Further Remarks about Hypothesis Tests

Engineering Statistics
Section 8.5

Josh Engwer

TTU

18 April 2016

PART I:

STATISTICAL SIGNIFICANCE VS PRACTICAL SIGNIFICANCE

Statistical Significance

Definition

Given a population with parameter θ.
Take a sample $\mathbf{x} := (x_1, x_2, \ldots, x_n)$ and compute unbiased estimate $\widehat{\theta}$ for θ.

When a hypothesis test rejects the null hypothesis $H_0 : \ \theta = \theta_0$,
one says the sample departure from θ_0, $|\widehat{\theta} - \theta_0|$ is **statistically significant**.

Recall the large-sample test statistic value used for testing $H_0 : \ \mu = \mu_0$

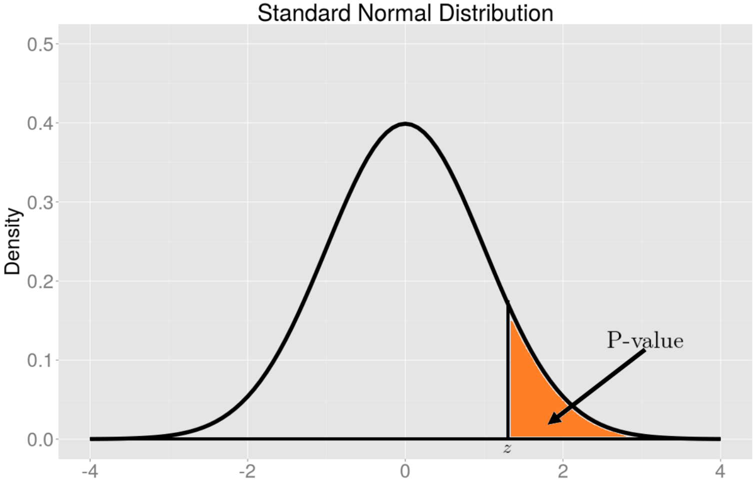
$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$$

Then, notice that as the sample size n increases, the test statistic gets <u>more extreme</u>, which means the resulting P-value <u>decreases</u>.

This means that as the sample size n increases,
it's more likely that a given hypothesis test will reject the null hypothesis H_0.

Statistical Significance

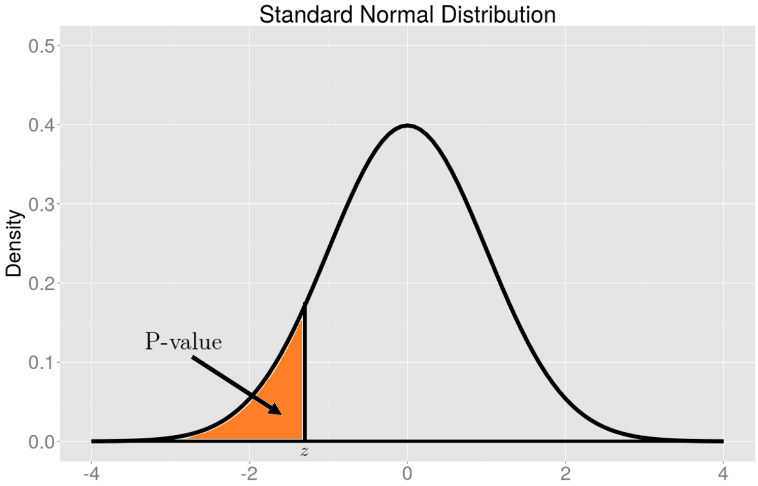
Hypothesis Test: $H_0 : \mu = \mu_0$, $H_A : \mu > \mu_0$ (One-Sided z-Test)



P-value $:= \mathbb{P}(\mathbf{W}(X;\mu_0) \geq \mathbf{W}(x;\mu_0)) \approx \mathbb{P}(Z \geq z) = 1 - \mathbb{P}(Z \leq z) = 1 - \Phi(z)$

Statistical Significance

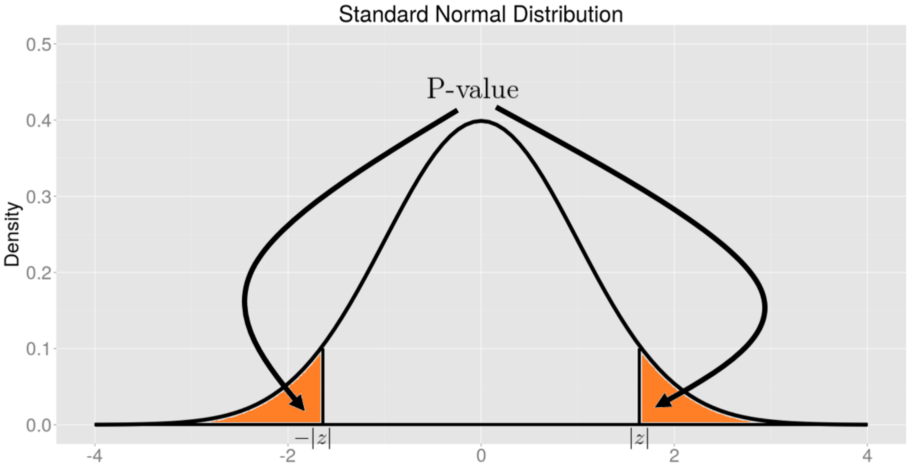
Hypothesis Test: $H_0: \mu = \mu_0$, $H_A: \mu < \mu_0$ (One-Sided z-Test)



