

Statistics 104

Introduction to Quantitative Methods

Mark E. Irwin
Department of Statistics
Harvard University

Autumn Term

Monday, September 20, 2004 -
Wednesday, January 19, 2005



Personnel

Instructor: Mark Irwin
Office: 611 Science Center
Phone: 617-495-5617
E-mail: irwin@stat.harvard.edu
Web-site: <http://www.courses.fas.harvard.edu/~stat104/>

Lectures: MWF 11:00 - 12:00, Science Center Lecture Hall A
Office Hours: Tuesday, 2:30 - 3:30, Friday 1:00 - 2:00,
or by appointment

Teaching Fellows: Chris Chiambalero, Soujit Ghosh, Hui Jin,
Taeyoung Park
Office Hours: Science Center 600 or 601 - Times to be announced

Sections

Weekly 1 hour session with a Teaching Fellow. Pick your section choices at

<http://www.register.fas.harvard.edu/section/>

Please do so by 3:00 pm on Monday September 27th.

Sections will start the 3rd week of the term.

Text Books

Required Text: Moore DS and McCabe (2002). Introduction to the Practice of Statistics, 4th Edition. Freeman.
Available from the COOP, <<http://redlinetextbooks.com/>>
and on reserve in Cabot Library

Suggested
References: Freedman D, Pisani R, and Purves R (1998). Statistics.
Norton (Less mathematical treatment)

(Available
in Cabot) Devore J and Peck R (2004). Statistics: The Exploration
and Analysis of Data, 5th edition. Brooks/Cole

Tanur J et al (1989). Statistics: A Guide to the
Unknown. Wadsworth (full of examples)

Against All Odds: Inside Statistics. A videotaped series
loosely tied to an earlier edition of Moore and McCabe.
(On reserve in Cabot)

Computing

The suggested statistics package for the course is Stata. It is available in Macintosh and Windows versions in the Science Center computer labs. It is also available from

<<http://www.fas.harvard.edu/cgi-bin/software/download.pl>>

if you wish to download it to your own computer.

No experience with Stata is required. The use of the package will be discussed during the term and documentation will be available from the course web site.

Note that this is a suggested package. If you are familiar with another package you may use it. However do not expect to be able to get help with these other packages.

Grading

- Homework (20%): Around 10 during the term. No late homework will be accepted. The lowest homework grade will be dropped when computing your final grade.
- Midterms (30%): Tentatively scheduled for Monday, October 25th and Wednesday, December 1 in lecture.
- Term project (15%): A brief (5 pages) paper on anything statistical which interests you. More information will be given in November. Will be due during reading period.
- Final (35%): Tentatively scheduled for Wednesday, January 19th. Location to be announced. (Exam group 4).

Web Site

<http://www.courses.fas.harvard.edu/~stat104/>

Available on the site:

- PDF files of the lecture notes
- Data files, in text and Stata format, used in class
- Assignments, solutions, data files for them
- Stata information
- Class announcements, calendar, etc

Course Goals

- Basic concepts and language of Statistics.
- Techniques: their use and interpretation.
- Familiarity with computer software (Stata).
- Data analysis.

A more detailed schedule, with suggested readings from Moore and McCabe and viewings from *Against All Odds*, is given in the syllabus available on line

Examples of Statistics

Medicine:

The image is a screenshot of a JAMA (Journal of the American Medical Association) article page. At the top left is the JAMA logo. To the right is a search bar with the text "SEARCH THIS JOURNAL:" and a "GO" button. Below the search bar is a navigation menu with links for "HOME", "CURRENT ISSUE", "PAST ISSUES", "COLLECTIONS", "CONTACT US", and "HELP". The main content area shows the article title "Sugar-Sweetened Beverages, Weight Gain, and Incidence of Type 2 Diabetes in Young and Middle-Aged Women" and the authors: Matthias B. Schulze, DrPH; JoAnn E. Manson, MD; David S. Ludwig, MD; Graham A. Colditz, MD; Meir J. Stampfer, MD; Walter C. Willett, MD; Frank B. Hu, MD. The article is from JAMA, 2004;292:927-934. The abstract includes sections for Context, Objective, Design, Setting, and Participants, Main Outcome Measures, Results, and Conclusion. On the right side, there are several sidebar sections: "Featured Link" with "E-mail Alerts"; "Article Options" with links for "Full text", "PDF", "Send to a Friend", "Related articles in this issue", and "Similar articles in this journal"; "Literature Track" with links for "Add to File Drawer", "Download to Citation Manager", "PubMed citation", and "Articles in PubMed by" (listing Schulze MB and Hu FB); "Articles that cite this article" (listing "ISI Web of Science (1)"); "Contact me when this article is cited"; and "Topic Collections" with links for "Obesity", "Diabetes Mellitus", "Women's Health, Other", and "Collection E-mail Alerts".

JAMA

SEARCH THIS JOURNAL:

GO TO ADVANCED SEARCH >

The Journal of the American Medical Association — To Promote the Science and Art of Medicine and the Betterment of the Public Health

HOME CURRENT ISSUE PAST ISSUES COLLECTIONS CONTACT US HELP

Vol. 292 No. 8, August 25, 2004 [TABLE OF CONTENTS >](#)

Original Contribution

Sugar-Sweetened Beverages, Weight Gain, and Incidence of Type 2 Diabetes in Young and Middle-Aged Women

Matthias B. Schulze, DrPH; JoAnn E. Manson, MD; David S. Ludwig, MD; Graham A. Colditz, MD; Meir J. Stampfer, MD; Walter C. Willett, MD; Frank B. Hu, MD

JAMA. 2004;292:927-934.

