

PC7 – Hypothesis testing

Octobre 22 2018

Unlike point estimators or confidence interval which give numerical values, we may be interested in answering yes/no questions or choose between two contradictorial theories. Hypothesis testing answer these kind of questions. The two theories (or the yes and the no), called hypotheses, won't have the same role. One will be called the null hypothesis H_0 and the other the alternative hypothesis H_1 . The way we choose between the two hypotheses is conceptually similar to the way a jury deliberates in a court trial. H_0 is going to be the defendant: just as the latter is presumed innocent until "proven" guilty, so is the null hypothesis "accepted" unless the data argue overwhelmingly to the contrary. In short, H_1 should be the hypothesis you want to prove.

Introduction to hypothesis testing

Exercise 1: Hypothesis testing and doping controls

During a sport meeting, J.C. and S.R. are subject to an unannounced doping control. Doctors measure their hematocrit levels in their blood. Normally this rate is equal to $\tau_0 = 45\%$ but it can be increased by taking some drug. The measure of this rate is assumed to be Gaussian with a standard deviation of 2. The observed value of J.C. is 48 and the one of S.R. is 50. We want to know if these values are abnormal, i.e. that they have taken a drug, or if they just are the result of the imprecision of the measurements.

1. The **first step** for solving a hypothesis testing is to write the **statistical model**. Precise the statistical model for the doping control of J.C. and S.R..
2. The **second step** is to choose and write the **hypotheses**. They are formally written as equations on the parameters of the model. To choose between H_0 and H_1 , we apply the same rule as in a court trial with H_0 being typically a statement reflecting the status quo. And we give to H_0 the benefit of the doubt. Write the hypotheses for the doping control.
3. The hypotheses are written thanks to parameters of the model, but as always in statistics, these parameters are not known. We then use an estimator to approximate these parameters. The **third step** consists in choosing an **estimator** (a test statistic) for the parameter which is tested and precise its distribution under H_0 (and H_1 if you can). Propose an estimator for the parameter you want to test for J.C. and S.R..
4. The **fourth step** is to choose the shape of the **rejection region**. Looking at the hypotheses, you can choose in which qualitative case an estimator can be considered as too extreme and tilts the balance to the side of the culpability of the defendant. Propose a shape for the rejection region.
5. The **fifth step** is to compute the **boundaries** of the rejection region using the chosen **significance level**. In the following, we want to choose τ_c such that we reject H_0 if the measured rate is larger than this threshold τ_c .
 - a. You propose to reject H_0 as soon as the measured rate is larger than 45. What is the probability of wrongly concluding that an athlete is doped?
 - b. The previous probability of accusing an innocent person is far too huge: no jury would convict a defendant knowing it had such a chance of sending an innocent person to jail! So you decide to change the rule and propose to reject H_0 as soon as the measured rate is larger than 60. What is the probability of wrongly concluding that an athlete is doped?
 - c. Setting τ_c that large err in the other direction by giving the null hypothesis too much benefit of the doubt. We then decide to choose τ_c such that the probability of wrongly concluding that an athlete is

doped, i.e. the probability of the type II risk, equals 0.05. Find such threshold τ_c . 0.05, usually denoted α is called the level of significance of the test, it the probability to reject H_0 while it's true.

6. The **sixth step** is to write the test, i.e. summarizing the previous steps by precisising the **rule of rejection for H_0** and then **conclude considering the observed values**.
 - a. Summarize the previous conclusions into a decision rule.
 - b. What are your conclusions for J.C. and S.R.
 - c. When you observe a measure leading to the rejection of H_0 , it doesn't prove that the athlete is doped. In other words, it must be remembered that rejecting H_0 does not prove that H_0 is false, any more than a jury's decision to convict guarantees that the defendant is guilty. It just means that if the true rate is 45, measurements of this rate as large or larger than τ_c are expected to occur only 5% of the time. Because of that small probability, a reasonable conclusion when the measurement of this rate is as large or larger than τ_c is that the true rate is larger than 45. Check that with your rule you wrongly reject H_0 with an average frequency of 0.05.
7. As a senventh step, you could check the power of the test at the end. You could also compute the p-value (PC8) for the observed value.
 - a. Plot the power function associated to your hypotheses test.
 - b. What is the probability of detecting an abnormal hematocrit level when it is equal to 50?

Note that the choices for steps 3 and 4 are optimal (best power) using Neyman-Pearson theory in the case of simple hypotheses. We will discuss it in PC8.

