

# Physical Activity and Body Mass Index in Sweden's National March Cohort

Scott Schwartz

Department of Statistical Science  
Duke University

June 11, 2010

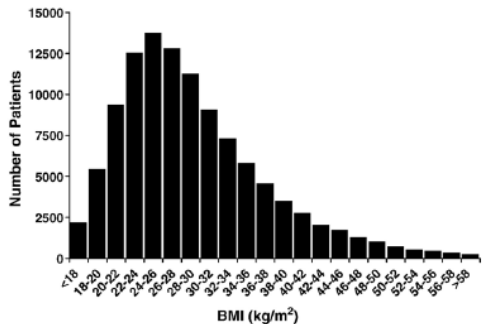
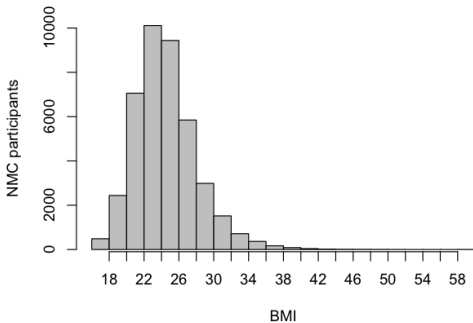


# Sweden is a little different

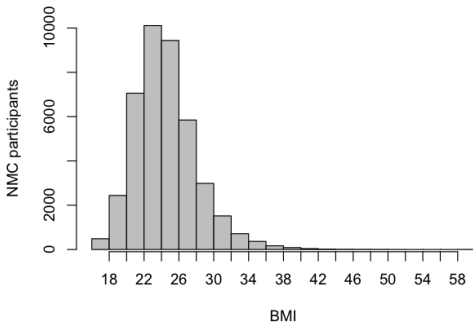


- ▶ Half of Sweden's border is coastline
- ▶ South is predominantly agricultural
- ▶ 65% forests, 15% in arctic circle
- ▶ temperate climate due to gulf stream
  
- ▶ Export economy: Timber/hydropower/iron
- ▶ Engineering (50%) pharmaceutical (25%)
- ▶ Germany/US/Norway/UK/Denmark/Finland
  
- ▶ Feudalism never developed in Sweden
- ▶ Swedes love newspaper and Smörgåsbords
- ▶ Fish (herring), meat, potatoes (sparse spices)
- ▶ Third largest music exporter in the world
- ▶ Sport activities are a national movement
- ▶ Half of the population actively participating

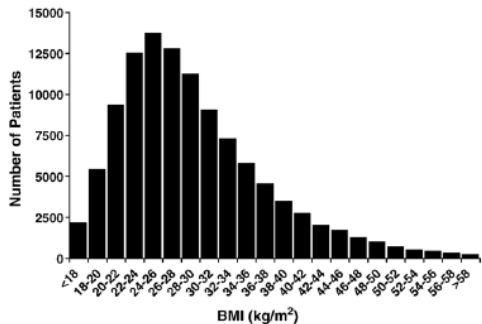
# Half the populations plays sports!??



# Half the populations plays sports!??




Sweden



Us!

# Swedish National March Cohort (NMC)

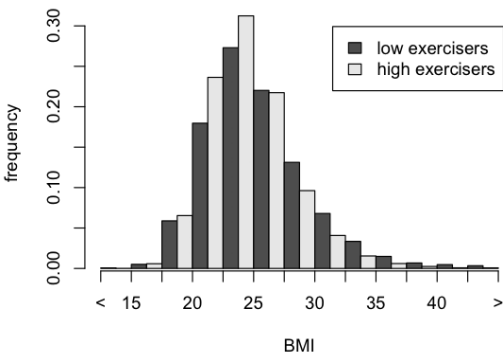
- ▶ NMC: In 1997, 300,000 (9/300)  Swedes participated in a Swedish Cancer Society national fund-raising event
- ▶ 43,880 Swedes returned questionnaires including items on risk factors for cancer & cardiovascular disease:
  - ▶ Age
  - ▶ Physical activity (PA)
  - ▶ Body mass index (BMI)
- ▶ Data issues:
  - ▶ Observational: Self-selection into study + non-response bias
  - ▶ Confounding: Self-selection into high physical activity
  - ▶ Incomplete Information: Missing many important covariates
  - ▶ Generalizability: Swedish people are Swedish people



(about 1 in 300 of US participate in the race for the cure)

## NMC data

Self-reported: 38,340 “high exercisers”; and 2,956 “low exercisers”



$$\begin{aligned}\bar{x}_1 &= 24.43 \\ s_1/\sqrt{38,340} &= 0.018 \\ \mu_1 &\in 24.43 \pm 1.96(0.018) \\ &= 0.035 \\ \bar{x}_0 &= 25.45 \\ s_0/\sqrt{2,956} &= 0.079\end{aligned}$$

- $H_0 : \mu_0 \leq 24.5$  ( $24.43 + 2.58 * 0.018 = 24.48$ )
- $H_1 : \mu_0 > 24.5$

