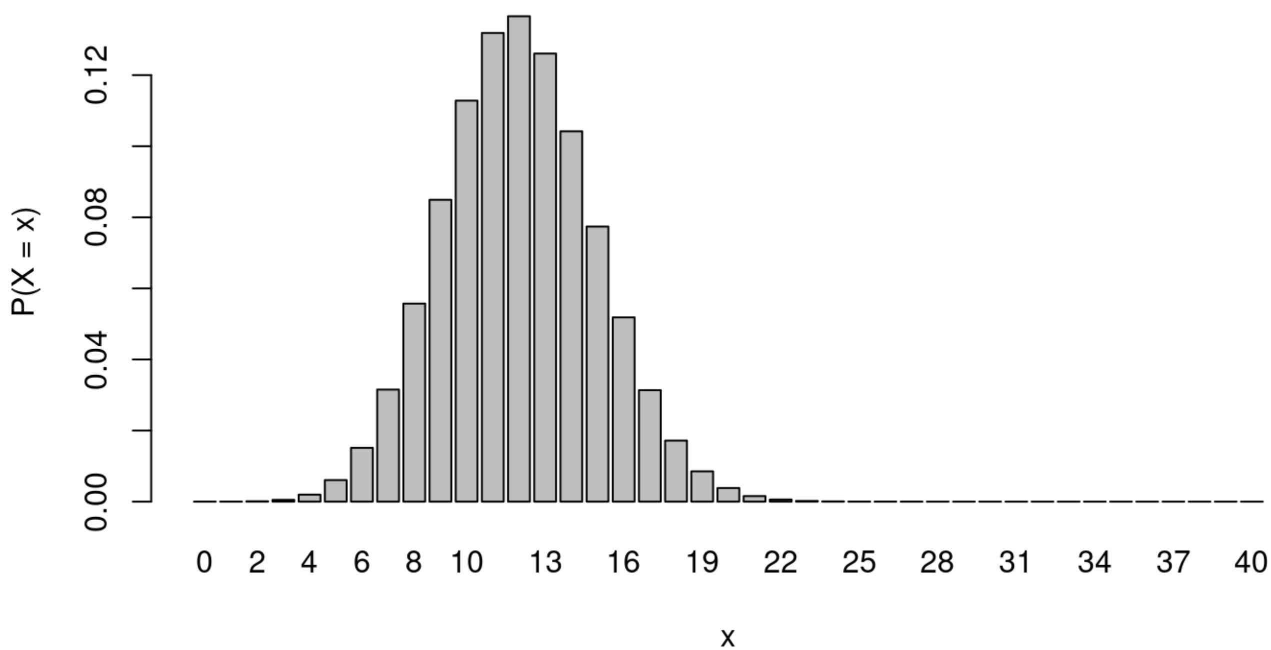
MATH 1051H-A: Lecture #10

Binomial & Normal - Recap

Binomial Distribution

The binomial distribution is usually visualized using a barplot:



Binomial - Formula

The binomial distribution has a complicated formula:

$$P(x) = \frac{n!}{(n-x)!x!} \cdot p^x \cdot (1-p)^{n-x}$$

where

- n - the fixed number of trials/experiments
- x - a specific number of successes we are interested in (must be between 0 and n, obviously)
- p - probability of success in any one trial/experiment
- $P(X)$ - the probability of getting exactly x successes in n trials

