

# Class 3c: Review of concepts in Probability and Statistics

Business Forecasting



# Hypothesis Testing

# Hypothesis Testing

Let's go back to our question. Can we get a higher price if cleanliness level is above 4.5?

This type of questions can be answered with hypothesis testing where we try to see which one of competing hypothesis is more likely to be true.

- **Hypothesis** is a claim about a parameter or a distribution.

## 1. $H_0$ **Null Hypothesis:**

- Claim to be tested, skeptical perspective
  - Prices of clean and dirty apartments are equal
  - $H_0 : \mu_c = \mu_d$

## 2. $H_A$ **Alternative Hypothesis:**

- An alternative idea under consideration
  - Prices of clean apartments are higher than prices of dirty apartments
  - $H_A : \mu_c > \mu_d$
- Assume that the null is true and see if there is enough of data to reject the null in favor of the alternative.

# Hypothesis Testing

Some examples of hypotheses testing:

- You released a new promotion to some customers (10% off) and you want to test if people who got the promotion spend different amount of money than those who didn't
  - $H_0$ : People who got the promotion spend the same as people who didn't get the promotion
    - $H_0 : \mu_p = \mu_n$
  - $H_A$ : People who got the promotion spend different amount than people who didn't get the promotion
    - $H_A : \mu_p \neq \mu_n$
- You designed a new healthy snack and want to test whether it needs "excessive sugar" sticker. It gets the sticker if it exceeds 100g of sugar. You take a sample and test.
  - $H_0$ : The average sugar content is 100 or less
    - $H_0 : \mu_s = 100$
  - $H_A$ : The average sugar content is more than 100
    - $H_A : \mu_s > 100$

# Hypothesis Testing

- Suppose we calculated the sample mean for sugar content. It's either below or above 100g
- Why would we need a fancy test?
- Because we only have a sample, but we want to learn about the population parameter!
- Maybe in our sample it's below 100g, but in the whole population it's above
- Hypothesis are always about parameters! Never about sample statistics

# Testing procedure

**Test procedure** is a rule based on a sample data whether to reject the null or not.

It usually relies on two elements:

- **Test statistics**
  - Function of the sample data summarizing evidence for or against a hypothesis
  - Usually we know how it should be distributed under the null. Ex: standardized mean
- **Rejection region**
  - Values of the test statistics such that we would reject the null if we observe them
  - Basically, values which are very unlikely under the null hypothesis
  - Typically threshold based:
  - Such as: reject  $H_0$  if the test statistic  $> k$ , where  $k$  is some threshold

# Testing procedure: Example

Let's go back to the sugar example. Suppose we have a sample  $n = 36$ .

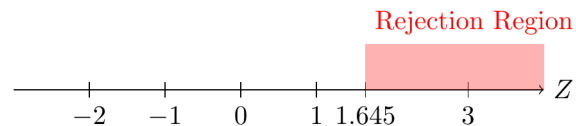
- Test statistics

- Our test statistic could be:

$$z = \frac{\bar{x} - 100}{\underbrace{s_x / \sqrt{36}}_{SE}}$$

- Rejection region

- For a 5% significance level, our rejection region would be:
- $\{1.645, \infty\}$





# Testing procedure: Example

- Suppose  $s_x = 20$ . When would we reject?

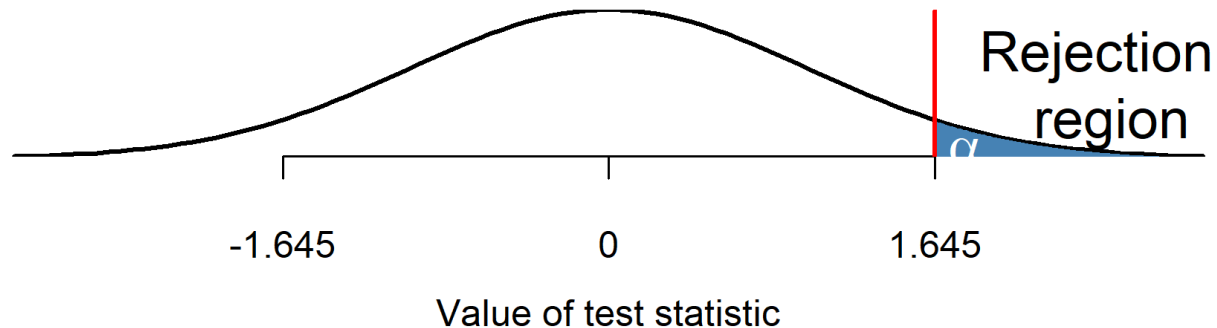
Reject if:

- $\frac{\bar{x}-100}{20/\sqrt{36}} > 1.645$
- Or equivalently if  $\bar{x} > 105.483$
- Why don't we reject even if  $\bar{x} = 103$ ?
- Because there could be randomness in the sample:
- So even if the  $\mu_s = 100$  we could get a sample with  $\bar{x} > 100$
- Why  $\bar{x} > 105.483$  though?
- Formally about it later, but it's very unlikely to get a sample mean that large if  $\mu_s = 100$

# Decision errors

- Suppose that the null is true and  $\mu_s = 100$
- You take a sample of 36.
- But you were very unlucky in your sampling, so by chance you actually got a sample with those rare packages with a lot of sugar and your  $\bar{x} = 106$
- According to the test procedure, you reject the null
- You make an error
- What's the probability of making such error?

$$\alpha = P(\text{Type 1 error}) = \underbrace{P_{H_0}(\bar{X} > 105.483)}_{\text{Pr. if null is true}} = P(Z > 1.645) = 1 - \underbrace{\Phi(1.645)}_{\text{Normal CDF}}$$

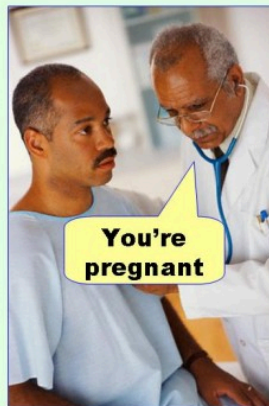


# Decision errors

	Do not reject $H_0$	Reject $H_0$ in favor of $H_A$
$H_0$ true	Okay	Type 1 Error
$H_A$ true	Type 2 Error	Okay

- $H_0$ : You are not pregnant
- $H_1$ : You are pregnant

**Type I error**  
(false positive)



**Type II error**  
(false negative)



# Decision errors

- **Type 1 error**: reject null when  $H_0$  is true
  - Conclude that people spend more when they got free shipping, while in reality they spend the same as those who didn't
- **Type 2 error**: don't reject null when  $H_A$  is true
  - Conclude that people who got free shipping spend the same as those without free shipping, while those with free shipping spend more
- Imagine a trial of a murderer and suppose  $H_0$  is innocent, and  $H_1$  is guilty.
- Describe error type 1 and 2 in these scenarios
  - Type 1: Convict an innocent
  - Type 2: Do not convict a murderer
- Probability of making type 1 error is  $\alpha$ 
  - It's also called **size**, or **significance level** of the test
  - We don't want to incorrectly reject null more than  $\alpha \times 100$  percent of times.
- Probability of making type 2 error is  $\beta$ 
  - To calculate  $\beta$ , you need to know the true  $\mu$

# Trade-off between type 1 and 2 errors

- When you decrease probability of error type 1, you increase the probability of error type 2
- (APP)

# Power of a test

- Power of a test is the probability of rejecting if alternative is true
- Power will be different for any possible value of the alternative
- Notice that

$$Power = P_{H_A}(Reject) = 1 - \underbrace{P_{H_A}(\text{Not Reject})}_{\text{type 2 error}} = 1 - \beta$$

