PSC 202
SYRACUSE UNIVERSITY

INTRODUCTION TO POLITICAL ANALYSIS

EXAM REVIEW, HYPOTHESIS TESTING WHEN USING SAMPLES, PART 2

TODAY

- Exam review
- More on hypothesis testing with samples

TODAY

- Exam review
- More on hypothesis testing with samples

EXAM

- Monday: Exam #2
  - Bring a calculator (no phone etc.)
  - Allowed to bring one single-page letter-size (8.5x11) sheet with you. Front page only. What you put on it is up to you, but it has to be your own sheet (we'll collect it)
- If you take exam at CDR, please sign up now!

- No new problem set this week
  - Problem set 7 due on Friday
  - No quiz for Monday

STUDENT HOURS

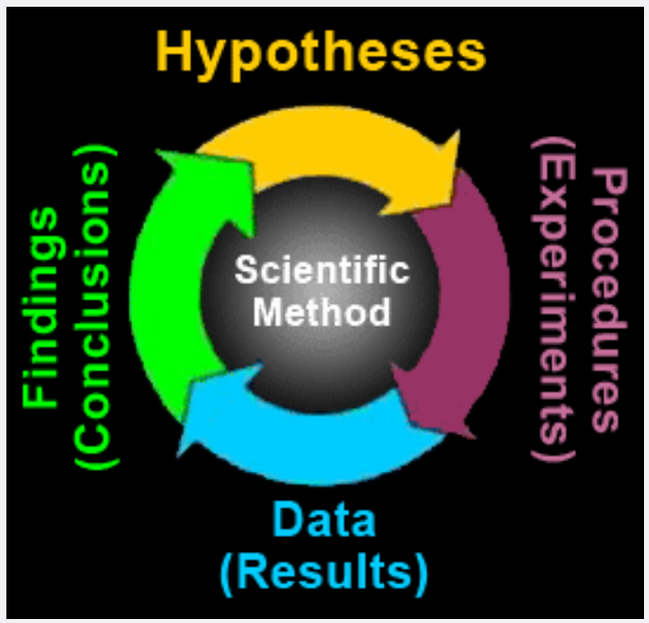
- **Next Monday: 11-1**
- 530 Eggers or Zoom
  - Zoom info on syllabus

EXAM

- Material covered
  - Everything from Feb 15 (slides: 09-sampling_2) to Mar 27 (17-hyptest_sample_1)

RESEARCH PROCESS

- Formulate research question
- Propose explanation/theory, hypotheses
- Data collection process
- Use data to evaluate hypotheses
- Reassess explanation



STUDY GUIDE

- Random sampling error
  - What is the standard error? Where does it come from and how can we compute it?
  - What is a 95% confidence interval? How can we compute it? How do we interpret it?

RECAP

$$\mathrm{SE} = \frac{s}{\sqrt{n}}$$

- SE: Standard error of the sample mean
  - A measure of how much random sampling error we have
- s: sample standard deviation
- n: sample size

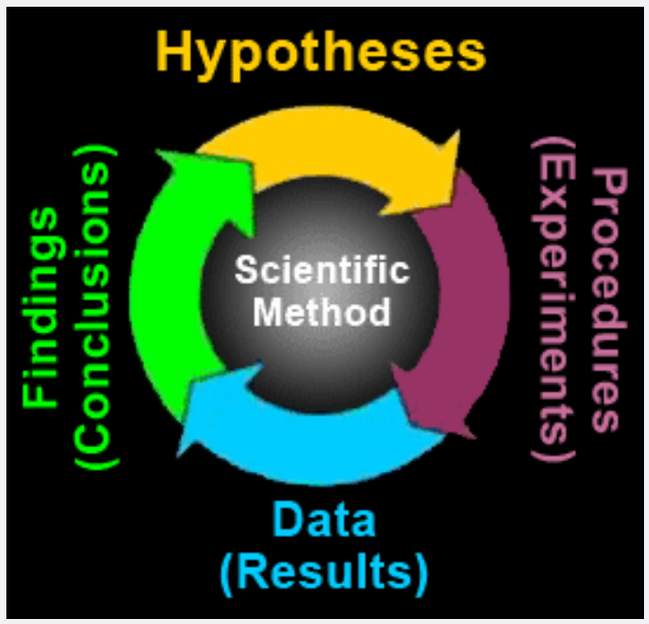
RECAP

$$95\%\ \text{CI} = \bar{x} \pm (1.96 \times \text{SE})$$

- We draw a large number of random samples from population
- Do the confidence interval for each
- 95% of those intervals will contain the population mean μ

