$$DSE\_HW5\_Day5$$
$$Worksheet\ 12$$

1

$n = 900$
$P_{red} = 9/10$
$P_{blue} = 1/10$

$\mu = nP_{red} = 900 \times 9/10 = 810$
$\sigma^2 = nP_{red}(1 - P_{red}) = 900 \times 9/10 \times 1/10 = 81$

$Answer:$
$N(810, 81)$

3

(a)

$n = 100$

$p = 1/20$

$E(X_i) = np = 100 \times 1/20 = 5$

$var(X_i) = np(1 - p) = 100 \times 1/20 \times 19/20 = 19/4 = 4.75$

$Answer:$

4.75

(b)

$N(5, 4.75)$

$95\%\ confidence\ interval: 5 \pm 2\sqrt{4.75} = 5 \pm 4.4$

$Answer:$

$Upper\ bound: 5 + 4.4 = 9.4$

(c)

$Answer:$

$E(Y_i) = 1 \times 1/2 + (-1) \times 1/2 = 0$

$var(Y_i) = 1$

(d)

$$E(Z_r - Z_b) = E(Y_1 + ... + Y_{100}) = E(Y_1) + ... + E(Y_{100}) = 0$$

$$var(Z_r - Z_b) = var(Y_1 + ... + Y_{100}) = var(Y_1) + ... + var(Y_{100}) = 100$$

Answer :

$Normal\ approximation : Z_r - Z_b \sim N(0, 100)$

(e)

Answer :

$99\%\ confidence\ interval\ for\ Z_r - Z_b : [-30, 30]$

$99\%\ confidence\ interval\ for\ Z : [0, 30]$

5

$p = 0.5$

(a)

$N(p,\ p(1-p)/100) = N(0.5,\ 0.5 \times (1-0.5)/100) = N(0.5,\ 1/400)$

$std(x) = \sqrt{1/400} = 1/20$

$Answer:$

1/20

(b)

$N(p,\ p(1-p)/2500) = N(0.5,\ 0.5 \times (1-0.5)/2500) = N(0.5,\ 1/10000)$

$std(x) = \sqrt{1/10000} = 1/100$

$Answer:$

1/100

7

$\sigma = 1\%$

$N(p,\ p(1-p)/n) = N(0.2,\ 0.16/n)$

$0.16/n = 0.01^2$

$n = 1600$

$\sigma = 1\%$

$N(p,\ p(1-p)/n) = N(0.4,\ 0.24/n)$

$0.24/n = 0.01^2$

$n = 2400$

Answer :

2500

8

$\widehat{p} = 194/500 = 0.388$

$n = 500$

$\widehat{p} : N(p,\ p(1-p)/n)$

$std(\widehat{p}) = \sqrt{p(1-p)/n} = \sqrt{0.388 \times 0.612/500} \approx 0.022$

$95.5\%\ confidence\ interval:$

$[\widehat{p} - 2std(\widehat{p}),\ \widehat{p} + 2std(\widehat{p})] = [0.344,\ 0.432]$

Answer :

$[0.344,\ 0.432]$

9

The distribution is the same.

What matters is the sample size and not the overall population size.

Answer :

1000

10

The distribution is the same.

What matters is the sample size and not the overall population size.

Answer :

nationwide average score : 307

standard deviation of this estimate : $30/\sqrt{1000} = 0.9487$

Worksheet 13

1

Answer :

No, this data does not show that the public's health got worse over the period 1960-1990. The overall population size may change or other factors that will also impact the results.

2

(a)

Answer :

Observational study

(b)

Answer :

It is possible that gender and age are confounding factors to health. Thus, by separating into groups, we can determine the effect that smoking has on health. For example, men are more likely to smoke than women, and more likely to get heart disease. Older people have different smoking habits and more at risk for heart disease.

(c)

Answer :

The conclusion does not follow as the number of people is not the same. People who have died from smoking are not included in the study, and thus, it seems like quitting smoking has an adverse effect on health.

3

(a)

$Answer:$

Observational study

(b)

$Answer:$

Age, education, and marital status could be confounding factors, as they contribute to the use of contraceptives themselves.

(c)

$Answer:$

No, as the observational group is not random, and there are more confounding factors that are not taken into account.

5

Answer :

Since parents were much more tolerant of left-handedness in mid-century, we would expect much more left-handed people in mid-century compared to the early part of the century. However, this would then also imply that most left-handed people who died could not have been very old as most left-handed people will be born at the very least in mid-century and thus this could explain the observed discrepancy in average of left-handed and right-handed people in the 1990s. There is not sufficient evidence to conclude that left-handedness is the cause of the smaller average death for left-handed people.

