STT200
Prep for Final Exam
Wednesday, April 30, 3 - 5 pm, B106 Wells Hall

All needed paper and tables will be provided.
As usual, no electronics, headphones, notes or books.

The exam will be comprehensive over:

Probability

Rules of Multiplication and Addition
Total probability
Conditional probability, Independence, Bayes
Trees, Venn, Balls from boxes, Dice, OIL

Random varibles

R.v. as a numerical function on outcomes
Numerical functions of r.v. are again r.v.

Expectation, variance and sd of r.v.

E X = sum of value times probability $\sum_x x\ p(x)$

E f(X) = $\sum_x f\ p(f) = \sum_x f(x)\ p(x)$

E(aX+bY+c) = aEX + bEY + c (always)

Var(aX+b) = a^2 Var X

$\sigma_{aX+b} = |a|\ \sigma_X$

Var(a X + bY +c) = a^2 Var X + b^2 Var Y (indep r.v.)

$\sigma_{aX+bY+c} = \sqrt{a^2 \text{ Var } X + b^2 \text{ Var } Y} = \sqrt{a^2\ \sigma_X^2 + b^2\ \sigma_Y^2}$ (indep)

$\sigma_{\overline{x}} = \frac{\sigma_{\text{population}}}{\sqrt{n}}$ for independent (e.g. with-repl) samples

Calculations of E (net return) in OIL type problems

Concept of standard score = $\frac{\text{score} - \text{mean}}{\text{sd}}$

Regression and correlation.

Scatterplot of (x, y) points in relation to the means of x, y.

Points "upper right" or "lower left" of $(\overline{x},\ \overline{y})$ give + assoc.

Other points give negative association.

Correlation R = $\frac{\overline{x\,y} - \overline{x}\,\overline{y}}{\sqrt{\overline{x^2} - \overline{x}^2}\,\sqrt{\overline{y^2} - \overline{y}^2}}$

R(ax+b, cy+d) = R(x, y) (if a, c have the same sign)

R^2 is the fraction of variance y explained by regression on x *

*actually, this is if we use n, not n-1, in calculating variance x

Plot of the regression line thru $(\overline{x}, \overline{y})$ with slope R $\frac{s_y}{s_x}$

Elliptical plot = joint normal

Predicting y from x using regression line*

*insert x, read off predicted y from line

Central Limit (normal) Approximations

Binomial (case when np and nq are large)

n indep trials,

probability p of "success" each trial,

X = number of successes in n trials

X ~ normal w/ mean np and sd $\sqrt{n\,p\,q}$, q = 1 - p

Poisson (when it applies)

expect μ rare events to occur (i.e. on avg)

sd X = $\sqrt{\mu}$ (if Poisson applies)

X ~ normal w/ mean μ and sd $\sqrt{\mu}$ if $\mu > 3$

Sample mean of independent samples from a population

population mean μ and sd σ, sample of n

$\overline{X}$ ~ normal with mean μ and sd $\frac{\sigma}{\sqrt{n}}$

Difference of independent sample means (large n's)

$\overline{x}$- $\overline{y}$ ~ normal

mean μ_x- μ_y

sd $\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$

Data calculations

sample mean (balance point of distribution)

sample sd
quartiles
inter-quartile range
box plot
$\overline{ax + by + c} = a\overline{x} + b\overline{y} + c$
s from a list x
$s_{ax+b} = |a|\, s_x$

Topics emphasized on Exam 4.

t-based and z-based CI and test for a population mean
t-plot (**requires normal population dist**); getting t from confidence C
z-plot is t-plot for df = infinity*
*often used for large df when population dist may not be **normal**

$\overline{x} \pm t\ \frac{s}{\sqrt{n}}$; claim made for CI; test statistic $t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$; df = n-1

H0 and H1; types I and II errors
P-value = 1-C (2-sided) or $\frac{1-C}{2}$(1-sided)
Reject H0 if P-value < α = desired type I error rate

z-based CI and test for population proportion p (p is NOT P-value)

$\hat{p} \pm z\ \frac{\sqrt{\hat{p}\,\hat{q}}}{\sqrt{n}}$; test statistic z = $\frac{\hat{p} - p_0}{\frac{\sqrt{p_0\, q_0}}{\sqrt{n}}}$(notice that p_0 is used)

z-based CI and test for μ_x- μ_y based on large indept samples of n_x, n_y

$\overline{x}$ - $\overline{y} \pm z \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$

No companion z - test on the exam except questions of interpreatation

z-based CI and test for $\hat{p}_x - \hat{p}_y$ based on large indep samples of n_x, n_y

$\hat{p}_x - \hat{p}_y \pm z \sqrt{\frac{\hat{p}_x \hat{q}_x}{n_x} + \frac{\hat{p}_y \hat{q}_y}{n_y}}$

No companion z-test on exam except questions of interpretation.

Margins of error; effect of increasing n

P-value, methods of calculation, use to reach decision on H0*
*see the prep exercises of Exam 4 for more details

Some prep exercises for the Final Exam will be posted Monday, April 21.