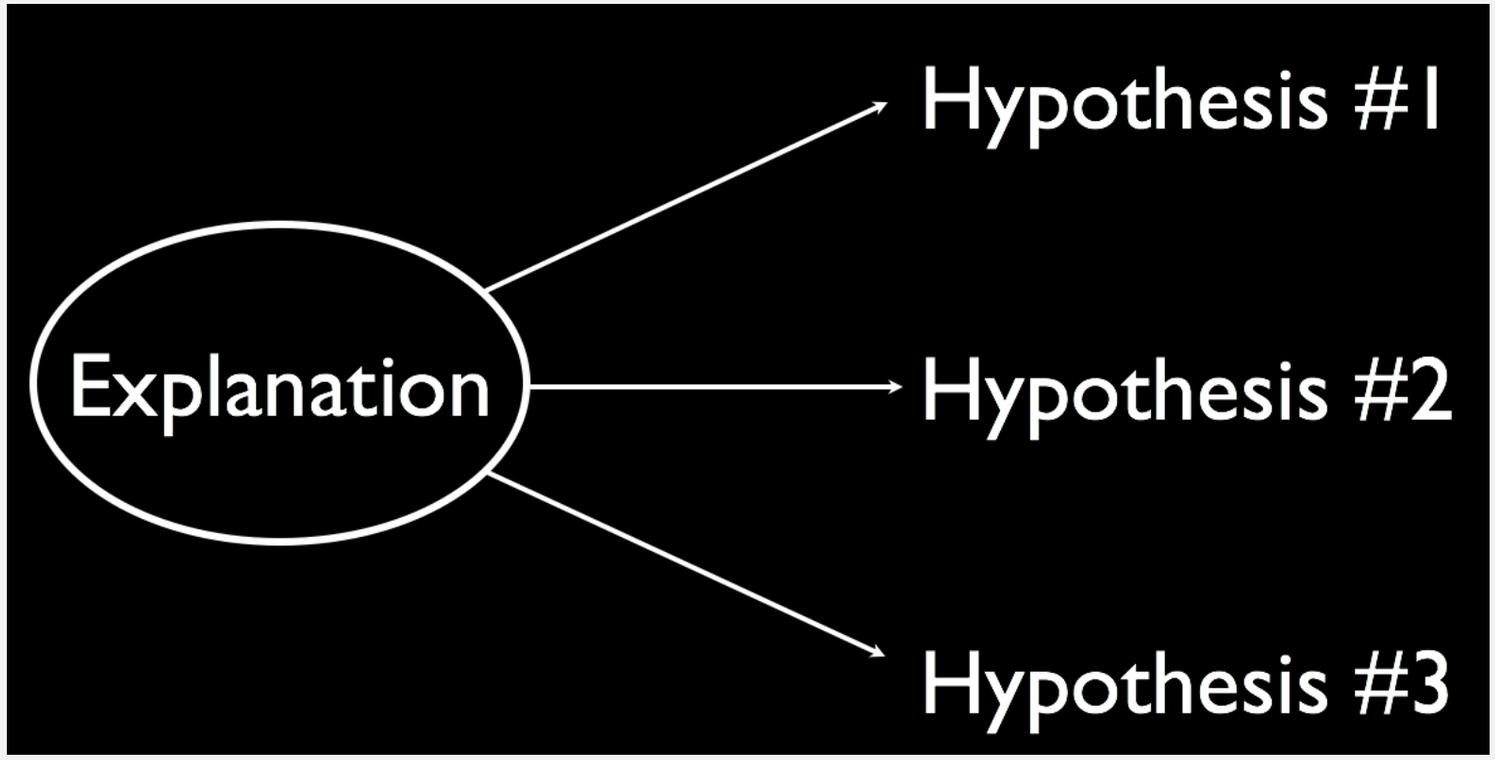
STUDY GUIDE

- Formulate research question
- Propose explanation/theory, hypotheses
- Data collection process
- Use data to evaluate hypotheses
- Reassess explanation



HYPOTHESES AND THEORY

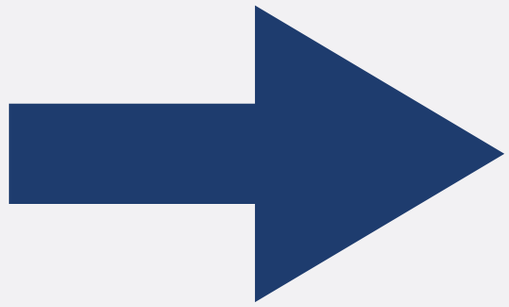
- **Explanation/Theory: (Simplified) description of how social reality works**
- **Hypotheses: Statements what, if the theory is true, we should observe in *our* data**



GOOD HYPOTHESES

- Involves two variables
  - dependent and independent variable
- Relationship between the variables is clearly specified and measurable
- Unit of analysis is clear
- Hypothesis is testable
  - falsifiable

CAUSALITY

Independent variable  Dependent variable

- Most of our theories: relationship between a single cause (independent variable) and a single effect (dependent variable)
- simple“bivariate” relationship (involves 2 variables)

HURDLES TO CAUSALITY

- Is there a credible causal mechanism that connects X to Y?
- Can we rule out the possibility that Y could cause X?
- Is there covariation between X and Y?
- Have we controlled for all confounding variables (Z) that might make the association between X and Y spurious?

