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?

   We have

   $$\begin{aligned} P(\bar{X}_5 < 500) &= P(Z < (500 - 510)/(20/\sqrt{5})) \\ &= P(Z < -\sqrt{5}/2) \\ &= \texttt{pnorm(-sqrt(5)/2)} \\ &= 0.1317762. \end{aligned}$$

   (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?

   We have

   $$\begin{aligned} P(\bar{X}_5 < 500) \leq 0.01 &\iff P(Z < (500 - \mu)/(20/\sqrt{5})) \leq 0.01 \\ &\iff (500 - \mu)/(20/\sqrt{5}) \leq q^Z_{0.01} \quad \text{(draw a picture to see this)} \\ &\iff \mu \geq 500 - q^Z_{0.01} \cdot 20/\sqrt{5} = 520.8075, \end{aligned}$$

   where $q^Z_{0.01} = \texttt{qnorm(0.01)} = -2.326348$. So we need to have

   $$\mu \geq 520.8075.$$

   (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?

We have

$$\begin{aligned} P(\bar{X}_n < 500) \leq 0.01 &\iff P(Z < (500 - 510)/(20/\sqrt{n})) \leq 0.01 \\ &\iff P(Z < -\sqrt{n}/2) \leq 0.01 \\ &\iff -\sqrt{n}/2 \leq q^Z_{0.01} \quad \text{(draw a picture to see this)} \\ &\iff n \geq 4 \cdot (q^Z_{0.01})^2 = 21.64758. \end{aligned}$$

So the regulator would need to draw a sample of size $n \geq 22$.

(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?

We have

$$\begin{aligned} P(\bar{X}_5 < 500) \leq 0.01 &\iff P(Z < (500 - 510)/(\sigma/\sqrt{5})) \leq 0.01 \\ &\iff P(Z < -10\sqrt{5}/\sigma) \leq 0.01 \\ &\iff -10\sqrt{5}/\sigma \leq q^Z_{0.01} \\ &\iff \sigma \leq -\frac{10\sqrt{5}}{q^Z_{0.01}} = 9.611924. \end{aligned}$$

So we need $\sigma \leq 9.61$.

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?

      If X represents a single measurement, then $X \sim \text{Normal}(\mu = 299774, \sigma^2 = 14^2)$. So we have

$$\begin{aligned} P(X > 299792) &= P((X - 299774)/14 > (299792 - 299774)/14) \\ &= P(Z > 9/7) \\ &= \texttt{1 - pnorm(9/7)} \\ &= 0.0992714. \end{aligned}$$

      ii. Give the value such that 95% of their measurements will lie above it.

If $Z \sim \text{Normal}(0,1)$, then Z lies above $q^Z_{0.05}$ with probability 0.95, where $q^Z_{0.05} =$ `qnorm(0.05)` $= -1.645$. We obtain the corresponding quantile for X as

$$q_{0.05} = 14 \cdot q^Z_{0.05} + 299774 = 299751.$$

iii. Give two values such that 80% of the measurements will lie between them.

We have $P(q^Z_{0.10} \le Z \le q^Z_{0.90}) = 0.80$, where $-q^Z_{0.10} = q^Z_{0.90} = 1.282$. To obtain the corresponding quantiles for X as

$$299774 \pm 14 \cdot 1.282 = (299756.1, 299791.9).$$

iv. If they take 50 measurements, give the probability that 10 or more exceed 299,792 km/s.

If Y represents the number of measurements among the 50 that exceed 299792, then $Y \sim \text{Binomial}(50, 0.0992714)$, where the success probability comes from our earlier work. We have

$$P(Y \ge 10) = 1 - P(Y \le 9) = \texttt{1 - pbinom(9,50,0.0992714)} = 0.02344993.$$

(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?

We have

$$\begin{aligned} P(\bar{X}_n > 299792) &= P((\bar{X}_n - 299774)/(14/\sqrt{50}) > (299792 - 299774)/(14/\sqrt{50})) \\ &= P(Z > 9.09) \\ &\approx 0. \end{aligned}$$

ii. With approximately what probability does $\bar{X}_n$ lie within 20 km/s of the true speed of light?

We have

$$\begin{aligned} P(|\bar{X}_n - 299792| \le 20) &= P(299772 \le \bar{X}_n 299812) \\ &= P((299772 - 299792)/(14/\sqrt{50}) \le Z \le (299812 - 299792)/(14/\sqrt{50})) \\ &= P(-1.01 \le Z \le 19.1929) \\ &= \texttt{1 - pnorm(-1.01)} \\ &= 0.8437524. \end{aligned}$$

iii. Give an interval within which $\bar{X}_n$ will lie with probability approximately 95%.

