Today:

a) a last look at X^2

b) LCQ on Correlation

_____

Have your Pink X^2
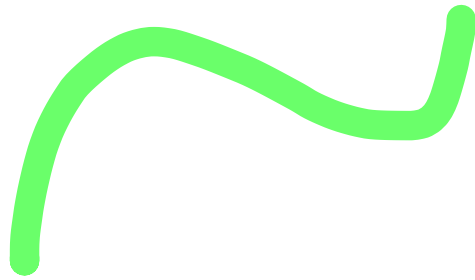packet available

HW
Questions
?

●

check
HW ⟶ along
with me

Pick up the Warm Up

You'll need your Pink Chi-Square packet from yesterday as a reference

do #1 and #2 only
for now

Warm Up

Use your notes as needed.

Practicing Using the Chi-Square Test of Independence

(5-1)

	red	green	blue	black	silve
Male					
female					

(2-1)

1. A researcher consulted 500 men and women to see if the colour of the car they drove was independent of gender. The colours were red, green, blue, black, and silver. A χ^2 test was conducted at the 5% significance level and the value found was 8.73.

$\leftarrow \chi^2_{calc}$

a. Write down the null hypothesis

H_0: **Color is independent of gender** ✓

b. Find the number of degrees of freedom. $(2-1)(5-1) = 4$

c. Write down the critical value for this test. $9.488 \leftarrow$

d. Is car colour independent of gender? Give a clear reason for your answer.

Yes

χ^2 of $8.73 < 9.488$

---

2. Suppose a similar test was conducted on a different population of 300 people. A 5% significance level is used. It was found that the p-value was 0.04. Is colour independent of gender ?

4 %

$P = .04$

Since the p-value < 0.05, we must reject H_0. Thus gender and car color preference are associated.

Aim for today

The limitations of the x^2 Test of Independence

A. Not enough data ✓

B. 2×2 adjustment
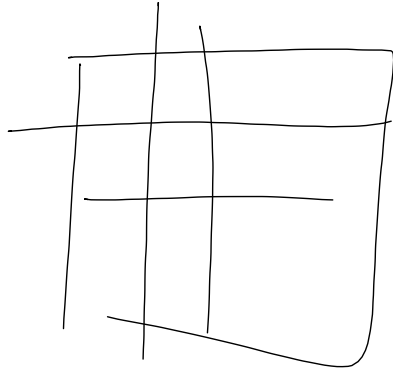

3.

Consider the contingency table alongside:

a Construct the expected frequency table. ✓

b Are any of the expected frequencies less than 5?

c Combine the data so that none of the cells have an expected frequency less than 5.

	Own a pet?	
Age	Yes	No
0 - 19	5	3
20 - 29	2	22
30 - 49	42	58
50+	39	34

a)

	Yes	No
0 - 19	4	3
20 - 29		
30 - 49		
50+		

Age	Yes	No
0 - 29	37	25
30 - 49	42	58
50+	39	34

In a 2 by 2 contingency table:

-- Yate's continuity correction must be used when calculating X^2

If $df = 1$, we use

$$\chi^2_{calc} = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}$$

where $|f_o - f_e|$ is the **absolute value** or **modulus** of $f_o - f_e$

The following table shows the results from a random sample carried out so that the question about the relationship between education and job satisfaction could be analysed.

Completed University

		YES	NO	
Satisfied in job	YES	272	618	890
	NO	238	292	530
		510	910	1420

1. Calcuate the expected freq

2. Set up a table to organize.

---

f_o	f_e	$|f_o - f_e| - .5$	$\dfrac{(|f_o - f_e| - 0.5)^2}{f_e}$
272	319.65		
618	570.35		
238	190.35		
292	339.65		

$$\chi^2_{calc} = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}$$

$$\chi^2 \doteq 29.1$$

$$\chi^2_{calc} = \sum \frac{(|f_o - f_e| - 0.5)^2}{f_e}$$

B.B.