Source: <https://shiny.rit.albany.edu/stat/betaprob/>

# Power of a test

- Let's go back to sugar example. Our rule was reject  $H_0$  if  $z > 1.645$ .
- Suppose  $\mu_s = 110$  and  $\sigma = 20$ . Let's calculate power of the test:

$$P_{H_A}(Z_{test} > 1.645) = P_{H_A}\left(\underbrace{\frac{\bar{X} - 100}{\sigma_x/\sqrt{36}}}_{SE} > 1.645\right) = P_{H_A}\left(\underbrace{\frac{\bar{X} - \mu_0}{\sigma_x/\sqrt{n}}}_{SE} > z_\alpha\right) = P_{H_A}\left(\bar{X} > \mu_0 + z_\alpha \frac{\sigma_x}{\sqrt{n}}\right) = P_{H_A}(\bar{X} > 105.483)$$

- To calculate this probability, let's standardize  $\bar{X}$ :
  - We subtract the true mean of the distribution and the standard deviation

$$P_{H_A}\left(\frac{\bar{X} - \mu_s}{\sigma_x/\sqrt{n}} > \frac{\mu_0 - \mu_s}{\sigma_x/\sqrt{n}} + z_\alpha\right) = P_{H_A}\left(\frac{\bar{X} - \mu_s}{\sigma_x/\sqrt{n}} > \frac{-10}{20/\sqrt{36}} + 1.645\right) = P_{H_A}\left(\underbrace{\frac{\bar{X} - \mu_s}{\sigma_x/\sqrt{n}}}_{st.normal} > -1.355\right) = 0.9115$$

Then:  $\beta = 1 - 0.9115 = 0.0885$

Power increases if:

- $n$  increases
- Alternative is further away from the null
- $\sigma$  decreases
- $\alpha$  increases

# General procedure:

1. Determine the the null and the alternative hypothesis
2. Collect your sample of independent observations
3. Choose the appropriate test
4. Calculate test statistic
5. Compare it to the rejection region
6. Reject or not your null hypothesis

Let's talk about **appropriate tests**!



# Appropriate tests

Choice of the test statistic and rejection region will depend on:

- **Type of data we are testing?**
  - **Single sample**
    - Example of hypothesis:  $H_0 : \mu = \mu_0$  vs  $H_A : \mu \neq \mu_0$
  - **Two independent samples**
    - Example of hypothesis:  $H_0 : \mu_{s1} = \mu_{s2}$  vs  $H_A : \mu_{s1} \neq \mu_{s2}$
  - **Paired data**
    - Example of hypothesis:  $H_0 : \mu_{d1} = \mu_{d2}$  vs  $H_A : \mu_{d1} \neq \mu_{d2}$
- **Can we estimate standard deviation well?**
  - **Yes**
    - Large sample
    - Known standard deviation
  - **No**
    - Small sample but normal distribution

# Hypothesis Testing: Single Sample

# Single Sample Test for Mean

## General idea:

- We test if the parameter is equal/larger/smaller than some concrete value
  - Ex:  $H_0 : \mu = 3$  vs  $H_A : \mu \neq 3$
- **Test statistic**: normalized sample mean

$$\text{test statistic} = \frac{\bar{X} - \mu_0}{SE}$$

- Where SE is standard error and depends on estimate of standard deviation
  - If standard deviation known:  $SE = \frac{\sigma}{\sqrt{n}}$
  - If standard deviation not known:  $SE = \frac{s}{\sqrt{n}}$
- **Rejection region**: depends on the distribution
  - If large sample or known variance test statistic =  $Z_{test} \sim N(0, 1)$ 
    - Critical values come from standard normal distribution
  - If small sample with normal distribution test statistic =  $T_{test} \sim t(n - 1)$ 
    - Critical values come from student t with n-1 degrees of freedom

# Single Sample - Mean - Two Sided Test

**Hypothesis:**  $H_0 : \mu = \mu_0$  and  $H_A : \mu \neq \mu_0$

**Rejection Region:** For a test of significance level  $\alpha$ , reject

- If large sample or known variance from normal

$$\frac{\bar{X} - \mu_0}{SE} < -z_{\frac{\alpha}{2}} \quad \text{or} \quad \frac{\bar{X} - \mu_0}{SE} > z_{\frac{\alpha}{2}}$$

- If small sample from normal

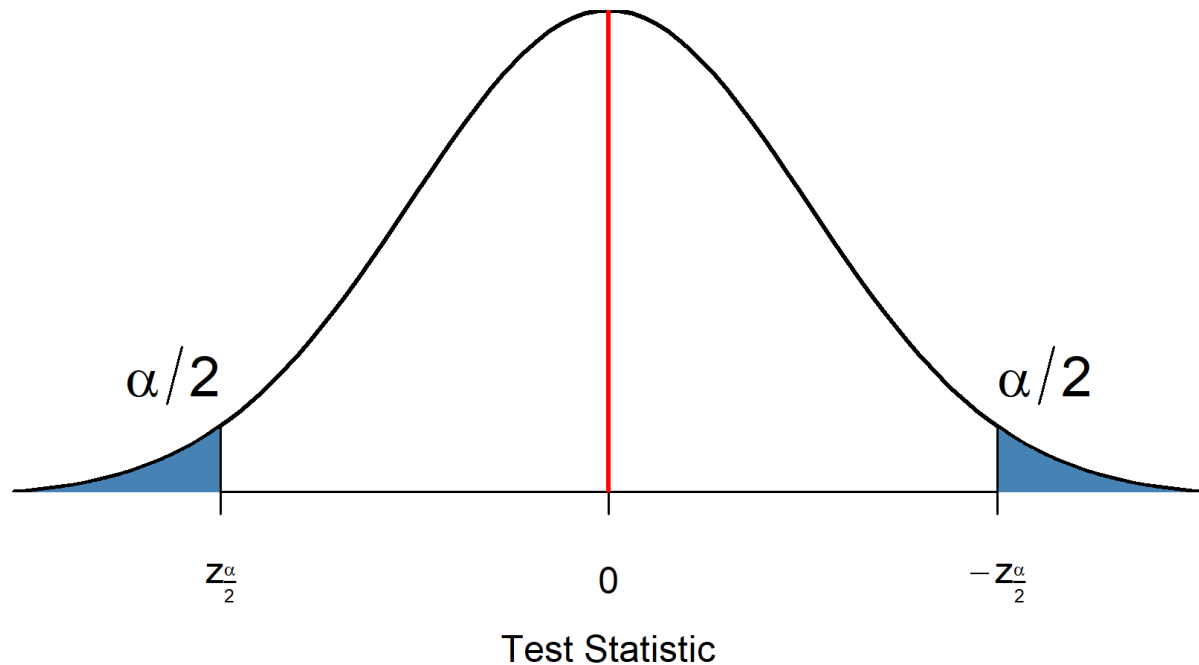
$$\frac{\bar{X} - \mu_0}{SE} < -t_{(n-1), \frac{\alpha}{2}} \quad \text{or} \quad \frac{\bar{X} - \mu_0}{SE} > t_{(n-1), \frac{\alpha}{2}}$$

Where

- $z_{\frac{\alpha}{2}}$  and  $t_{(n-1), \frac{\alpha}{2}}$  are  $(1 - \frac{\alpha}{2})$  quantiles of standard normal and t-students with n-1 degrees of freedom

# Single Sample - Mean - Two Sided Test

**Distribution of the statistic under the null (standard normal case)**



# Link between Tests and Confidence Intervals

Imagine a test for a parameter  $\mu$  with hypothesis  $H_0 = 4$  and  $H_A \neq 4$ .

