Worksheet 12

Prashant Kolkur

1) The data does not clearly show that public health got worse since over the period of 1960-1990 population also increased

2) a) Observational study.
Study was based on observation and not based on investigators dividing the group into control and treatment group

b) Because the observation needs to be made on age and gender so that age and gender also needs to be taken into consideration
Different age group men and women can have different effects of smoking

(c) Its a wrong conclusion. Because people might have stopped smoking since their health is severly deteriorated after so many years of smoking. so current smokers health may be better than people who recently stopped smoking

3) a) Observational study

b) Because different age group can have different sexual activities and different needs.
If people are married and not married then their sexual partner can be wife in case of married and many persons in case of unmarried.
Educated people may try to have safe sex than uneducated people.

c) People who use pills are more sexually active and may have sex with many people. So chances of having Cervical cancer are higher in pill users than non users.

(6) H = 5400

Null hypothesis = outcome should be
H = 5000 heads
T = 5000 tails

$p = \frac{1}{2}$

Alternate hypothesis = coin is biases with probability p

$$Z = \frac{\text{observed} - \text{Expected}}{\text{std}}$$

$$\text{observed} = \frac{5400}{10000} = 0.54$$

$$\text{Expected} = \frac{5000}{10000} = 0.5$$

$$\text{std} = \sqrt{\frac{p(p-1)}{n}} = \sqrt{\frac{0.5 \times 0.5}{10000}} = 5 \times 10^{-3}$$

$$Z = \frac{0.54 - 0.5}{5 \times 10^{-3}} = \frac{0.04 \times 10^{3}}{5} = \frac{40}{5}$$

$$\boxed{Z = 8}$$

P-value with significance 0.05

≤ 0.00001

This is strong evidence against the null

(7) Die is rolled 100 times

observed = 368

Expected = 350

Variance σ^2 for mean = 3.5 of one die

$= \sum x_i^2 P(x_i) - \mu^2$

$= \frac{1}{6}[1^2+2^2+3^2+4^2+5^2+6^2] - 3.5^2$

$= \frac{91}{6} - 12.25$

$= 15.1667 - 12.25$

$= 2.9167$

$N(3.5, 2.9167)$

For 100 rolls, distribution $= N(350, 291.67)$

$\therefore Z = \frac{\text{observed} - \text{expected}}{\sigma}$

$= \frac{368 - 350}{\sqrt{291.67}} = \frac{18}{17.07}$

$= 1.054$

p-value with significance 0.05 is $= 0.14594$
$= 14.59\%$

So die is not loaded. It is just chance variation
high p-value indicates weak evidence against null hypothesis, so we fail to reject null hypothesis

(8) Null hypothesis is what we try to discredit hence lower p-value is better to indicate strong evidence against null hypothesis

(9) $n = 700$

(a) Null hypothesis – No change in marijuana uses from 1985 to 1992

~~observed = 21.9% − 11.0%~~

X_A = marijuana users in 1985 → 21.9% = 0.219

X_B = marijuana users in 1992 – 11.0% = 0.11

$$\sigma_A = \sqrt{\frac{0.219(1-0.219)}{700}} = \sqrt{0.0002443} = 0.0156$$

$$\sigma_B = \sqrt{\frac{0.11(1-0.11)}{700}} = \sqrt{0.000139} = 0.0118$$

$$\sigma = \sqrt{\sigma_A^2 + \sigma_B^2} = \sqrt{0.015^2 + 0.0118^2}$$

$$\sigma = \sqrt{0.0003826} = 0.0195$$

$$Z = \frac{0.110 - 0.219}{0.0195} = -5.589$$

p-value with significance level 0.05 is < 0.0001

So Difference is real

(b) Null hypothesis — No change in percentage of cigarette smokers between 1985 and 1992

X_A = cigarette smokers in 1985 = 36.9% = 0.369

X_B = cigarette smokers in 1992 = 31.9% = 0.319

$$\sigma_A = \sqrt{\frac{0.369 \times (1-0.369)}{700}} = 0.018$$

$$\sigma_B = \sqrt{\frac{0.319 \times (1-0.319)}{700}} = 0.0176$$

$$\sigma = \sqrt{\sigma_A^2 + \sigma_B^2} = 0.0252$$

$$z = \frac{0.319 - 0.369}{0.0252} = -1.98$$

P value with significance level 0.05 is 0.0238
= 2.38%.

So difference is real

(10) $n = 1000$

Public

$\mu_1 = 12.2 \quad \sigma_1 = 10.5$

Private universities :-

$\mu_2 = 9.2 \quad \sigma_2 = 9.9$

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$$

$$= \sqrt{\left(\frac{10.5}{\sqrt{1000}}\right)^2 + \sqrt{\left(\frac{9.9}{1000}\right)^2}}$$

$$= 0.456$$

$$z = \frac{12.2 - 9.2}{0.456}$$

$$z = 6.579$$

P-value with significance level 0.05 is < 0.0000

So difference between the two averages are real and not due to chance