Statistics Applied to Economics

Degree in Economics

F.Tusell

Dpto. Economía Aplicada III (Estadística y Econometría)

Curso 2012–2013

eman ta zabal zazu

Universidad del País Vasco   Euskal Herriko Unibertsitatea

Índice I

Índice II

Logically equivalent statements (I)

- "If an animal is a whale, it lives in the water."
- What can be inferred for animals which live in the water?
- And for animals which do **not** live in the water?
- $\underbrace{\text{Is a whale}}_{p} \Longrightarrow \underbrace{\text{Lives in the water}}_{q}$
- $\underbrace{\text{Does not live in water}}_{\neg q} \Longrightarrow \underbrace{\text{Is not a whale}}_{\neg p}$

Logically equivalent statements (II)

Quite generally,

- $p \implies q$ and $\neg q \implies \neg p$ are logically equivalent. ($\neg$ above stands for negation:)
- Both are true or false.
- When testing hypothesis, we rely on a softened versions of this equivalence.

Statements probabilistically related (I)

- Consider $p \implies$ **most of the time** q.
- Then $\neg q \implies \neg p$ **is likely** (or p is unlikely).
- Same structure, only now the implications are not required to hold all times.
- $\neg q$ is no longer proof of $\neg p$, *but can be taken as evidence in favour of it.*

Statements probabilistically related (II)

Example:

- $\underbrace{\text{Coin is regular}}_{p} \implies$ **most of the time** $\underbrace{\text{about 50\% of heads}}_{q}$.
- $\underbrace{\text{Far from 50\% of heads}}_{\neg q} \implies \underbrace{\text{Coin not regular}}_{\neg p}$ **is likely**.
- $\underbrace{\text{Far from 50\% of head}}_{\neg q}$ is taken as evidence in favour of $\neg p$ (and therefore against p).

Hypothesis testing (I)

- A **null hypothesis** is an statement which we hold to be true.
- If it is indeed true (p), a given experiment should very likely produce a result in a certain range (q).
- If it so happens that the result is not observed in the very likely range ($\neg q$), either:
  1. Something very strange has happened (should not be the case very often)...
  2. ...or else the null hypothesis is not true to begin with.
- As statisticians, we go with the second option.

Hypothesis testing (II)

- Empiricism!
- If the experiment does not quite agree with the hypothesis, we scrap the hypothesis.
- *However,* we cannot completely rule out the possibility that something strange happened. We are bound to make errors!
- But we try to keep those to a minimum.

The anatomy of a hypothesis test (I)

- As already mentioned, a hypothesis is a conjecture.
- A **statistical hypothesis** is usually phrased in terms of the values of one or more parameters.
  1. The mean of a distribution is $m = 0$, (one parameter).
  2. Two distributions have the same mean: $m_1 = m_2$, (two parameters).
  3. Two characters are independent: $p_{ij} = p_{i.} \times p_{.j}$.
- Equivalently, a hypothesis is phrased by stating that a parameter vector belongs to a subset Θ_0 of the entire feasible space Θ.

How would you phrase the hypothesis in items 1 and 2 above?

1) $\Theta_0 = 0$, $\Theta = \mathcal{R}$. 2) $\Theta_0 = \{(x, y) : x = y\}$, $\Theta = \mathcal{R}^2$

The anatomy of a hypothesis test (II)

- In order to test the *null hypothesis* H_0, we use as evidence the information contained in a sample. We usually condense that information using a *test statistic*, $S = S(\vec{X})$.
- We better use a sufficient statistic!
- To be useful, that test statistic must have a known distribution under H_0. This is required, so that we can tell when a sampled value is "rare" under H_0.
- The decision procedure then is:

  Reject H_0 if the sampled value of S is "rare", do not reject otherwise.
- What is "rare"? Problem dependent.

