Math 218, Mathematical Statistics

D Joyce, Spring 2016

Assignment. Chap. 5, p. 189, exercises 4, 5, 8, 14, 16, 17, 23 .

Selected answers.

4. A soft drink company uses a filling machine to fill cans. Each 12 oz. can is to contain 355 milliliters of beverage. In fact, the amount varies according to a normal distribution with mean $\mu = 355.2$ ml and standard deviation $\sigma = 0.5$ ml.

a. What is the probability that an individual can contains less than 355 ml?

Our random variable X is normal with mean $\mu = 355.2$ and standard deviation $\sigma = 0.5$. Here are three slightly different ways you can get the answer. You'll probably find one of these methods a little more to your liking.

Method 1: The mean is 355.2, so the value 355 is 0.2 ml below the mean. Since a standard deviation is 0.5 ml, that means 0.2 ml is $0.2/0.5 = 0.4$ standard deviations. That means that we need to find the probability that a standard normal random variable Z has a value less than -0.4. In the notation of our text, that's $\Phi(-0.4)$. Look that up in Table A.3 of standard normal curve areas $\Phi(z) = P(Z \leq z)$. You get $\Phi(-0.4) = 0.3446$.

Comment: Thus about $\frac{1}{3}$ of the cans are too light.

Method 2: Normalize X to get a standard normal distribution Z. Then $Z = (X - \mu)/\sigma$. We want to know $P(X \leq 355)$. But $X \leq 355$ iff $X - \mu \leq 355 - \mu$ iff $Z = (X - \mu)/\sigma \leq (355 - \mu)/\sigma$ iff $Z \leq (355 - 355.2)/0.5$ iff $Z \leq -0.4$. Therefore, we need $P(Z \leq -0.4)$. Proceed as in method 1 to get 0.3446.

Method 3: Normalize X. Then $Z = (X - \mu)/\sigma$. We need $P(X \leq 355)$. But $X = \sigma Z + \mu$, so that means we need $P(\sigma Z + \mu \leq 355)$. That simplifies to $P(Z \leq -0.4)$, so proceed as before.

b. What is the probability that the mean content of a six-pack of cans is less than 355 ml?

The last part was about the capacity X of 1 can, but in this part, we take the average $\overline{X} = \frac{1}{6}(X_1 + \cdots + X_6)$ of 6 cans. The mean of $\overline{X}$ is the same as the mean of X, namely $\mu = 355.2$, but the standard deviation of $\overline{X}$ is different from that of X. In fact,

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{0.5}{\sqrt{6}} = 0.2041.$$

Now, 355 is 0.2 below the mean of 355.2, and since one standard deviation for $\overline{X}$ is 0.2041, that means 0.2 is $0.2/0.2041 = 0.9798$ standard deviations. That means that we need to find the probability that a standard normal random variable Z has a value less than -0.9798. Look up $\Phi(-0.9798)$ in the table to get 0.1635.

Comment. Even though about $\frac{1}{3}$ of the cans are that light, the probability that six cans will average out to be that light is only about $\frac{1}{6}$.

5. A random sample $X_1, \ldots, X_{150}$ is drawn from a population with mean $\mu = 40$ and standard deviation $\sigma = 15$ but from an unknown distribution. Let $U = (X_1 + \cdots + X_{50})/50$ represent the sample mean of the first 50 observations and $V = (X_{51} + \cdots + X_{150})/100$ the sample mean of the last 100 observations.

a. What are the approximate distributions of U and V?

Since they're large samples, the central limit theorem says they'll be approximately normal. They'll have the same mean as the population mean, that is, $\mu_U = \mu_V = \mu = 40$. But their variances will be smaller than the population variances by factors of their sizes, that is, $\sigma_U^2 = \sigma^2/50 = 15^2/50 = 4.5$ and $\sigma_V^2 = \sigma^2/100 = 15^2/100 = 2.25$

(Alternatively, you could answer this question in terms of standard deviations. Their standard deviations will be smaller than the population standard deviations by factors of the square roots of their sizes, that is, $\sigma_U = \sigma/\sqrt{50} = 15/\sqrt{50} = 2.121$ and $\sigma_V = \sigma/\sqrt{100} = 15/\sqrt{100} = 1.5$.)

b. Which would you expect to be larger, $P(35 \leq U \leq 45)$ or $P(35 \leq V \leq 45)$? Why?

