Tutorial 2: Solutions

January 24, 2018

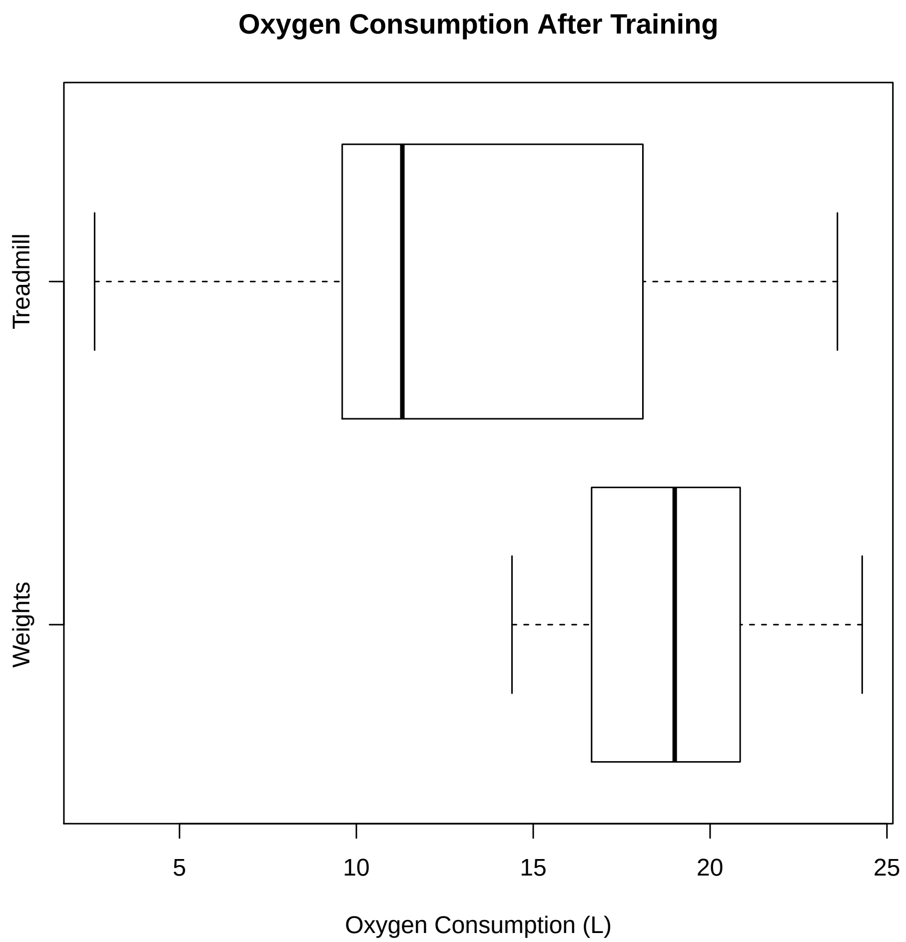
Question 1.S.70, Page 48-49

(a) Construct a comparative boxplot of the weight and treadmill observations, and comment on what you see.

- The median of the x values is 19.0
- Q_1 is the average of 16.3 and 17.0 which gives 16.65
- Q_3 is the average of 19.6 and 22.1 which gives 20.85
- $\text{IQR}_x = Q_3 - Q_1 = 20.85 - 16.65 = 4.2$
- $Q_1 - 1.5\text{IQR} = 10.35$ (don't need to calculate $Q_1 - 3\text{IQR}$ since nothing is below 10.35)
- $Q_3 + 1.5\text{IQR} = 27.15$ (don't need to calculate $Q_3 + 3\text{IQR}$ since nothing is above 27.15)
- No mild outliers here (and no extreme outliers)
- Mean is 19.01. St. dev is 3.09.

- The median of the y values is 11.3
- Q_1 is the average of 9.1 and 10.1 which gives 9.6
- Q_3 is the average of 16.6 and 19.6 which gives 18.1
- $\text{IQR}_y = Q_3 - Q_1 = 18.1 - 9.6 = 8.5$
- $Q_1 - 1.5\text{IQR} = -3.15$ (Can stop here)
- $Q_3 + 1.5\text{IQR} = 30.85$ (Can stop here)
- No mild outliers here (and no extreme outliers)
- Mean is 12.95. St. dev is 6.58.