Context Sugar-sweetened beverages like soft drinks and fruit punches contain large amounts of readily absorbable sugars and may contribute to weight gain and an increased risk of type 2 diabetes, but these relationships have been minimally addressed in adults.

Objective To examine the association between consumption of sugar-sweetened beverages and weight change and risk of type 2 diabetes in women.

Design, Setting, and Participants Prospective cohort analyses conducted from 1991 to 1999 among women in the Nurses' Health Study II. The diabetes analysis included 91 249 women free of diabetes and other major chronic diseases at baseline in 1991. The weight change analysis included 51 603 women for whom complete dietary information and body weight were ascertained in 1991, 1995, and 1999. We identified 741 incident cases of confirmed type 2 diabetes during 716 300 person-years of follow-up.

Main Outcome Measures Weight gain and incidence of type 2 diabetes.

Results Those with stable consumption patterns had no difference in weight gain, but weight gain over a 4-year period was highest among women who increased their sugar-sweetened soft drink consumption from 1 or fewer drinks per week to 1 or more drinks per day (multivariate-adjusted means, 4.69 kg for 1991 to 1995 and 4.20 kg for 1995 to 1999) and was smallest among women who decreased their intake (1.34 and 0.15 kg for the 2 periods, respectively) after adjusting for lifestyle and dietary confounders. Increased consumption of fruit punch was also associated with greater weight gain compared with decreased consumption. After adjustment for potential confounders, women consuming 1 or more sugar-sweetened soft drinks per day had a relative risk [RR] of type 2 diabetes of 1.83 (95% confidence interval [CI], 1.42-2.36; $P < .001$ for trend) compared with those who consumed less than 1 of these beverages per month. Similarly, consumption of fruit punch was associated with increased diabetes risk (RR for ≥ 1 drink per day compared with < 1 drink per month, 2.00; 95% CI, 1.33-3.03; $P = .001$).

Conclusion Higher consumption of sugar-sweetened beverages is associated with a greater magnitude of weight gain and an increased risk for development of type 2 diabetes in women, possibly by providing excessive calories and large amounts of rapidly absorbable sugars.

Featured Link

- E-mail Alerts

Article Options

- Full text
- PDF
- Send to a Friend
- Related articles in this issue
- Similar articles in this journal

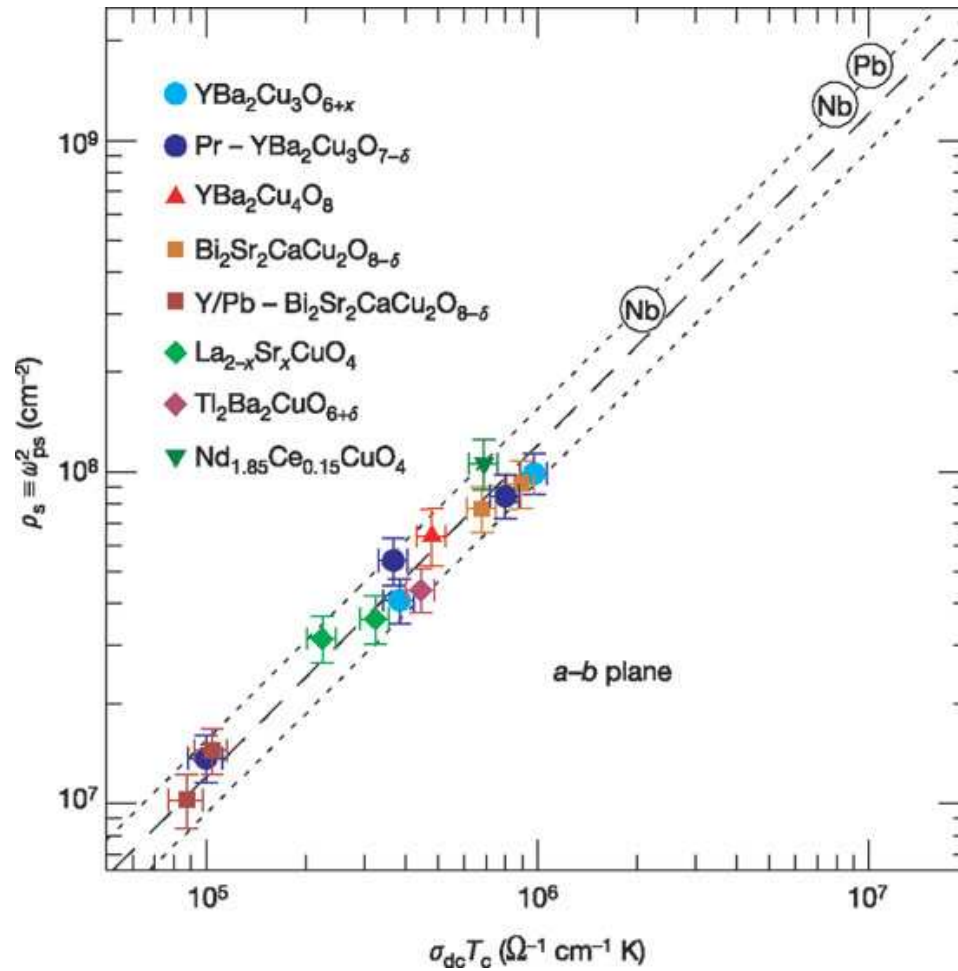
Literature Track

- Add to File Drawer
- Download to Citation Manager
- PubMed citation
- Articles in PubMed by
 - Schulze MB
 - Hu FB
- Articles that cite this article
 - ISI Web of Science (1)
- Contact me when this article is cited

