Sampling Distributions – Solutions

STAT-UB.0103 – Statistics for Business Control and Regression Models

Normal Random Variables (Review)

1. Suppose that X is a normal random variable with mean $\mu = 26$ and standard deviation $\sigma = 4$. What is the probability that X will take a value greater than 34?

> **Solution:** Define $Z = (X - \mu)/\sigma$, a standard normal random variable.
>
> $$\begin{aligned} P(X > 34) &= P\left(\frac{X-\mu}{\sigma} > \frac{34-\mu}{\sigma}\right) \\ &= P\left(Z < \frac{34-26}{4}\right) \\ &= P(Z > 2) \\ &= 1 - \Phi(2.00) \\ &= 1 - .9772 \\ &= .0228. \end{aligned}$$

Normal Inverse CDF

2. Suppose that Z is a standard normal random variable. Find the value w so that $\mathrm{P}(|Z| \leq w) = 0.60$.

> **Solution:** The problem is asking for w such that $\mathrm{P}(-w \leq Z \leq w) = 0.60$. Note that
>
> $$\mathrm{P}(Z < -w) + \mathrm{P}(-w \leq Z \leq w) + P(Z > w) = 1.$$
>
> Also, $\mathrm{P}(Z < -w) = \mathrm{P}(Z > w)$ (draw a picture if this is not obvious to you). In this case, we must have that $\mathrm{P}(Z < -w) = 0.20$. Therefore, $\mathrm{P}(Z < w) = 0.80$.
>
> Now, $\Phi(w) = 0.80$. The normal CDF table tells us that $\Phi(0.84) = .7795$ and $\Phi(0.85) = .8023$; we take w to be the closer of these two values, i.e. $w \approx 0.84$.

3. A machine that dispenses corn flakes into packages provides amounts that are approximately normally distributed with mean weight 20 ounces and standard deviation 0.6 ounce. Suppose that the weights and measures law under which you must operate allows you to have only 5% of your packages under the weight stated on the package. What weight should you print on the package?

Solution: Let X be the weight in ounces of a typical package; this is approximately normally distributed with mean $\mu = 20$ and standard deviation $\sigma = 0.6$. We seek a printed weight, w, such that $P(X < w) = .05$. Define $Z = (X - \mu)/\sigma$, a standard normal random variable. We have the following relation:

$$.05 = \mathrm{P}(X < w) = \mathrm{P}\left(\frac{X-\mu}{\sigma} < \frac{w-20}{0.6}\right) = \mathrm{P}\left(Z < \frac{w-20}{0.6}\right) = \Phi\left(\frac{w-20}{0.6}\right).$$

Thus,

$$\frac{w-20}{0.6} = \Phi^{-1}(.05).$$

With a normal table, we compute $\Phi(-1.64) = .0505$ and $\Phi(-1.65) = .0495$, so $\Phi^{-1}(.05) \approx -1.645$. Finally,

$$\frac{w-20}{0.6} = 1.645,$$

so $w = 20 + 0.6 \times 1.645 = 19.01$. We would probably round this to 19.01 and print "19 ounces" on the box.

4. Suppose X is a normal random variable with mean $\mu = 10$ and standard deviation $\sigma = 3$.

(a) Find the value M such that $P(X \geq M) = 0.75$.

Solution:

$$\begin{aligned}
P(X \geq M) &= .75 \\
P\left(Z \geq \frac{M-\mu}{\sigma}\right) &= .75 \\
P\left(Z < \frac{M-10}{3}\right) &= .25 \\
\frac{M-10}{3} &= \Phi^{-1}(.25) \\
\frac{M-10}{3} &= -0.67 \\
M &= 10 + 3(-.67) \\
M &= 7.99
\end{aligned}$$

(b) Find the value K such that $P(|X - \mu| \leq K) = 75\%$.

