Stat 306:
Finding Relationships in Data.
Lecture 13
Section 3.11 Interpretations

Section 3.11 Interpretations

- 4 categories of study

- Three issues:
  - 1. **Regression to the mean**
  - 2. **Unobserved confounding**
  - 3. **Multiple comparisons**

- **Look through some papers**

Four categories of scientific study

	Observational	Experimental
Goal is Explanation	1.	2.
Goal is Prediction	3.	4.

Studies with the goal of explaining a phenomenon

1. Observational studies are defined by having **no intervention** by researchers.

The exploratory variables in the model (X) are not determined by the researchers. Often data comes from surveys or databases.

Observational studies are important in the following fields:

- macro-economics
- epidemiology/public health
- public policy
- political science
- sociology
- criminology

Studies with the goal of explaining a phenomenon

2. Experimental studies are defined by having **a specific intervention** by researchers.

At least one exploratory variable in the model (X) is determined for each observation by the researchers. This is often done by **randomization**.
Data is collected by researchers.

Experimental studies are important in the following fields:

- medicine (clinical trials)
- educational research
- psychology

Studies with the goal of predicting future events

3. Observational studies for prediction are important in the following fields:

- Economics
- transportation research
- real estate
- financials
- insurance

4. Experimental studies for prediction are important in the following fields:

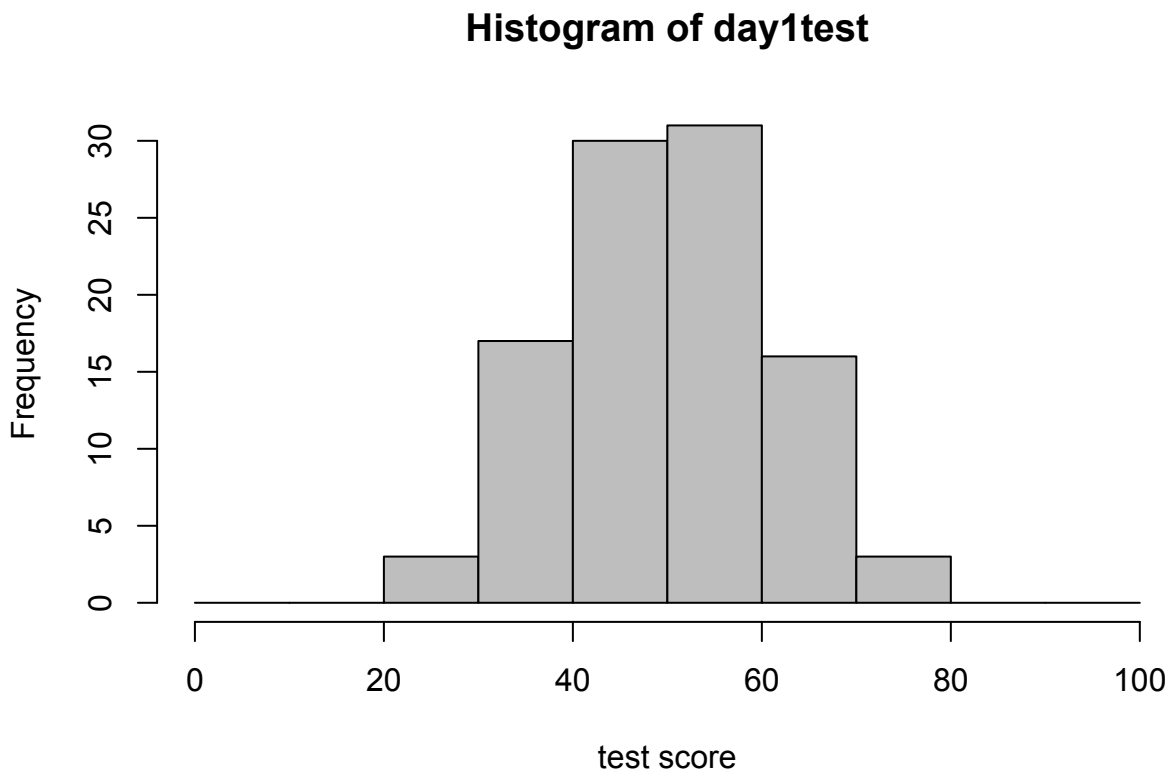
- AB testing (online advertising, website optimization)

1. Regression to the mean

Example:

On Day 1: Students take a multiple choice test and fill out the answers randomly.

We look at the results, a histogram of test scores:

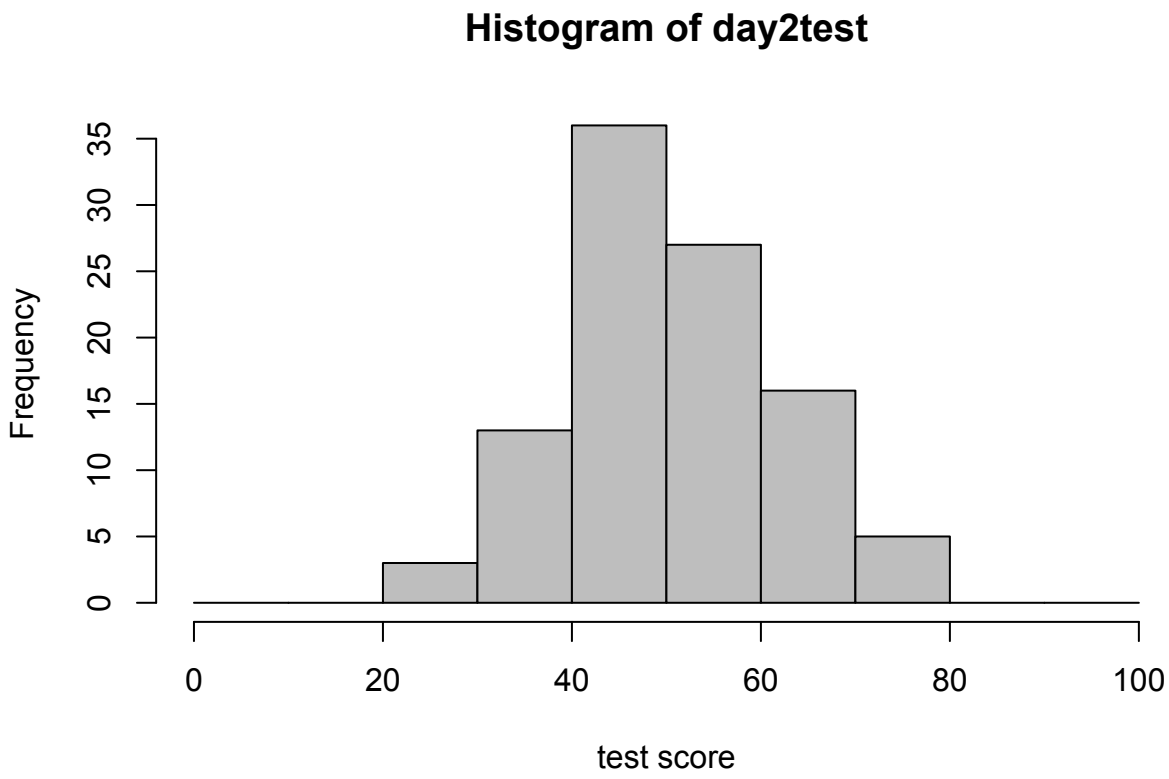


1. Regression to the mean

Example:

On Day 2: Students take another multiple choice test and fill out the answers randomly.