Topic Collections

- Obesity
- Diabetes Mellitus
- Women's Health, Other
- Collection E-mail Alerts

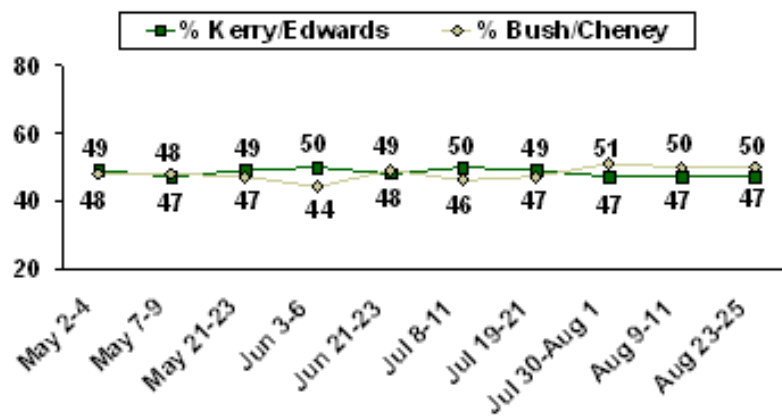
Physics:



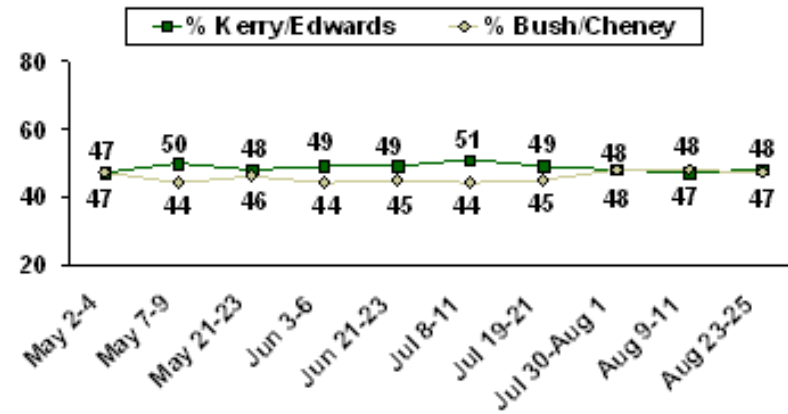
(Holmes CC et al (2004). A universal scaling relation in high-temperature superconductors. Nature **430**: 539-541.)

Politics:

**Trial Heat Trends:
John Kerry vs. George W. Bush
May 2004 – August 2004
among likely voters**



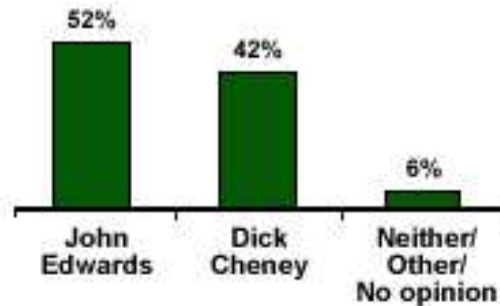
**Trial Heat Trends:
John Kerry vs. George W. Bush
May 2004 – August 2004
among registered voters**



<<http://www.gallup.com/content/?ci=12823>>

GALLUP GLANCE

Vote for Vice President?



When asked whom they would vote for if they could vote separately for vice president, 52% of registered voters say John Edwards and 42% say Dick Cheney.

If you could vote separately for vice president, would you be more likely to vote for: John Edwards, the Democrat, or Dick Cheney, the Republican?

±5 pct. pt. margin of error
Aug. 23-25, 2004
Sample size = 460
Registered voters

August 31, 2004

Overtime Rules Dispute Is a Numbers Game

By STEVEN GREENHOUSE (New York Times)



“The data isn’t there for the Labor Department to say 1.3 million workers will gain eligibility,” said John R. Fraser, the federal government’s top wage-and-hour official under the first President Bush and President Bill Clinton. “By my analysis, under the new regulations, 3, 4 or 5 million Americans could easily lose overtime coverage.”



Mr. Fraser, who headed the Wage and Hour Administration of the Labor Department in the 1990’s, estimates that 20 percent of the nation’s 72 million white-collar workers are already exempt from overtime pay, including many executives and learned professionals like lawyers.

“I would fearlessly predict that under these rules, within three years or five years, instead of 20 percent of workers being exempt, it will be 25 or 30 percent,” Mr. Fraser said.

Jay Whitehead, publisher of HRO Today, one of the leading magazines for human resources executives, agreed that many more workers would be exempted under the new rules. He said that because of the more expansive criteria, 130,000 more cooks would be exempted as well as 90,000 computer workers and 160,000 mortgage loan officers.

“The administration’s numbers are clearly just wrong or they forgot to include a bunch of people,” he said. “There’s no question that fewer people will be eligible. About 4 million people will lose eligibility, 1.3 million will gain, so the net will be 2.7 million more ineligible.”



Published on Friday, December 12, 2003

UHS Alcohol Visits Up From Last Year

By KIMBERLY A. KICENUIK Crimson Staff Writer

The number of students treated for alcohol poisoning by University Health Services (UHS) in October and November nearly doubled from the same period last year, continuing a trend of a drastic increase in alcohol-related admissions this year.

UHS admitted 22 students for alcohol-related illnesses in October and 22 in November, up from 4 last October and 19 last November.