Solution:

$$
\begin{aligned}
P(|X-\mu| \le K) &= 0.75 \\
P(-K \le X-\mu \le K) &= .75 \\
P\left(-\frac{K}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{K}{\sigma}\right) &= .75 \\
P\left(-\frac{K}{3} \le Z \le \frac{K}{3}\right) &= .75 \\
P\left(Z \le \frac{K}{3}\right) &= .75 + 0.5(.25) \\
\frac{K}{3} &= \Phi^{-1}(.875) \\
\frac{K}{3} &= 1.15 \\
K &= 3.45
\end{aligned}
$$

Design-Based Inference

5. In the following scenarios, identify the sample, a relevant population, and the relevant population parameter or parameters. All parameters will either be proportions or means.

   (a) You survey your classmates and ask them whether or not they are international students.

   > **Solution:** Sample: surveyed classmates, and whether or not they are international students.
   >
   > Population: all NYU Stern freshman, and whether or not they are international students.
   >
   > Parameter: the proportion of students in the population who are international students.

   (b) You want to quickly gauge the performance of internet company stocks over the past week. You check last week's returns for Facebook, Google, and Yahoo.

   > **Solution:** Sample: last week's returns for Facebook, Google, and Yahoo.
   >
   > Population: the returns of all internet company stocks over the past week.
   >
   > Parameters: the mean return of all internet company stocks over the last week; the standard deviation of the return of all internet company stocks over the last week.

   (c) A department store manager selects 10 receipts from yesterday's purchases and records the purchase amounts.

   > **Solution:** Sample: the purchase amounts for the 10 selected receipts.
   >
   > Population: the purchase amounts for all of yesterday's receipts.
   >
   > Parameters: the mean purchase amount for all of yesterday's receipts; the standard deviation of the purchase amount for all of yesterday's receipts.

   (d) A garment factory is ready to make a shipment of 100 new sweaters. A quality control inspector checks 10 of these sweaters for defects.

   > **Solution:** Sample: the 10 inspected sweaters, and whether or not they are defective.
   >
   > Population: all 100 new sweaters, and whether or not they are defective.
   >
   > Parameter: the proportion of sweaters in the population that are defective.

Model-Based Inference

6. In the following scenarios, it is more natural to think of the population as a random process. Identify the sample, then describe the population as a random process, and identify the relevant population parameters. In each case, give at least one example of where the random process perspective could be useful. All parameters will either be probabilities or expectations.

   (a) The NYU admissions office reports that the average SAT Math score for the current Freshman class (5625 students) is 685, and the standard deviation is 80.

   Solution: Sample: SAT scores of all 5625 students.

   Population: the SAT scores of the students that got admitted and chose to attend after the random admissions and acceptance processes.

   Parameter: the mean and standard deviation of the scores of students who come to NYU via the random admissions and acceptance processes.

   How the random process view is useful: The random process perspective could be useful to forecast average SAT score next year, or to compare the admissions processes from different years.

   (b) You are considering buying a new Vizio LED TV. You look at customer ratings of the product on Amazon.com and find the average star rating.

   Solution: Sample: recorded ratings on Amazon.com.

   Population: ratings made by all people who, through some random process, decide to purchase the product on amazon.com, decide to leave a review, and come up with an overall star rating.

   Parameter: the expected star rating for reviews generated through the random population process.

   How the random process view is useful: you want to know what rating *you* would give the product if you purchase it in the future.

   (c) A bakery wants to know how many sales to expect this year. They look at weekly sales data from last year.

   Solution: Sample: weekly sales numbers for last year.

   Population: The sales generated from a random week's activity.

   Population parameter: The mean, μ of the weekly sales.

   How the random process view is useful: you want to say something about future sales.

(d) You own and operate a shoe store in SoHo. Some proportion of customers who make purchases eventually return their merchandise. Suppose you want to estimate this proportion based on all purchases made in January, of this year.

> **Solution:**
>
> Sample: Return indicators for all purchases made in January, 2012.
>
> Poplution: The random decisions customers make on whether to return or keep their merchandise.
>
> Population parameter: The probability that a purchase will get returned.
>
> How the random process view is useful: you want to say something about future returns.