AP Stats Ch 11,12,13 - Required Practice **Name:** Solutions **Per:** ____

Ch11

Simulation: Senior Parking

Suppose that 100 seniors at a particular school (including 19 members of Student Council signed up for "senior parking," which gives each student the right to a particular parking space close to the campus for the year. Also, suppose that each student's spot is determined by a lottery. The student body became suspicious, however, when Student Council members were awarded 5 of the 10 best spots. Is the suspicion warranted or could this have occurred by chance? Using the random digits below, design and conduct a simulation to estimate the probability of Student Council members getting at least 5 of the 10 best spots, assuming that the lottery is fair.

12975 13258 13048 45144 72321 81940 00360 02428

96767 35964 23822 96012 94951 65194 50842 55372

37609 59057 66967 83401 60705 02384 90597 93600

Component: assign one of the good spots

Model: 00-99: 00-18 19-99
council other students

Outcome: at least 5 are council or not (make it "categorical")

RV

trials

outcomes (#council related)	= Y/N ≥5?
3	= no
3	= no
1	= no
1	= no
1	= no
1	= no

The suspicion is warranted because it is very unlikely for 5 of the best 10 spots to go to student council members randomly.

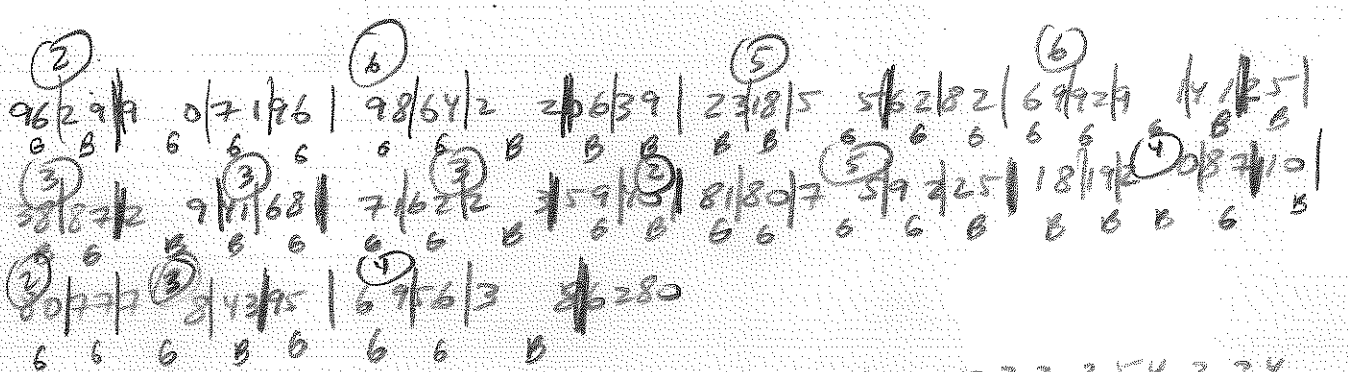
our experimental $P(\geq 5 \text{ council of } 10) = \frac{0}{6} = 0$

(theoretical: binomial model:

x	0 1 2 3 4 5 6 7 8 9 10
p	1-binomcdf(10, .19, 4) = .0266 (2.7%)

25. The family. Many couples want to have both a boy and a girl. If they decide to continue to have children until they have one child of each gender, what would the average family size be? Assume that boys and girls are equally likely.

