- Pick up the solutions. Use only pens of a different color to mark your HW.
- Let me know if there are questions.

ch. 11 p. 324...5
11B.1

a) r ≈ 0.917
b) strong, positive, correlation

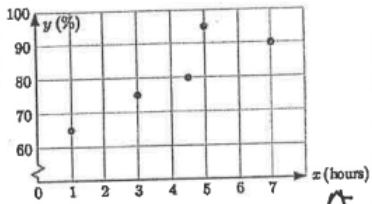
In general, the greater the young athletete's age, the further they can throw the discus

Summarizing statement!

p. 326 ... #3

11.B.2

(a)



$\bar{x} = 4.1$

$\bar{y} = 81$

(b) $r = \dfrac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \cdot \sum (y-\bar{y})^2}} = \dfrac{94.5}{\sqrt{(20.2)(570)}} \approx 0.881$

(c) strong, positive, linear correlation

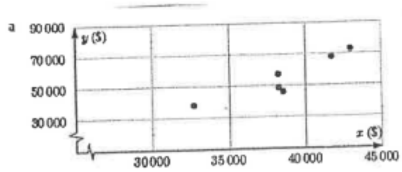
As students increase their preparation time, their grade increases

Summary Statement

p. 332 .... #1

11D

a



b $r \approx 0.921$

c There is a strong, positive, linear association between the starting salaries for Bachelor degrees and the starting salaries for PhDs.

d $y \approx 3.44x - 78300$

e i $59300

ii This is an interpolation, so the prediction is likely to be reliable.

You should be able to get this by using the formula!

Since at $40,000, the prediction would be an interpolation. Also there is a strong correlation.

p. 333.... #3
11D

a $r \approx -0.924$
b There is a strong, negative, linear correlation between the *petrol price* and the *number of customers*.
c $y \approx -4.27x + 489$
d gradient ≈ -4.27, for every 1 cent per litre increase in the price of petrol, a service station will lose 4.27 customers.
e -5.10 customers
f It is impossible to have a negative number of customers. This extrapolation is not valid.

Pick up the Warm Up

WARM UP

The probability of flipping a head when a coin is tossed is 0.5 or 50%

What is the probability of flipping two coins and getting a head on each?

Now we'll flip one coin and roll one die. What is the probability of flipping a tail and getting a six?

(1) The probability of flipping a head when a coin is tossed is 0.5 or 50%

What is the probability of flipping two coins and getting a head on each?

$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ or .25 or 25%

(2) Now we'll flip one coin and roll one die. What is the probability of flipping a tail and getting a six?

$\frac{1}{2} \bullet \frac{1}{6} = \frac{1}{12}$

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

③ The probability of being in the experimental group is ______

The probability of someone graduating in the sample is ____________

The probability of being BOTH in the experimental group and a graduate is:

$\frac{73}{167} \approx$

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

③ The probability of being in the experimental group is ______

The probability of someone graduating in the sample is ____________

The probability of being BOTH in the experimental group and a graduate is:

$\frac{73}{167} \approx 43.7\%$

if independent...

$\frac{85}{167} \cdot \frac{116}{167} \approx 35.4\%$

4

Now that we know the probability of being in both groups is 43.7%, how many students would we expect to be a graduate from the experimental group?

IB Math Studies SL

Statistical Applications

- The Normal Distribution
- Correlation
- Regression Line
- X^2 - test

Types of statistics

1. Descriptive (which *summarize some characteristic* of a sample)
   - Measures of central tendency - Mean, Median, Mode
   - Measures of dispersion - Range, Interquartile Range, Standard Deviation
   - Measures of skewness
2. Inferential (which test for significant *differences* between groups and/or significant *relationships* among variables within the sample
   - t-ratio, chi-square, beta-value

Two Variable Statistics

using pairs of numerical data

- *Scatter Plots*
- *Correlation*
- *Line of Best Fit*

using two categories of numbers.

Chi-Square Test of Independence

We don't use the *term*

"correlation"

with data that is categorical.

χ^2 (Chi-Squared) Test of Independence

next 3 days

- The chi-square test of independence is one of the most frequently used hypothesis test in the social sciences because it can be used with variables at any level of measurement.

today

Part 1
Background and Skills

I will be giving you a hand out with some partially filled notes.

Contingency Tables

		Completed University	
		YES	NO
Satisfied in job	YES	272	618
	NO	238	292

	Wore hat and sunscreen	Wore hat or sunscreen	Wore neither
Sunburnt	3	5	13
Not sunburnt	36	17	1
sum			

	Likes chicken	Dislikes chicken	sum
Likes fish			60
Dislikes fish			40
sum	75	25	100

Another Example

In the the following data set, students in grades 4-6 were asked whether good grades, athletic ability, or popularity was most important to them.

A two-way contingency table separating the students by grade and by choice of most important factor is shown :

Is there an association between the type of school area and the students' choice of good grades, athletic ability, or popularity as most important?