Binomial - Computing Probabilities

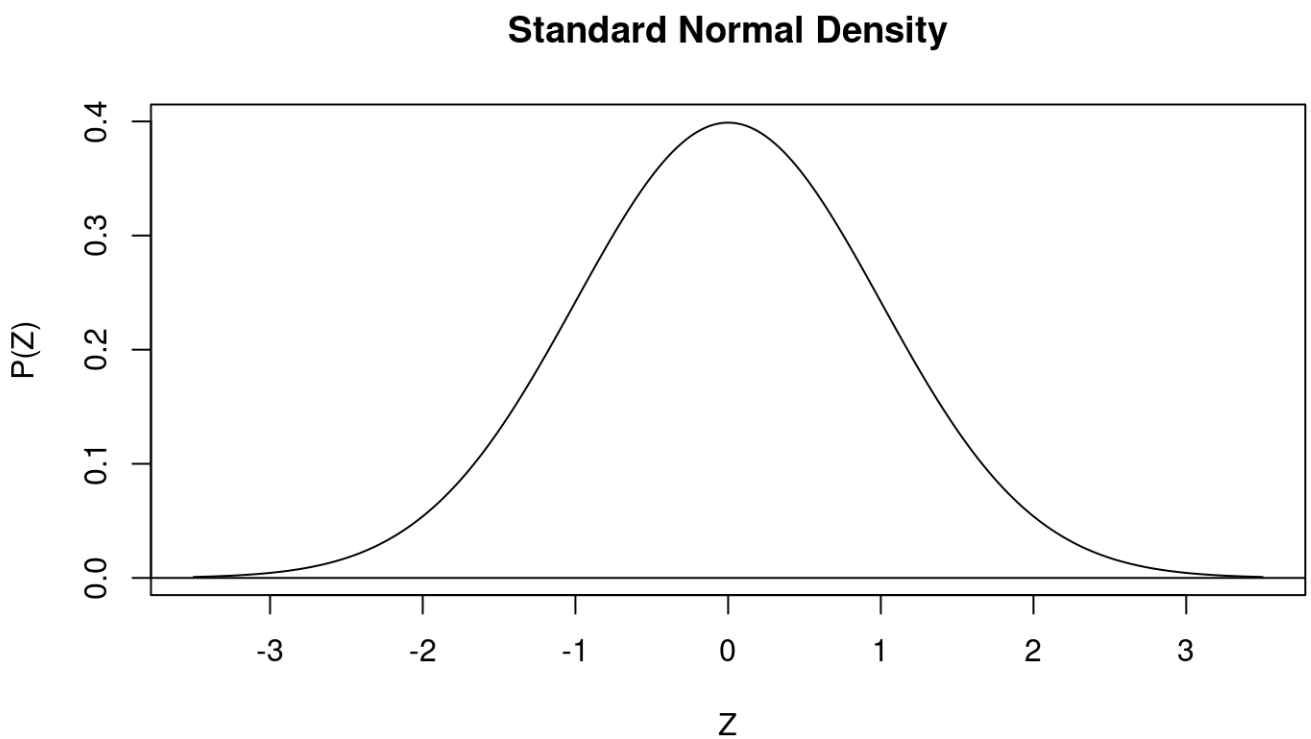
We never use the previous formula - we don't use calculators, so what's the point?

Instead, we use the following relationships:

- $P(X = x) \Leftrightarrow$ dbinom(q = x, size = n, prob = p)
- $P(X \leq x) \Leftrightarrow$ pbinom(q = x, size = n, prob = p, lower.tail = TRUE)
- $P(X < x) \Leftrightarrow$ pbinom(q = x - 1, size = n, prob = p, lower.tail = TRUE)
- $P(X \geq x) \Leftrightarrow$ pbinom(q = x - 1, size = n, prob = p, lower.tail = FALSE)
- $P(X > x) \Leftrightarrow$ pbinom(q = x, size = n, prob = p, lower.tail = FALSE)

Normal Distribution

The normal distribution is **continuous**, and the single most used thing in statistics. It looks like this:



Normal - Formula 1

The normal density formula is complicated, and we will never use it. However, the **Z-score** formula gets used often. Recall its purpose: it takes a normally distributed random variable, X, with mean μ and variance σ^2, and converts it into a *standard normal* random variable Z:

