - \mu_2 + \frac{\sigma_1^2 + \sigma_2^2}{2}}$$

↑ moment generating function

$e^{Y_1 - Y_2} \sim$ log-normal

R_i = diagonal $(\phi h(\mu_{ii})$ . . . . $)$ ← this is the case for mathematical convenience

Can't use GEE b/c mean structure is complicated

Mixed 2

$R : cov(Y|b)$ $\quad$ Y: all subjects stacked
b: all random effect stacked

Bernoulli
- common in applications
- demonstrate difficulty when bring in link function

b/c of independence

$$P(Y_i = y_i | b_i) = \prod_{j=1}^{n_i} P(Y_{ij} = y_{ij} | b_i) = \prod_{j=1}^{n_i} \nu_{ij}^{y_{ij}} (1-\nu_{ij})^{1-y_{ij}}$$

when I know the marginal mean, I know the variance

Interpret σ^2
1) subjects w/ same covariates then heterogeneity in subject specific risk between subjects is σ^2

BIOS 663

$$\text{logit } P(Y_i = 1) = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + \cancel{\varepsilon_i}$$ ← marginal, not conditional
$= \log E[Y_{ij}]$

$$\text{logit } P(Y_i = 1 | b_i) = b_i + x_i^T \underset{\sim}{\beta}$$

$$= \text{logit } E[Y_i] = \frac{1}{1 + \exp\{-x_i^T \beta\}} + \varepsilon_i = Y_i$$
$E[\varepsilon_i] = 0$

$Y_{i1} | b_i \perp Y_{iu} | b_i$

conditioned on random effect, the observations are independent

by double expectation

$E[Y_{ij}] = E[\mu_{ij}]$ ← simulate the μ_{ij}

ICC

books give $\dfrac{\sigma^2}{\sigma^2 + \underbrace{\pi^2/3}_{\text{within}}}$ → based on the latent variables only b_i and a latent logistic regression variable

→ the actual response is not involved at all

$var(Y_{ij}) = \mu_{ij}(1-\mu_{ij})$ ← Bern

$= E[\underbrace{var(Y_{ij}|b_i)}] \Big\} \leftarrow$ Double expectation

Still Bernoulli, mean is μ_{ij}

$+ var(E(Y_{ij}|b_i))$

$= \underbrace{E[\mu_{ij}(1-\mu_{ij})]}_{\text{avg w/in subject variance}} + var(\mu_{ij})$

= within + between

$$ICC: \frac{\text{between}}{\text{total}} = \frac{var(\mu_{ij})}{\mu_{ij}(1-\mu_{ij})}$$

Mixed one → ICC is correlation ??? in linear

ICC → not a correlation in Bernoulli, Bi, etc.

__Numerical Example:__

only random intercept

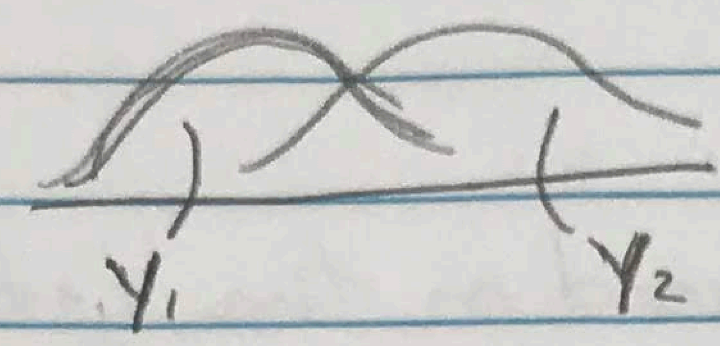
- in this example, treating values as true parameters, not estimates

Picked Y_{i1}, Y_{iu} at random

Odds of RI has gone down + the ratio is $\exp\{-0.1756\}$

- log of ratio of odds of RI of a child at a given age relative to a year younger (negative: risk of infection goes down)

"mother smoking"
↳ have to consider random pairs of subjects



$Y_1 \perp Y_2$

$$\text{logit } \nu_{Aj} - \text{logit } \nu_{Bj} = (b_A - b_B) + \beta_2 (MS_A - MS_B)$$

same occassion
so age cancels out

$\underbrace{\text{log odds of 2 subject-specific}}_{\star\star}$ odds of RI at the same occassion

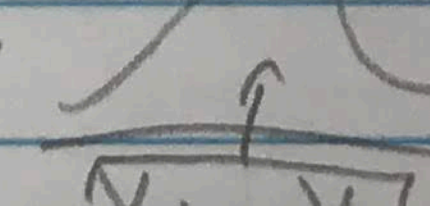
0.3986 is the $\underbrace{\text{expected value}}_{\star \text{ the avg value}}$ ...

