MATH 4939
Sample Final Exam Questions
March 20, 2018

1. State and explain the principle of marginality. Discuss how it is an instance of a principle of invariance.

2. [10] Consider the following models where Y, X, Z1, Z2, Z3 are numerical variables. All but one of these models will produce the same regression coefficient for X Xr but they will produce different standard errors. Identify the model that produces a different coefficient. Rank the others where you can according to the se of the estimated coefficient stating which would be equal if any (assume a very large n and ignore the effect of slight differences in degrees of freedom for the error term). Explain your reasoning.
    a. Y ~ X + Z1 + Z2 + Z3
    b. Yr ~ Xr where Yr is the residual of Y regressed on Z1, Z2, Z3 and Xr is the same for X
    c. Y ~ Xr
    d. Y ~ X + Xh where Xh is the predictor of X in the regression of X on Z1, Z2 and Z3
    e. Y ~ X
    f. Y ~ X + Xh + Z1

propensity score

$\hat{\beta}_X$ is same as a
SSE bigger - not as big as (c)

3. Let Y and X be a numerical variables and let G be a factor. Consider the following models. All but one of these models will produce the same regression coefficient for X or Xr but they will produce different standard errors. Identify the model that produces a different coefficient. Rank the others where you can according to the se of the estimated coefficient stating which would be equal if any (assume a very large n and ignore the effect of slight differences in degrees of freedom for the error term). Explain your reasoning.
    a. Y ~ X + G
    b. Y ~ X
    c. Yr ~ Xr where Yr is the residual of Y regressed on G and Xr is the same for X
    d. Y ~ Xr
    e. Y ~ X + Xh where Xh is the predictor of X in the regression of X on G
    f. Y ~ X + Xh + Zg where Zg is a 'G-level' numerical variable, i.e. it has the same value for all observations within any value of G.

$Y \sim X_r$

Lemma 1: $\hat\beta(Y|X, Z_1, Z_2, Z_3) = \hat\beta(Y - \hat Y_2 | X - \hat X_2)$

of or adjust for

$SSE(Y|X, Z_1, Z_2, Z_3) = SSE(Y - \hat Y_2 | X - \hat X_2)$

AVP theorem

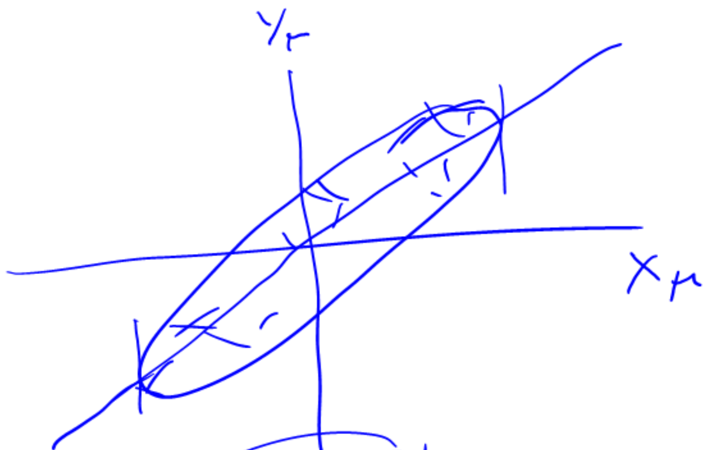
FWL theorem

Lemma 2: $\delta \perp X \quad \delta \perp 1$

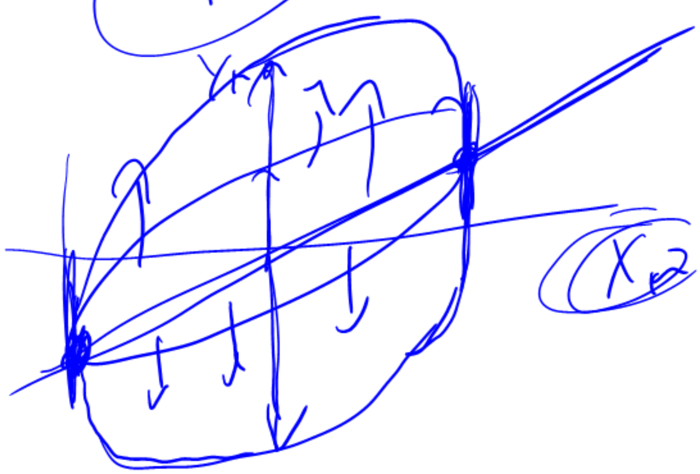
$\hat\beta(Y + \delta | X) = \hat\beta(Y|X)$

