Hypothesis Testing

Ex: An election with two candidates, A and B.

- Candidate A receives 42% of the vote
- Candidate B receives 58% of the vote

} true population percentages

- Candidate A is convinced more people voted for him, so he wants an investigation to see if the vote was rigged
- Candidate A hires a consulting agency to randomly sample 100 voters to record whether or not each person voted for him
- Suppose that 53 out of the 100 voted for him.
- 53% clearly exceeds 42%
- Was there fraud? We cannot be certain
- Even if the true pop. percentage voting for Candidate A is 42%, it is possible that out of a particular sample of 100 that 53 of the 100 voted for A.
- How strong is the sample evidence against the officially reported 42%?

One way to proceed is to set up an Hypothesis Test

- Let θ denote the true proportion of the population voting for A

- The hypothesis can be stated as

$$H_0 : \theta = .42$$

- This is an example of a Null Hypothesis (presumed innocent until proven guilty)
  ↳ assume H_0 true until rejected

- The alternative hypothesis is

$$H_1 : \theta > .42$$

- In order to conclude that H_0 is false and that θ_1 is true we must have evidence "beyond a reasonable doubt" against H_0

  - 43 out of 100 would not be enough to reject H_0

In hypothesis testing we can make 2 kinds of mistakes

- Type I: rejecting the Null Hypothesis when it is true (e.g. if we reject $H_0: \theta = .42$ when in fact it is true)
- Type II: failing to reject H_0 when it is in fact false (e.g. $\theta > .42$ but we fail to reject H_0)

- After we decide whether or not to reject H_0 we never know with certainty if we have committed an error

- We can compute the probability of making either a type I or type II error

- Hypothesis testing rules are constructed to make the probability of committing a type I error small.

- We define the significance level (or simply the level) of a test as the probability of a type I error. Typically denoted by α

$$\alpha = P(\text{Reject } H_0 \mid H_0)$$

- Classical statistics requires specifying a significance level for a test

- When we set α we are quantifying our tolerance for a type I error

- Common values are $\alpha = .10$, $\alpha = .05$, $\alpha = .01$

- Once we have chosen α we would like to minimize the probability of a type II error
- OR maximize the power of a test against all alternatives
- Power of a test

$$\pi(\theta) = P(\text{Reject } H_0 \mid \theta) = 1 - P(\text{Type II} \mid \theta)$$

where θ denotes the actual value of the parameter

Testing Hypotheses a/b the mean in a Normal Population

- A test statistic, T, is some function of the random sample
- A particular outcome is t (for a given sample)
- Given a test statistic we can define a rejection rule that determines when H_0 is rejected in favor of H_1
- Rejection rules are based on comparing the value of the test statistic, t, to a critical value, C

- The values of t that result in rejection are collectively known as the <u>rejection region</u>

- To determine the critical value, we must first decide on a significance level of the test

  - Then given α the critical value associated with α is determined by the distribution of T, assuming that H_0 is true

- Testing hypotheses about μ from a $N(\mu, \sigma^2)$ is straightforward

  - The Null

$$H_0 : \mu = \mu_0$$

  * where μ_0 is a value that we specify

  ** in most applications $\mu_0 = 0$

  - The rejection rule depends on the nature of the alternative. These alternatives of interest are:

$$\left.\begin{array}{l} H_1 : \mu > \mu_0 \\ H_1 : \mu < \mu_0 \end{array}\right\} \text{(one-sided alternatives)}$$

$$H_1 : \mu \neq \mu_0 \quad \text{(two-sided alternatives)}$$

- Consider $H_1: \mu > \mu_0$. We should reject H_0 when $\bar{y}$ is "sufficiently" greater than μ_0. How to what "sufficiently large" means?

- This requires knowing the probability of rejecting the null when it is true.

- $$t = \frac{\sqrt{n}(\bar{y} - \mu_0)}{s} = \frac{(\bar{y} - \mu_0)}{se(\bar{y})}$$

Where $se(\bar{y}) = s/\sqrt{n}$ is the standard error of $\bar{y}$

- Given a sample of data it is easy to obtain t

- Under the null hypothesis

$$T = \sqrt{n}(\bar{Y} - \mu_0)/S$$

has a t_{n-1} distribution

- Suppose we have chosen a 5% significance level. Then the critical value c is chosen so that

$$P(T > c \mid H_0) = .05$$

i.e. the probability of a Type I error is 5%

- Once we have c, the rejection region is

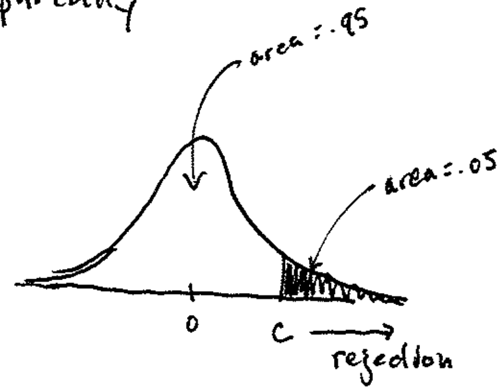
$$t > c$$

where c is the $100(1-\alpha)$ percentile in a t_{n-1} distribution.

* as a percent, the significance level is 100α%

* This is an example of a one-tailed test, because the rejection region is in one tail of the distribution

- Graphically



- $t = \dfrac{(\bar{y} - \mu_0)}{se(\bar{y})}$ is the t statistic.