Exercise 2: Hypothesis testing and biased coin

You are arguing with a friend about the movie you are about to watch. This friend decides to make this choice by tossing a coin. Before doing so, you want to check that the coin is not biased.

1. You propose to flip the coin 10 times and keep the coin if the number of heads is 5. Write the statistical model and the hypotheses you are testing. What is the rejection region and the acceptance region? Compute the significance level of this test and plot its power function.
2. You repeated the previous strategy twice on two of your four coins and you rejected the two first coins. Since the rejection rate under H_0 (0.75) is too high, you decide to change your strategy. You propose another rule for the two last coins: if the number of heads among 10 tosses is smaller or equal to 1 or larger or equal to 9 then you use another coin. What is the rejection region and the acceptance region? Compute the significance level of this test. Plot the power function. What is the probability that you accept the coin if it is actually biased with a probability of 3/4 (and 1/4) to obtain a head?
3. Your friend disagrees with your rule because he is superstitious and hates number 2. He would prefer to accept the coin if there are 1, 3, 4, 5, 6, 7, 8 or 9 heads. What is the rejection region and the acceptance region? Compute the significance level of this test. Plot the power function. What is the probability that you accept the coin if it is actually biased with a probability of 3/4 (and 1/4) to obtain a head?
4. You finally propose to throw 30 times the coin and reject the coin if the number of heads X is such that $|X - 15| > \delta$ for some $\delta > 0$. Propose a hypothesis testing at level smaller than and as close as possible to 0.05. Plot its power. What is the probability that you accept the coin if it is actually biased and with a probability of 3/4 (and 1/4) to obtain a head? What is your conclusion if you obtain 4 heads?
5. Comparing the power function of the four previous decision rules, which test would you choose and why?

Choice of the hypotheses

Exercise 3: Test for an organic certification

To obtain an “organic” certification, a manufacturer of organic products has to guarantee that each batch contains less than 1% of GMO. Then he collects $n = 25$ products by batch and test if they contain less than 1% of GMO. Let X_i be the logarithm of the GMO level (in %) of the product i .

Model : We assume that the X_i are independent and are distributed from a Gaussian distribution $\mathcal{N}(\theta, 1)$.

1. As to the manufacturer, the GMO level is less than 1% unless the contrary is proved. He wants to test the hypothesis $H_0 : \theta \leq 0$ against $H_1 : \theta > 0$. The two hypotheses are composite and not simple. He wants that the test fails with probability less than 5% when $\theta \leq 0$. Compute a threshold $a_{25,5}$ such that

$$\sup_{\theta \leq 0} \mathbb{P}_\theta(\bar{X}_{25} > a_{25,5}) = 0.05,$$

where $\bar{X}_{25} = \frac{1}{25} \sum_{i=1}^{25} X_i$. You can use that $\mathbb{P}(Z > 1.645) \simeq 0.05$, when $Z \sim \mathcal{N}(0, 1)$.

2. An anti-GMO organization wants to be sure that the GMO level in organic products are indeed less than 1%. In particular, it wants to know if the test succeeds in eliminating products which contains more than 50% the legal maximum. What is the probability that the test does not reject H_0 when the GMO level is 1.5%?
3. Shocked by the previous result, the organization advocates for a test which really proves that the GMO level is less than 1%. The organization supporters think that the GMO level is greater than 1% unless the contrary is proved. Then H_0 is $\theta > 0$ and H_1 is $\theta \leq 0$. Propose a hypothesis test H_0 against H_1 such that the probability that the test wrongly rejects H_0 is less than 0.05.
4. Regulatory agencies accept the hypothesis test proposed by the manufacturer ($H_0 : \theta \leq 0$ against $H_1 : \theta > 0$) but require to detect at least 80% of the cases where the GMO level exceeds 10% of the legal maximum. The significance level is still 5%, what should the manufacturer do to adhere to the law?