If $Z \sim \text{Normal}(0,1)$ then $P(q^Z_{0.025} < Z < q^Z_{0.975}) = 0.95$, where $-q^Z_{0.025} = q^Z_{0.975} = 1.96$. We obtain the corresponding quantiles of $\bar{X}_n$ as

$$299774 \pm 1.96 \cdot 14/\sqrt{50} = (299770.1, 299777.9).$$

iv. Give the value such that $\bar{X}_n$ exceeds it with probability approximately 0.01.

We have $P(Z > q^Z_{0.99}) = 0.01$, where $q^Z_{0.99} = 2.326$. We obtain the corresponding quantile of $\bar{X}_n$ as

$$299774 + 2.326 \cdot 14/\sqrt{50} = 299778.6.$$

v. What theorem did you invoke to justify your calculations?

the central limit theorem.

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?

Letting Y be the number of acceptable crystals from the batch, we have $Y \sim \text{Binomial}(112, 0.80)$, so $\mathbb{E}Y = 112(0.80) = 89.6$.

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

We have $\text{Var}\, Y = 112(0.80)(1 - 0.80) = 17.92$, so the standard deviation is $\sqrt{17.92} =$ 4.233202.

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

We have

$$P(Y \geq 84) = 1 - P(Y \leq 83) = \texttt{1 - pbinom(83,112,0.80)} = 0.9221304.$$

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

We have

$$\begin{aligned} P(\hat{p} \geq 0.75) &= P((\hat{p} - 0.80)/\sqrt{0.80(1-0.80)/112} \geq (0.75 - 0.80)/\sqrt{0.80(1-0.80)/112}) \\ &\approx P(Z \geq -1.322876) \\ &= \texttt{1 - pnorm(-1.322876)} \\ &= 0.9070616. \end{aligned}$$

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

Beginning with $P(q^Z_{0.05} \leq Z \leq q^Z_{0.95}) = 0.90$, where $-q^Z_{0.05} = q^Z_{0.95} = 1.644854$, we have that $\hat{p}$ will fall between the values

$$0.80 \pm 1.645 \cdot \sqrt{0.80(1-0.80)/112} = (0.7378304, 0.8621696)$$

with probability approximately 0.95. This means that the number of acceptable crystals will lie in the interval $(82.64, 96.563)$, which we may round to

$$(83, 97)$$

with probabability approximately 0.95. It would also be okay to round the lower bound down and the upper bound up to get the interval $(82, 97)$.

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

We have

$$\begin{aligned} P(83 \leq Y \leq 97) &= P(Y \leq 97) - P(Y \leq 82) \\ &= \texttt{pbinom(97,112,0.80) - pbinom(82,112,0.80)} \\ &= 0.9233741. \end{aligned}$$

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?

Let X be the number of pieces of litter you encounter along 100 yards of the path. Then $X \sim$ Poisson(2) distribution. We have

$$P(X = 0) = \frac{e^{-2} \cdot 2^0}{0!} = e^{-2} = 0.1353353.$$

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

> This is given by
>
> $$P(X \geq 4) = 1 - P(X \leq 3) = 1 - \sum_{x=0}^{3} \frac{e^{-2} \cdot 2^x}{x!} = \texttt{1 - ppois(3,2)} = 0.1428765.$$

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

> Our unit of distance travelled along the path is 100 yards. If we travel 3,000 yards, we have travelled 30×100 yards, so that if Y is the number of pieces of litter we encounter in 3,000 yards, $Y \sim \text{Poisson}(60)$, where $60 = 30 \times 2$. We have $\mathbb{E}Y = 60$, so we expect (in the sense of the expected value) to encounter 60 pieces of litter.

(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?

> This is given by $P(Y \geq 50)$, where the random variable Y is defined in the answer to the previous part. We have
>
> $$P(Y \geq 50) = 1 - P(Y \leq 49) = \texttt{1 - ppois(49,60)} = 0.9155933.$$

(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?

> If the random variable W is the distance from one piece of litter to the next, then W has the Exponential distribution with mean $1/2$, since the number of pieces X of litter encountered per each unit of distance has the Poisson(2) distribution. To convert our answer to yards, we must multiply $1/2$ by 100 (since the unit of distance is 100 yards), which gives an expected distance of 50 yards between one piece of litter and the next.

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?

It should be right-skewed.

(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?

Should be right-skewed.

(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?

Should look pretty Normal.

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

$$\boxed{\text{The central limit theorem.}}$$

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