Lecture 25 More On Hypothesis Testing

BIO210 Biostatistics

Xi Chen

Spring, 2022

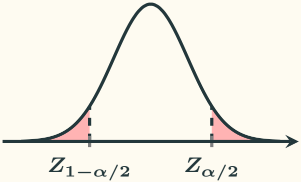
School of Life Sciences
Southern University of Science and Technology

南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

A step by step hypothesis testing

1. Specify what you are comparing.
2. Formulate hypotheses
3. Check assumptions
4. Determine significance level α
5. Compute the test statistic
6. Check significance
7. Make a decision about whether to reject H_0
8. Interpret findings

6a)



6b) Calculate the p-value.

6c) Construct $(1-\alpha) \times 100\%$ confidence interval to see if it covers the H_0 value.

The Null And Alternative Hypotheses

	Two-tailed test	One-tailed test	
Hypotheses style #1	$H_0 : \pi = \pi_0$ $H_1 : \pi \neq \pi_0$	$H_0 : \pi \leqslant \pi_0$ $H_1 : \pi > \pi_0$	$H_0 : \pi \geqslant \pi_0$ $H_1 : \pi < \pi_0$
Hypotheses style #2	$H_0 : \pi = \pi_0$ $H_1 : \pi \neq \pi_0$	$H_0 : \pi = \pi_0$ $H_1 : \pi > \pi_0$	$H_0 : \pi = \pi_0$ $H_1 : \pi < \pi_0$

Both are okay and used in statistics!

One-sample Hypothesis Testing For Proportion

Test statistic for proprotion

$$z = \frac{p - \pi}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

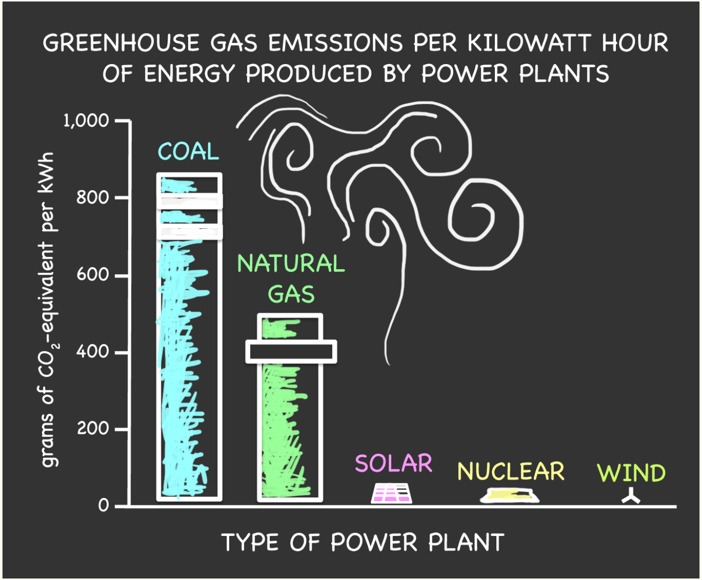
$P($data or more extreme $\mid H_0$ is true$)$

$$z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}}$$

consistent with$(1-\alpha) \times 100\%$ CI

Both are okay and used in statistics!

Nuclear Power Plant



Risk of Nuclear Power Plant

- **Occupational Health, Cancer**

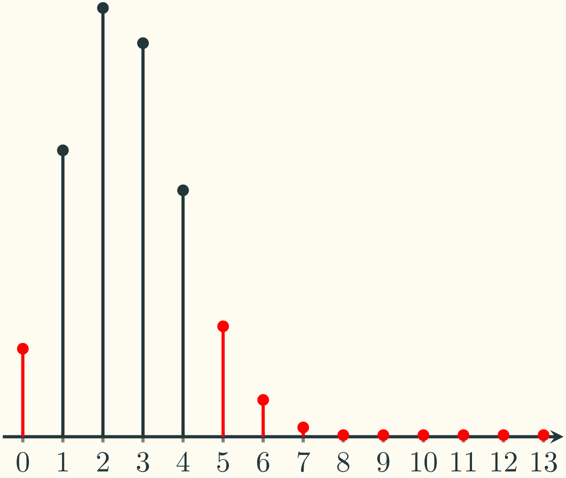
  - The safety of people who work at or live close to nuclear-power plants has been the subject of widely publicised debate in recent years. One possible health hazard from radiation exposure is an excess of cancer deaths among those exposed. One problem with studying this question is that because the number of deaths attributable to either cancer in general or specific types of cancer is small, reaching statistically significant conclusions is difficult except after long periods of follow-up. An alternative approach is to perform a proportional-mortality study, whereby the proportion of deaths attributed to a specific cause in an exposed group is compared with the corresponding proportion in a large population. Suppose, for example, that 13 deaths have occurred among 55- to 64-year-old male workers in a nuclear-power plant and that in 5 of them the cause of death was cancer. Assume, based on vital-statistics reports, that approximately 20% of all deaths can be attributed to some form of cancer. Is this result significant different from the reports?

