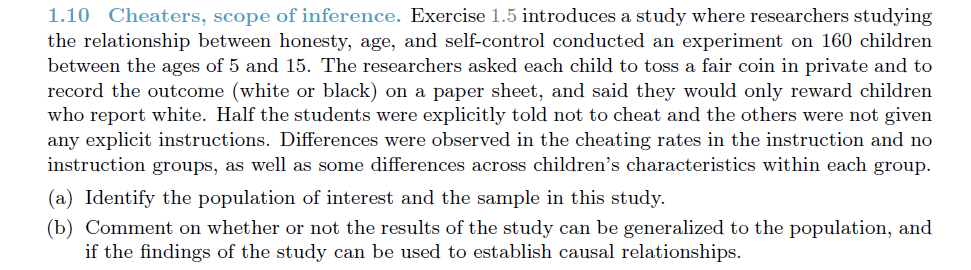


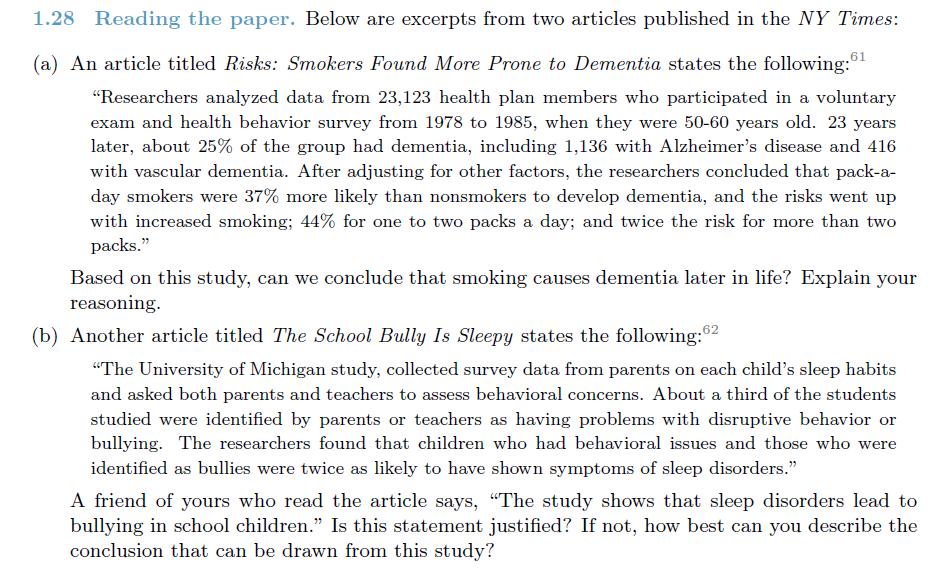
**Ans:**

1. Each row of the data matrix represents a UK resident.
2. 1691 participants were included in the survey
3. Variables
   1. Sex – Categorical
   2. Age – Numeric, Discrete
   3. Marital – Categorical
   4. Gross Income – Categorical, Ordinal
   5. Smoke – Categorical
   6. Amount Weekends – Categorical, Ordinal
   7. Amount Weekdays – Categorical, Ordinal



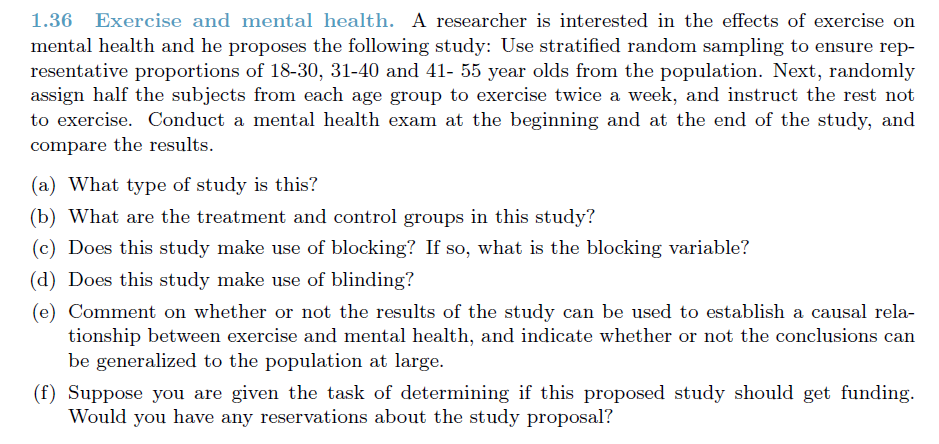
**Ans:**

1. **Population of interest:** All children between ages 5 and 15. **Sample**: 160 children
2. The results of the study can be generalized to the population and the findings can be used to establish causal relationships since this is an experiment that was conducted and it was randomized.