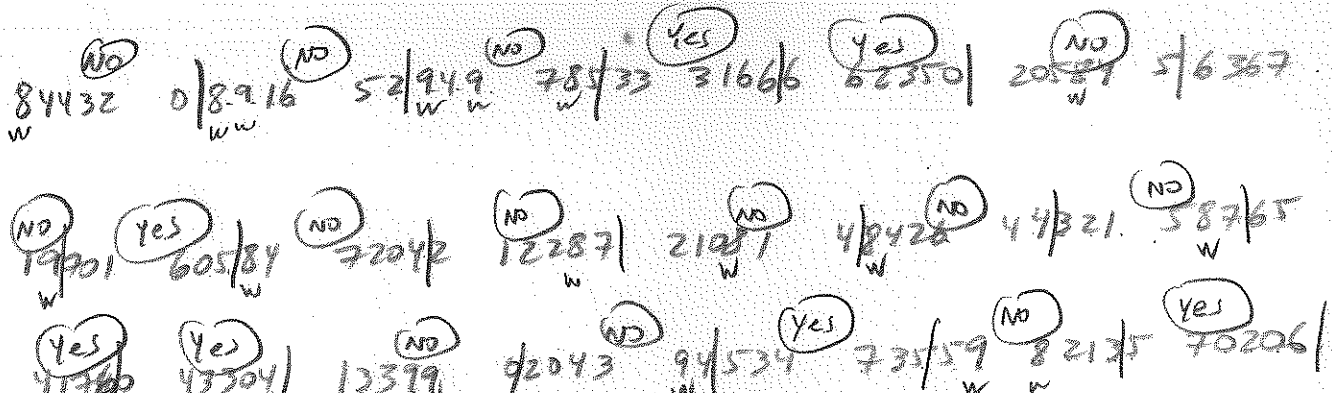
component: 1 child
outcome: boy, girl
model: 0–49, 50–99
boy girl
trial: 2 digit # = 1 child
continue until 1 of each
RV: # trials to get 1 of each



number of children required for 1 of each gender: 2, 6, 5, 6, 3, 3, 3, 2, 5, 4, 2, 3, 4
range: 2–6 kids
$\bar{x}$ = 3.7 kids

13. Multiple choice. You take a quiz with 6 multiple choice questions. After you studied, you estimated that you would have about an 80% chance of getting any individual question right. What are your chances of getting them all right? Your simulation should use at least 20 runs.

component: one multiple-choice question
outcome: right, wrong
model: 0–7 right, 8–9 wrong
RV: all correct (or not)



total trials = 20
all correct = 7
simulated P(all correct) = $\frac{7}{20}$ = 0.35

** SEED any #, STO, rand

Ch 12 – "The River Problem"

Suppose we wanted to estimate the yield of our corn field. The field is square and divided into 16 equally sized plots (4 rows x 4 columns). A river runs along the eastern edge of the field. We want to take a sample of 4 plots.

Using a random number generator, pick a simple random sample (SRS) of 4 plots. Place an X in the 4 plots that you choose.

use calc randInt (1, 16, 4)
= 8, 4, 3, 6

1	X	3	X
5	X	7	X
9	10	11	12
13	14	15	16

river

Now, randomly choose one plot from each horizontal row. This is called a stratified random sample.

use calc randInt (1,4) four times
= 4, 2, 1, 2

1	2	3	X
1	X	3	4
X	2	3	4
1	X	3	4

river

Finally, randomly choose one plot from each vertical column. This is also a stratified random sample.

use calc randInt (1,4) four times
= 3, 2, 4, 3

1	1	1	1
2	X	2	2
X	3	3	X
4	4	X	4

river

Which method is most likely to produce a sample mean closest to the population mean? Explain.

Not sure, are the plots all the same?

It's time for the harvest! The numbers below are the yield for each of the 16 plots. For each of your three samples above, calculate the average yield.

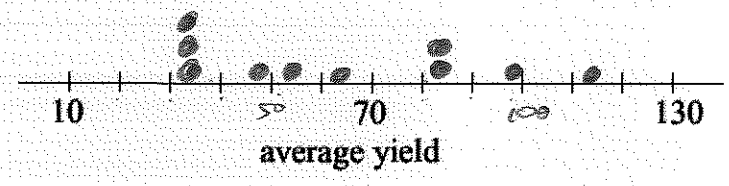
4	29	94	150
7	31	98	153
6	27	92	148
5	32	97	147

whole class results

my $\bar{x} = \frac{29+150+31+153}{4}$

$= 90.25 \approx 91$

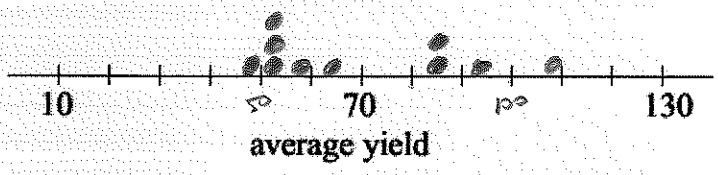
SRS:



