STUDY GUIDE FOR EXAM 1
EXPERIMENTAL DESIGN
Chap 1: Controlled Experiments
Main Idea: Treatment and Control groups should be as much alike as possible.
- Randomized, double-blind design is ideal.
- Non-randomized and historical controls may introduce systematic bias.

Chap 2: Observational Studies
Main Idea: Treatment and Control groups are likely to be different in important ways since they chose themselves.
- Difficult to conclude causation from association.
- Adjust for confounders by dividing into subgroups
- Simpson's Paradox--clear-cut case of confounding where the overall percentages are misleading and are reversed when divided into sub-groups.

From a description of a study you should be able to:
- Identify whether the study was a controlled experiment or an observational study.
- Identify what type of controls (if any) were used, i.e., randomized, non-randomized or historical.
- Determine whether the study was double-blind. Was a placebo used? Did the evaluators know who was in the treatment and who was in the control group?
- Determine what conclusions are valid.
- What are the likely confounders and how would you adjust for them?

Sample Questions on Experimental Design:

1. Two experiments were done comparing the effects of listening to classical music versus hip hop music while studying. All the students in both experimental designs were given an identical 2-hour lesson and then allowed time to study for a short exam.
In Design A students themselves chose to study either listening to Mozart or Nelly.
In Design B the students were randomly assigned to study either listening to Mozart or Nelly.
Design A found that the Mozart study group scored significantly higher on the exam than the Nelly group did.
Design B found no significant difference in exam scores between the 2 groups.
The overall exam average in both designs was the same.

a) Which design had controls?	A only	B only	Both	Neither
b) Which design had randomized controls?	A only	B only	Both	Neither

c) Which design is more likely to have confounders?	A	B	Both are equally likely

d) Which conclusion is best supported by the evidence?   <u>Circle only one.</u>

i)   Students learn better when they are able to choose their own music while studying.
ii)   Students who choose Mozart are different in more ways than just their musical tastes than students who choose Nelly

iii)   Classical music seems to enhance learning better than hip hop music.

2. Two anti-depressant drugs were tested on both moderately depressed and severely depressed people. The table below gives the results after 12 weeks of treatment. Drug B has a better overall improvement rate since 55% of their users had significant improvement compared to only 41.67% of Drug A users.

	Drug A			Drug B		
	# Improved	# Not-Improved	% Improved	#Improved	# Not-Improved	% Improved
Severely Depressed	40	160	20%	30	170	15%
Moderately Depressed	85	15	85%	300	100	75%
Total	125	175	41.67%	330	270	55%

a) Which **2 percentages** on the table are the most relevant for doctors to compare when considering giving Drug A or Drug B to **moderately depressed** patients?   _____%   vs _____%

Which drug works best for moderately depressed people?   Drug A   Drug B   Cannot be determined

b) Which **2 percentages** on the table are the most relevant for doctors to compare when considering giving Drug A or Drug B to **severely depressed** patients?   _____%   vs _____%

Which drug works best for severely depressed people?   Drug A   Drug B   Cannot be determined

c) Drug B had a higher improvement rate for moderate depression that it did for severe depression (75% vs. 15%) because more moderately depressed than severely depressed took Drug B (400 moderate vs. 200 severe).

i)   True, the more people taking a drug, the higher the improvement rate.

ii)   False, measuring **rates** makes the difference between 400 and 200 irrelevant.

d) If Drug A has a better improvement rate for both severe and moderate depression, then why does it have a lower overall improvement rate?   Circle only one.

i)   It doesn't- there must have been an arithmetic error made in calculating the overall percents.

ii)   It's lower because Drug A had a total of only 300 subjects whereas Drug B had a total of 600.

iii)   It's lower because Drug A has a relatively higher percentage of severely depressed subjects, and the severely depressed have a lower overall improvement rate.