EMPIRICAL STUDIES

- Two ways to do empirical studies:
- Qualitative, small N
- Quantitative, large N
- N=number of observations

QUALITATIVE STUDIES

- Talked about two forms of qualitative studies
  - Case study
  - Comparative case study

CASE STUDY

- Key technique: “Process tracing”
- Method to identify the causal relationship in a particular case though detailed examination of each step in the causal chain

COMPARATIVE CASE STUDY

- Method of difference
  - Cases where dependent variable is *different* between cases
  - Identify independent variable that is different among cases in the same way as DV is
- Method of agreement
  - Cases where dependent variable is *same* between cases
  - Identify independent variable that is also the same among cases

QUANT AND QUAL

- Strength and weakness of small-n studies relative to large-n studies
  - Internal validity
  - External validity

QUANT: BIVARIATE RELATIONSHIPS

Dependent Variable \ Independent Variable	Nominal/Ordinal	Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	Correlation Coefficient, Linear Regression

CROSS-TABULATIONS

Independent Variable

Dependent Variable	I V Value 1	I V Value 2	Total
D V Value 1	% In Column (# Cases)	% In Column (# Cases)	% Of Total (# In Row)
D V Value 2	% In Column (# Cases)	% In Column (# Cases)	% Of Total (# In Row)
Total	100% (# In Column)	100% (# In Column)	100% (# Total)

CROSS-TABULATIONS

Approve of Biden	Gender: Male	Gender: Female	Total
Approve	43.2% (19)	60.3% (38)	53.2% (57)
Do Not Approve	56.8% (25)	39.7% (25)	46.7% (50)
Total	100% (44)	100% (63)	100% (107)

TERMINOLOGY

- Zero-order relationship: relationship between two variables, without controlling for any other factors

ZERO-ORDER RELATIONSHIP

Gender

Approve of Biden	Male	Female	Total
Approve	43.2% (19)	60.3% (38)	53.2% (57)
Do Not Approve	56.8% (25)	39.7% (25)	46.7% (50)
Total	100% (44)	100% (63)	100% (107)

17.1%

BIVARIATE RELATIONSHIPS

Dependent Variable	Independent Variable: Nominal/Ordinal	Independent Variable: Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	Correlation Coefficient, Linear Regression

MEAN COMPARISON TABLE

	Average of DV	Frequency
IV Value 1	Mean of DV for IV Value 1	# Cases IV Value 1
IV Value 2	Mean of DV for IV Value 2	# Cases IV Value 2
Total	Mean of DV overall	# Cases overall

DEMOCRATIC PARTY

	Mean Thermometer Score	Frequency
Female	62.7	79
Male	43.9	50
Total	55.1	129

ZERO-ORDER RELATIONSHIP

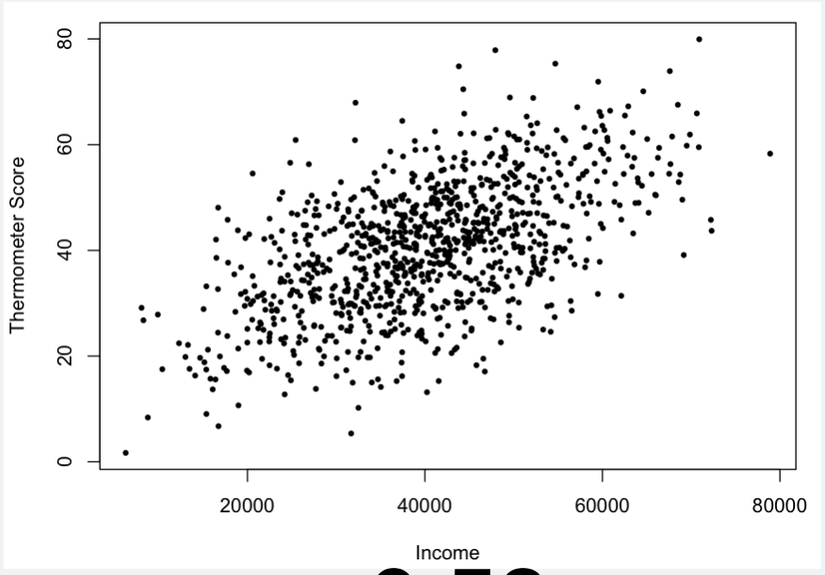
	Mean Thermometer Score	Frequency
Female	62.7	79
Male	43.9	50
Total	55.1	129