$\overbrace{\text{on avg,}}^{\star}$ the subjects who mother is smoker have odds exp(.3986) of disease

Dist of contrast $N(\beta_2, 2\sigma^2)$ → from earlier
$Y_1 - Y_2 = N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

55% of pairs: subject w/mother smoker will have higher RI than subject w/mother notsmoker

We're thinking about random pairs of subjects and their distribution!
of their difference →



$Y_1 - Y_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

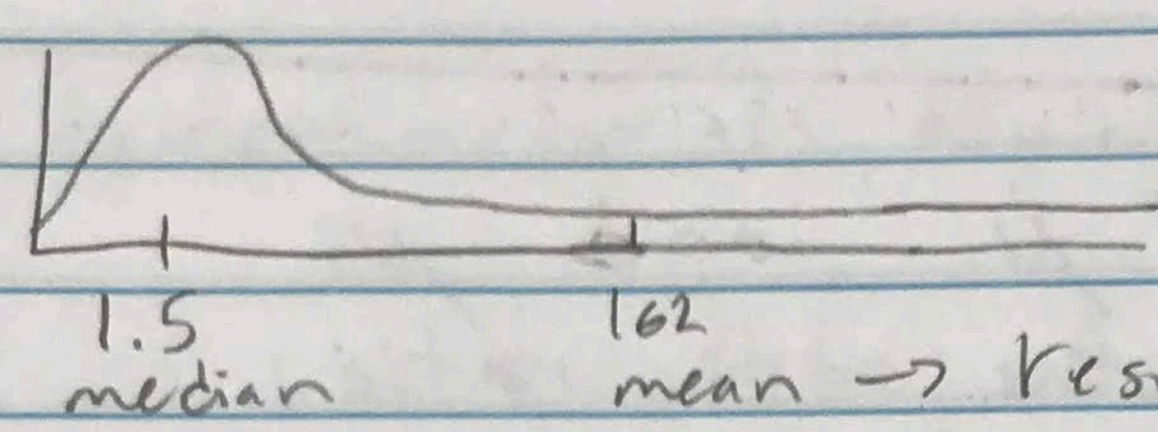
These contrasts are on the scale of log odds

$$E\left[\exp\{b_A - b_B + 0.39\}\right]$$

$$= e^{\underbrace{\mu_1+\mu_2}_{\downarrow} + \underbrace{\frac{\sigma_1^2+\sigma_2^2}{2}}}$$

$$= e^{\beta_2 + \frac{2\sigma^2}{2}} \qquad \approx 162$$

log-normal has very high skewness



→ result of extremely long tail

when you transform to logs → any monotone transformation → the quantiles transform like that but expectations don't

To interpret mother smoking → have to form contrast between 2 subjects
To interpret age → you have 2 choices: contrast w/in subject OR pick 2 children w/in same mother smoking group at different ages. Can do these comparisons + calibrate age effect w/ the heterogeneity between subjects

Give median, quantiles, in addition to mean
↳ long tail gives you large mean

If these were estimates, insert into every sentence, the quantiles "are estimated to be"

Take Home Final

→ main program → use program + output + that's it

→ second program → modify small things and re-run

- based on dataset from Ch. 13. clinic × treatment × time had difficulty w/ interactions

