1. Consider a the distribution with the following probability mass function

x	30	35	40
p(x)	0.2	0.3	0.5

Let X_1, X_2 be two independent random sample from this distribution, and $T = X_1 - X_2$.

(a) (10 points) Compute the expected value and the standard deviation of X_1.

$E[X_1] = 30 \times 0.2 + 35 \times 0.3 + 40 \times 0.5 = 36.5$

$E[X_1^2] = 30^2 \times 0.2 + 35^2 \times 0.3 + 40^2 \times 0.5 = 1347.5$

$Var[X_1] = E[X_1^2] - (E[X_1])^2 = 1347.5 - 36.5^2 = 15.25$

$\sigma(X_1) = \sqrt{Var(X_1)} = \sqrt{15.25}$

(b) (10 points) Compute the expected value and the standard deviation of T.

$E[X_2] = E[X_1] = 36.5 \quad Var(X_2) = Var(X_1) = 15.25$

$E[T] = E[X_1 - X_2] = E[X_1] - E[X_2] = 36.5 - 36.5 = 0$

$Var[T] = Var[X_1 - X_2] = Var[1 \times X_1 + (-1) \times X_2]$

$= 1^2\, Var(X_1) + (-1)^2\, Var(X_2)$

$= 15.25 + 15.25 = 30.5$

$\sigma(T) = \sqrt{Var(T)} = \sqrt{30.5}$

(c) (20 points) Compute

$$P[T = 5]$$

$P[T=5] = P[X_1 = 35 \text{ and } X_2 = 30] + P[X_1 = 40 \text{ and } X_2 = 35]$

$= P[X_1 = 35]\, P[X_2 = 30] + P[X_1 = 40]\, P[X_2 = 35]$

(since X_1, X_2 are independent)

$= 0.3 \times 0.2 + 0.5 \times 0.3 = 0.21$

2. (a) (20 points) Let Y be a continuous random variable with the following probability density function

$$f(y) = \begin{cases} \frac{3}{7}y^2, & \text{for } \ y \in [1,2] \\ 0 & \text{otherwise} \end{cases}$$

Compute $P[1.25 \le Y \le 1.75]$.

$$P[1.25 \le Y \le 1.75] = \int_{1.25}^{1.75} \frac{3}{7} y^2 \, dy$$

$$= \left. \frac{y^3}{7} \right|_{1.25}^{1.75}$$

$$\approx \frac{1.75^3}{7} - \frac{1.25^3}{7} = 0.4866$$

(b) (20 points) Assume that the joint probability of X (receive values 0, 1) and Y (receives values 0, 1, 2) is represented by the following table

X \ Y	0	1	2
0	1/18	1/18	4/18
1	2/18	5/18	5/18

Are X and Y independent?

Marginal distribution

X	0	1
p(X)	$\frac{1}{3}$	$\frac{2}{3}$

Y	0	1	2
	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

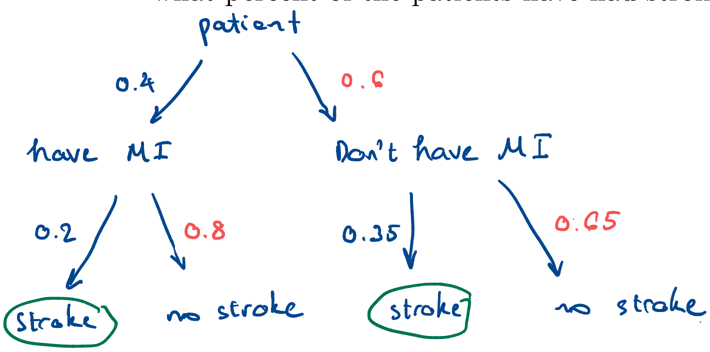
$$\frac{1}{18} = P(0,0) = P(X=0) \times P(Y=0) = \frac{1}{3} \times \frac{1}{6}$$

$$\frac{1}{18} = P(0,1) \neq P(X=0) \times P(Y=1) = \frac{1}{3} \times \frac{1}{3}$$

Since $P(0,1) \neq P(X=0) \times P(Y=1)$,

X and Y are not independent

3. (a) (20 points) Of the patients in a hospital, 20% of those with, and 35% of those without myocardial infarction have had strokes. If 40% of the patients have had myocardial infarction, what percent of the patients have had strokes?



$$P[\text{stroke}] = P[\text{have MI}]\,P[\text{stoke} \mid \text{have MI}] + P[\text{don't have MI}]\,P[\text{stroke} \mid \text{don't have MI}]$$

$$= 0.4 \times 0.2 + 0.6 \times 0.35 = 0.29$$

(b) (20 points) In a region, the correlation coefficient between corn yield and peanut yield (planted in the same soil, in MT/ha) is 0.9. We also know that

- The mean and the standard deviation of corn yield is 3.2 and 2.4, respectively.
- The mean and the standard deviation of peanut yield is 1.8 and 0.72, respectively.