$SSE(Y + \delta | X) = SSE(Y|X) + \|\delta\|^2$

$\delta \perp X_r$



$Y_r + \delta$



Pf: $(X'X)^{-1}X'(Y+\delta)$

$= (X'X)^{-1}X'Y$ because $X'\delta = 0$

SSE part exercise

---

b) $Y_r \sim X_r$

c) $Y \sim X_r$

$Y = Y_r + \underbrace{(Y - Y_r)}$

$Y = \hat{Y}_2 + Y_r$

$Y - Y_r = \hat{Y}_2$

$X_r = \underline{X - \hat{X}_2}$ i.e resid of reg. on Z

$\therefore X_r \perp Z$

and $\hat{Y}_2 \in \mathcal{L}(Z)$

$\therefore \hat{Y}_2 \perp X_r$

$\hat{\beta}(Y|X_r) = \hat{\beta}(Y_r|X_r)$

$SSE(Y|X_r) = SSE(Y_r|X_2) + \|\hat{Y}_2\|^2$

$\qquad\qquad\qquad \| $

$\qquad\qquad SSE(Y|X, Z)$

$\underline{Y \sim X + X_n}$ $\qquad$ X_n pred of X on Z

Donald Rubin

AVP? Y axis $Y - \hat{Y}_{X_n}$

X axis $X - \hat{X}_{X_n} = X - \hat{X}_2$

$$Y = \underbrace{Y - \hat{Y}_{X_n}}_{} + \underbrace{\hat{Y}_{X_n}}_{\in \mathcal{L}(X_n) \subset \mathcal{L}(Z)}$$

$\perp \mathcal{L}(X - \hat{X}_2)$

$$\hat{\beta}(Y \mid X, X_n)$$
$$= \hat{\beta}(Y - \hat{Y}_2 \mid X - X_n) = \hat{\beta}(Y \mid X, Z_1, Z_2, Z_3)$$

$$SSE(a) \le SSE(d) \le SSE(c)$$

(d) $Y \sim X + X_n$

(f) $Y \sim X + X_n + Z_1$

doubly robust propensity score

$$\hat{\beta}(Y \mid X, X_n, Z_1) = \hat{\beta}(Y \mid X, Z, \ldots Z)$$

$$SSE(a) \le SSE(f) \le SSE(d) \le SSE(c)$$
$$\overset{\shortparallel}{SSE(b)}$$

Consider X_n is contextual variable in MXM where Z is id.

#3 G = id variable in MXM

X_n is cvar(X, id)

4. Longitudinal data analysis with mixed models: Consider a mixed model with random intercept and slope with respect to time, T. Suppose that the G matrix is $\begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix}$.

   a. Find a the value of T for which the variance of Y is minimized.
   b. Show that recentering T on this value (if known) turns the G matrix into one with only two free parameters.
   c. Explain why recentering and rescaling T can thus help with convergence problems.

5. Consider the following output:

```
> head(hs)
  school mathach    ses    Sex Minority Size   Sector PRACAD DISCLIM
1   1317  12.862  0.882 Female       No  455 Catholic   0.95  -1.694
2   1317   8.961  0.932 Female      Yes  455 Catholic   0.95  -1.694
3   1317   4.756 -0.158 Female      Yes  455 Catholic   0.95  -1.694
4   1317  21.405  0.362 Female      Yes  455 Catholic   0.95  -1.694
5   1317  20.748  1.372 Female       No  455 Catholic   0.95  -1.694
6   1317  18.362  0.132 Female      Yes  455 Catholic   0.95  -1.694
> fit <- lme( mathach ~  ses * cvar(ses,school), hs,
+             random = ~ 1 + ses|school)
> summary(fit)
Linear mixed-effects model fit by REML
 Data: hs
       AIC      BIC    logLik
  12846.85 12891.54 -6415.423

Random effects:
 Formula: ~1 + ses | school
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 1.6293867 (Intr)
ses         0.6614903 -0.469
Residual    6.1109156

Fixed effects: mathach ~ ses * cvar(ses, school)
                          Value Std.Error   DF  t-value p-value
(Intercept)           12.681917 0.3054760 1935 41.51526  0.0000
ses                    2.243374 0.2416545 1935  9.28339  0.0000
cvar(ses, school)      3.687892 0.7699000   38  4.79009  0.0000
ses:cvar(ses, school)  0.873953 0.5771829 1935  1.51417  0.1301
 Correlation:
                      (Intr) ses    cv(,s)
ses                   -0.188
cvar(ses, school)      0.022 -0.261
ses:cvar(ses, school) -0.258  0.065  0.014
```

```
Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-3.2291287 -0.7433282  0.0306118  0.7770370  2.6906899

Number of Observations: 1977
Number of Groups: 40
```

a. Sketch the estimated response function for a school with mean ses of 0 and for a school with a mean ses of 1. Assume that the range of ses is from -2 to 2.
b. Show clearly where is each of the linear regression coefficients estimated in the the model are reflected on the graph.
c. For what value of ses is the variance of mathach estimated to be minimized.

6. Suppose you are studying how some measure of health is related to weight. You are looking at a regression of health on height and weight but you observe that what you are really interested in is the relationship between health and excess weight relative to height.What happens if compute the residuals of weight on height and replace weight in the model with this new variable? Compare the results you would get by doing this with 1)the results you would get by using a multiple regression of health on height and weight; or using a regression of health on height and 'excess weight'? How will these methods vary in the estimation of the effect of 'weight'?

7. You are studying observational data on the relationship between Health and Coffee (measured in grams of caffeine consumed per day). Suppose you want to control for a possible confounding factor 'Stress'. In this kind of study it is more important to make sure that you measure coffee consumption accurately than it is to make sure that you measure 'stress' accurately? What are the consequences of measurement error in Coffee? What are the consequences of measurement error in Stress? Which consequences are more consequential?

