

Lecture 17

Nancy Pfenning Stats 1000

In general, if we want to know the probability that any normal variable X falls in a given interval, we rewrite this as a problem about the standardized normal variable $Z = \frac{X-\mu}{\sigma}$, then use Table A.1, which gives the probability of a standard normal variable Z being less than any value from -3.49 to +3.49.

Example

If X is normal with mean 80, standard deviation 10, find...

1. $P(X \leq 93) = P\left(\frac{X-80}{10} \leq \frac{93-80}{10}\right) = P(Z \leq 1.3) = .9032$.
2. $P(X < 93) = P\left(\frac{X-80}{10} < \frac{93-80}{10}\right) = P(Z < 1.3) = .9032$. Note that for a finite data set like female class members' heights, represented with a histogram, whether or not we had strict inequality was important: the proportion with height less than *or equal to* 65 could be quite a bit more than the proportion with height strictly less than 65. For an abstract "infinite" population of values represented with a density curve, the area under the curve $\leq X$ is the same as the area $< X$, and so we need not be concerned with strict inequality or not for normal distributions.
3. $P(X \leq 80) = P\left(\frac{X-80}{10} \leq \frac{80-80}{10}\right) = P(Z \leq 0) = .5$ This makes sense because the normal curve is symmetric, so the mean is the middle, and the proportion below the mean must equal the proportion above and together they equal 1, so each is .5.
4. $P(65 < X < 100) = P\left(\frac{65-80}{10} < Z < \frac{100-80}{10}\right) = P(-1.5 < Z < 2) = .9772 - .0668 = .9104$.
5. $P(X > 70) = P\left(Z > \frac{70-80}{10}\right) = P(Z > -1) = P(Z < +1) = .8413$.
6. $P(X < 35) = P(Z < -4.5) = 0$.
7. $P(X > 35) = P(Z > -4.5) = 1$.
8. $P(X < 120) = P(Z < 4) = 1$.
9. $P(X > 120) = P(Z > 4) = 0$.

Just as we worked standard normal problems going in both directions, we have two kinds of non-standard normal problems. In the previous example, we were given a non-standard normal value x and were asked to find the corresponding probability. In the following example, we will be given a probability, and must find the corresponding non-standard normal value. The best approach is in two steps: first find the z value corresponding to the given probability, then "unstandardize": Note that since $z = \frac{x-\mu}{\sigma}$, it follows that $x = \mu + z\sigma$.

Example

1. Suppose exam scores X are normal with $\mu = 80$, $\sigma = 10$. What is the 40th percentile of X ? The probability of .4000 (approximately) has $z = -.25$ and so we unstandardize to $x = 80 - .25(10) = 77.5$. In other words, 40% of the scores were below 77.5.
2. Suppose blood cholesterol levels X in middle-aged men are normal with $\mu = 222$, $\sigma = 37$. The highest 3% are above what level? 3% above corresponds to 97% below; the probability of .9700 has $z = 1.88$ and so we unstandardize to $x = 222 + 1.88(37) = 291.56$, which we may round to 292.
3. The distribution of heights X of men in the U.S. is normal with mean 69, standard deviation 3.
 - (a) Suppose I want to classify the shortest 10% as being "short", the tallest 10% as being "tall", and the rest as being "medium". What should be the cut-off heights S and T ? A probability of .1000 has $z < -1.28$, so $S = 69 - 1.28(3) = 65.16$. Similarly a probability of .1000 above has .9000 below, which has $z = +1.28$ and $T = 69 + 1.28(3) = 72.84$.

- (b) If a man is in the 85th percentile for height, can he pass barefoot through a 72-inch doorway without stooping? $.8500 = P(Z < 1.04)$, so $x = 69 + 1.04 * 3 = 72.12$. No, he can't quite pass through the doorway without stooping.
4. The time X in minutes required for deliveries by a certain pizza franchise has $\mu = 21.6$ and $\sigma = 3.6$. The owners want to offer their customers free pizza if delivery takes longer than a certain amount of time. What should the time limit be if they want to award free pizza 1% of the time?
A right-tail probability of .01 corresponds to the same z that has a left-tail probability of .99: $z = 2.33$ and we unstandardize to $x = 21.6 + 2.33(3.6) = 29.988$, or just about 30 minutes.

Example

IQ's X of children at a large gifted center are normal with $\mu = 128$ and $\sigma = 1.8$. Is 126 unusually low? $P(X < 126) = P(Z < \frac{126-128}{1.8}) = P(Z < -1.11) = .1335$. If more than 13% are below 126, we shouldn't consider 126 to be unusually low.

Example

A class of mine in a previous semester had course grades with mean 79, standard deviation 14. Assume scores were approximately normally distributed.

1. What proportion failed (i.e., scored below 60)? $P(X < 60) = P(Z < \frac{60-79}{14}) = P(Z < -1.36) = .0869$, or about 9%. [In fact, 7 out of 69, or about 10%, failed.]
2. The lowest 3% scored below what value? $.0300 = P(Z < -1.88)$ so $x = 79 - 1.88(14) = 52.68$. In fact, 3 out of 69, or about 4%, scored below 52.68.