$$Z = \frac{X - \mu}{\sigma}.$$

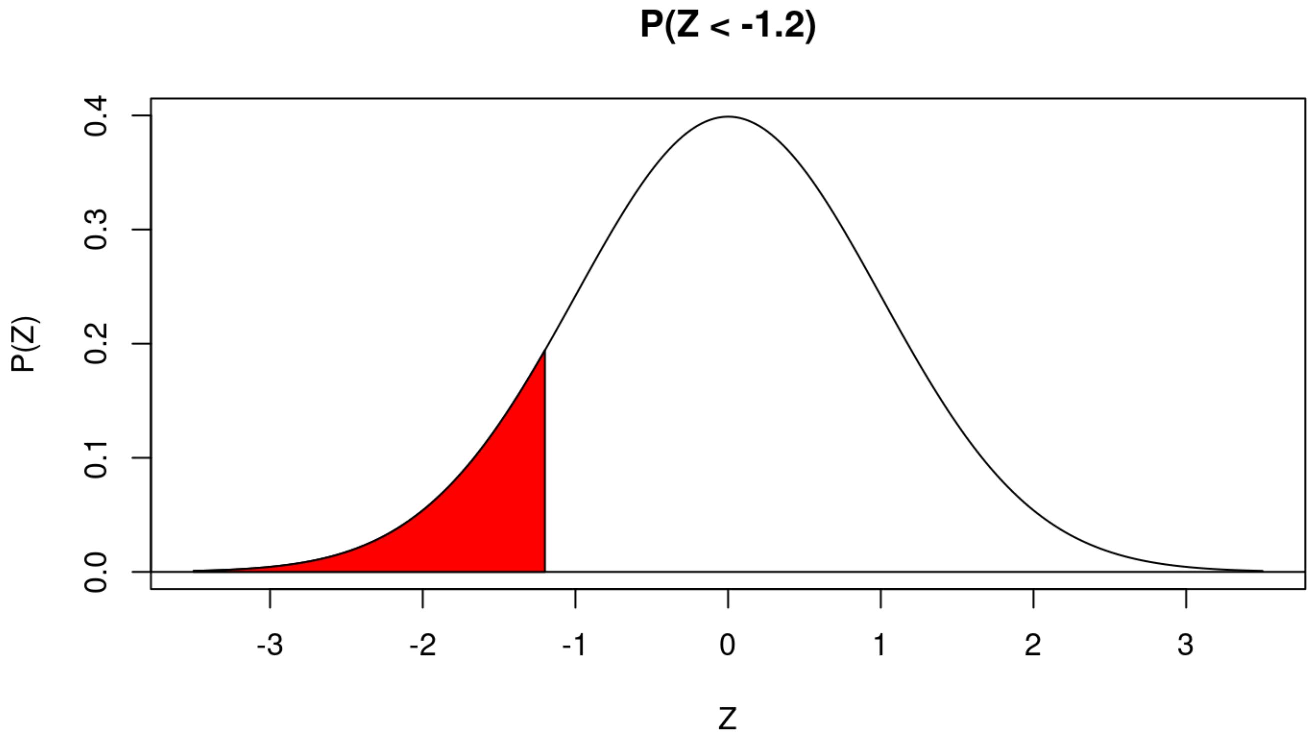
This Z is a reserved variable: any time we use Z in statistics, it means this Z in particular!

Normal - Computing Probabilities (from Z-score / quantile)

Given a Z-score z_0 (which is a quantile), we can compute the probability under the normal curve as:

- $P(Z \leq z_0) \;\; \Leftrightarrow$ pnorm(z0, mean = 0, sd = 1, lower.tail = TRUE)
- $P(Z < z_0) \;\; \Leftrightarrow$ pnorm(z0, mean = 0, sd = 1, lower.tail = TRUE)
- $P(Z \geq z_0) \;\; \Leftrightarrow$ pnorm(z0, mean = 0, sd = 1, lower.tail = FALSE)
- $P(Z > z_0) \;\; \Leftrightarrow$ pnorm(z0, mean = 0, sd = 1, lower.tail = FALSE)

Normal - Computing Probabilities - Picture



Normal - Computing Quantiles (from probabilities)

Another problem we often do for normals (and one we will need later this semester!) is to find the quantile that goes with a probability. So, if we are given a description of a probability statement, with an unknown z_0, and the probability answer, we have to invert it.

- $P(Z \leq z_0) = p \;\Leftrightarrow\;$ qnorm(p, mean = 0, sd = 1, lower.tail = TRUE)
- $P(Z \geq z_0) = p \;\Leftrightarrow\;$ qnorm(p, mean = 0, sd = 1, lower.tail = FALSE)

Normal Approximation to Binomial

When $np \geq 10$ and $n(1-p) \geq 10$, we can approximate a binomial distribution by a normal. This is useful to avoid having to compute many many binomial probabilities, and as a connection between the two.