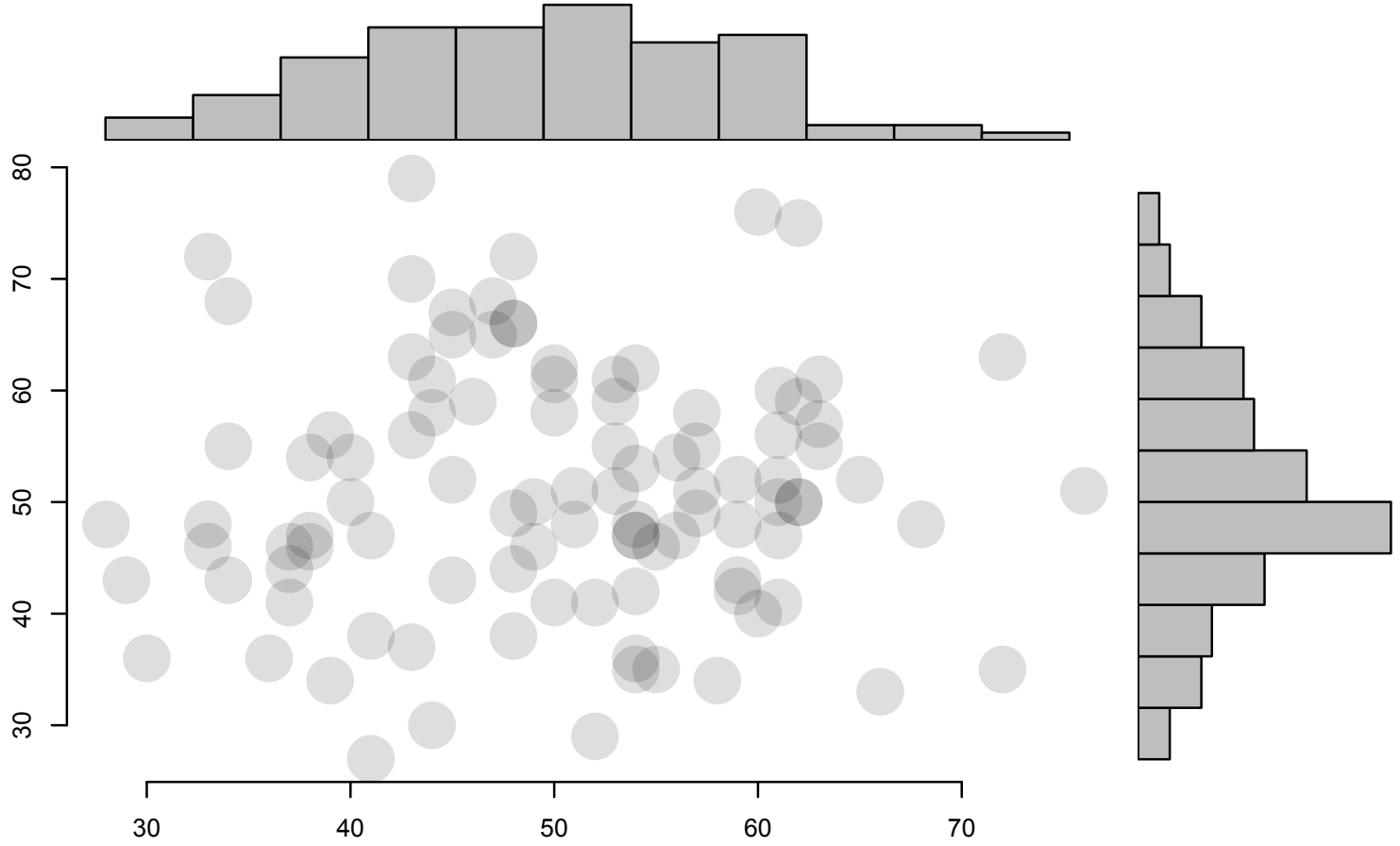
We look at the results, a histogram of test scores:



1. Regression to the mean

Example:

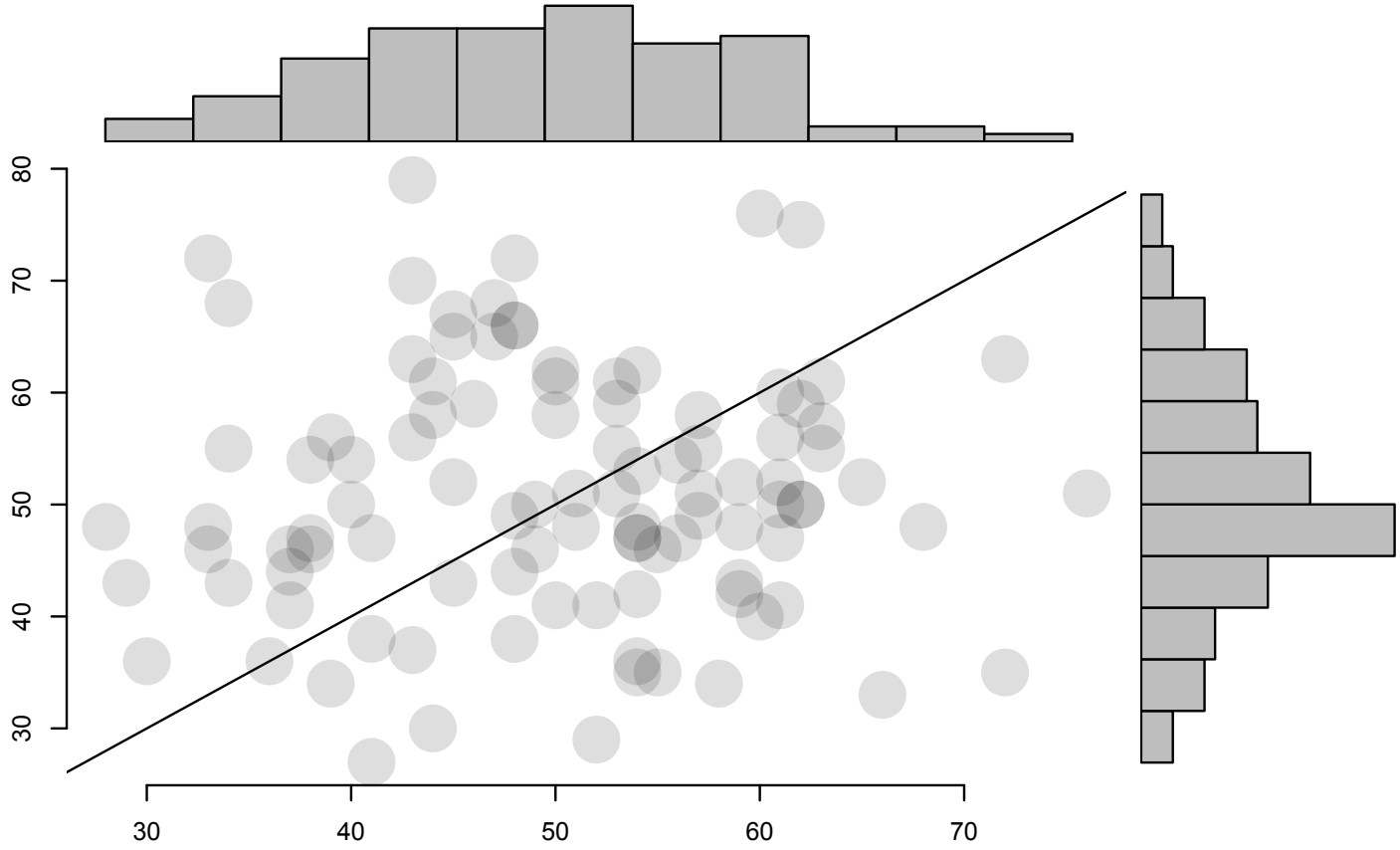
Let's look at a scatterplot of each student's two test scores:



1. Regression to the mean

Example:

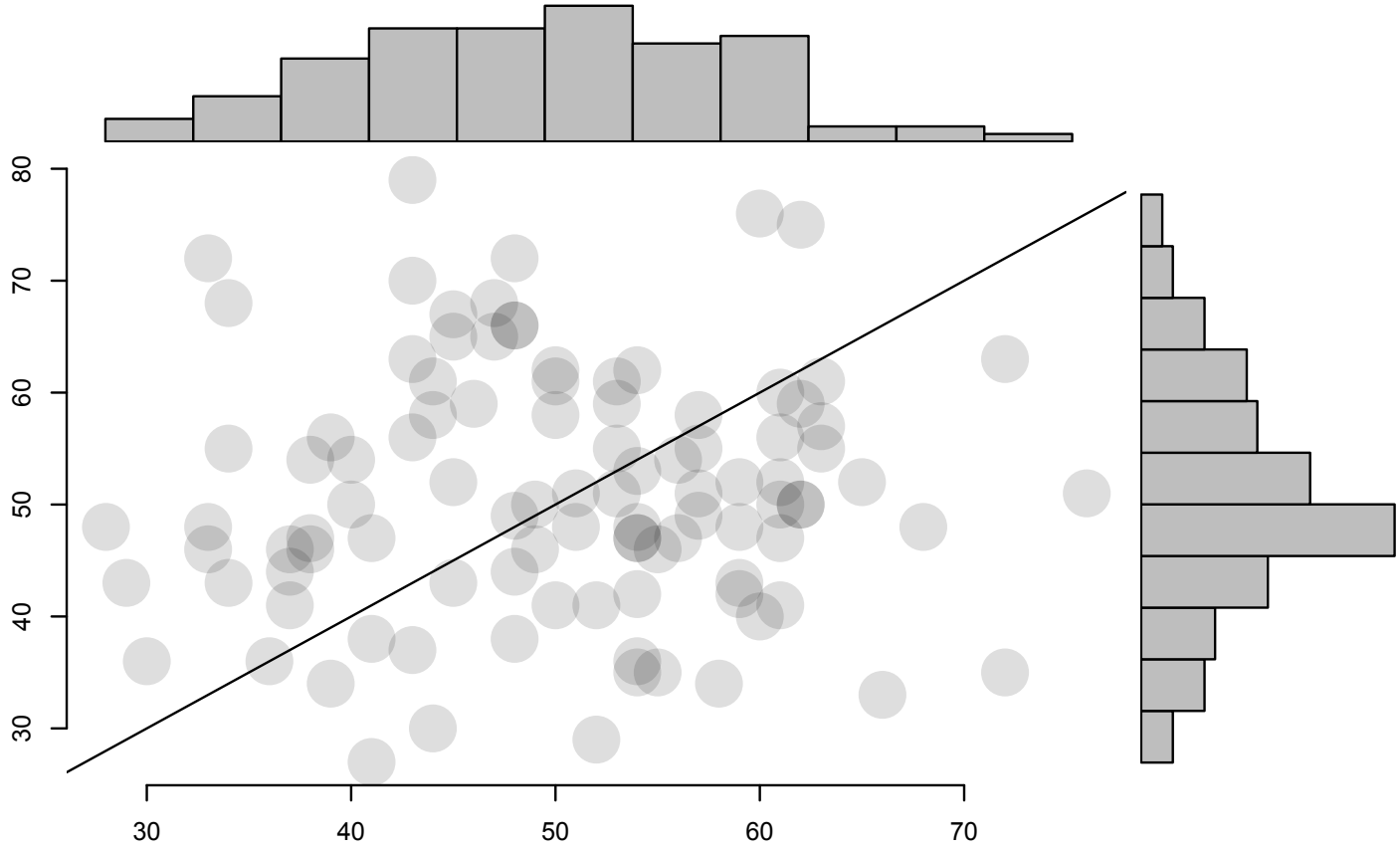
Those who did worst on the first test, tended to improve their score on the second test.



1. Regression to the mean

Let's imagine someone has a new "treatment" to help students who do poorly on multiple choice tests get better grades.