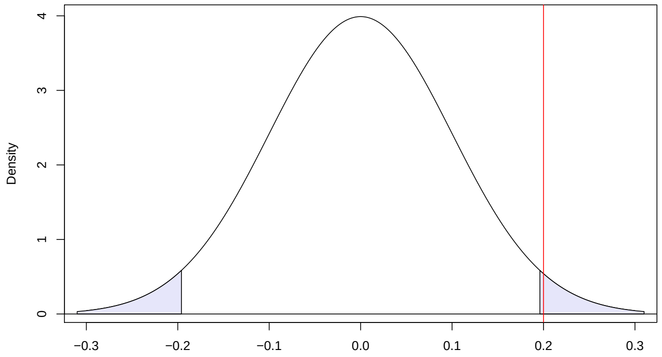
The anatomy of a hypothesis test (III)

Example:

- We believe the mean of a $N(m, \sigma^2 = 1)$ distribution to be zero (H_0). A sample of $n = 100$ observations gives $\overline{X} = 0.20$.
- We are willing to reject the hypothesis if the evidence found is among the 5% "rarest" events that could happen under H_0. What will be our decission?
- The events that we decide constitute evidence against H_0 is called the critical region.
- The probability of the critical region when H_0 is true, is called the significance level.

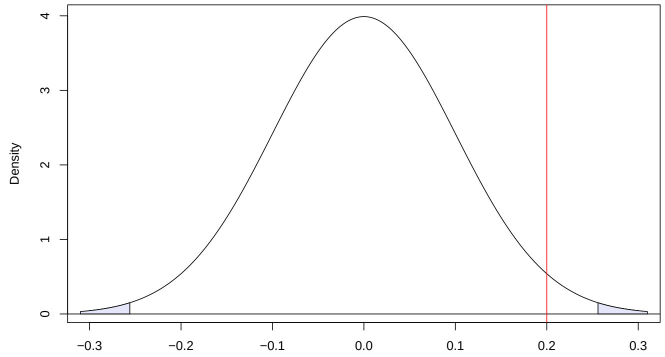
The anatomy of a hypothesis test (IV)

At the stated level of significance (5%), we would reject H_0.



The anatomy of a hypothesis test (V)

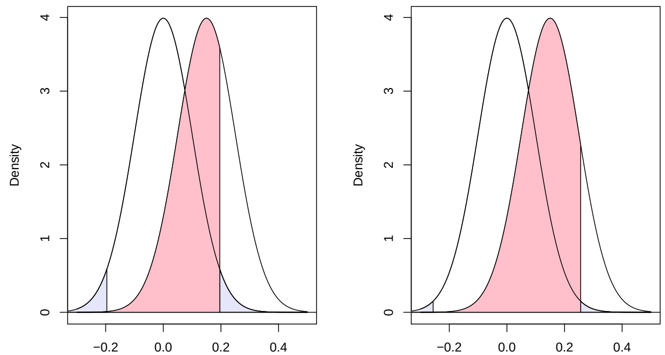
With a different level of significance (1%), we would **not** reject H_0.



The trade-off between Type I and Type II errors

- The significance level α is the probability of unduly rejecting H_0.
- We should choose α considering how "grave" or "costly" is such an error, called *Type I error*.
- If we make α very small (an hence the critical region very small also), we will almost never reject H_0 . . .
- . . .even when we would like to, because it is false!
- Not rejecting H_0 when it is false is called *Type II* error, and its probability is denoted by β.

Trade-off between Type I and II errors - Ilustration

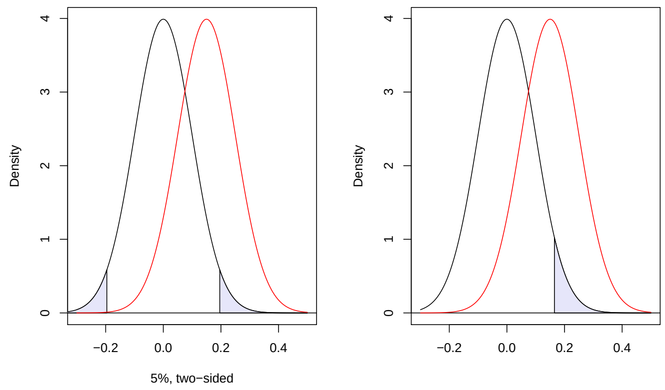


Pure significance tests

- We are only considering so far H_0.
- We are looking at empirical evidence to see it it “contradicts” H_0.
- When it does, we reject H_0.
- Sometimes, we have a clear idea of what the “competing” hypothesis is, and in this case we want to use that information.