Soon after UHS released statistics detailing a 60 percent rise in students seeking treatment for alcohol poisoning this September, the College formed a committee to investigate campus alcohol abuse.



First-Year Hangovers				
The number of first-years admitted to UHS for alcohol abuse during the first month of the school year has nearly doubled each year since 2000.				
Month	First Year	Soph.	Junior	Senior
Sept. 1998	2	1	0	0
Sept. 1999	2	1	1	0
Sept. 2000	2	5	0	1
Sept. 2001	4	4	2	2
Sept. 2002	8	3	0	4
Sept. 2003	13	4	5	2
Oct. 2003	9	4	7	2
Nov. 2003	11	3	2	6

Courtesy of UHS

CUE Guide

cueguide@fas.harvard.edu

2000		2001-2002 applications
2001	by course name by department by instructor	writer systems analyst
1999		introduction
2000		faq
		credits
1000		addenda/errata

Factoids

- There are more doughnut shops per capita in Canada than in any other country. (Snapple cap)
- The Library of Congress has 600 miles of shelves. (Snapple cap)
- Saginaw, MI has 20,000,000 gallons of wastewater treated daily (City of Saginaw website)
- South Dakota's Pierre Capital Journal reports (Mar. 1) that "an average of 650 beer cans and bottles are tossed per mile of road annually."

Given that the state "has 83,472 miles of highways, roads, and streets," this implies that approximately 54,256,800 bottles or cans every year. Given the Census Bureau's population figure of 754,844, this implies that each resident throws at least 71 bottles or cans to the side of the road every year.

(<http://www.stats.org/record.jsp?type=news&ID=74>)

Improbable Statistics (*Useful, if unverifiable information*)

Percentage of computer programmers who have ever, even once, used calculus in their work	3%
Percentage of former college English majors who have ever, even once, used calculus in their work	3%
Percentage of condoms, purchased in public restrooms, which are actually used	12%
Average number of bacteria in an hour-old hamburger	970,000,000,000
Average number of bacteria that survive for at least three days in a can of diet soda	zero
Number of transistors accidentally swallowed in 1994 by a typical employee in a microprocessor chip factory	3200
Hour at which the maximum amount of coffee is consumed each weekday in Prague	10 p.m.
Hour at which the maximum amount of coffee is consumed each weekday in New York	10 a.m.

(Annals of Improbable Research, Volume 2, Number 5)

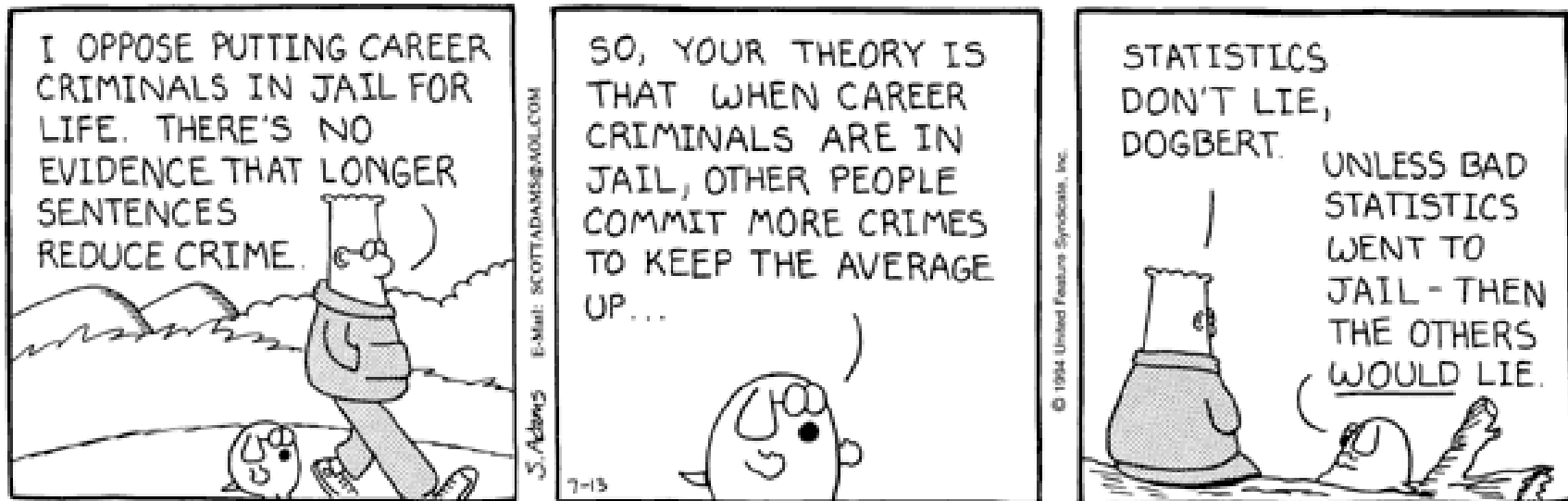
Plausible Statistics (*Improbable information from unknown sources*)

Average number of glass fragments (1 square cm or larger) in a broken window	26
Average number of glass fragments (1 square cm or larger) in a container of grain carried on the freighter <i>Thronebury</i> in August, 1995	13
Percentage of Japanese citizens who, during the course of a year, misspell their names at least once	23%
Percentage of US citizens who, during the course of a year, misspell their names at least once	26%
Average number of chews made by an adult woman while eating dinner	1610
Average number of chews made by an adult woman while eating dinner	412
Number of beakers stolen from the undergraduate chemistry laboratory at the University of Leeds on September 28, 1994	3
Number of beakers stolen from the undergraduate chemistry laboratory at the University of Leeds on September 29, 1994	4

What Are We Making Small Talk About?