The conversion is:

$$Z = \frac{X_{\text{correct}} - np}{\sqrt{np(1-p)}}$$

where X_{correct} is the corrected binomial value - either $X + 0.5$ or $X - 0.5$, depending on the setup and picture. This is made more clear by examples, which we will do at the end of today's lecture.

Examples

Example 1

Suppose we toss a fair coin 20 times. What is the probability of getting between 9 and 11 heads?

Example 2

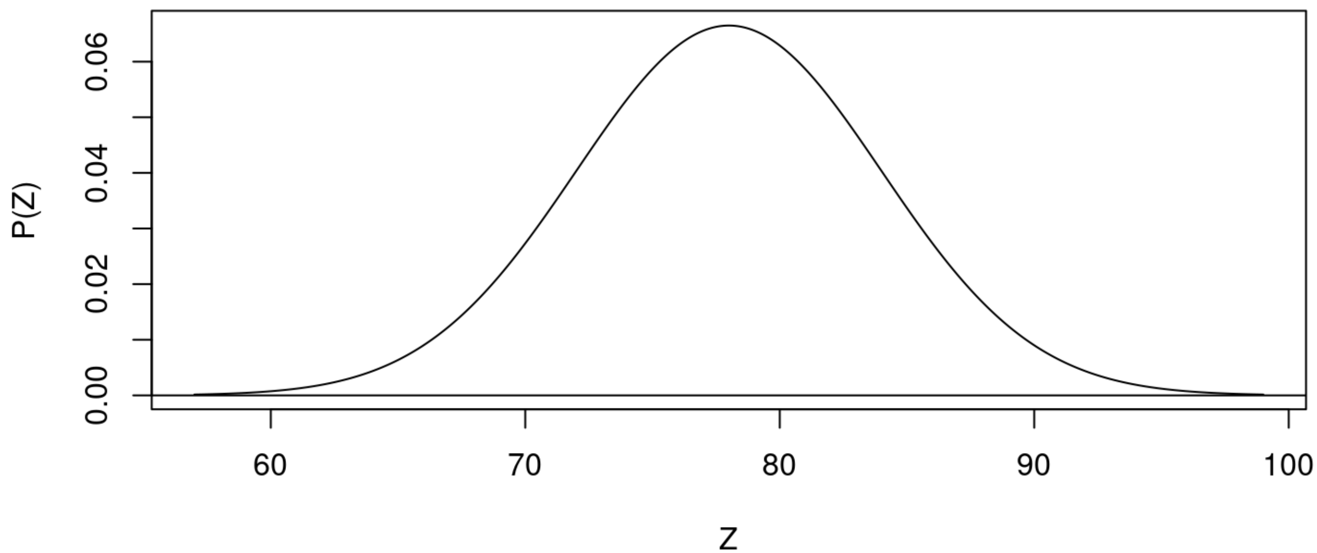
In a particular program at Trent, 60% of students are men and 40% are women. In a random sample of 50 students what is the probability that more than half are women?

Example 3

Skip the Dishes finds that 70% of people order through them and ask for Chinese food, and 30% for Italian food. Last week 426 orders were made through the local Peterborough, unit. Find the probability that at least 200 orders were for Italian food.