my $\bar{x} = \frac{150+31+6+32}{4}$

$= 54.25 \approx 55$

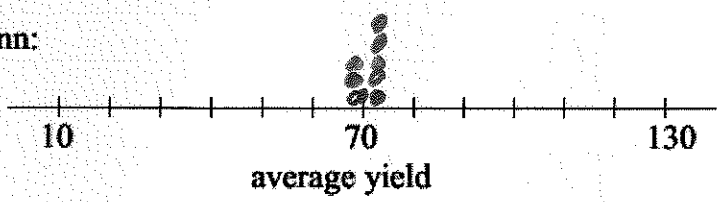
Stratified by Row:



my $\bar{x} = \frac{31+6+148+92}{4}$

$= 69.25 \approx 70$

Stratified by Column:



Which method is best, and why? Stratified by column. The yields in each column are similar w/in the column, but different from other columns. The most accurate representation comes from including one plot from each column.

1–10. What did they do? For the following reports about statistical studies, identify the following items (if possible). If you can't tell, then say so—this often happens when we read about a survey.

a) The population
b) The population parameter of interest
c) The sampling frame
d) The sample
e) The sampling method, including whether or not randomization was employed
f) Any potential sources of bias you can detect and any problems you see in generalizing to the population of interest

5. Researchers waited outside a bar they had randomly selected from a list of such establishments. They stopped every 10th person who came out of the bar and asked whether he or she thought drinking and driving was a serious problem.

a) population: adults

b) parameter: whether or not they think drinking and driving is a problem.

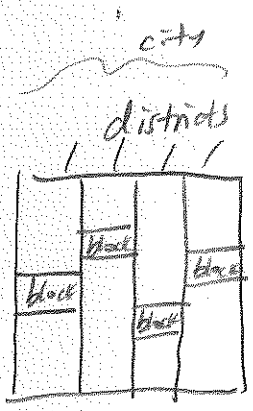
c) sampling frame: adults leaving this particular bar.

d) sample: every 10th person leaving this bar

e) method: systematic (and convenience)

f) undercoverage bias: sample does not include people who don't drink or don't drink in bars.

6. Hoping to learn what issues may resonate with voters in the coming election, the campaign director for a mayoral candidate selects one block from each of the city's election districts. Staff members go there and interview all the residents they can find.



a) population: voters in city election

b) parameter: issues of concern

c) frame: residents in election districts

d) sample: as many residents as they could find in the selected blocks

e) multistage: 1) stratified by district
2) clustered by block
3) convenience sample within the blocks

f) undercoverage bias: due to the convenience sampling residents not available to be found are underrepresented.

21. Survey questions. Examine each of the following questions for possible bias. If you think the question is biased, indicate how and propose a better question.

a) Should companies that pollute the environment be compelled to pay the costs of cleanup?

b) Given that 18-year-olds are old enough to vote and to serve in the military, is it fair to set the drinking age at 21?

a) Suggests prejudgement about companies and suggested action.
"If a test of soil/water near a company indicates contamination, do you believe that the company should pay for the cost of cleanup efforts?"

b) This question leads respondents to the conclusion that the drinking age is unfair.
"Do you believe the drinking age should be 21?"

27. Accounting. Between quarterly audits, a company likes to check on its accounting procedures to address any problems before they become serious. The accounting staff processes payments on about 120 orders each day. The next day, the supervisor rechecks 10 of the transactions to be sure they were processed properly.

a) Propose a sampling strategy for the supervisor.

b) How would you modify that strategy if the company makes both wholesale and retail sales, requiring different bookkeeping procedures?

a) Assign numbers 1–120 to the orders, and write these numbers on identical slips of paper. Put the slips in a box and mix thoroughly, then randomly draw 10 slips to form the sample.

b) Stratify the orders first (wholesale and retail) and number each strata separately, with separate slips of paper. Randomly select 5 from the wholesale orders box and 5 from the retail orders box.

