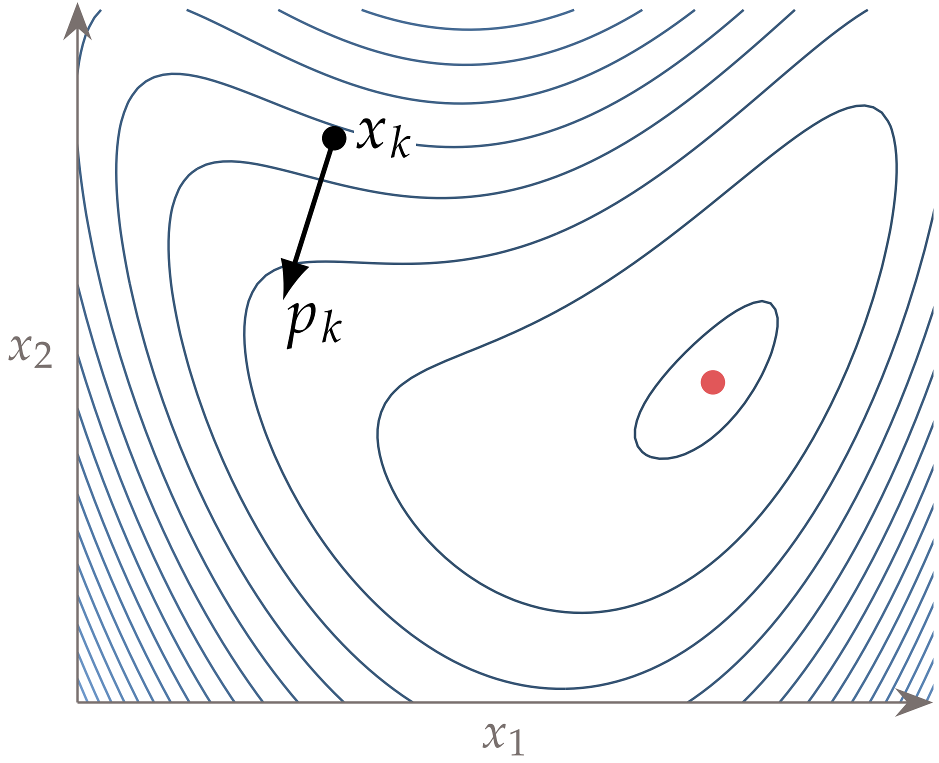
Hypothesis Tests (single sample)



ME EN 275
Andrew Ning
aning@byu.edu

Motivation

You work in a factory that produces that produces widgets that weigh 5 kg on average. You suspect something is miscalibrated and widgets now weigh more than they should.

Motivation

You work in a factory that produces that produces widgets that weigh 5 kg on average. You suspect something is miscalibrated and widgets now weigh more than they should.

You measure 30 random widgets and find them to be 5.1 kg. Has the mean shifted or is that just random variation? What if you measure it to be 5.5 kg or 6 kg or 7 kg? At what point is the shift significant enough that you can reasonably conclude that something has changed?

Definitions

null hypothesis H_0: no difference. status quo baseline.

alternative hypothesis H_a: opposite of null

significance level α: ~(1 − confidence interval)

p-value.

