Final Exam

- MC and True/False
  - 30 questions
    - Worth 30% of final exam grade
    - Concepts, definitions
    - Simple questions interpreting a test statistic or asking you to choose the appropriate one
- Computational
  - 8 questions, you answer 7
    - Each question will include one or more of the following:
      - Preliminary examinations
      - Tests of association
      - Tests of significance

      Planned to be about the same 'size' each as on the midterm

- Think of it as 20 minutes per question, with 40 minutes for the multiple choice

A few things to think about:

Level of measurement is always relevant!! (nominal/ordinal/interval-ratio)
You should always be considering this, in answering questions.

The word “significance” is important, and it always involves some sort of formal test (always, 5 steps)
Either:
-> Tests of difference, 2 samples (means & proportions)
-> association (cross tab)
If at least one variable is “nominal” -> X^2 test of independence
If both are “ordinal” -> use the Gamma based test of significance.
If both varaibles are “interval/ratio” use the regression based test of significance

Association?
Strength of the association?
If both or one variable is nominal -> cramer’s v, phi, lambda (and also Maximum difference method)
If both are ordinal -> gamma
If both are interval -> Pearson’s R

If both variables are “ordinal” or both are “interval ratio”, we can speak of the direction of a relationship (positive or negative)..
Again: -> level of measurement involved is relevant in deciding which statistics/formula to work with..

$$Z(\text{obtained}) = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{x}-\bar{x}}} \qquad \sigma_{\bar{x}-\bar{x}} = \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}$$

$$t(\text{obtained}) = \frac{(\bar{X}_1 - \bar{X}_2)}{\sigma_{\bar{x}-\bar{x}}} \qquad \sigma_{\bar{x}-\bar{x}} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

$$Z(\text{obtained}) = \frac{(P_{s1} - P_{s2})}{\sigma_{p-p}} \qquad \sigma_{p-p} = \sqrt{P_u(1-P_u)} \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \qquad P_u = \frac{n_1 P_{s1} + n_2 P_{s2}}{n_1 + n_2}$$

$$G = \frac{N_s - N_d}{N_s + N_d} \qquad z(\text{obtained}) = G\sqrt{\frac{N_s + N_d}{N(1 - G^2)}}$$

$$y = a + bx \qquad b = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2} \qquad t_{obtained} = r\sqrt{\frac{n-2}{1-r^2}}$$

$$a = \bar{Y} - b\bar{X}$$

$$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}}$$

$$f_e = \frac{\text{Row marginal} \times \text{Column marginal}}{N} \qquad \chi^2(\text{obtained}) = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$\phi = \sqrt{\frac{\chi^2}{N}} \qquad V = \sqrt{\frac{\chi^2}{(N)(\min r - 1, c - 1)}} \qquad \lambda = \frac{E_1 - E_2}{E_1}$$

Confidence Level	Alpha (α)	α/2	Z score
90%	0.10	0.05	± 1.65
95%	0.05	0.025	± 1.96
99%	0.01	0.005	± 2.58

Cut points for indexed tests of association			
	Weak	Moderate	Strong
Phi/Cramer's V	<.1	.1 to .2999	.3 or greater
Gamma	<.3	.3 to .5999	.6 or greater
Peason's r	<.3	.3 to .5999	.6 or greater

a. Is there a gender gap in use of the Internet? Random samples of men and women have been questioned about the average number of minutes they spend each week on the Internet for any purpose. Is the difference significant?

Women	Men
$\bar{X}_1 = 55$	$\bar{X}_2 = 60$
$s_1 = 2.5$	$s_2 = 2.0$
$N_1 = 520$	$N_2 = 515$

We are working with two samples:

A sample of King's students
N=150
12% smoke

A sample of Brock students
n= 125
20% smoke

Are King's students significantly less likely to smoke than Brock students?

The following table demonstrates the crosstabulation of marital status & where people work. Is there a significant association between the two variables? If so, using the appropriate measure of association, how strong is the relationship? What can you tell me about the pattern of the relationship??

	Marital Status			
Do You Work from Home?	Married/ Common-law	Separated/ Divorced/ Widowed	Single	Totals
Yes, but not regularly	2	0	3	5
Yes, 1–2 days each week	8	4	0	12
Yes, 3 or more days each week	0	3	3	6
No	10	8	4	22
Totals	20	15	10	45

Are the two variables significantly associated with each other?
What's the direction/pattern of the relationship?
Is there a strong association?

	Low	Medium	High
1.Strongly Disagree	24	20	18
2. Disagree	22	22	33
3. Agree	21	22	18
4. Strongly Agree	20	7	22

Calculate an appropriate measure of association. What does this tell us in terms of direction and strength of the association?

Is the association significant?

b. There is some evidence that people's involvement in their communities (membership in voluntary organizations, participation in local politics, and so forth) has been declining, and television has been cited as the cause. Do the data below support the idea that TV is responsible for the decline?

Hours of Community Service	Television Viewing			
	Low	Moderate	High	Totals
Low	5	10	18	33
Moderate	10	12	10	32
High	15	8	7	30
Totals	30	30	35	95

Calculate the "appropriate measure of association" and determine whether the Relationship is significant or not..

What's the direction of the relationship?

Is this consistent with what you would conclude if you used the maximum % difference approach??

For each situation, compute and interpret the appropriate measure of association. Describe relationships in terms of the strength and pattern or direction.

Consider crime as dependent variable

What can you tell me about this relationship? Is it significant??

a. For 10 Canadian cities data has been gathered on rates of property crime (theft, breaking and entering, fraud) per 100,000 population and the percentage of people who are new immigrants (arrived in Canada within the past five years). Are the variables related? i.e. is the relationship significant?

City	Property Crime Rate	Percent Immigrants
A	15	10
B	12	18
C	20	9
D	17	11
E	16	15
F	10	20
G	17	9
H	13	22
I	9	10
J	7	15