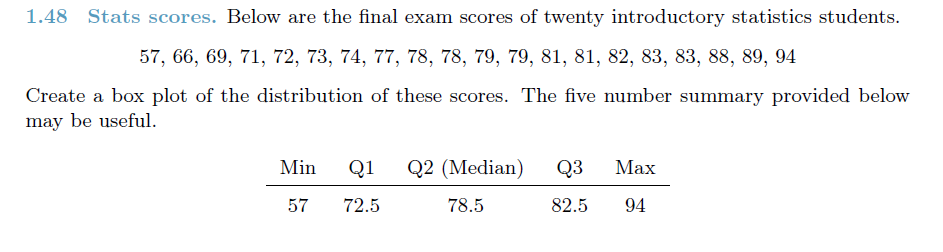
**Ans:**

1. This is not an experimental design. It is just an observational study. Hence the findings of this study cannot be used for any causal conclusions.
2. Again this is an observational study and hence we cannot make any causal connections.



**Ans:**

1. This is an **Experiment**
2. **Treatment group:** Asked to exercise twice a week, **Control group:** No Exercise.
3. The **age group** is the blocking variable.
4. There is no indication of the use of Blinding in the above explanation.
5. Since this is an experiment, we can make causal observations. However, we can only generalize for the population between 18 – 55 years. The experiment does not deal with other age groups above 55 and below 18.
6. As mentioned above the sampling strategy does not cover all age groups and hence the study would not be conclusive.



