1 Review

- Transform $N(\mu, \sigma)$ to $N(0,1)$. If $X \sim N(\mu, \sigma)$ then $Z$ becomes standard normal where,

$$Z = \frac{X - \mu}{\sigma}.$$  (1)

eg) If $X \sim N(\mu = -5, \sigma = 3)$ then $Z$ becomes standard normal where,

$$Z = \frac{X - (-5)}{3}.$$  (2)

- Return your mid 1 at the end of the lecture. 14/35 problems from lecture/R (40%). 21/35 from HW/suggested (60%). Possible Max=105, min=0. If your score =84/105, your mid1 contribution to the total grade is

$$\frac{84}{105} \cdot 0.20 = 16 \text{ points}.$$  (3)

- Remember to return after- Mid 1 feedback by email to get 1 point. These are explained in the syllabus.

DUE: Oct 15 (Fri) midnight

2 Ch6.6 Binomial Approximation

What does “approximation” mean? When you have $49.89, you just say “I have $50.” $49.89 is different from $50. But in reality, it does not matter. Same idea here. Binomial distribution is different from Normal distribution, but you can think they are the same under some conditions.

- When $X \sim Bin(n, p)$ and $np \geq 15$, $n(1-p) \geq 15$, then you can approximate $X \sim N(np, np(1-p))$

In other words, new random variable,

$$Z = \frac{X - np}{\sqrt{np(1-p)}},$$

is approximately $N(0,1)$.

- Remember the fact: $X \sim Binomial(n, p)$ (or Bin(n,p)), then $E[X] = np$, $Var[X] = np(1-p)$.

- If one of $np$ or $n(1-p)$ is less than 15, then binomial probability histogram is not symmetry, skewed to right or left. Then it is not well approximated by the normal distribution which is symmetry. The condition, $np \geq 15$, $n(1-p) \geq 15$ guarantees binomial has almost bell-shaped, symmetry distribution.

eg1) The unemployment rate in a city is 7.9%. A sample of 300 persons is selected from the labor force. Approximate the probability that

(Q1) Fewer than 18 unemployed persons are in the sample.

(Q2) More than 30 unemployed persons are in the sample.

(A1) Let $X =$ number of unemployed persons among 300. Then $X \sim Bin(n = 300, p = 0.079)$. We can compute it by

This is correct formula but takes long time to calculate. Hence we try to use Normal approximation. To do that, we need to check 2 conditions.
eg2) Let the number of successes $X$ have a binomial distribution with $n = 200$ and $p = 0.65$. Use the normal distribution to approximate the probability of
(a) $X \leq 150$.
(b) $137 \leq X \leq 152$.

3 Ch7 Sampling distributions

• **Parameter** (p293) : a numerical feature of a population.
  eg1) If $X \sim Binomial(n,p)$, then $(n,p)$ are parameters. (We can just say “n and p are parameters”)
  eg2) If $X \sim Bernoulli$ with success rate =p, then
  eg3) If $X \sim Bernoulli$ with success rate =s, then

• **Statistic** (p293) : a numerical value function of the sample observations.
  eg1) Suppose you cast a dice 4 times, and set $X_i = i$-th result. Then
  $V_1 = X_1$
  $V_2 = (X_1 + X_2)/5$
  $V_3 = (X_1 + X_2)/X_3$
  $V_4 = (X_1 + X_3 + 4) * 2$
  $V_5 = (X_1 + X_2 + X_3 + X_4)/4$
  $V_6 = (X_1 - V_5)^2 + (X_2 - V_5)^2 + (X_3 - V_5)^2 + (X_4 - V_5)^2$
  are all statistics. Usually, $V_5$ is called
  and $V_6$ is called

• **Sampling Distribution** (p294): the probability distribution of a statistic.
  eg1) Flip a fair coin (1 or 0) twice. Define $X_1$ = 1st result, $X_2$=2nd result. Then, find sampling distribution of $V = X_1 + X_2$
  ANS: Remember,
  $P(X_1 = 0) = P(X_1 = 1) = 0.5$
  $P(X_2 = 0) = P(X_2 = 1) = 0.5$
  Then possible $V =$?

eg4) If $X \sim N(\mu, \sigma^2)$, then

eg1) Flip an **unfair** coin (1 or 0) twice. Success rate=0.7
Define $X_1$ = 1st result, $X_2$=2nd result. Then, find sampling
distribution of $V = X_1 + X_2$
ANS: Remember,
$P(X_1 = 0) = 0.3, P(X_1 = 1) = 0.7$
$P(X_2 = 0) = 0.3, P(X_2 = 1) = 0.7$
Then possible $V =$?
• The observations \(X_1, X_2, \ldots, X_n\) are called i.i.d or random sample of size \(n\) from the population distribution, if they result from independent selections and each observation has the same distribution as the population. (i.i.d means independent and identically distributed.)

• The Distribution of sample mean \(\bar{X}\) (i.i.d observation). Suppose \(E X_1 = \mu\) and \(Var X_1 = \sigma^2\), then

\[
E[\bar{X}] = \mu,
\]

\[
Var[\bar{X}] = \frac{\sigma^2}{n},
\]

\[
SD[\bar{X}] = \frac{\sigma}{\sqrt{n}}.
\]

eg1) Flip a fair coin 20 times (0 or 1). Define \(X_i = \) ith result. Remember,

\[
E[X_i] = 0.5,
\]

\[
Var[X_i] = 0.25,
\]

\[
E[\bar{X}] = 0.5,
\]

\[
Var[\bar{X}] = \frac{0.25}{20},
\]

\[
SD[\bar{X}] = \frac{0.5}{\sqrt{20}}
\]

4 Ch7.3 \(\bar{X}\) (Sample mean) distribution

• case1: \(X_1, X_2, \ldots, X_n \sim N(\mu, \sigma)\). Then,

\[
\frac{\sum_{i=1}^{n} X_i}{n} = \bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}),
\]

or, equivalently,

\[
\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1).
\]

Note: This is exact.

• case2: (Central Limit Theorem (CLT)) \(X_1, X_2, \ldots, X_n \sim unknown\) AND \(n \geq 30\). Then,

\[
\frac{\sum_{i=1}^{n} X_i}{n} = \bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}),
\]

or, equivalently,

\[
\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1).
\]

Note: This is an approximation.

We need these equations (approximation) soon. For example, Testing of statistical hypotheses (Ch8, p321) More details coming soon.

eg1)(p307) Example 5

Population mean = \(\mu = 82\), \(\sigma = 12\) are given.

Q1) If 64 samples are selected randomly, what is the probability that the sample mean will lie between 80.8 and 83.2?

A1) \(64 \geq 30\), use Central Limit Theorem (CLT).

\[
\mu = 82, \sigma = 12
\]

\[
P(80.8 < \bar{X} < 83.2) =
\]
Q2) If 100 samples are selected randomly, what is the probability that the sample mean will lie between 80.8 and 83.2?

\[ \mu = \]
\[ \sigma / \sqrt{n} = \]
\[ P(80.8 < \bar{X} < 83.2) = \]