8. A survey of Canadian families yielded average 'equity' (i.e. total owned in real estate, bonds, stocks, etc. minus total owed) of $48,000. Aggregate government data of the total equity in the Canadian population shows that this figure must be much larger, in fact more than three times as large. This shows that respondents much tend to dramatically underreport their equity.

9. Discuss situations when it would be important to include a variable that is not significant. When would it be important to exclude a variable that is highly significant? When are fitting criteria (e.g. AIC) suitable for model selection and when are they not?

10. In a regression model with two predictors X1 and X2, and an interaction term between the two predictors, we know that it is dangerous to interpret the `main' effects of X1 and

X2 when the model includes an interaction term but is it safe to do so provided the interaction term is not significant. Discuss in a way a client would understand.

11. Discuss the relationship between the concept of interaction and Simpson's Paradox. How could the idea behind Simpson's Paradox be applied to a situation in which there is interaction. What would it say about conditional relationships?

12. You need to impute a mid-term grade for a student who missed the mid-term with a valid excuse. You plan to use the grade on the final exam. Discuss the relative consequences of using 1) the regression equation of the mid-term on the final, 2) the raw grade on the final, 3) using the z-score on the final to compute the score on the mid-term with the same z-score, and 4) the regression equation of the final on the mid-term to compute the mid-term mark that would have predicted the mark on the final obtained by the student. If you had to choose one of these four, which would you use?

13. What are the differences between Lord's Paradox, Simpson's Paradox and Suppression Effects? What are the similarities?

14. A researcher studying a schizophrenia medication in a clinical population discovers that the dosage is positively correlated with strength of symptoms. She is about to begin a recall because the drug appears to be making patients worse, when it occurs to her that perhaps there is another variable in play which restores the good name of her drug. What might that variable be? How could this variable have this effect (sketch!) and would you describe it as a 'confounding' or 'mediating' variable?

15. A client comes to a consulting session with a study looking at depression as an outcome. The depression measure is continuous, but the hypothesis that there was a difference between 2 groups on depression didn't pan out, because the t-test was not significant. Their supervisor has instructed them to score the depression items such that they have 3 levels - not depressed, somewhat depressed, depressed. The supervisor suggests that this method of scoring may eliminate some of the white noise in the scale. What would you say to the client?

16. In a multiple regression, if you drop a predictor whose effect is not significant is it true that the p-values of the other predictors should not change very much. If this is not true, describe the circumstances under which you expect to be true, or not to be true.

17. [10] Discuss the role of possible confounding variables and of possible mediating variable in an analysis of observational data with the goal of estimating whether a

variable X has a causal effect on a response Y.

18. [10] Daniel Kahneman is a Nobel prize winning psychologist. Early in his career he was training air force instructors in psychological methods to improve their own training methods for flight personnel. Kahneman told his class that praise was much more effective than criticism to encourage better performance among students. The members of Kahneman's class strongly disagreed and told Kahneman that, in their experience, student performance tended to improve after criticism and deteriorate after praise. Can you reconcile Kahneman's claim that praise is better than criticism with the experience of his students that suggests the contrary?

19. [10] Consider a "full" model in which Y depends on X, Z1, Z2, Z3 linearly, where all variables are continuous numerical variables. Let Xh be the least-squares predictor of X based on Z1, Z2, Z3. Discuss the relative advantages and disadvantages of using the model Y ~ X + Xh versus Y ~ X + Z1 + Z2 + Z3 if your main goal is to estimate the coefficient for X in the full model.

20. [10] Explain why a random intercept model, lme(Y ~ 1, random = ~ 1 | id) could also be fitted by using a 'compound symmetry' error structure, i.e. no random effect but the within 'id' covariance of the error has the form

$$\operatorname{Var}(\varepsilon)=\sigma^2\begin{bmatrix}1 & \rho & \cdots & \rho\\ \rho & 1 & \cdots & \rho\\ \vdots & \vdots & \ddots & \vdots\\ \rho & \rho & \cdots & 1\end{bmatrix}.$$

