Test #1 Answers

STAT 880

Spring 2021

This is an “open book, open note” test. You may use the course textbooks, course notes, your own notes, R programs, Sage code/output in Jupyter Notebook documents, data from the course website, any R programs and Sage code/output in Jupyter Notebook documents that you created prior to the test, projects, and help within R or Sage. You cannot communicate with your classmates or anyone outside of our class to complete the test. Also, you cannot use the Internet other than to download the test, turn in the test, or to communicate with me. E-mail to me ([bilder@unl.edu](mailto:bilder@unl.edu)) your completed test prior to 2PM. You will receive an e-mail response from me to confirm receipt. Do not move on from taking the test until you receive this confirmation.

Complete the problems below. Make sure to fully explain all answers and show your work to receive full credit. Use R or Sage for all calculations and plots.

1. (21 total points) Which brand of paper towels is more absorbent? Some graduate students a few years ago decided to investigate this on their own by performing an experiment. Below is how the experiment was performed:
   * 1. A large glass is filled with 32 ounces of water.
     2. A paper towel is submerged into the water for 20 seconds.
     3. After the paper towel is removed, the amount of water remaining in the glass is recorded.
     4. The difference between the beginning amount of water (32 ounces) and the ending amount of water is found, and this represents the amount of water absorbed by the paper towel.

A number of paper towels from two brands (“A” and “B”) were used in the experiment. The data from the experiment is available in the set1.csv file located where you downloaded this test. Below is an example of how to read in the data.

> # Set the appropriate file location on your computer

> paper.towel <- read.csv(file = "set1.csv")

> head(paper.towel)

type response

1 A 3.665608

2 A 3.313991

3 A 3.936691

4 A 3.997216

5 A 4.508679

6 A 4.734902

* 1. (8 points) Construct a box and dot plot of the data so that one can compare the two paper towel brands. Make sure the dot plot overlays the box plot and use a yellow color inside the boxes.

> set.seed(8912)

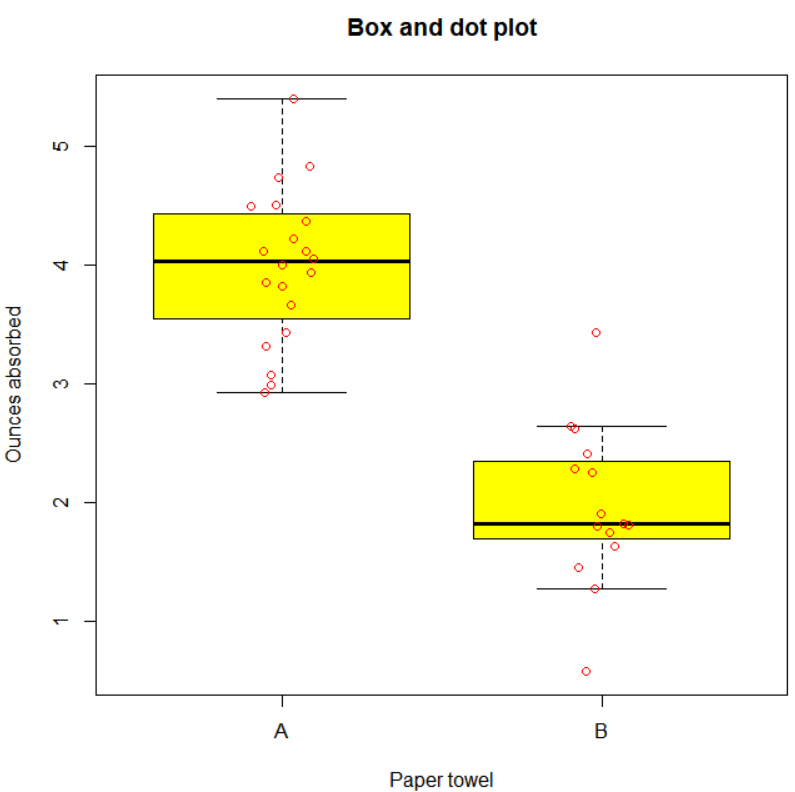
> boxplot(formula = response ~ type, data = paper.towel, main = "Box and dot plot",

ylab = "Ounces absorbed", xlab = "Paper towel", pars = list(outpch=NA), col =

"yellow")

> stripchart(x = paper.towel$response ~ paper.towel$type, lwd = 1, col = "red",

method = "jitter", vertical = TRUE, pch = 1, add = TRUE)



* 1. (6 points) Were there any “unusual” observed absorbency values? Explain.

Yes, brand B has two outliers that are outside of the whiskers for the box plot.

* 1. (7 points) Is there any preliminary evidence that one paper towel brand absorbs more water than the other? Explain.

The absorbency values for brand A are all almost above those for brand B. This shift of the distribution is indicative that there is preliminary evidence that A is more absorbent than B.

