

Question 1: (Marked out of 22)

(HW_1_1) Consider the following four studies:

Study 1: A researcher was interested in whether pain tolerance levels were related to hair color. 80 people were selected from a group of volunteers, 20 with light blonde hair, 20 with dark blond hair, 20 brunettes and 20 redheads. The subjects underwent a series of tests and pain tolerance scores (on a scale of 0 to 100) were assessed.

Study 2: A technician is interested in the effects of using different baking temperatures on the impact strength of particle board. 20 boards are randomly allocated to 20 different baking temperatures. After the boards are baked, they are sent to a laboratory where the impact strengths are measured.

Study 3: The manager of an auto repair shop is interested in whether using a new diagnostic machine will speed up the regular servicing of cars. There are two mechanics working on regular services, one with 8 years experience while the other mechanic had only 2 years experience. One mechanic was told to use the diagnostic machine on the next 10 cars she serviced, but not use it on the following 10 cars. The other mechanic was told not to use the diagnostic machine on the next 10 cars he serviced, but to use it on the following 10 cars. Each mechanic recorded the time it took to complete the services for each of these jobs.

Study 4: A sociologist is interested in comparing the exam results for male and female students on 10 different subjects. The proper authority was contacted to obtain the numbers of male and female students who took the exam and the numbers of male and female students who got each of the grades A, B, C and D.

- For each study, describe what *treatment* is being compared and what *response* is being measured to compare the treatments.

Study 1: Treatment - hair color (1)
Response - pain tolerance level (1)

Study 2: Treatment - baking temperature (1)
Response - impact strength (1)

Study 3: Treatment - use of new diagnostic machine (1)
Response - service time (1)

Study 4: Treatment - gender (1)
Response - exam result (1)

- Which of the studies would be described as *experiments* and which would be described as *observational* studies?

Study 1: Observational study (1)

Study 2: Experiment (1)

Study 3: Experiment (1)

Study 4: Observational study (1)

- For the studies that are observational, could an experiment have been carried out instead? If not, briefly explain why not.

Study 1: No. We have no control over which hair color we assign to each subject (assuming we are talking about natural hair color). (2)

Study 4: No. We have no control over which gender we assign to each student. (2)

[1 point for answer, 1 point for explanation.]

- For the studies that are experiments, briefly discuss what *forms of blinding* would be possible to be used.

Study 2: The person testing the impact strength of the particle boards in the laboratory does not know which temperature each board has been baked at. (2)

Study 3: The person analyzing the data is given a list of service times for each car, but does not know which list corresponds to each service method (i.e. using new diagnostic machine or not). (2)

or:

The mechanic does not know whether they are using the new diagnostic machine or the old machine, as the machines have been disguised so that they both look the same.

[There are a wide variety of other possible correct answers for this question.]

- In which of the studies has *blocking* been used? Briefly describe **what** was blocked and why it was blocked.

Study 3: Blocking by mechanic has been used here. This was done because we think that the difference in work experience between mechanics may also have an effect on the time for service. (2)

[Depending on how you interpret the description of each study, you could also say that study 4 contains blocking by subject. Students should explain their reasoning if they decide this is the case.]

Question 2: (Marked out of 18)

(HW_1_2) The following data represent the daily number of parking tickets given out on UCLA campus over a period of 29 days.

42, 47, 46, 35, 43, 39, 38, 40, 50, 37, 68, 37, 47, 44, 49
41, 34, 38, 41, 36, 42, 38, 38, 58, 34, 32, 42, 49, 52.