21. [20 total] Consider the following output using a familiar data set:

```
> summary(hs)
     school        mathach             ses                Sex        Minority
 5619   :  66   Min.   :-2.832   Min.   :-2.49800   Female:1074   No :1433
 4292   :  65   1st Qu.: 7.529   1st Qu.:-0.55800   Male  : 903   Yes: 544
 3610   :  64   Median :13.095   Median :-0.02800
 4530   :  63   Mean   :12.783   Mean   :-0.02684
 2208   :  60   3rd Qu.:18.336   3rd Qu.: 0.52200
 9586   :  59   Max.   :24.993   Max.   : 1.65200
 (Other):1600
      Size           Sector          PRACAD           DISCLIM
 Min.   : 215   Catholic:1144   Min.   :0.0500   Min.   :-2.4160
 1st Qu.: 545   Public  : 833   1st Qu.:0.3200   1st Qu.:-0.9390
 Median :1114                   Median :0.5800   Median :-0.2820
 Mean   :1106                   Mean   :0.5513   Mean   :-0.3022
 3rd Qu.:1415                   3rd Qu.:0.7600   3rd Qu.: 0.3360
 Max.   :2713                   Max.   :1.0000   Max.   : 1.7420
```

```
> fit <- lme( mathach ~ dvar(ses,school) + cvar(ses,school), hs,
+ random = ~ 1 + dvar(ses,school) | school)
> summary(fit)
Linear mixed-effects model fit by REML
 Data: hs
       AIC      BIC    logLik
  12847.45 12886.57 -6416.726

Random effects:
 Formula: ~1 + dvar(ses, school) | school
 Structure: General positive-definite, Log-Cholesky parametrization
                  StdDev    Corr
(Intercept)       1.5769381 (Intr)
dvar(ses, school) 0.8592066 -0.349
Residual          6.1085959

Fixed effects: mathach ~ dvar(ses, school) + cvar(ses, school)
                      Value Std.Error   DF  t-value p-value
(Intercept)       12.837129 0.2867591 1936 44.76625       0
dvar(ses, school)  2.212561 0.2569591 1936  8.61056       0
cvar(ses, school)  5.966283 0.6891302   38  8.65770       0
 Correlation:
                  (Intr) dv(,s)
dvar(ses, school) -0.162
cvar(ses, school)  0.081 -0.004

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.17563491 -0.74858727  0.03066125  0.78283631  2.71975197

Number of Observations: 1977
Number of Groups: 40
```

a) [9] On the same graph sketch the fitted population response function for mathach as a function of individual ses for a school with average ses equal to 0 and for a school with average ses equal to 1. Show unambiguously where the fixed effects coefficients above appear on the graph.

b) [6] Given the same output, specify the coefficients linear hypothesis matrices for Wald tests to test each of the following (or state that this is not possible if that is the case):
   a. Whether there is evidence of a contextual effect of ses.
   b. Whether there is evidence of a within-school effect of ses
   c. Whether there is evidece of compositional effect of ses.

c) [5] Is the following statement correct (if not, correct it)? In this model, if there were no compositional effect of ses then it would be okay to fit the model whose fixed effect formula is "mathach ~ ses".

22. [15] Referring to the same output as above:
   a. [5] Compute the G matrix

b. [5] Draw a rough sketch (but it must be clear to me) that shows the interpretation of the estimated coefficients of the random effects model. You don't need to represent the numbers accurately on your sketch but it must be clear that you understand the concepts.

c. [5] Find the standard deviation of the response relative to the population for a student with ses = 1 in a school with mean ses = 0.

23. [10] Discuss the taxonomy of outliers presented in class. How do you identify the three archetypal types of outliers? What are the consequences for statistical inference of including each type of outlier when the data point is not, in fact, an observation from the target population?

24. [5] Explain how Simpson's Paradox exemplifies the problem of causal inference with observational data.

25. [5] A news item on the radio says that new research shows that people who use sunscreen are at a higher risk of developing skin cancer than people who don't. A friend who heard the item tells you that they plan to stop using sunscreen when they go out into the sun. What advice do you give your friend? What would you do to determine whether you should stop using sunscreen?

26. [5] Explain Lord's Paradox in language a client would understand.

27. [5] Explain Suppression Effects in language a client would understand.

28. [15] Consider the attached output from the 'lme' function in R.

a. Sketch the fitted population response function over a suitable range of values of 'ses'.

b. How would you go about estimating the difference in predicted ses between Minority students and non-Minority students whose ses = -1.5. Be as specific as you can, preferably showing the code you would use in R.

c. Find the value of the within school deviation of ses for which the variance of the response is minimized. Does the analysis provided evidence that this minimizing value of the within school deviation is different from 0?

```
> summary(dd)
     school         mathach            ses           Minority
 Min.   :1317   Min.   :-2.832   Min.   :-2.49800   No :1433
 1st Qu.:3013   1st Qu.: 7.529   1st Qu.:-0.55800   Yes: 544
 Median :5650   Median :13.095   Median :-0.02800
 Mean   :5507   Mean   :12.783   Mean   :-0.02684
 3rd Qu.:7345   3rd Qu.:18.336   3rd Qu.: 0.52200
 Max.   :9586   Max.   :24.993   Max.   : 1.65200
```