**Ans:**

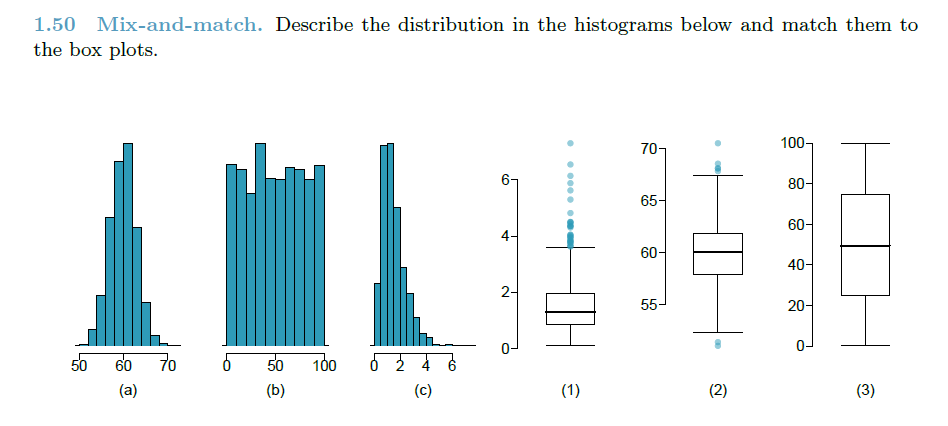
94

82.5

78.5

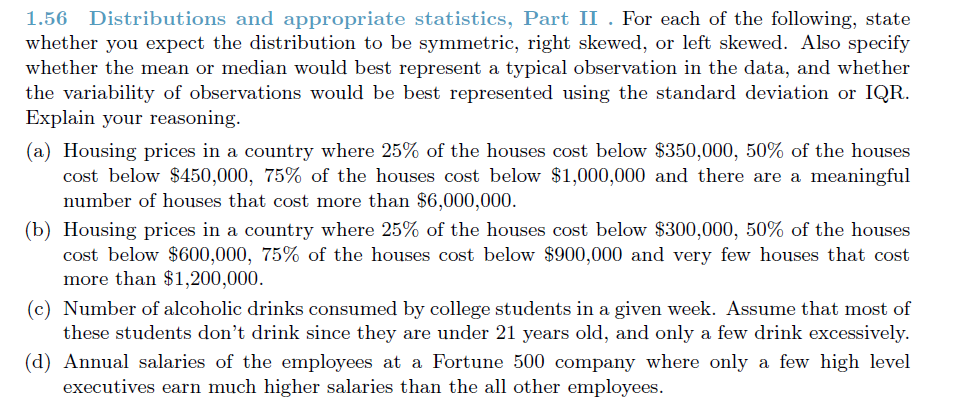
72.5

57



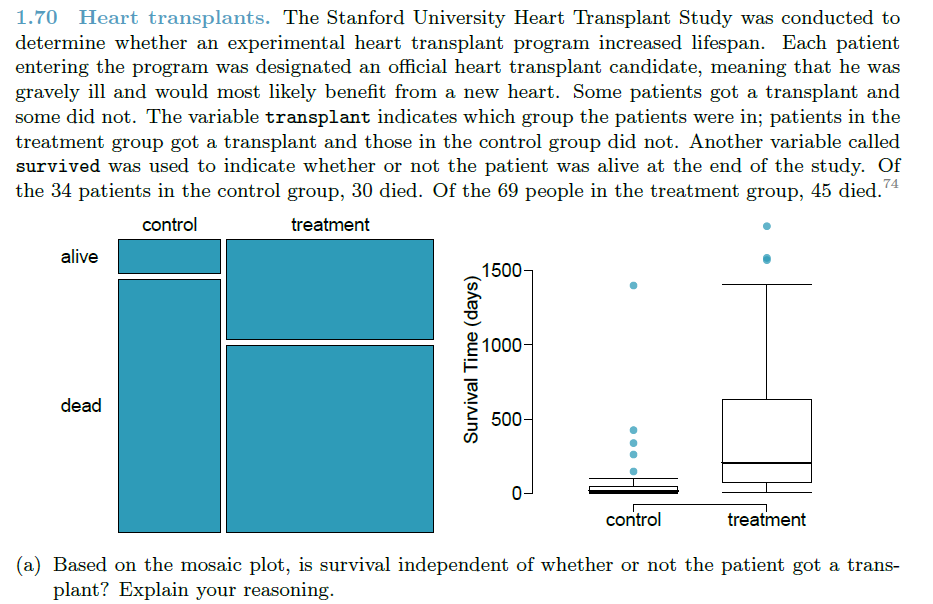
**Ans:**

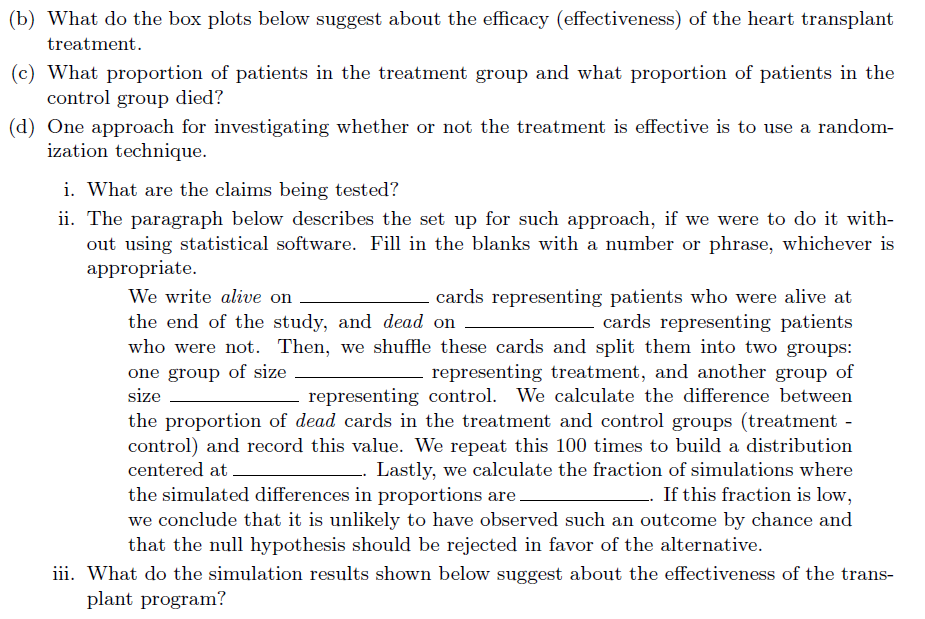
1. This looks like a normal unimodal distribution. It matches with (2)
2. This looks like a multimodal distribution. It matches with (3)
3. This looks like a bi-modal right skewed distribution and it matches with (1)

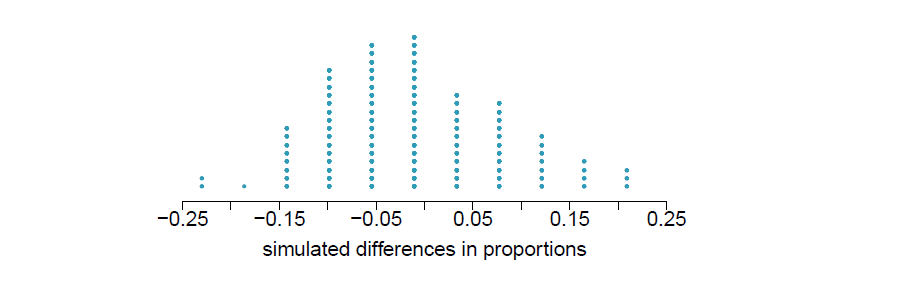


**Ans:**

1. This is a Left Skewed distribution. The Median would better represent an observation. The Variability would be better represented by the IQR as there are outliers.
2. This is again a Left Skewed distribution. The Median would better represent an observation. The Variability would be better represented by the IQR as there are outliers.
3. This is a right skewed distribution. The Median would better represent an observation. The Variability would be better represented by the IQR as there are outliers.
4. This would tend to be a normal distribution with a slightly right skew because of the high executive salaries. Since we have outliers, again here the Median and IQR would be better statistic.







**Ans:**

1. Since the treatment group has a better survival rate, It would mean that the transplant was a causal factor for survival.
2. Those patients who received the treatment have a better number of survival days as compared to the control group.
3. 88.23% of the control group and 65.22% of the treatment group died.
   1. The claim being tested is that the new heart transplant would lead to better survival days.
   2. We write alive on **28** cards representing patients who were alive at the end of the study, and dead on **75** cards representing patients who were not. Then, we shuffle these cards and split them into two groups: one group of size **69** representing treatment, and another group of size **34** representing control. We calculate the difference between the proportion of dead cards in the treatment and control groups (treatment - control) and record this value. We repeat this 100 times to build a distribution centered at **0**. Lastly, we calculate the fraction of simulations where the simulated differences in proportions are above 83.3%. If this fraction is low, we conclude that it is unlikely to have observed such an outcome by chance and that the null hypothesis should be rejected in favor of the alternative.