Mathematical Statistics II

Statistical Computing Activity: Module 1

One purpose of these Statistical Computing Activities (SCAs) is to give you a chance to explore statistics when the random variables do not follow a Normal distribution. Another purpose is to give you more skills in thinking about the randomness that is life.

Usually, like here, these SCAs will have a theme and several problems dealing with that theme or purpose. The reason for that extra layer of complexity is to tie what we do in the class with what we can use these techniques for in our lives as statisticians and/or consultants and/or full members of a democratic society.

⋆ ⋆ ⋆

What to submit. On the due date, please submit a typed paper (Word or LaTeX). Each problem starts on a new page. The code is either all at the end of the submission in an appendix or is at the end of each problem — your choice on this. I have provided sample solutions for the first three problems. You should follow those.

Problem 1: Simple

Let us model the volume of a US Quarter. First, let us start with the unrealistic assumption that both dimensions, height and radius, are Normally distributed. Specifically, in millimeters,

$$H \sim \mathcal{N}(\mu = \ \ 1.75;\ \sigma = 0.1)$$
$$R \sim \mathcal{N}(\mu = 12.13;\ \sigma = 0.2)$$

Estimate the following

a) What is the expected volume of a US Quarter?
b) What is a 95% prediction interval for that volume?
c) What proportion have a volume less than 750mm^3?
d) What proportion have diameter less than 24mm?

Solution: To estimate the distribution of the volume, we generate a large number of heights and radii, and apply the formula to those values.

```
n = 1000000

h = rnorm(n, m= 1.75, s=0.1)
r = rnorm(n, m=12.13, s=0.2)

v = pi*h*r^2
```

At the end of this code, the variable `v` will contain a million simulated US Quarter volumes. This incredibly large number ensures that this sample will be *extremely* similar to the population. This allows us to estimate population parameters from this sample.

Thus, answers to the above questions (with `R` code following) are

a) The expected volume of a US Quarter is 809.3mm^3.
   `mean(v)`
b) A 95% prediction interval for that volume is from 707 to 916 mm^3.
   `quantile(v, c(0.025,0.975))`
c) The proportion that have a volume less than 750mm^3 is 0.132.
   `mean(v<750)`
d) The proportion that have diameter less than 24mm is 0.257.
   `mean(r<12)`

Note that these are not mathematically correct. However, because of the randomness inherent in studies of probability, rounding will be necessary. Also, using a computer forces rounding to take place. Also, note that my random sample will not be the same as yours. It will be very close, however.

Problem 2: Truncating the Normals

The US Quarters go through a quality-control process. If any dimension is too small or large, the coin is rejected. There are many way of dealing with truncation, including something called a Truncated Normal distribution. Here, we will just do what the US Mint does: reject the coins are too big or too small. Estimate the following for the coins that meet specifications

a) What is the expected volume of a US Quarter?

b) What is a 95% prediction interval for that volume?

c) What proportion have a volume less than 750mm^3?

Solution: We first need to generate the volume of coins that meet the specifications. To do this, let's start with most of the previous code

```
n = 1e6
h = rnorm(n, m= 1.75, s=0.1)
r = rnorm(n, m=12.13, s=0.2)
```

Now, let's find out which coins are too thin or too thick ($h < 1.5$ or $h > 2.0$).

```
badh = which(h<1.5 | h>2.0)
```

After running this, the variable `badh` holds the identifiers for every coin that is too thin or too thick (over 12,000 of them!).

Let's now do similar for too narrow or too wide ($r < 11.5$ or $r > 12.75$):

```
badr = which(r<11.5 | r>12.75)
```

The variables `badh` and `badr` contain the identifiers for the coins that should be rejected. Now, we determine those coins that are the wrong size:

```
badCoins = c(badh,badw)
```

Here is the volume calculation. Note that there are two lines. The first calculates the volume of all of the coins. The second returns just those coins that are *not* of an "inappropriate" size.

```
v = pi*h*r^2
vGood = v[-badCoins]
```

At the end of this code, the variable `v` will contain fewer than a million simulated US Quarter volumes. That is because we created 1,000,000 coins and threw out those that were too big or small.

With this, the answers to the above questions (with `R` code following) are

a) The expected volume of a US Quarter is 809.1mm^3.

   ```
   mean(vGood)
   ```

b) A 95% prediction interval for that volume is from 711 to 911 mm^3.

   ```
   quantile(vGood, c(0.025,0.975))
   ```

c) The proportion that have a volume less than 750mm^3 is 0.128.

   ```
   mean(vGood<750)
   ```

Problem 3: Another Distribution

Instead of using the Normal distribution, perhaps the coin dimensions are Uniformly distributed. That is, let's consider

$$H \sim \mathcal{U}\,(min = 1.5;\ max = 2.0)$$
$$R \sim \mathcal{U}\,(min = 11.5;\ max = 12.75)$$

Estimate the following

a) What is the expected volume of a US Quarter?
b) What is a 95% prediction interval for that volume?
c) What proportion have a volume less than 750mm^3?
d) What proportion have diameter less than 24mm?

Solution: Again, to estimate the distribution of the volume, we generate a large number of heights and radii, and apply the formula to those values.

```
n = 1000000

h = runif(n, min=1.5, max=2.0)
r = runif(n, min=11.5, max=12.75)

v = pi*h*r^2
```

Again, at the end of this code, the variable `v` will contain a million simulated US Quarter volumes. Thus, answers to the above questions are

a) The expected volume of a US Quarter is 809.1mm^3.
b) A 95% prediction interval for that volume is from 661 to 971 mm^3.
c) The proportion that have a volume less than 750mm^3 is 0.260.

Note that I did not provide the `R` code. I leave it up to you to do this. If you have questions, check back through previous problems to see what should work. By the way, your answers may slightly differ from mine because of randomness.

Now, let's have you do the rest without help from me.

⋆⋆⋆

Problem 4

My can of Monster "Mean Bean" is essentially a cylinder. The height of the cylinder is a random variable whose distribution in inches is $H \sim \mathcal{N}(\mu = 6; \sigma = 0.06)$. The diameter of the cylinder is a random variable whose distribution is $D \sim \mathcal{N}(\mu = 2; \sigma = 0.02)$.

Estimate the following quantities:

(1) the expected volume of the can
(2) the variance of the volume
(3) the expected area of metal used in its construction
(4) the variance of the area

Now, if the metal used weighs 0.00325 pounds per square inch, then calculate

(1) the expected weight of the can
(2) the variance of the weight
(3) the proportion of cans less than 0.14 pounds
(4) the proportion of cans more than 0.15 pounds

Problem 5

A page in a book is also an interesting thing to study. The thickness of the page is a random variable whose distribution in inches is $T \sim \mathcal{N}(\mu = 0.0037; \sigma = 0.0001)$. The height of the page is a random variable whose distribution in inches is $H \sim \mathcal{N}(\mu = 11; \sigma = 0.05)$. The width is a random variable whose distribution in inches is $W \sim \mathcal{N}(\mu = 8.5; \sigma = 0.05)$.

Estimate the following quantities:

(1) the expected volume of the page
(2) the variance of the volume
(3) the expected area of the page
(4) the variance of the area

If there are 275 pages in the book. Ignoring the cover, estimate

(1) the expected thickness of the book
(2) the variance of this thickness
(3) the expected volume of the book
(4) the variance of the volume