solution-1.R

wickhamc

Wed Oct 15 16:50:59 2014

library(openintro)

# 1

## 4.35

1. = Plot B, (2) = Plot A, (3) = Plot C

A single sample from the population will look like the population and hence have a similar spread. The distribution of sample means from samples of size 5 will be narrower than the population, and the distribution of sample means from samples of size 25 will be **even** narrower. Since, Plot B has the largest spread it corresponds to (1). Plot A has the next largest spread and corresponds to (2), finally the narrowest plot is C so it corresponds to (3).

## 4.36

1. = Plot B, (2) = Plot C, (3) = Plot A

Same reasoning as above.

# 2

## a.

Sample 1: this is a very small sample (n = 4), there are three observations at about 1, 2 and 3 and a fourth outlying observation near 11.

Sample 2: this sample contains only positive values and is right skewed. The majority (~75%) of the data lies between 0 and 1, but there are a couple of values higher than 3.

Sample 3: this sample is unimodal (one peak) and appears roughly symmetric around a center of zero. Most values fall between about -2 and 2.

Sample 4: this sample is roughly symmetric around a center of zero but has two distinct peaks, one at -2 and one at 2. Most values fall between about -5 and 5.

## b.

### Dotplot

**Advantages:** concise (i.e. doesn't take up much room) display of small samples.
**Disadvantages:** large samples result in lots of overplotting and it is hard to determine the density of points (i.e. sample 3 and sample 4 look very similar in the dotplots)

### Boxplot

**Advantages:** concise (i.e. doesn't take up much room) display of samples, the simplicity of the display handles large samples well, easy to see symmetry versus skew.
**Disadvantages:** completely obscures interesting features in the center of data (i.e. sample 4, can't see bimodality.), can also obscure the fact that a sample is very small.

### Histogram

**Advantages:** pretty complete picture of the data, no problems dealing with very large sample sizes, easy to evaluate symmetry, look for outliers etc.
**Disadvantages:** can take up a lot of space, which means comparing many samples can be hard, sometimes the jaggedness can be distracting, the choice of binwidth can obscure or mislead.

In general the best plot will depend on the sample size, distribution shape and purpose of plot.

*Charlotte's note: Often I start with a histogram, if the distributions are well behaved I might decide to switch to a boxplot (only if I'm sure they are not obscuring something interesting). Dotplots I reserve for very small samples.*

# 3

library(openintro)
gifted$fatheriq - gifted$motheriq

## [1] -2 4 -3 -18 1 4 -1 -3 -17 2 -6 -8 -7 6 -15 -7 -4
## [18] 4 15 6 -10 -7 -3 -9 -9 -1 -12 -5 -7 -6 -5 4 10 6
## [35] -10 -9

### A. Using R, construct a 95% confidence interval for the mean difference in IQ between the mother's and father's IQ score.

diffs <- gifted$fatheriq - gifted$motheriq
xbar <- mean(diffs)
sd <- sd(diffs)
n <- length(diffs)

se <- sd/sqrt(n)
df <- n-1

xbar + qt(0.975, df)\*se

## [1] -0.8669

xbar - qt(0.975, df)\*se

## [1] -5.911

### B. Write a one sentence summary of the interval in the context of the data.

For gifted children in this city, with 95% confidence, the mother's IQ is between 0.87 and 5.91 points higher on average than the father's IQ.