Lord's Paradox

Treat Y pre & post

pre Y_1 post Y_2 treat T

Model: (1) $Y_2 \sim T + Y_1$

$D = Y_2 - Y_1$ – gain score

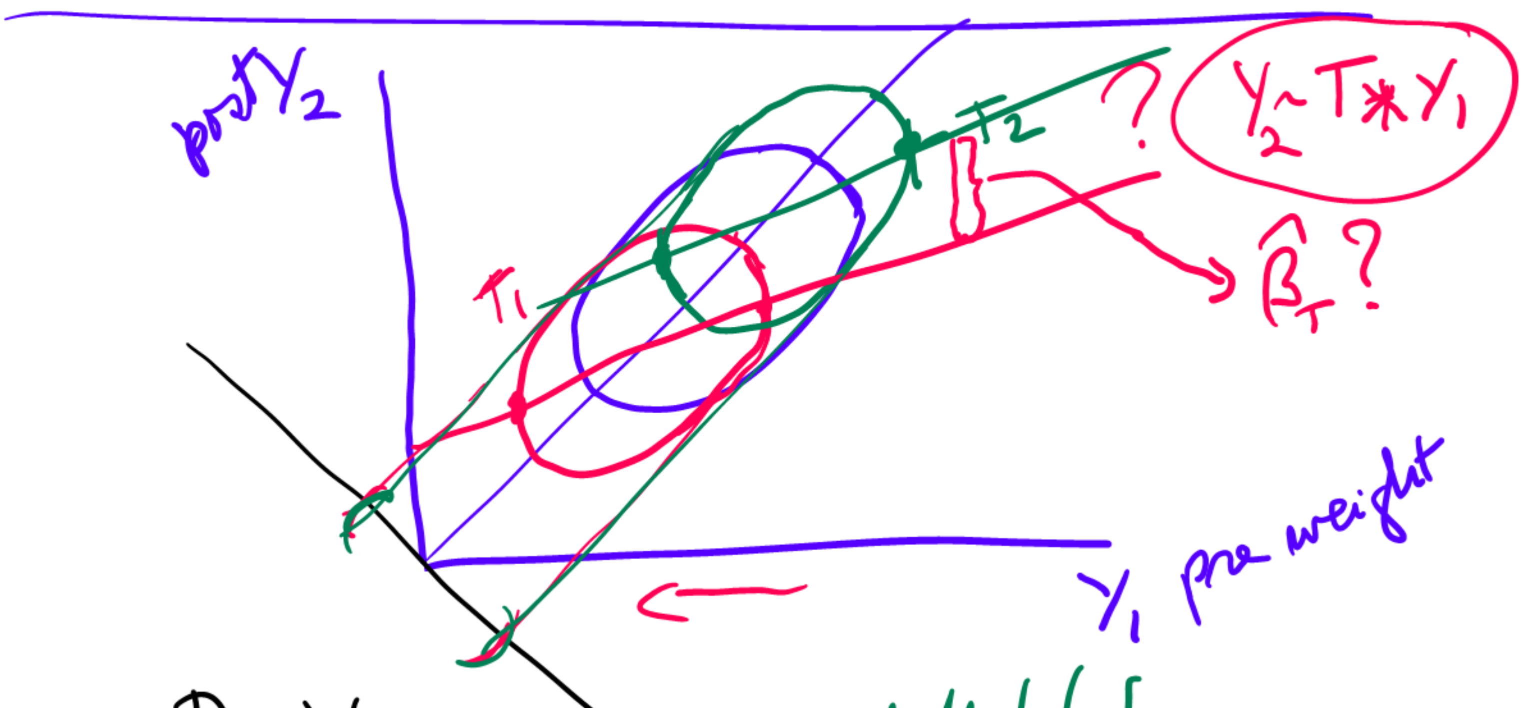
(2) $D \sim T$

(3) $D \sim T + Y_1$

(4) Residual of Y_2 on Y_1

residualized gain score

$Y_{2\cdot1} \sim T$

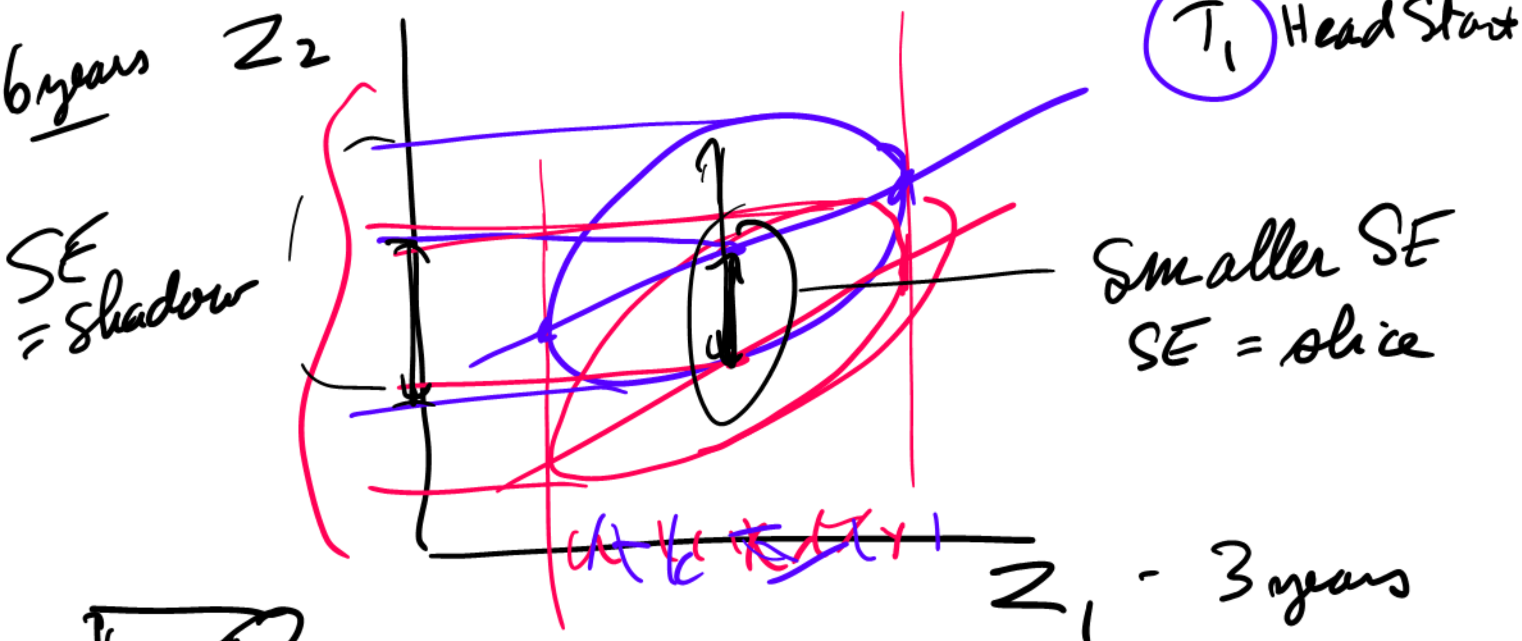


$D = Y_2 - Y_1$ — no diff !!!

Head Start 3-year-old