- Suppose you draw a sample of 49 units with a mean  $\bar{x} = 3$  and standard deviation is  $s = 3$
- Our test statistic is  $z_{test} = \frac{3-4}{3/7} = -2.34$
- Our critical values at  $\alpha = 0.05$  are -1.96 and 1.96, hence we reject
- We can also calculate the 95% confidence interval for our sample mean:

$$\left\{3 - 1.96 \frac{3}{\sqrt{49}}, 3 + 1.96 \frac{3}{\sqrt{49}}\right\} = \{2.16, 3.84\}$$

- It does not contain the null hypothesis

# Link between Tests and Confidence Intervals

- More generally: the  $1 - \alpha$  confidence interval doesn't contain null = null would be rejected with a test of significance  $\alpha$
- Mathematically:
  - Let  $\mu_0$  be the null hypothesis. We reject (in two sided test) the null at  $\alpha$  if

$$\left| \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \right| > z_{\frac{\alpha}{2}}$$

- Which can be rewritten as:

$$\mu_0 < \bar{X} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \quad \text{or} \quad \mu_0 > \bar{X} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

# Single Sample - Mean - One sided - Case 1

**Hypothesis:**  $H_0 : \mu = \mu_0$  and  $H_A : \mu < \mu_0$

- This includes any null hypothesis like  $\mu \geq \mu_0$ . Why?
- If you rejected  $\mu_0$  at  $\alpha$ , you would reject for sure anything larger than  $\mu_0$

**Rejection Region:** For a test of significance level  $\alpha$ , reject

- If large sample or known variance from normal

$$\frac{\bar{X} - \mu_0}{SE} < -z_\alpha$$

- If small sample from normal

$$\frac{\bar{X} - \mu_0}{SE} < -t_{(n-1),\alpha}$$

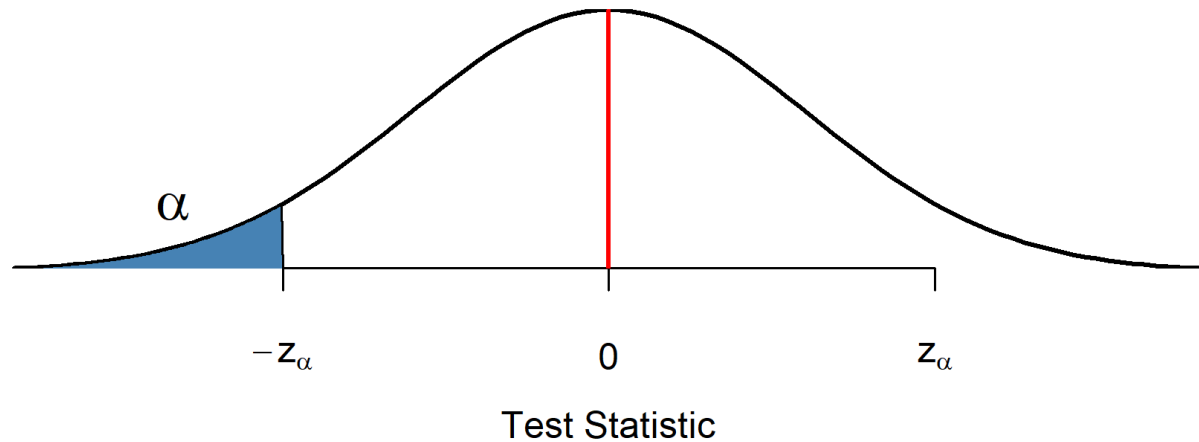
Where

- $z_\alpha$  and  $t_{(n-1),\alpha}$  are  $(1 - \alpha)$  quantiles of standard normal and t-students with  $n-1$  degrees of freedom



# Single Sample - Mean - One sided - Case 1

**Distribution of the statistic under the null (standard normal case)**



- **Intuition:**

- We never reject  $H_0$  if  $\bar{X} \geq \mu_0$ .
- We reject  $H_0$  if  $\bar{X}$  is sufficiently smaller than  $\mu_0$
- If null is true (  $\mu = \mu_0$  ) then we reject with probability

$$P\left(\frac{\bar{X} - \mu_0}{s/\sqrt{n}} < -z_\alpha\right) = P(Z < -z_\alpha) = \alpha$$

# Single Sample - Mean - One sided - Case 2

**Hypothesis:**  $H_0 : \mu = \mu_0$  and  $H_A : \mu > \mu_0$

- This includes any null hypothesis like  $\mu \leq \mu_0$ . Why?
- If you rejected  $\mu_0$  at  $\alpha$ , you would reject for sure anything smaller than  $\mu_0$

**Rejection Region:** For a test of significance level  $\alpha$ , reject

- If large sample or known variance from normal

$$\frac{\bar{X} - \mu_0}{SE} > z_\alpha$$

- If small sample from normal

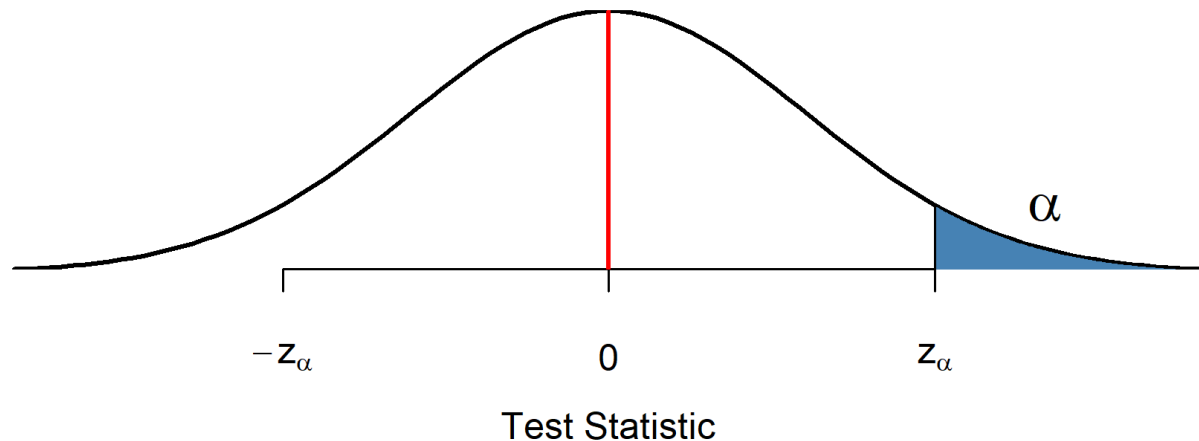
$$\frac{\bar{X} - \mu_0}{SE} > t_{(n-1),\alpha}$$

Where

- $z_\alpha$  and  $t_{(n-1),\alpha}$  are  $(1 - \alpha)$  quantiles of standard normal and t-students with  $n-1$  degrees of freedom

# Single Sample - Mean - One sided - Case 2

**Distribution of the statistic under the null (standard normal case)**



- **Intuition:**

- We never reject  $H_0$  if  $\bar{X} \leq \mu_0$ .
- We reject  $H_0$  if  $\bar{X}$  is sufficiently larger than  $\mu_0$
- If null is true (  $\mu = \mu_0$  ) then we reject with probability

$$P\left(\frac{\bar{X} - \mu_0}{s/\sqrt{n}} > z_\alpha\right) = P(Z > z_\alpha) = \alpha$$

# Tests for Variance

Test for variance has the same idea, but different distribution of the test statistic. Suppose you have an IID sample from [normal distribution](#)

**Test statistic** for  $H_0 : \sigma = \sigma_0$  and its distribution if null is true is:

$$\chi_{test} = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}$$

It is distributed as a Chi-square with  $n-1$  degrees of freedom if null is true:

## Rejection regions:

1.  $H_A : \sigma \neq \sigma_0$ 
    - Reject null if  $\chi_{test} < \chi_{\frac{\alpha}{2}, n-1}$  or  $\chi_{test} > \chi_{1-\frac{\alpha}{2}, n-1}$
  2.  $H_A : \sigma \leq \sigma_0$ 
    - Reject null if  $\chi_{test} < \chi_{\alpha, n-1}$
  3.  $H_A : \sigma \geq \sigma_0$ 
    - Reject null if  $\chi_{test} > \chi_{1-\alpha, n-1}$
- $\chi_{\alpha, n-1}$  is  $\alpha$  quantile of  $\chi_{n-1}$
  - $\chi_{1-\alpha, n-1}$  is  $1 - \alpha$  quantile of  $\chi_{n-1}$
  - $\chi_{df}$  distribution is not symmetric around 0!