What would happen if we **tested this treatment? Would we see any improvement?**



1. Regression to the mean

More Examples from Wikipedia

https://en.wikipedia.org/wiki/Regression_toward_the_mean

- The hottest place in the country today is more likely to be cooler tomorrow than hotter, as compared to today.
- The best performing mutual fund over the last three years is more likely to see relative performance decline than improve over the next three years.
- The most successful Hollywood actor of this year is likely to have less gross than more gross for his or her next movie.
- The baseball player with the greatest batting average by the All-Star break is more likely to have a lower average than a higher average over the second half of the season.

1. Regression to the mean

- "Regression to the mean" can be a problem for **observational studies** depending on which observations are included in the analysis.

- "Regression to the mean" can be a problem for **experimental studies** if subjects are used as their own control. In other words if you simply compare The outcome post-treatment to pre-treatment, you will likely see "regression to the mean" and could mistake this for treatment effect.

- The best way to avoid this problem in **experimental studies** is to randomize subjects to two groups: a treatment group and a control group.

1. Regression to the mean

The **measurement of blood pressure** serves as a good example. If blood pressure is initially measured in a group of patients and then re-measured after a period of time, people with extreme blood pressure at Time 1 will tend to be closer to the average level at Time 2. This improvement is not due to any treatment, only due to random error.

People usually seek treatment when their symptoms are particularly severe. If treatment is sought when these symptoms are at their worst, these symptoms should be less severe simply by random fluctuations and natural recovery, even when no treatment is used

Yu and Chen (2015)
https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01574/full

1. Regression to the mean

Regression to the mean can be a problem for **observational studies** and **experimental studies** (that have no control group).

```
day1test<-round(rnorm(100,50,10))
hist(day1test, col="grey", xlab="test score", breaks=seq(0,100,10))

day2test<-round(rnorm(100,50,10))
hist(day2test, col="grey", xlab="test score", breaks=seq(0,100,10))

plot(day1test, day2test, col="grey", cex=5)

# Let's test our "treatment" on those students that
# failed the first test. Do we see any improvement?
change<-day2test-day1test
change[day1test<50]

# Using linear regression or equivalently a t-test:
summary(lm(change[day1test<50]~1))
t.test(change[day1test<50])
```

2. Unobserved Confounding

Most famous example:

The Nurses' Health Study (NHS) was one of the largest and most influential observational studies in health.

The NHS began in 1976 and subsequently followed more than 120,000 married female registered nurses.

The NHS published results in the 1991 and found that **hormone therapy** in post-menopausal women was associated with a substantial reduction in the development of **heart disease**.

In 1998, the Heart and Estrogen-progestin Replacement Study (HERS) randomized 2,763 women to receive either **hormone therapy** or **placebo**. It concluded that **hormone therapy** increased, not decreased, the risk of **heart disease**.

POSTMENOPAUSAL ESTROGEN THERAPY AND CARDIOVASCULAR DISEASE

Ten-Year Follow-up from the Nurses' Health Study

MEIR J. STAMPFER, M.D., GRAHAM A. COLDITZ, M.B., B.S., WALTER C. WILLETT, M.D., JOANN E. MANSON, M.D., BERNARD ROSNER, PH.D., FRANK E. SPEIZER, M.D., AND CHARLES H. HENNEKENS, M.D.

Abstract *Background.* The effect of postmenopausal estrogen therapy on the risk of cardiovascular disease remains controversial. Our 1985 report in the *Journal,* based on four years of follow-up, suggested that estrogen therapy reduced the risk of coronary heart disease, but a report published simultaneously from the Framingham Study suggested that the risk was increased. In addition, studies of the effect of estrogens on stroke have yielded conflicting results.

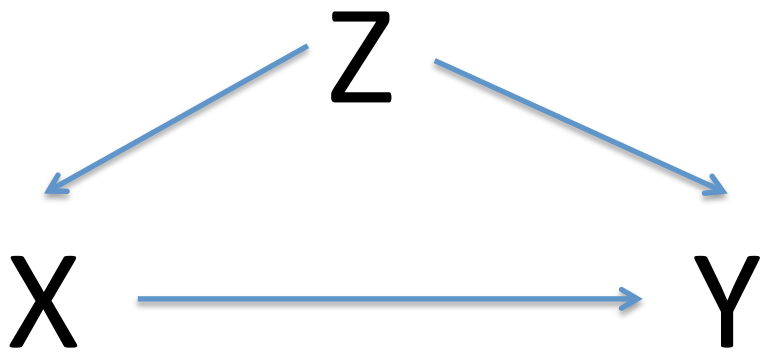
Methods. We followed 48,470 postmenopausal women, 30 to 63 years old, who were participants in the Nurses' Health Study and who did not have a history of cancer or cardiovascular disease at base line. During up to 10 years of follow-up (337,854 person-years), we documented 224 strokes, 405 cases of major coronary disease (nonfatal myocardial infarctions or deaths from coronary causes), and 1263 deaths from all causes.

Results. After adjustment for age and other risk factors, the overall relative risk of major coronary disease in women currently taking estrogen was 0.56 (95 percent confidence interval, 0.40 to 0.80); the risk was significantly reduced among women with either natural or surgical menopause. We observed no effect of the duration of estrogen use independent of age. The findings were similar in analyses limited to women who had recently visited their physicians (relative risk, 0.45; 95 percent confidence interval, 0.31 to 0.66) and in a low-risk group that excluded women reporting current cigarette smoking, diabetes, hypertension, hypercholesterolemia, or a Quetelet index above the 90th percentile (relative risk, 0.53; 95 percent confidence interval, 0.31 to 0.91). The relative risk for current and former users of estrogen as compared with those who had never used it was 0.89 (95 percent confidence interval, 0.78 to 1.00) for total mortality and 0.72 (95 percent confidence interval, 0.55 to 0.95) for mortality from cardiovascular disease. The relative risk of stroke when current users were compared with those who had never used estrogen was 0.97 (95 percent confidence interval, 0.65 to 1.45), with no marked differences according to type of stroke.

Conclusions. Current estrogen use is associated with a reduction in the incidence of coronary heart disease as well as in mortality from cardiovascular disease, but it is not associated with any change in the risk of stroke. (N Engl J Med 1991; 325:756-62.)