T_0 - no

T_1 Head Start



$Z_2 \sim T + Z_1$? Conclusion? - T_1 worse than T_0,

Green - Blue = $\hat{\beta}_T$

$Z_2 \sim T$? $\hat{\gamma}_T \downarrow$ T_1 way worse than T_0

4) D ~ T

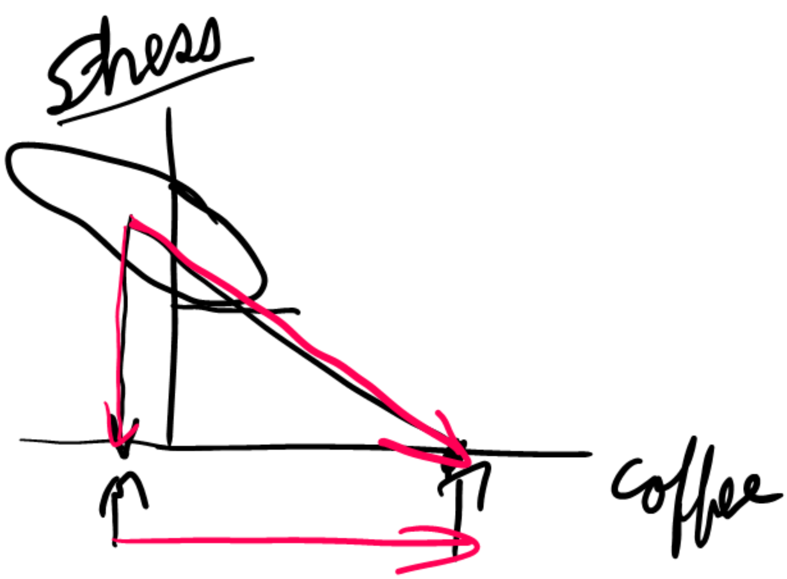
?5) D ~ T + Z_1 $\quad$ Z_2 ~ T + Z_1

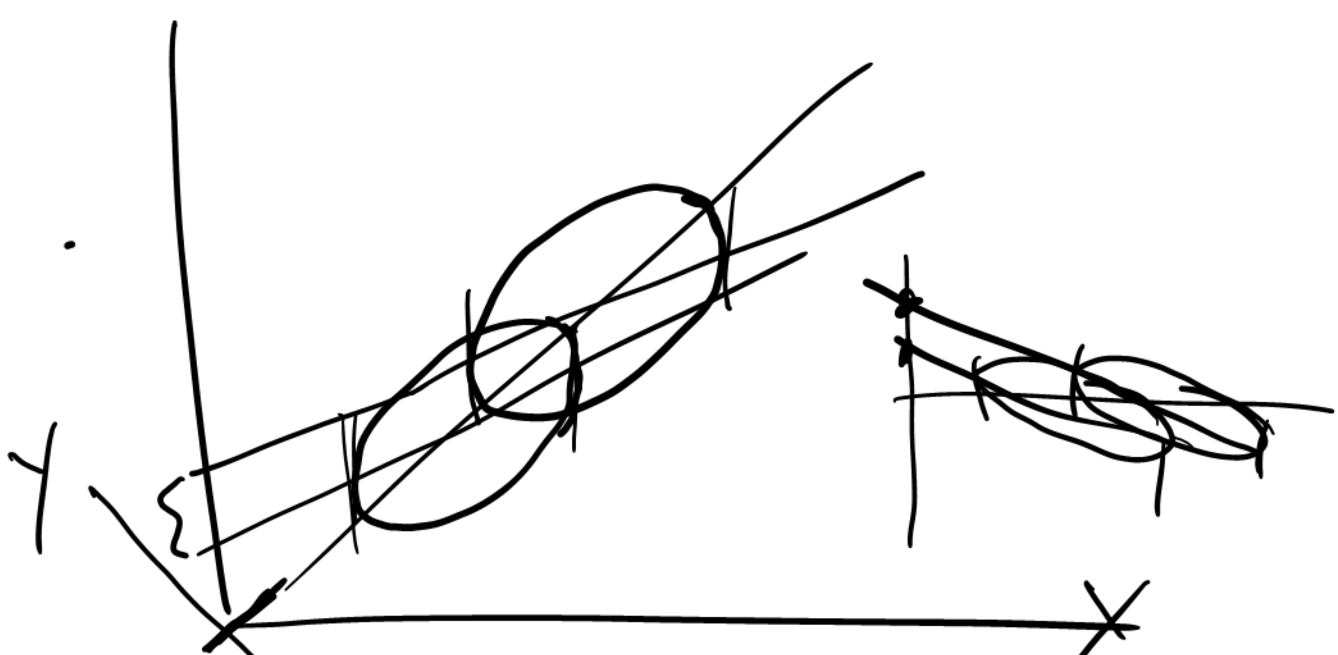
β_T $\quad$ identical $\quad$ $\hat{\beta}_T$