1. (39 total points) The waiting time, in hours, between successive speeders spotted by a police radar unit is a continuous random variable X with CDF of

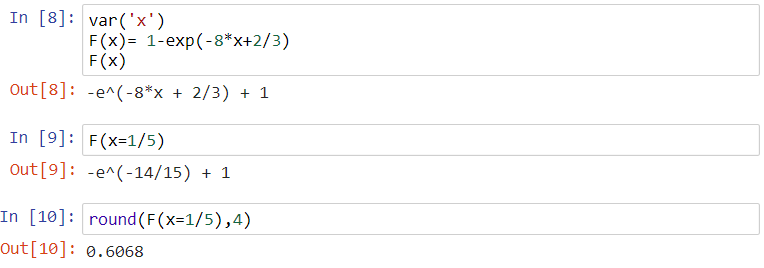


Note that x = 1/5 is the equivalent to 12 minutes. Answer the questions below and be VERY careful about the units that X is measured in.

* 1. (6 points) Find the probability of waiting less than 12 minutes (12/60 = 1/5 hours) before successive speeders using the CDF.

P(X < 1/5) = 

Sage:



R:

> 1 - exp(-8 \* 1/5 + 2/3)

[1] 0.6067593

* 1. (7 points) Find the PDF.

 and f(x) = 0 otherwise



* 1. (8 points) Plot the PDF and show how the probability of waiting less than 12 minutes before successive speeders is represented.

> curve(expr = 8 \* exp(-8\*x + 2/3), xlim = c(1/12, 1), col = "red", ylab = "f(x)")

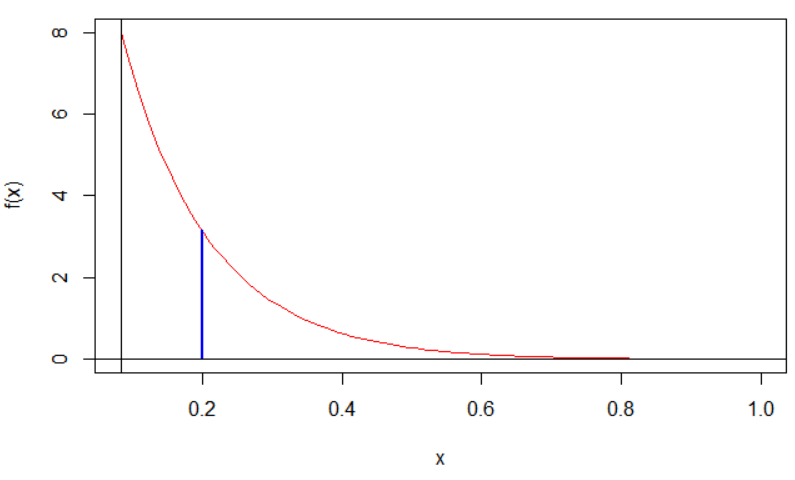
> segments(x0 = 1/5, x1 = 1/5, y0 = 0, y1 = 8 \* exp(-8\*1/5 + 2/3), col = "blue",

lwd = 2)

> abline(h = 0)

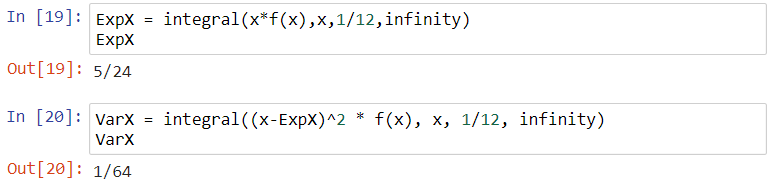
> abline(v = 1/12)

One way to show this on a plot:



* 1. (10 points) Find the standard deviation of the waiting time and interpret it in the context of the problem.

 and 

Sage:

R:

> pdf.x <- function(x) {

x \* 8 \* exp(-8\*x + 2/3)

}

> # E(X)

> mu <- integrate(f = pdf.x, lower = 1/12, upper = Inf)

> mu

0.2083333 with absolute error < 1.9e-07

> # E[(X-mu)^2]

> pdf.var <- function(x, mu) {

(x-mu)^2 \* 8 \* exp(-8\*x + 2/3)

}

> integrate(f = pdf.var, lower = 1/12, upper = Inf, mu = mu$value)

0.015625 with absolute error < 9.4e-07

The standard deviation is 1/8 of an hour. Using the rule of thumb for the number of standard deviations all data lies from mean,

μ ± 2σ = (5/24-2×1/8, 5/24+2×1/8) = (-1/24, 11/24) or

μ ± 3σ = (5/24-3×1/8, 5/24+3×1/8) = (-4/24, 14/24) = (-1/6, 7/12)

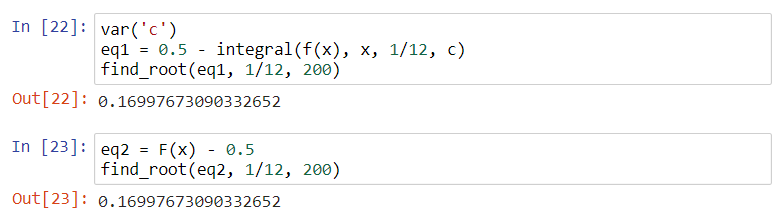
I would expect all waiting times to be somewhere between 0 and 11/24 of an hour using two standard deviations.