---

Marginal mean: μ_1 and μ_2 you can form the odds ratio:

$$\frac{\mu_1(1-\mu_1)}{\mu_2(1-\mu_2)}$$

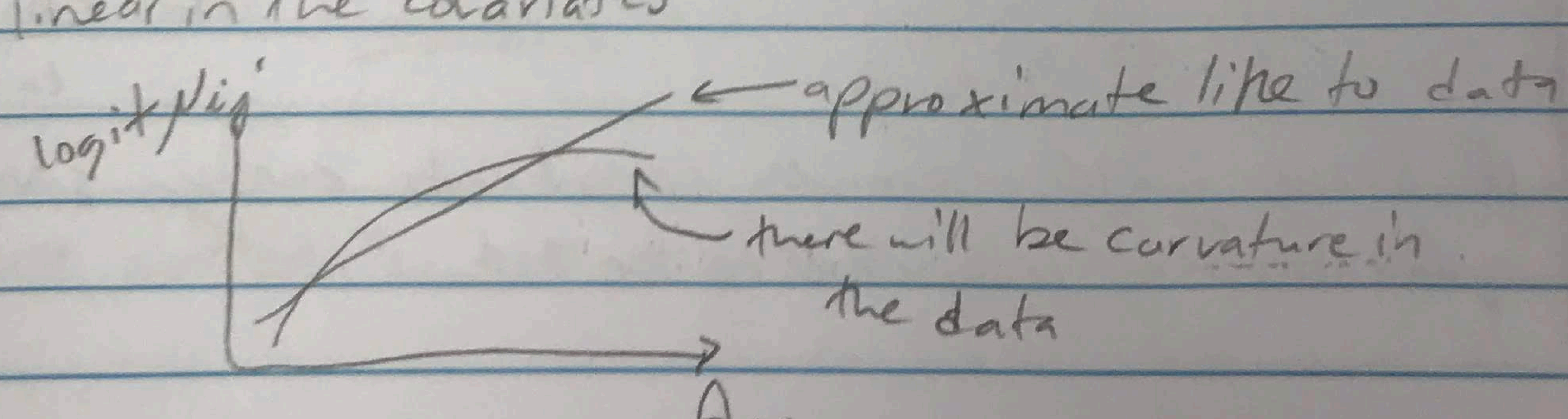
There isn't an odds ratio marginally. There is an approximation

Cluster-level variables like mother-smoking → have to compare different subjects

---

How does the marginal model compare to mixed model? Does one model imply the other?

log μ_{ij} → not linear. complicated function of σ^2 and covariates. Then take logit + won't be linear in the covariates



not exact

approximate line

$$\text{logit } \mu_{ij} \approx \frac{x_{ij}^T \beta}{\sqrt{1+t^2\sigma^2}}$$

rely on similarity between normal + logistic dist.

$\pi^2/3 \rightarrow$ variance of logistic dist.

The denominator is always larger than 1.

slope of covariate in absolute value is smaller than β. The approximate marginal effect is smaller than conditional $\rightarrow$ attenuation

σ is a guage of how much smaller the approx marginal effect is than the conditional

(0.1614 , 0.1467 , 0.132 , 0.12)

marginal means

$\hookrightarrow$ can take ~~them~~ logits of them and do the regression and get

$$\text{logit } \mu_{ij} = -1.876 - 0.1149\,(\text{Age} - 9)$$

$$E[\underset{\sim}{Y_i} \mid \underset{\sim}{b_i}] = X_i \underset{\sim}{\beta} + Z_i \underset{\sim}{b_i}$$

$$E[\underset{\sim}{Y_i}] = X_i \underset{\sim}{\beta}$$

b_i has mean 0 : If b_i doesn't have mean 0 $\rightarrow$ has mean δ

$$X_i\beta + Z_i b_i = X_i\beta + Z_i(b_i - \delta) + Z_i \underset{\sim}{\delta} = X_i \beta^* + Z_i \underset{\sim}{b_i^*}$$

In linear mixed model do predictions:

BLUP : $\hat{E}[\underset{\sim}{b_i} \mid \underset{\sim}{Y_i}]$

Too much work for general mixed models.