Conduct The Hypothesis Testing

1. $p = \dfrac{x}{n} = \dfrac{5}{13}$
2. $H_0: \ \pi = \pi_0 = 0.2, \ H_1: \ \pi \neq \pi_0 = 0.2$
3. Check assumptions: randomisation, independence, $np = 5, \ nq = 8$
4. $\alpha = 0.05$
5. Compute the test statistic ?
6. Check significance ?
7. Make the decision ?
8. Interpret findings ?

$$
\begin{aligned}
&P(\text{data or more extreme} \mid H_0 \text{ is true}) \\
&= P(x \geqslant 5 \mid \pi = 0.2) \\
&= P(x = 5 \mid \pi = 0.2) + \\
&\quad P(x = 6 \mid \pi = 0.2) + \\
&\quad P(x = 7 \mid \pi = 0.2) + \\
&\quad \vdots \\
&\quad P(x = 13 \mid \pi = 0.2) \\
&= \sum_{k=5}^{13} \binom{13}{k} 0.2^k 0.8^{13-k} = 0.099
\end{aligned}
$$

Tow-tailed Exact Binomial Test



One tailed p-value :

$$\sum_{k=5}^{13} \binom{13}{k} 0.2^k 0.8^{13-k} = 0.099$$

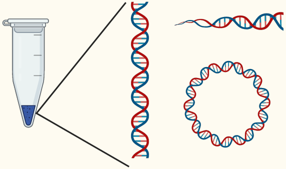
Two-tailed Exact Binomial p-values:

$$\begin{aligned} p &= \sum_{i \in \mathcal{I}} P(X = i) \\ &= \sum_{i \in \mathcal{I}} \binom{n}{i} p^i (1-p)^{n-i} \end{aligned}$$

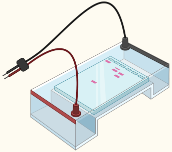
where $\mathcal{I} = \{i \mid P(X = i) \leqslant P(X = k)\}$

Gel Extraction

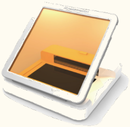
Target DNA preparation



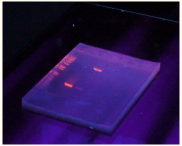
DNA separation



Transilluminator



Check bands



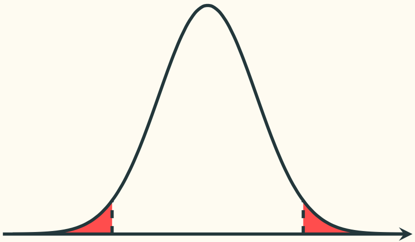
Cut the band

UV Safety Visor

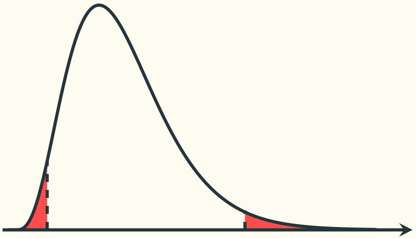
- Suppose you have a production line to make UV safe visors for scientists in the lab. Their quality is measured by a score indicating the percentage of UV they can block. Based on your past experience, you know their scores follow a normal distribution with a mean of 950 and a variance of $1,600$. To check the quality of visors from a particular batch, tests were run on a random sample of $n = 40$ visors, and the sample mean and sample variance were found to be 945 and $2,352.25$, respectively.
- Do the data provide sufficient evidence, at the $\alpha = 0.05$ level, to conclude that the population variance exceeds $1,600$?

One-sample Hypothesis Testing For Variance

Sampling distribution of the sample mean/proportion



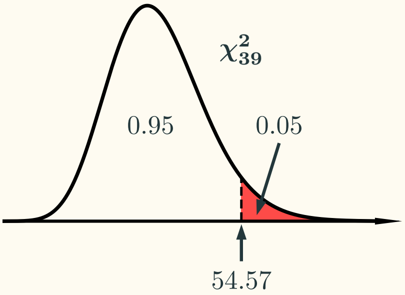
Sampling distribution of the sample variance



One-sample Hypothesis Testing For Variance

1. $s^2 = 2352.25$
2. $H_0 : \sigma^2 \leqslant \sigma_0^2 = 1600,\ H_1 : \sigma^2 > \sigma_0^2 = 1600$
3. Check assumptions: Randomisation, independence, population normally distributed
4. $\alpha = 0.05$
5. $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2} \geqslant \dfrac{39 \times 2352.25}{1600} = 57.34$
6. Check significance
7. Reject H_0.
8. The population variance do not exceed $1,600$.

6a) Rejection/critical region:



6b) P-value: $P(\chi^2 \geqslant 57.34) = 0.029$

6c) Construct $(1-\alpha) \times 100\%$ **CI:**

$$\left[\frac{39 \times 2352.25}{54.57},\ +\infty\ \right)$$