Testing against an alternative H_a

If we test H_0 against an alternative H_a, a one-sided critical region makes more sense.



Optimal critical regions for H_0 vs. H_a

The usual procedure is:

- Fix α, the probability of unduly rejecting H_0.
- Among all critical regions of size α, find the one which minimizes β (or, equivalently, maximizes $1-\beta$, the *power*).
- When both H_0 and H_a are *simple* ($=$ fix completely the distribution of the test statistic), a simple procedure exists, base on Neyman-Pearson's theorem.
- In other cases, a unique most powerful test may not exist.

The Neyman-Pearson theorem (I)

- After fixing the significance level α, what critical region would give better power against a simple alternative?
- Let's consider testing $H_0 : \theta = \theta_0$ vs. $H_a : \theta = \theta_a$:

x	0	1	2	3	4	5
$P(x;\theta_0)$	0.60	0.26	0.05	0.04	0.04	0.01
$P(x;\theta_a)$	0.10	0.15	0.10	0.25	0.30	0.10

How would you choose a critical region of size $\alpha = 0.05$ with maximum power?

Picking $x = 4$ and $x = 5$, for a total power of 0.40.

The Neyman-Pearson theorem (II)

- The intuition is that we want our critical region to be made of points x with high ratio

$$\frac{f(x;\theta_a)}{f(x;\theta_0)}$$

where $f(x;\theta_0)$ is the density under the null and $f(x;\theta_a)$ is the density under the alternative.
- Neyman-Pearson theorem: *The most powerful test of given size α for $H_0 : \theta = \theta_0$ against the alternative $H_a : \theta = \theta_a$ has critical region of the form:*

$$C_\alpha = \left\{ \vec{x} : \frac{f(\vec{x};\theta_a)}{f(\vec{x};\theta_0)} > k_\alpha \right\}$$

for a constant k_α which depends on α.

The Neyman-Pearson theorem - Proof (I)

- Consider the critical region

$$C_\alpha = \left\{ \vec{x} : \frac{f(\vec{x};\theta_a)}{f(\vec{x};\theta_0)} > k_\alpha \right\}$$

and any other α-size region A_α.
- C_α and A_α will in general overlap. Dropping the α subscript:

$$\int_C f(\vec{x};\theta_0)d\vec{x} = \int_A f(\vec{x};\theta_0)d\vec{x} = \alpha$$

- Subtracting $\delta = \int_{C\cap A} f(\vec{x};\theta_0)d\vec{x}$ in both sides:

$$\int_{C\cap A^c} f(\vec{x};\theta_0)d\vec{x} = \int_{A\cap C^c} f(\vec{x};\theta_0)d\vec{x} = \alpha - \delta \geq 0$$

How do we know $\alpha - \delta \geq 0$?

Because $C \cap A \subseteq C$.

The Neyman-Pearson theorem - Proof (II)

- The difference of powers of the two critical regions is:

$$\int_C f(\vec{x};\theta_a)d\vec{x} - \int_A f(\vec{x};\theta_a)d\vec{x}$$

- Inside C we have $f(\vec{x};\theta_a) > kf(\vec{x};\theta_0)$ and outside $f(\vec{x};\theta_a) \leq kf(\vec{x};\theta_0)$. The difference of powers is:

$$\begin{aligned}
\int_C f(\vec{x};\theta_a)d\vec{x} \quad - & \quad \int_A f(\vec{x};\theta_a)d\vec{x} \\
= & \quad \int_{C\cap A^c} f(\vec{x};\theta_a)d\vec{x} - \int_{A\cap C^c} f(\vec{x};\theta_a)d\vec{x} \\
\geq & \quad k\int_{C\cap A^c} f(\vec{x};\theta_0)d\vec{x} - k\int_{A\cap C^c} f(\vec{x};\theta_0)d\vec{x} \\
= & \quad k(\alpha-\delta) - k(\alpha-\delta) = 0
\end{aligned}$$

Neyman-Pearson example (I)

