1-sample Hypothesis Tests

Statistics 251: Statistical Methods

Module 9

Updated 2020

Introduction

We have learned about estimating parameters by point estimation and interval estimation (specifically confidence intervals). More often than not, the objective of an investigation is not to estimate a parameter but to decide which of two (or more) contradictory claims about the parameter is correct.

This part of statistics is called *hypothesis testing*

Terms

Statistical hypotheses is a claim or assertion about

(1) The value of a single parameter
(2) The values of several parameters
(3) The form of an entire probability distribution

Hypotheses

(1) Null hypothesis, denoted by H_0, is the claim that is initially assumed to be true (the "prior belief" or "historical" claim)
(2) Alternative hypothesis, denoted by H_a, is the assertion that is contradictory to H_0; it is a researcher's claim, what they are trying to prove (thus the reason behind the study)

Hypothesis Testing Checklist

All tests include the following four steps:

(1) State hypotheses, check assumptions
(2) Calculate the test statistic
(3) Find the rejection region
(4) Results and conclusion of the test

Hypotheses

When stating the hypotheses, the notation used is always population parameter notation; inferences upon populations need population notation (the Greek letters)

μ for the mean and π for the proportion

Hypotheses for μ

Hypotheses for inferences concerning means (regardless of whether or not σ is known

$$H_0 : \mu = \mu_0 \text{ vs. } H_a : \mu \neq \mu_0$$
$$H_0 : \mu \geq \mu_0 \text{ vs. } H_a : \mu < \mu_0$$
$$H_0 : \mu \leq \mu_0 \text{ vs. } H_a : \mu > \mu_0$$

Most often the null hypothesis will have = while the alternative will be one of either $\neq$, $>$, or $<$. μ_0 is a specified value (a number that is given in the problem)

Hypotheses for π

Hypotheses for inferences concerning proportions:

$$H_0 : \pi = \pi_0 \text{ vs. } H_a : \pi \neq \pi_0$$
$$H_0 : \pi \geq \pi_0 \text{ vs. } H_a : \pi < \pi_0$$
$$H_0 : \pi \leq \pi_0 \text{ vs. } H_a : \pi > \pi_0$$

Most often the null hypothesis will have = while the alternative will be one of either $\neq$, $>$, or $<$. π_0 is a specified value (a number that is given in the problem)

Assumptions

(1) Independence: observations are independent from one another
(2) Randomization: proper randomization was used
    - Takes care of independence issue if there is one
(3) Normality
    (a) Means need an *approximate* normal distribution ($n \geq 30$ should take care of it)
    (b) Proportions need $n \geq 60$ (via CLT)

If assumptions are violated, the results from the analyses are not valid nor reliable

Test Statistic

1-sample test of the mean μ when σ is known: Use Z

$$z = \frac{\overline{X} - \mu_0}{se_{mean}} \; ; \; se_{mean} = \frac{\sigma}{\sqrt{n}}$$

1-sample test of the proportion p: Use Z

$$z = \frac{\hat{\pi} - \pi_0}{se_{\pi}} \; ; \; se_{\pi} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

1-sample test of the mean μ when σ is unknown: Use t

$$t = \frac{\overline{X} - \mu_0}{se_{mean}} \; ; \; se_{mean} = \frac{s}{\sqrt{n}}$$

Rejection Region

Is based on significance level α. $\alpha = 1 - CL$ where CL is the confidence level

Always assume $\alpha = 0.05$ unless specified otherwise)