2×2

5 × 4
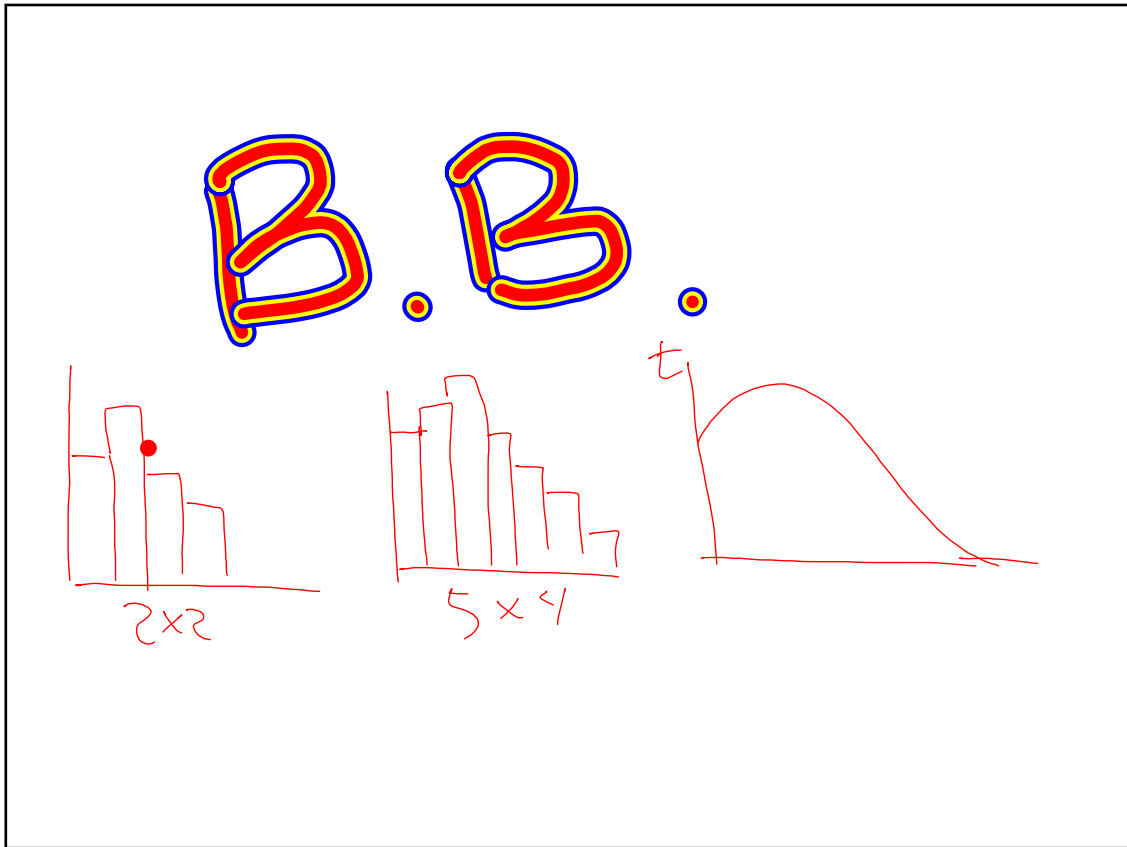
~ Tipping Point

~ Blink

TIMSS
international math study

Outliers

THE STORY OF SUCCESS

MALCOLM GLADWELL
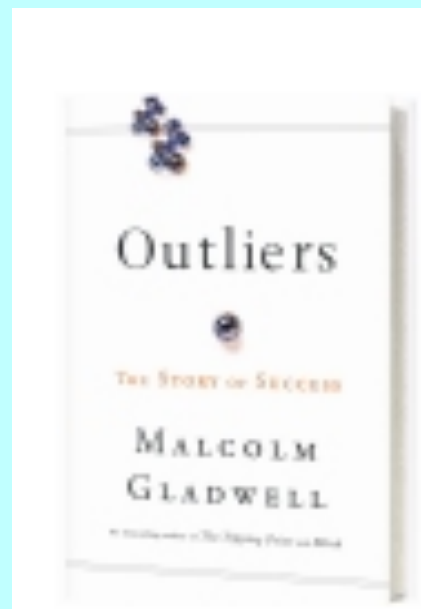
- Every 4 years
- Comprehensive test science/math
- Compare educational achievement

- Before Test - Questionnaire

Ed.
of
Parent

Pers.
Views
of
math

Income

120 questions

Brainstorm:

Think back to all of the statistical graphs, statistics, measurements so far in this course:

Correlation btwn amount spent vs length of stay

Relationship btwn occupation and amt spent
" " gender and occupation

What gender would spend more. $\frac{\bar{x}}{y}$

m
r

Coffee Shop Brainstorm

In your group :

Now that you have some statistical tools, What type of questions could be investigated and answered?

Ideas

Ideas

relationship between age & time spent
relation   between age/spent
HISTOGr / occupation
        and $

Relation between $spent vs time
Stem/leaf - Extent of age
Avg of Age in Coffee

relationship btwn occupation
vs length of
stay

Project Scoring Guide

- a student friendly version

-the acutal one is posted on the class blog if /
when you want to look at it.

Definitely get a folder to house all project materials and keep separate from the regular course materials.

LCQ on Correlation

You can use ....