Using this information, predict the expected corn yield of a filed with peanut yield of 1.76.

x : peanut

y : corn

① $$\hat{x}_0 = \frac{x_0 - \text{mean}(\{x\})}{\text{std}(\{x\})} = \frac{1.76 - 1.8}{0.72} = \frac{-0.04}{0.72} = \frac{-1}{18}$$

② $$y_0 = \text{std}(y) \cdot r \cdot \hat{x}_0 + \text{mean}(y)$$

$$= 2.4 \times 0.9 \times \left(\frac{-1}{18}\right) + 3.2$$

$$= 3.08$$

4. Urban storm water can be contaminated by many sources, including discarded batteries. When ruptured, these batteries release metals of environmental significance. The paper “Urban Battery Litter” (J. Environ. Engr., 2009: 46–57) presented summary data for characteristics of a variety of batteries found in urban areas around Cleveland. A sample of 51 Panasonic AAA batteries gave a sample mean zinc mass of 2.06 (g) and a sample standard deviation of 0.141 (g).

$n = 51 \quad \bar{x} = 2.06 \quad s = 0.141$

$n > 40$ ✓

Normal ✗

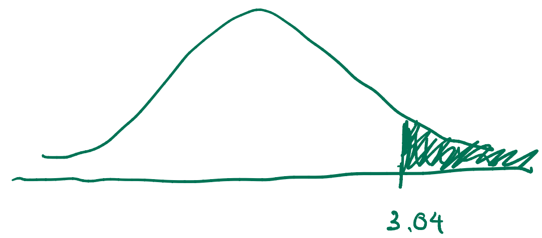
σ known ✗

(a) (20 points) With a significance level of $\alpha = 0.001$, does this data provide compelling evidence for concluding that the population mean zinc mass exceeds 2.0 g.?

$H_0 : \mu = 2.0$

$H_a : \mu > 2.0$

$z = \dfrac{\bar{x} - 2.0}{s/\sqrt{n}} = 3.04$



$p\text{-value} = 1 - \phi(3.04) = 1 - 0.9988$
$= 0.0012 > \alpha = 0.001$

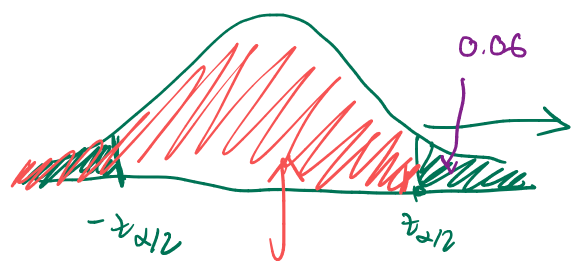
Conclusion: Fail to reject H_0

(b) (20 points) Construct a 88% confidence interval of the population mean zinc mass.

$\alpha = 0.12$

$\alpha/2 = 0.06 \rightarrow$ look up 0.94

$z_{\alpha/2} = 1.55$



$1 - 0.06 = 0.94$

88% confidence interval:

$$\left(\bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}} \,,\, \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}} \right)$$

$(2.03, 2.09)$

5. A study was carried out in an attempt to improve student performance in a low-level university mathematics course. The study involved assigning the students to sections based on odd or even Social Security number. Half of the sections were taught traditionally, whereas the other half were taught in a way that hopefully would keep the students involved.

   The final exam scores for the 79 students taught traditionally (the control group) and for the 85 students taught with more involvement (the experimental group) are summarized as follows:

Group	Sample Size	Sample Mean	Sample SD
Control	79	23.87	11.60
Experimental	85	27.34	8.85

   (a) (20 points) With a significance level of $\alpha = 0.05$, does this data provide compelling evidence that that the experimental method of instruction is an improvement?

μ_1 : mean score experimental

μ_2 : mean score control

H_0 : $\mu_1 - \mu_2 = 0$

H_a : $\mu_1 - \mu_2 > 0$

Test statistic

$$z = \frac{(\bar{x} - \bar{y}) - 0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = 2.14$$

$$\text{p-value} = 1 - \Phi(2.14) = 1 - 0.9838 = 0.0162 < \alpha$$

Conclusion : Reject H_0

   (b) (20 points) Construct a 92% confidence interval of the true average difference in performance of the two approaches ($\mu_{experimental} - \mu_{control}$).

$X_1, \dots, X_m$: experimental $m = 85$ $\bar{x} = 27.34$ $s_1 = 8.85$

$Y_1, \dots, Y_n$: control $n = 79$ $\bar{y} = 23.87$ $s_2 = 11.6$

92% CI:

$$\left((\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}} \, , \, (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}} \right)$$

$$(1.76 \, , \, 5.18)$$

$\alpha = 0.08$ $\alpha/2 = 0.04 \rightarrow$ look 0.96 $z_{\alpha/2} = 1.75$