Ch13

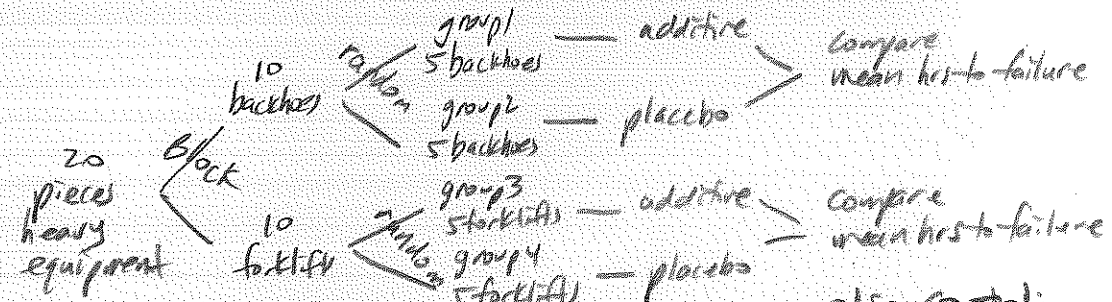
Backhoes & Forklifts

A heavy equipment manufacturer introduces new models of their backhoes and forklifts, and sells several of these machines. A few months later they discover a problem. An unacceptably high number of their customers needed to seek repairs of the machines' hydraulic systems within the first few weeks of operation.

The company's engineers go to work on this problem, and soon think that they have found a solution. They believe that an additive poured into the hydraulic oil may greatly extend the number of hours these machines can be used before repairs become necessary. A few tests conducted under laboratory conditions indicate that this solution shows promise, but "real world" customer experience is needed before they can be sure. Impressed by these preliminary results, the company's management gives the research team the green light to test their theory using up to 20 newly manufactured machines that will be sold during the next few weeks.

It is your job to design the experiment by specifying the procedure the company should use. Be sure to use the appropriate vocabulary throughout your description.

(a) Draw a design diagram showing how you would design this experiment.



(b) For your experiment identify…

…the subjects: the backhoes and forklifts

also control:
- how the machines are driven (same drivers, same driving/use conditions)

…explain your randomization procedure, showing the resulting assignments:
Number the 10 backhoes 1–10, and use a single random digit to select one of the backhoes to include in group1. Use another digit to select another backhoe (ignore repeats) and keep selecting until 5 are selected as group 1. The rest of the backhoes are placed in group 2. Repeat this procedure to select the forklifts for groups 3 and 4.

…the factors and numbers of levels for each: 1 factor: additive with 2 levels: additive, no additive

…the number of treatments: 2 treatments (additive, no additive)

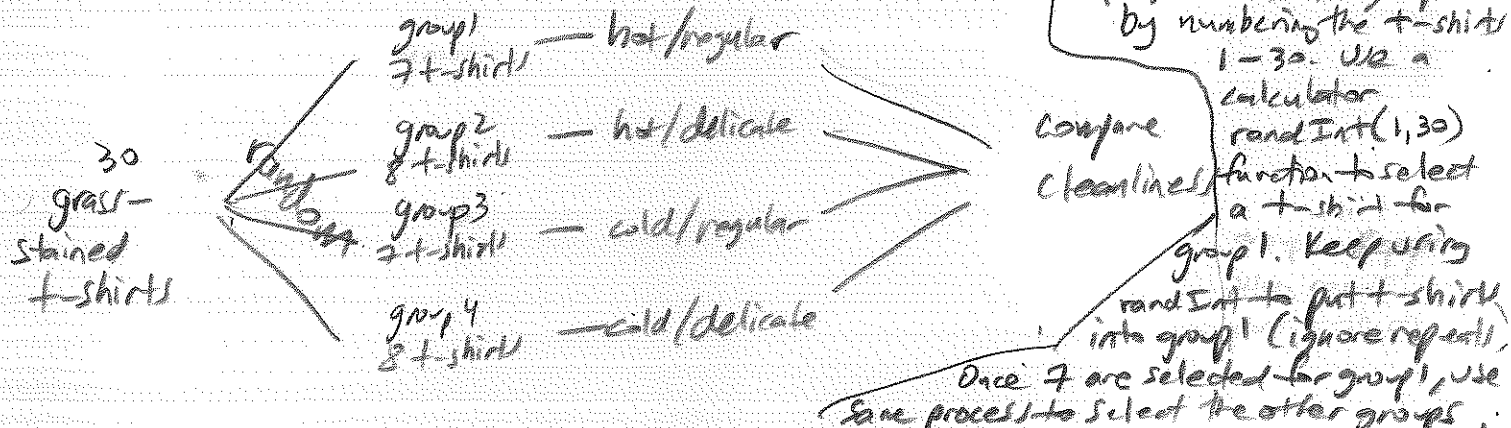
…whether or not the experiment is blind or double-blind: could be double blind:
- make a placebo oil additive so mechanics don't know which grp equip is in.
- make sure people driving the equip don't know.

