Module 5: Hypotheses Testing and Significance Tests

Dr. Dancik's favorite study

Consider the following experiment published in the *International Journal of Cosmetic Science* in 2009.

Male volunteers were randomly given either a cologne or a placebo (an identical aerosol spray lacking active ingredients). The males then completed a set of tasks that were videotaped. These videotapes were shown to a panel of females who rated the males according to their attractiveness.

Although the females could not smell the males, the males who received the cologne were rated higher, on average, than the males who received the placebo.

Does using cologne really make males more attractive?

Or can this difference between the groups be explained by *chance*?

Is the difference between groups *statistically significant?*

Hypothesis testing is the procedure used to answer these questions.

Study link: https://www.ncbi.nlm.nih.gov/pubmed/19134127

To conduct a hypothesis test, we first define the *null* and *alternative* hypothesis.

Null and Alternative Hypotheses

The **null hypothesis**, denoted H_0, is a statement that either a parameter takes on a particular value, or that there is no difference between parameters (i.e., there is no effect).

The **alternative hypothesis**, denoted H_A, states that either the parameter falls within an alternate range of values, or that there *is* a difference between parameters (i.e., there is an effect). The alternative hypothesis corresponds to a claim being made (e.g., that a treatment has an effect).

The logic beind hypothesis testing is similar to the logic in *proof by contradiction*.

Proof by contradiction: An observation that would be *impossible* if H_0 is true is evidence that H_0 is not true (in which case we accept H_A).

Silly example: I claim that everyone in this class is 19 years old, e.g. $H_0 : \mu = 19$ and $\sigma = 0$. I randomly select 5 people, one of them is 20. What do you conclude?

Hypothesis testing: If an observation would be *very unlikely* when H_0 is true, then this is evidence against H_0 and in support of H_A.

Basic idea

Testing whether or not a coin is biased.

1. State the null hypothesis:
2. State the alternative hypothesis:
3. If H_0 is true, and we flip a coin $n = 100$ times, 95% of the time the number of times the coin lands heads up would be between (approximately) 40 and 60 times.
4. Getting less than 40 heads or more than 60 heads would be unlikely (this only happens 5% of the time). Therefore, if we get, e.g., 72 heads, this provides evidence against H_0. We would reject H_0 and accept H_A. In practice, we calculate the probability of getting more than 72 heads (or less than 28 heads), which gives us a *P-value*, which is the probability of observing something more extreme (in the direction of H_A) than what we observe if H_0 is true.The smaller the p-value, the more evidence against H_0.

Steps of a Signficance Test About a Population Proportion

1. Verify the assumptions:
    1. The variable is categorical
    2. The data are obtained from randomization
    3. The sampling distribution of $\hat{p}$ under H_0 is approximately normal (we expect at least 15 successes and at least 15 failures under H_0).

Steps of a Signficance Test About a Population Proportion

2. State the null and alternative hypothesis
   - The null hypothesis will always be of the form $H_0 : p = p_0$.
   - The alternative hypothesis in general can have one of three forms:
     - Two-tailed: $H_A : p \neq p_0$
     - Left-tailed: $H_A : p < p_0$
     - Right-tailed: $H_A : p > p_0$

     We will only consider the *two-tailed* form in this class.

Steps of a Signficance Test About a Population Proportion

3. Calculate the test statistic, which measures how many standard deviations the sample statistic (i.e., $\hat{p}$) is from its center if H_0 is true.

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Steps of a Signficance Test About a Population Proportion

4. Calculate the p-value, which is the probability that the test statistic is *more extreme* than what was observed, in the direction of the alternative hypothesis H_A. This p-value represents the area
   - in both tails, with respect to $\pm z$, if H_A is two-tailed
   - in the left tail, with respect to z, if H_A is left-tailed
   - in the right tail, with respect to z, if H_A is right-tailed

Steps of a Signficance Test About a Population Proportion

5. If the p-value is sufficiently small, reject H_0 and accept H_A. Otherwise, fail to reject H_0 and conclude that there is insufficient evidence to reject H_0 (you NEVER conclude that H_0 is true!).

   Whether or not the p-value is sufficiently small depends on the chosen significance level. The **significance level**, denoted by α, is a number such that we reject H_0 if the p-value is less than or equal to that number. Typically, α is taken to be 0.05.