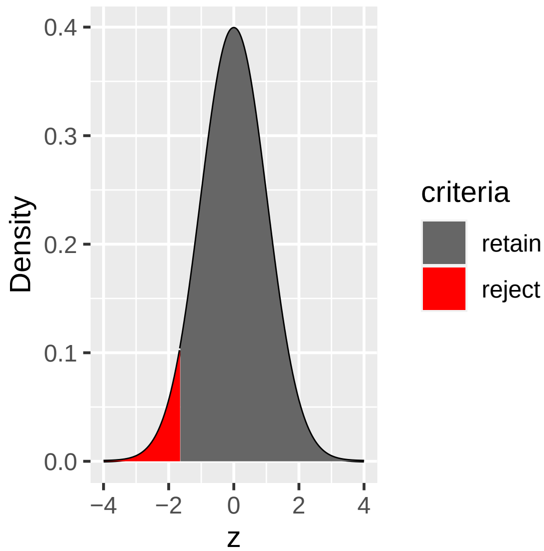
Two methods for rejection:

(1) Critical value approach
(2) *pvalue* approach

The alternative hypothesis (H_a) determines rejection based on where you are at on the curve

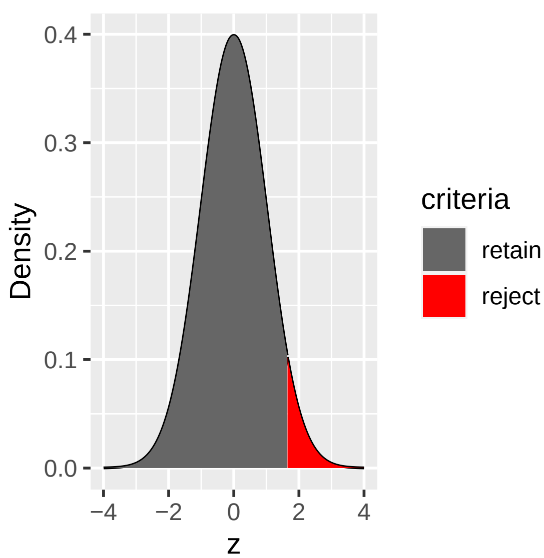
Critical Value Approach $H_a :<$

Reject H_0 iff (if and only if) $z_{calc} \leq z_\alpha$ (z_{calc} will most likely be a negative value and z_α *must* be negative)



Critical Value Approach $H_a :>$

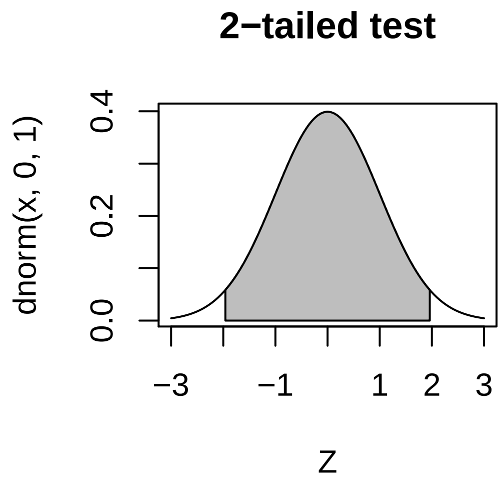
Reject H_0 iff $z_{calc} \geq z_\alpha$ (z_{calc} will most likely be a positive value and z_α *must* be positive)



Critical Value Approach $H_a :\neq$

Reject H_0 iff $|z_{calc}| \geq |z_{\alpha/2}|$ (z_{calc} and z_α can both be either positive or negative, but we will deal with absolute values)

(white area is the rejection region, and yes there are two area here that are *both* the rejection region)



Results and Conclusion

- Results: we either
  - Reject H_0 (rejecting the null hypothesis in favor of the alternative)
  - Fail to reject H_0 (we are not rejecting the null hypothesis so that means that the null hypothesis gives a reasonable explanation of the question at hand) Conclusion: explain what the results did in relation to the actual data

Example test of μ with known σ

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130°F. It is known from previous studies that the temperatures are normally distributed with standard deviation 1.5°F. A sample of $n = 9$ systems, when tested yields a sample average activation temperature of 131.08°F. Is there sufficient evidence that the true mean activation temperature is more than what the manufacturer claims?