3. According to a recent study published in the Journal of the American Medical Association (JAMA) infants breast-fed for nine or more months grew up to be significantly more intelligent than infants breast-fed for one month or less. The study examined IQ scores of over 3,000 young men and women born in Copenhagen, Denmark and found that those who had been breast fed for at least nine months scored an average of six IQ points higher than those who had been breast fed for less than 1 month. (When the babies were 1 year old the mothers were questioned on how long they had chosen to breast feed their babies.)

a) Is this an observational study or a designed experiment?   Observational Study   Designed Experiment

b) Does the study show that breast feeding causes higher intelligence?

i)   Yes, the study shows definite causation although the causal mechanism can only be inferred.
ii)   No, it only shows that there is an association between breast feeding and higher intelligence. It does not show that breast feeding causes higher intelligence.

c) Which of the following could confound the results?   Circle only one.

i)   There could be nutrients in human milk that may help brain development.
ii)   Nursing could cause better mother-infant bonding which could help improve children's learning.
iii)   Breast-feeding mothers could be wealthier, better-educated and more concerned about infant development, all factors that could contribute to their children scoring higher on IQ tests.

Question 4 and 5 pertain to the following study:
A study was done to test the effectiveness of a new diet pill.
The subjects were 1000 adult volunteers. Half were randomly selected to take the diet pill daily and half were randomly selected to take a placebo daily. Neither the subjects nor those who evaluated their health knew who was in which group. All the subjects were followed for 1 year. **There was no difference in average weight loss between the 2 groups.**

4a) Is this observational study or a designed experiment?   Observational Study   Designed Experiment

4b) Were there controls? _____   If so, were they randomized?_____

4c) Was the study double-blind?_____

5) Even though everyone was encouraged to take the medicine daily, not everyone did. Those who faithfully took their medicine were called "adherers" and those that didn't were called "non-adherers". The adherers lost an average of 8.1 lbs over the year and the non-adherers lost an average of 2.3 lbs over the year as shown in the table below:

	Ave # of pounds lost per person per year on Diet Pill	Ave # of pounds lost per person per year on Placebo
Adherers	8.1	8.2
Non-Adherers	2.3	2.2
Total	5.2	5.2

a) Since those who faithfully took the diet pill lost more weight than those who neglected to take it, do the results from the table above show that the diet pill works better than a placebo?   Yes   No

b) Which one of the following could best explain why the adherers lost more weight than the non-adherers? Circle exactly two.

i) Those who took the diet pill lost more weight because the diet pill helped to boost their metabolism.

ii) Adherers are different than non-adherers; those that were responsible about taking their medicine may also be more responsible about their health in general (i.e. have better eating and exercise habits.)

        iii) Those who weren't losing much weight figured either the pill wasn't working or they must be in the placebo group
and   quit taking it.

iv) The diet pill cannot work if you don't take it. Non-adherers didn't lose as much weight because they didn't take the diet pill.

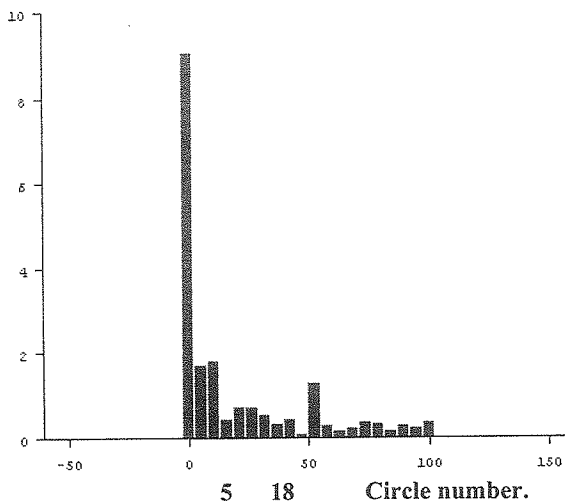Study Guide for Exam 1

Chap 3: Histograms
- Be able to read and interpret histograms.
- Be able to draw histograms
- Be able to tell how the data is distributed from the shape of a histogram.

Sample Questions on Histograms
1. Below are 3 histograms representing 3 variables in the Beginning Survey data: number of pets owned, **percent** of time spent dieting over the last 2 years, and number of hours of sleep per night. **Fill in the blanks below with the correct letters to match the histograms with the variables they represent.**