$\alpha = 0.05$

$p = 0.1$ $(p > \alpha)$ → don't reject null

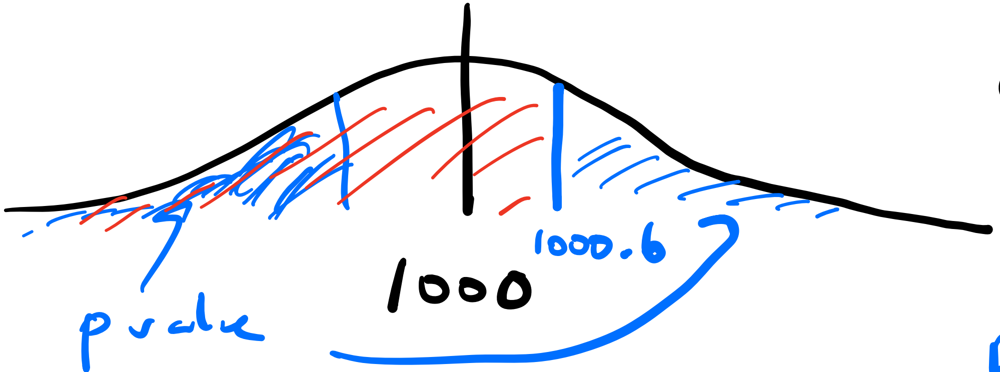
$p = 0.01$ $(p < \alpha)$ → reject null

Example

A scale is to be calibrated by weighing a 1000 g test weight 60 times. The 60 scale readings have mean 1000.6 g and standard deviation 2 g.

$H_0: \mu = 1000g$ $\qquad$ $H_a: \mu \neq 1000g$

$\sigma = \frac{s \leftarrow 2g}{\sqrt{n}}$



tail = $1 - \text{norm.cdf}\left(1000.6, 1000, \frac{2}{\sqrt{60}}\right)$

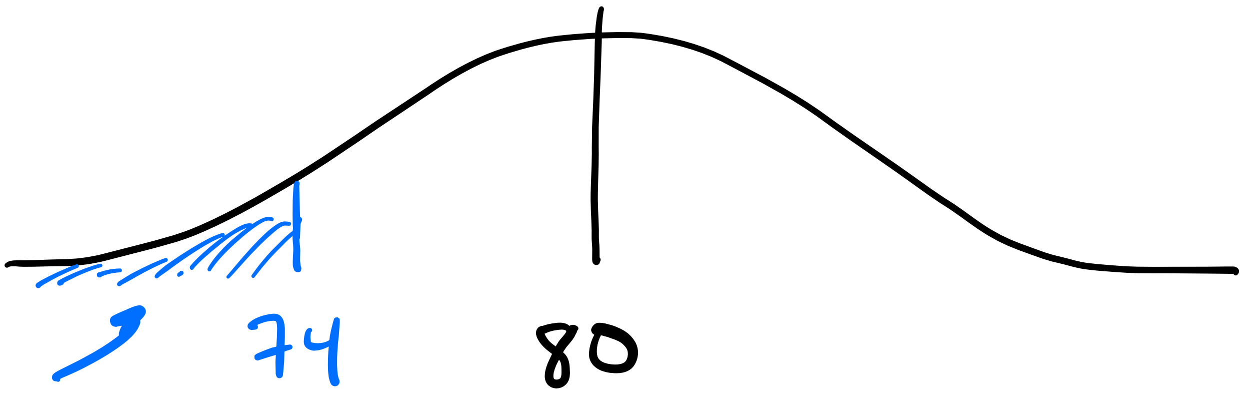
$p = 2 \cdot \text{tail} = 0.02$

Example

A new coating has become available that is supposed to reduce the wear on a certain type of rotary gear. The mean wear on uncoated gears is known from long experience to be 80 μm per month. Engineers perform an experiment to determine whether the coating will reduce the wear. They apply the coating to a simple random sample of 60 gears and measure the wear on each gear after one month of use. The sample mean wear is 74 μm, and the sample standard deviation is s = 18 μm.