Assignment

Ch.11 Packet

p.341....#2 (use the x^2 statistic) ←·—·—·—·
p.344.... #1abcd
p.348.....#4 (use probability) ←·—·—·—·

clearly show all steps

---

3. The veterinarian has gathered the following data about the weight of dogs and the weight of their puppies.

		Dog		Total
		Heavy	Light	
	Heavy	36	27	63
Puppy	Light	22	35	57
	Total	58	62	120

The veterinarian wishes to test the following hypotheses.

H_0: A puppy's weight is independent of its parent's weight.
H_1: A puppy's weight is related to the weight of its parent.

(a) The table below sets out the elements required to calculate the χ^2 value for this data.

(a) The table below sets out the elements required to calculate the χ^2 value for this data.

	f_o	f_e	$f_e - f_o$	$(f_e - f_o)^2$	$(f_e - f_o)^2 / f_e$
heavy/heavy	36	30.45	−5.55	30.8025	1.012
heavy /light	27	32.55	5.55	30.8025	0.946
light/heavy	22	27.55	5.55	30.8025	1.118
light/light	35	a	b	c	d

4.12

(i) Write down the values of a, b, c, and d.

$a = 29.45$ $b = {}^-5.55$ $c = 30.8025$ $d = 1.046$ (4)

(ii) What is the value of χ^2_{calc} for this data?

$X^2 \doteq 4.12$ by adding up values in last column (1)

(iii) How many degrees of freedom exist for the contingency table?

$df = (2-1)(2-1) = 1$ (1)

(iv) Write down the critical value of χ^2 for the 5% significance level.

3.841 (1)

(b) Should H_0 be accepted? Explain why. (2)

since $4.12 > 3.841$ we reject H_0

(Total 9 marks)

---

④ ✗

Let $x = 7.94$.

