STAT 515 hw 5

Sampling distribution of mean from Normal population, Poisson, exponential, CLT

Attach a sheet with the R plots and R code printed on it. You may write out your other answers by hand if you want. Just try to make it easy for me grade!!

1. Your company sells bags of a boutique nutritious grain labelled as weighing 500 grams. The bags are filled such that their weights have the Normal distribution with mean $\mu = 510$ grams and standard deviation $\sigma = 20$ grams. A regulator over the commerce of boutique nutritious grains will visit your bagging facility, take a sample of $n = 5$ bags, and fine you if the average of the weights of the sampled bags is less than 500 grams.
   (a) With what probability will the regulator fine you?
   (b) You decide to alter your bagging process by increasing μ until the probability that you are fined is no more than 0.01. To what value should you increase μ?
   (c) You decide to keep $\mu = 510$ and to try convincing the regulator to sample a larger number of bags and fine you if the average of the bag weights is less than 500 grams. What sample size should you ask him to take if you wish him to fine you with probability no greater than 0.01?
   (d) The regulator is too busy to weigh more than $n = 5$ bags, and you wish to keep $\mu = 510$. Your chief engineer of boutique grain bagging suggests a way to reduce the standard deviation of the bag weights. To what value must you reduce σ in order that the regulator fines you with probability no greater than 0.01?

2. Some scientists will measure the speed of light many times with a method resulting in measurements having an expected value of 299,774 kilometers per second (km/s) and standard deviation 14 km/s. The true speed of light is 299,792 km/s (you will notice I have rounded if you fact-check this).
   (a) Assume the measurements are Normally distributed.
      i. What proportion of their measurements will exceed 299,792 km/s?
      ii. Give the value such that 95% of their measurements will lie above it.
      iii. Give two values such that 80% of the measurements will lie between them.
      iv. If they take 50 measurements, give the probability that 10 or more exceed 299,792 km/s.
   (b) Assume the measurements are not Normally distributed. The scientists take 50 measurements of the speed of light and record the average as $\bar{X}_n$.
      i. With approximately what probability will $\bar{X}_n$ exceed the true speed of light?
      ii. With approximately what probability does $\bar{X}_n$ lie within 20 km/s of the true speed of light?
      iii. Give an interval within which $\bar{X}_n$ will lie with probability approximately 95%.
      iv. Give the value such that $\bar{X}_n$ exceeds it with probability approximately 0.01.
      v. What theorem did you invoke to justify your calculations?

3. A batch of 112 zinc oxide crystals will be grown in an autoclave under conditions such that each full-grown crystal will be acceptable (have a low enough density of defects) with probability 0.80, and the outcome for each crystal will be independent of that of the other crystals.

(a) What is the expected number of acceptable crystals from the batch?

(b) What is the standard deviation of the number of acceptable crystals from the batch?

Give the exact probability that at least three-quarters of the crystals are acceptable.

(c) Give an approximation based on the Normal distribution to the probability that at least three-quarters of the crystals are acceptable.

(d) Use the Normal distribution to find an interval within which the number of acceptable crystals will lie with probability approximately 90%.

(e) Compute the exact probability that the number of acceptable crystals falls within the interval from your answer to part (d).

4. Pieces of litter are strewn along a path according to a Poisson process such that every 100 yards you expect to encounter 2 pieces of litter.

(a) With what probability do you not encounter any pieces of litter in the first 100 yards?

(b) With what probability do you encounter four or more pieces of litter in the first 100 yards?

(c) If you follow the path for 3,000 yards, how many pieces of litter do you expect to encounter?

(d) You decide to collect each piece of litter you encounter in a trash bag which can hold 50 pieces of litter. If you follow the path for 3,000 yards, with what probability will you fill the bag?

(e) If you were to average the distances, in yards, from each piece of litter to the next, to what number would you expect the average to be close?

5. Do a simulation with R.

(a) Use the following commands to set `n` equal to 500 and to generate `n` values of an exponential random variable with mean equal to 1/2 and to make a histogram of the values:

```
n <- 500
x <- rexp(n=500,rate=2)
hist(x)
```

Turn in the histogram. (Here “rate” corresponds to our λ, and the mean is $1/\lambda = 1/2$.)

(b) Based on your histogram, is the distribution of the 500 values Normal, right-skewed, left-skewed, or heavy-tailed?

(c) The following code draws 200 samples of size 10 from the exponential distribution with mean 1/2. For each sample of size 10, the sample mean is computed and is stored in the vector `x.bar`. At the end, a histogram of the 200 sample means is generated.

Note that the code is annotated with *comments*. In R, you can put a `#` symbol and then write non-code after it. When R sees the `#`, it ignores everything after it on that line. You can type this code into R omitting the comments.

```
n <- 10 # set the size of each sample
x.bar <- numeric(200) # create an empty vector of length 200
for(i in 1:200){ # start a ‘‘loop’’ of commands which are to be executed 200 times

    x <- rexp(n=n,rate=2)   # take a sample of size n
    x.bar[i] <- mean(x)     # store sample mean in position i of the vector x.bar
```

```
}
hist(x.bar) # make a histogram of the sample means.
```

Turn in the histogram of sample means.

(d) Make a QQ plot of the sample means using the commands

```
qqnorm(scale(x.bar))
abline(0,1)
```

Turn in this plot (the `scale` command gives the vector `x.bar` a mean of zero and a variance of 1, so that we can compare its quantiles to those of the Normal$(0, 1)$ distribution. This makes the scales of the X and Y axes the same so that we can draw the reference line $y = x$, which is what the command `abline(0,1)` does.).

(e) Is the distribution of the 200 sample means with $n = 10$ Normal, right-skewed, left-skewed, or heavy-tailed?

(f) Repeat part (c) with $n = 100$. You just have to change the first line of the code to `n <- 100`. (turn in the histogram).

(g) Repeat part (d) with the $n = 100$ output (turn in the QQ plot).

(h) Is the distribution of the 200 sample means with $n = 100$ Normal, right-skewed, left-skewed, or heavy-tailed?

(i) To what phenomenon is owed the difference between your answers to parts (e) and (h)? (A theorem we learned in class).

Optional (do not turn in) problems for additional study from McClave, J.T. and Sincich T. (2017) *Statistics*, 13th Edition: 6.30, 6.38, 6.45, 6.50, 6.56, 6.61, 6.80, 6.88