* Worksheet 10 :

problem 1 : No, because the population might have increased. There could be many other confounding factors for this.

problem 2 : a) observational

b) to compare similar population with the same characteristics which can potentially avoid confounding factors in conclusions.

c) The conclusion is not true. The current smokers might have been the healthy people who have just started smoking.
To draw proper conclusions, we will need to compare the people who have smoked for the same duration of time.

problem 6 : a) The coin is not biased

b) X = number of heads $= X_1 + \cdots + X_{10000}$

under null hypothesis : $E[X] = \frac{1}{2} \times 10000 = 5000$

$Var[X] = \frac{1}{2} \times \frac{1}{2} \times 10000 = 2500$

$\Rightarrow Z = \frac{5400 - 5000}{50} = 8$ $\Rightarrow X \sim N(5000, 2500)$

$p = N(5400, \mu=5000, \sigma^2=2500) = 0.00003$

c) large z & small p-value are strong evidence against our null hypothesis $\Rightarrow$ the coin is biased towards head

Problem 7 : Null hypothesis: die is fair

$X = X_1 + \cdots + X_{100}$ $\quad X_i$: i with prop $1/6$

under null $\Rightarrow E[X_i] = 3.5 \quad \Rightarrow Var[X_i] = \frac{1}{6}(1^2 + \cdots + 6^2) - 3.5^2 = 2.91$

$\Rightarrow E[X] = 100 \times 3.5 = 350 \quad Var[X] = 100 \times 2.91 = 291$

$\Rightarrow z = \frac{368 - 350}{\sqrt{291}} = 1.055 \Rightarrow$ not very strong evidence against null hypothesis

$p = 0.145$

$\Rightarrow$ This can be explained as chance variation

Problem 8 : a higher p-value is better for proving null hypothesis / lower p-value will be against null hypothesis

Problem 9 : X_1 : drug abuse percentage in 1985
X_2 : " " " " 1992

Null hypothesis: $X_1 = X_2$ (or X_1 & X_2 has the same mean $= M$)

a) $E[X_1] = M$ $\quad Var[X_1] = 700 \times \frac{21.9}{100} \times (1 - \frac{21.9}{100}) = 119.72$

$E[X_2] = M$ $\quad Var[X_2] = 700 \times \frac{11}{100} \times \frac{89}{100} = 68.53$

$E[X_1 - X_2] = 0$ $\quad Var[X_1 - X_2] = \sigma_1^2 + \sigma_2^2 = 188.25$

$\Rightarrow Z = \frac{(\frac{21.9 - 11}{100}) \times 700 - 0}{\sqrt{188.25}} = 5.56$ $\Rightarrow$ Strong evidence against null

$\Rightarrow$ This is a real difference

b) $Var[X_1] = 700 \times \frac{36.4}{100}(1 - \frac{36.4}{100}) = 162.98$

$Var[X_2] = 700 \times \frac{31.9}{100}(1 - \frac{31.9}{100}) = 152.06$

$\Rightarrow X_1 - X_2 \sim N(0, 315.04)$

$\Rightarrow Z = \frac{(\frac{36.9 - 31.9}{100}) \times 700}{\sqrt{315.04}} = 1.97$ $\Rightarrow p = 0.0242$

Stray evidence against null hypothesis

$\Rightarrow$ This is real difference

Problem 10

X_1: average # hours in public school

X_2: " " " " private "

Null hypothesis $\Rightarrow$ X_1 & X_2 has same mean μ

$$\Rightarrow X_1 \sim N\left(\mu, \frac{10.5^2}{1000}\right)$$

$$X_2 \sim N\left(\mu, \frac{9.9^2}{1000}\right)$$

$$\Rightarrow X_1 - X_2 \sim N\left(0, \frac{10.5^2 + 9.9^2}{1000}\right) = N(0, 0.2082)$$

$$\Rightarrow Z = \frac{12.2 - 9.2}{\sqrt{0.2082}} = \frac{3}{0.456} = 6.57$$

p-value = 0.00003 $\Rightarrow$ strong evidence against null hypothesis

$\Rightarrow$ The difference is not due to chance