Normal Practice Exercises

Assume all the data sets in these exercises follow a normal curve closely enough that a normal approximation may be used.

1. μ is 100 points, σ is 10 points.
 - (a) Find the proportion of values below 70 points.
 - (b) The middle 95% of the values are between what two point values?
2. μ is 6 inches, σ is 1.5 inches.
 - (a) Find the proportion of values below 6 inches.
 - (b) The shortest 16% are shorter than how many inches?
3. μ is 20 cigarettes, σ is 5 cigarettes.
 - (a) Find the proportion of values above 22.5 cigarettes.
 - (b) The top 10% are greater than how many cigarettes?
4. μ is 165 lbs., σ is 12 lbs.
 - (a) Find the proportion of values below 148 pounds.
 - (b) The lightest 2% are lighter than how many pounds?
5. μ is 4 people, σ is 1.3 people. Find the proportion of values less than 2 people.
6. μ is 11 years, σ is 2 years.
 - (a) Find the proportion of values less than 17 years.

- (b) The longest 7% are longer than how many years?
- μ is \$30,000; σ is \$8,000. Find the proportion of values between \$20,000 and \$22,000.
 - μ is 120 mm., σ is 20 mm. Find the proportion of values greater than 112 mm.
 - μ is 6 feet, σ is .2 feet. Find the proportion of values greater than 6.5 feet.
 - μ is 60 degrees, σ is 16 degrees. Find the proportion of values below 79 degrees.
 - μ is 300 ml., σ is 3 ml. Find the proportion of values (a) less than 280 ml (b) greater than 280 ml (c) less than 315 ml (c) greater than 315 ml.

Approximating Binomial Distribution Probabilities

Binomial R.V.'s can be approximated with normal R.V.s having the same mean and standard deviation, as long as the sample size n is large enough relative to the shape of the population (determined by p): we require that $np \geq 10$ and $n(1-p) \geq 10$. This requirement ties in with the **Central Limit Theorem**, which will be discussed in more detail in the next chapter. It is worth noting at this point that in general, when we are interested in a single categorical variable which fits the binomial model and the sample size is large, our approach will utilize normal probabilities.

Another important detail is the fact that we are shifting now from a discrete distribution (binomial count) to a continuous one (the normal distribution). By using what's called the **continuity correction**, it is possible to accurately approximate any binomial probability, as long as the above requirement holds. For simplicity's sake, we will forego the continuity correction, and only solve problems which can easily be expressed in terms of **cumulative probability**, that is, the probability that our binomial or normal random variable takes a value *less than or equal to* a particular value k .

Example

The proportion of Pitt students who take four years to graduate is .33. Use a normal approximation to find the probability that in a group of 90 students, at least 45 take four years to graduate. The count X of those students who take four years to graduate is binomial with $n = 90$, $p = .33$. A normal approximation is appropriate because $np = 30$ and $n(1-p) = 60$ are both greater than 10. Our binomial X has $\mu = np = 90(.33) = 30$, $\sigma = \sqrt{np(1-p)} = \sqrt{90(.33)(.67)} = 4.46$. (Of those 90 students, we'd expect about 30, give or take 4 or 5, to take four years to graduate.) If we subtract μ from X and divide by σ , we call the resulting variable Z because it is approximately standard normal:

$$P(X \geq 45) \approx P(Z \geq \frac{45 - 30}{4.46}) = P(Z \geq 3.36) = P(Z \leq -3.36) = .0004$$

Thus, it is almost impossible for as many as 45 (half) of a group of 90 Pitt students to take four years to graduate.

Example

Assume that the proportion of all college students who can be classified as "binge drinkers" is .44. Suppose a class has 100 students.

- Find μ and σ for X :

$$\mu = np = 100 * .44 = 44; \quad \sigma = \sqrt{np(1-p)} = \sqrt{100 * .44 * .56} \approx 5$$

- Suppose 38 of the 100 students claim to be binge drinkers. Is this surprisingly low? We'll find

$$P(X \leq 38) = P(Z \leq \frac{38 - 44}{5}) = P(Z \leq -1.2) = .1151$$

In general, we wouldn't call something "surprising" if its probability is .1151; such an occurrence is not unusual.

Example

A university has found that out of all the students who are offered admission, the proportion who accept is .70. Suppose they offer admission to 1700 students.

1. Find μ and σ for X :

$$\mu = np = 1700 * .70 = 1190; \quad \sigma = \sqrt{np(1-p)} = \sqrt{1700 * .7 * .3} = 18.9$$

2. Find $P(1150 < X < 1250)$

$$= P\left(\frac{1150 - 1190}{18.9} < Z < \frac{1250 - 1190}{18.9}\right) = P(-2.12 < Z < 3.17) = .9992 - .0170 = .9822$$

Note: We will not cover sums, differences, and combinations of random variables.