…the response variable: number of hours until failure.

…how you would decide whether the additive has solved the problem:
Additive is successful if the number of hours until failure is significantly increased.

42. Washing clothes. A consumer group wants to test the effectiveness of a new "organic" laundry detergent and make recommendations to customers about how to best use the product. They intentionally get grass stains on 30 white T-shirts in order to see how well the detergent will clean them. They want to try the detergent in cold water and in hot water on both the "regular" and "delicates" wash cycles. Design an appropriate experiment, indicating the number of factors, levels, and treatments. Explain the role of randomization in your experiment.

Factors/levels:
water temp / hot, cold
cycle / regular, delicate
(2 factors, 2 levels each)
Treatments/groups: (2)(2) = 4
Randomization: we will randomly assign t-shirts to the groups by numbering the t-shirts 1–30. Use a calculator randInt(1,30) function to select a t-shirt for group 1. Keep using randInt to put t-shirts into group 1 (ignore repeats). Once 7 are selected for group 1, use same process to select the other groups.



29. Frumpies. The makers of Frumpies, "the breakfast of rug rats," want to improve their marketing, so they consult you:

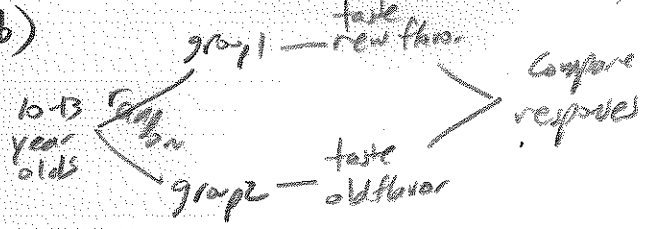
a) They first want to know what fraction of children, ages 10 to 13, like their celery-flavored cereal. What kind of study should they perform?

b) They are thinking of introducing a new flavor, maple-marshmallow Frumpies, and want to know whether children will prefer the new flavor to the old one. Design a completely randomized experiment to investigate this question.

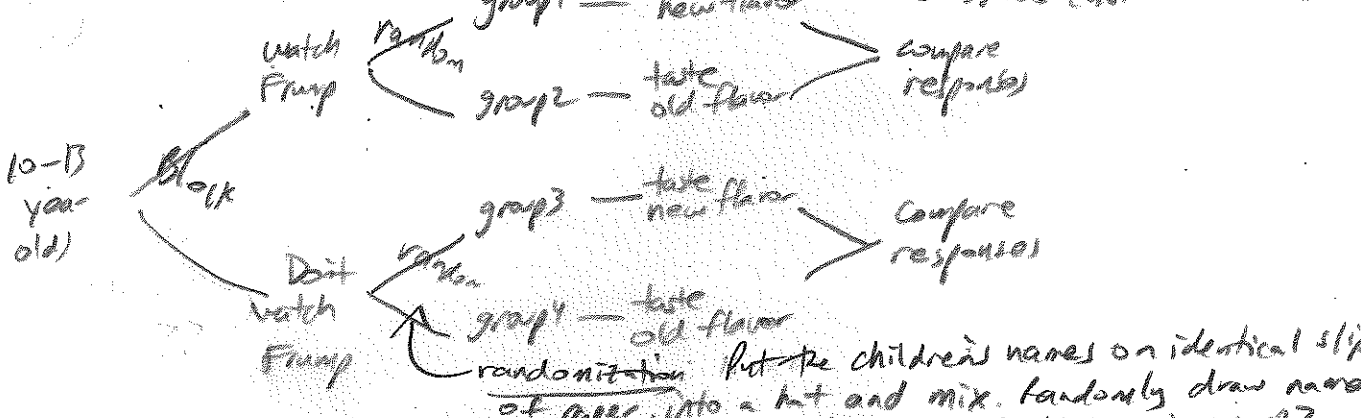
c) They suspect that children who regularly watch the Saturday morning cartoon show starring Frump, the flying teenage warrior rabbit who eats Frumpies in every episode, may respond differently to the new flavor. How would you take that into account in your design?

