CHAPTER SIX: NORMAL DISTRIBUTIONS

Review January 25 ⚘ Test February 3

As opposed to the discrete distributions in the previous chapter, the normal distribution is a continuous distribution, and its graph is a curve. Real world data centered around a mean, such as weights of people, normally follow normal distributions. By converting raw scores into standardized (z) scores representing the number of standard deviations a raw score is above the mean, probabilities and other percentages can be calculated regarding where a score fits in a distribution. For example, a score of 12 in a normal distribution with $\mu = 10$ and $\sigma = 2$ would have a z score of 1 and would be higher than 84% of other scores in the distribution. This value of 84% is the same for $z = 1$ (that is, one standard deviation above the mean) in any normal distribution, not just $\mu = 10$ and $\sigma = 2$. Even if a distribution is not normal, sample means from the distribution tend to be normally distributed.

6-A The Normal Curve — Monday • 1/9

z score • normal curve

❶ Convert a raw score x to a standardized score z.
❷ Sketch and label a normal curve for a given μ and σ and use it to calculate probabilities or frequencies.

6-B Normal Probabilities — Wednesday • 1/11

❶ Find the area under the normal curve above or below a z score.
❷ Find the area under the normal curve between two z scores.
❸ Calculate normal probabilities or frequencies from raw scores.

6-C Percentiles and the Normal Curve — Friday • 1/20

❶ Find the z score needed to achieve a given percentile.
❷ Find the z scores needed to achieve a given range centered about the mean.
❸ Convert a standardized score z to a raw score x.
❹ Calculate the raw score needed to be higher or lower than a specified percentage of the population.

6-D The Central Limit Theorem — Wednesday • 1/25

law of large numbers • sampling distribution • standard error • central limit theorem

❶ Use the law of large numbers to determine whether a statistic is more likely to fall within a given range for a small sample or for a large sample.
❷ Calculate sampling probabilities or frequencies of sample means, or calculate sample means from percentiles.

6-A The Normal Curve

A Z SCORE, or Standardized Score, is the number of standard deviations a raw score x is above the mean: $z = \frac{x - \mu}{\sigma}$.

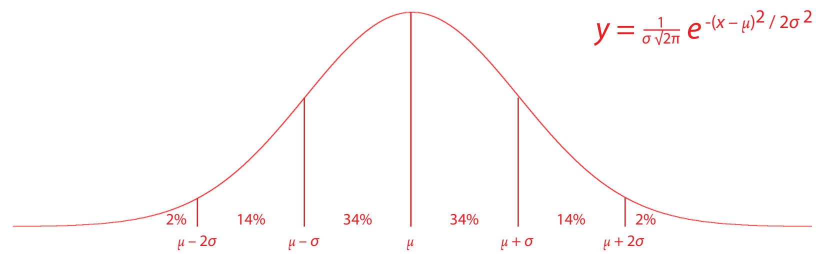
❶ Convert a raw score x to a standardized score z.

1. Subtract the mean from the raw score.
2. Divide the difference by the standard deviation.

❶ Calculate the z score of a 25-year-old man weighing 75 kg, given $\mu = 77$ kg and $\sigma = 13$ kg.

$z = \frac{75 - 77}{13} \approx -0.15$

The Standard NORMAL Curve at right shows the Normal Probability Distribution in terms of z scores. As shown by the curve, most data in a normal distribution are close to the mean ($z = 0$), and very few are more than two standard deviations away from the mean ($z = \pm 2$). The curve is called normal because real-world data normally are distributed in this manner.



The percentages at right show the area under the curve in a given region, which is the same as the probability of a data value falling within that region.

Because areas under the normal curve represent probabilities, they are typically written using probability notation. For example, "15.9% of the normal curve is above $z = 1$" can be expressed "$P(z > 1) = 15.9\%$".

Unlike binomial distributions which use discrete data, normal distributions use continuous data. The larger the sample size is for a binomial distribution with $p = ½$, the more the distribution will smooth out into a normal curve.

❷ Sketch and label a normal curve for a given μ and σ, and use it to estimate probabilities or frequencies.