- By hand, construct a stem-and-leaf plot of the data using an appropriate scale.

```
3* | 244
3. | 567788889
4* | 01122234
4. | 67799
5* | 02
5. | 8
6* |
6. | 8
```

(3)

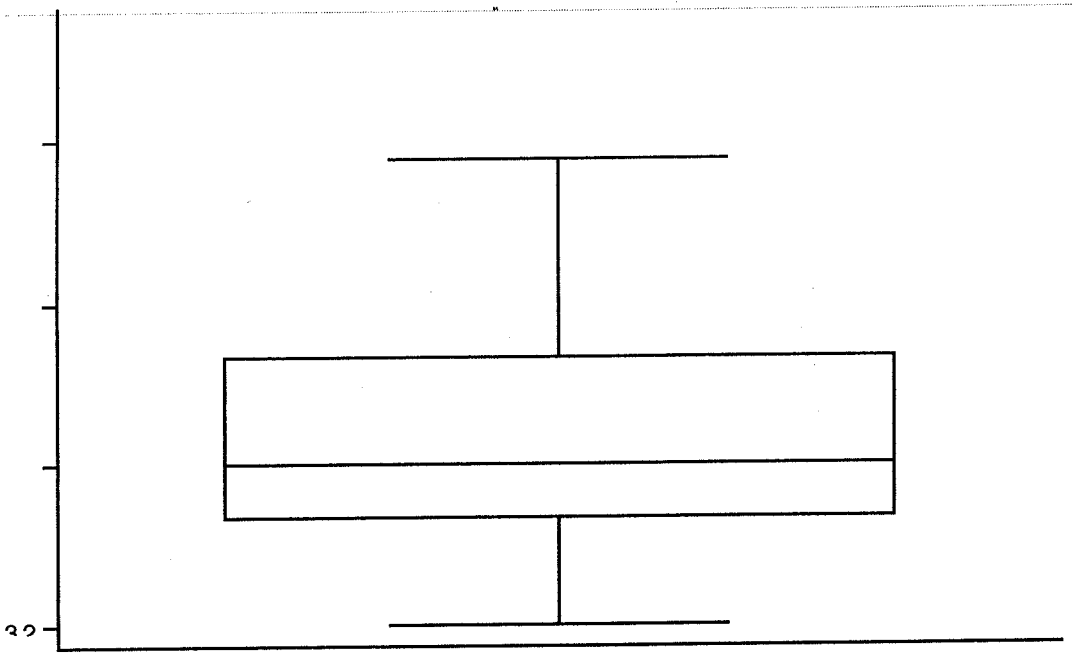
[The above stem-and-leaf plot is the one produced by Stata. Students may have a different number of rows in theirs, but each row should be the same “width” as the others. However, I think 4 rows here is too few for this example.]

- Calculate the five number summary for the data (**min**, **lower quartile**, **median**, **upper quartile** and **max**). Note: If there are x indices of numbers below the median then the $(x+1)/2$ -st number in the ordered observations, see stem-and-leaf plot, gives you the value of the lower quartile. The **upper quartile** is computed identically as the lower quartile, except that you take the value of the observation half-way between the median and the max-observation.

Min:	32	(1)
Lower quartile:	37.5	(1)
Median:	41	(1)
Upper quartile:	47	(1)
Max:	68	(1)

[Note that *median* was changed from *mean* after it had been posted on the internet. So some students who didn't hear about the change through lectures or sections may have calculated the mean instead (the mean here is 42.66).]

- By hand, draw a box plot for this set of data. Show your working. (5)



[The above box plot is the one produced by Stata. Students could have either used the computer to do the graph, or constructed it by hand.]

- Using your plots, in plain English, briefly comment on the data and the story the above summary/plots convey.

From our analysis we see that the number of parking tickets given out at UCLA over a period of 29 days is about 41 per day. The highest number of parking tickets given out on a single day during that period was 68, and the least number was 32. Most days had between 37 and 47 tickets given out. (5)

[Again, students may have slightly different answers, but they should explain themselves clearly. It is important to make sure to answer the original question. Because the question asks for an explanation *in plain English*, they shouldn't really be using terms such as "skewed", "outlier" or (my personal favorite!) "p-value". However, maybe we should be a bit lenient for this first assignment.]