We will carry out hypothesis tests on proportions using the *prop.test* function in *R*.

Can dogs detect cancer by smell?

Fifty four trials were conducted as follows: 1 sample from a bladder cancer patient was placed among 6 control urine samples. In the 54 trials, the dogs correctly identified the bladder cancer sample 15 times. Did this study provide strong evidence that the dogs' predictions were better or worse than random guessing?

Example

A survey is carried out to determine whether or not a majority of American adults approve of the legalization of marijuana. In October of this year, Gallup conducted a survey of 1,028 adults, and found that 596 of them favor legalizing marijuana.

1. Specify the population parameter
2. Specify the point estimate (sample proportion)

Example (continued)

A survey is carried out to determine whether or not a majority of American adults approve of the legalization of marijuana. In October of this year, Gallup conducted a survey of 1,028 adults, and found that 596 of them favor legalizing marijuana.

4. Is there evidence to suggest that a majority of Americans favor the legalization of marijuana? Carry out a hypothesis test by completing the following steps, and using a level of significance of 0.05:
   (a) State the null and alternative hypotheses
   (b) Calculate the test statistic
   (c) Calculate the p-value
   (d) State the conclusion

Blind taste test

In a class at the University of Florida, 49 students are given a blind taste test comparing Coke and Pepsi. Approximately half of the subjects are randomly chosen to receive Coke first and the rest randomly receive Pepsi first. Why? Coke was preferred 29 times.

1. Is there evidence to suggest that students have a taste preference between Coke or Pepsi? Carry out a hypothesis test by completing the following steps, and using a level of significance of 0.05:
   (a) State the null and alternative hypotheses
   (b) Calculate the test statistic
   (c) Calculate the p-value
   (d) State the conclusion

Decisions and Errors in Significance Tests

	Decision	
Truth	Reject H_0	Fail to Reject H_0
H_0 is True		
H_0 is False		

Note that

- if H_0 is true, the probability of a Type I error is the significance level α
- a Type II error can result if the sample size is too small. Why?

Theorem

If $X \sim N(\mu_X, \sigma_X)$, *and* $Y \sim N(\mu_Y, \sigma_Y)$, *and* X *and* Y *are independent, then*

$$aX + bY \sim N\left(a\mu_X + b\mu_Y, \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}\right)$$

In particular,

$$X - Y \sim N\left(\mu_X - \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2}\right)$$

Difference between proportions

Consider the null hypothesis $H_0 : p_1 - p_2 = 0$, where p_1 and p_2 are the population proportions for two independent populations.

Then under H_0, the test statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions of the two groups, n_1 and n_2 are the sample sizes, and p is the common population proportion, which is estimated in pratice by $p = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}$, and $Z \sim N(0,1)$ under H_0.

Example

In a survey of 2600 individuals, 110 out of 1120 males were red-green color-blind and 10 out of 1480 females were red-green colorblind. Is there a significant relationship (at $\alpha = 0.05$) between sex and red-green color-blindness? Answer this question by (1) stating the null and alternative hypotheses, finding the test statistic, finding the p-value, and stating the conclusion.

Central Limit Theorem

Recall that if a population has mean μ and standard deviation σ, then the sample mean $\overline{X_n} \sim \mathrm{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Standardization of the sample mean

Following from the CLT, let

$$Z = \frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Then $Z \sim \mathrm{N}(0, 1)$.

However, we cannot calculate Z in practice because we do not know σ.

If we replace σ with an estimate using the sample standard deviation s, and let

$$T = \frac{\overline{X_n} - \mu}{\frac{s}{\sqrt{n}}}.$$

Then T follows the Student's t distribution with $n - 1$ degrees of freedom.

The t distribution has the same shape as the standard normal distribution but has more area in its tails (i.e., a larger standard deviation). As the degrees of freedom increases, the t distribution approaches the standard normal distribution.

Hypothesis tests regarding population means are carried out by calculating t-statistics and their corresponding $p-$values.

General hypothesis testing steps

1. State the null (H_0) and alternative (H_1) hypotheses.
2. Calculate the appropriate test statistic, whose distribution is known under H_0.
3. Calculate the p-value.
4. Make a conclusion about whether there is sufficient evidence to reject H_0 and accept H_1 based on the p-value, and state this conclusion in the context of the problem.