Exam 2 is Wed. March 7. **You are allowed 4 sheets of notes and a calculator.** The exam covers ch. 3, 5, 8, 10, 11, and 12. The pull out section of the text, colored boxes and bold face text have important ideas. Also see the chapter summaries. Memorize $\overline{x}$, sd s, zscore, and how to use tables A and B.

Types of problems:
Problems 9) and 11) were also fair game on exam 1.
The following problem is very important for the midterm and the final.

9) Know how to do a **forwards calculation using table A.** See E1 5, 6, 14, Q2 5, Q3 3bc, HW2 C), E), HW 6 Ea, Fa.

11) Given a density that is box shaped with base from a to b, know that the height of the density is $1/(b-a)$ and that the chance that X is between c and d where $a \le c < d \le b$ is given by (base)(height) = $(d-c)/(b-a)$. A triangle has area = 0.5 (base) (height). See HW2 B bcd, HW5 Dbc

The following 3 problems are very important for the midterm and the final.

13) Be able to find the least squares line $\hat{y} = a + bx$ from Minitab output. A typical table and a table with numbers are shown below. See E1 11, Q4 6, HW3 H.

predictor	coef	stdev	T	Pvalue
Constant	a			
x	b	unimportant	numbers	for exam 2

predictor	coef	stdev	T	Pvalue
constant	272.819	63.4963	4.297	0.0000
sternal height	1.01482	0.04537	22.370	0.0000

14) Be able to find the least squares line $\hat{y} = a + bx$ given 2 means, 2 SD's and the correlation r. Recall that $b = rs_y/s_x$ and $a = \overline{y} - b\overline{x}$. Remember that **the response variable** y is what you want to predict. The explanatory variable x is used to help predict y. See Q4 1,2, HW3 E.

15) After being given or finding the slope and intercept or the line $\hat{y} = a + bx$, be able to predict y for a given value of x. See Q4 7, HW3 Eb, H.

16) Know that the slope is b and the intercept is a. Know the meaning of these terms.

17) Given the least squares line $\hat{y} = a + bx$, be able to put it on a scatterplot. HW3 Da

18) Know that least squares should only be used if the scatterplot is linear (football shaped). A residual is $y - \hat{y}$ and a residual plot should be football shaped with zero slope. Know that extrapolation is risky. See HW3 Db.

19) Know the difference between an individual and a population. HW4 A, B.

20) Know that association does not imply causation.

21) Know what a lurking variable is. HW3 F

The following problem is very important for the midterm and the final.

22) Know how to get a SRS using table B. See Q4 4, HW4 D, E.

23) Know that voluntary response samples and samples of convenience are bad **regardless of the sample size** while probability samples are good. Q4 3, HW4 Cb.

24) Know that SRS's are too expensive so multistage samples are used when interviewers are sent out. Random digit dialing is used for many opinion polls.

25) Know that the accuracy of the probability sample depends on the size of the sample. Two SRS's of the same size have the same accuracy (if the sample size is small compared to both population sizes). **Bigger samples have greater accuracy.** Some times you will be given the sample size, sometimes a percentage of the population sampled (then you need to figure out which sample is larger). See Q4 5, HW4 G.

The following problem is very important for the midterm and the final.

26) Know that the normal approx for $\overline{x}$ holds for $n \geq 5$ if the population of x is approximately normal. Know that the CLT **does not apply** if $n \leq 30$ and x comes from a highly skewed population. Unless told otherwise, assume CLT holds for $n \geq 100$ even for a highly skewed population. See Q5 2,3?.

27) Know that for the CLT to apply, the data needs to be a SRS or observations from a randomized experiment (eg coin tossing). If the data comes from a sample of convenience or a voluntary response sample, you can not find probabilities such as $P(\overline{x} < a)$.

28) Know that $\mu_{\overline{x}} = \mu$ and $\sigma_{\bar{x}} = \sigma/\sqrt{n}$.

The following problem is very important for both the midterm and the final.

29) Know how to do a forwards calculation involving $\overline{x}$. See Q5 3?, HW6 Eb, Fb, and H.

30) Law of large numbers. Figure out the mean μ. If μ is favorable (eg stock market, number of questions likely to get right if you are a good student) larger sample sizes n are better than smaller. If μ is not favorable (eg casino gambling or guessing on a multiple choice exam) smaller sample sizes are better. See Q5 4?, HW6 D, G.

31) Know that for any event A, $0 \leq P(A) \leq 1$.

