Homework #6 Solutions

3.12, 3.~~23~~, 3.28, 3.37, 3.52 3.53

Problem 3.12

(a) response – colon cancer rate

explanatory – vitamin/anti-oxidant usage

(b) Using random assignment assign 216 subjects to each group ie

daily beta carotine	daily vitamin C and E	All three vitamins	daily Placebo
216 subjects	216 subjects	216 subjects	216 subjects

At the conclusion of the study compare colon cancer rates across each of the four different groups

(c) Label the subjects with numbers 1 through 864. Then the first 5 subjects chosen will be 731, 253, 304, 470, 296

"Double blind" means:

(d) Neither the subject nor the person treating the subject knows which treatment the subject is receiving.

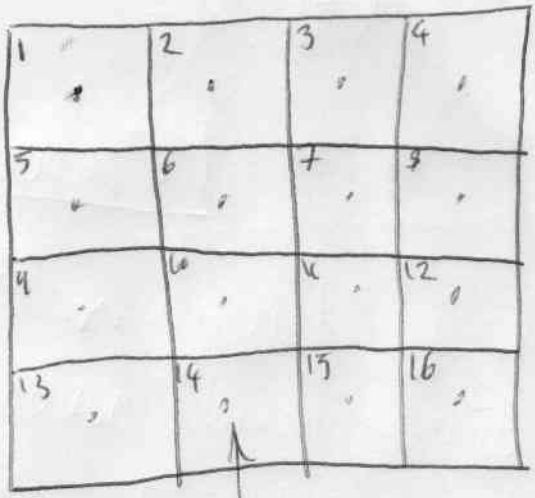
This is done to avoid potential bias.

(e) "No significant difference" means that when variability is taken into account the colon cancer rates between each of the four different treatments does not change.

(f) One potential lurking variable would be quantity of dietary fiber consumed. Since, along with many vitamins, fruit and vegetables have large amounts of fiber.

Another potential lurking variable could be related to quantity of other foods eaten. For instance, perhaps people who substitute fruit and vegetables for high salt-high fat diets are removing potential disease causes from their diet.

Problem 3.23



a pole location

the field

the numbering and grids will help us assign colors to the poles

The important thing about this problem is how you assign the colors. Note that "randomly" does not mean helter-skelter or just how you feel like doing it. You have to have a method of assigning colors using random number or similar.

First I decide to assign the 4 blue, then 4 green, 4 white and 4 yellow. How do I do this? Make tickets with numbers 1, 2, ..., 16 on them. Put all the tickets into a box and shake vigorously. Then 1 by 1 I draw tickets from the box. The first four numbers correspond to the four poles which get blue. The next four correspond to green, the next four to white and the remaining four to yellow.

My drawing order:

3, 13, 4, 15, 14, 6, 5, 2, 10, 16, 7, 11, 8, 1, 12, 9

So

blue: 3, 13, 4, 15
green: 14, 6, 5, 2
white: 10, 16, 7, 11
yellow: 8, 1, 12, 9

Y	G	B	B
G	G	W	Y
W	W	Y	Y
B	G	B	W

Problem 3.28

(a) Factors: 1. Time question presented
2. Type of question

levels

Time question presented: before, after (2)

Type of question asked: simple fact, computation, word problem (3)

There are (2)(3)=6 different treatments

trt 1: before ; simple fact
trt 2: before ; computation
trt 3: before ; word problem
trt 4: after ; simple fact
trt 5: after ; computation
trt 6: after ; word problem

} the six possibilities listed.

(b) Assume we start with 12 classes. Assign each class a number between 1 and 12. Using a computer we draw numbers between 1 and 12 in particular we get

1, 9, 4, 5, 2, 6, 8, 3, 11, 10, 7, 12

Use these to assign treatments as follows.

	Simple fact	Computation	word problem
Before	1, 9	4, 5	2, 6
After	8, 3	11, 10	7, 12

After applying the treatments measure the average math score for all students in the classes who received that treatment.

Problem 3.37

The population is the words used in this particular Tom Clancy novel.

The sample is the first 250 words on the randomly chosen page.

The variable measured is word length.

Problem 3.52

The period with higher "ring-no-answer" is probably the one where there are fewer people at home.

plausible reasons for July 1 – August 31
- summertime months. people might be away on vaction or out due to the longer daytime hours

Plausible reasons for Jan 1 – Easter
- Religous holidays?

In either case a high non response is bad because it may lead to a bias in the results of the survey. For instance, people who are more likely to out (eg young people) might well have different opinions than those more likely to be home (eg elderly, homebodies)

Problem 3.53

a) The sample size is

$$13147 + 15182 + 1448 = 29777$$

b) There is a number of reasons why the results of this poll can not be trusted

- self selection bias - respondents choose whether or not to answer the poll.
- undercoverage bias - only potential users of news.excite.com are likely to be reached

For these reasons and others it is very unlikely that the sample is at all representative of the US population at large or even the set of all excite users

c) If more men than women use the web the difference between genders in opinions on this issue then the results will be skewed towards the male opinion.