18.8

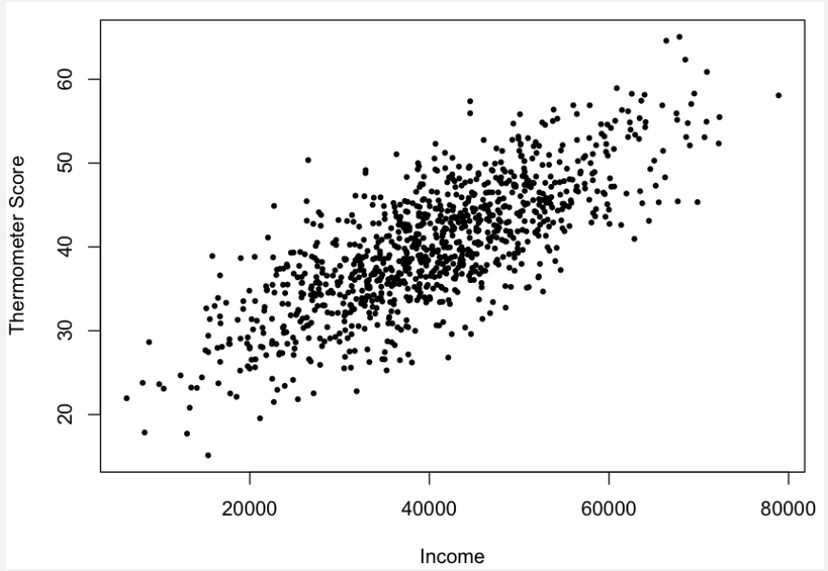
BIVARIATE RELATIONSHIPS

Dependent Variable	Independent Variable: Nominal/Ordinal	Independent Variable: Interval
Nominal/Ordinal	Cross-Tabulation	Not In This Class…
Interval	Mean Comparison	Correlation Coefficient, Linear Regression