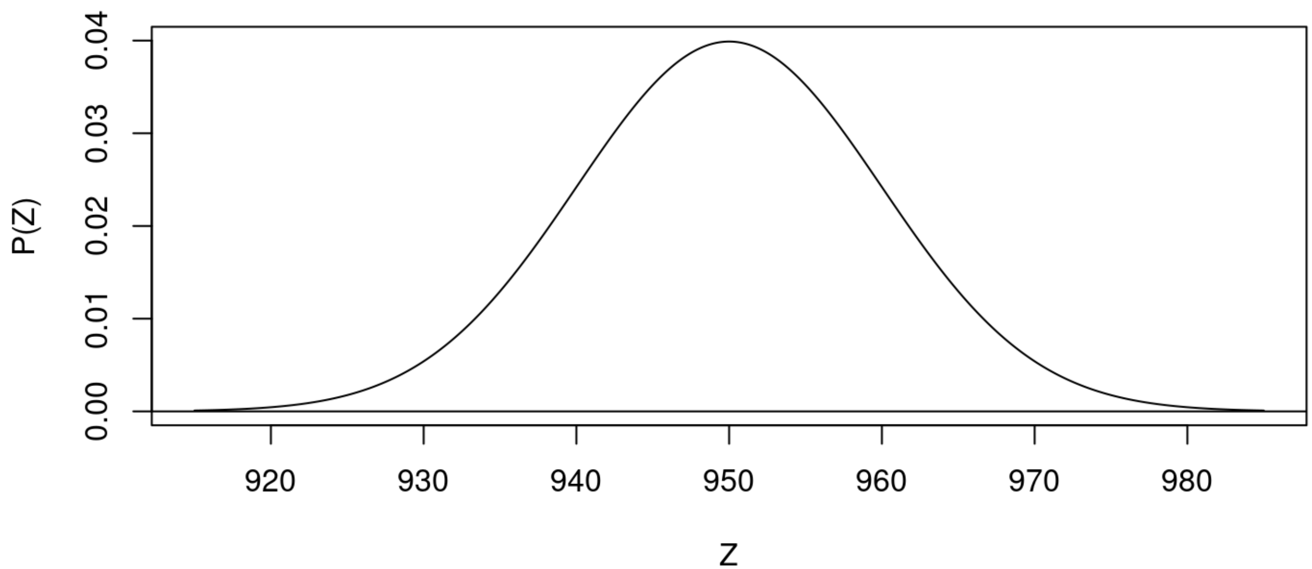
Example 4

Scores on a final examination are assumed to be normally distributed with mean 78 and variance 36. What is the probability that a person taking the examination scores higher than a 72?



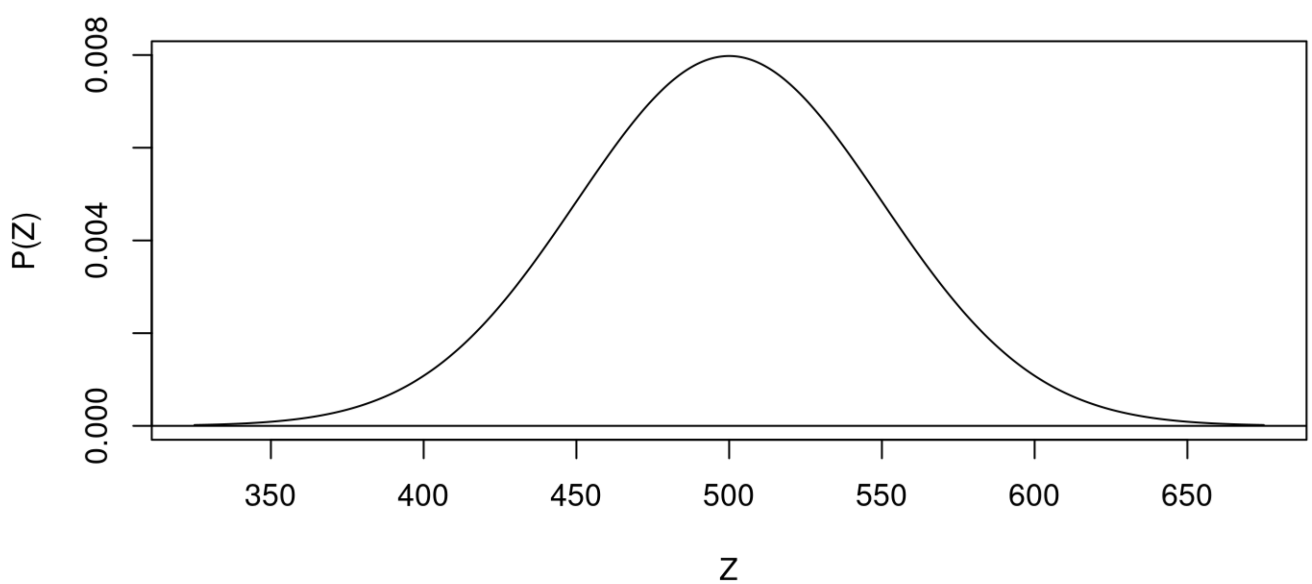
Example 5

The width of bolts of fabric is normally distributed with mean 950mm and standard deviation 10mm. What is the probability that a randomly chosen bolt has width between 947 and 958mm?