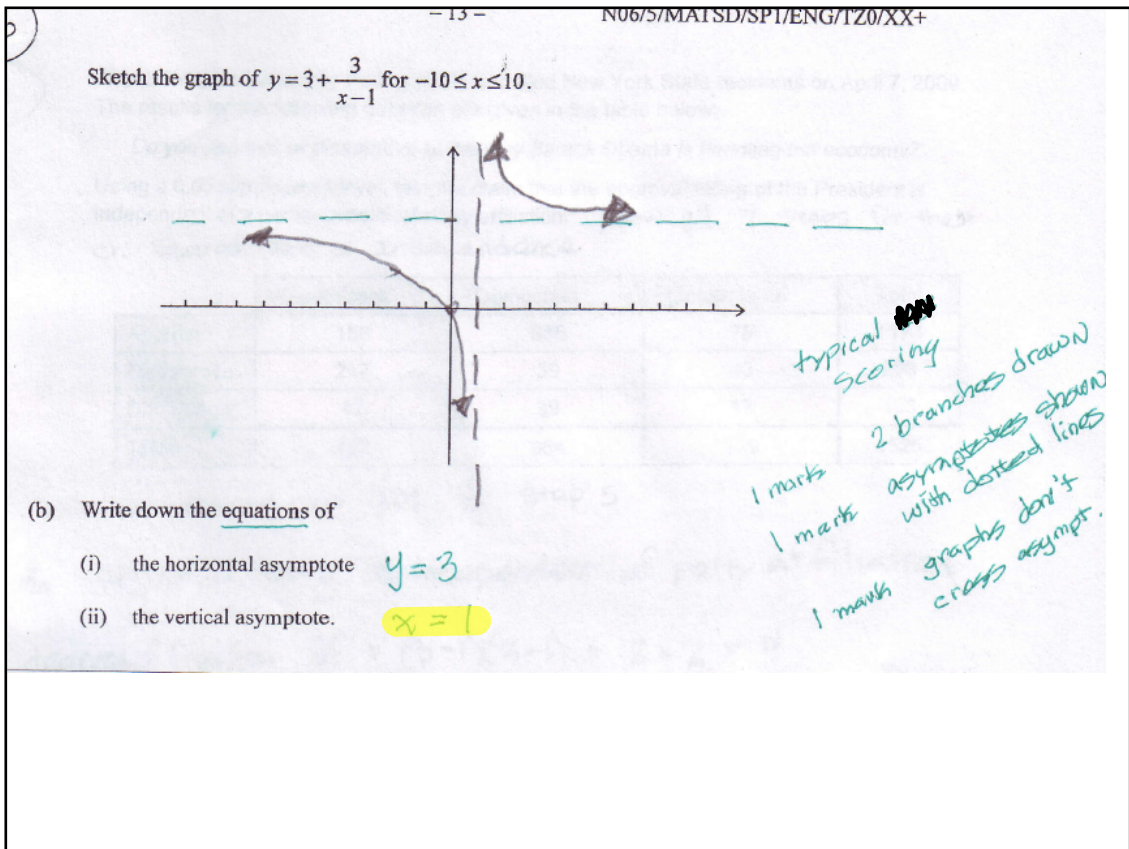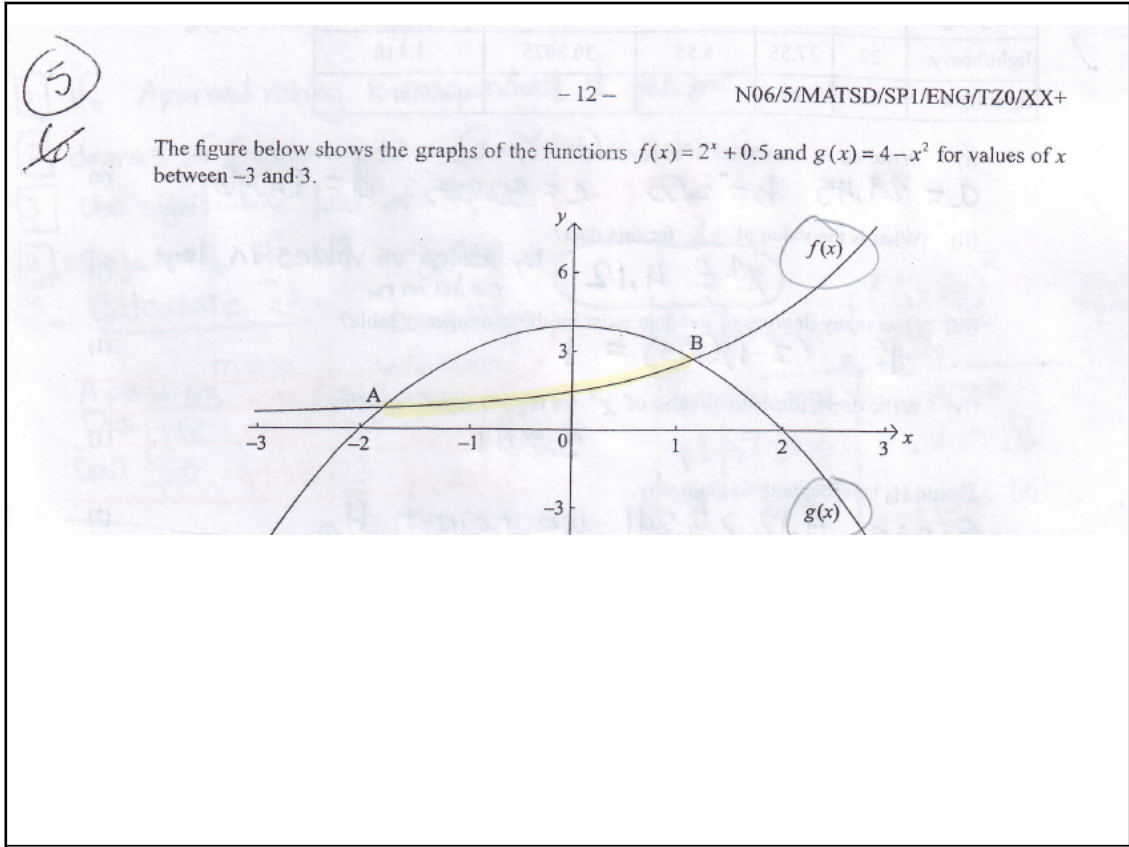(a) Calculate the value of $\dfrac{2x+1}{x^3}$.