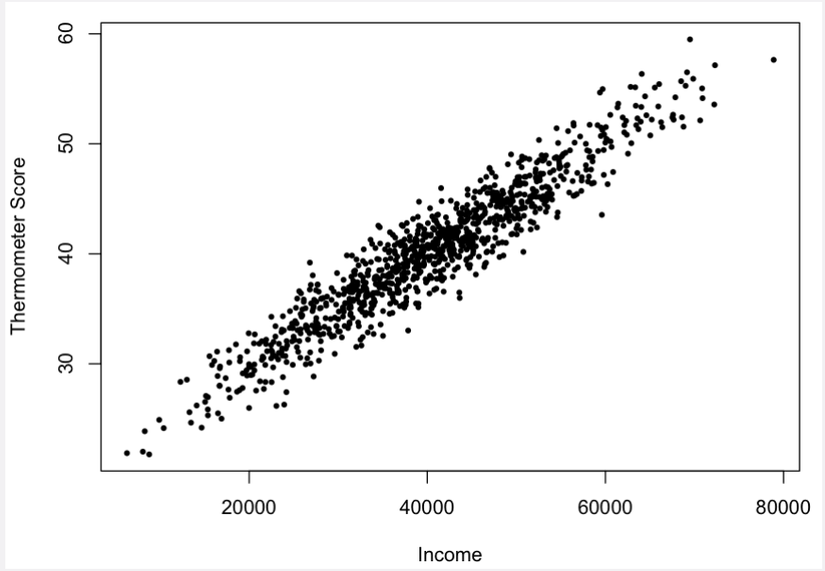
CORRELATION



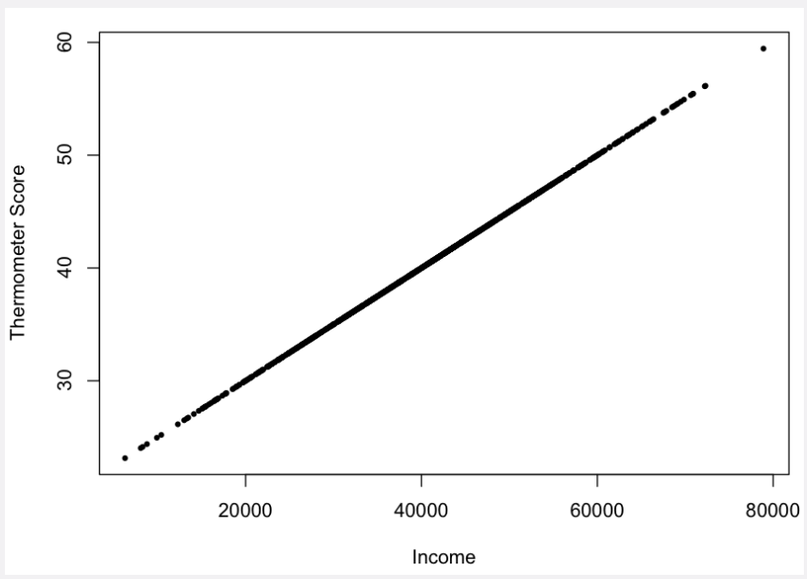
r=0.53



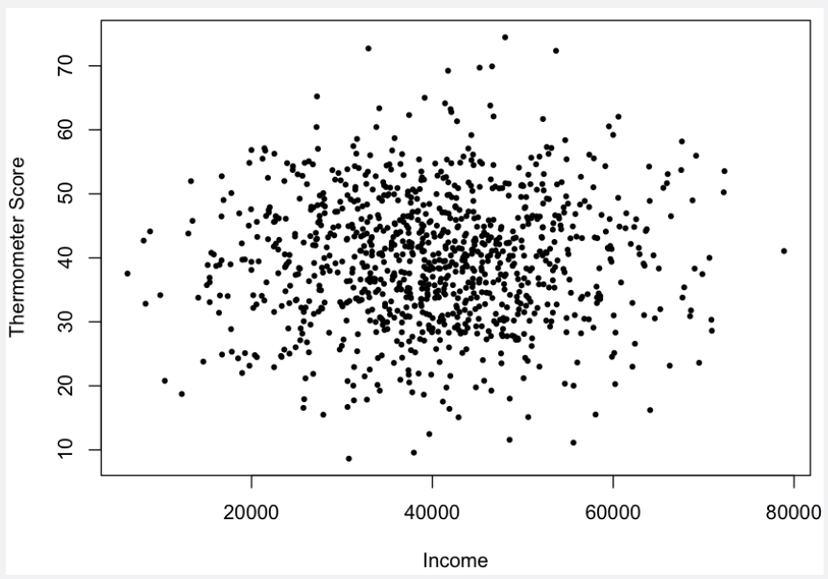
r=0.77



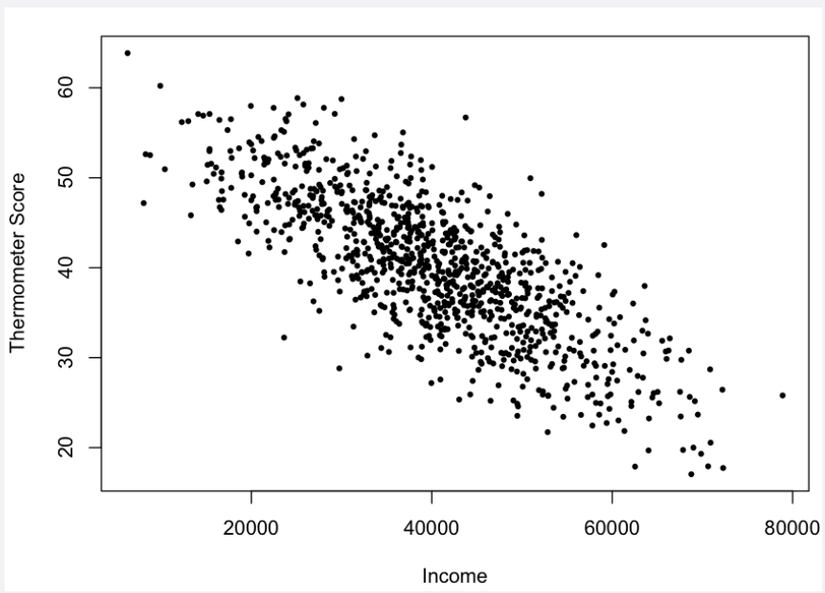
r=0.95



r=1

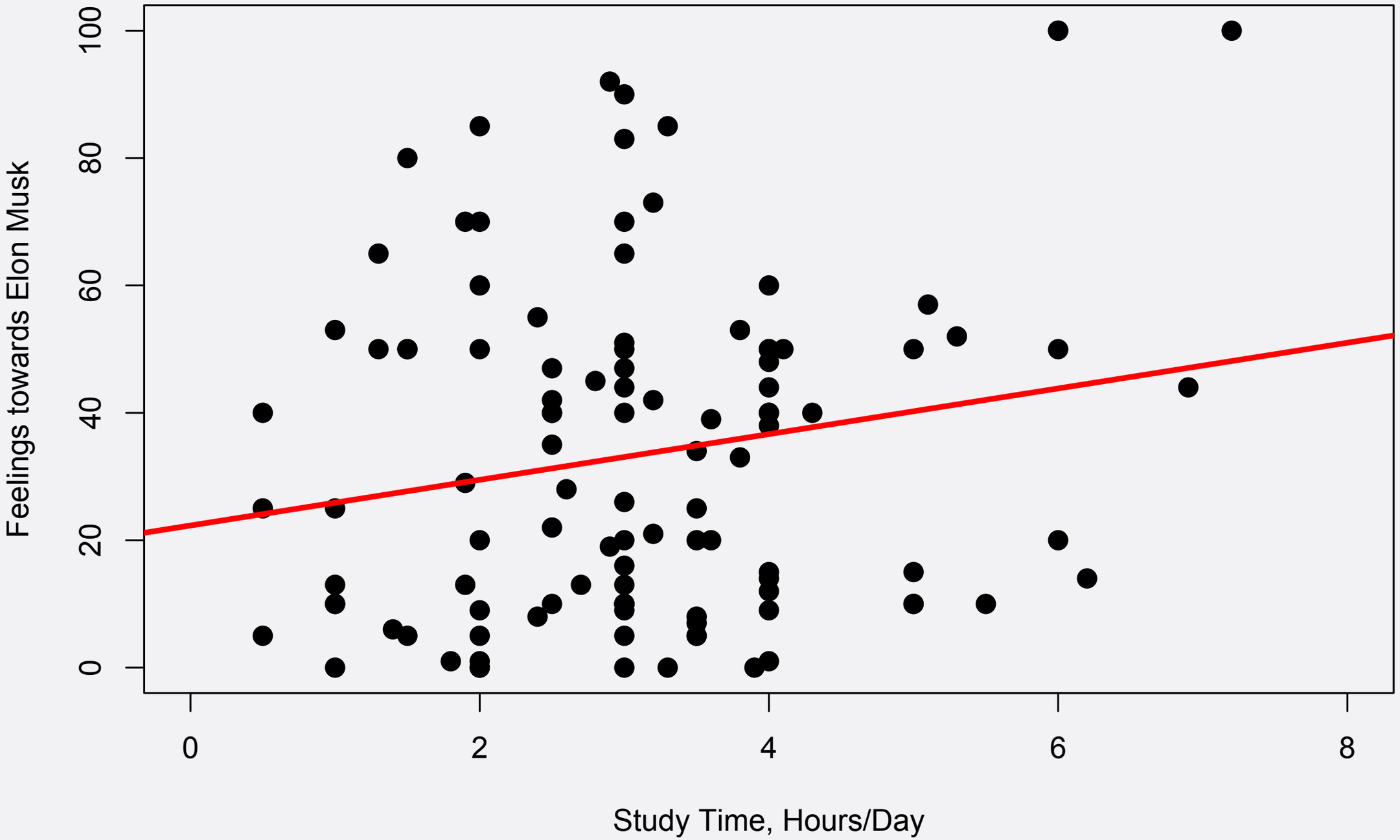


r=0

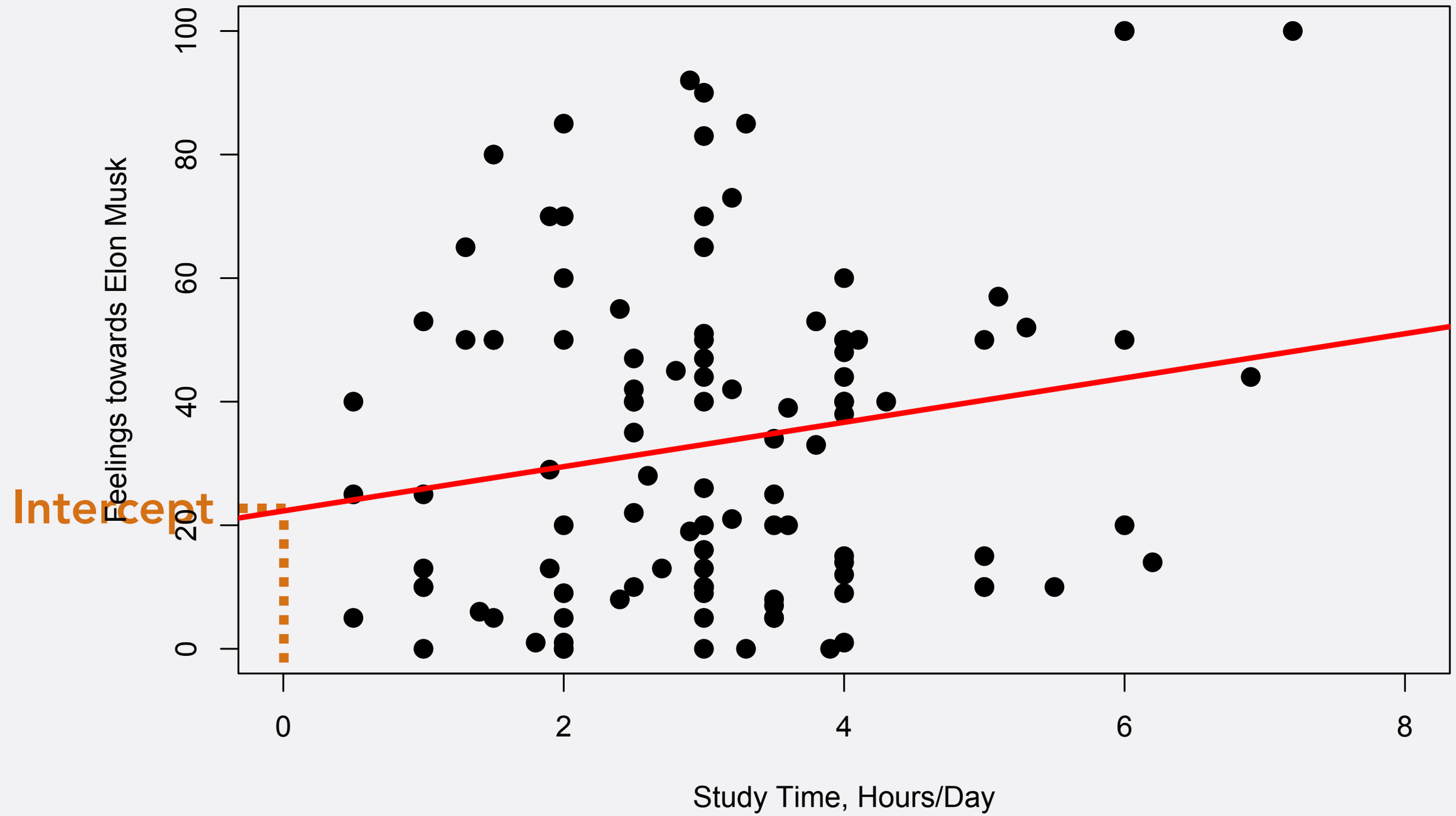


r=-0.77

REGRESSION LINE

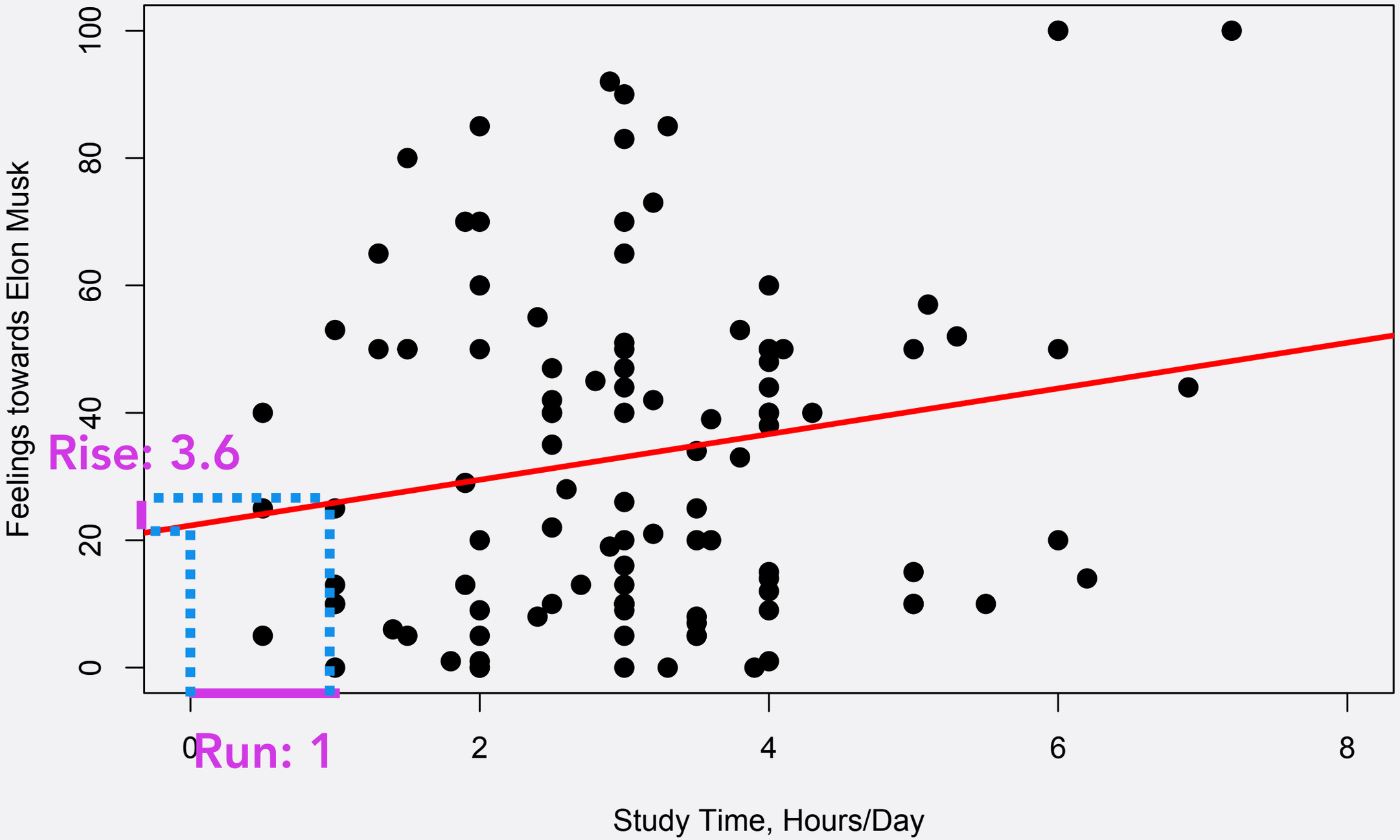


REGRESSION LINE



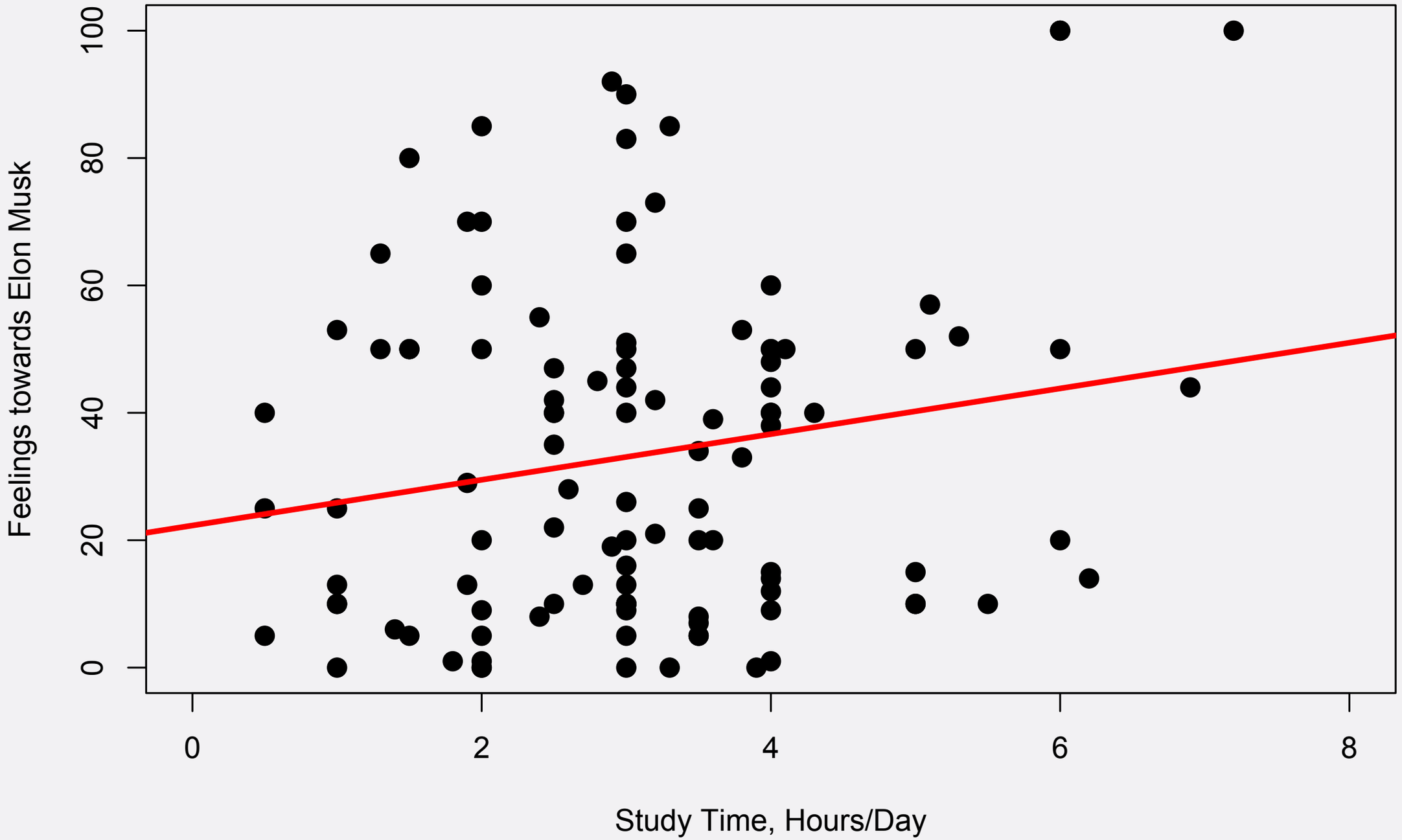
- Intercept: 22

REGRESSION LINE



- Slope = Rise over run = 3.6/1 = 3.6

REGRESSION LINE



- Thermometer Score = 22 + 3.6 * Hours/Day

REGRESSION EQUATION

- Thermometer score = 22 + 3.6 * Hours/Day
- General form: y = a + b * x
- y: dependent variable
- a: intercept
- b: slope
- x: independent variable

SLOPE

- y = a + b * x