1. Copy the red normal curve above.
2. Label μ in the middle.
3. Add or subtract σ to get from each boundary line to the next.
4. Add the areas between the boundaries stated in the problem. For areas not bounded by integer values of z, use the areas shown to estimate areas, keeping in mind that the outside half of each section of the curve is smaller than the inside half.

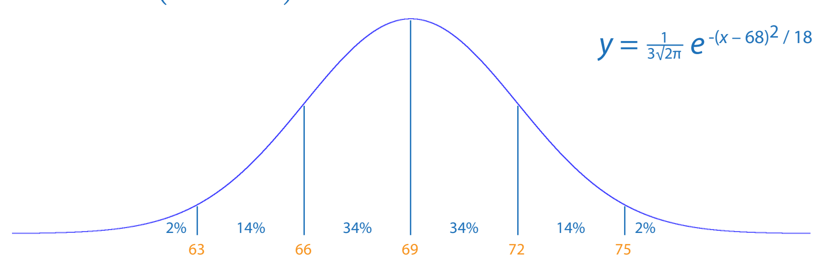
❷ Given the mean height of college men is 69" with standard deviation 3", estimate the following.

a) the percentage of college men between 66" and 75" tall

4. 34% + 34% + 14% = 82%

b) the probability that a random one will be shorter than 67"

2% + 14% + 8% (estimated) = 24%



6-B Normal Probabilities

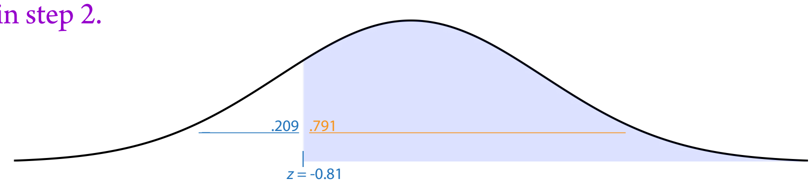
The Z table (Table 5) in the book gives areas under the standard normal curve below z, so that probabilities can be calculated rather than estimated. However, it is a good idea to sketch a curve, approximately to scale, and label it. This can help with identifying what the problem is asking, and it can make clear when an answer is unreasonable, such as if an answer is 35% but more than half the curve is shaded.

❶ Find the area under the normal curve above or below a z score.

1. On the left side of Table 5 look up z, rounded down to the next tenth.
2. Move across to the appropriate hundredths place, as shown at the top. This is the area under the curve below z.
3. If finding the area under the curve above z, take the complement of the answer in step 2.

❶ Find the area under the normal curve above $z = -0.81$.

1. Go down to the row for $z = -0.8$.
2. Go over to the column for .01: $P(z < -0.81) = .209$
3. $P(z > -0.81) = 1 - .209 = .791$

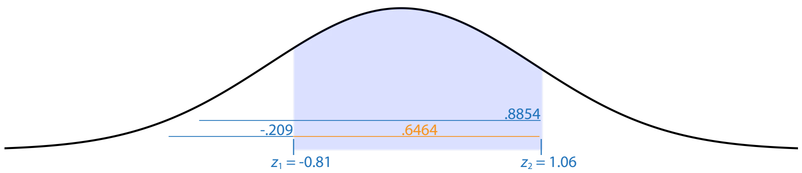


❷ Find the area under the normal curve between two z scores.

1 Use the z table to find the area under the normal curve below the first z score.
2. Use the z table to find the area under the normal curve below the second z score.
3. Subtract the smaller area from the larger area.

❷ Find the area under the normal cuve between -0.81 and 1.06.

1. $P(z < -0.81) = .209$
2. $P(z < 1.06) = .8554$
3. $P(-0.81 < z < 1.06) = .8554 - .2090 = .6464$

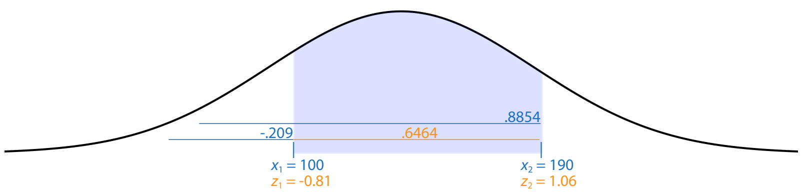


❸ Calculate normal probabilities or frequencies from raw scores.