Here is the comparative boxplot:



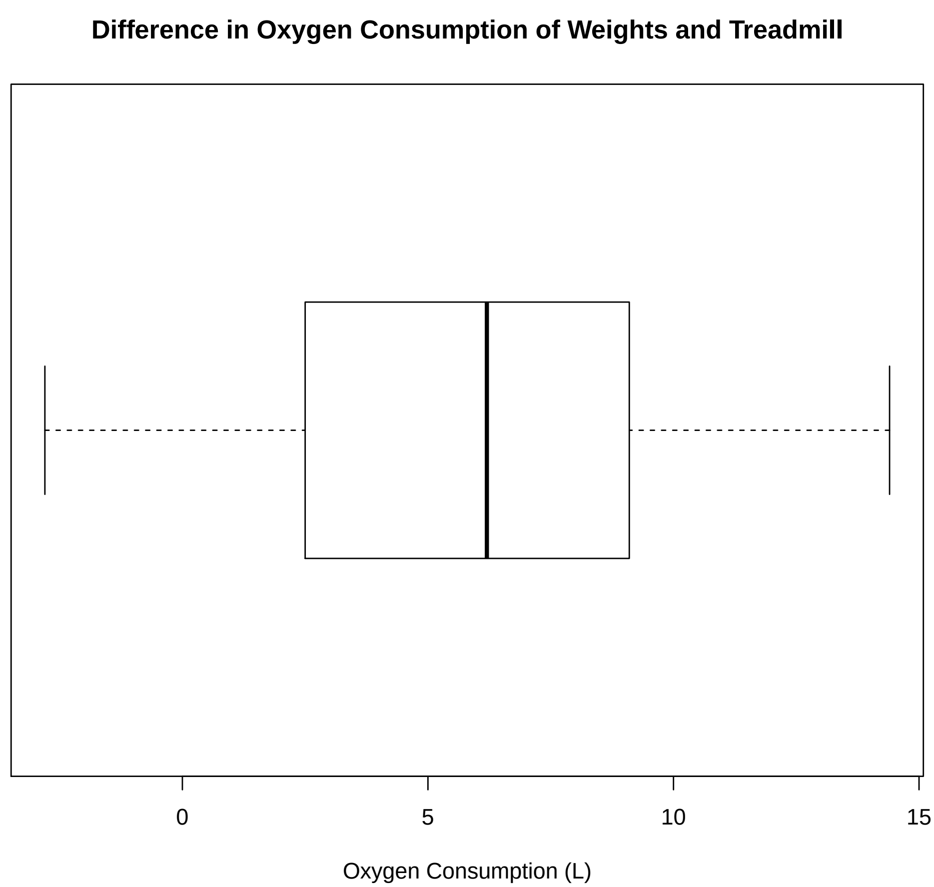
Comments:

- All weight values are higher than the median for treadmill values. In other words, oxygen consumption after weight training is greater than at least 50% of the oxygen consumption values after treadmill training.
- We can see that the range for weight values is much smaller than for treadmill values. The values for weight are much more concentrated about the mean compared to treadmill values (weight has a smaller standard deviation).
- Median and mean for weights are very close, so we expect its distribution to be approximately symmetrical
- Median for treadmill is smaller than the mean, so we expect the distribution to be right-skewed (positively skewed)

(b) Construct a boxplot of the sample differences. What does it suggest?

- The median is 6.2
- Q_1 is the average of 2.5 and 2.5 which gives 2.5
- Q_3 is the average of 9.1 and 9.1 which gives 9.1
- $\text{IQR} = Q_3 - Q_1 = 9.1 - 2.5 = 6.6$
- $Q_1 - 1.5\text{IQR} = -7.4$
- $Q_3 + 1.5\text{IQR} = 19$
- All our values are contained within -7.4 and 19 so there are no outliers (mild and extreme).
- The mean is 6.067. St. dev is 4.76.

Here is the boxplot:



Comments:

- Recall that we are the looking at the change in oxygen consumption from weights to treadmill. In other words, a positive difference means a higher oxygen consumption after weight training than treadmill training. A negative difference means a higher oxygen consumption after treadmill training than weights training.
- Looking at Q_1 to Q_3 we can see that 50% of our participants experienced a decrease between 2.5 L and 9.1 L in oxygen consumption from weights to treadmill exercises. For the majority of our observed values (13/15 observations), we observe a decrease in oxygen consumption moving from weights to treadmill exercises. There are two observations where the opposite has occurred.
- It is very plausible from this data that weight training requires a higher oxygen consumption compared to treadmill training due the majority of the data values being positive.
- The standard deviation (and the range) are quite large so there does not seem to be much concentration about the mean.
- The median is quite close to the mean so we expect that the data is somewhat symmetric about the mean.