- In a large company, the number of workers not showing up for work is Poisson-distributed. Workers claim that $\lambda = 1$, while management claims $\lambda = 2$. They check four days and obtain 1, 0, 2, and 2 workers not showing up for work.
  1. Obtain the most powerful critical region to test the workers hypothesis (H_0) against the management's at a 0.05 significance level.
  2. What is the power of the test?
- We have:

$$\begin{aligned}
f(\vec{x};\lambda=1) &= \prod_{i=1}^{4} \frac{e^{-1}1^{x_i}}{x_i!} = \frac{e^{-4}}{\prod_{i=1}^{4} x_i!} \\
f(\vec{x};\lambda=2) &= \prod_{i=1}^{4} \frac{e^{-2}2^{x_i}}{x_i!} = \frac{e^{-8}2^{\sum_{i=1}^{4} x_i}}{\prod_{i=1}^{4} x_i!}
\end{aligned}$$

Neyman-Pearson example (II)

- From Neyman-Pearson, the most powerful critical region of size α is of the form:

$$\begin{aligned} C_\alpha &= \left\{ \vec{x} : \frac{f(\vec{x}; \lambda = 2)}{f(\vec{x}; \lambda = 1)} > k_\alpha \right\} \\ &= \left\{ \vec{x} : \frac{e^{-8} 2^{\sum_{i=1}^4 x_i}}{e^{-4}} \right\} \\ &= \left\{ \vec{x} : e^{-4} 2^{\sum_{i=1}^4 x_i} > k_\alpha \right\} \end{aligned}$$

- Taking logs and bringing all constants into k'_α:

$$C_\alpha = \left\{ \vec{x} : \sum_{i=1}^4 x_i > k'_\alpha \right\}$$

Neyman-Pearson example (III)

- We now know **the form** of C_α

$$C_\alpha = \left\{ \vec{x} : \sum_{i=1}^4 x_i > k'_\alpha \right\}$$

- Have no clue about what the value of k'_α is, but know $\sum_{i=1}^4 x_i \sim \mathcal{P}(\lambda = 4)$ when H_0 is true.
- For C_α to have size $\alpha = 0.05$, the constant must be a value exceeded with probability no greater than α when sampling a $\mathcal{P}(\lambda = 4)$ distribution. Resorting to tables (or R) gives us:

```
> ppois(0:8, lambda = 4)
[1] 0.01832 0.09158 0.23810 0.43347 0.62884
[6] 0.78513 0.88933 0.94887 0.97864
```

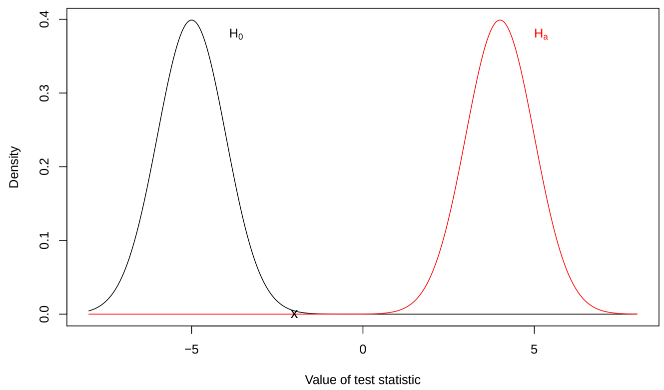
- $[8, \infty)$ would be a critical region for $S = \sum_{i=1}^4 x_i$ quite close to $\alpha = 0.05$; $[9, \infty)$ would have $\alpha = 0.02136$.

Some quirks of hypothesis testing (I)

- Very non symmetric role of null and alternative hypothesis.
- Management could have replied the worker's representative: "Why don't we test as null *our hypothesis* and not yours?
- If evidence is not strong, the null is the surviving hypothesis, whichever it happens to be!
- The null should be provisionally established knowledge, put to test. How we arrive to that knowledge, there is no telling.
- Alternative approaches (like bayesian inference) treat conjectures in a more symmetric way.

Some quirks of hypothesis testing (II)

- That H_0 is rejected **does not mean that H_a should be accepted.**



- An observation at X is evidence against H_0 but much more so against H_a. In such situation, we should revise our hypothesis and admit that other possibilities might exist.