2. Unobserved Confounding

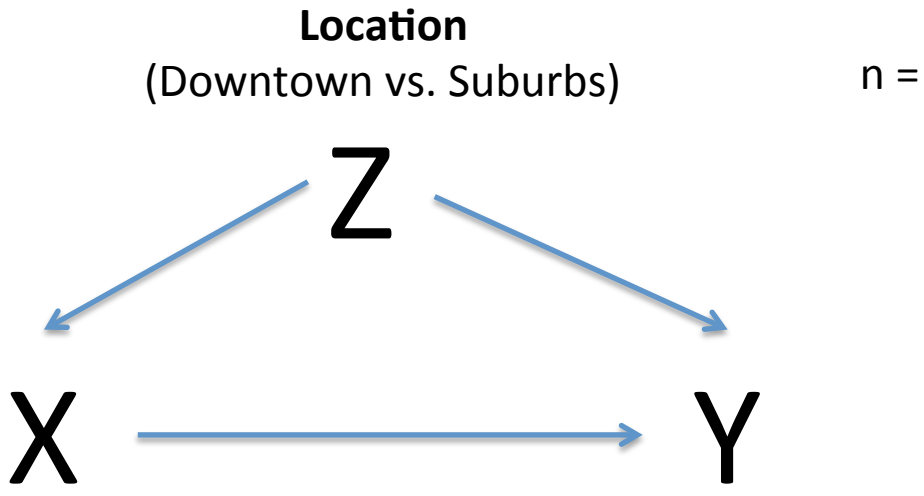
Variable that is unknown and/or unmeasured



Variable that is known and measured

Outcome variable

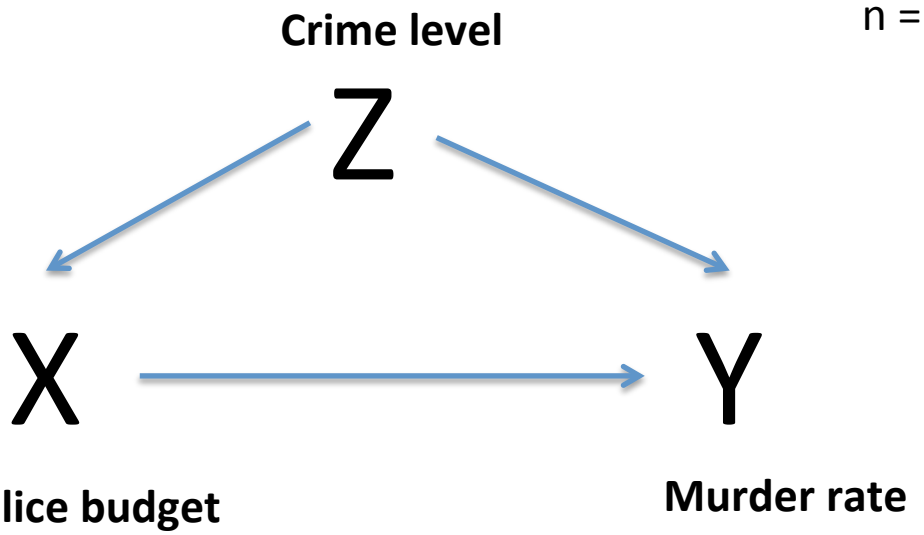
2. Unobserved Confounding



n = 200 houses for sale

What is the effect of having a larger garage on the sale price of the house?

2. Unobserved Confounding



n = 60 cities

What is the effect of increasing or decreasing the police budget on the murder rate?

3. "Multiple comparisons"

Type 1 error = Pr(reject H_0 | H_0 is true)

For linear regression:

Type 1 error = Pr(β_j is not zero | $\beta_j = 0$)
= Pr(*p*-value for β_j is small| $\beta_j = 0$)

We want to control type 1 error:

0.05 > Pr(*p*-value for β_j<0.05| $\beta_j = 0$)

3. "Multiple comparisons"

We want to control type 1 error:

$$0.05 \quad > \quad \Pr(\ p\text{-value for } \beta_j < 0.05 \mid \beta_j = 0)$$

This means that 1 out of every 20 times we claim that a β_j is significant (i.e. reject H_0: $\beta_j = 0$), we will have made a mistake.

```
> for(i in 1:20){
+ x <- rnorm(50)
+ y <- rnorm(50)
+
+ pvalue<-summary(lm(y~x))$coef[2,4]
+
+ # print the p-value for the slope:
+ print(i)
+ print(pvalue)
+ print(pvalue<0.05)
+ }
```

3. "Multiple comparisons"

We want to control type 1 error:

$$0.05 \;>\; \Pr(\, p\text{-value for } \beta_j < 0.05 \mid \beta_j = 0)$$

What happens if we have 20 variables in our model?

In other words, for one model, we test whether each β_j is significant (i.e. reject H_0: $\beta_j = 0$, for j=1,…20).

```
> y <- rnorm(50)
>
> mydata<-
data.frame(y,x1,x2,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12,x13,x14,
+ x15,x16,x17,x18,x19,x20)
>
> summary(lm(y~., data=mydata))
```

linear regression

Age vs. Money

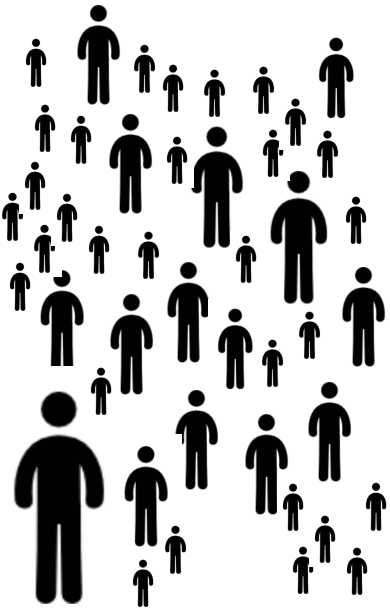
PREDICTOR *variable*

χ ⟶ Age in Years

RESPONSE *variable*

γ ⟶ cash ($) on hand

Population



Population parameters

β_0 , β_1 , σ^2

Hypothesis Test

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

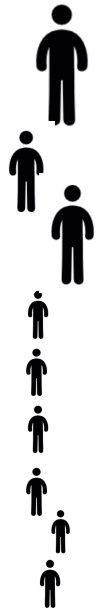
Sample statistics

b_0 = 17.7
b_1 = 0.55
s = 15.5
R^2 = 0.49

For statistic β_1 :

95% C.I. = [0.05, 1.05]
p-value = 0.036

Sample, n=9



χ	y
82	71
45	54
71	43
22	45
29	21
9	11
12	30
18	45
24	10

Age vs. Money

Objective: The purpose of this **observational study** was to demonstrate if, and to what extent, age is associated with use of cash.

Design and Methods: We collected a random sample of individuals and for each determined their age (recorded in years) and the amount of cash (in dollars) they had on hand. Analysis of the data was done using linear regression.