1. Convert the raw scores to z scores (see 6-A).
2. Use the z scores to calculate the probability (see ❷).

❸ On a test with normally distributed scores with $\mu = 139$ and $\sigma = 48$, what percent of scores are between 100 and 190?

1. $z_1 = \frac{100 - 139}{48} \approx -0.81$ $z_2 = \frac{190 - 139}{48} \approx 1.06$
2. $P(100 < x < 190) = P(-0.81 < z < 1.06) \approx .8554 - .2090 \approx 64.6\%$



6-C Percentiles and the Normal Curve

The z table can be used in reverse, that is, to find a z score from a known percentile rather than to find a percentile from a known z score.

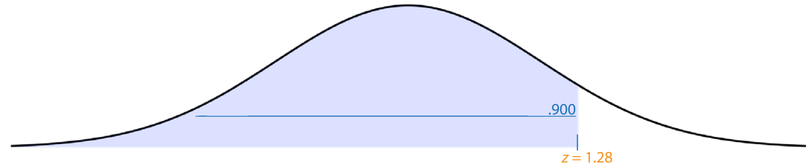
❶ Find the z score needed to achieve a given percentile.

1. On the inside of the z table, find the number closest to the desired percentile.
2. From this number, move to the left of the table and to the top of the table.
3. Add the number on the top to the number on the left.

❶ What z score is at the 90th percentile?

$P(z < 1.28) = .8997 \approx 90\%$

$z = 1.28$

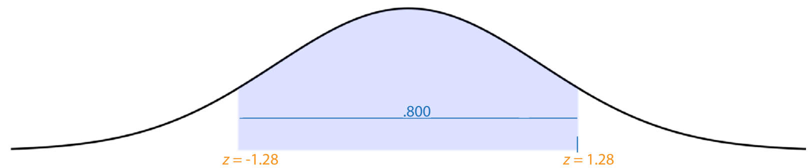


❷ Find the z scores needed to achieve a given range centered about the mean.

1. Find the top of the range by adding half of its size to 50%.
2. Find the z score for this percentile (see ❶). Call it z_0.
3. The z scores needed are $\pm z_0$.

❷ The middle 80% of the normal curve lies between which z scores?

1. $50\% + (80\% \div 2) = 90\%$
2. $P(z > 1.28) = 90\%$
3. $z = \pm 1.28$



Z-scores can be converted to raw scores by using the z formula in 6-B backward: $x = \mu + z\sigma$.

❸ Convert a standardized score z to a raw score x.

1. Add z standard deviations to the mean.

❸ Find the height of a college woman with a z score of -1.08 given $\mu = 65$” and $\sigma = 2.5$”.

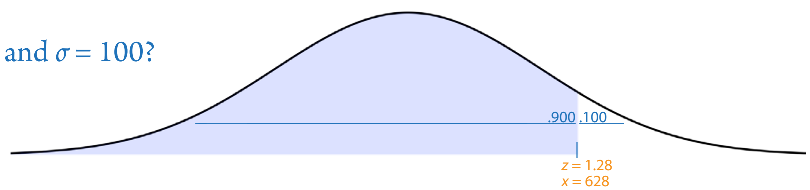
$x = 65 - 1.08(2.5) = 62.3$”

❹ Calculate the raw score needed to be higher or lower than a specified percentage of the population.

1. Identify the percentile specified.
2. Find the z score for this percentile (see ❶).
3. Convert the z score to a raw score (see ❸).

❹ How high of a score is needed on an SAT I to be in the top 10%?, given $\mu = 500$ and $\sigma = 100$?

1. The top 10% starts at the 90th percentile.
2. $P(z < 1.28) = 90\%$
3. $x = 500 + 1.28(100) = 628$



6-D The Central Limit Theorem

The LAW OF LARGE NUMBERS states that the larger a sample is, the more accurately statistics calculated from the sample tend to represent the actual population parameters. As a result, larger samples are more likely to have results that turn out about as expected.