- 14% How new water cooler differs from old one
- 19% Giant warehouse fire across the street
- 22% Today's small-talking points
- 11% Whitey
- 16% That baby-drowning thing from last month
- 18% Oh, this and that





© UFS, Inc.

“There are three kinds of lies: lies, damned lies, and statistics.”
Attributed to Disraeli in Mark Twain’s autobiography

“Computers are useless. They can only give you answers.”
Pablo Picasso

“Statistics are no substitute for judgment”
Henry Clay

What is Statistics?

1. Descriptive Statistics

The use of graphical and numerical summaries of data to describe the results of an experiment or an observational study.

Examples:

- How much do Harvard graduates earn?
- How many eggs do Humpback turtles lay?
- How popular is each presidential candidate?

2. Inferential Statistics

The use of a sample to draw a conclusion about a population

- Do Harvard male graduates earn more than female graduates?
- Are the Massachusetts lotteries fair?
- Should the Challenger shuttle have launched?

Example: Illinois Lotteries

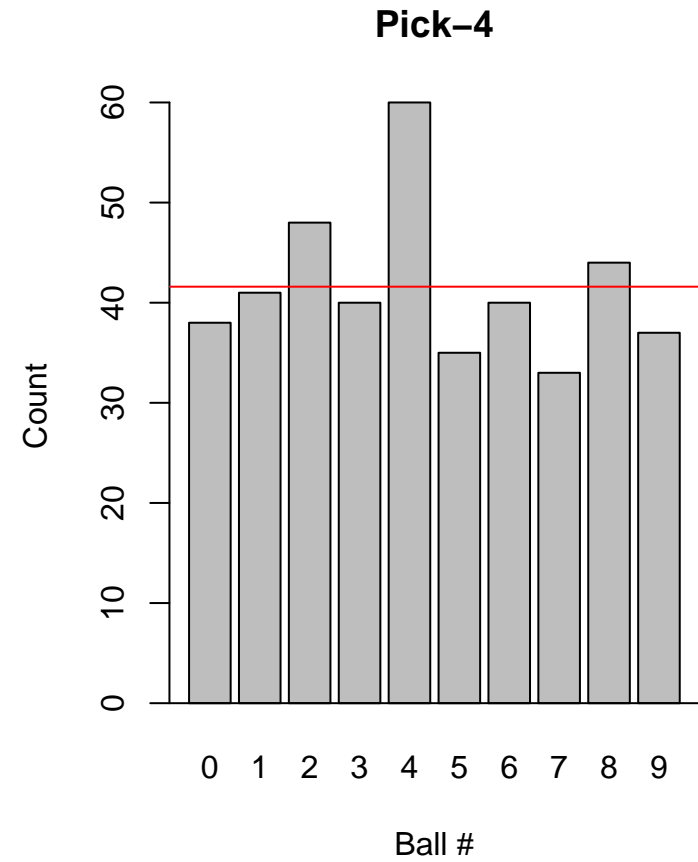
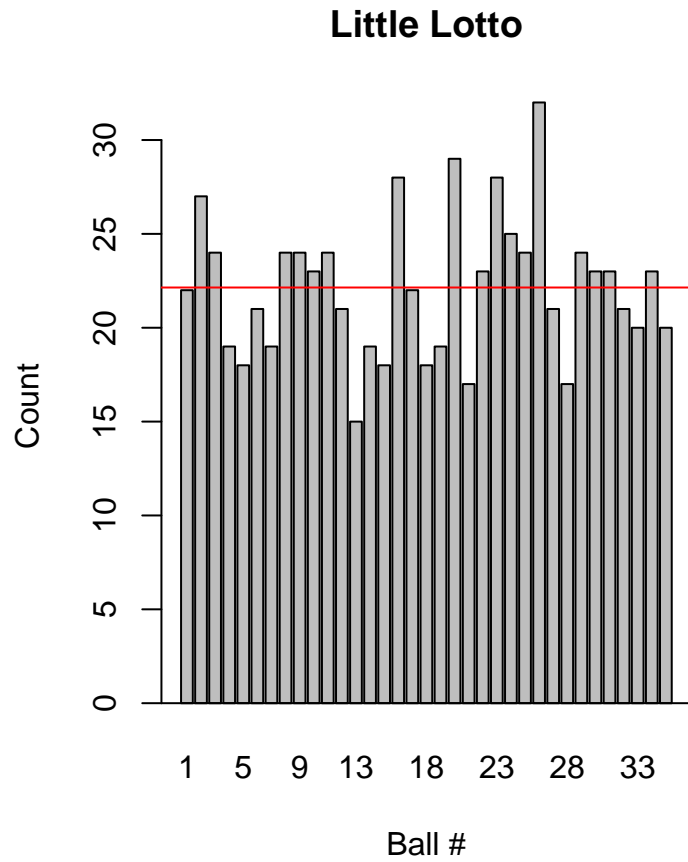
1. Little Lotto

- Balls numbered 1 to 35
- Draw 5

2. Pick-4

- 4 sets of balls numbered 0 to 9
- Draw 1 ball from each set
- Similar to Massachusetts the Numbers Game

	Little Lotto	Pick-4
Games	155	104
Expected	22.14	41.6
Max Picked	32 (ball 26)	60 (ball 4)
Min Picked	15 (ball 13)	33 (ball 7)