  • It measures the distance of $\bar{y}$ from μ_0 relative to the $se(\bar{y})$

- Ex: In the population of cities granted enterprise zones in a particular state, let Y denote the percentage change in investment from the year before to the year after a city became an enterprise zone. Assume $Y \sim N(\mu, \sigma^2)$

$H_0: \mu = 0$ (enterprise zones have no effect on business investment)

$H_1: \mu > 0$ (enterprise zones have a positive effect)

- Want to test at the 5% level

$$t = \frac{\bar{y}}{s/\sqrt{n}} = \frac{\bar{y}}{se(\bar{y})}$$

- Suppose $n = 36$ cities. The critical value is $c = 1.69$
- We reject H_0 in favor of H_1 when $t > 1.69$
- Suppose $\bar{y} = 8.2$ and $s = 23.9$ then $t \approx 2.06$
- H_0 is rejected at the 5% level
- The 1% critical value is $c = 2.44$, so H_0 is not rejected at the 1% level

- The rejection region is similar for $H_1: \mu < \mu_0$. A test with a significance level of $100\alpha\%$ rejects H_0 against H_1 whenever

$$t < -c$$

- For two sided alternatives we must be careful so that the significance level is still α. If H_1 is given by $H_1: \mu \neq \mu_0$, then we reject H_0 if $\bar{y}$ is sufficiently far from zero in absolute value.

- A $100\alpha\%$ level test is obtained from the rejection rule

$$|t| > c$$

where $|t|$ is the absolute value of the t statistic

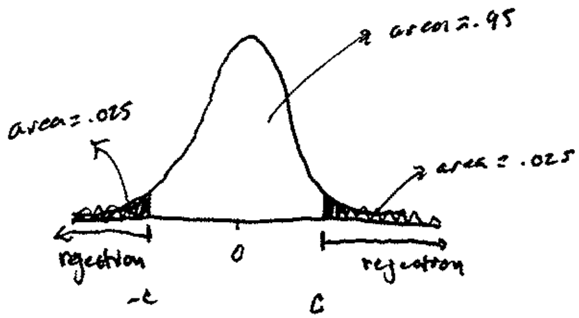
- C is the $100(1-\alpha/2)$ percentile in the t_{n-1} distribution

  * For example if $\alpha = .05$ then the critical value is the 97.5^{th} percentile in the t_{n-1} distribution

  * If $n = 22$ then $c = 2.08$ is the 97.5^{th} percentile in a t_{21} distribution

  * abs(t) must exceed 2.08 in order to reject H_0 against H_1 at the 5% level

- Graphically



- Asymptotic tests for non-normal populations
- Under the null

$$T = \sqrt{n}(\bar{Y} - \mu_0)/S \overset{a}{\sim} N(0,1)$$

- for large n we can compare the t-statistic with critical values from a standard normal distribution

Computing and Using P-values

* The requirement of choosing a significance level ahead of time means that different researchers at different times, using the same data and procedure to test the same hypothesis, could wind up with different conclusions.

  - Reporting our significance level helps this

Ask ourselves the question:

Q: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?

This value is known as the p-value of a test

Consider the problem of testing $H_0: \mu = 0$ in a $N(\mu, \sigma^2)$ population

Our test statistic is $T = \sqrt{n}\,\overline{Y}/S$ and we assume n is large enough to have a standard normal distribution under H_0:

Suppose the observed T, $t = 1.52$ (notice we skipped choosing α)

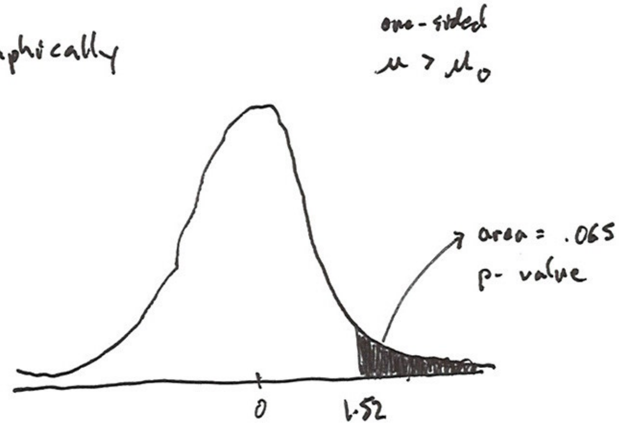
Now we can ask what is the largest significance ~~level~~ level at which we would fail to reject H_0:

This is the significance level associated with using t as our critical value.

$$\text{P-value} = P(T > 1.52 \mid H_0) = 1 - \Phi(1.52) \approx .065$$

where $\Phi(\cdot)$ denotes the standard normal CDF

Graphically



If we carry out the test at a significance level below 6.5% (~~e.g.~~ say 5%) we fail to reject. If we carry out the test for a significance level above 6.5% we reject. With the p-value in hand we can carry out the test for any significance level.

Generally small p-values are evidence against H_0; since they indicate that the outcome of the data occurs with small probability if H_0 is true.

If had been $t = 2.85$ then the p-value $= 1 - \Phi(2.85) = .002$.

This means that when H_0 is true we observe a value of T as large as 2.85 with probability .002