

Statistics 200 Honors
Homework Solutions
Solution Set 1: Due dates August 27 – Sept. 3

p. 87, exercise 1.76 Use the 68-95-99.7 rule to find intervals centered at the mean that will include 68%, 95%, and 99.7% of the weights of the elite runners described in Exercise 1.74 who have weights normally distributed with mean 63.1 kg and standard deviation 4.8 kg.

The 68-95-99.7 rule states that the intervals $\mu \pm \sigma$, $\mu \pm 2\sigma$, and $\mu \pm 3\sigma$ contain roughly 68%, 95%, and 99.7% of the a normally distributed population, respectively. Thus, we conclude that for the given population of runners, the three intervals

$$(58.3, 67.9), (53.5, 72.7), \text{ and } (48.7, 77.5)$$

contain approximately 68%, 95%, and 99.7% of the population, respectively.

p. 89, exercise 1.88 The length of human pregnancies from conception to birth varies according to a distribution that is approximately normal with mean 266 days and standard deviation 16 days.

(a) What percent of pregnancies last fewer than 240 days?

We are given $\mu = 266$ and $\sigma = 16$ for a normal distribution. To find the proportion of this distribution less than 240, we standardize (that is, convert 240 into a z-score):

$$240 \mapsto \frac{240 - 266}{16} = -1.625.$$

In Table A, we learn that the proportion of the standard normal curve less than -1.625 is between .0516 and .0526. Let's call it 5.2%. This is the desired answer.

(b) What percent of pregnancies last between 240 and 270 days?

As above, we convert 240 and 270 into z-scores:

$$240 \mapsto \frac{240 - 266}{16} = -1.625 \quad \text{and} \quad 270 \mapsto \frac{270 - 266}{16} = 0.25$$

From Table A, we find $P(Z < -1.625) \approx 5.2\%$ and $P(Z < 0.25) \approx 59.9\%$. Therefore, the proportion between these z-values is about $59.9\% - 5.2\% = 54.7\%$.

(c) How long do the longest 20% of pregnancies last?

Since the cutoff for which 20% of pregnancies are longer is the same as the 80th percentile, we look first for the z-value z which satisfies the equation $P(Z < z) = 80\%$. (Note the difference between Z and z here: Z is a standard normal random variable and z is a constant we must find.) From Table A, we see that $z \approx 0.84$. Finally, we convert the z-value of 0.84 back to the original scale by multiplying by σ and adding μ : $0.84(16) + 266 = 279.4$ days.

p. 391, exercise 5.4

(a) You are interested in attitudes toward drinking among the 75 members of a fraternity. You choose 25 members at random to interview. One question is “Have you had five or more drinks at one time during the last week?” Suppose that in fact 20% of the 75 members would say “Yes.” Explain why you *cannot* safely use the $B(25, 0.2)$ distribution for the count X in your sample who say “Yes.”

The population of 75 is too small in relation to the sample of 25 in order that the binomial assumption of independence is approximately satisfied. The text recommends that the population be at least 10 times as large as the sample in order that we may safely ignore the effect on the rest of the population of drawing the individuals for the sample.

(b) The National AIDS Behavioral Surveys found that 0.2% of adult heterosexuals had both received a blood transfusion and had a sexual partner from a group at high risk of AIDS. Suppose that this national proportion holds for your region. Explain why you *cannot* safely use the normal approximation for the sample proportion who fall in this group when you interview an SRS of 500 adults.

With a sample size of 500 and a population proportion of 0.2%, the number of people in the sample who have the characteristic of interest DOES follow a binomial (500, .002) distribution, but the text recommends that np and $n(1 - p)$ should both be at least 10 before the normal approximation is very good. In this case, with $n = 500$ and $p = .002$, np is only 1.

p. 395, exercise 5.18 According to government data, 21% of American children under the age of six live in households with incomes less than the official poverty level. A study of learning in early childhood chooses an SRS of 300 children.

(a) What is the mean number of children in the sample who come from poverty-level households? What is the standard deviation of this number?

In this case, the sample of 300 from the enormous U.S. population of children will behave much like a binomial (300, .21) distribution. The mean of a binomial variable is np , which in this case is $300(.21) = 63$. The standard deviation of a binomial variable is $\sqrt{np(1 - p)}$, which in this case is $\sqrt{300 \cdot 0.21 \cdot 0.79} = 7.05$.

(b) Use the normal approximation to calculate the probability that at least 80 of the children in the sample live in poverty. Be sure to check that you can safely use the approximation.