Sample statistics

$b_0 = 17.7$
$b_1 = 0.55$
$s = 15.5$
$R^2 = 0.49$

For parameter β_1 :

$95\%\ \text{C.I.} = [0.05, 1.05]$
$p\text{-value} = 0.036$

Results: We obtained a random sample of *n* = 9 subjects. There is a statistically significant association between age and money (*p*-value =0.036). For every additional year in age, an individual's amount of money increases on average by an estimated of $0.55 (95% C.I. = [$0.05, $1.05]).

Conclusions: We found that, as hypothesized, age is associated with cash use. In our sample age accounted for about half of the variability observed in money (R^2=0.49). We predict that a 50 year old will have $45.1 (95% P.I. = [$5.6, $84.5]), whereas a 40 year old will have $39.6 (95% P.I. = [$0.8, $78.4]).

Small Print: The analysis rests on the following assumptions:

- the observations are independently and identically distributed.
- the response variable, money, is normally distributed.
- Homoscedasticity of residuals or equal variance.
- the relationship between response and predictor variables is linear.

Gender Differences in the Salaries of Physician Researchers

Reshma Jagsi, MD, DPhil
Kent A. Griffith, MS
Abigail Stewart, PhD
Dana Sambuco, MPPA
Rochelle DeCastro, MS
Peter A. Ubel, MD

Context It is unclear whether male and female physician researchers who perform similar work are currently paid equally.

Objectives To determine whether salaries differ by gender in a relatively homogeneous cohort of physician researchers and, if so, to determine if these differences are explained by differences in specialization, productivity, or other factors.

Design and Setting A US nationwide postal survey was sent in 2009-2010 to assess the salary and other characteristics of a relatively homogeneous population of physicians. From all 1853 recipients of National Institutes of Health (NIH) K08 and K23 awards in 2000-2003, we contacted the 1729 who were alive and for whom we could identify a mailing address.

Participants The survey achieved a 71% response rate. Eligibility for the present analysis was limited to the 800 physicians who continued to practice at US academic institutions and reported their current annual salary.

Main Outcome Measures A linear regression model of self-reported current annual salary was constructed considering the following characteristics: gender, age, race, marital status, parental status, additional graduate degree, academic rank, leadership position, specialty, institution type, region, institution NIH funding rank, change of institution since K award, K award type, K award funding institute, years since K award, grant funding, publications, work hours, and time spent in research.

Results The mean salary within our cohort was $167 669 (95% CI, $158 417-$176 922) for women and $200 433 (95% CI, $194 249-$206 617) for men. Male gender was associated with higher salary (+$13 399; $P=.001$) even after adjustment in the final model for specialty, academic rank, leadership positions, publications, and research time. Peters-Belson analysis (use of coefficients derived from regression model for men applied to women) indicated that the expected mean salary for women, if they retained their other measured characteristics but their gender was male, would be $12 194 higher than observed.

Conclusion Gender differences in salary exist in this select, homogeneous cohort of mid-career academic physicians, even after adjustment for differences in specialty, institutional characteristics, academic productivity, academic rank, work hours, and other factors.

JAMA. 2012;307(22):2410-2417 www.jama.com

STUDIES HAVE REVEALED GENDER differences in physicians' pay,[1-8] but experts continue to debate the magnitude and cause of these differences. Some evidence suggests that disparities in pay are explained by specialization, work hours, and productivity,[9] leading some to believe that they are justifiable outcomes of different choices made by men and women. Debate persists in part because most studies of physicians' pay have included relatively heterogeneous groups, are now dated, or are limited by lack of information on key factors such as specialty[10] or family characteristics.[11]

Given the lack of conclusive evidence to answer whether male and female physicians who do similar work are paid differently in the present day, we sought to determine whether there were gender differences in salary of a relatively homogeneous sample: phy-

Accident Analysis and Prevention 41 (2009) 820–828

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Collision prediction models using multivariate Poisson-lognormal regression

Karim El-Basyouny*, Tarek Sayed[1]

Dept. of Civil Engineering, University of British Columbia, 2002-6250 Applied Science Lane, Vancouver, BC, Canada V6T 1Z4

ARTICLE INFO

Article history:
Received 24 December 2008
Received in revised form 20 March 2009
Accepted 2 April 2009

Keywords:
Collision prediction models
Full Bayes estimation
Markov Chain Monte Carlo
Multivariate lognormal distribution
Multivariate identification of hot spots

ABSTRACT

This paper advocates the use of multivariate Poisson-lognormal (MVPLN) regression to develop models for collision count data. The MVPLN approach presents an opportunity to incorporate the correlations across collision severity levels and their influence on safety analyses. The paper introduces a new multivariate hazardous location identification technique, which generalizes the univariate posterior probability of excess that has been commonly proposed and applied in the literature. In addition, the paper presents an alternative approach for quantifying the effect of the multivariate structure on the precision of expected collision frequency. The MVPLN approach is compared with the independent (separate) univariate Poisson-lognormal (PLN) models with respect to model inference, goodness-of-fit, identification of hot spots and precision of expected collision frequency. The MVPLN is modeled using the WinBUGS platform which facilitates computation of posterior distributions as well as providing a goodness-of-fit measure for model comparisons. The results indicate that the estimates of the extra Poisson variation

VERSITA

Information Technology and Management Science

doi: 10.2478/itms-2013-0014

2013 / 16

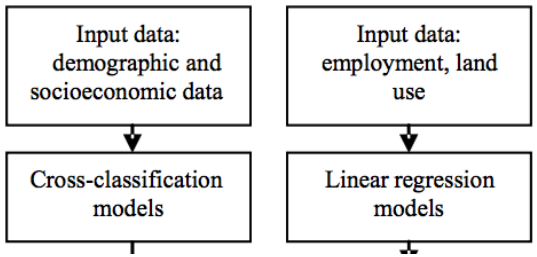
Regression Analysis for Transport Trip Generation Evaluation

Nadezda Zenina[1], Arkady Borisov[2], [1-2] *Riga Technical University*

Abstract **– The paper focuses on transportation trip generation models based on mixed-use and transport infrastructure near the site. Transport trip generation models are considered with an aim to improve the accuracy of transport generated trips. Information systems are reviewed, and "smart growth" criteria that could affect the accuracy of trip generation models are also identified. Experimental results of transport generated trips based on linear regression equations and "smart growth" tools are demonstrated.**

Keywords **– Information systems, linear regression equations, transport trip generation models**

For each type of generated trips, there are various calculation methods (Fig. 2), but, as a result, the number of trip attraction should be the same as the number of trip production.