Example test of μ with unknown σ (use s)

New York City, NY is known as the "city that never sleeps". A random sample of 25 New Yorkers was taken and they were asked how much sleep they get per night. Given the data, is there evidence that New Yorkers sleep is different from the "norm"; a full 8 hours of sleep? From the sample, the mean was 7.73 hours with standard deviation 0.77 hours.

Example test of π

Ingots are huge pieces of metal often weighing more than 10 tons (20,000 lbs.). They must be cast in one large piece for use in fabricating large structural parts for cars and planes. If they crack while being made, the crack can propagate into the zone required for the part, compromising its integrity; metal manufacturers would like to avoid cracking if at all possible. In one plant, only about 80% of the ingots have been defective-free. In an attempt to reduce the cracking, the plant engineers and chemists have tried some new methods for casting the ingots and from a sample of 500 ingot cast in the new method, 16% of the casts were found to be defective (cracked). Is there sufficient evidence that the defective rate has decreased?

pvalue logistics I

The *pvalue* of a test is the probability that, *given* the null hypothesis (H_0) is true, the results from another random sample will be as or more extreme as the results we observed from our sample.

The *pvalue* of the test is dependent on the type of test you are doing, as in one-tail upper, one-tail lower, or two-tail. The sign of the alternative hypothesis is the determining factor in calculation of the *pvalue*.

$pvalue$ **logistics II**

The pvalue approach; the null hypothesis can be rejected iff (if and only if) $pvalue \leq \alpha$ (with $\alpha = 0.05$ most often). This does not change, regardless of the sign of the alternative hypothesis. However, the calculation of the $pvalue$ is dependent on the sign of the alternative hypothesis. The $pvalue$ will be the $P($ the results of the test $|H_0$ is correct), in other words, it is the probability that the results would occur by random chance if the null hypothesis is actually correct.

Assume that $\alpha = 0.05$ unless specified; any rejection of H_0 means that the results (of experiment, survey, etc.) are significant.

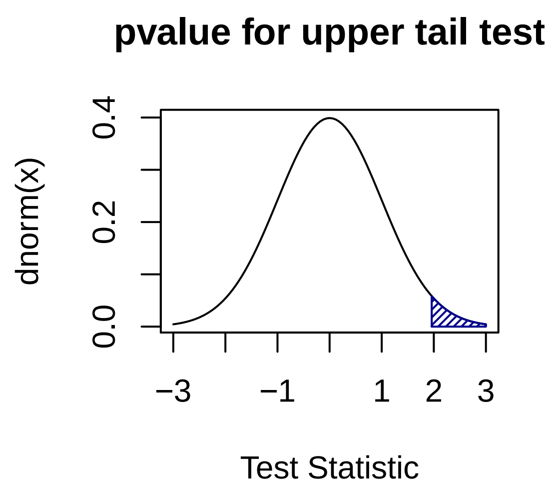
$$pvalue \leq \alpha \Rightarrow Reject\ H_0$$

H_a : $>$ **upper tail test**

Note that while all examples are with z, it is interchangeable with t (df is needed)

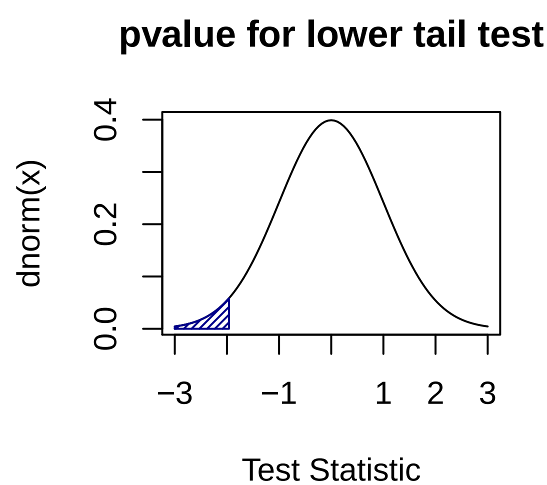
In this case, $pvalue$ represents the rejection region in the right tail of the distribution.