$H_0: \mu \geq 80$ µm $\qquad H_a: \mu < 80$ µm

what is null hypothesis and alternative hypothesis?



p-value

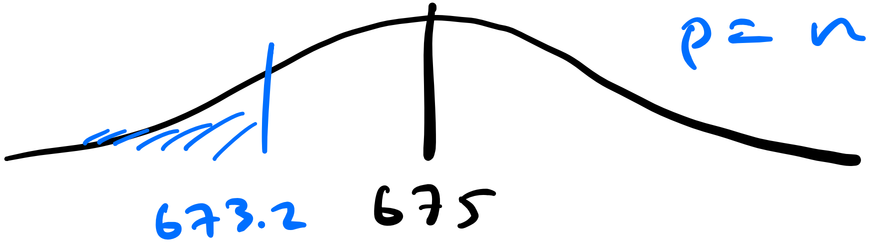
$$p = \text{normcdf}\left(74, 80, \frac{18}{\sqrt{60}}\right) = 0.0049$$

Example

$H_0: \mu \le 675$ $H_a: \mu > 675$

In one ~~experiment~~, 45 steel balls lubricated with purified paraffin were subjected to a 40 kg load at 600 rpm for 60 minutes. The average reduction in diameter was 673.2 μm, and the standard deviation was 14.9 μm. Assume that the specification for a lubricant is that the mean reduction be less than 675 μm.

$H_0: \mu \ge 675\ \mu m$ $H_a: \mu < 675\ \mu m$



$$p = \text{norm.cdf}\left(673.2,\ 675,\ \frac{14.9}{\sqrt{45}}\right)$$

$$= 0.209$$

Example

A more formal procedure

Choose null hypothesis appropriately

Specifications for steel plate to be used in the construction of a certain bridge call for the minimum yield (Fy) to be greater than 345 MPa.

They will select a random sample of steel plates, measure their breaking strengths, and perform a hypothesis test. The steel will not be used unless the engineers can conclude that $\mu > 345$.

$H_0: \mu \leq 345$ or $H_0: \mu \geq 345$

$H_a: \mu > 345$ $\quad$ $H_a: \mu < 345$

reject H_0: I use steel $\quad$ reject null: don't use steel

don't reject H_0: don't use steel $\quad$ don't reject null: don't use steel

Significance level

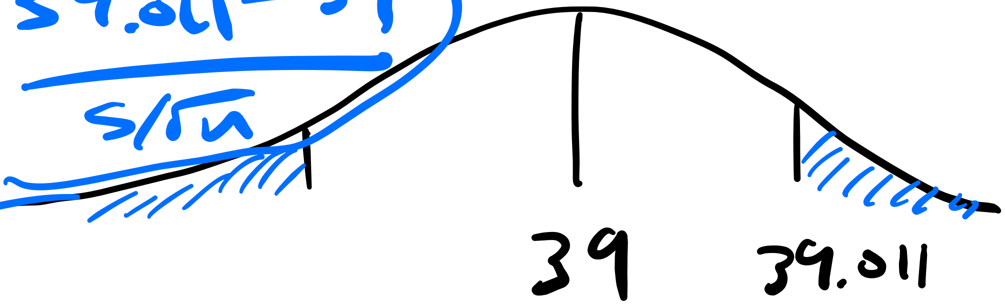
The process that manufactures spacer collars for a transmission countershaft are supposed to be calibrated so that the mean thickness is 39.00 mm. A sample of six collars is drawn and measured for thickness. The six thicknesses are 39.030, 38.997, 39.012, 39.008, 39.019, and 39.002. Assume that the population of thicknesses of the collars is approximately normal. Can we conclude that the process needs recalibration?

$x = [39.030,\ 38.997,\ \ldots]$

$xbar = np.mean(x) = 39.0113\ldots$

$s = np.std(x, ddof=1) = 0.011927\ldots$

$tscore = \dfrac{39.011 - 39}{s/\sqrt{n}}$



$H_0: \mu = 39$

$H_a: \mu \neq 39$

$t.cdf(tscore, df)$

$tail = 1 - t.cdf(xbar, 5, 39, s/\sqrt{6})$

statistic   df   mean   std.

$p = 2 \cdot tail = 0.0674$

Type I and Type II Errors

		reality	
		H_0 true	H_0 false
decision	reject H_0	type I error	correct
	don't reject H_0	correct	type II error

Type I and Type II Errors