Homework 5 solutions

S320/520

Due at the beginning of class, Thursday 1st October

Please write "S320" or "S520" at the top of your homework. All students should answer all questions. Trosset question numbers refer to the hardcover textbook. Show all working.

1. Trosset exercise 6.4.1

   For $1 < y < 2$, the cdf $F(y)$ is the area of a triangle with base $y-1$ and height $f(y) = 2(y-1)$:

$$F(y) = \begin{cases} 0 & y < 1 \\ (y-1)^2 & 1 \le y < 2 \\ 1 & y \ge 2. \end{cases}$$

   To find the α-quantile, set $F(q_\alpha) = \alpha$:

$$\begin{aligned} (q_\alpha - 1)^2 &= \alpha \\ q_\alpha - 1 &= \sqrt{\alpha} \\ q_\alpha &= 1 + \sqrt{\alpha}. \end{aligned}$$

   To find the median, substitute $\alpha = 0.5$, giving $q_2 = 1 + \sqrt{0.5} \approx 1.71$.

   To find the IQR, substitute $\alpha = 0.25$ and 0.75 to find the quartiles, then take the difference:

$$IQR = (1 + \sqrt{0.75}) - (1 + \sqrt{0.25}) = \sqrt{0.75} - 0.5 \approx 0.37.$$

2. Trosset exercise 6.4.2

   (a) By geometry, the area under $g(x) = 2$. We require the area under $f(x)$ to be 1, so $c = 1/2$.

   (b) The probability is the area between $x = 1.5$ and $x = 2.5$. By symmetry, the area between $x = 1.5$ and $x = 3$ is $1/2$.

   The area between $x = 2.5$ and $x = 3$ is a triangle with base $1/2$ and height $f(2.5) = 1/4$, giving area $1/16$.

   The area between $x = 1.5$ and $x = 2.5$ is thus $1/2 - 1/6 = 7/16$. This is the probability we want.

   (c) The pdf is symmetric about $x = 1.5$, so $EX = 1.5$.

   (d) This is the area between $x = 0$ and $x = 1$. This is a triangle with base 1 and height $f(1) = 1/2$, giving area 1.4.

(e) We need to find q such that the area to the left of the vertical line at q is 0.9, and the area to the right is 0.1. By inspection of the pdf, this will happen for some x-value between 2 and 3. The area to the right is a triangle with base $3 - q$ and height $f(q) = (3 - q)/2$. Setting this area equal to 0.1 gives:

$$\begin{aligned} 0.1 &= \frac{1}{4}(3-q)^2 \\ 0.4 &= (3-q)^2 \\ \sqrt{0.4} &= 3 - q \\ q &= 3 - \sqrt{0.4} \approx 2.37. \end{aligned}$$

3. Trosset exercise 6.4.6

(a) X is uniformly distributed in the range $(5, 15)$, so its quartiles are 7.5, 10, and 12.5. The ratio of interquartile range to standard deviation is

$$\frac{12.5 - 7.5}{\sqrt{25/3}} = \frac{5}{5/\sqrt{3}} = \sqrt{3} \approx 1.73.$$

(b) Using R,

```
> sd = 5/sqrt(3)
> iqr = qnorm(.75, 10, sd) - qnorm(.25, 10, sd)
> iqr/sd
[1] 1.34898
```

(See also section 6.3 of Trosset.)

4. Trosset exercise 7.7.1 parts (a)–(e)

Here's some R code for the question.

```
x = scan("http://mypage.iu.edu/~mtrosset/StatInfeR/Data/sample771.dat")
plot.ecdf(x)
mean(x)
# Variance
mean(x^2) - mean(x)^2
median(x)
quantile(x, c(0.25, 0.75))
433 / sqrt(mean(x^2) - mean(x)^2)
boxplot(x, main="Boxplot for Trosset exercise 7.7.1", ylab="x")
```

(a) See Figure 1.

(b) Mean is 494.6, plug-in variance is 91079, or 94974 if you use `var(x)` (which technically is not a plug-in estimate.)

(c) Median is 462, IQR is $658 - 225 = 433$

(d) The ratio is 1.43 (or 1.41 if you use the `sd` function.)

(e) See Figure 2.

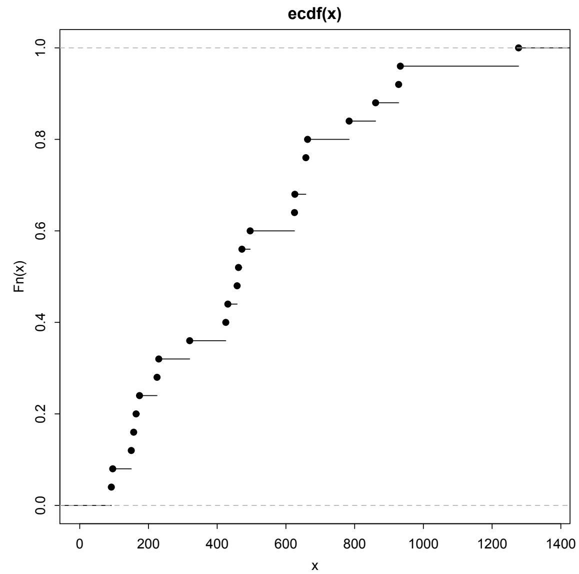

Figure 1: ECDF plot for Trosset exercise 7.7.1.

Figure 2: Boxplot for Trosset exercise 7.7.1.

5. Trosset exercise 7.7.3

```
data = read.table("http://mypage.iu.edu/~mtrosset/StatInfeR/Data/sample773.dat")
Sample1 = data$V1
Sample2 = data$V2
Sample3 = data$V3
Sample4 = data$V4
boxplot(Sample1, Sample2, Sample3, Sample4,
  main = "Boxplot for data from Trosset exercise 7.7.3")
# Compare to random normal data
boxplot(rnorm(10), rnorm(10), rnorm(10), rnorm(10),
  main = "Boxplot for random normal samples")
# Repeat this a few times if you wish.
```

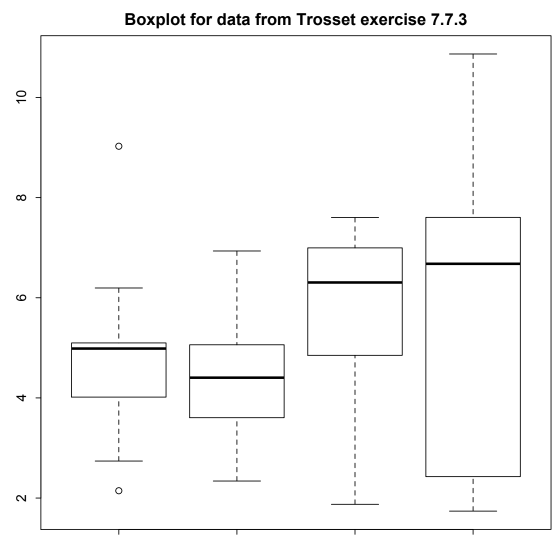

Figure 3: Boxplots for four samples from four populations (Trosset exercise 7.7.3.)

On one hand, boxplot of the four samples of data provided by Trosset look very different from each other, and a couple look highly skewed. On the other hand, boxplots of four independent standard normal samples of the same size ($n = 10$) also look very different from each other, and a couple look highly skewed. It's possible that Samples 1–4 were all drawn from the same normal distribution, and it's possible they were not. With small sample sizes, boxplots are highly variable.