

Instructions

Upload your own solutions (handwritten or typed) to the corresponding Canvas Assignment on or before the due date: 26/11/2021.

Grading scheme

a	b	c	d	e	f	g	h	Total
2	4	4	4	2	1	2	1	20

Part 1

a. $\bar{x} = 59\,400$, $s = 34\,000$.

b. $m = 54\,000$, which is smaller than the mean. This suggests skewness to the right.

c. $Q_1 = 33\,000$, $Q_3 = 77\,000$ (these are the 8th and 23rd order statistics respectively), thus the length of the IQR is $Q_3 - Q_1 = 44\,000$.

d. The $1.5 \times \text{IQR}$ interval is $(-33\,000, 143\,000)$, and the $3 \times \text{IQR}$ interval is $(-99\,000, 209\,000)$. There is just one outlier: $x_{20} = 179\,000$ and no extreme observations.

e. $(-42\,600, 161\,400)$

f. There is just one outlier, 3.5 standard deviations larger than the mean. Under normality, this wouldn't be common for such a small sample. But given what we know, there is absolutely no reason to believe that a modest 179 000 was due to some "error". In fact, the outlier is reminding us that the number of views of a YouTube video could get quite big. It's not a mistake but a warning sign!

g. Here we continue to measure views by the thousand (eg. 600 means 600 000 views).

$$\mathbf{P}(Y \geq 1500) = \mathbf{P}\left(\frac{Y - 600}{\sqrt{9000}} \geq \frac{1500 - 600}{\sqrt{9000}}\right) = \mathbf{P}(Z \geq 9.48), \quad \text{where } Z \sim N(0, 1).$$

Observing a value of 9.48 for a standard normal random variable is incredible rare! Using an online high-precision normal distribution calculator or statistical software we get

$$P(Z \geq 9.48) \approx 10^{-21}.$$

h. This is a personal question, so everyone will have their own reasons.

In reality, the true distribution of 'number of views' is far from normal. The startup was unlucky to get a misleading sample (misleading because it looks somewhat normal – usually it is even more skewed to the right!), but that happens sometimes. We know that videos can go viral.

Maybe we don't know the probability of a video going viral, but at least we can foresee that if we pay 1€ for every thousand views and one of the videos goes viral we are going to be in trouble. Our suggested analysis would be this: Based on the outlier and skewness of the data, plus what we know about YouTube videos, it is reasonable to doubt the normality assumption. Without a model for the number of views, it is not possible to compute exact probabilities, so the computations in part **g.** are useless. Moreover, it is not unreasonable to expect that every now and then a new video will get a lot of views, perhaps even go viral. Therefore, no matter how tempting Offer 2 is, it is better to chose the safer Offer 1.