32) **Probability rules:** i) $P(S) = 1$
ii) **Complement rule:** P(not A) = 1− P(A).
iii) A and B are **disjoint events** if A and B have no outcomes in common: P(A and B) = 0. Hence if A occurs, B did not occur and vice verca. If A and B are disjoint, then the **addition rule for 2 disjoint events** is P(A or B) = P(A) + P(B).

iv) Finite S. If $S = \{e_1, ..., e_k\}$ then $0 \leq P(e_i) \leq 1$, $\sum_{i=1}^k P(e_i) = 1$. If e_i is a sample point, then $P(A) = \sum_{i:e_i \in A} P(e_i)$. That is, $P(A)$ is the sum of the probabilities of the sample points in A. If all of the outcomes e_i are *equally likely,* then $P(e_i) = 1/k$ and $P(A) =$ (number of outcomes in $A)/k$ if S contains k outcomes.

v) Two events A and B are **independent** if knowing that one occurs does not change the probability that the other occurs. If events are not independent, then they are dependent. Two events A and B are independent if P(A and B) = P(A) P(B). The events $A_1, ..., A_n$ are independent if knowing any subset of one to n - 1 events occurred does not change the probabilities of the other events.

vi) **Multiplication rule for 2 independent events:** If A and B are independent, then P(A and B) = P(A) P(B).

vii) **General Addition Rule:** P(A or B) = P(A) + P(B) − P(A and B).
Notice that if A and B are disjoint, then P(A or B) = P(A) + P(B).
Notice that if A and B are independent, then P(A or B) = P(A) + P(B) − P(A) P(B).

viii) **Addition rule for n disjoint events:** If $A_1, ..., A_n$ are disjoint, then $P(A_1 \text{ or } A_2 \text{ or } \cdots \text{ or } A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$. This is the probability that at least one of the n events occurs.

ix) **Multiplication rule for n independent events:** If $A_1, ..., A_n$ are independent, then $P(A_1 \text{ and } A_2 \text{ and } \cdots \text{ and } A_n) = P(A_1)P(A_2) \cdots P(A_n)$. This is the probability that all n events occur.

The following problem is very important for the midterm and the final.

33) Table of probabilities with some outcomes blank. Use the fact that all of the probabilities add to 1. See Q5 4?, HW 5 C, and HW6 C.

34) Given a story problem, list the outcomes that make up an event (especially for die problems). Often you can use order to find S. Using a table to find S if two die are tossed or if a die is tossed twice and to find S if a coin is flipped 2, 3, or 4 times are typical examples. After listing all outcomes in S, use these outcomes to find $P(A)$.

35) **Toss two die** (eg red or green) (or toss a die twice with a 1st die, 2nd die). Find the probability that the sum of the two die = k. Solution: fill a table with 36 entries and find the number of entries where the sum is equal to k. These entries lie on a diagonal. Let E_k = "sum of the dice is k". Then $P(e_k)$ = P(sum of the dice is equal to k) = (number of table entries where the sum is k)/(number of table entries). Frequently a 4, 5, or 6-sided die will be used. For a 6–sided die the number of table entries is (6)(6) = 36 and

k	2	3	4	5	6	7	8	9	10	11	12
P(sum of two dice = k)	1/36	2/36	3/36	4/26	5/36	6/36	5/36	4/36	3/36	2/36	1/36

36) Given P(A), find P(not A). Given P(not A), find P(A). Use the complement rule: P(not A) = 1 − P(A).

37) **General Addition Rule:** P(A or B) = P(A) + P(B) − P(A and B).
Notice that if A and B are disjoint, then P(A or B) = P(A) + P(B).
Notice that if A and B are independent, then P(A or B) = P(A) + P(B) − P(A) P(B).
Given any three of the above probabilities, use the general additive rule to find the fourth probability. USING A PROBABILITY VENN DIAGRAM CAN BE USEFUL. The probabilities in 4 regions are P(A and not B), P(A and B), P(not A and B) and P(not A and not B). The four regions are disjoint.

38) Given P(A), P(B), and that A and B are disjoint, find P(A and B) or find P(A and B). If A and B are disjoint, P(A and B) = 0 while P(A or B) = $P(A) + P(B)$.

39) Given P(A), P(B), and that A and B are independent, find P(A and B) or find P(A or B). If A and B are independent, P(A and B) = P(A) P(B) while
P(A or B) = $P(A) + P(B) - P(A)P(B)$.

40) Know P(x was at least k) = $P(x \geq k)$ and P(x at most k) = $P(x \leq k)$.

41) Suppose there are n independent identical trials and x counts the number of successes (outcome D) and the p = probability of success for any given trial. Then
i) P(x=0) = P(**none** of the n trials were successes) = $(1 - p)^n$.
ii) $P(x \geq 1)$ = P(**at least one** of the trials was a success) = $1 - P(x = 0) = 1 - (1 - p)^n$.
iii) P(x=n) = P(**all** n trials were successes) = p^n.
iv) $P(x < n)$ = P(**not all** n trials were successes) = $1 - P(x = n) = 1 - p^n$.
See HW6 J, K.