Introduction to Statistics - Homework #10

- Use Appendix C: Distribution tables (e.g. Z-table) from page 408 to 417 in the textbook if necessary.

Exercise 8.12 - Crawling babies, Part I

A study recorded data on 12 groups of infants born throughout the year to investigate whether crawling age (x) is associated with environmental temperature (y). For each group, researchers measured the average crawling age (in weeks) and the average temperature (in °F) when the babies were six months old. Conduct a hypothesis test to determine if there is a **negative** correlation between temperature and crawling age. (The data come from a bivariate normal distribution.)

(1) State the null and alternative hypotheses. (Use one-sided test)

Solution:

$$H_0 : \rho = 0, \qquad H_A : \rho < 0$$

(2) Find the test statistic and its null distribution.

Solution: Under H_0, the test statistic is:

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} = \frac{R\sqrt{10}}{\sqrt{1-R^2}} \sim t(10)$$

(3) Find the rejection region.

Solution: Since $H_A : \rho < 0$ is a left-sided alternative and $T \sim t(10)$ under H_0, we reject H_0 at significance level $\alpha = 0.05$ if

$$t < -t_{0.05}(10) = -1.812.$$

Thus, the rejection region is $t < -1.812$.

(4) Compute the correlation coefficient and the observed test statistic using the following summary statistics.

n	s_{xx}	s_{yy}	s_{xy}
12	34.0961	2762.25	-214.735

Solution:

$$r = \frac{s_{xy}}{\sqrt{s_{xx} \cdot s_{yy}}} = \frac{-214.735}{\sqrt{34.0961 \cdot 2762.25}} \approx -0.6998$$

$$t = \frac{-0.6998 \cdot \sqrt{10}}{\sqrt{1-(-0.6998)^2}} \approx \boxed{-3.096}$$

(5) Using the rejection region, complete the hypothesis test. State the conclusion in the context of the data.

Solution: The observed test statistic is $t \approx -3.096$. Since

$$-3.096 < -1.812,$$

the observed statistic falls in the rejection region. Therefore, we reject H_0.

Conclusion: At the significance level $\alpha = 0.05$, there is statistically significant evidence of a negative correlation between temperature and crawling age.

Exercise 8.9 - Speed and height, revised

50 UCLA students were asked to fill out a survey where they were asked about their height, fastest speed they have ever driven, and gender. We are interested in investigating whether height (x) is associated with fastest driving speed (y). Conduct a hypothesis test to determine if there is a correlation between height and fastest speed. (Assume the data come from a bivariate normal distribution.)

(1) State the null and alternative hypotheses. (Use two-sided test)

Solution:

$$H_0 : \rho = 0, \qquad H_A : \rho \neq 0$$

(2) Find the test statistic and its null distribution.

Solution:

Under H_0, the test statistic is:

$$T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

With a sample of $n = 50$ students, we have:

$$T = \frac{R\sqrt{48}}{\sqrt{1-R^2}} \sim t(48)$$

(3) Compute the correlation coefficient and the observed test statistic using the following summary statistics.

n	s_{xx}	s_{yy}	s_{xy}
50	727.245	18792.32	399.78

Solution:

First, compute the sample correlation coefficient:

$$r = \frac{s_{xy}}{\sqrt{s_{xx} \cdot s_{yy}}} = \frac{399.78}{\sqrt{727.245 \cdot 18792.32}} \approx 0.1081$$

Then compute the observed test statistic:

$$t = \frac{0.1081 \cdot \sqrt{48}}{\sqrt{1-(0.1081)^2}} \approx \boxed{0.7534}$$

(4) Compute the p-value and complete the hypothesis test.

Solution:

The p-value for a two-sided test is

$$p\text{-value} = 2 \times P(T > 0.7534) \approx \boxed{0.4548}.$$

Since the p-value is larger than $\alpha = 0.05$, we do not reject the null hypothesis.

Conclusion: There isn't statistically significant evidence of a correlation between height and fastest speed driven among UCLA students.