## Exercise

Suppose you are a ham producer. As you need to stick to the nutritional guidelines, the standard deviation of the fat content in your product needs cannot be larger than 0.1. You take a sample of 16 hams and you want to test this. Suppose  $s = 0.2$  and  $\alpha = 0.05$

1. Write down the hypotheses of the test
  - Remember test statistic is about variance
1. What is type 1 and type 2 error in this context
2. Test the claim at 5% significance level. State assumptions you need to implement it
3. What is 95% confidence interval for variance

# P-values

**Technically:** The P-value is the probability (so between 0 and 1), under the null hypothesis, of obtaining a value of the test statistic at least as contradictory to  $H_0$  as the value calculated from the available sample.

Depends on:

- Your hypothesis
- Calculated test statistic
- Suppose the test statistic under  $H_0$  is  $Z_{test} \sim N(0, 1)$
- One sided tests:
  - For  $H_A : \mu > \mu_0$   
 $p - value = P(Z_{Test} \geq (computed - test - statistic))$
  - For  $H_A : \mu < \mu_0$   
 $p - value = P(Z_{Test} \leq (computed - test - statistic))$
- Two sided test:
  - For  $H_A : \mu \neq \mu_0$   
 $p - value = 2P(Z_{Test} \geq |(computed - test - statistic)|)$

### Example:

- We test for  $H_0 : \mu = 5$  and  $H_A : \mu > 5$
- Suppose in sample a of 36 we have  $\bar{x} = 5.5$  and  $s = 2$ .
- Test statistic in this case is distributed normally

$$p - value = P(Z > \frac{5.5 - 5}{2/\sqrt{6}}) = P(Z > 1.5) = 0.066$$

**Intuitively:** The probability that a statistic we calculated (or more extreme value) could arise just by chance if null is true

# P-value in two-sided test

R  PSYCHOLOGIST

About

Posts

Visualizations



## Understanding $p$ -values Through Simulations

**An Interactive Visualization**

Created by [Kristoffer Magnusson](#)

Source: <https://rpsychologist.com/pvalue/>



# P-Values and test significance

- In our example we calculated the p-value of 0.066
- Would the test at  $\alpha = 0.05$  reject the null?
- Any test with the significance level  $\alpha > p - value$  would reject the null

## Another way to think about p-values

- The P-value is the smallest significance level  $\alpha$  at which the  $H_0$  can be rejected

# Hypothesis Testing: Two samples

# Two Samples and the Difference in Means

We have 2 iid samples **independent** of each other from two populations  $X_1, X_2, \dots, X_n$  and  $Y_1, Y_2, \dots, Y_m$ .

## General idea:

- We test the difference between **population means** in two populations
  - Ex:  $H_0 : \mu_X - \mu_Y = 100$  vs  $H_A : \mu_X - \mu_Y \neq 100$
  - Ex:  $H_0 : \mu_X - \mu_Y = 0$  vs  $H_A : \mu_X - \mu_Y > 0$
  - Test for no difference:
    - Ex:  $H_0 : \mu_X - \mu_Y = 0$  vs  $H_0 : \mu_X - \mu_Y \neq 0$
- **Test statistic**: normalized difference in sample means

$$\text{test statistic} = \frac{\bar{X} - \bar{Y} - \overbrace{(\mu_{X,0} - \mu_{Y,0})}^{\Delta_0}}{SE} = \frac{\bar{X} - \bar{Y} - \Delta_0}{SE}$$

- Where SE is standard error and depends on the estimate of standard deviation
  - If standard deviation known:  $SE = \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}$
- Why? Because  $Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) = \frac{\sigma_X}{n} + \frac{\sigma_Y}{m}$
- If standard deviation not known:  $SE = \sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}$
- **Rejection region**: depends on the distribution
  - If large sample or known variance from normal  
test statistic =  $Z_{test} \sim N(0, 1)$ 
    - Critical values come from standard normal distribution
  - If small sample with normal distribution test statistic =  $T_{test} \sim t(v)$ 
    - Critical values come from student t with v degrees of freedom

## Side note on the number of degrees of freedom for difference in means

Two ways to do it:

1. Proper, complex (and annoying):

$$\frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\left(\frac{s_X^2}{n}\right)^2 / (n-1) + \left(\frac{s_Y^2}{m}\right)^2 / (m-1)}$$

Round down to nearest integer

1. Simple, approximate shortcut:

$$v = \min(n-1, m-1)$$

# Difference in means - Two Sided Test

**Hypothesis:**  $H_0 : \mu_x - \mu_Y = \Delta_0$  and  $H_A : \mu_x - \mu_Y \neq \Delta_0$

**Rejection Region:** For a test of significance level  $\alpha$ , reject

- If large sample or known variance from normal

$$\frac{\bar{X} - \bar{Y} - \Delta_0}{SE} < -z_{\frac{\alpha}{2}} \quad or \quad \frac{\bar{X} - \bar{Y} - \Delta_0}{SE} > z_{\frac{\alpha}{2}}$$

- If small sample from normal

$$\frac{\bar{X} - \bar{Y} - \Delta_0}{SE} < -t_{v, \frac{\alpha}{2}} \quad or \quad \frac{\bar{X} - \bar{Y} - \Delta_0}{SE} > t_{v, \frac{\alpha}{2}}$$

Where

- $z_{\frac{\alpha}{2}}$  and  $t_{v, \frac{\alpha}{2}}$  are  $(1 - \frac{\alpha}{2})$  quantiles of standard normal and t-students with v degrees of freedom

# Difference in means - Two Sided Test

If CLT kicks in, the distributions are:

$$\bar{X} \sim N(\mu_X, \sigma_X/\sqrt{n}) \quad \text{and} \quad \bar{Y} \sim N(\mu_Y, \sigma_Y/\sqrt{m})$$

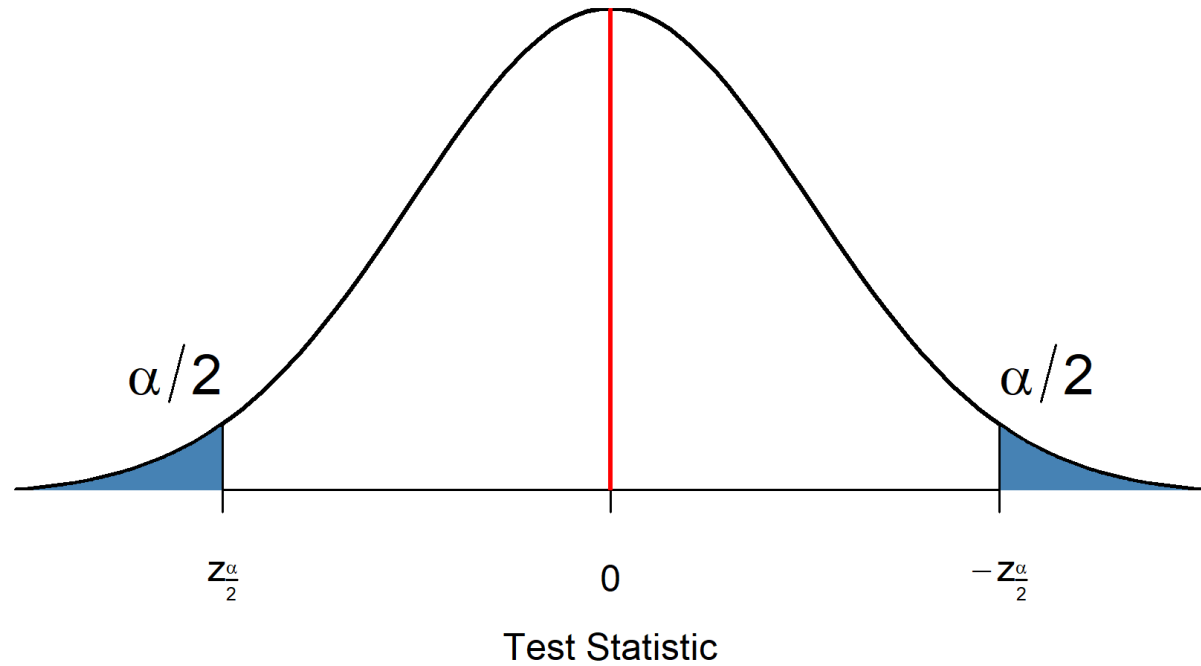
The difference of two normal (independent) distributions under:

$$\bar{X} - \bar{Y} \sim N(\mu_X - \mu_Y, \sqrt{\sigma_X^2/n + \sigma_Y^2/m})$$

So if null is true, then standardized statistic:

$$\frac{\bar{X} - \bar{Y} - (\mu_{X,0} - \mu_{Y,0})}{\sqrt{s_X^2/n + s_Y^2/m}} \sim N(0, 1)$$