6

(a)

Answer :

$Null\ hypothesis:\ \ The\ coin\ is\ fair\ (p = 0.5).$

$Alternative\ hypothesis:\ \ The\ coin\ is\ biased\ in\ favor\ of\ heads\ (p > 0.5).$

(b)

$X:\ \ number\ of\ heads$

$E(X) = np = 10000 \times 0.5 = 5000$

$var(X) = np(1-p) = 10000 \times 0.5 \times 0.5 = 2500$

$std(X) = 50$

$Under\ the\ Null,\ X\ has\ a\ N(5000, 2500)\ distribution.$

$$z - statistics = \frac{observed - expected}{std(X)} = \frac{5400-5000}{50} = 8$$

$p - value = P_r(being\ more\ than\ 8\ standard\ deviations\ above\ the\ mean)$

$= Tiny.$

Answer :

$z - statistics = 8$

$p - value = Tiny$

(c)

Answer :

$Conclude:\ \ strong\ evidence\ that\ the\ coin\ is\ not\ fair!$

7

$Null\ hypothesis: \ The\ die\ is\ fair\ (p = 1/6).$

$X: observed\ sum\ of\ 100\ die\ rolls.$

$X = X_1 + ... + X_{100}$

$E(X_i) = (1+2+3+4+5+6) \times 1/6 = 3.5$

$var(X_i) = ((1-3.5)^2 + ... + (6-3.5)^2)/6 = 17.5/6 = 35/12$

$std(X_i) = \sqrt{\frac{35}{12}}$

$E(X) = nE(X_i) = 100 \times 3.5 = 350$

$var(X) = nvar(X_i) = 100 \times 35/12 = 875/3$

$std(X) = \sqrt{var(X)} = \sqrt{\frac{875}{3}}$

$Under\ the\ Null,\ X\ has\ a\ N(350, 875/3)\ distribution.$

$$z - statistics = \frac{observed - expected}{std(X)} = \frac{368-350}{\sqrt{\frac{875}{3}}} \approx 1.05$$

$Answer:$

Since it is not very unusual to obtain an observed value that is 1.05 standard errors from the expected value, the difference between the total number of spots and the expected value can be explained by chance variation in this case.

9

(a)

Null hypothesis : the two distributions have the same mean.

X : *fraction in* 1985.

Y : *fraction in* 1992.

$$std(X) = \sqrt{p(1-p)/n} = \sqrt{21.9\% \times (1-21.9\%)/700} = 0.0156$$

$$std(Y) = \sqrt{p(1-p)/n} = \sqrt{11.0\% \times (1-11.0\%)/700} = 0.0118$$

$$std(Y-X) = \sqrt{(std(X))^2 + (std(Y))^2} = 0.0196$$

Under the Null, $Y - X$ *is normal with mean* 0 *and* $(std(Y-X))^2$.

$$z - statistics = \frac{observed - expected}{std(Y-X)} = \frac{(21.9\% - 11.0\%) - 0}{0.0196} \approx 5.56$$

Answer :

difference is real

(b)

Null hypothesis : the two distributions have the same mean.

X : *fraction in* 1985.

Y : *fraction in* 1992.

$$std(X) = \sqrt{p(1-p)/n} = \sqrt{36.9\% \times (1-36.9\%)/700} = 0.01824$$
$$std(Y) = \sqrt{p(1-p)/n} = \sqrt{31.9\% \times (1-31.9\%)/700} = 0.01762$$
$$std(Y-X) = \sqrt{(std(X))^2 + (std(Y))^2} = 0.02536$$

Under the Null, $Y - X$ *is normal with mean* 0 *and* $(std(Y-X))^2$.

$$z - statistics = \frac{observed - expected}{std(Y-X)} = \frac{(36.9\% - 31.9\%) - 0}{0.02536} \approx 1.97$$

Answer :

chance variation

10

Let X be number of hours worked at public universities

Let Y be number of hours worked at public universities

Null hypothesis : *they have the same averages*

$n = 1000$

$mean(X) = 12.2$

$std(X) = 10.5/\sqrt{1000}$

$mean(Y) = 9.2$

$std(Y) = 9.9/\sqrt{1000}$

$mean(X - Y) = mean(X) - mean(Y) = 3$

$std(X - Y) = \sqrt{10.5^2/1000 + 9.9^2/1000} = \sqrt{0.20826}$

Under the Null, $X - Y$ *has a* $N(0, \sqrt{0.20826})$ *distribution.*

$z - statistics(X - Y) = \frac{observed - expected}{std(X)} = \frac{3-0}{\sqrt{0.20826}} \approx 6.57$

Answer :

No, the difference between these two averages isn't due to chance.