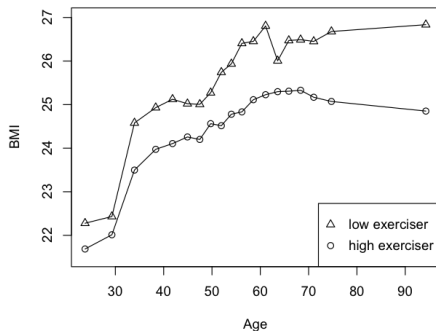
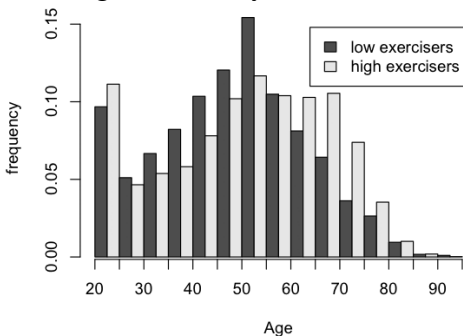
$$Z = \frac{25.45 - 24.5}{0.079} = 12.03$$

# Age is a confounding variable

$$\bar{x}_0 - \bar{x}_1 \pm 1.96 \sqrt{s_0^2/n_0 + s_1^2/n_1}$$

$$\mu_0 - \mu_1 \in (0.87, 1.19)$$

This may underestimate effects of physical activity on BMI since Age is unevenly distributed across treatments and may affect BMI.



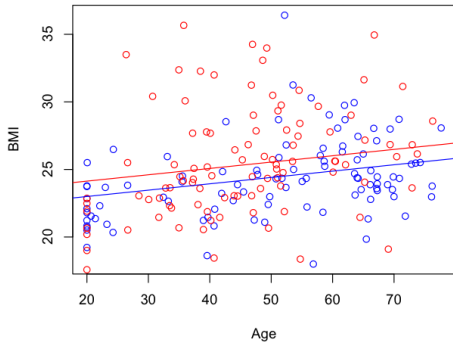
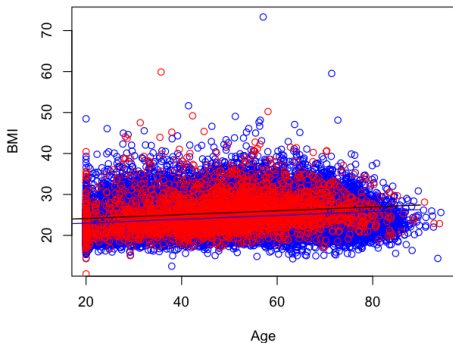
# Regression Model Fit

lm(formula = BMI ~ AGE + PA)

Coefficients:	Estimate	Std. Error	t value	$Pr(>  t )$
(Intercept)	22.06	0.06	402.64	$<2e-16$
AGE	0.047	0.001	45.71	$<2e-16$
PA	1.14	0.07	17.12	$<2e-16$

Residual standard error: 3.479 on 41293 degrees of freedom

Multiple R-squared: 0.05337, Adjusted R-squared: 0.05333



## Model Prediction

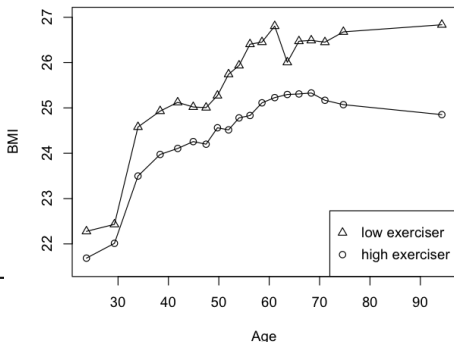
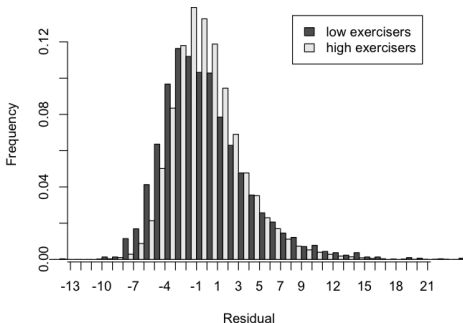
$$\hat{Y} = 22.06 + 1.14PA + 0.047Age$$

High PA	Age	estimate	estimate interval	prediction interval
Yes	30	23.47	(23.41, 23.52)	(16.65, 30.29)
No	30	24.60	(24.47, 24.73)	(17.78, 31.42)
Yes	60	24.88	(24.84, 24.92)	(18.06, 31.70)
No	30	26.01	(25.89, 26.14)	(19.19, 32.83)

- ▶ There is evidence that high physical activity reduces BMI on average.
- ▶ There is a lot of variation of BMI in the population:
- ▶ A person who exercises a lot will not necessarily have a lower BMI than someone who does not exercise a lot.
- ▶ Many other factors (genetic, etc.) likely influence BMI.

# Model Critique

- ▶ Normality? Um, no, doesn't look like it.
- ▶ Fix  $\sigma$  and homoscedasticity? No, doesn't look fixed...
- ▶ Independent observations? What about families?
- ▶ Known fixed X's? Self reported... measurement error?
- ▶  $Y = \alpha + \beta_P PA + \beta_A Age + \epsilon$ ? Is the world this simple?  
Should physical activity reduce BMI always?



# Generalizability, again

- ▶ Observational: Self-selection into study + non-response bias
- ▶ Confounding: Self-selection into high physical activity
- ▶ Incomplete Information: Missing many important covariates
- ▶ Generalizability: Swedish people are Swedish people

26

4

00:34:03  
Television

SWEDEN  
SUCKS!