The interval is symmetric about the mean 40, from $40 - 5$ to $40 + 5$, so the question is which of U and V has a smaller variation. Since V has the smaller variation, it's more likely to be inside this interval. Therefore, $P(35 \leq V \leq 45)$ is larger.

c. Find $P(35 \leq U \leq 45)$ and $P(35 \leq V \leq 45)$ using the normal approximation.

For U, a standard deviation is 2.121, so 5 is $5/2.121 = 2.357$ standard deviations. We need $P(-2.357 \leq Z \leq 2.357)$. That's the same as $1 - 2\Phi(-2.357) = 1 - 2 \cdot 0.0091 = 0.9818$.

For V, a standard deviation is 1.5, so 5 is $5/1.5 = 3.333$ standard deviations. We need $P(-3.333 \leq Z \leq 3.333)$. That's the same as $1 - 2\Phi(-3.333) = 1 - 2 \cdot 0.0004 = 0.9992$.

8. The lifetime of disc brake pads varies according to a normal distribution with mean $\mu = 50,000$ miles and standard deviation $\sigma = 3000$ miles. Suppose that a sample of nine break pads is tested.

a. What is the distribution of the sample mean $\overline{X}$? Give the mean and standard deviation of this distribution.

It's a normal distribution. Its mean is $\mu_{\overline{X}} = \mu = 50,000$ miles. Its standard deviation is $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 3000/\sqrt{9} = 1000$ miles.

b. Does your answer in part a require the use of the central limit theorem? Why or why not?

No. Since the population distribution is normal, a sample mean $\overline{X}$ is normal, too. Not approximately normal as in the conclusion to the central limit theorem, but exactly normal.

c. Suppose that a sample mean less than 47,000 miles is considered evidence that $\mu < 50,000$ miles. What is the probability that this will happen even when the true mean $\mu = 50,000$, leading to an incorrect conclusion?

We need to compute $P(\overline{X} < 47,000)$. Now, 47,000 miles is 3,000 miles below the mean, and 3,000 miles is 3 standard deviations for $\overline{X}$. Therefore, $P(\overline{X} < 47,000) = P(Z < -3) = \Phi(-3) = 0.0013$. So the probability of an incorrect conclusion under the given assumptions is 0.13%, almost 0.

14. In an ESP experiment, five choices are offered for each question. Assume that a person without ESP guesses randomly and thus correctly answers with probability $\frac{1}{5}$. Further assume that the responses are independent. Suppose that 100 questions are asked.

a. What are the mean and standard deviation of the number of correct answers?

Each Bernoulli trial X_i has mean $\mu = p = \frac{1}{5}$ and variance $\sigma^2 = pq = \frac{1}{5} \cdot \frac{4}{5} = \frac{4}{25}$, so the standard deviation is $\sigma = \frac{2}{5}$. Therefore the sample sum, $S_n = \sum_{i=1}^n X_i$, which is a binomial distribution, has mean $\mu_{S_n} = n\mu = npq = 20$ and variance $\sigma^2_{S_n} = npq = 16$, and so its standard deviation is $\sigma_{S^n} = 4$.

b. What are the mean and standard deviation of the proportion of correct answers?

The proportion of correct answer is the same as the sample mean $\overline{X}$. Therefore, $\mu_{\overline{X}} = \mu = \frac{1}{5}$ and $\sigma_{\overline{X}} = \sigma/\sqrt{n} = \frac{2}{5} \cdot \frac{1}{10} = 0.04$.

c. What is the probability that a person without ESP will correctly answer at least 30 of the 100 questions? Check if the normal approximation will give accurate results. If so use it with a continuity correction.

We need the probability $P(S_n \geq 30)$ with $n = 100$. That's the sum

$$\sum_{k=30}^{n} \binom{n}{k} p^k q^{n-k} = \sum_{k=30}^{100} \binom{100}{k} \left(\tfrac{1}{5}\right)^k \left(\tfrac{4}{5}\right)^{100-k}.$$

There is a table to find cumulative sums like this, Table A.1 page 669, but it only goes up to $n = 25$. Beyond that the normal approximation is pretty good.

Now, S_n is approximately normal with mean 20 and standard deviation 4. With the continuity correction, we need to compute $P(S_n \geq 29.5)$. That's 9.5 units above the mean, which is $9.5/4 = 2.375$ standard deviations above the mean. We need the probability $P(Z \geq 2.375)$ which is $1 - \Phi(2.375) = 1 - 0.9912 = 0.0088$.