Giuli et al. BMC Public Health 2014, 14:35
http://www.biomedcentral.com/1471-2458/14/35

Giuli et al. (2014)

BMC Public Health

RESEARCH ARTICLE Open Access

Correlates of perceived health related quality of life in obese, overweight and normal weight older adults: an observational study

Cinzia Giuli[1*], Roberta Papa[2], Roberta Bevilacqua[3], Elisa Felici[3], Cristina Gagliardi[4], Fiorella Marcellini[3], Marco Boscaro[5], Marco De Robertis[5], Eugenio Mocchegiani[6], Emanuela Faloia[5] and Giacomo Tirabassi[5]

Abstract

Background: Obesity is a complex multifactorial disease, which also has an impact on quality of life. The aim of this paper is to identify the correlates of perceived health related quality of life in obese, overweight and normal weight Italians older adults.

Methods: 205 subjects at the age ≥ 60 yrs. were recruited into the Division of Endocrinology of the Polytechnic University of Marche Region, Ancona (Italy). A protocol of questionnaires was constructed for data collection, and included domains such as physical activity, quality of life, socio-psychological aspects. The association of the latter variables with SF-36 Health Survey physical component (PCS-36) were evaluated in the whole sample. Multiple linear regression models were used to assess the effect of independent variables on PCS-36 and the physical subscales of SF-36.

Results: PCS-36 showed a lower score in the obese and overweight subjects than the normal weight group (post-hoc test, $p < 0.001$ and $p < 0.05$ respectively). Age, gender (male), Body Mass Index, years of education, Physical Activity Scale for the Elderly (PASE) total score, Hospital Anxiety and Depression Scale anxiety, Hospital Anxiety and Depression Scale depression, number of medications prescribed and number of diseases were included in the model. Negative and significant PCS-associated variables included depression ($p = 0.009$), BMI ($p = 0.001$), age in years ($p = 0.007$), whereas positive and significant PCS-associated independent variables were years of education ($p = 0.022$), physical activity ($p = 0.026$). BMI was negatively associated with all the physical subscales of SF-36 ($p < 0.05$).

Conclusions: Research funding should be invested in the study of the benefits accruing from reducing obesity in the elderly.

Keywords: Obesity, Older adults, Health related quality of life

Health Promotion International, Vol. 27 No. 2
doi:10.1093/heapro/dar036
Advance Access published 16 June, 2011

Haby et al. (2011)

Future predictions of body mass index and overweight prevalence in Australia, 2005–2025

MICHELLE M. HABY[1*], ALISON MARKWICK[1], ANNA PEETERS[2], JONATHAN SHAW[3] and THEO VOS[4]

[1]Department of Health, Prevention and Population Health Branch, 50 Lonsdale Street, Melbourne, VIC 3000, Australia [2]Department of Epidemiology and Preventive Medicine, Monash University, VIC, Australia [3]Baker IDI Heart and Diabetes Institute, 99 Commercial Road, Melbourne, VIC 3004, Australia [4]School of Population Health, University of Queensland, Herston, QLD, Australia
**Corresponding author. E-mail: michelle.haby@health.vic.gov.au*

SUMMARY

To predict current and future body mass index (BMI) and prevalence of overweight and obesity in Australian children and adults based on sex, age and year of birth (cohort). These predictions are needed for population health planning and evaluation. Data were drawn from 11 cross-sectional national or state population surveys conducted in Australia between 1969 and 2004. These included representative population samples of children (n = 27 635) and adults (n = 43 447) aged 5 years or older with measured height and weight data. Multiple linear regression analyses of measured log-transformed BMI data were conducted to determine the independent effects of age and year of birth (cohort) on ln(BMI) for males and females, respectively. Regression coefficients for cohort obtained from these analyses were applied to the National Nutrition Survey 1995 data set to predict mean BMI and prevalence of overweight (BMI 25–29.99 kg/m^2) and obesity (BMI ≥30kg/m^2) in 2005, 2015 and 2025. Based on past trends, BMI is predicted to continue to increase for both males and females and across the age span. This would result in increases in the prevalence of overweight and obesity of between 0.4 and 0.8% per year, such that by 2025 around one-third of 5–19 year olds will be overweight or obese as will 83% of males and 75% of females aged 20 years and over. The increases in prevalence and mean BMI predicted in this study will have significant impacts on disease burden, healthcare costs and need for prevention and treatment programmes.

Key words: body mass index; forecasting; overweight; prevalence; Australia

American Journal of Epidemiology

Vol. 155, No. 7
Printed in U.S.A.

Protein Consumption and Bone Mineral Density in the Elderly

The Rancho Bernardo Study

Promislow et al. (2002)

Joanne H. E. Promislow,[1] Deborah Goodman-Gruen,[2] Donald J. Slymen,[3] and Elizabeth Barrett-Connor[2]

The role of dietary protein in osteoporosis is unclear, with previous studies having suggested both protection and harm. The associations of total, animal, and vegetable protein with bone mineral density (BMD) and the variations in these associations with calcium intake were studied in a community-dwelling cohort of 572 women and 388 men aged 55–92 years (Rancho Bernardo, California). Multiple linear regression analyses adjusted for standard osteoporosis covariates showed a positive association between animal protein consumption, assessed by food frequency questionnaires in 1988–1992, and BMD, measured 4 years later. This association was statistically significant in women. For every 15-g/day increase in animal protein intake, BMD increased by 0.016 g/cm^2 at the hip ($p = 0.005$), 0.012 g/cm^2 at the femoral neck ($p = 0.02$), 0.015 g/cm^2 at the spine ($p = 0.08$), and 0.010 g/cm^2 for the total body ($p = 0.04$). Conversely, a negative association between vegetable protein and BMD was observed in both sexes. Some suggestion of effect modification by calcium was seen in women, with increasing protein consumption appearing to be more beneficial for women with lower calcium intakes, but evidence for this interaction was not consistently strong. This study supports a protective role for dietary animal protein in the skeletal health of elderly women. *Am J Epidemiol* 2002;155:636–44.

aged; bone density; dietary proteins; osteoporosis

Fallowfield et al. (2002)

British Journal of Cancer (2002) 87, 1341 – 1353

www.bjcancer.com

Multivariate regression analyses of data from a randomised, double-blind, placebo-controlled study confirm quality of life benefit of epoetin alfa in patients receiving non-platinum chemotherapy

Clinical

L Fallowfield*[,1], D Gagnon[2], M Zagari[3], D Cella[4], B Bresnahan[2], TJ Littlewood[5], P McNulty[2], G Gorzegno[6] and M Freund[7] for the Epoetin Alfa Study Group[8]