* 1. (8 points) Find the waiting time such that 50% are less than this value (i.e., find the median). Interpret this value in the context of the problem.

Find c in . Note that e-8c+2/3 = 0.5. Then c = -1/8 × log(0.5) + 2/24 = 1/8 × log(2) + 1/12 = 0.1700 of an hour.

Thus, the median waiting time between speeders is 0.1700 of an hour (about 10.2 minutes). This means that waiting times are 0.17 hours or less 50% of the time.

Sage:



R:

> find.root <- function(c.val) {

1 - exp(-8 \* c.val + 2/3) - 0.5

}

> uniroot(f = find.root, interval = c(1/12,200))

$root

[1] 0.1699776

$f.root

[1] 3.621971e-06

$iter

[1] 15

$init.it

[1] NA

$estim.prec

[1] 6.103516e-05

1. (20 total points) Two electronic components of a system work together for the success of the total system. Let X and Y be random variables for the life in hours of the two components. The joint PDF of X and Y is



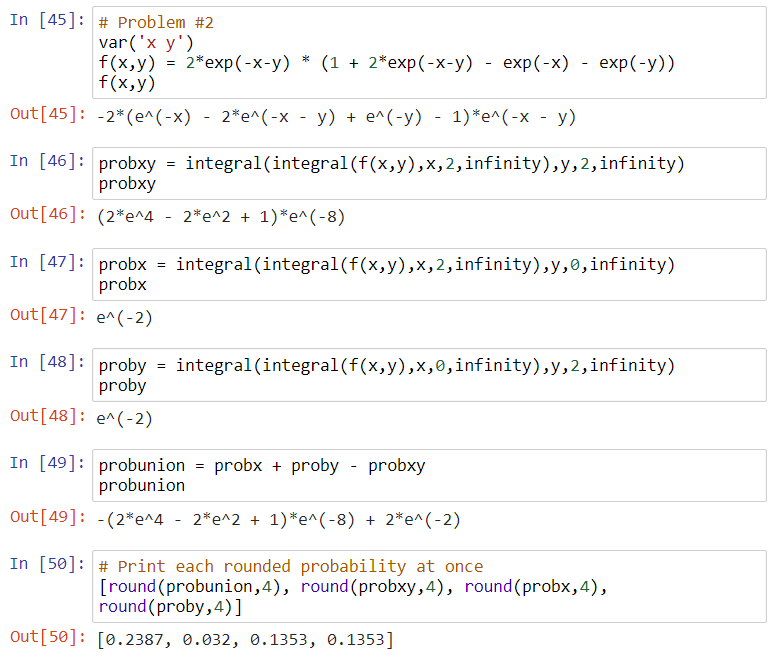
* 1. (12 points) What is the probability that at least one of the components will exceed 2 hours of life?







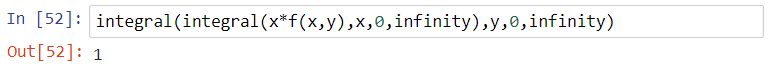




* 1. (8 points) Find average number of hours for the X component. Interpret it in the context of the problem



One expects the component to last for 1 hour on average.



1. (20 total points) Answer the questions below.
   1. (7 points) Explain why P(A|B) = P(A∩B)/P(B) for two events A and B. You may use a two row and two column contingency table to help with your explanation if needed.

From the Section 2 notes:

* Suppose the event B occurs and it had a particular probability (P(B)) of occurring. This now limits the possibility of what other events occur.
* To determine the probability that A occurs, we must examine P(A∩B) because B occurs.
* To find the probability that A occurs given the B occurred, we use P(A∩B)/P(B). This gives us the probability of A occurring out of all possibilities where B occurred.
  1. (7 points) Suppose F(x) is a CDF of a random variable X. What are F(-∞) and F(∞)? Why?

F(-∞) = 0 and F(∞) = 1

F(x) cumulates probabilities so F(∞) is the accumulation of all probabilities associated with all possible values of X. Thus, F(∞) = 1. Also, F(-∞) is the accumulation of no probabilities.

* 1. (6 points) Suppose a medical professional informs you that you have tested positive for an infectious disease. Why would knowing the positive predictive value for the test be important for you know?

Because infectious disease tests are not 100% accurate, you would want to know the probability that you are truly positive.