HW6 Solution

- (HW6_1)

  (a) Each trail has only two outcomes: success or failure. $p = P(success)$ is the same for every trail. Trails are independent.

  (b) If we take a sample of size n from a much larger population (of size N) in which a proportion p have a characteristic of interest, then the distribution of X, the number in the sample with that characteristic, is approximately $Binominal(n, p)$. (from slide 83, class notes, week 6).

  (c) X = number of successes in 10 trails. $P(success) = 0.3$.

$$P(X = x) = \binom{10}{x} 0.3^x (1 - 0.3)^{10-x}, x = 0, 1, \cdots, 10.$$

  (d)

$$\begin{aligned} P(A) &= P(1 \le X < 4) \\ &= P(X = 1) + P(X = 2) + P(X = 3) \\ &= \binom{10}{1} 0.3^1 0.7^{10-1} + \binom{10}{2} 0.3^2 0.7^{10-2} + \binom{10}{3} 0.3^3 0.7^{10-3} \\ &= 0.62. \end{aligned}$$

$$\begin{aligned} P(B) &= P(4 < X \le 10) \\ &= P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) - P(X = 4) \\ &= 0.15. \end{aligned}$$

  So, $P(A)$ is bigger.

- (HW6_2) Let Y = outcome of a single octahedral (8 face) die. $E(Y) = \sum_{x=1}^{8} x \times P(X = x) = 4.5$, And $Var(Y) = \frac{1}{N-1} \sum_{x=1}^{8} (x - E(Y))^2 = (approx.)6$.

  Then, if X = outcome of 6 dice rolled twice (total of 12 independent observations from the same distribution), we have $E(X) = E(\sum_{k=1}^{12} Y_k) = \sum_{k=1}^{12} E(Y_k) = 12 \times 4.5 = 54$.

  $Var(X) = Var(\sum_{k=1}^{12} Y_k) = \sum_{k=1}^{12} Var(Y_k) = 12 \times 6 = 72$.

$SD(Y) = 6\sqrt{2}$.

Let, now, $\overline{Y}$ be the average of 11 such experiments. Then $\overline{Y} = \frac{1}{11}\sum_{k=1}^{11} X_k$. $E(\overline{Y}) = E(X) = 54$ and $Var(\overline{Y}) = Var(\frac{1}{11}\sum_{k=1}^{11} X_k) = \frac{1}{11^2}\sum_{k=1}^{11} Var(X_k) = \frac{1}{11^2} \times 11 \times 72$.

Lastly, $SD(\overline{Y}) = \sqrt{Var(\overline{Y})} = \sqrt{72/11}$.

The distribution of sample average, $\bar{Y}$ is normal, $\overline{Y} \sim N(54, 72/11)$.

- (HW6_3)

  (a) Let $\bar{X}_1$, S_1, and n_2 be the mean, SD and the sample size of the mnemonic group, respectively.

$$\bar{X}_1 = \frac{\sum X_{1i}}{n_1} = 14.1$$
$$S_1 = \sqrt{\frac{\sum(X_{1i} - \bar{X}_1)^2}{n_1 - 1}} = 2.47$$

  Let $\bar{X}_2$, S_2, and n_2 be the mean, SD and the sample size of the control group, respectively.

$$\bar{X}_2 = \frac{\sum X_{2i}}{n_2} = 9.22$$
$$S_2 = \sqrt{\frac{\sum(X_{2i} - \bar{X}_2)^2}{n_2 - 1}} = 2.90$$

  (b) A 95% C.I. for the difference in the mean number of words recalled between the normal and mnemonic instruction methods is

$$\bar{X}_1 - \bar{X}_2 \pm t\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$= 14.1 - 9.22 \pm 2.11\sqrt{\frac{2.47^2}{20} + \frac{2.9^2}{18}}$$
$$= [3.03, 6.73].$$

  where $df = min\{20 - 1, 18 - 1\} = 17$, $t_{17,0.25} = 2.11$.
  We have 95% confident that the difference in the mean number of words recalled between the normal and mnemonic instruction methods is between 3.03 to 6.73.

(c) To approximately halve the size of the CI we need to quadruple the (largest) sample size. That is we'll need about 20x4 = 80 observations in every group.

(d) The C.I. above may or may not contain the true difference in means between the two groups. We have no knowledge about the true population differences between the control and mnemonic groups. We only have a sample *estimate of this difference. Hence, with the given knowledge, we can not affirmatively answer, or reject the fact that the true population difference is in the interval. The only statement we can make is that we are 95the true groups difference in memory performance is within the range of our CI(95%).*