STAT 221/231 Midterm Test 1
Thursday October 13, 4:40-6:10 pm

Seating for Midterm Test 1 is predetermined so please check your seat assignment at https://odyssey.uwaterloo.ca/teaching/schedule **after 1pm on Friday October 7**.

Bring your Watcard and a ruler. Only Pink Tie or Blue Goggles Calculators may be used.

You may bring one (1) double-sided, letter sized (8.5 x 11 inches), handwritten page of notes to the exam (no photocopies).

For efficiency of marking your final numerical answers must be rounded to 3 decimal places.To avoid round off errors carry as many decimal places as possible while making your calculations.

Midterm Test 1 covers the material in Chapters 1 to 3.

In preparation for the test you should do the following end-of-chapter problems:
- Chapter 1: 1-20
- Chapter 2: 1-19
- Chapter 3: 1-7
- See also the Sample Midterm Test 1 on pages 389-398 of the Course Notes.

One of the questions on the midterm will be similar to the end of Chapter 3 problems 1-5. The context will be drawn from one of the articles posted on Learn under Content -> Additional Resources -> Statistics in the Media.

There will also be multiple choice or short answer questions on the statistical software R.

To aid you in creating your sheet of notes here is a list of the ideas and definitions you should know for Midterm Test 1:

Empirical Studies (Sections 1.1-1.2)
- units, populations and processes (page 1)
- variate and types of variates (page 3)
- response versus explanatory variates (page 24)
- attributes (apge 4)
- types of studies (sample surveys, observational studies, experimental studies)

Numerical Summaries (Section 1.3)
- measures of location: sample mean, median, mode (pages 7-8)
- measures of variability: sample variance, sample standard deviation, range, IQR (page 8 and Definition 3, page 12)
- measures of shape: skewness, kurtosis (page 8)
- sample percentiles and quantiles (Definition 1, page 11)
- lower or first quartile, upper or third quartile (Definition 2, page 11)
- five number summary (Definition 4, page 12)
- numerical summaries for bivariate data: sample correlation (Definition 5, page 14), relative risk (Definition 6, page 15)

Graphical Summaries (Section 1.3)
- relative frequency histograms (page 16)
- empirical cumulative distribution function (page 19)
- boxplots (page 20-21)
- scatterplots (page 21)

Data Analysis and Statistical Models (Sections 1.5 and 2.1)

descriptive statistics, statistical inference (inductive versus deductive reasoning) (page 25)
discrete statistical models: Binomial(n,θ), NegativeBinomial(k,θ), Poisson(θ), Geometric(θ)
continuous statistical models: Exponential(θ), G(μ,σ), Multinomial$(n,\theta_1,\theta_2,\ldots,\theta_k)$ (pages 45-46)

Estimation and the Method of Maximum Likelihood (Sections 2.2-2.5)

estimate of a parameter (Definition 7, page 47)
likelihood functions for discrete distributions (Definition 8, page 48)
maximum likelihood estimate (Definition 9, page 48)
relative likelihood function (Definition 10, page 51)
log likelihood function (Definition 11, page 52)
likelihood function for a random sample (page 53)
likelihood functions for continuous distributions (Definition 12, page 57)
likelihood function for multinomial distribution (page 59)
invariance property of maximum likelihood estimates (Theorem 13, page 61)

You should know how to construct the likelihood function and find the maximum likelihood estimate of θ for each of the following situations:

1) *Binomial*(n,θ) and *Negative Binomial*(k,θ) for observed data y

2) *Poisson*(θ), *Geometric*(θ), *Exponential*(θ) based on an observed random sample $y_1,y_2,\ldots,y_n$

3) $G(\theta,\sigma)$ (σ known) and $G(\mu,\theta)$ (μ known) based on an observed random sample $y_1,y_2,\ldots,y_n$

4) $f(y;\theta)$ (for given p.f./p.d.f. $f(y;\theta)$) based on an observed random sample $y_1,y_2,\ldots,y_n$

Check the Fit of a Model (Section 2.6)

You should know how to check the fit of a model by:

1) calculating expected frequencies using the assumed model (discrete and continuous distributions) and comparing them with observed relative frequencies

2) examining a graph of the probability density function of the assumed model superimposed on a relative frequency histogram for the observed data (continuous distributions)

3) examining a graph of the cumulative distribution function of the assumed model superimposed on a graph of the empirical cumulative distribution function for the observed data (continuous distributions)

4) examining a Normal or Gaussian qqplot

Planning and Conducting an Empirical Study (Chapter 3)

the steps of PPDAC (page 87)
the elements of each step of PPDAC (Section 3.2)
types of problems (decriptive, causative, predictive) (page 88)
units, target population and target process (Definition 14, page 89)
study units, study population and study process (Definition 17, page 90)
study error (Definition 18, page 90)
sampling protocol and sample size (Definition 19, page 91)
sample error (Definition 20, page 91)
measurement error (Defintion 21, page 91)
fishbone diagram (page 95)