Mixed model. Y_1 Y_2 T

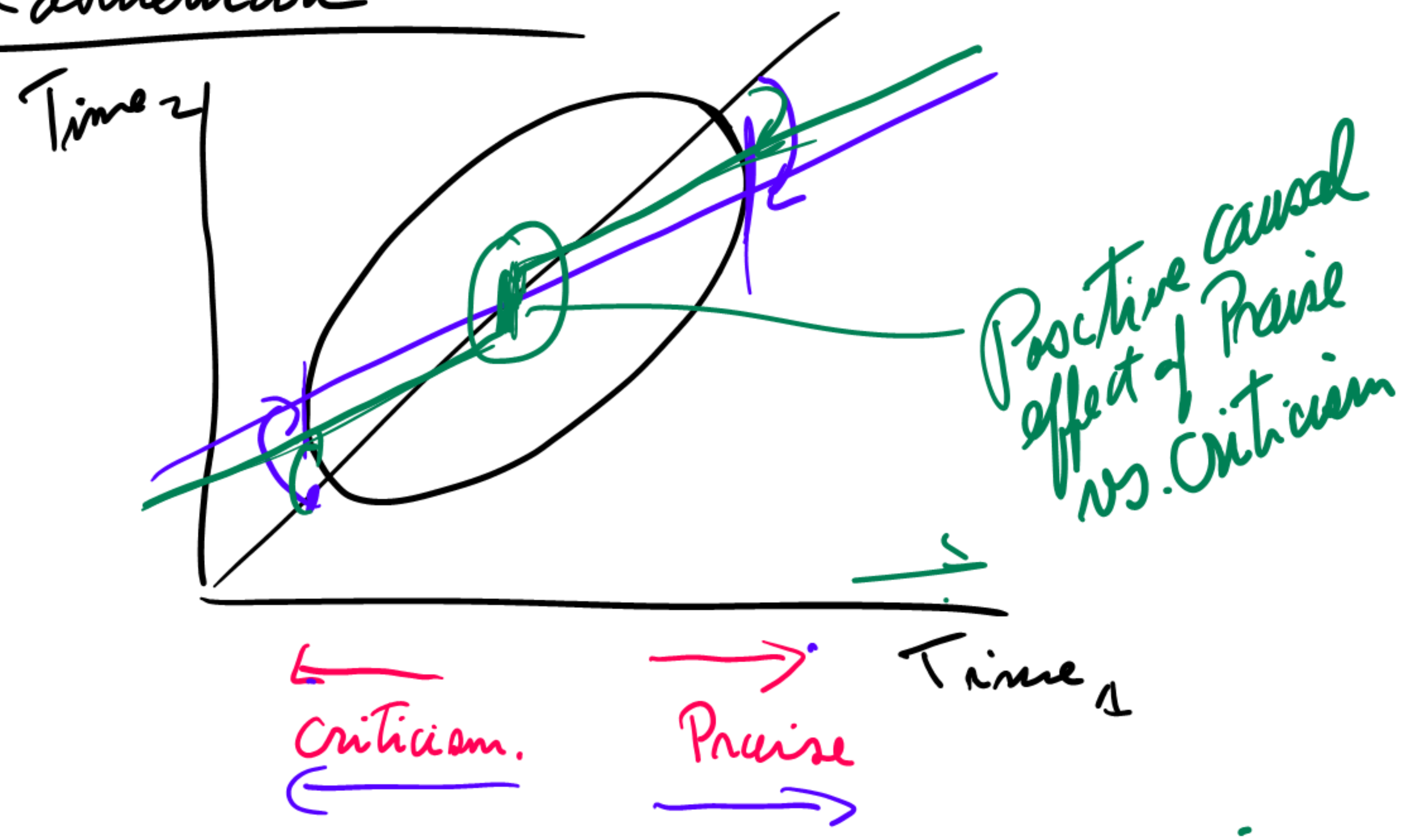
Gain score

Y ~ T random ~ 1 | id





Kahneman



Time 2
Positive causal effect of Praise vs. Criticism
Time 1
Criticism.
Praise

Y_1 , Y_2 T Change score $D = Y_2 - Y_1$

pre   post

(1) $D \sim T$

(2) $D \sim T + Y_1$ (3) $Y_2 \sim T + Y_1$

$\hat{\beta}_T$ $\hat{\beta}_T$

Prove $\hat{\beta}_T^{(3)} = \hat{\beta}_T^{(2)}$

$$\boxed{Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \underset{\sim}{e}}$$