a) Histogram _____ represents sleep, histogram_____ represents pets, and histogram _____ represents diet.
b) Below each graph are 2 numbers. (They're not in any particular order.) One is the average and the other is the median. For each histogram, **circle the number which is the median**.
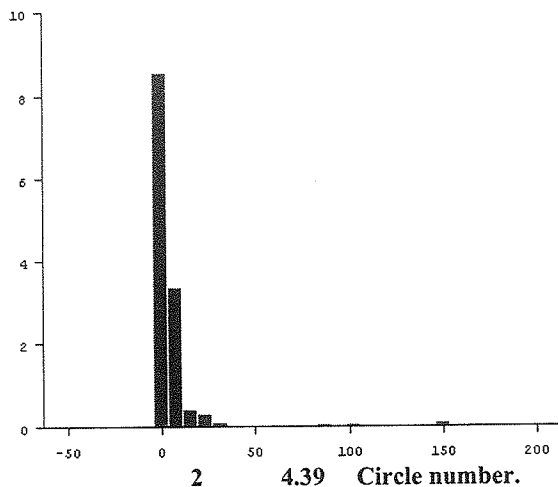
Histogram A

5   18   Circle number.

Histogram B

6.754   7   Circle number.

Histogram C

2   4.39   Circle number.
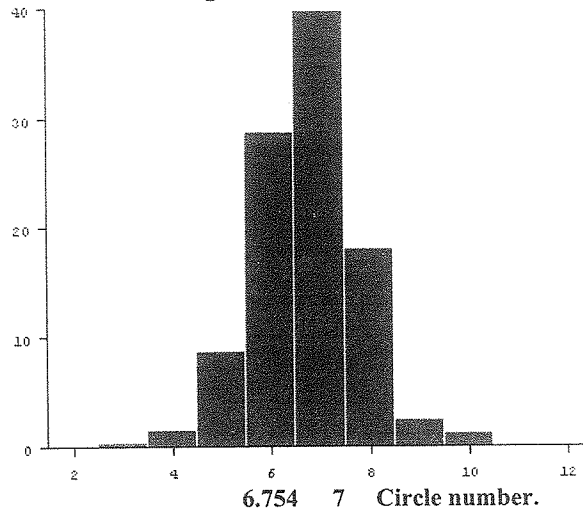
2. The figure below is a histogram for the number of alcoholic drinks consumed per week by about 600 Stat 100 students (roughly based on a past semester's survey data). Class intervals include the left-endpoint but not the right. (For example someone who drinks 10 alcoholic beverages per week would fall in the 10-20 block not the 5-10 block.) The height of each block is given in parentheses.
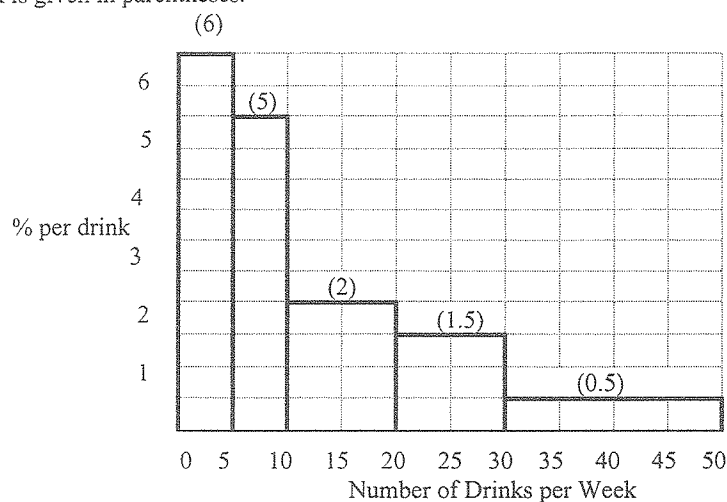
a)   20% of the subjects consumed 10-20 drinks. What percentage of the subjects fell in the following intervals?

0-5 drinks_____%   5-10 drinks_____%   20-30 drinks_____%   30-50 drinks_____%

b)   The median is closest to   5   6   9   15   20

c)   Is the median >, <, or = to the average? _____

d)   What percent of the subjects drank 10 or more drinks per week?   (Circle answer)

    45%   70%   75%   80%

e) Which interval has more people 0-5 or 20-50 or are they the same?   0-5   20-50   Same

f) The percent of subjects in the study who consumed 15 drinks per week is closest to (Assume an equal distribution throughout the interval)
       1%   2%   5%   10%   20%

g) If everyone in the 30-50 block reduced their drinking to 10-20 drinks per week, how would it affect the median, average and SD? (Fill in the blanks below with "increase", "decrease" or "stay the same".)