## Distribution of the statistic under the null (standard normal case)





# Two Samples - Mean - One sided - Case 1

**Hypothesis:**  $H_0 : \mu_X - \mu_Y = \Delta_0$  and  $H_A : \mu_X - \mu_Y < \Delta_0$

- This includes any null hypothesis like  $\mu_X - \mu_Y \geq \Delta_0$ . Why?
- If you rejected  $\mu_X - \mu_Y = \Delta_0$  at  $\alpha$ , you would reject for sure anything larger than  $\Delta_0$

**Rejection Region:** For a test of significance level  $\alpha$ , reject

- If large sample or known variance from normal

$$\frac{\bar{X} - \bar{Y} - (\Delta_0)}{SE} < -z_\alpha$$

- If small sample from normal

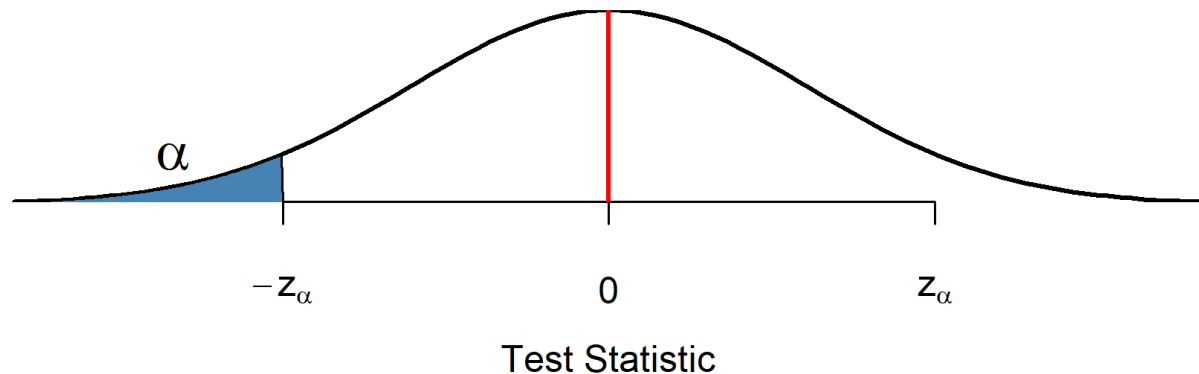
$$\frac{\bar{X} - \bar{Y} - (\Delta_0)}{SE} < -t_{v,\alpha}$$

Where

- $-z_\alpha$  and  $-t_{v,\alpha}$  are  $\alpha$  quantiles of standard normal and t-students with  $v$  degrees of freedom

# Two Samples - Mean - One sided - Case 1

**Distribution of the statistic under the null (standard normal case)**



- **Intuition:**

- We never reject  $H_0$  if  $\bar{X} - \bar{Y} \geq \Delta_0$ .
- We reject  $H_0$  if  $\bar{X} - \bar{Y}$  is sufficiently smaller than  $\Delta_0$
- If null is true (  $\bar{X} - \bar{Y} = \Delta_0$  ) then we reject with probability

$$P\left(\frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{s_X^2/n + s_Y^2/m}} < -z_\alpha\right) = P(Z < -z_\alpha) = \alpha$$

# Two Samples - Mean - One sided - Case 2

**Hypothesis:**  $H_0 : \mu_X - \mu_Y = \Delta_0$  and  $H_A : \mu_X - \mu_Y > \Delta_0$

- This includes any null hypothesis like  $\mu_X - \mu_Y \leq \Delta_0$ . Why?
- If you rejected  $\mu_X - \mu_Y = \Delta_0$  at  $\alpha$ , you would reject for sure anything smaller than  $\Delta_0$

**Rejection Region:** For a test of significance level  $\alpha$ , reject

- If large sample or known variance with normal

$$\frac{\bar{X} - \bar{Y} - (\Delta_0)}{SE} > z_\alpha$$

- If small sample from normal

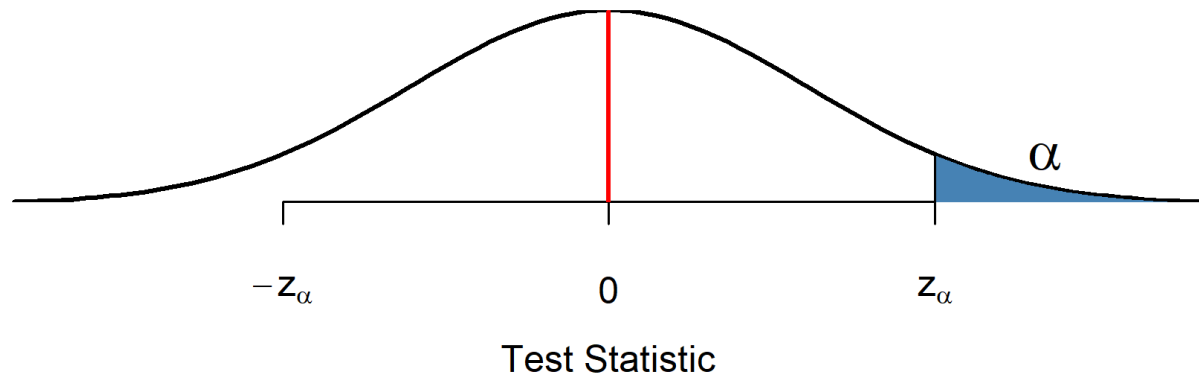
$$\frac{\bar{X} - \bar{Y} - (\Delta_0)}{SE} > t_{v,\alpha}$$

Where

- $z_\alpha$  and  $t_{v,\alpha}$  are  $(1 - \alpha)$  quantiles of standard normal and t-students with  $v$  degrees of freedom