$$\boxed{\underset{\sim}{e} \perp \mathcal{L}(X_1, X_2)}$$

then $\hat{\beta}_1$, $\hat{\beta}_2$ are OLS reg coeff

From (3)

$$\boxed{Y_2 = 1\hat{\beta}_0 + T\hat{\beta}_T + Y_1\hat{\beta}_1 + \underset{\sim}{e}}$$

$$\underset{\sim}{e} \perp \mathcal{L}(1, T, Y_1)$$

$$Y_2 - Y_1 = 1\hat{\beta}_0 + T\hat{\beta}_T + Y_1(\hat{\beta}_1 - 1) + \underset{\sim}{e} \quad \text{(sub. } Y_1 \text{ from both sides)}$$

$$D = 1\hat{\beta}_0 + T\hat{\beta}_T + Y_1\hat{\beta}_N + \underset{\sim}{e}$$

$$\underset{\sim}{e} \perp \mathcal{L}(1, T, Y_1)$$

$\therefore$ are reg. coeff of reg of D on $T + Y_1$

```
> fit <- lme( mathach ~ ses * Minority,
+           dd, random = ~ 1 + dvar(ses,school) | school)
```

$$-\frac{g_{01}}{g_{11}} = ?$$

```
> summary(fit)
Linear mixed-effects model fit by REML
 Data: dd
      AIC      BIC    logLik
  12814.02 12858.72 -6399.009

Random effects:
 Formula: ~1 + dvar(ses, school) | school
 Structure: General positive-definite, Log-Cholesky parametrization
                   StdDev   Corr
(Intercept)        1.926901 (Intr)
dvar(ses, school)  0.487619 -0.164
Residual           6.044891
```

$$\sqrt{g_{00}} = 1.92 \qquad \frac{g_{01}}{\sqrt{g_{00}g_{11}}} = -.164$$

$$\sqrt{g_{11}} = .48$$

```
Fixed effects: mathach ~ ses * Minority
                    Value Std.Error   DF   t-value p-value
(Intercept)     13.415306 0.3513888 1934 38.17795  0.0000
ses              2.588397 0.2591204 1934  9.98917  0.0000
MinorityYes     -2.908193 0.3847775 1934 -7.55811  0.0000
ses:MinorityYes -1.067539 0.4379882 1934 -2.43737  0.0149
 Correlation:
                (Intr) ses    MnrtyY
ses             -0.085
MinorityYes     -0.292  0.043
ses:MinorityYes  0.054 -0.519  0.141
```

$$\hat{\sigma} = 6.044$$

```
Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.20286150 -0.74361944  0.03123766  0.75904715  2.69807950

Number of Observations: 1977
Number of Groups: 40

> intervals(fit)
Approximate 95% confidence intervals

 Fixed effects:
                     lower      est.      upper
(Intercept)     12.726165 13.415306 14.1044464
ses              2.080212  2.588397  3.0965813
MinorityYes     -3.662815 -2.908193 -2.1535704
ses:MinorityYes -1.926518 -1.067539 -0.2085604
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: school
                                        lower       est.
sd((Intercept))                     1.4548363  1.9269006
sd(dvar(ses, school))               0.0558892  0.4876190
cor((Intercept),dvar(ses, school)) -0.9480924 -0.1642253
                                       upper
sd((Intercept))                    2.5521400
sd(dvar(ses, school))              4.2543511
cor((Intercept),dvar(ses, school)) 0.9016774

 Within-group standard error:
   lower     est.    upper
5.855011 6.044891 6.240929
```

28. [15] Consider the attached output from the 'lme' function in R.

a. Sketch the fitted population response function over a suitable range of values of 'ses'.

b. How would you go about estimating the difference in predicted ses between Minority students and non-Minority students whose ses = -1.5. Be as specific as you can, preferably showing the code you would use in R.

c. Find the value of the within school deviation of ses for which the variance of the response is minimized. Does the analysis provided evidence that this minimizing value of the within school deviation is different from 0?

```
> summary(dd)
     school          mathach              ses              Minority
 Min.   :1317   Min.   :-2.832   Min.   :-2.49800   No :1433
 1st Qu.:3013   1st Qu.: 7.529   1st Qu.:-0.55800   Yes: 544
 Median :5650   Median :13.095   Median :-0.02800
 Mean   :5507   Mean   :12.783   Mean   :-0.02684
 3rd Qu.:7345   3rd Qu.:18.336   3rd Qu.: 0.52200
 Max.   :9586   Max.   :24.993   Max.   : 1.65200
```

L = [ ]

```
Fixed effects:
                     lower       est.      upper
(Intercept)      12.726165  13.415306  14.1044464
ses               2.080212   2.588397   3.0965813
MinorityYes      -3.662815  -2.908193  -2.1535704
ses:MinorityYes  -1.926518  -1.067539  -0.2085604
attr(,"label")
```