Thus, about 1% of people without ESP will guess 30 or more questions correctly.

16. Using table A.5, find the following values: $\chi^2_{5,0.01}$, $\chi^2_{10,0.05}$, $\chi^2_{10,0.95}$, $\chi^2_{10,0.75}$.

The value $\chi^2_{5,0.01}$ is the value of x such that a χ^2 random variable X with 5 degrees of freedom has the probability 0.01 of being greater than x, that is, $P(X \geq x) = 0.01$. That's found in the fifth row of the table, in the next to the last column. $\chi^2_{5,0.01} = 15.085$.

Likewise, $\chi^2_{10,0.05} = 18.307$ and $\chi^2_{10,0.95} = 3.940$.

There is no column for $\alpha = 0.75$ in the table, so $\chi^2_{10,0.75}$ can't be determined from it.

17. Consider the χ^2 distribution with 8 degrees of freedom.

a. What are the mean and variance of a χ^2_8 random variable?

The mean of a χ^2 distribution with ν degrees of freedom is ν, which is 8 in this exercise, and the variance is 2ν, which is 16 in this exercise.

b. Using table A.5, find constants a, b, c, d, and e so that
$P(\chi^2_8 > a) = 0.05$,
$P(\chi^2_8 > b) = 0.99$,
$P(\chi^2_8 < c) = 0.90$, and
$P(d < \chi^2_8 < e) = 0.95$.

See part **c**.

c. Express the constants from part b in terms of the notation $\chi^2_{\nu,\alpha}$.

$P(\chi^2_8 > a) = 0.05$ when $a = \chi^2_{8,0.05} = 15.507$.
$P(\chi^2_8 > b) = 0.99$ when $b = \chi^2_{8,0.99} = 1.646$.
$P(\chi^2_8 < c) = 0.90$ when $P(\chi^2_8 > c) = 0.10$, and that's when
$c = \chi^2_{8,0.10} = 13.362$.

We can make $P(d < \chi^2_8 < e) = 0.95$ by making the two tails the same, that is, $P(\chi^2_8 > e) = 0.025$ and $P(\chi^2_8 > d) = 0.975$. That happens when $d = \chi^2_{8,0.975} = 2.180$ and $e = \chi^2_{8,0.025} = 17.534$. (You may have reason to make the tails different sizes, in which case there are other values for d and e.

d. Sketch each of the probabilities from part b as an area under the χ^2_8 probability density curve.

23. An engineer suspects that the temperature inside an oven is not as uniform as when it was new, at which time the temperature varied $\pm 10°$ F around its setting. (Taking the range of a normal distribution to be roughly $\pm 2\sigma$, this translates into $\sigma = 5°$ F.) To verify his suspicion, he takes 20 measurements in different parts of the oven. He wants a rule which decides that the true $\sigma > 5$ if the sample standard deviation of the measurements exceeds $5c$ where $c > 0$ is a suitably chosen constant. The rule must not have more than a 10% chance of making a wrong decision, i.e., deciding that $\sigma > 5$ when, in fact, $\sigma = 5$.

a. Find the value of c.

In this exercise, $n = 20$ and $\sigma = 5$.

We need to find a value c so that $P(S > \sigma c) = 0.10$. That's the same as $P(S^2 > c^2\sigma^2) = 0.10$. Now, $\dfrac{(n-1)S^2}{\sigma^2}$ is a χ^2 random variable with $n - 1 = 19$ degrees of freedom. Convert the condition $S^2 > c^2\sigma^2$ so the left hand side becomes that variable. The condition becomes

$$\frac{(n-1)S^2}{\sigma^2} > (n-1)c^2.$$

Thus, we need $P(\chi^2_{19} > 19c^2) = 0.10$. That happens when $19c^2 = \chi^2_{19,0.10} = 27.203$. So $c^2 = 27.203/19 = 1.4317$, and therefore $c = 1.197$. Since the original specification of that σ was 5° F was really only accurate to one or two digits, c should only have that accuracy, $c = 1.2$.

b. Based on this value of c, does the rule decide that $\sigma > 5$ if the sample standard deviation of the engineer's measurements is $s = 7.5°$ F?

Is $7.5 > 5c$? Since $5c = 6$, yes it is. That is, if the measured standard deviation of 20 measurements is higher than 6° F, then this test concludes that the temperature standard deviation is not 5° F.

Math 218 Home Page at
`http://math.clarku.edu/~djoyce/ma218/`