	Little Lotto	Pick-4
χ^2	21.6	13.0
p -value	0.95	0.16

Example: Challenger Disaster

- 25th shuttle launch - January 28, 1986
- night before launch had meeting to decide whether to go ahead with launch as scheduled
- predicted temperature at launch time: 31°F.
- lowest previous launch temperature: 53°F.
- one concern was the effect of temperature on the O-Rings of the solid rocket motors
- during meeting looked at data of previous flights on temperature and O-Ring failure

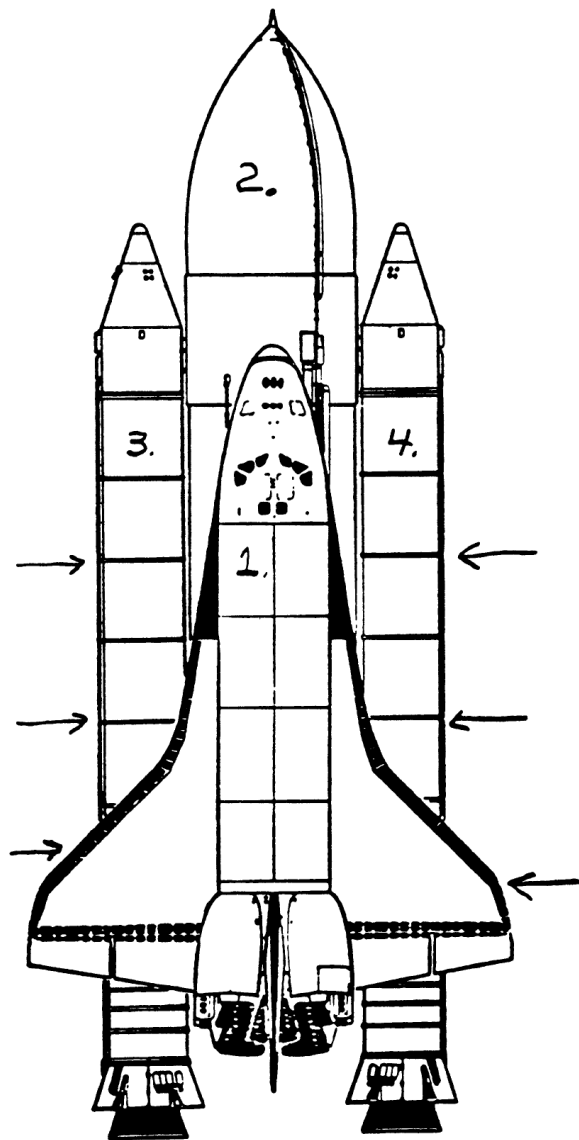


Figure 2. Space Shuttle: Orbiter, External Tank, Solid Rocket Motors, and Field Joints.

