Exam Format
• This examination will feature 30 multiple choice questions (worth 1 point each) and 5 short answer questions (worth 2 points each). No calculators will be needed; the short answer questions will require you to define relevant terms, interpret results that are provided for you, or apply statistical tables (which will be provided for you) to answer questions, and won't require advanced math like square roots.
• All needed exam materials (such as Scantron sheets) will be provided for you, except for your preferred pencil (and eraser, if needed).
• No books, notes, or other materials will be allowed during the exam; copies of the relevant formulas and tables will be included with the exam.
• Note that no bathroom breaks are permitted during the exam -- once you leave the room, you can’t return without a five letter grade penalty.
• Also, be on time -- once the first person finishes and leaves the room, anybody else who enters will suffer an automatic deduction of five letter grades; be aware that students sometimes finish exams like this in as little as 5-10 minutes.

Advice on Studying
• The number of questions on each topic in the exam will (approximately) reflect the relative time spent on each topic in the course. A topic that was covered over two lectures should thus have approximately (but perhaps not exactly) twice as many exam questions as a topic that only lasted for one lecture.
• The exam will draw from both the assigned readings (including both textbook and assigned journal articles/book chapters) and the lecture notes. As many as 1/3 of the exam questions will be drawn from materials in the readings that were not covered at all in lecture, with the rest reflecting topics that were only covered in lecture or were covered in both lecture and the readings. If you missed one or more days of class, be sure to get a copy of those notes from somebody who was there.
--Note that in the past, questions drawn from the readings (even on concepts or topics specifically listed on the review sheet) have produced the lowest scores of any questions on the test, so you should take special care to study these topics on the review sheet.
• The exam will not be written with the intention of fooling students with trick questions or with the goal of failing as many students as possible. The main goal of this course is to provide students with an understanding of how Political Science research works, so the exam questions will reflect this goal.
• The list of topics on this review sheet is not legally binding; these are just suggestions for the most important topics that are most likely to be on the test (some of which may not actually appear on the test). If you understand all of these topics you are much more likely to do well on the test.
Examples of Possible Short Answer Questions
• Design and interpretation of 2x2 tables/crosstabulation (if some of the information in the table is provided for you, you will need to calculate the remaining information, such as marginals or N)
• Interpretation of $X^2$ (if a statistic is provided for you, you will need to interpret it)
• Interpretation of Phi (if a statistic is provided for you, you will need to interpret it)
• Interpretation of Cramer’s V (if a statistic is provided for you, you will need to interpret it)
• Interpretation of odds ratio (if a statistic is provided for you, you will need to interpret it)
• Interpretation of scatterplots (if a plot is provided for you, you will need to interpret it)
• Interpretation of correlation (if a statistic is provided for you, you will need to interpret it)
• Interpretation of regression coefficients (if a regression equation is provided for you, you will need to interpret one or more of the coefficients)
• Interpretation of $R^2$

Topics Covered in This Portion of the Course
Hypothesis Testing for Nominal/Ordinal Variables
• Crosstabulation:
  --Dependent & independent variables
  --Cell frequencies, marginals, N
  --Column percentages and direction/strength of relationships
• $X^2$ test of independence: main goal (what is being tested?, $H_0$ vs. $H_1$)
  --Basic approach: how to predict cell frequencies, compare observed & predicted frequencies
  --Limits (indep. but not strength/pattern, not causation, less accurate for small #s, inflated by n)

Measures of Association for Nominal/Ordinal Variables
• Idea of measuring association (and difference from statistical significance)
• Phi: main goal (association in 2x2 tables of nominal or ordinal variables)
  --Interpretation (larger=more assoc., rule of thumb for weak/moderate/strong)
• Cramer’s V: main goal (association in larger tables of nominal or ordinal variables)
  --Interpretation (larger=more assoc., rule of thumb for weak/moderate/strong)
• PRE measures (Lambda, Somers’ d): main goal of PRE, types of variables used for each, interpretation
• Odds, Odds Ratio: main goal (substantive significance for two nominal/ordinal variables)
  --General meaning, calculation, interpretation

Correlation and Regression
• Scatterplot: main goal (display relationship between two interval/ratio variables)
  --Setup (dependent & independent variables)
  --Possible shapes – linear relationships: positive, negative, no relationship
  --Some possible nonlinear shapes: curvilinear (U or inverted U), exponential
• Correlation: main goal (association between two interval/ratio variables)
  --Basic calculation (shared variability / separate variability)
  --Interpretation (difference between 0 and +/- 1, difference between positive and negative values)
  --Significance testing (t-test, interpretation of significant result)
  --Benefits (symmetric, scale doesn't matter)
  --Limits (only linear, distorted by extreme outliers, no direct interpretation, not causation)
• Regression lines: the basic idea (mathematical formula summarizing scatterplots)
• Interpreting bivariate regression formulas: $Y = a + bX + e$
  --a: Y intercept ("constant term"); meaning and interpretation
  --b: slope; meaning and interpretation
Applications: using the regression equation to make predictions