❶ Use the law of large numbers to determine whether a statistic is more likely to fall within a given range for a small sample or for a large sample.

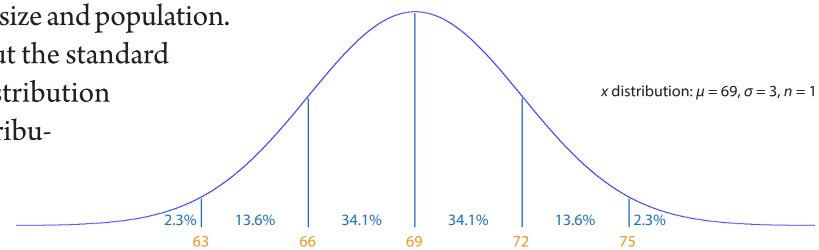
1. Identify whether or not the range includes the population parameter.
2. If so, the larger sample is more likely to include the stated statistic. Otherwise, the smaller sample is more likely.

❶ Is Trump more likely to have a favorable approval rating in a random sample of 20 California voters or in a random sample of 200 California voters?

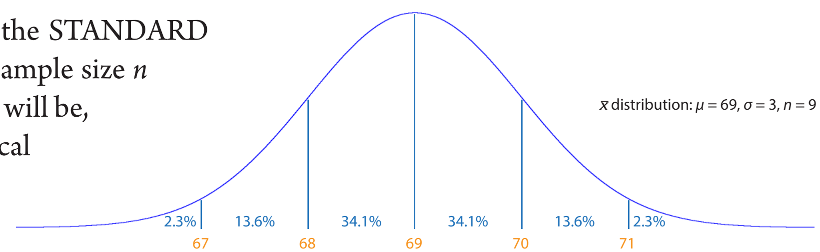
1. Trump's favorability rating is low in California, perhaps $p = 30\%$. It is clearly not in the range $\hat{p} > 50\%$.
2. He is not likely to have a favorable rating in either sample, but it could happen by coincidence in either sample. Coincidences are more likely in smaller samples, which in this case is the sample of 20 California voters.

A SAMPLING Distribution is the distribution of a statistic, such as $\bar{x}$, for a given sample size and population.

The mean of an $\bar{x}$ distribution is the same as the mean of the original x distribution, but the standard deviation is smaller. For example, the normal curve at right from 6-A, shows the distribution of heights of heights (x) of individual men, while the curve below it shows the distribution of averages heights of groups of nine men. Both distributions have an average of $\bar{x} = 69$. However, due to the law of large numbers, the distribution with the larger samples (9 instead of 1) will have values closer to the expected average of 69, thus reducing the standard deviation.



The standard deviation of a sampling distribution is labeled $\sigma_{\bar{x}}$ (or $s_{\bar{x}}$) and is called the STANDARD ERROR of the Mean. The CENTRAL LIMIT Theorem states that the larger the sample size n is, and the more normal the x distribution is, the closer to normal the $\bar{x}$ distribution will be, with a standard error of $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, especially if the distribution is somewhat symmetrical and $n \geq 30$.



For sampling distributions, use $\frac{\sigma}{\sqrt{n}}$ instead of σ in the z formulas: $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ and $x = \mu + z\frac{\sigma}{\sqrt{n}}$.

❷ Calculate sampling probabilities or frequencies of sample means, or calculate sample means from percentiles.

1. Follow the appropriate steps in 6-B or 6-C, except replace σ with $\frac{\sigma}{\sqrt{n}}$ (or s with $\frac{s}{\sqrt{n}}$).

❷ Four random students take a test that has normally distriubted scores with $\mu = 140$ and $\sigma = 55$. What is the probability that their average will be between 100 and 190?

1. $z_1 = \frac{100 - 139}{48 / \sqrt{4}} \approx -1.62$ $\qquad z_2 = \frac{190 - 139}{48 / \sqrt{4}} \approx 2.12$
2. $P(100 < \bar{x} < 190) = P(-1.62 < z < 2.12) \approx .9830 - .0526 \approx 93.0\%$