

10: HYPOTHESIS TESTING

Larsen & Marx 6.1-6.4

Prof Amanda Luby

So far, we've treated *inference* as either *point estimation* or *interval estimation*. In some experimental settings, however, we don't want to draw a numerical conclusion but rather evaluate two competing theories. For instance, we may wish to know whether a candidate for political office is likely to win or lose; whether a new vaccine is effective or ineffective; or whether a policy intervention improves or does not improve quality of life for citizens.

The process of dichotomizing possible conclusions from an experiment and using probability theory to choose one over the other is called *hypothesis testing*. We have two competing hypotheses:

- 1.
- 2.

Parameter Space

Simple and Composite Hypotheses

1 Decision Rule

Example: A high school was chosen to participate in the evaluation of a new geometry and algebra curriculum. In the recent past, the school's students were considered "typical", receiving scores on standardized tests that were very close to the nationwide average. In the year of the study, 86 sophomores were randomly selected to participate in a special set of classes that interated geometry and algebra. Those students averaged 502 on the SAT-I math exam; the nationwide average was 494 with a standard deviation of 124. Did the curriculum improve scores?

Decision Rule:

What if?

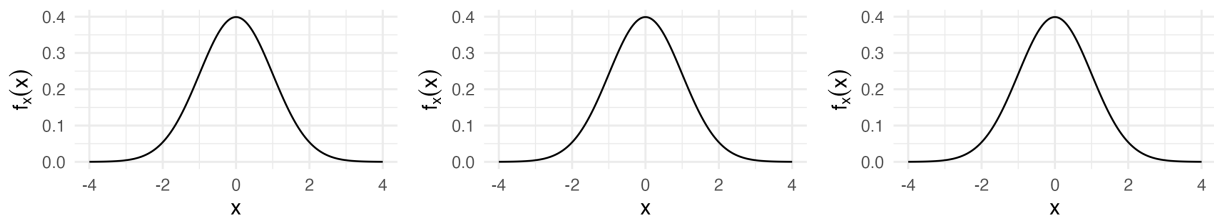
Mapping to Z-scores:

Test statistic

Critical Region

Critical Value

2 One-sided vs Two-sided Alternatives



3 P-Values

There are two ways to quantify the amount of evidence against H_0 in a given dataset. The first involves defining a *level of significance* (α), identifying a corresponding *critical region*, and reject H_0 if the *test statistic* falls within the critical region. Another strategy is to calculate the p-value:

P-value

Example:

4 Non-normal data

Up to this point, we've assumed that we're working with the normal distribution and setting up a hypothesis test for a mean. Decision rules for other probability distributions are rooted in the same basic principles.

In general, to test $H_0 : \theta = \theta_0$, where θ is an unknown parameter in $f_x(x; \theta)$, we define the decision rule in terms of $\hat{\theta}$, a sufficient statistic for θ . We want to set up the decision rule such that the probability of rejection if the null hypothesis is true is equal to α .

Example: Four measurements (k_1, \dots, k_4) are taken of a Poisson random variable X (so $p_x(k; \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$) and we wish to test $H_0 : \lambda = 0.8$ against $H_1 : \lambda > 0.8$.

K	P(X = K)
0	0.041
1	0.130
2	0.209
3	0.223
4	0.178
5	0.114
6	0.061
7	0.028
8	0.011
9	0.004
10	0.001
11	0.000
12	0.000
13	0.000
14	0.000
15	0.000