- Calculating regression equations: OLS
  - $e_i$: error term (residual, disturbance); $e_i = Y_i - \bar{Y}_i$
  - Ordinary Least Squares basic idea (minimize sum of squared errors)
  - Basic idea of calculating $a$, $b$

- $R^2$ (“coefficient of determination”)
  - Basic idea: measure model fit (how well does regression equation fit the data?)
  - Interpretation: PRE (variation in DV explained by IV)
  - Basic idea of calculation

- Hypothesis testing with regression:
  - $H_0$ and $H_1$
  - Tested with t-ratio; interpretation of results

- Multiple regression: basic idea (more than one IV in model, statistical control for other IVs)
  - $a$, $b$: differences in interpretation from bivariate regression (partial slope)
  - Hypothesis testing: done separately for each IV
  - $R^2$: difference in interpretation from bivariate regression (effect of all variables)
  - Adjusted $R^2$: what is different and why?
  - F-ratio and model significance; $H_0$ vs. $H_1$, interpretation of results

- Dummy variables: definition, interpretation of results, how to handle more than 2 categories

- Interaction terms: definition, interpretation of results

- Multicollinearity: nature of the problem
  - How to detect the problem
  - How to try to fix the problem

- Model specification: nature of the problem (omitted variable bias, irrelevant/extraneous var.s)
  - How to detect the problem
  - How to try to fix the problem

More Advanced Analyses

- Logistic regression: basic idea (regression with dummy DV)
  - MLE rather than OLS calculation
  - Calculation: DV measured in model as log of the odds that DV=1
  - Interpreting coefficients (direct interpretation, exponentiate to get odds ratio)
  - Substantive sig. (book: “MEMs and MERs”)
  - Model significance: $-2LL$ (based on logged likelihood for null model, main model)
  - Model fit: “pseudo-$R^2$”

- Multinomial logit/probit: basic idea (nominal DV with more than 2 categories)

- Ordered logit/probit: basic idea (ordinal DV with more than 2 categories)

- Selection models (a.k.a. Heckman model, Heckman probit, censored probit):
  - Basic idea (2-stage model, selection stage before outcome stage)

- Survival/duration models (including Cox regression, Weibull model, etc.):
  - Basic idea (DV = time until event / time between events)

- Event count models (incl. Poisson, negative binomial, etc.):
  - Basic idea (DV = count of events in time period)
Chi-Square
\[ X^2 = \sum \left( \frac{(f_o - f_e)^2}{f_e} \right) \]
- \( f_o \) = observed cell frequencies
- \( f_e \) = frequencies expected if the variables were indep.
- \( df = (r - 1)(c - 1) \) -- where \( r \) = # of rows in table, \( c \) = # of columns

Phi
\[ \phi = \sqrt{\frac{\chi^2}{n}} \]
(where \( \chi^2 \) is calculated as above, and \( n \) is the # of cases in the table)

Cramer’s V
\[ V = \left( \frac{\chi^2}{\min(r-1,c-1)} \right) \]
(where \( \chi^2 \) is calculated as above, \( n \) = # cases, \( r \) = # rows in table, \( c \) = # columns in table)

Odds Ratio
Odds = \( \frac{\text{Prob. of event}}{\text{Prob. of non-event}} \)

Odds Ratio = relative odds of DV under 2 conditions (values of IV) = \( \frac{\text{Odds of DV in condition 1}}{\text{Odds of DV in condition 2}} \)

Pearson’s r (Correlation)
\[ r = \frac{\text{Shared variability of X&Y}}{\text{Separate variability of X&Y}} \]

Regression
**Basic Regression Formula:**
- Bivariate Regression: \( Y = a + bX + e \)
- Multiple Regression: \( Y = a + b_1X_1 + b_2X_2 + \ldots + e \)

\[ e_i = Y_i - \bar{Y}_i \]

\[ R^2 = \frac{\text{Explained variance in DV}}{\text{Total variance in DV}} = \frac{\text{Sum of Squares–Regression}}{\text{Sum of Squares–Total}} \]

\[ t = \frac{b}{\text{S.E.}(b)} \]