## Two Samples - Mean - One sided - Case 2

**Distribution of the statistic under the null (standard normal case)**



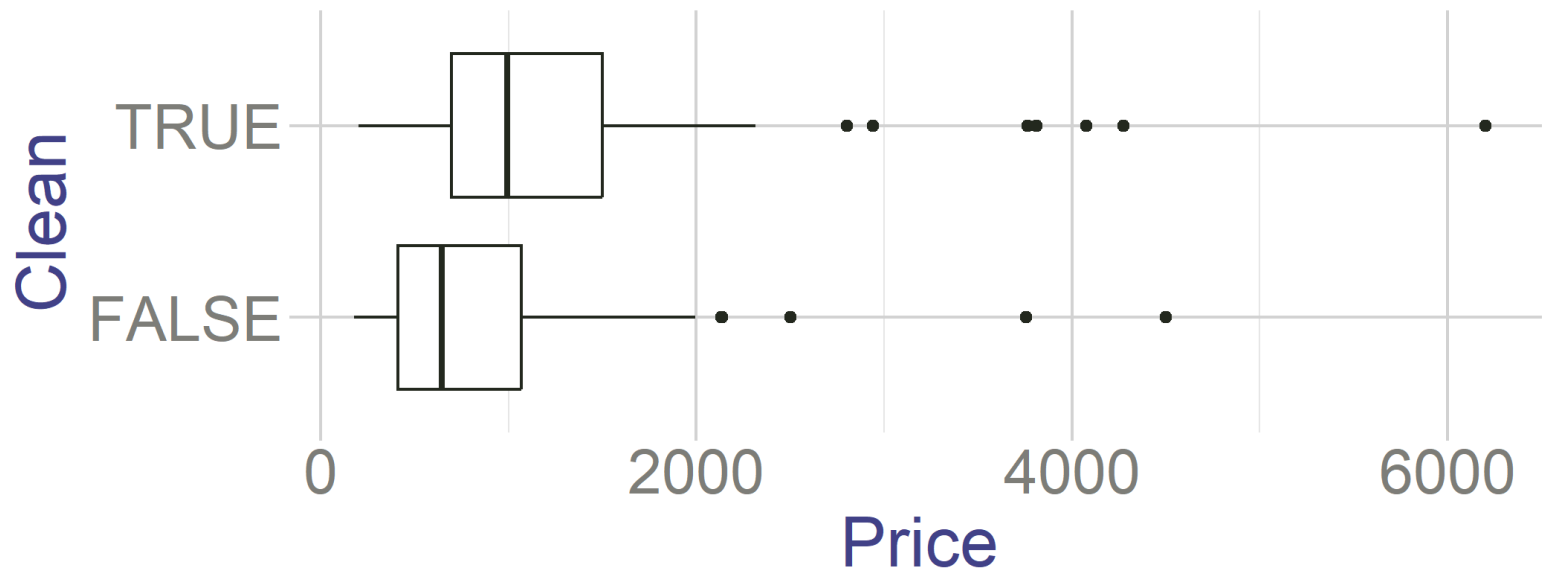
- **Intuition:**

- We never reject  $H_0$  if  $\bar{X} - \bar{Y} \leq \Delta_0$ .
- We reject  $H_0$  if  $\bar{X} - \bar{Y}$  is sufficiently larger than  $\Delta_0$
- If null is true (  $\bar{X} - \bar{Y} = \Delta_0$  ) then we reject with probability

$$P\left(\frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{s_X^2/n + s_Y^2/m}} > z_\alpha\right) = P(Z > z_\alpha) = \alpha$$

Is the price in dirty apartments the same as price in clean apartments?

Sample	n	Sample Mean	Sample Standard Deviation
Clean	100	1245	962
Dirty	100	869	693



- State the null and alternative hypothesis
- Calculate the value of the test statistic
- Calculate the p value
- At which  $\alpha$  we reject the null?

We suspect Burger King might have larger caloric content for kids' meal. Let's test it.

Sample	n	Sample Mean	Sample Standard Deviation
McDonald	9	475	230
Burger King	16	520	180

- State the null and alternative hypothesis
- Calculate the value of the test statistic
- Determine the appropriate number of degrees of freedom
- Calculate the p value
- At which  $\alpha$  we reject the null?

# Comparing variances in two samples

**Example Problem 1:** is price dispersion the same in clean or dirty apartments?

**Example Problem 2:** is variance of prices of two stocks the same?

We need a test for variances in two different samples.

# Hypothesis testing with variances

- Suppose you have an iid sample  $X_1, X_2, \dots, X_n$  and iid sample  $Y_1, Y_2, \dots, Y_m$ , both from **normal distributions**
- Denote  $s_1^2$  and  $s_2^2$  their sample variances and  $\sigma_1^2$  and  $\sigma_2^2$  their population variances

**Test Statistic** and its distribution under the null is:

Null hypothesis:  $H_A : \sigma_1 = \sigma_2$

$$F_{test} = \frac{s_1^2}{s_2^2} \sim F_{n-1, m-1}$$

## Rejection region

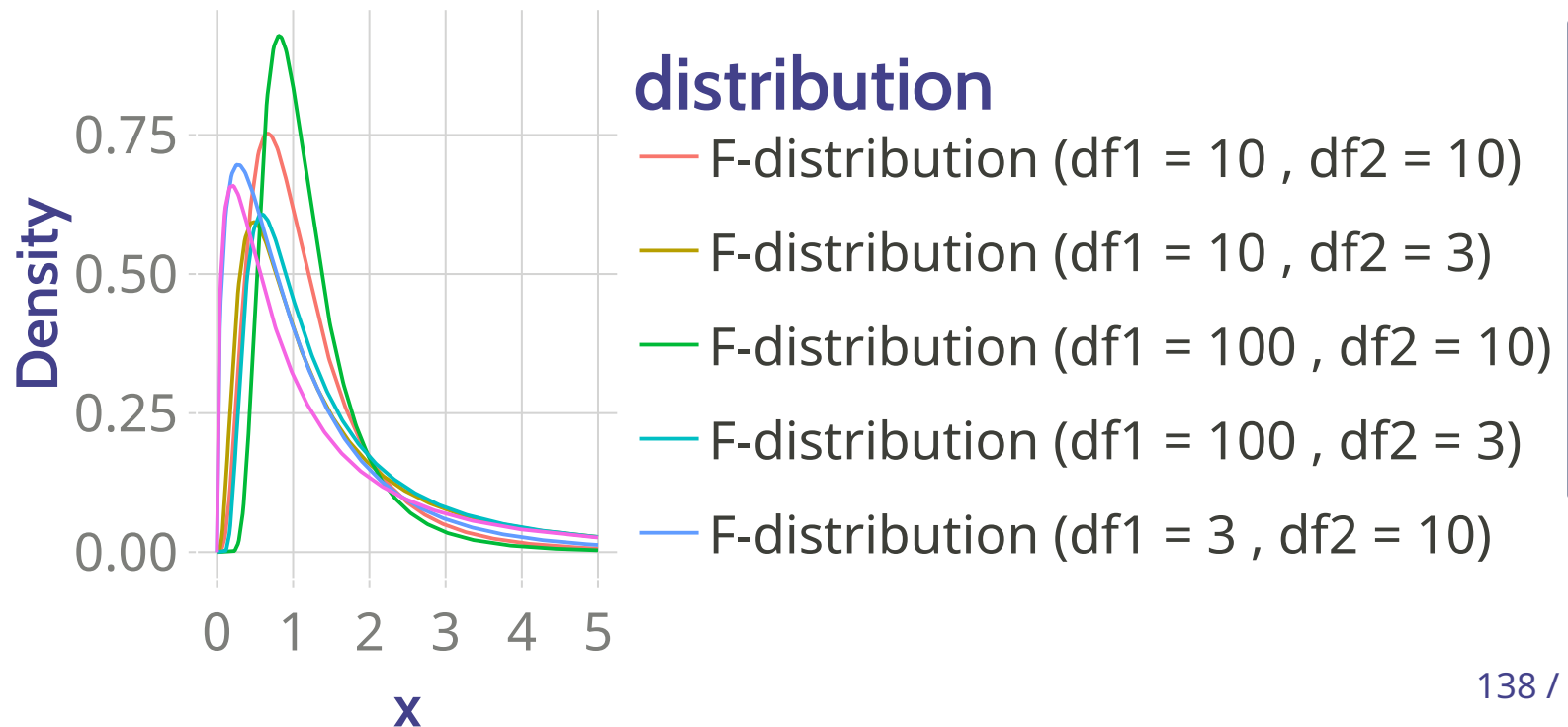
1.  $H_A : \sigma_1 \neq \sigma_2$ 
  - Reject null if  $F_{test} < F_{\frac{\alpha}{2}, n-1, m-1}$  or  $F_{test} > F_{1-\frac{\alpha}{2}, n-1, m-1}$
2.  $H_A : \sigma_1 \leq \sigma_2$ 
  - Reject null if  $F_{test} < F_{\alpha, n-1, m-1}$
3.  $H_A : \sigma_1 \geq \sigma_2$ 
  - Reject null if  $F_{test} > F_{1-\alpha, n-1, m-1}$



