WORKSHEET 13 Hypothesis testing

1. In the US in 1990, there were 2.1 million deaths from all causes, compared to 1.7 million in 1960: nearly a 25% increase. Does this data show that the public's health got worse over the period1960–1990?

Solution:

Not enough information to conclude this result since death could be attributed to many factors.

2. The Public Health Service studied the effects of smoking on health, in a large sample of representative households. For men and women in each age group, those who had never smoked were on average somewhat healthier than the current smokers, but the current smokers were on average much healthier than those who had recently stopped smoking.
(a) Was this a controlled experiment or an observational study?
(b) Why did they study men and women and the different age groups separately?
(c) The conclusion seems to be that you shouldn't start smoking, but once you've started, you shouldn't stop.Comment.

Solution:

(a) This is an observational study.
(b) This will ensure limiting confounding factors to the study.
Note: **Confounding** is defined in terms of the data generating model (as in the figure above). Let X be some independent variable, and Y some dependent variable. To estimate the effect of X on Y, the statistician must suppress the effects of extraneous variables that influence both X and Y. We say that X and Y are confounded by some other variable Z whenever Z causally influences both X and Y. — Wikipedia
(c) The conclusion ignores the effect of other confounding factors. People might quit smoking for different reasons.

3. According to a study done by Kaiser Permanente in California, users of oral contraceptives have a higher rate of cervical cancer than non-users, even after adjusting for age, education, and marital status.
(a) Was this a controlled experiment or an observational study?
(b) Why did the investigators adjust for age, education, and marital status?
(c) The investigators concluded that the pill causes cervical cancer. Was this justified?

Solution:

(a) This is an observational study.
(b) This will ensure limiting confounding factors to the study.
(c) This is Not justified, since there is other significant confounding factors like genetic factors or eating habits.

5. Studies of death certificates in the 1990s showed the average age of death was smaller for left-handed people than for right-handers.
During the 20th century, there were big changes in child-rearing practices. In the early part of the century, parents insisted that their children be right-handed. By mid-century, parents were a lot more tolerant of left-handedness. Could this explain the observed discrepancy in average age at death of left- and right-handed people in the 1990s?

Solution:

Since parents were more tolerant of left-handedness in the mid-century compared to the early century, the total amount of death certificates for left-handed people would increase, and also, those also imply that most left-handed people who died could not have been very old. Hence this could explain the observed discrepancy in average age at death of left- and right-handed people in the 1990s.

6. In 10,000 tossings, a coin came up heads 5,400 times. Should we conclude that the coin is biased?
(a) Formulate the null hypothesis and alternative hypothesis.
(b) Compute the z statistic and the p-value.
(c) What do you conclude?

Solution:

(a) Null hypothesis: The coin is fair.
Alternative hypothesis: The coin is biased with $Alpha = 0.05$ (95% Confidence Interval)

(b) $\sigma = 10000 * 0.5 * (1 - 0.5) = 50$
$Z - statistic = \frac{Observed - Expected}{\sigma} = \frac{5400-5000}{\sigma} = 400/50 = 8$
The p-value is $\leq .0001$.

(c) Since the result is significant when $p - value \leq 0.05$, the null hypothesis has been rejected, the coin is biased.

7. A die is rolled 100 times. The total number of spots is 368 instead of the expected 350. Can this be explained as chance variation, or is the die loaded?

Solution:

Null hypothesis: the die is not loaded
Alternative hypothesis: the die is loaded(Alpha=0.05)
$\mu = 3.5$
$\sigma^2 = 1/6(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - 3.5^2 = 2.92$
$\sigma = 1.71$
$\hat{\sigma} = 100 * \sigma$
$Z - statistic = \frac{Observed - Expected}{\hat{\sigma}} = \frac{368-350}{1.71*100} \simeq 1.054$
p-value is 0.145941 which is greater than 0.05, hence the difference is not significant, null hypothesis will not be rejected, the die is not loaded.

9. The National Household Survey on Drug Abuse was conducted in 1985 and 1992. In each year, a simple random sample of 700 people was used.
(a) Among persons age 18 to 25, the percentage of marijuana users dropped from 21.9% to 11.0%. Is this difference real, or a chance variation?
(b) Among persons age 18 to 25, the percentage of cigarette smokers dropped from 36.9% to 31.9%. Is this difference real, or a chance variation?

Solution:

(a) Null hypothesis: there is no difference, $\mu_{1985} = \mu_{1992}$
Alternative hypothesis: $\mu_{1985} \neq \mu_{1992}$
Alpha =0.05
$Var_1 = 0.219 * (1 - 0.219)/700$
$Var_2 = 0.11 * (1 - 0.11)/700$
$Sd = \sqrt{Var(\mu_{1985} - \mu_{1992})} = \sqrt{Var_1 + Var_2}$
$Z - statistic = \frac{Observed - Expected}{Sd} = \frac{0.219-0.11-0}{sd} \simeq 5.561$
p-value is less than 0.0001, hence the null hypothesis has been rejected. This difference is real.

(b) Null hypothesis: there is no difference, $\mu_{1985} = \mu_{1992}$
Alternative hypothesis: $\mu_{1985} \neq \mu_{1992}$
Alpha =0.05
$Var_1 = 0.369 * (1 - 0.369)/700$
$Var_2 = 0.319 * (1 - 0.319)/700$

$Sd = \sqrt{Var(\mu_{1985} - \mu_{1992})} = \sqrt{Var_1 + Var_2}$
$Z - statistic = \frac{Observed - Expected}{Sd} = \frac{0.369 - 0.319 - 0}{sd} \simeq 1.972$
p-value is 0.04861 which still less than 0.05, hence the null hypothesis has been rejected. This difference is real.

10. A random sample of 1000 freshmen at public universities were asked how many hours they worked each week (for pay). The average number of hours turned out to be 12.2, with a standard deviation of 10.5. A similar survey at private universities had an average of 9.2 hours, with a standard deviation of 9.9. Is the difference between these two averages due to chance?

Solution:

Null hypothesis: there is no difference, $\mu_{public} = \mu_{private}$
Alternative hypothesis: $\mu_{public} \neq \mu_{private}$
Alpha =0.05
$\sigma_{public} = 10.5/\sqrt{1000}$
$\sigma_{private} = 9.9/\sqrt{1000}$
$\sigma = \sqrt{\sigma^2_{public} + \sigma^2_{private}}$
$Z - statistic = \frac{Observed - Expected}{\sigma} = \frac{12.2 - 9.2 - 0}{\sigma} \simeq 6.21$
p-value is less than 0.0001, hence the null hypothesis has been rejected. This difference is real not due to chance.