Question 5.3.38, Page 229

(a) Determine the pmf of $T = X_1 + X_2$.

Let $T = X_1 + X_2$. Below is a table showing the probabilities of each pair of values.

		x_2			
		0	1	2	Totals
	0	0.04	0.10	0.06	0.2
x_1	1	0.10	0.25	0.15	0.5
	2	0.06	0.15	0.09	0.3
Totals		0.2	0.5	0.3	1

From the table we can see that the probability mass function will be:

$$f(t) = \begin{cases} 0.04 & t = 0 \\ 0.20 & t = 1 \\ 0.37 & t = 2 \\ 0.30 & t = 3 \\ 0.09 & t = 4 \\ 0 & \text{otherwise} \end{cases}$$

(b) Calculate μ_T . How does it relate to μ, the population mean?

Classical Method:

$$\begin{aligned}
\mathrm{E}(T) &= \sum_{t \in \mathcal{S}} t \cdot \mathrm{P}(T = t) \\
&= (0 \cdot 0.04) + (1 \cdot 0.20) + (2 \cdot 0.37) + (3 \cdot 0.30) + (4 \cdot 0.09) \\
&= 2.2 \\
&= 2 \cdot \mu
\end{aligned}$$

Using the properties discussed in Ch. 5.5:

$$\begin{aligned}
\mathrm{E}(T) &= \mathrm{E}(X_1 + X_2) \\
&= \mathrm{E}(X_1) + \mathrm{E}(X_2) \\
&= 1.1 + 1.1 \\
&= 2.2 \\
&= 2 \cdot \mu
\end{aligned}$$

(c) Calculate σ_T^2. How does it relate to σ^2, the population variance?

Using the classical method:

$$\begin{aligned}
\mathrm{Var}(T) &= \sum_{t \in \mathcal{S}} (t - \mu_T)^2 \cdot \mathrm{P}(T = t) \\
&= ((0 - 2.2)^2 \cdot 0.04) + ((1 - 2.2)^2 \cdot 0.20) + ((2 - 2.2)^2 \cdot 0.37) \\
&+ ((3 - 2.2)^2 \cdot 0.30) + ((4 - 2.2)^2 \cdot 0.09) \\
&= 0.98 \\
&= 2 \cdot \sigma^2
\end{aligned}$$

Using the properties discussed in Ch. 5.5

$$\begin{aligned}
\mathrm{Var}(T) &= \mathrm{Var}(X_1 + X_2) \\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) \\
&= 0.49 + 0.49 \\
&= 0.98 \\
&= 2 \cdot \sigma^2
\end{aligned}$$

(d) Let X_3 and X_4 be the number of lights at which a stop is required when driving to and from work on a second day assumed independent of the first day. With $T = X_1 + \ldots + X_4$, what now are the values of $\mathrm{E}(T)$ and $\mathrm{Var}(T)$?

$\mathrm{E}(T) = \mathrm{E}(X_1 + X_2 + X_3 + X_4) = \sum_{i=1}^{4} \mathrm{E}(X_i) = 4(1.1) = 4.4$ (Expectation is a linear operator)

$\mathrm{Var}(T) = \mathrm{Var}(X_1 + X_2 + X_3 + X_4) = \sum_{i=1}^{4} \mathrm{Var}(X_i) = 4(0.49) = 1.96$ (The X_is are independent here)

(e) Referring back to (d), what are the values of $\mathrm{P}(T = 8)$ and $\mathrm{P}(T \geq 7)$ [Hint: Don't even think of listing all possible outcomes!]

$\mathrm{P}(T = 8) = \mathrm{P}(X_1 = 2,\ X_2 = 2,\ X_3 = 2,\ X_4 = 2) = (0.3)^4 = 0.0081$
We notice that since our random variables take on values from 0 to 2, the only way to get a sum of 7 with four of such random variables is if we have a three 2s and one 1. There are a total of four ways of getting this:

$$\begin{aligned} & \mathrm{P}\left(\vec{x} = (1,2,2,2) \text{ or } \vec{x} = (2,1,2,2) \text{ or } \vec{x} = (2,2,1,2) \text{ or } \vec{x} = (2,2,2,1)\right) \\ = \; & 4\left((0.3)^3(0.5)\right) \\ = \; & 0.054 \end{aligned}$$

$$\begin{aligned} \mathrm{P}(T \geq 7) &= \mathrm{P}(T = 7) + \mathrm{P}(T = 8) \\ &= 0.054 + 0.0081 \\ &= 0.0621 \end{aligned}$$