Aarhus university

Computer-Science

Introduction to probability and statistics

Handin 12

Author
Søren M. Damsgaard

Student number
202309814

December 20, 2025

Last Assignment

We want to test the null hypothesis that 'at least 10% of the students suffer from allergies'. We have collected data from 225 students, where 21 of them suffer from allergies. Consider the following statistical model for our data: Let $X_1, \ldots, X_{225}$ denote the sample, where $X_i \sim \text{Bernoulli}(\theta)$ and $\theta \in (0, 1)$ is an unknown parameter. Consider the following hypotheses:

$$H_0 : \theta \geq 0.1 \quad \text{vs} \quad H_1 : \theta < 0.1$$

—

(a) Argue that the above model corresponds to a sample with unknown mean and variance, and likewise argue that H_0 corresponds to the statement that at least 10% of the students suffer from allergies.

Solution for (a):

i. **Unknown Mean and Variance:** For a Bernoulli distribution $X_i \sim \text{Bernoulli}(\theta)$, the mean is $E[X_i] = \theta$ and the variance is $\text{Var}[X_i] = \theta(1 - \theta)$. Since θ is an **unknown parameter**, both the mean and the variance of the underlying distribution are unknown.

ii. **H_0 Interpretation:** θ represents the true proportion (or probability) of students in the population who suffer from allergies. The null hypothesis $H_0 : \theta \geq 0.1$ states that this true proportion is greater than or equal to 0.1 (or 10%). This directly translates to the statement that **at least 10% of the students suffer from allergies**.

—

(b) Perform an asymptotic hypothesis test with a significance level $\alpha = 0.05$ for the null hypothesis.

Solution for (b): We perform a Z-test for a population proportion, θ.

i. **Hypotheses and Significance Level:**

$$H_0 : \theta \geq 0.1 \quad \text{vs} \quad H_1 : \theta < 0.1 \quad \text{(Left-tailed test)}$$

$$\alpha = 0.05$$

ii. **Observed Sample Proportion ($\hat{\theta}$):** The number of successes is $Y = 21$, and the sample size is $n = 225$.

$$\hat{\theta} = \frac{Y}{n} = \frac{21}{225} \approx 0.0933$$

iii. **Test Statistic (Z):** The asymptotic test statistic uses the value under the null hypothesis, $\theta_0 = 0.1$.

$$Z = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$$

$$Z = \frac{0.0933 - 0.1}{\sqrt{\frac{0.1(1-0.1)}{225}}} = \frac{-0.0067}{\sqrt{\frac{0.09}{225}}} = \frac{-0.0067}{\sqrt{0.0004}} = \frac{-0.0067}{0.02}$$

$$Z \approx -0.335$$

iv. **Critical Value (Z_α):** Since this is a left-tailed test with $\alpha = 0.05$, the critical value $Z_{0.05}$ is found such that $P(Z \leq Z_{0.05}) = 0.05$.

$$Z_{0.05} \approx -1.645$$

v. **Conclusion:** We compare the calculated test statistic ($Z \approx -0.335$) with the critical value ($Z_{0.05} \approx -1.645$). Since $Z > Z_{0.05}$ (i.e., $-0.335 > -1.645$), the test statistic does not fall into the rejection region. We **do not reject H_0**. *Conclusion in context:* There is not enough evidence at the 5% significance level to conclude that the true proportion of students suffering from allergies is less than 10%.

—

(c) Calculate the corresponding P-value.

Solution for (c): The P-value is the probability of observing a result as extreme as, or more extreme than, the one observed, assuming H_0 is true. Since this is a left-tailed test, we calculate $P(Z \leq Z_{\text{obs}})$ where $Z_{\text{obs}} \approx -0.335$:

$$\text{P-value} = P(Z \leq -0.335)$$

Using the standard normal distribution (Z-table or calculator):

$$\text{P-value} \approx 0.3687$$

Since the P-value (0.3687) is much larger than the significance level ($\alpha = 0.05$), we confirm the decision from part (b) to **not reject the null hypothesis H_0**.

$$f_{X,Y}(x,y) = \frac{2}{\sqrt{\pi}} e^{-x^2-2y} \text{ for } y > 0$$

find distribution for X and say if X and Y are independent?

$$f_X(x) = \int_0^\infty f_{X,Y}(x,y)dy = \int_0^\infty \frac{2}{\sqrt{\pi}} e^{-x^2-2y} dy = \frac{2}{\sqrt{\pi}} e^{-x^2} \int_0^\infty e^{-2y} dy = \frac{2}{\sqrt{\pi}} e^{-x^2} \cdot \frac{1}{2} = \frac{1}{\sqrt{\pi}} e^{-x^2}$$

Now for Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{-\infty}^{\infty} \frac{2}{\sqrt{\pi}} e^{-x^2-2y} dx = \frac{2}{\sqrt{\pi}} e^{-2y} \int_{-\infty}^{\infty} e^{-x^2} dx = \frac{2}{\sqrt{\pi}} e^{-2y} \cdot \sqrt{\pi} = 2e^{-2y}$$

To check for independence, we see if $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$:

$$f_X(x) \cdot f_Y(y) = \left(\frac{1}{\sqrt{\pi}} e^{-x^2}\right) \cdot \left(2e^{-2y}\right) = \frac{2}{\sqrt{\pi}} e^{-x^2-2y} = f_{X,Y}(x,y)$$

Thus, X and Y are independent.

Find the distribution of $Z = e^{\frac{Y}{3}} - 1$. To find the distribution of $Z = e^{\frac{Y}{3}} - 1$, we first need to find the cumulative distribution function (CDF) of Z.

$$F_Z(z) = P(Z \leq z) = P\left(e^{\frac{Y}{3}} - 1 \leq z\right) = P\left(Y \leq 3\ln(z+1)\right)$$

Now, we need to find the CDF of Y:

$$F_Y(y) = \int_0^y f_Y(t)dt = \int_0^y 2e^{-2t} dt = \left[-e^{-2t}\right]_0^y = 1 - e^{-2y}$$

Substituting back into the CDF of Z:

$$F_Z(z) = F_Y(3\ln(z+1)) = 1 - e^{-2\cdot 3\ln(z+1)} = 1 - e^{-6\ln(z+1)} = 1 - (z+1)^{-6}$$

To find the probability density function (PDF) of Z, we differentiate the CDF:

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz}\left(1 - (z+1)^{-6}\right) = 6(z+1)^{-7}$$

Thus, the distribution of Z is given by the PDF:

$$f_Z(z) = 6(z+1)^{-7} \text{ for } z > -1$$