Mode($\underset{\sim}{b_i} \mid \underset{\sim}{Y_i}$)

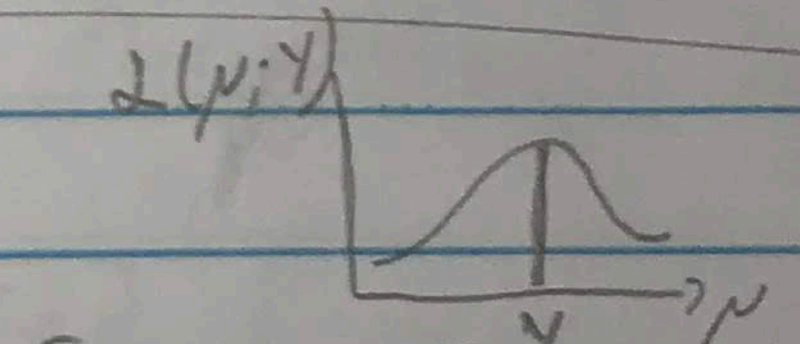
↑ for predictions

Posterior mode (borrowed from Bayesian language, but not actually Baysian)

Rule: If you have model w/ $q+1$ random effects (G is $(q+1) \times (q+1)$), want to compare smaller model w/ this model : as before 50-50 mixture of χ^2 dist.'s

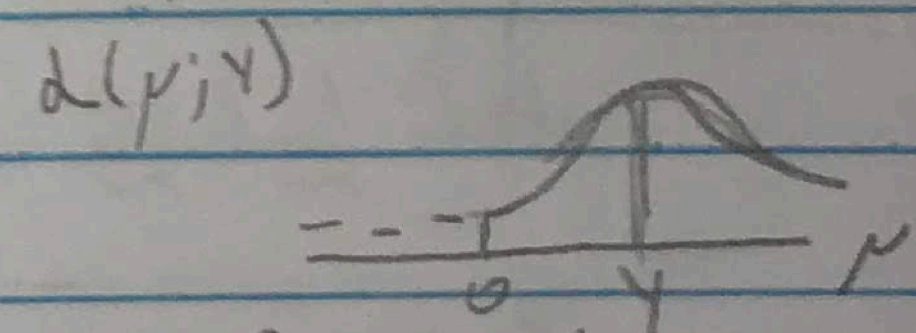
If comparing random-intercept only model w/ no random effect, then q is 0, so χ^2_0, p-value is 0. So the p-value is χ^2_1 divided by 2.

$Y \sim N(\mu, 1)$ ← one obs

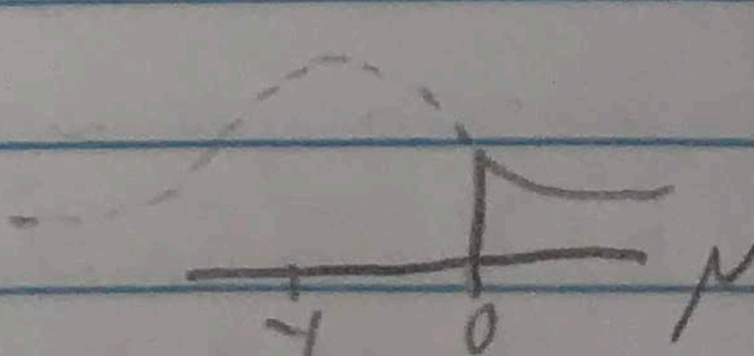


$\hat{\mu} = Y$

$Y \sim N(\mu, 1)$, $\mu > 0$



$$\hat{\mu} = \begin{cases} \text{if } Y > 0, & \hat{\mu} = Y \\ \text{if } Y < 0, & \hat{\mu} = 0 \end{cases}$$



$P(\hat{\mu} = 0) = P(Y < 0) = P(Y - \mu < -\mu) = \Phi(-\mu)$ mixture probability depends on the unknown variable

Good paper for this: Cui & Li
↳ Proper paper, Doctoral Level

6city06.sas

proc nlmixed
↳ very different from other procs we've seen

"random u ~normal(0, sigmasq) subject=id"
↳ u is RV $\sim N(0, \sigma^2)$. subject specific
↳ this is b_i

parms → you're supposed to supply initial values
↳ fit a marginal model using genmod

"replicate count"
↳ dataset was summarized. (summed) use "replicate" and "count" is the name of the variable
→ If 2 subjects have the same response pattern + covariates → then don't need to compute individually. do it once + multiply by how many there are
→ speeds it up

Lots of computations going on → for each obs have 4 things → then empirically integrate?