Example 6

A manufacturing plant utilizes 3000 light bulbs whose length of life is normally distributed with mean 500 hours and standard deviation 50 hours. To minimize the number of bulbs that burn out during operating hours, all bulbs are to be replaced after a given period of operation. How often should the bulbs be replaced if we want no more than 1% of bulbs to burn out between replacement periods (that is, before the bulb would have been replaced)?

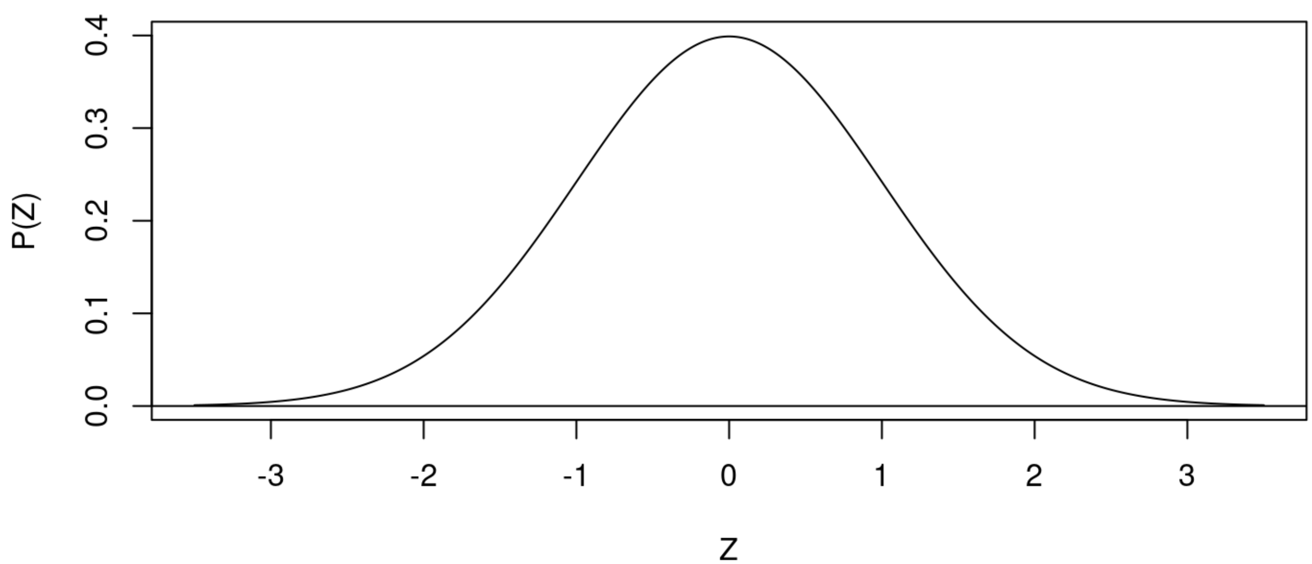


Example 7

A soft-drink machine can be regulated so that it discharges an average of μ ounces per cup. If the ounces of fill are normally distributed with standard deviation 0.3 ounce, find the setting for μ so that 8-ounce cups will only overflow 1% of the time.

Example 8

A machining operation produces bearings with diameters that are normally distributed with mean 3.0005 inches and standard deviation 0.0010 inch. Specifications require the bearing diameters to lie in the interval 3.000 ± 0.0020 inches. Those outside the interval are considered scrap, and must be remachined. With the existing production, what percentage of bearings would be scrap?



Example 9

A manufacturer of floor wax has developed two new brands, A and B, which they wish to subject to homeowners' evaluation to determine which of the two is considered superior. Both waxes, A and B, are applied to floor surfaces in each of 65 homes. Assume that there is actually no difference in the composition of the brands. What is the probability that 40 or more homeowners would state a preference for Brand A?

Compute this solution using both the binomial approach, and the normal approximation approach, if possible.

Example 10

A missile protection system consists of n radar sets operating independently, each with a probability of 0.9 of detecting a missile entering a zone covered by all units. If $n = 110$, use the normal approximation to the binomial to compute the probability that at least 50% of the radar sets will detect a given missile.