F-test for significance of model:
\[ F = \frac{\text{Variance explained by regression}}{\text{Unexplained variance}} \]
<table>
<thead>
<tr>
<th>Degrees of freedom</th>
<th>.10</th>
<th>.05</th>
<th>.025</th>
<th>.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.706</td>
<td>3.841</td>
<td>5.024</td>
<td>6.635</td>
</tr>
<tr>
<td>2</td>
<td>4.605</td>
<td>5.991</td>
<td>7.378</td>
<td>9.210</td>
</tr>
<tr>
<td>3</td>
<td>6.251</td>
<td>7.815</td>
<td>9.348</td>
<td>11.345</td>
</tr>
<tr>
<td>4</td>
<td>7.779</td>
<td>9.488</td>
<td>11.143</td>
<td>13.277</td>
</tr>
<tr>
<td>5</td>
<td>9.236</td>
<td>11.070</td>
<td>12.833</td>
<td>15.086</td>
</tr>
<tr>
<td>7</td>
<td>12.017</td>
<td>14.067</td>
<td>16.013</td>
<td>18.475</td>
</tr>
<tr>
<td>8</td>
<td>13.362</td>
<td>15.507</td>
<td>17.535</td>
<td>20.090</td>
</tr>
<tr>
<td>10</td>
<td>15.987</td>
<td>18.307</td>
<td>20.483</td>
<td>23.209</td>
</tr>
<tr>
<td>11</td>
<td>17.275</td>
<td>19.675</td>
<td>21.920</td>
<td>24.725</td>
</tr>
<tr>
<td>13</td>
<td>19.812</td>
<td>22.362</td>
<td>24.736</td>
<td>27.688</td>
</tr>
<tr>
<td>14</td>
<td>21.064</td>
<td>23.685</td>
<td>26.119</td>
<td>29.141</td>
</tr>
<tr>
<td>15</td>
<td>22.307</td>
<td>24.996</td>
<td>27.488</td>
<td>30.578</td>
</tr>
<tr>
<td>16</td>
<td>23.542</td>
<td>26.296</td>
<td>28.845</td>
<td>32.000</td>
</tr>
<tr>
<td>17</td>
<td>24.769</td>
<td>27.587</td>
<td>30.191</td>
<td>33.409</td>
</tr>
<tr>
<td>18</td>
<td>25.989</td>
<td>28.869</td>
<td>31.526</td>
<td>34.805</td>
</tr>
<tr>
<td>19</td>
<td>27.204</td>
<td>30.144</td>
<td>32.852</td>
<td>36.191</td>
</tr>
<tr>
<td>20</td>
<td>28.412</td>
<td>31.410</td>
<td>34.170</td>
<td>37.566</td>
</tr>
<tr>
<td>21</td>
<td>29.615</td>
<td>32.671</td>
<td>35.479</td>
<td>38.932</td>
</tr>
<tr>
<td>22</td>
<td>30.813</td>
<td>33.924</td>
<td>36.781</td>
<td>40.289</td>
</tr>
<tr>
<td>23</td>
<td>32.007</td>
<td>35.172</td>
<td>38.076</td>
<td>41.638</td>
</tr>
<tr>
<td>24</td>
<td>33.196</td>
<td>36.415</td>
<td>39.364</td>
<td>42.980</td>
</tr>
<tr>
<td>25</td>
<td>34.382</td>
<td>37.652</td>
<td>40.646</td>
<td>44.314</td>
</tr>
<tr>
<td>26</td>
<td>35.563</td>
<td>38.885</td>
<td>41.923</td>
<td>45.642</td>
</tr>
<tr>
<td>27</td>
<td>36.741</td>
<td>40.113</td>
<td>43.195</td>
<td>46.963</td>
</tr>
<tr>
<td>28</td>
<td>37.916</td>
<td>41.337</td>
<td>44.461</td>
<td>48.278</td>
</tr>
<tr>
<td>29</td>
<td>39.087</td>
<td>42.557</td>
<td>45.722</td>
<td>49.588</td>
</tr>
<tr>
<td>30</td>
<td>40.256</td>
<td>43.773</td>
<td>46.979</td>
<td>50.892</td>
</tr>
<tr>
<td>40</td>
<td>51.805</td>
<td>55.758</td>
<td>59.342</td>
<td>63.691</td>
</tr>
<tr>
<td>50</td>
<td>63.167</td>
<td>67.505</td>
<td>71.420</td>
<td>76.154</td>
</tr>
<tr>
<td>60</td>
<td>74.397</td>
<td>79.082</td>
<td>83.298</td>
<td>88.379</td>
</tr>
<tr>
<td>70</td>
<td>85.527</td>
<td>90.531</td>
<td>95.023</td>
<td>100.425</td>
</tr>
<tr>
<td>80</td>
<td>96.578</td>
<td>101.879</td>
<td>106.629</td>
<td>112.329</td>
</tr>
<tr>
<td>90</td>
<td>107.565</td>
<td>113.145</td>
<td>118.136</td>
<td>124.116</td>
</tr>
<tr>
<td>100</td>
<td>118.498</td>
<td>124.342</td>
<td>129.561</td>
<td>135.807</td>
</tr>
</tbody>
</table>