(?) What's happening in this process?

1 core, 1 sec $\rightarrow 10^9$ multiplications 3.14159×2.71828

predict u out = U
↳ predictions

3: GEE 1+ MStage
↳ how he got the ratios

see proc nlmixed in take-home exam

Dec 5 Review

133

$g(\underset{\sim}{\nu_i}) = X_i \underset{\sim}{\beta} + Z_i \underset{\sim}{b_i}$

$C\underset{\sim}{\beta} = \underset{\sim}{0}$

$\rightarrow (C\hat{\underset{\sim}{\beta}})^T (C V C^T)^{-1} (C\hat{\underset{\sim}{\beta}})$
$\widehat{cov}(\hat{\beta})$

hyp testing is not a procedure for model selection
(Akaike - AIC - original paper)

Can't test $H_0: b_{11} = 0$ Random variable
hyp test is a statement about the parameters

$Y \sim Bin(2, \theta)$ $H_0: \theta = \frac{1}{2}$ ↳ use Y to test this, but not testing Y itself

4a) $\cancel{\mu_i = E[Y_i] = EE[Y_i \mid b_i] = E[\nu_{ij}] = -1(0.5) + 1(0.5) = 0} \to ??$

b) $\log \nu_{i1} = b_i + 2$

$\log \nu_{i2} = b_i + 4$

$var(Y_{i1} \mid b_i) = 2\nu_{i1}$

$var(Y_{i2} \mid b_i) = 2\nu_{i2}$

Value	Prob
e^{-1}	$\frac{1}{2}$
e^{1}	$\frac{1}{2}$

$\nu_{i1} = e^{b_i} \cdot e^2$

$\nu_{i2} = e^{b_i} \cdot e^4$

$\mu_{i1} = E[\nu_{i1}] = e^2 \cdot \frac{1}{2}(e + e^{-1})$

$Var(Y_{i1}) = E(var(Y_{i1} \mid b_i)) + var(E(Y_{ij} \mid b_i))$

$= 2\mu_{i1} + e^4 + \ldots ??$

$E[(e^{b_i})] = \frac{e^{-2} + e^2}{2}$

$\nu_{i1} = e^2 e^{b_i}$

$\nu_{i2} = e^4 e^{b_i}$

$cov(Y_{i1}, Y_{i2}) = E(\underset{\text{ind}}{cov(Y_{i1}, Y_{i2} \mid b_i)}) + cov(E(Y_{i1}) E(Y_{i2}))$

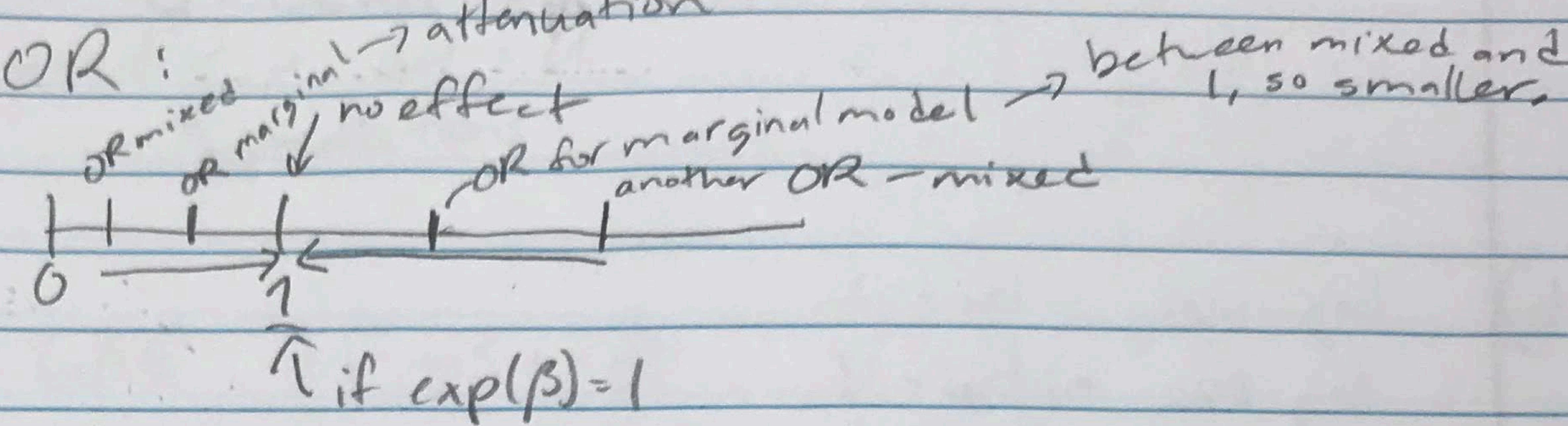
$= E[0] + cov(e^2 e^{b_i}, e^4 e^{b_i})$