The median would _____, the average would _____and the SD would _____

3. A distribution table for the yearly wages of part-time student employees is shown below. Class intervals include the left endpoint but not the right. For example, the second row says that 20% of the employees earned $1 thousand, but less than $2 thousand. Fill in the column for the height of the blocks.

a)

Wages (in thousands)	%	Height(% per $1000)
0-1	15	
1-2	20	
2-3	15	
3-5	20	
5-11	30	

b) Draw the histogram below.

% per $1000

Wages (in thousands)
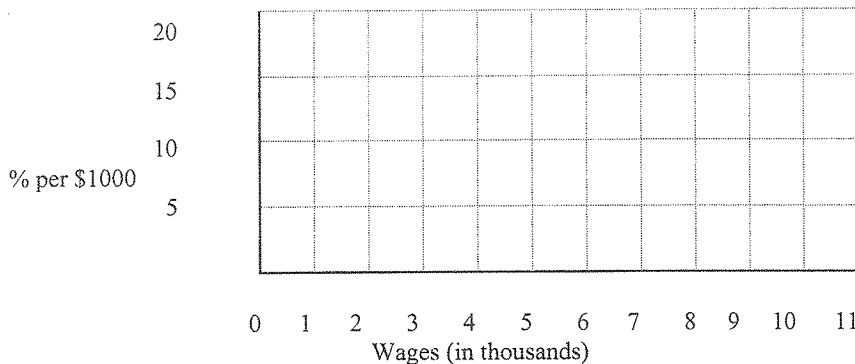
c) What is the median income? _____

d) Is the median >, <, or = the average? _____

Chap 4: Average, Median and Standard Deviation

- Compute average and median from a list of numbers.
- Know that average is sensitive to extreme values, median is not.
- Locate the median (50% of the area is above and 50% is below the median) and the average (where the histogram "balances") on a histogram.
- Long right-hand tail--average>median, long left-hand tail--average<median.
- Be able to calculate the SD for a list of numbers.
- How does the SD and the average change when a constant is added to all numbers on a list? How about when all the numbers are multiplied by the same value? What if that value is negative?

Sample Questions on the Average, Median and SD

1. Consider this list of numbers: 2, 5, 9, 11, 13

a) The average is _____

b) The median is _____ .

c) Compute the SD. Show work. Circle answer.

2. A list of numbers has an average of 300 and a SD of 30. Fill in the blanks below with the correct **numbers. (DO NOT** write "increased" or "decreased" in the blanks.)

a) If 1 is subtracted from each number on the list, the new average is _____ and the new SD is _____

b) If each number in the original list is tripled, then the new average is _____ and the new SD is_____

c) If each number in the original list is increased by 10%, then the new average is_____ and the new SD is_____

d) If each number in the original list is multiplied by -2, then the new average is _____ and the new SD is_____.

3. If the SD of a list of numbers is zero, which of the following statements is true? Circle true or false for each statement.

a)	The average must be zero.	True	False
b)	All the numbers on the list must be the same.	True	False
c)	All the numbers on the list must be zero	True	False

Chap 5: Normal Approximation
- Convert values to standard units.
- Find area under normal curve using table
- Do normal approximation
- Find the percentile given the score, and the score given the percentile.

Sample Questions on the Normal Approximation

1. According to the survey data, the histogram for the heights of the 346 women in this class is close to the normal curve with an **average of about 65" and a SD of about 3". (You may round z scores to fit the closest line on the table and you may round percents on the table to the nearest whole number.)**

a) About 68% of the women are between _____ inches and _____ inches.

b) About 95% of the women are between _____ inches and _____ inches.

c) One student listed her height as 70". How many SD's above average is her height? (In other words, what is her height in Standard Units?)

d) What percent of the females in the class are **over 72"?**   (Show work, draw a picture. Circle answer.)

e) One female listed her height as 60". What percentile is she in? (In other words, what percent of the females in the class is she taller than?) (Show work, draw a picture. Circle answer.)

f) What height corresponds to the 96th percentile? In other words, how tall is a student if she is taller than 96% of the other females in the class? (Show work, draw a picture. Circle answer.)

g) If all the women gave their heights wearing 1" inch heels, what would the new average and SD be? (The original heights are with no shoes on.)

New average_____ inches   New SD _____ inches

h) If the original heights (without shoes on) were all converted from inches to centimeters by multiplying by 2.54 cm/inch what would the average and SD in centimeters be?

Average _____cm   SD _____cm