a) They should perform an experiment on randomly selected 10–13 year olds using a questionnaire, after they taste it: "do you like this cereal?"

b)



c) block on viewing Frump cartoon — they already watch or don't watch Frump, so this is an aspect of the subjects (not a treatment)



Put the children's names on identical slips of paper, into a hat and mix. Randomly draw names to form group 1 — the rest of children in group 2 (repeat for groups 3 & 4)

40. SAT Prep. Can special study courses actually help raise SAT scores? One organization says that the 30 students they tutored achieved an average gain of 60 points when they retook the test.

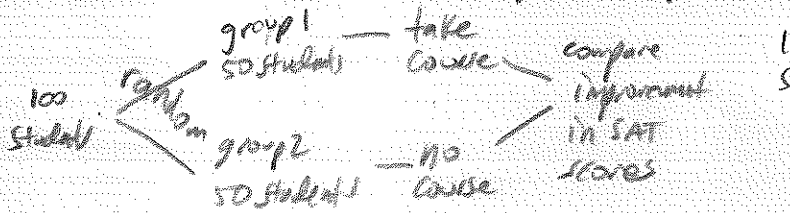
a) Explain why this does not necessarily prove that the special course caused the scores to go up.

b) Propose a design for an experiment that could test the effectiveness of the tutorial course.

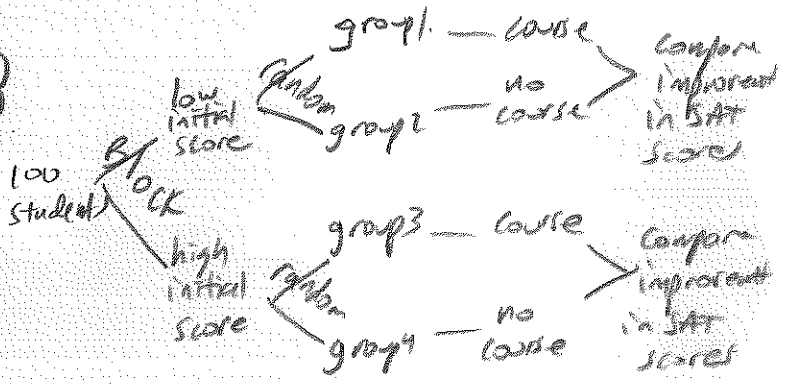
c) Suppose you suspect that the tutorial course might be more helpful for students whose initial scores were particularly low. How would this affect your proposed design?

a) There was no random assignment to groups — those students in the study group may have differences with the non-study group (parents w/higher expectations, better self-motivation, etc.) So this is not really an experiment and can't establish cause-and-effect.

b) Select 100 students at random from one HS who had taken the SAT. 50 are randomly assigned to take the course and the others don't (by using calculator randInt(1,100) and ignoring repeats)



c) Could block on initial score:



43. Skydiving, anyone? A humor piece published in the *British Medical Journal* notes ("Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomized control trials," Gordon, Smith, and Pell, *BMJ*, 2003:327) that we can't tell for sure whether parachutes are safe and effective because there has never been a properly randomized, double-blind, placebo-controlled study of parachute effectiveness in sky diving. (Yes, this is the sort of thing statisticians find funny . . .) Suppose you were designing such a study:

a) What is the factor in this experiment?

b) What experimental units would you propose?[6]

c) Explain what would serve as a placebo for this study.

d) What would the treatments be?

e) What would be the response variable for such a study?

f) What sources of variability would you control?

g) How would you randomize this "experiment"?

h) How would you make the experiment double-blind?

a) 1 factor: parachute: working/not working

b) volunteers to skydive

c) fake parachutes which don't work

d) one group gets working parachutes the other group gets non-working parachutes

e) condition of the jumper on the ground.

f) – same altitude
– same surface
– same weather & time of day

g) names on slips of paper to randomly select some volunteers for one group

h) – jumpers and those working with the jumpers don't know which group each jumper is in.
– the people evaluating the medical condition on the ground after jump aren't told which group each jumper was in.