# F-distribution

If  $X_1$  and  $X_2$  are chi-square distributed and independent, then the ratio of  $X_1$  by its degrees of freedom to  $X_2$  by its degrees of freedom is distributed as F.

$$F = \frac{X_1/v_1}{X_2/v_2} \sim F_{v_1, v_2}$$



# Variance test and F-distribution

How come under the null we get F statistics?  $s^2$  is not chi square.... but!

$$\frac{(n-1)s_1^2}{\sigma_1^2} \sim \chi_{n-1}$$
$$F_{test} = \frac{\frac{(n-1)s_1^2}{\sigma_1^2(n-1)}}{\frac{(m-1)s_2^2}{\sigma_2^2(m-1)}} \sim F_{n-1, m-1}$$

Under null we have that  $\sigma_1 = \sigma_2$ , so

$$F_{test} = \frac{\frac{(n-1)s_1^2}{\sigma_1^2(n-1)}}{\frac{(m-1)s_2^2}{\sigma_2^2(m-1)}} = \frac{s_1^2}{s_2^2}$$

## Example:

Let's test the equality of variances of clean and dirty apartments.

- $H_0 : \sigma_c^2 = \sigma_d^2$
- $H_A : \sigma_c^2 \neq \sigma_d^2$
- Variances in each samples are:  $s_c^2 = 925255.5$  and  $s_d^2 = 480681.9$
- We have 100 observations in each sample, so 99 degrees of freedom on each side

$$F_{test} = \frac{s_c^2}{s_d^2} = \frac{925255.5}{480681.9} = 1.924$$

Critical regions at 5% are:

- $F_{\frac{\alpha}{2}, n-1, m-1} = F_{0.025, 99, 99} = 0.67284$
- $F_{1-\frac{\alpha}{2}, n-1, m-1} = F_{0.975, 99, 99} = 1.48623$
- We reject the equality at 5%

# Hypothesis Testing: Paired Data

# Paired Data

- Sometimes we have only one set of individuals or objects, but two (or more) observations per each.
- **Example:** We look at employees productivity. For each employee we have two observations: their productivity when working from home and their productivity when working from the office.
- **Example:** We look at sales at Walmart and Soriana in each neighborhood. We have two observations per neighborhood: sales at local Walmart and sales at local Soriana

# Paired Data

We have a set of independent pairs of data  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$

Observation within the pair are not necessarily independent!

- Some productive workers are productive both in the office and at home, some unproductive are unproductive both in the office and at home.
- We are interested in the difference between  $X$  and  $Y$ . We observe that difference for each individual in the sample:  $X_i - Y_i = d_i$
- We want to look at the population value of that difference:  $\mu_X - \mu_Y = \Delta$ 
  - We can check if the means are the same:  $\Delta = 0$ 
    - No difference in productivity between work at home and at the office
  - Or mean  $X$  is bigger than mean  $Y$ :  $\Delta > 0$

# Paired Data

Null hypothesis:  $H_0 : \Delta = \Delta_0$

**Test Statistic** is:

$$\text{test statistic} = \frac{\overbrace{\bar{X} - \bar{Y}}^{\bar{d}} - \Delta_0}{SE} = \frac{\bar{d} - \Delta_0}{SE}$$

- Note on the Standard Error - it's not the same formula as for the two independent samples:  $Var(X_i - Y_i) \neq Var(X_i) + Var(Y_i)$
- We just calculate it from the differences.

- If we know the variance,  $SE = \frac{\sqrt{var(d_i)}}{\sqrt{n}} = \frac{\sigma_d}{\sqrt{n}}$
- If we don't:  $SE = \frac{s_d}{\sqrt{n}}$

# Paired Data

**Rejection Region:** For a test of significance level  $\alpha$ :

- If large sample or known variance with normal distribution:
- for  $H_A : \bar{d} \neq \Delta_0$ , reject if:

$$\frac{\bar{d} - \Delta_0}{SE} < -z_{\frac{\alpha}{2}} \quad \text{or} \quad \frac{\bar{d} - \Delta_0}{SE} > z_{\frac{\alpha}{2}}$$

- for  $H_A : \bar{d} > \Delta_0$ , reject if:

$$\frac{\bar{d} - \Delta_0}{SE} > z_{\alpha}$$

- for  $H_A : \bar{d} < \Delta_0$ , reject if:

$$\frac{\bar{d} - \Delta_0}{SE} < -z_{\alpha}$$

- If small sample from normal, replace critical values with **t-distribution with n-1 degrees of freedom**



# Paired or Unpaired Data?

1. We would like to know if Intel's stock and Southwest Airlines' stock have similar rates of return. To find out, we take a random sample of 50 days, and record Intel's and Southwest's stock on those same days

- Paired

1. We randomly sample 50 items from Target stores and note the price for each. Then we visit Walmart and collect the price for each of those same 50 items.

- Paired

1. A school board would like to determine whether there is a difference in average SAT scores for students at one high school versus another high school in the district. To check, they take a simple random sample of 100 students from each high school

- Not paired

A psychologist thinks that age influences IQ. They take a random sample of 100 people of age 40. For each person we know their IQ at age 16 and now. On average, in this sample, IQ at young age was 8 points higher than at age 40. Standard deviation of that difference was 7 points. Using  $\alpha = 0.01$  test the hypothesis that IQ decreases with age.

## Hypothesis Testing for Correlation

**Example:** are cleanliness score (C) and price (P) correlated in the population?

Assume you have **normally distributed variables** in **independent pairs**.

$$\{(C_1, P_1), (C_2, P_2), \dots\}$$

**Null hypothesis:**  $H_0 : \rho(C, P) = \rho_0$

**Test statistic** and its distribution under the null:

$$T_{test} = \frac{\hat{\rho}(C, P) - \rho_0}{\sqrt{\frac{1 - \hat{\rho}(C, P)^2}{n-2}}} \sim t_{n-2}$$