[1]Cancer Research UK Psychosocial Oncology Group, Brighton and Sussex Medical School, University of Sussex, Brighton, BN1 9QG, UK; [2]Johnson & Johnson Pharmaceutical Research & Development, L.L.C., 700 Route 202, Raritan, New Jersey, NJ 08869, USA; [3]Ortho Biotech UKI, PO Box 829, Saunderton High Wycombe, Bucks, HP14 4HJ, UK; [4]Institute for Health Services Research and Policy Studies, Northwestern University, 1001 University Place, Evanston, Illinois, IL 60201, USA; [5]John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK; [6]Ospedale S. Luigi, Orbassano, Via Regione Gonzole 10, Orbassano, 10126 Italy; [7]University of Rostock, Postsach 100888, Rostock, 18055 Germany

Cancer-related anaemia is associated with a wide spectrum of symptoms that can negatively affect quality of life. Because epoetin alfa has demonstrated efficacy in correcting cancer-related anaemia, the impact of this treatment on quality of life was evaluated in a multinational, randomised, double-blind, placebo-controlled trial in 375 anaemic cancer patients receiving non-platinum-based chemotherapy. The cancer-specific measures of quality of life included the general scale (FACT-G Total) and fatigue subscale (FACT-An Fatigue subscale) of the Functional Assessment of Cancer Therapy-Anaemia and the Cancer Linear Analogue Scales measuring energy, ability to do daily activities, and overall quality of life. These measures were also used to examine the relationship between haemoglobin levels and quality of life. Both univariate and multiple linear regression analyses of quality of life data were performed. Results of the univariate analysis have been reported previously. The *a priori*-planned multiple linear regression analysis, which accounted for the effects of disease progression and several other possibly confounding variables on quality of life, showed a significant advantage for epoetin alfa over placebo for the five scales (all, $P<0.05$), and confirmed the results of the univariate analysis. For cancer-specific measures, significant correlations were demonstrated between baseline haemoglobin and quality of life (r, range: 0.14 – 0.26, all $P<0.05$) and between change in haemoglobin and change in quality of life (r, range: 0.26 – 0.34, all $P<0.01$). These findings provide evidence that increasing haemoglobin levels by epoetin alfa administration can significantly improve cancer patients' quality of life.
British Journal of Cancer (2002) **87,** 1341 – 1353. doi:10.1038/sj.bjc.6600657 www.bjcancer.com

Keywords: anaemia; cancer; epoetin alfa; haemoglobin; non-platinum chemotherapy; quality of life

Ethnic Differences in Blood Pressure Response to First and Second-Line Antihypertensive Therapies in Patients Randomized in the ASCOT Trial

Ajay K. Gupta[1], Neil R. Poulter[1], Joanna Dobson[1], Sandra Eldridge[2], Francesco P. Cappuccio[3], Mark Caulfield[2], David Collier[2], J. Kennedy Cruickshank[4], Peter S. Sever[1] and Gene Feder[5] on behalf of ASCOT investigators

BACKGROUND
Some studies suggest that blood pressure (BP)–lowering effects of commonly used antihypertensive drugs differ among ethnic groups. However, differences in the response to second-line therapy have not been studied extensively.

METHODS
In the BP-lowering arm of the Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT-BPLA), BP levels of European ($n = 4,368$), African (203), and South-Asian- (132) origin patients on unchanged monotherapy (atenolol or amlodipine) and/or on second-line therapy (added thiazide or perindopril) were compared. Interaction between ethnicity and BP responses (defined as end BP minus start of therapy BP) to both first- and second-line therapies were assessed in regression models after accounting for age, sex, and several other potential confounders.

RESULTS
BP response to atenolol and amlodipine monotherapy differed among the three ethnic groups (interaction test $P = 0.05$). Among those allocated atenolol monotherapy, black patients were significantly less responsive (mean systolic BP (SBP) difference +1.7 (95% confidence interval: −1.1 to 4.6) mm Hg) compared to white patients (referent). In contrast, BP response to amlodipine monotherapy did not differ significantly by ethnic group. BP responses to the addition of second-line therapy also differed significantly by ethnic group (interaction test $P = 0.004$). On adding a diuretic to atenolol, BP lowering was similar among blacks and South-Asians as compared to whites (referent). However, on addition of perindopril to amlodipine, BP responses differed significantly: compared to whites (SBP difference −1.7 (−2.8 to −0.7) mm Hg), black patients had a lesser response (SBP difference 0.8 (−2.5 to 4.2) mm Hg) and South-Asians had a greater response (SBP difference −6.2 (−10.2 to −2.2) mm Hg).

CONCLUSIONS
We found important differences in BP responses among ethnic groups to both first- and second-line antihypertensive therapies.

Keywords: *antihypertensive agents; blood pressure; ethnicity; hypertension; race; treatment*

American Journal of Hypertension, advance online publication 19 August 2010; doi:10.1038/ajh.2010.105

American Journal of Epidemiology

Vol. 155, No. 7
Printed in U.S.A.

Protein Consumption and Bone Mineral Density in the Elderly

The Rancho Bernardo Study

Joanne H. E. Promislow,[1] Deborah Goodman-Gruen,[2] Donald J. Slymen,[3] and Elizabeth Barrett-Connor[2]

The role of dietary protein in osteoporosis is unclear, with previous studies having suggested both protection and harm. The associations of total, animal, and vegetable protein with bone mineral density (BMD) and the variations in these associations with calcium intake were studied in a community-dwelling cohort of 572 women and 388 men aged 55–92 years (Rancho Bernardo, California). Multiple linear regression analyses adjusted for standard osteoporosis covariates showed a positive association between animal protein consumption, assessed by food frequency questionnaires in 1988–1992, and BMD, measured 4 years later. This association was statistically significant in women. For every 15-g/day increase in animal protein intake, BMD increased by 0.016 g/cm^2 at the hip ($p = 0.005$), 0.012 g/cm^2 at the femoral neck ($p = 0.02$), 0.015 g/cm^2 at the spine ($p = 0.08$), and 0.010 g/cm^2 for the total body ($p = 0.04$). Conversely, a negative association between vegetable protein and BMD was observed in both sexes. Some suggestion of effect modification by calcium was seen in women, with increasing protein consumption appearing to be more beneficial for women with lower calcium intakes, but evidence for this interaction was not consistently strong. This study supports a protective role for dietary animal protein in the skeletal health of elderly women. *Am J Epidemiol* 2002;155:636–44.

aged; bone density; dietary proteins; osteoporosis

The art of linear regression

- Categorical predictors
- Quadratic (polynomial) relationships
- Outliers
- How to fix heterogeneity
- Regression to the mean
- Simpsons Paradox
- Unobserved Confounding