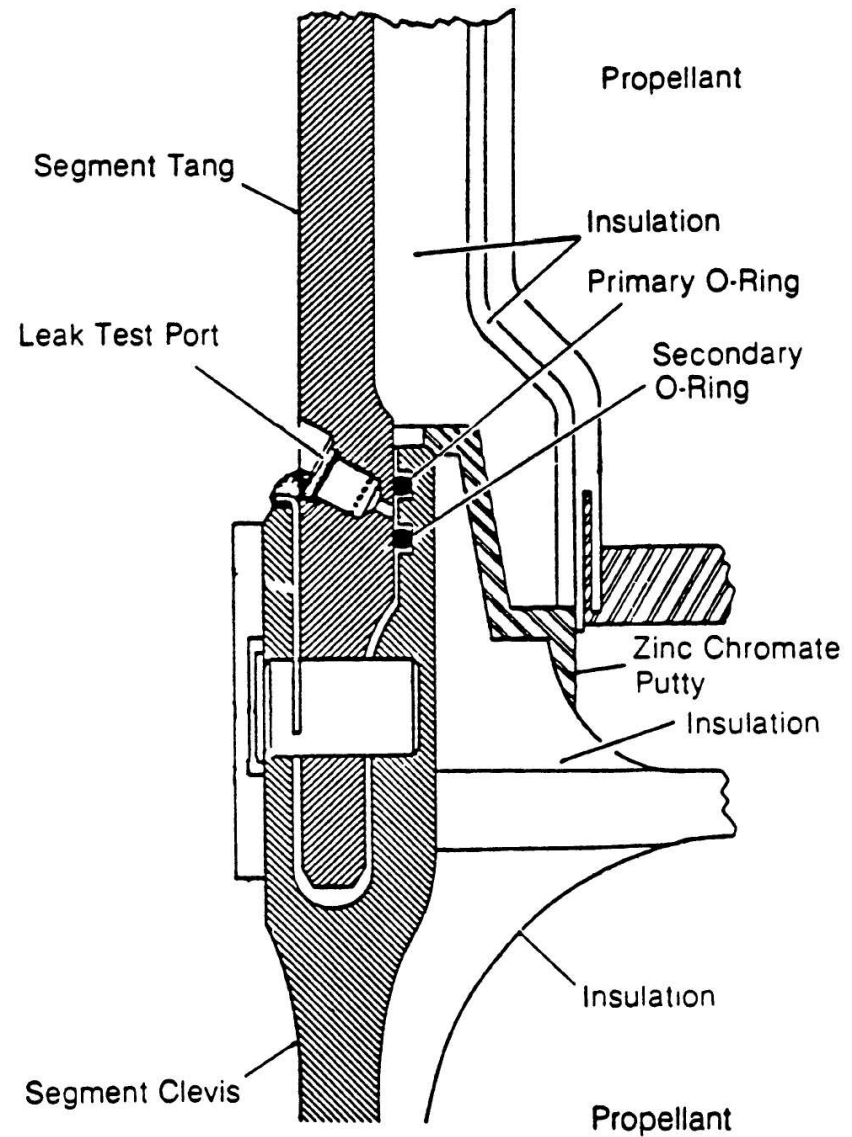
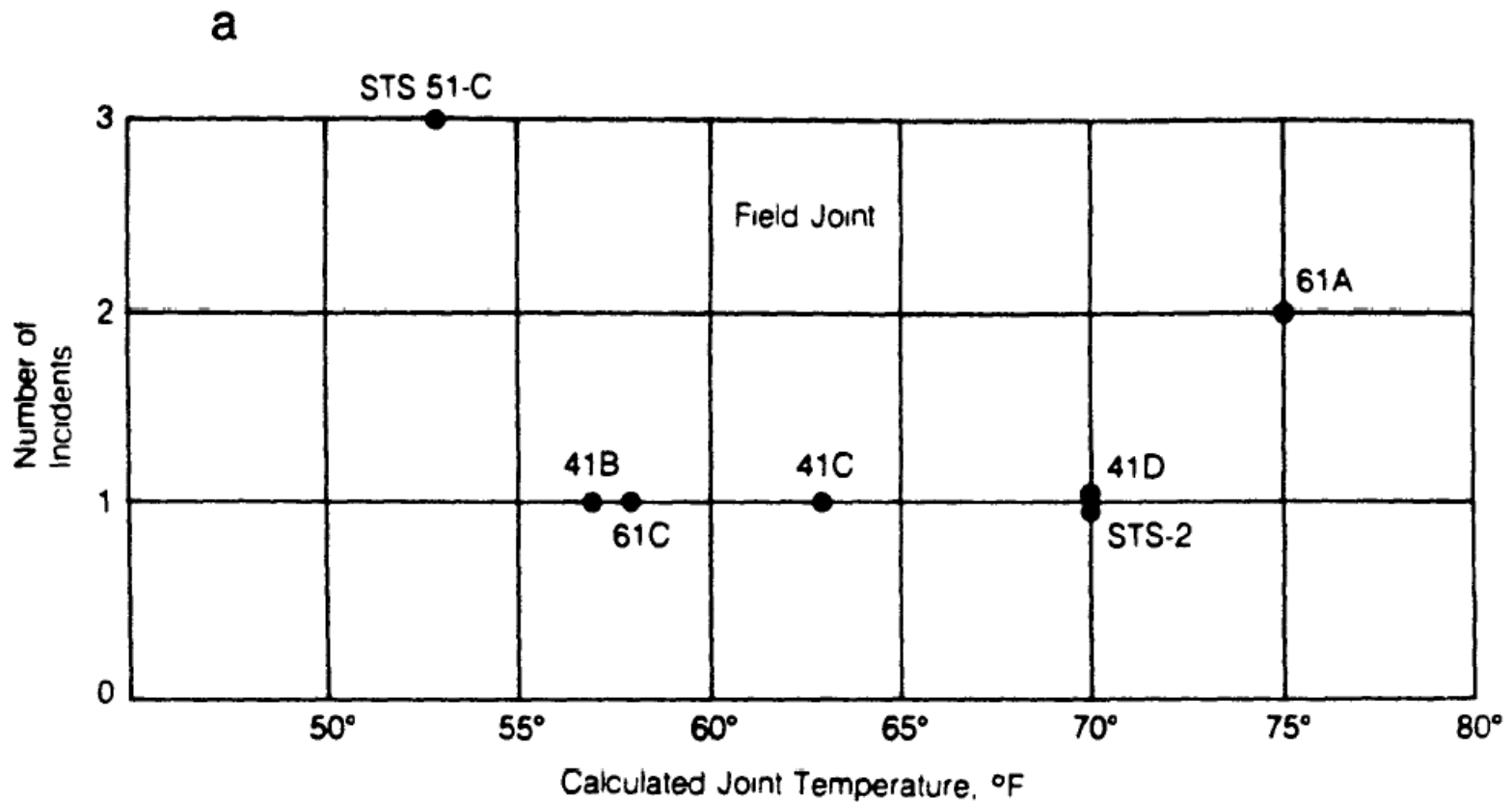


Figure 3. Solid Rocket Motor Cross Section: Tang, Clevis, and O-Rings.