Chapter 11,12 Practice Quiz ***(for Ch13, use the 'Backhoes & Forklifts' problem to review)***

AP Statistics Quiz B – Chapter 11 ***Name*** _____

On January 1 of every year, many people watch the Rose Parade on television. The week before the parade is very busy for float builders and decorators. Roses, carnations, and other flowers are purchased from around the world to decorate the floats. Based on past experience, one float decorator found that 10% of the bundles of roses delivered will not open in time for the parade, 20% of the bundles of roses delivered will have bugs on them and be unusable, 60% of the bundles of roses will turn out to be beautiful, and the rest of the bundles of roses delivered will bloom too early and start to discolor before January 1. Conduct a simulation to estimate how many roses the float decorator will need to purchase to have 15 good bundles of roses to place on the float.

1. **Describe how you will use a random number table to conduct this simulation.**

component: selecting a bundle of roses
outcomes: good, bad
Model: 0–59, 66–99
trial: keep selecting bundles until we have 15 good bundles
response variable: # of bundles it took to get 15 good ones

2. **Show three trials by clearly labeling the random number table given below. Specify the outcome for each trial.**

trial1 (21)
37542 04805 64894 74296 24805 24037 20636 10402 00822

trial2 (22)
08422 68953 19645 09303 23209 02560 15953 34764 35080

trial3 (24)
99019 02529 09376 70715 38311 31165 88676 74397 04436

12807 99970 80157 36147 64032 36653 98951 16877 12171

range: 21–24
$\bar{x}$ = 22.3 bundles

3. **State your conclusion** This simulation shows you will need to purchase between 21 and 24 (on average 22.3) bundles to get 15 good ones.

Statistics Quiz A – Chapter 12 ***Name*** _____

1. A statistics teacher wants to know how her students feel about an introductory statistics course. She decides to administer a survey to a random sample of students taking the course. She has several sampling plans to choose from. Name the sampling strategy in each.
   a. There are four ranks of students taking the class: freshmen, sophomores, juniors, and seniors. Randomly select 15 students from each class rank. — Stratified random sample
   b. Randomly select a class rank (freshmen, sophomores, juniors, and seniors) and survey every student in that class rank. — cluster
   c. Each student has a nine-digit student number. Randomly choose 60 numbers. — Single random sample (SRS)
   d. Using the class roster, select every fifth student from the list. — Systematic

2. Explain why the second plan suggested above, sampling all students from one class rank, might be biased. Be sure to name the kind(s) of bias you describe.

(name) → Undercoverage bias

(desribe) → The selected rank may not be representative of all ranks.

3. Listed below are the names of 20 students who are juniors. Use the random numbers listed below to select five of them to be in your sample. Clearly explain your method.

0,1	2,3	4,5	6,7	8,9 ← column
Adam	Chris	Dave	Deirdre	Dick
Ellen	Eric	Joan	John	Judi
Joy	Kenny	Laura	Mary	Paul
Peter	Rachel	Rob	Sara	Stacey

row 0,1 / 2,3 / 4,5 / 6,7 / 8,9

Sample will be: Dave, Chris, Laura, Joan

(this is just one possibility)

digits represent a row (or column) use them to select a row then a column to include that student in the sample (and ignore repeated students)

(ignore 5/9) Dave Chris Laura Joan

04028 55259 81183 40754 60209 06765 27306

RC RC RC C

28370 82669 83236 77479 90618 43707 78695

4. Name and describe the kind of bias that might be present if the statistics teacher decides that instead of randomly selecting students to survey on how they feel about the course she just..
   a. asks students to volunteer for the survey.

   volunteer response bias : bias towards more gregarious students or students who like the course.

   b. gives the survey during class one day.

   undercoverage bias : students who are absent are underrepresented.