## Sets 25 and 26: Section 6.2 and 6.3, Hypothesis testing

**The logic of hypothesis testing:**

1. **The experimenter forms a *null hypothesis* $\mathrm{H}_0$ to test against an *alternative hypothesis* $\mathrm{H}_1$ (theory, guess, hope, research hypothesis).**
2. **The experimenter collects data.**
3. **"Are the data compatible with $\mathrm{H}_0$?" If yes, $\mathrm{H}_0$ is not rejected. If no, $\mathrm{H}_0$ is rejected.**

**Example: Consider a court of law where a defendent is accused of a crime.**

$\mathrm{H}_0$: **Defendant is innocent**
$\mathrm{H}_1$: **Defendant is guilty**

**Example: A new pain medication is being tested. The conjecture (theory, guess, hope, research hypothesis) is that its effective time is longer than 12 hours, which is the mean effective time of the old drug. 60 patients take the drug and the average effective time is 13.1 hours with $s = 3$.**

**Let $\mu$ = unknown mean of new drug**

**Conjecture:** $\mathrm{H}_1$: $\mu > 12$ **hours (research hypothesis)**

$\mathrm{H}_0$: $\mu = 12$ **hours, the effective time of the old drug**

**Two types of errors are possible:**

| Decision | $H_0$ True | $H_0$ False |
|---|---|---|
| Reject $H_0$ | Type I error | correct |
| Do not reject $H_0$ | correct | Type II error |

**Example: Pain medication example, errors.**

**Example: Court of law, errors.**

**We use the $p$-value approach to measure the weight of evidence against $H_0$.**

**The $p$-value is the probability of seeing results as extreme or more extreme (in the direction of $H_1$) than our actual observations if the null hypothesis were true.**

**The $p$-value is calculated using a test statistic.**

**If our observations are very unlikely to occur assuming $H_0$ is true, then we reject the null hypothesis. Otherwise, we keep the null hypothesis.**

## The p-value approach

1. **Define the parameters to be tested.**
2. **Define $H_0$ and $H_1$.**
3. **Specify the test statistic and the distribution assuming $H_0$ is true.**
4. **Find the observed value of the test statistic.**
5. **Find the $p$-value.**
6. **Report the strength of evidence against $H_0$:**

- ***Very strong*** **if** $p$**-value**$\leq 0.01$
- ***Strong*** **if** $0.01 < p$**-value**$\leq 0.05$
- ***Moderate*** **if** $0.05 < p$**-value**$\leq 0.1$
- ***Little or none*** **if** $0.1 < p$**-value.**

**Example: Consider the pain medication example.**

[**1. Define the parameter of interest**] $\mu$**, the mean effective time.**

[**2. Define** $\mathrm{H}_1$, $\mathrm{H}_0$] $\mathrm{H}_1$**:** $\mu > 12$, $\mathrm{H}_0$**:** $\mu = 12$

**Note: Although technically our** $\mathrm{H}_0$ **should be** $\mu \leq 12$**, we use** $\mu = 12$ **for simplicity; if there is strong evidence against** $\mu = 12$**, then there would certainly be strong evidence against for values of** $\mu$ **which are smaller than 12.**

[**3. Define the test statistic and distribution**]

**Since we have a large number of patients, and computed a sample mean** $\overline{x} = 13.1$ **and a sample standard deviation of** $s = 3$**, our test statistic is:**

$$z = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

**This test statistic has approximately a standard normal distribution.**

**[4. Find the observed value of the test statistic]**

**If the null hypothesis were true, then $\mu = 12$. The observed value of the test statistic is:**

$$z = \frac{13.1 - 12}{3/\sqrt{60}} \approx 2.84$$

**[5. Find the p-value, and draw conclusions]**

**The $p$-value is the probability of having a test statistic value as extreme or more as the one we observed:**

$$p\textbf{-value} = P(Z \geq 2.84)$$

**Using the standard normal table, we find $p$-value$= 0.002$. We conclude that there is strong evidence against $\mathrm{H}_0$.**

**[6. Report the strength of evidence against $\mathrm{H}_0$] There is very strong evidence against $\mathrm{H}_0$ in favour of $\mathrm{H}_1$. There is very strong evidence against the hypothesis that the new drug has the same mean effective life as the old drug.**

**The test we just performed is called a one-tailed test, since $\mathrm{H}_1$ specifies that $\mu$ is in one region. If our $\mathrm{H}_1$ specified that the parameter is in one of two regions, then we call the test a two-tailed test.**