P-value $:= \mathbb{P}(\mathbf{W}(X;\mu_0) \leq \mathbf{W}(x;\mu_0)) \approx \mathbb{P}(Z \leq z) = \Phi(z)$

Statistical Significance

Hypothesis Test: $H_0 : \mu = \mu_0$ $H_A : \mu \neq \mu_0$ (Two-Sided z-Test)



P-value $\approx \mathbb{P}(|Z| \geq |z|) = \mathbb{P}(Z \leq -|z|) + \mathbb{P}(Z \geq |z|) = 2 \cdot [1 - \Phi(|z|)]$

Statistical Significance vs Practical Significance

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

So how do you decide if this claim is reasonable or not??
TAKE A SAMPLE!

You randomly select a sample of <u>100 bottles</u> from your warehouse stock.
You measure the amount of soda in each bottle and compute the mean:

$$\bar{x}_{soda} = 16.6 \text{ oz}$$

Would you believe the stores' claim?? Yes, probably so!

$$\begin{matrix} H_0: \ \mu = 20 \\ H_A: \ \mu < 20 \end{matrix} \quad \begin{matrix} \text{would likely} \\ \text{reject } H_0 \end{matrix} \quad \left(\begin{matrix} \text{3.4 oz change } \underline{\text{is}} \text{ statistically significant} \\ \text{3.4 oz change } \underline{\text{is}} \text{ practically significant} \end{matrix} \right)$$

Statistical Significance vs Practical Significance

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

So how do you decide if this claim is reasonable or not??
TAKE A SAMPLE!

Suppose instead you randomly select a sample of 100 bottles from your stock.
You measure the amount of soda in each bottle and compute the mean:

$$\bar{x}_{soda} = 19.8 \text{ oz}$$

Would you believe the stores' claim?? No, probably not!

$H_0: \ \mu = 20$ would likely accept H_0 (0.2 oz change is not statistically significant)
$H_A: \ \mu < 20$ (0.2 oz change is not practically significant)

Statistical Significance vs Practical Significance

You run a manufacturing plant that bottles soda and sells them to stores.
You have an assembly line & a soda dispenser that is calibrated to dispense 20 ounces of soda into a 20 oz bottle.
In 2014, the dispenser breaks down beyond repair, so you buy a new one.
However, by the second quarter of 2015, you receive several complaints from stores that most of the bottles have noticeably less than 20 ounces of soda!!

Since complaints never occur prior to this, $\mu_{soda} = 20$ oz
But these irate stores are claiming that $\mu_{soda} = 15$ oz (!!)

So how do you decide if this claim is reasonable or not??
TAKE A SAMPLE!

Suppose instead you randomly select a sample of 100,000 bottles from stock.
You measure the amount of soda in each bottle and compute the mean:

$$\bar{x}_{soda} = 19.99 \text{ oz}$$

Would you believe the stores' claim?? No, probably not!

$H_0: \mu = 20$ would likely
$H_A: \mu < 20$ reject H_0

(0.01 oz change is statistically significant
0.01 oz change is not practically significant)

Statistical Significance vs Practical Significance

THE MORAL OF THE STORY:

Taking far too large of a sample may cause the hypothesis test to be too sensitive to an insignificant sample departure, $|\widehat{\theta} - \theta_0|$, from the hypothesized value θ_0 (!!!)

PART II:

POWER FUNCTION OF A HYPOTHESIS TEST

Power Function of a Hypothesis Test

So, how to determine how "good" a hypothesis test is for testing parameter θ?

Definition

Given one of the these hypothesis tests:

$$H_0: \theta = \theta_0 \quad H_0: \theta = \theta_0 \quad H_0: \theta = \theta_0$$
$$H_A: \theta < \theta_0 \quad H_A: \theta \neq \theta_0 \quad H_A: \theta > \theta_0$$

Then the **power function** of the test, denoted $K(\theta)$, is defined by:

$$K(\theta^*) := \mathbb{P}(\text{Reject } H_0 \text{ when in reality } \theta = \theta^*) = 1 - \beta(\theta^*)$$

where $\beta(\theta^*) := \mathbb{P}(\text{Accept } H_0 \text{ when in reality } \theta = \theta^*)$

Textbook Logistics for Section 8.5

- Skip "The Relationship between CI's & Hypothesis Tests" (pg 353-354)
  - Remember that α in CI's & α in Hypothesis Tests are the same α.
  - $\alpha = 0.05$-test of $\begin{array}{l} H_0: \ \theta = \theta_0 \\ H_A: \ \theta \neq \theta_0 \end{array}$ is equiv to whether 95% CI for θ contains θ_0
  - In practice, nobody ever "converts" a hypothesis test to a CI or vice versa.
  - Read this section after the course if interested.
- Ignore "Simultaneous Testing of Several Hypotheses" (pg 354-355)
- Ignore the "Likelihood Ratio Principle" section (pg 355-356)
  - The test statistics used in this Chapter can be formally found using the **Likelihood Ratio Test** (**LRT**) statistic.
  - The LRT statistic is formed using the **Maximum Likelihood Estimator** (**MLE**), which was discussed in section 6.2

Fin

Fin.