Hypothesis testing

Suppose that a bulb manufacturer claims that its bulbs have an average lifetime of 1000 hours. In statistical terms this is equivalent to say tha the population mean equals 1000 hours. Since it's not possible to check all the bulbs produced to prove this claim we just take a sample and test it. Suppose an extreme example in which only 2 bulbs are tested and their average lifetime (sample mean) is found to be 900. Does this prove that the manufacturer's claim is false? On the contrary, if the lifetime of these 2 bulbs is found to be 1000, does this prove the claim true? Well, in both case the answer is probably not, because the sample is too small to make a reasonable decision based on it.

To analyze this problem we specify two hypothesis:

H_0, called the null hypothesis

H_1, called the alternative hypothesis

In the example above $H_0: \mu = 1000$
$H_1: \mu < 1000$

Typically we suppose H_0 is true, unless we find proof to the contrary. Now, some concepts arise by using this approach (independently on how we proceed).

Definition. Type I error. Is the error comitted when H_0 is true and we reject it.

Definition Significance level. Is the probability of type I error. Usually denoted by α.

Definition Type II error. Is the error comitted when H_0 is false and we accept it. Its probability is usually denoted by β.

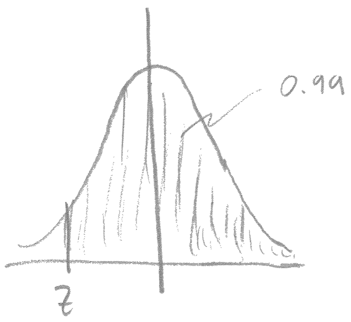
Definition Power. Is the probability of rejecting the null hypothesis when it's false. It's denoted by $1-\beta$.

	Accept	Reject
True	✓	I, α
False	II, β	✓

Back to the example of the bulbs, the idea would be to take a reasonable sized sample, compute the sample mean and reject H_0 if the sample mean is too small, but what exactly is too small? That depends on the distribution of the sample mean.

To be concrete suppose the lifetime X of the bulbs is normally distributed. If H_0 is true, then $X \sim N(1000, 100)$ (suppose we know $\sigma^2 = 100$ from previous studies). If from a sample of size 10 we determine that $\bar{X} = 990$, then

$$z_{obs} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{990 - 1000}{10/\sqrt{10}} = -3.16$$



Now, the value z for which $N(0,1)$ (recall $z \sim N(0,1)$) has accumulated 0.99^* of the area is $z = -2.33$ and since we observed -3.16 we reject the hypothesis that $\mu = 1000$ (the null hypothesis)

Observe that if $\bar{X} = 993 \Rightarrow z = -2.21$ and then the null hypothesis is not rejected

Also, if we choose an area of 0.9993 ($\alpha = 0.0007$) we don't reject.

$* \Rightarrow \alpha = 0.01$ significance

Definition. Test statistic. It's a statistic (a function of the sample) $T = T(X_1, X_2, .. X_n)$ that is used to decide if we reject the null hypothesis. In example $z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

Definition Rejection region. Is the set of values of the test statistic that lead to rejection of the null hypothesis. In the example $(-\infty, -2.33)$

Definition. Acceptance region. Is the set of values of the test statistic that lead to acceptance of the null hypothesis. In the example $(-2.33, \infty)$

Definition Null distribution. Is the probability distribution of the test statistic when we consider the null hypothesis true. In the example $z \sim N(1000, 100)$

Now we can outline a procedure:

1) Define H_0 and H_1
2) Fix significance level α and define the test statistic $T = T(X_1, X_2, \dots, X_n)$
3) With help of $\alpha = P_r(T \in C \mid H_0 \text{ is true})$ find the rejection region C (this requires knowing the null distribution).
4) Compute the test statistic and reject H_0 if $T \in C$

Example. It's claimed that more than 50 % of all the cars have wrongly adjusted front lights. Let p = real proportion of cars with wrongly adjusted front light in the population.

1) $H_0: p \leq 0.5$

$H_1: p > 0.5$

2) $\alpha = 0.05$. If we take a sample of n cars $X_1, X_2, ..., X_n$ and test for the front headlight then they're either righ or wrong, that is $X_i = \{0, 1\}$ $i = 1, 2, ..., n$ so we can use the test statistic $T = T(X_1, ..., X_n) = \sum_{i=1}^{n} X_i$ which is the number of cars with wrongly adjusted lights. (observe $\frac{T}{n}$ will give the proportion

3) Assume now H_0 is true, that means we assume $p \leq 0.05$ but it's indeed enough to assume $p = 0.5$ because, if we can reject the claim for this value, then we can reject it for any $p < 0.5$. This way we can assign a concrete distribution to the random variables in the sample, that is $T \sim Bin(n, 0.5)$

Suppose $n = 20$, we need to find C such that $Pr(T \in C \mid H_0 \text{ true}) = 0.05$

If we tabulate the binomial probabilities, or look them up in a table or use a computer, we can find that $Pr(T \leq 13) = \sum_{i=0}^{13} \binom{20}{i} 0.5^i (1-0.5)^{n-i} = 0.9423$

$\Rightarrow C = \{14, 15, ..., 20\}$ is the rejection region.

4) If $T_{obs} = 14, 15 ...$ or 20 then we say that we reject H_0 at 0.05 significance level

Since different significance levels define different rejection regions (that's why it's important to decide it in advance), sometimes the information is summarized as a p-value

Definition. p-value. Is the smallest significance level at which the null hypothesis would be rejected, that is p-value $= Pr(T \geq t \mid H_0 \text{ true})$

In the previous example, $Pr(T > 13 \mid H_0 \text{ true}) = 0.05$, as we knew it.

In the first example, $Pr(Z \leq -3.16 \mid H_0 \text{ true}) = 0.0007$

Example. To reduce car's fuel consumption some additives like aluminium are used. A certain car modell can run 26 miles/gallon (mpg) in average with std of $\sigma = 5$ mpg with the current type of fuel. It's claimed that a new fuel type with more aluminium can decrease fuel consumption.

Let μ be the average distance the car can run per gallon. Then we test $H_0: \mu \leq 26$ with a sample of size $n = 36$ which yields $\bar{X} = 28.04$

$H_1: \mu > 26$

It is enough to reject for $H_0: \mu = 26$. Under this null hypothesis

$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$. $T_{obs} = \frac{28.04 - 26}{5/\sqrt{36}} = 2.45$

At significance level $\alpha = 0.05$ $z(1-\alpha) = z(1-0.05) = z(0.95) = 1.65$

Since $T_{obs} > 1.65$ we reject the null hypotesis

p-value $= Pr(T > 1.65 \mid H_0) = 1 - Pr(T < 1.65 \mid H_0) = 1 - \Phi(2.45)$

$= 1 - 0.9929 = 0.0071$

which means we can reject $H_0: \mu = 26$ for $\alpha = 0.0071$ and larger, therefore we reject (the smaller the p-value the more evidence we have against the null hypothesis)