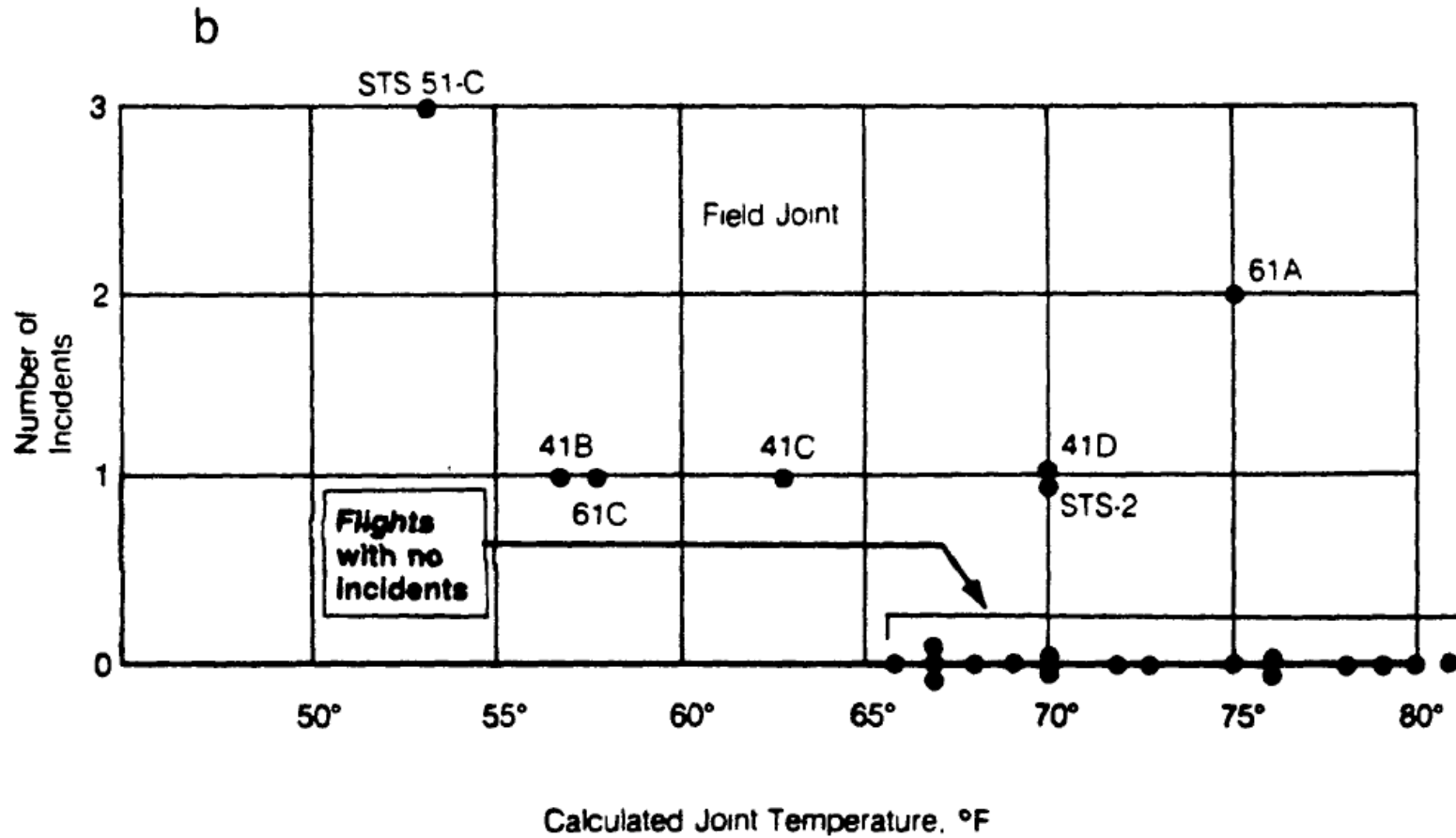
1. Observational Study - participants looked at past data. Factors such as launch temperature were not controlled
2. Wanted to find relationship between launch temperature and number of failures
3. For a given temperature, what different number of O-Ring problems could occur
4. Is it safe to launch?



Based on plot (a), they concluded

“Temperature data are not conclusive on predicting primary O-ring blowby”

Should have considered plot (b), which had data on 23 of the previous 24 flights.



Analysis suggested

- 4 or 5 blowbys should have been expected.
- Probability of at least one blowby around 67%

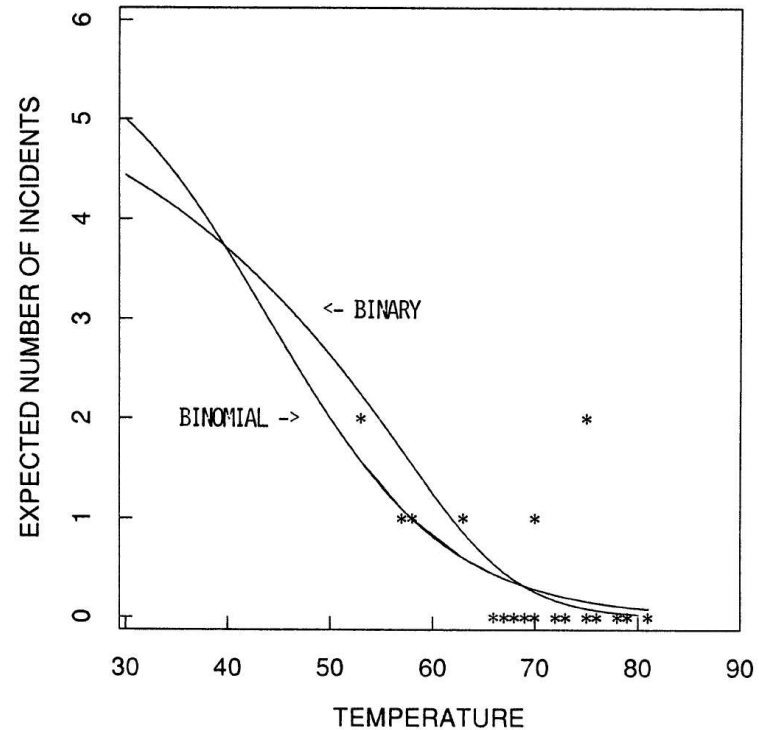


Figure 4. O-Ring Thermal-Distress Data: Field-Joint Primary O-Rings, Binomial-Logit Model, and Binary-Logit Model.

Figures taken from Dalal, Fowkes, and Hoadley (1989), Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure. JASA **84**: 945-957.