Exercise 4: Various experiments

Formalize the hypotheses and statistical models in the following cases:

1. Defeated in his most recent attempt to win a congressional seat because of a large gender gap, a politician has spent the last two years speaking out in favor of women’s rights issues. A newly released poll claims to have contacted a random sample of 120 of the politician’s current supporters and found that 72 were men. In the election that he lost, exit polls indicated that 65% of those who voted for him were men. The politician wants to know if his campaign has has a positive effect on women.
2. A herbalist is experimenting juices extracted from berries and roots that may have the ability to affect the Stanford-Binet IQ scores of students afflicted with mild cases of attention deficit disorder (ADD). A random sample of twenty-two children diagnosed with the condition have been drinking Brain-Blaster daily for two months. Past experience suggests that children with ADD score an average of 95 on the IQ test with a standard deviation of 15.
3. A company sold a sampler producing uniform random digits. Its client suspects the sampler to be biased towards small digits and wants to complain. The client pays an external control society to test this sampler. They obtain the following sample 1, 2, 1, 2, 0, 8, 6, 1, 2, 4, 5, 1, 2, 4, 8, 4, 4, 3, 0, 2.

Gaussian models

Exercise 5: Gaussian hypothesis testing

Let

$$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$$

be Gaussian observations.

1. We first assume that σ is known and we want to test $H_0: \mu = \mu^*$ versus $H_1: \mu \neq \mu^*$ for $\mu^* \in \mathbb{R}$.

- a. Propose a test at level 0.05 (steps 1 to 6 of Exercise 1) and create an R function **rej_reg_1** that takes as input a vector of observations $x = (x_1, \dots, x_n)$, the variance σ^2 , the parameter μ^* and a level of significance $\alpha \in (0, 1)$, and outputs a vector of size four containing the lower and upper bound of the acceptance region, the statistic of the test and the result of the test: 0 if H_0 is not rejected and 1 if H_0 is rejected.
- b. Check the significance level of the test through simulations when $\mu^* = 0$, $\sigma = 2$, $\alpha = 0.05$ and $n = 10$.
- c. Plot the power function of this test under the same condition.
2. We do not assume that σ is known anymore and we want to test $H_0: \mu = \mu^*$ versus $H_1: \mu \neq \mu^*$ for $\mu^* \in \mathbb{R}$.
 - a. Propose a test at level 0.05 (steps 1 to 6 of Exercise 1) and create an R function **rej_reg_2** that takes as input a vector of observations $x = (x_1, \dots, x_n)$, the parameter μ^* and a level of significance $\alpha \in (0, 1)$, and outputs a vector of size four containing the lower and upper bound of the acceptance region, the statistic of the test and the result of the test: 0 if H_0 is not rejected and 1 if H_0 is rejected.
 - b. Check the significance level of the test through simulations when $\sigma = 2$, $\alpha = 0.05$ and $n = 10$.
3. The athlete S.R. of the first exercise contests the control. He goes through another control where 10 measures of some other level of his blood are done. A usual measure is 0, and doping product can increase or decrease this level. The measures are assumed to be Gaussian but its standard deviation is not known. The following measurements are observed: 2.1, 1.4, -0.1, -0.2, 1.2, -0.7, 0.7, 1.5, 0.5, 0.3. What is the conclusion of the control for S.R.?

Other models

Exercise 6: Urn

A urn contains N balls, numbered from 1 to N where $N \in \{99, 101\}$ is unknown. The following balls, which were placed back in the urn after each drawing, were drawn:

12	52	48	60	17	03	25	98	15	51	12	38	76
25	64	87	23	19	05	82	46	73	09	50	48	34

Propose a rule to test $H_0: N = 99$ versus $H_1: N = 101$, following all the steps described in Exercise 1. Besides compute the probability of type II error of your decision rule.

Exercise 6: Are financial professionals unscrupulous?

Roger read in the newspaper that 20% of financial professionals think they need to break the law to succeed: <http://www.slate.fr/story/101785/wall-street-enfeindre-loi-reussir>

He thinks that this percentage is underestimated and wants to investigate and to test his hypothesis. He only has two financial professionals in his address book; He is going to ask them the following question: “Do you think you need to break the law to succeed?”.

1. Assuming that Roger’s conclusion only depends on the answers at this question, which decision rule can he establish? What is the probability that Roger is wrong if the percentage of financial professionals is indeed 20%?
2. Roger wants to wrongly conclude that the percentage is strictly greater than 20% with probability 5%. He realizes that he does not need to survey anyone to realize this constraint: he just has to put 19 black chips and 1 red chip in his pocket. Why? What is the drawback of this method?
3. Roger now chooses a card in a deck of 32 cards and write the result before surveying the two financial professionals he knows. Why did he choose this method? What is probability that Roger’s conclusion is wrong if the percentage of financial professionals is indeed 20%?
4. Roger feels that such a decision should not be taken on a random basis. What should he do?