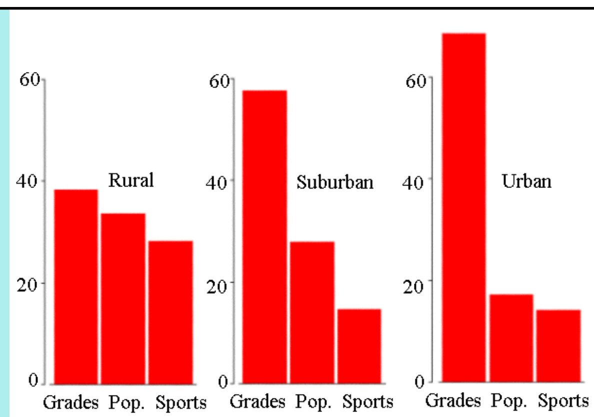
A two-way table for student goals and school area appears as follows:

Goals	School Area: Rural	Suburban	Urban	Total
Grades	57	87	24	168
Popular	50	42	6	98
Sports	42	22	5	69
Total	149	151	35	335

2

The corresponding column percentages are the following:

Goals	School Area: Rural	Suburban	Urban
Grades	38	58	69
Popular	34	28	17
Sports	28	14	14
Total	100	100	100



Sometimes these differences in can be "eyeballed" by observing split bar graphs...

From the table and corresponding graphs, *it appears* that the emphasis on grades increases as the school areas become more urban, while the emphasis on popularity decreases.

To be sure, we can use a **statistic** to measure if there is truly a relationship (or asscociation) between the two variables.

In the last example, the statistic would tell us if there is a LINK between living area and emphasis on grades.

X^2 *is a statistic that measures the difference between values that we observe in a contingency table and values that we would expect to see given the overall totals.*

Age of voter

	18 to 35	36 to 59	60+
Party A	85	95	131
Party B	168	197	173

3 Contingency tables are used to examine the relationship between two qualitative or categorical variables.

For example,
consider the hypothetical experiment on the effectiveness of early childhood intervention programs.

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

In the experimental group, 73 of 85 students graduated from high school. In the control group, only 43 of 82 students graduated.

4 The table shows that people in the experimental condition were more likely to graduate than were subjects in the control condition.

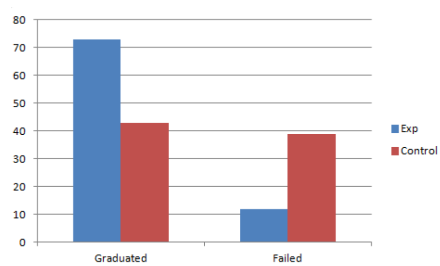
Thus, the **column** a person is in (graduated or failed to graduate) is ***contingent upon*** **(depends on)** the **row** the person is in (experimental or control)

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

5 **If** the columns are not contingent on the rows, then the rows then the column frequencies are **independent** from each other.

(which means no association or link between the two variables)

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167



6

The **TEST** of whether the columns are contingent on the rows is called the

Chi square test of independence.

When running the test, the assumption (or null hypothesis) is always made that:

there is no relationship between row and column frequencies.

7

The first task in computing the chi square test of independence is to compute the expected frequency for each cell with the assumption of independence.

In other words, we are assuming that the two variables (graduation and intervention) are *independent* from each other, or not linked.

8

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

probability of being both experimental AND Graduated

•

	Graduated	Failed to Graduate	Total
Experimental			85
Control			82
Total	116	51	167

Expected frequency

of people being both experimental AND Graduated =

9

Expected Frequency of Any Cell in a Contingency Table

$$f_e = \frac{r \cdot c}{n}$$

f_e = expected freq

where, r = row total

c = column total

n = total number of observations

10

Once the expected cell frequencies are computed, it is convenient to enter them into the original table as shown below. The expected frequencies are in parentheses.

	Graduated	Failed to Graduate	Total
Experimental	73 (59.042)	12 (25.958)	85
Control	43 (56.958)	39 (25.042)	82
Total	116	51	167

OR

Observed frequencies

	Graduated	Failed to Graduate	Total
Experimental	73	12	85
Control	43	39	82
Total	116	51	167

Expected frequencies

	Graduated	Failed to Graduate	Total
Experimental	59.042	25.958	85
Control	56.958	25.042	82
Total	116	51	167

Chi Square Statistic is :

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_e}$$

is a ***statistic*** *that measures the difference between observed values and expected values*

The next time we meet we will learn how to calculate this statistic in 3 different ways.

Project Stuff

-today you will get a copy of the Project Scoring Criteria

-and evaluate a project

(Wealth and Obsesity)

for Criteria A and B

Assignment in the Ch 11 packet:

1. read pp. 334-335

   do p.336..... 1E.1 abcd - in each problem, be sure to show at least one sample calculation

2. do p. 333.... #5 on parts b and d, show the calculation by hand.