Answers (Problem set 7)

Online S520

1. (Trosset exercise 8.4.6.)

   (a) The sum $X_1 + \cdots + X_{400}$ follows an approximately Normal distribution with expected value 4, variance 4, and SD 2. The probability of a profit is approximately `1 - pnorm(0, 4, 2)` or about 98%.

   (b) The sum $Y_1 + \cdots + Y_{400}$ follows an approximately Normal distribution with expected value 0, variance 100, and SD 10. The probability of a profit is approximately `1 - pnorm(0, 0, 10)` or about 50%.

   (c) Following part (a), the probability of a profit of at least \$20 is approximately `1 - pnorm(20, 4, 2)`, which is essentially zero.

   (d) Following part (b), the probability of a profit of at least \$20 is approximately `1 - pnorm(20, 0, 10)`, which is about 2%.

   (e) Let G_{400} be the price of the growth stock after 400 days, and V_{400} be the price of the value stock after 400 days. If both stocks start at a price of \$50, then G_{400} is approximately Normal(50, 100), and V_{400} is approximately Normal(54, 4). Assuming independence, this implies $G_{400} - V_{400}$ is approximately Normal$(-4, 104)$. The probability this difference is positive is `1 - pnorm(0, -4, sqrt(104))`, which is about 35%. That's the approximate probability the growth stock is worth more than the value stock. (The real-life implication here is that although the value stock gives a positive expected return and the growth stock doesn't, it might take years to actually realize this.)

2. *Let X be a discrete random variable with probability mass function*

$$P(X = x) = \begin{cases} 0.3 & x = -2 \\ 0.6 & x = -1 \\ 0.1 & x = 12 \\ 0 & \text{otherwise.} \end{cases}$$

   Let $X_1, \ldots, X_n$ be an iid sequence of random variables with the same distribution as X. Let $\bar{X}$ be the sample mean (of $X_1, \ldots, X_n$.)

   (a) *Find EX.*
   $EX = (-2 \times 0.3) + (-1 \times 0.6) + (12 \times 0.1) = 0.$

   (b) *Find $Var(X)$.*

$$\begin{aligned} EX^2 &= (4 \times 0.3) + (1 \times 0.6) + (144 \times 0.1) = 16.2 \\ VarX &= EX^2 - (EX)^2 = 16.2 \end{aligned}$$

   (c) *What is the expected value of $\bar{X}$?*
   This is the same as EX, i.e. zero.

   (d) *What is the variance of $\bar{X}$? (Note: This will depend on n.)*
   $\sigma^2/n = 16.2/n.$

(e) *Suppose $n = 100$. Use the R function* `pnorm()` *to find the approximate probability that $\bar{X}$ is greater than 0.5.*

`1 - pnorm(0.5, mean=0, sd=sqrt(16.2/100))` gives 0.107.

3. *I want to find out the average number of people per household in the U.S. I survey a simple random sample of U.S. households and obtain the results displayed in the following table.*

Household size	Number of households
1	27
2	34
3	16
4	13
5	6
6	3
7	1

(a) *Lacking any other information, our best estimate for the population mean household size is the sample mean. What is the sample mean of our data?*

To use R, we need to enter all 100 observations. Here's a quick way:

```
households = c(rep(1,27), rep(2,34), rep(3,16), rep(4, 13), rep(5, 6), rep(6,3), 7)
mean(households)
```

This gives a sample mean of 2.5. Alternatively, we could find the empirical distribution of the data and hence find the plug-in estimate of the mean, which would give the same answer.

(b) *What is our estimate for the standard deviation of household sizes?*

`sd(households)` gives 1.41. (Alternatively, the square root of the plug-in variance is 1.40.)

(c) *What is the estimated standard error of the sample mean? (That is, based on our answer to (b), what is our estimate for the standard deviation of the distribution of the sample mean?)*

$1.41/\sqrt{100} = 0.141$.

(d) *Our error is the difference between the sample mean and the population mean. Using the normal distribution, find the approximate probability that the absolute value of the error in a survey of this form and size is less than 0.5.*

`pnorm(0.5, 0, sd(households)/sqrt(100)) - pnorm(-0.5, 0, sd(households)/sqrt(100))`
gives a probability of 99.96%.

(e) *Can we be reasonably sure that the average household size for all U.S. households is between 2 and 3?*

Yes. Since the sample mean from this survey is 2.5, and the probability that the error is less than 0.5 is almost 100%, we can be reasonably confident that the population mean (average household size) is between 2 and 3.