$= e^6 \cdot \underbrace{cov(e^{b_i}, e^{b_i})}_{= var(e^{b_i})}$

Attenuation effect: approximation.

4d) $\alpha = 2$ (?) ⟶ no attenuation
should be

OR :



$$\nu_{ij} = \underbrace{\exp\{b_i + \underset{\sim}{x}_{ij}^T \underset{\sim}{\beta}\}}$$ ← log linear mixed model w/ random int only

Double Expectation → $\mu_{ij} = E[e^{b_i} \cdot e^{x_{ij}^T \beta}]$

$= \exp\{\underset{\sim}{x}_{ij}^T \underset{\sim}{\beta}\} \cdot E[e^{b_i}]$

$E[e^{b_i}] = \delta$

$\mu_{ij} = e^{x_{ij}^T \beta} \cdot \delta$

$\log \mu_{ij} = \underset{\sim}{x}_{ij}^T \beta + \log \delta$

new intercept: log $\overbrace{\text{intercept}}^{\text{from } x_{ij}}$ + δ

but slope from x_{ij} is the same

$\nu_{ij} = \exp\{\underset{\sim}{x}_{ij}^T \underset{\sim}{\beta} + \underset{\sim}{z}_{ij}^T \underset{\sim}{b}_i\}$

↑ age is in X, but not in Z,

so exp(age

any covariate in X not in Z, marginal coefficient exactly the same (not true if it is in Z → might not even be linear)

as conditional coefficient →

when b_i's normal, $E[e^{\underset{\sim}{z}_i^T \underset{\sim}{b}_i}]$ ← easy to get b/c it is the moment generating function

- Any power link function → won't same coefficient

reciprocal link function – canonical for gamma dist

BLUP (Pg 212)

This is true for linear combination of β, not each $\hat{\beta}$

$\hat{\beta}i = \qquad + (I_q - W_i) A_i \hat{\beta}$

if : $\hat{\beta}_i = \lambda + (1-\lambda)\hat{\beta}$ if $\lambda \in (0,1)$, then each component of β_i is an avg of these two

same for formula on pg 210

Final

1) Mixture: χ^2_0 χ^2_1 pvalue 0

3) a) Question: compute the $\underbrace{\text{BLUP}}$ (not the empirical BLUP)

the actual conditional Expectation → $E[b_i | Y_i]$

↳ bivariate normal → usual formula

↳ $\hat{E}[b_i | Y_i]$

Pg 211 → BLUP has $\hat{\beta}$ in it → This is wrong there is no hat in this. Same 1st formular Pg 210

formula for cov is questionable

mixed2.pdf

how to calculate $\mathrm{var}(\nu_{ij})$

?

how?

$$\nu_{ij} = \frac{1}{1 + e^{-(b_i + \underset{\sim}{x}_i{}^T \underset{\sim}{\beta})}}$$

$E[\nu_{ij}]$

$E[\nu_{ij}^2]$ $\leftarrow$ simulate + use sample mean

Or use numerical integration

$$\mathrm{logit}\, \nu_{ij} = b_i + x_{ij}^T \beta$$

$$\log\left(\frac{\nu_{ij}}{1 - \nu_{ij}}\right) = b_i + x_{ij}^T \beta$$

$$\nu_{ij}/(1-\nu_{ij}) = \exp(b_i + x_{ij}^T \beta)$$

$$\nu_{ij} = \exp(\quad) - \nu_{ij} \cdot \exp(\quad)$$