  - Interpretation of slope: For every one unit increase in x, y changes by b units
  - Interpretation of intercept: When x=0, y takes the value a

EXERCISE

- Thermometer Score = 22 + 3.6 * Hours/Day
- What is the expected thermometer score of someone who studies 6 hours per day?
  - Solution will be on last slides (don’t peek)

EXERCISE

- Income/year = 10000 + 2000 * Years of Education

- What does the 10000 tell us?
- What does the 2000 tell us?
- How much money is someone with 16 years of education expected to earn?
  - Solution on last slides

R-SQUARE

- R^2 tells us how much variation of the dependent variable is explained by the independent variable (in a linear regression)

USING SAMPLES

- Bivariate relationship between two variables in sample
- Is this a real relationship that we would find in the population as well, or is it something that only shows up in our sample?

HYPOTHESIS

- H_0: In the population, there is *no relationship* between dependent and independent variable
  - If there is a difference in the sample, it is due to random sampling error
- H_A: There *is* a relationship between the independent and dependent variable in the population

ERRORS

	There Is A Relation In The Population	There Is No Relation In The Population
We Conclude There Is A Relation	✔	✘ Type I
We Conclude There Is No Relation	✘ Type II	✔

ERRORS

	There Is A Relation In The Population	There Is No Relation In The Population
We Conclude There Is A Relation	✔	✘ Type I in at most 5% of cases
We Conclude There Is No Relation	✘ Type II	✔

IDEA

- We start out thinking H_0 is true
  - No relationship between X and Y in population
- We ask: If H_0 is true, how likely is it that a random sample would produce an effect as large (or larger) than the one we have observed in our sample?
  - If less than 5% ($p<0.05$): we reject H_0
  - If more than 5% ($p>0.05$): we don't reject H_0

EXAM

- What to do and bring
  - Be a few minutes early
  - Bring calculator and pen/pencil
  - Bring your cheat sheet
  - Show your steps
  - Remember time management
  - Write legibly

STUDY GUIDE

- Questions?

EXERCISE SOLUTION

- Thermometer Score = 22 + 3.6 * 6 = 43.6

EXERCISE SOLUTION

- Someone with 0 years of education is expected to have an annual income of $10,000
- For every additional year of education, annual income is expected to increase by $2,000
- Income/year = 10000 + 2000 * 16 = 42000