Since $np = 63 \geq 10$ and $n(1 - p) = 237 \geq 10$, the normal approximation to the binomial is safe. We know that the mean and standard deviation of the normal distribution should be the same as those of the binomial distribution it is approximating, namely (from part (a)) $\mu = 63$ and $\sigma = 7.05$. We are asked to find the proportion of this population greater than or equal to 80. (Note: for technical reasons, it is slightly better to find $P(X > 79.5)$ instead of $P(X > 80)$. See why? The reason is discussed in the text under the topic of “continuity correction”. We won’t worry about it here.) Converting to a z-score, we get

$$80 \mapsto \frac{80 - 63}{7.05} = 2.41.$$

Looking in Table A, we find a proportion 99.2% to the left, but we were asked for area to the right, so subtract from one to get the final answer of 0.8%. (Incidentally, the actual answer for the binomial distribution without the normal approximation is 1.1%, so our answer isn't far off.)

p. 408, exercise 5.26 The scores of students on the ACT college entrance examination in a recent year had the normal distribution with mean $\mu = 18.6$ and standard deviation $\sigma = 5.9$.

(a) What is the probability that a single student randomly chosen from all those taking the test scores 21 or higher?

With $\mu = 18.6$ and $\sigma = 5.9$, converting to a z -score gives

$$21 \mapsto \frac{21 - 18.6}{5.9} = 0.41.$$

Thus, Table A reveals that the desired proportion is $1 - 65.9\% = 34.1\%$.

(b) Now take an SRS of 50 students who took the test. What are the mean and standard deviation of the sample mean score \bar{x} of these 50 students?

For a SRS from a normal (μ, σ) population, the sample mean has a normal $(\mu, \sigma/\sqrt{n})$ distribution. Thus, the mean and standard deviation of \bar{x} are $\mu = 18.6$ and $\sigma/\sqrt{50} = 0.834$.

(c) What is the probability that the mean score \bar{x} of these students is 21 or higher?

Using the information in part (b), we convert to a z -score,

$$21 \mapsto \frac{21 - 18.6}{0.834} = 2.88,$$

then use Table A to conclude that the desired probability is $1 - 99.8\% = 0.2\%$. Compare this answer with the answer to part (a). Convince yourself that it makes sense that the sample mean is much less likely than a given individual score to be greater than 21.

p. 448, exercise 6.6 A test for the level of potassium in the blood is not perfectly precise. Moreover, the actual level of potassium in a person's blood varies slightly from day to day. Suppose that repeated measurements for the same person on different days vary normally with $\sigma = 0.2$.

(a) Julie's potassium level is measured once. The result is $x = 3.4$. Give a 90% confidence interval for her mean potassium level.

A confidence interval for the population mean is

$$\bar{x} \pm z^* \left(\frac{\sigma}{\sqrt{n}} \right).$$

In this case, we know $\bar{x} = 3.4$, $n = 1$, $\sigma = 0.2$, and the z^* value corresponding to a 90% confidence interval is 1.645. Plugging all of this in gives us the interval (3.07, 3.73).

(b) If three measurements are taken on different days and the mean result is $\bar{x} = 3.4$, what is a 90% confidence interval for Julie's mean blood potassium level?

We use the same formula as in part (a), except now $n = 3$. This gives a narrower interval, namely (3.21, 3.59).

p. 449, exercise 6.10 How large a sample of the hotel managers in Exercise 6.7, whose stay at their current company has a standard deviation of 3.2 years, would be needed to estimate the mean μ within ± 1 year with 99% confidence?

There is a direct formula for this in the text, but it's not hard to reason out: We know that the confidence interval for the mean μ will be

$$\bar{x} \pm z^* \left(\frac{\sigma}{\sqrt{n}} \right),$$

where in this case $\sigma = 3.2$ and $z^ = 2.576$ because that is the z^* associated with a 99% confidence interval. The thing we don't know is n , but we are told that our confidence interval should be $\bar{x} \pm 1$. Thus, we do a little algebra and solve for n :*

$$2.576 \left(\frac{3.2}{\sqrt{n}} \right) = 1 \implies n = \left(\frac{2.576 \cdot 3.2}{1} \right)^2 = 67.95.$$

Therefore, we need 68 subjects in the sample. Notice that we always round up to the nearest integer; since we cannot have a fraction of a person in the sample, we err on the side of caution by making the sample slightly larger than we need.