$$pvalue = P(Z \geq z_{calc}) = 1 - P(Z \leq z_{calc})$$



H_a : $<$ **lower tail test**

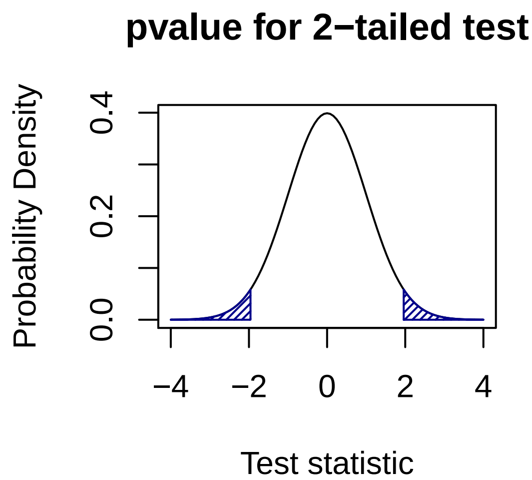
$$pvalue = P(Z \leq z_{calc})$$



H_a : $\neq$ **two tail test**

$$pvalue = 2[P(Z \leq z_{calc})] \ or \ 2[1 - P(Z \leq z_{calc})]$$

$$= 2[1 - P(Z \leq |z_{calc}|)]$$



pvalue rejection Examples

(1) $pvalue = 0.4$ with $\alpha = 0.05$. Since $pvalue = 0.4 \nleq \alpha(0.05)$, H_0 is not rejected (fail to reject H_0). There is a 40% chance that we would see these results due to random chance (dumb luck) if the null hypothesis is correct; results are not significant.

(2) $pvalue = 0.04$ with $\alpha = 0.05$. Since $pvalue = 0.04 \leq \alpha(0.05)$, H_0 is rejected. There is a 4% chance that we would see these results due to random chance (dumb luck) if the null hypothesis is correct; results are significant.

(3) $pvalue = 0.04$ with $\alpha = 0.01$. Since $pvalue = 0.04 \nleq \alpha(0.01)$, H_0 is not rejected. There is a 4% chance that we would see these results due to random chance (dumb luck) if the null hypothesis is correct; results are not significant.

pvalue Hand calculations I

(1) $z_{calc} = -1.99$, $H_a :<$, $pvalue = P(Z \leq z_{calc}) = 0.0233$. Since $pvalue = 0.0233 \leq \alpha(0.05)$, H_0 is rejected. There is a 2.33% chance that we would see these results due to random chance (dumb luck) if the null hypothesis is correct; results are significant.

(2) $z_{calc} = -1.99$, $H_a :\neq$, $pvalue = 2P(Z \leq z_{calc}) = 0.0466$. Since $pvalue = 0.0466 \leq \alpha(0.05)$, H_0 is rejected. There is a 4.66% chance that we would see these results due to random chance (dumb luck) if the null hypothesis is correct; results are significant. Note that the *pvalue* is almost equal to α, but as long as $pvalue \leq \alpha \Rightarrow Reject\ H_0$.

pvalue Hand calculations II

(3) $z_{calc} = 1.99$, $H_a :>$ and $\alpha = 0.01$. $pvalue = P(Z \geq z_{calc}) = 1 - P(Z \leq z_{calc}) = 1 - 0.9767 = 0.0233$. Since $pvalue = 0.0233 \nleq \alpha(0.10)$, H_0 is not rejected. There is a 2.33% chance that we would see these results due to random chance (dumb luck) if the null hypothesis is correct; results are not significant.

