Stat 145, Thu 18-Mar-2021 -- Thu 18-Mar-2021
Biostatistics
Spring 2021

Thursday, March 18th 2021

Wk 7, Th
Topic:: Central Limit Theorem
Read:: Lock5 5.2

Answers to yesterday's worksheet

Variables can
- have an association, or
- not have an association.

negation of associated

We also talk about independent variables, which is roughly the same as

Examples:

i.i.d. = independent and identically distributed

1. If we draw twice from a bag and take
   X = 1st outcome
   Y = 2nd outcome
   then X and Y are
   i) independent if sampling "with replacement"
      call this an i.i.d. random sample of size 2

   ii) approximately independent if the composition of the bag is little changed after the first draw

2. If we draw n times from a bag and take

   sample size

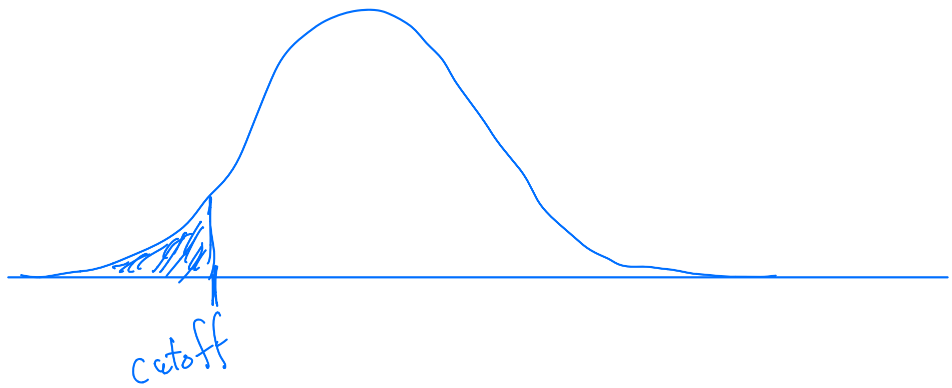
   X_1 = 1st outcome
   X_2 = 2nd outcome
   .
   .
   .
   X_n = nth outcome

   X_i = i^{th} outcome

   the X_i are
   i) independent if sampling "with replacement"

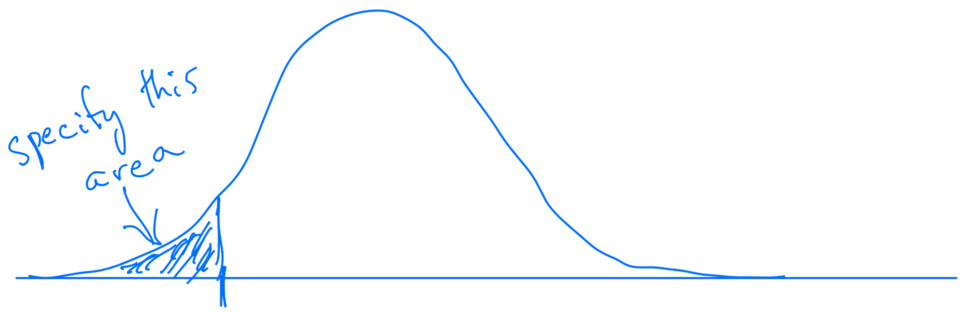
pnorm(cutoff, mean = ______, sd = ______)

results in shaded area



qnorm(desired left-tail area, mean = ______, sd = ______)

results in value of cutoff



New terms:

- independence of variables
- iid random sample
- random variable

call this an i.i.d. random sample of size n

ii) approximately independent if the composition of the bag is
little changed after by the draws

$X - Y \sim \text{Norm}(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

A random variable X is one that is numeric for each case
- sex: F/M we think of as categorical
- X(case) = 0 if case=female, 1 if case=male is a random variable

$X \sim \text{Norm}(\mu_1, \sigma_1)$

$Y \sim \text{Norm}(\mu_2, \sigma_2)$

Some facts about independent normal random variables
- If X and Y are independent normal random variables, with
  X ~ Norm(mu_1, sigma_1)
  Y ~ Norm(mu_2, sigma_2)

  $X + Y \sim \text{Norm}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

  then X+Y (their sum) is ~ Norm(mu_1 + mu_2, sqrt(sigma_1^2 + sigma_2^2)
  then X+Y (their difference) is ~ Norm(mu_1-mu_2, sqrt(sigma_1^2 + sigma_2^2)

  Ex.: Suppose Ray and Joan are bowlers. Their scores have normal dists
  R ~ Norm(142, 17)
  J ~ Norm(138, 22)

  $R + J \sim \text{Norm}(280, \sqrt{17^2 + 22^2})$

  How likely is it for them, in one game, to have a combined score > 350?

- If we draw an i.i.d. random sample of size n, each X_i ~ Norm(mu, sigma),
  then the
  sum = X_1 + ... + X_n is Norm(n mu, sigma sqrt(n))
  avg = (X_1 + ... + X_n) / n is Norm(mu, sigma / sqrt(n))

  If each $X_i \sim \text{Norm}(\mu, \sigma)$
  then
  sum $\sim \text{Norm}(n\mu, \sigma\sqrt{n})$
  mean $\sim \text{Norm}(\mu, \sigma/\sqrt{n})$

Central Limit Theorem tomorrow

Suppose a random sample of size n is drawn from the population either
- with replacement (so it is i.i.d.), or
- with n smaller than 10% of the full population.

$\sqrt{\underbrace{\sigma^2 + \sigma^2 + \sigma^2 + \cdots \sigma^2}_{n \text{ times}}} = \sigma\sqrt{n}$

If the variable of interest is quantitative and n is large enough, then
  the sum X_1 + ... + X_n is approximately normal
  the mean (X_1 + ... + X_n)/n is approximately normal
If the variable of interest is binary categorical and n is large enough,
then the sample proportion has approximately a normal distribution.

Say Ray $\sim$ Norm$(142, 17)$ and bowls 5 games

His composite score $X_1 + X_2 + \cdots + X_5 \sim \text{Norm}\left(710, \underbrace{\sqrt{17^2+17^2+17^2+17^2+17^2}}_{= 17\sqrt{5}}\right)$

```
Explorations
 - links from website
 - apps at
     https://shiny.calvin.edu:3838/scofield/samplingDists/
     http://www.lock5stat.com/StatKey/bootstrap_1_quant/bootstrap_1_quant.html
```