Hao-HW1

Bruce Hao

September 2, 2016

1.8

a) Each row represents an individual UK resident within the study.
b) 1691
c) Variable data types:

- sex: categorical
- age: numeric discrete
- marital: categorical
- grossIncome: categorical ordinal
- smoke: categorical
- amtWeekends: categorical (could be numeric but the 'cig/day' suggests text)
- amtWeekdays: categorical (could be numeric but the 'cig/day' suggests text)

1.10

a) Population: all children; Sample: 160 children between the ages of 5 and 15
b) If the children sampled are representative of all children, then the study can be generalized to the population. Since the study was an experiment, it potentially can be used to establish causal relationships.

1.28

a) No, this was an observational study; therefore, we cannot establish a causal relationship between smoking and dementia.
b) No, this was another observational study, and no causal relationship can be established. The best we can say is that there appears to be an association between behavior disorders/bullying and sleep disorders.

1.36

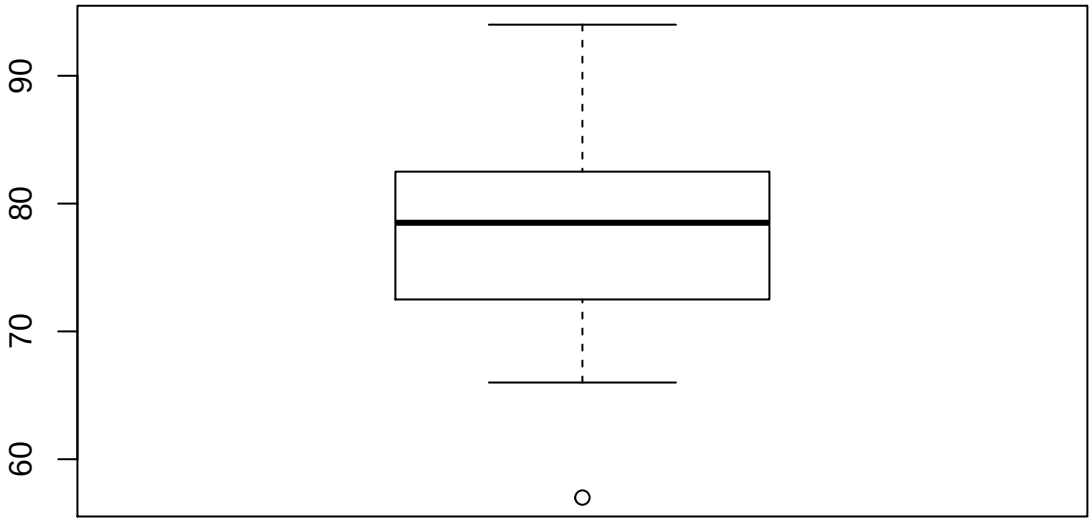
a) An experimental study based on stratified sampling.
b) Treatment: exercise twice a week; control: no exercise.
c) Yes, the blocking variables are the different age groups.
d) No, as both the experimenter and subjects are aware of the treatment and control groupings.
e) Yes, since this was an experiment, the results can potentially be used to establish a causal relationshiop between the treatment and response. Yes, the results can be generalized to the population at large (within the age ranges at least) because the samples were random.
f) Yes, I would have reservations in funding the study. There are likely many confounding factors that need to be considered, and additional details about the stud would need to be outlined, e.g. length of study, controls for other confounding factors, etc.

1.48

```r
v = c(57,66,69,71,72,73,74,77,78,78,79,79,81,81,82,83,83,88,89,94)
summary(v)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   57.00   72.75   78.50   77.70   82.25   94.00
```

```r
boxplot(v)
```



Note: I'm not sure why the 1st and 3rd quartiles are different than what's in the book.

1.50

a) is a symmetrical bell shape with a range from 50-70 and pairs with 2)
b) is symmetrical and more uniformly distributed across a range from 0-100 and pairs with 3)
c) is skewed to the right with a range from 0-6+ and pairs with 1)

1.56

No, the distribution is not symmetric. With a mean of 85, maximum of 100 and sd of 15, the distribution must be skewed to the left.

1.70

a) Optically, it appears that survial is not independent of whether or not the patient got a transplant; however, it is impossible to tell by 'eyeballing' if this is statistically significant or not.

b) The box plots paint a much clearer picture of the efficacy of the treatment. Survival time in the control group is tightly clustered near zero; wherease survival time for the treatent group is much higher with a median higher than the upper whisker of the control group and much higher upper quartile and whisker as well.

c) 88% of the control group died by the end of the study; 65% of the treatment group died by the end of the study.

d)

- i) H0: The claim being tested is that a heart transplant and increased lifespan are independent.

- ii) alive on 28 cards; dead on 75 cards; one group size of 69 representing treatment; one group size of 34 representing control; distribution centered at 0; lastly, we calculate the fraction of simulations where the difference in proportion are less than -23%.
- iii) Only 2% of the simulations had differences in proportion less than -23%, i.e. there is only a 2% chance that the results from the study were random.