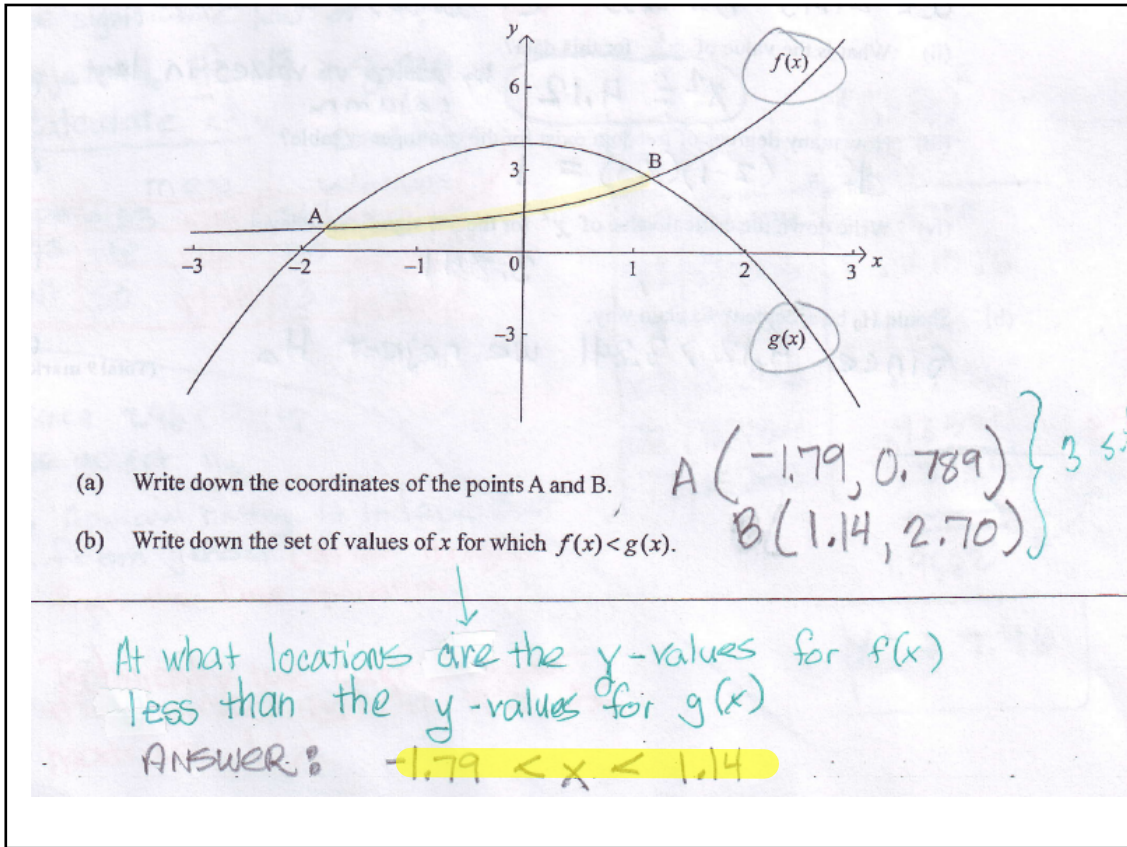(b) (i) Give your answer correct to **three** decimal places. $0.03372...$ 0.034

(ii) Write your answer to (b)(i) as a percentage. 3.4%

(c) Give your answer to part (b)(i) in the form $a \times 10^k$, where $1 \le a < 10, k \in \mathbb{Z}$. 3.4×10^{-2}

↳ this means scientific notation a.k.a "STANDARD FORM"

(5)

6

The figure below shows the graphs of the functions $f(x) = 2^x + 0.5$ and $g(x) = 4 - x^2$ for values of x between -3 and 3.

Sketch the graph of $y = 3 + \dfrac{3}{x-1}$ for $-10 \le x \le 10$.

typical ~~xxx~~
scoring

1 mark — 2 branches drawn

1 mark — asymptotes shown with dotted lines

1 mark — graphs don't cross asympt.

(b) Write down the equations of

(i)   the horizontal asymptote $y = 3$

(ii) the vertical asymptote. $x = 1$

(a) Write down the coordinates of the points A and B.

$A(-1.79, 0.789)$ } 3 s⧵

$B(1.14, 2.70)$ }

(b) Write down the set of values of x for which $f(x) < g(x)$.

At what locations are the y-values for $f(x)$
less than the y-values for $g(x)$

ANSWER: $-1.79 < x < 1.14$

A small Pennsylvania town.

Outliers

THE STORY OF SUCCESS

MALCOLM
GLADWELL
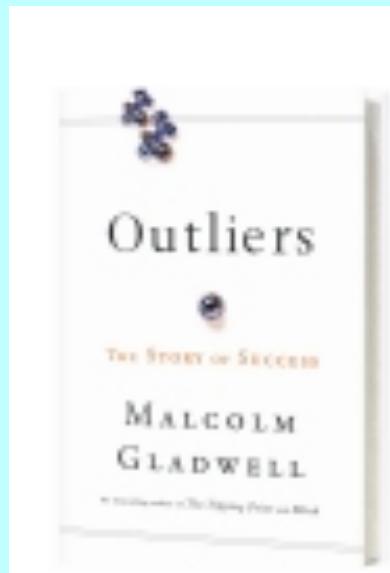
The rest of the day _today_ we will devote to learning what is involved in the course project.

- Get an overview (only)
- Hear about some past projects
- Start brainstorming for yourself.