STATS 100C Final Exam Information (Spring 2009)

- **Final exam**: Monday, June 8, 11:30-2:30pm in **BOELTER 2444**.
- This is a closed-book exam, but you can use a calculator and the formulas provided.
- The z, t and F tables will be provided if necessary.
- Material: Chapters 1-6, HW 1-9.

Formulas for STATS 100C Final Exam (Spring 2009)

<u>Chapter 2: Simple Linear Regression</u>

The pdf of $N(\mu, \sigma^2)$ is $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{(x-\mu)^2}{2\sigma^2}]$.
Model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. LSE: $\hat\beta_1 = \frac{s_{xy}}{s_{xx}}$, $\hat\beta_0 = \bar y - \hat\beta_1 \bar x$, $\hat\mu_0 = \hat\beta_0 + \hat\beta_1 x_0$.
Variance: $V(\hat\beta_1) = \sigma^2/s_{xx}$, $V(\hat\beta_0) = \sigma^2(\frac{1}{n} + \frac{\bar x^2}{s_{xx}})$, $V(\hat\mu_0) = \sigma^2(\frac{1}{n} + \frac{(x_0-\bar x)^2}{s_{xx}})$.
Notation: sample correlation coefficient $r = s_{xy}/\sqrt{s_{xx}s_{yy}}$, $s_{xx} = \sum(x_i - \bar x)^2 = \sum x_i^2 - \frac{1}{n}(\sum x_i)^2$
$s_{xy} = \sum(x_i-\bar x)(y_i - \bar y) = \sum(x_i - \bar x)y_i = \sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i)$
$SST = \sum(y_i - \bar y)^2$, $SSR = \sum(\hat\mu_i - \bar y)^2 = \hat\beta_1^2 s_{xx}$ and $SSE = \sum(y_i - \hat\mu_i)^2$.

<u>Chapter 3: Random Vectors</u> Suppose $E(y) = \mu$ and $V(y) = \Sigma$.

$$E(Ay+b) = A\mu + b, V(Ay+b) = A\Sigma A', V(y) = E[(y-\mu)(y-\mu)'], cov(a'y, b'y) = a'\Sigma b.$$

The pdf of a n-variate normal distribution $N(\mu, \Sigma)$ is

$$f(x) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}\exp[-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)]$$

<u>Chapter 4: Multiple Linear Regression Model</u>

Model: $y = X\beta + \epsilon,\ \epsilon \sim N(0, \sigma^2 I)$
LSE $\hat\beta = (X'X)^{-1}X'y$, $s^2 = S(\hat\beta)/(n-p-1) = (y - X\hat\beta)'(y - X\hat\beta)/(n-p-1)$.
Fitted values $\hat\mu = X\hat\beta = Hy$; residuals $e = y - \hat\mu = y - Hy = (I-H)y$; where $H = X(X'X)^{-1}X'$.
Variance $V(\hat\beta) = \sigma^2(X'X)^{-1}$, $V(\hat\mu) = \sigma^2 H$, $V(e) = \sigma^2(I-H)$, $V(a'\hat\beta) = \sigma^2 a'(X'X)^{-1}a$.
Properties: $\hat\beta$ and e are independent; $S(\hat\beta)/\sigma^2 = e'e/\sigma^2 \sim \chi^2_{n-p-1}$; LSE is BLUE.
CI for β_i: $\hat\beta_i \pm t(1-\alpha/2, n-p-1)s\sqrt{v_{ii}}$; CI for $a'\beta$: $a'\hat\beta \pm t(1-\alpha/2, n-p-1)s\sqrt{a'(X'X)^{-1}a}$
F statistic for testing $H_0: A\beta = 0$:

$$F = \frac{\|\hat\mu - \hat\mu_A\|^2/l}{S(\hat\beta)/(n-p-1)} = \frac{(S(\hat\beta_A) - S(\hat\beta))/l}{S(\hat\beta)/(n-p-1)}$$

Joint confidence region for β: $(\hat\beta - \beta)'(X'X)(\hat\beta - \beta)/[(p+1)s^2] \le F(1-\alpha, p+1, n-p-1)$.
$SST = \sum(y_i - \bar y)^2$, $SSE = \sum(y_i - \hat\mu_i)^2$, $SSR = \sum(\hat\mu_i - \bar y)^2 = \hat\beta'X'y - n\bar y^2 = \hat\beta'X'X\hat\beta - n\bar y^2$.
Overall significance test: $F = MSR/MSE$; Coefficient of determination $R^2 = SSR/SST = 1 - SSE/SST$
GLS $\hat\beta^{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}y$, $V(\hat\beta^{GLS}) = \sigma^2(X'V^{-1}X)^{-1}$, $s^2_{GLS} = S(\hat\beta^{GLS})/(n-p-1)$.

<u>Chapter 5: Specification Issues in Regression Models</u>

Comparison of k treatments: $SSR = \sum_i n_i(\bar y_i - \bar y)^2$, $SSE = \sum_i\sum_j(y_{ij} - \bar y_i)^2$, $SST = \sum_i\sum_j(y_{ij} - \bar y)^2$.
Variance inflation factor $VIF_j = 1/(1-R_j^2)$

<u>Chapter 6: Model Checking</u>

Studentized residual $d_i = e_i/(s\sqrt{1-h_{ii}})$; leverage h_{ii}; where $H = (h_{ij})$ is the hat matrix

$$\text{Cook's D statistic } D_i = \frac{(\hat\beta_{(i)} - \hat\beta)'(X'X)(\hat\beta_{(i)} - \hat\beta)}{(p+1)s^2} = \frac{h_{ii}}{1-h_{ii}}\frac{d_i^2}{p+1}$$