where  $\hat{\rho}(C, P)$  is sample correlation coefficient

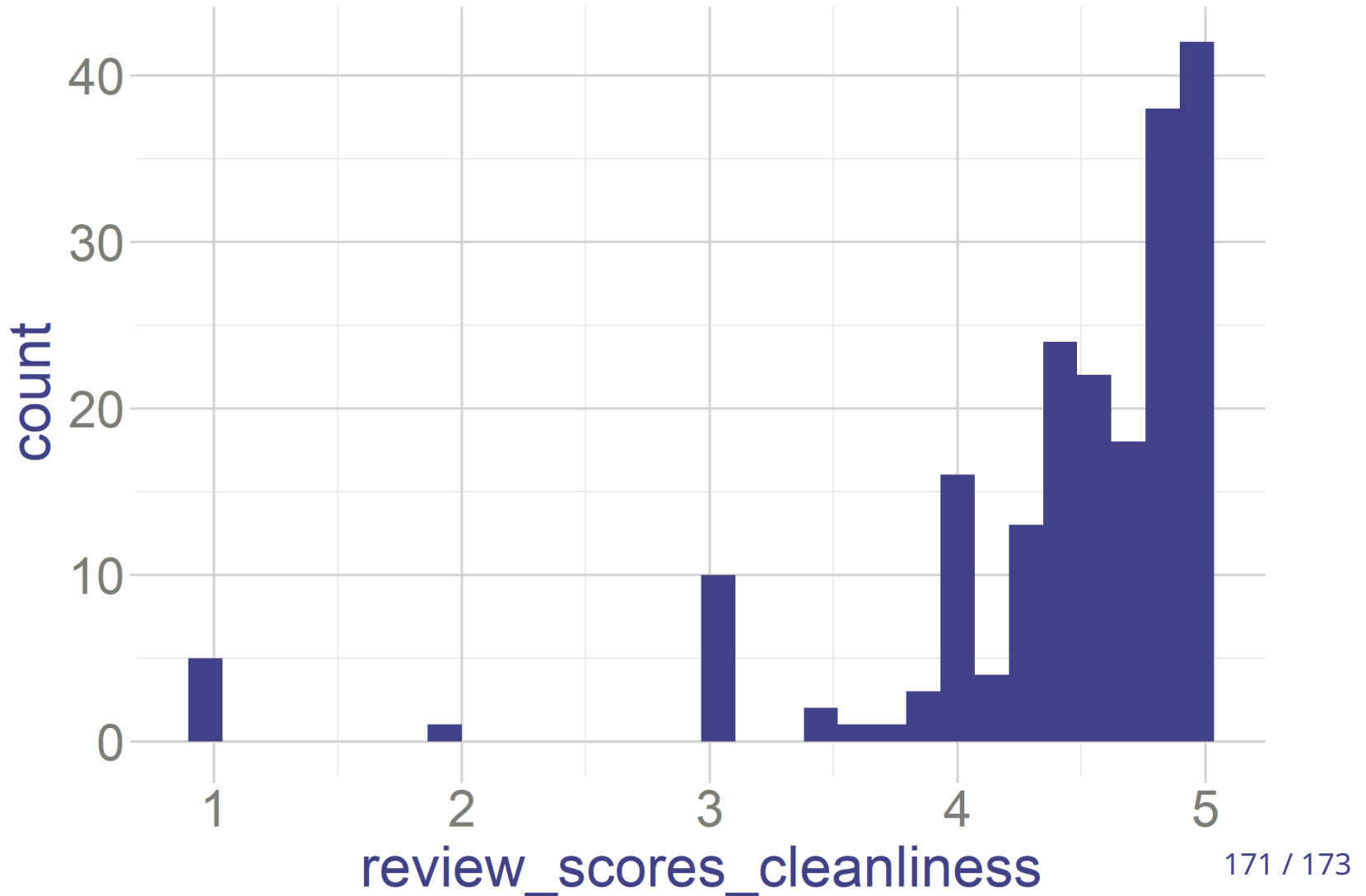
**Alternative hypothesis** and their **rejection regions**:

- $H_A : \rho(C, P) \neq \rho_0$  - reject  $H_0$  if  $t_{test} > t_{\frac{\alpha}{2}, n-2}$  or  $t_{test} < -t_{\frac{\alpha}{2}, n-2}$
- $H_A : \rho(C, P) > \rho_0$  - reject  $H_0$  if  $t_{test} > t_{\alpha, n-2}$
- $H_A : \rho(C, P) < \rho_0$  - reject  $H_0$  if  $t_{test} < -t_{\alpha, n-2}$

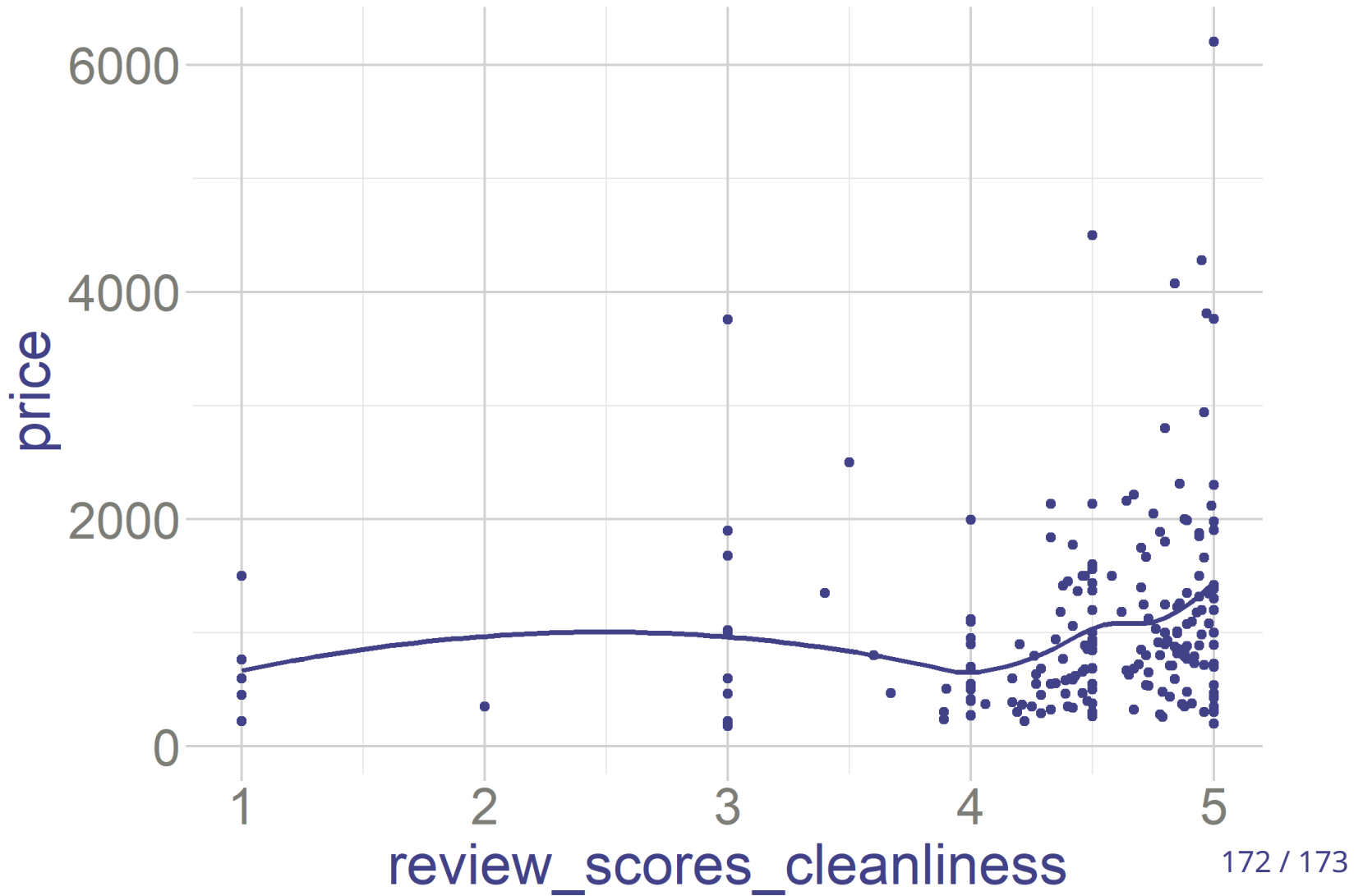
# Correlation test

- Suppose we test  $H_0 : \rho(C, P) = 0$  vs  $\rho(C, P) > 0$  in our Airbnb Data (clean score vs price)
- We **assume** that:
  - The variables are normally distributed
  - Pairs are independent
  - Relationship would be linear
- Sample correlation is: 0.159, we have  $n = 200$
- So  $t_{test} = \frac{\hat{\rho}(C,P)\sqrt{(n-2)}}{\sqrt{1-\hat{\rho}(C,P)^2}} = \frac{0.159\sqrt{98}}{\sqrt{1-0.025}} = 2.25$
- Since  $n = 200$ , t-student is identical to standard normal. We can either
  - Compare calculated statistic to critical regions at various significance levels
  - Compute its p-value
- $p - value = P(Z > t_{test}) = P(Z > 2.25) = 0.0126$

## Correlation Test



## Correlation Test



## Exercise

Suppose you have a random sample of 27 units (from a bivariate normal).  $X$  measures client's age and  $Y$  measures their spending. You calculated the correlation coefficient of  $\hat{\rho} = -0.45$ . Can you reject null of  $\rho = 0$  in favor of alternative  $\rho \neq 0$  at 5% significance level?