Errors

Type I=$\alpha = P(reject\ H_0|H_0\ true)$. This is a conditional probability statement that reads as "the probability of rejecting the null given that the null is true."
TLDR; we rejected a true null hypothesis (that's a bad thing).
Type I can only happen when H_0 is rejected

Type II=$\beta = P(Fail\ to\ reject\ H_0|H_0\ false)$. This is a conditional probability statement that reads as "the probability of not rejecting the null given that the null is false."
TLDR; we kept a false hypothesis (again, a bad thing).
Type II can only happen when H_0 is not rejected

Power=$1 - \beta = P(reject\ H_0|H_0\ false)$. This is a conditional probability statement that reads as "the probability that the null is rejected given that it is false."
TLDR; we correctly rejected H_0 when H_0 is false (a good thing. Finally!)

Error table

		The truth	
		H_0 true	H_0 false
My decision	Reject H_0	Type I (α)	:-)
	Fail to reject H_0	:-)	Type II (β)

Answers: Sprinklers

$H_0 : \mu = 130$ $H_a : \mu > 130$ (assumptions are met: independent, random sample with normal distribution so the sample size requirement is not needed here $\Rightarrow$ use z)

Provided information: $\overline{x} = 131.08$, $\sigma = 1.5$, $n = 9$, $se = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{9}} = 0.5$, $\mu_0 = 130$

Test statistic: $z = \frac{\overline{x}-\mu}{se} = \frac{131.08-130}{0.5} = 2.16$

Answers: Sprinklers con't

Rejection region (critical value): Since H_a :>, it is an upper-tail test (also called a one-tailed test). α will give the probability (area) and we need to find the z-score that is associated with that area. $\alpha = 0.05$, so we need to find $z_{0.05} = 1.645$. We can reject H_0 iff $z_{calc} \geq z_\alpha$.

Results and conclusion: $2.16 \geq 1.645$ so we will reject H_0. With H_0 rejected, there is evidence the mean activation temperature is higher than the manufacturer's claim of $130°F$ (maybe their manufacturing equipment need recalibration?).

Answers: Ingots

$H_0 : \pi = 0.2$ $H_a : \pi < 0.2$ (assumptions are met: independent, random sample with large sample $n = 500$, proportions always use z)

Provided information: $\hat{\pi} = 0.16$, $n = 500$, $se = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{(0.2)(0.8)}{500}} = 0.0179$, $\pi_0 = 0.20$

Test statistic: $z = \frac{\hat{\pi}-\pi_0}{se} = \frac{0.16-0.2}{0.0179} = -2.23$

Answers: Ingots con't

Rejection region (critical value): Since H_a :<, it is an lower-tail test (also called a one-tailed test). α will give the probability (area) and we need to find the z-score that is associated with that area. $\alpha = 0.05$, so we need to find $z_{0.05} = -1.645$. We can reject H_0 iff $z_{calc} \leq z_\alpha$.

Results and conclusion: $-2.23 \leq -1.645$ so we will reject H_0. With H_0 rejected, there is evidence the new casting method's defective rate is significantly less than the current method.

Answers: NY sleep

$H_0 : \mu = 8$ $H_a : \mu \neq 8$ (assumptions are met: independent, random sample with small sample and unknown $\sigma |Rightarrow$ use t).

Provided information: $\overline{x} = 7.73$, $s = 0.77$, $n = 25$, $se = \frac{s}{\sqrt{n}} = \frac{0.77}{\sqrt{25}} = 0.154$, $\mu_0 = 8$

Test statistic: $t = \frac{\overline{x}-\mu}{se} = \frac{7.73-8}{0.154} = 2.16$

Answers: NY sleep con't

Rejection region (critical value): Since H_a :$\neq$, it is an two-tailed test. $\alpha/2$ will give the probability (area) and we need to find the t-score that is associated with that area and its *df* (degrees of freedom). $\alpha/2 = 0.05/2 = 0.025$, so we need to find $t_{0.025,24} = 2.064$. We can reject H_0 iff $|t_{calc}| \geq |t_{\alpha/2,df}|$.

Results and conclusion: $|-1.753| \ngeq |2.064|$ so we will fail to reject H_0. With H_0 not rejected, there is not enough evidence that New Yorkers get something other than the "8 hours" of sleep.