SPRING 2019

TEXAS
The University of Texas at Austin

CE 311S : PROBABILITY AND STATISTICS

Discussion session

M 1:00 – 2:00 PM

PRIYADARSHAN PATIL
Teaching Assistant, The University of Texas at Austin

Discussion overview

- Administrative stuff
- Intervals – Confidence, Tolerance, Prediction
- Hypothesis testing

Administrative Stuff

- Online assignment 6 due this Friday

Learning goals

- Outline
  - Confidence Interval, Prediction interval, Tolerance interval (Review + Examples)

- Learning outcomes (at the end of this lab you should be able to):
  - **Identify the type of interval** asked in the problem
  - Use the **correct formula for the interval** using the property of the problem
    - Two sided intervals
    - One sided intervals
  - **Use standard tables** to evaluate chi-critical, t-critical, and tolerance critical values

Review part 1

- A confidence interval (CI) gives a **plausible range for a** _____ **parameter p** by using outcomes from a sample
  - Typical parameters of interest: Mean and Variance
- A prediction interval (PI) gives a **plausible range for** _____
- A tolerance interval (TI) gives a plausible range which _____ in the entire population

Example 1: Which interval to use?

A sample of 100 people are selected from UT campus. The weight of any person is normally distributed in the population

- Interval which contains weight of a single individual selected in the population with 95% probability
  (prediction)
- Interval which contains population variance with 95% probability
  (confidence)
- Interval $(-\infty, K)$ where K is the age above which only 5% of the population belongs with 95% probability
  (tolerance)
- Interval (K, ∞) which contains population mean with 90% probability
  (confidence)

Review part 2: formulae

Assumption	X_i's are from **normal distribution** with unknown mean and unknown variance	X_i's are from **normal distribution** with unknown mean and unknown variance	X_i's are from **normal distribution** with unknown mean and unknown variance and n<40
Interval	Confidence interval on population variance	Prediction interval for X of a single individual	Tolerance interval containing at least $k\%$ of population
Notations	n: $\chi^2_{\alpha,n-1}$: s:	$\bar{x}$: n: s: $t_{\frac{\alpha}{2},n-1}$:	$\bar{x}$: n: s: $C_{\alpha,k}$:
$(1-\alpha)100\%$ two-sided interval	Lower limit: $(n-1)s^2/\chi^2_{\frac{\alpha}{2},n-1}$ Upper limit:	Lower limit: $\bar{x} + t_{\frac{\alpha}{2},n-1} * s * \sqrt{1+\frac{1}{n}}$ Upper limit:	Lower limit: Upper limit:
$(1-\alpha)100\%$ one-sided upper bound interval	Lower limit: $-\infty$ Upper limit:	Lower limit: $-\infty$ Upper limit:	Lower limit: $-\infty$ Upper limit:
$(1-\alpha)100\%$ one-sided lower bound interval	Lower limit: Upper limit: $+\infty$	Lower limit: Upper limit: $+\infty$	Lower limit: Upper limit: $+\infty$

Given: $X_1, X_2, \ldots, X_n$ are random samples i.e. they are **independent and identically distributed**

A confidence interval (CI) gives a plausible range for a population parameter p by using outcomes from a sample

A prediction interval (PI) gives a plausible range for a single future prediction value

A tolerance interval (TI) gives a plausible range which contains at least k% of the entire population

Assumption	X_i's are from **normal distribution** with unknown mean and unknown variance	X_i's are from **normal distribution** with unknown mean and unknown variance	X_i's are from **normal distribution** with unknown mean and unknown variance and n<40
Interval	Confidence interval on population variance	Prediction interval for X of a single individual	Tolerance interval containing at least $k\%$ of population
Notations	n: $\chi^2_{\alpha,n-1}$: s:	$\bar{x}$: n: s: $t_{\frac{\alpha}{2},n-1}$:	$\bar{x}$: n: s: $C_{\alpha,k}$:
$(1-\alpha)100\%$ two-sided interval	Lower limit: $(n-1)s^2/\chi^2_{\frac{\alpha}{2},n-1}$ Upper limit: $(n-1)s^2/\chi^2_{1-\frac{\alpha}{2},n-1}$	Lower limit: $\bar{x} - t_{\frac{\alpha}{2},n-1} * s * \sqrt{1+\frac{1}{n}}$ Upper limit: $\bar{x} + t_{\frac{\alpha}{2},n-1} * s * \sqrt{1+\frac{1}{n}}$	Lower limit: $\bar{x} - C_{\alpha,k}s$ Upper limit: $\bar{x} + C_{\alpha,k}s$ (two sided C values)
$(1-\alpha)100\%$ one-sided upper bound interval	Lower limit: $-\infty$ Upper limit: $(n-1)s^2/\chi^2_{1-\alpha,n-1}$	Lower limit: $-\infty$ Upper limit: $\bar{x} + t_{\alpha,n-1} * s * \sqrt{1+\frac{1}{n}}$	Lower limit: $-\infty$ Upper limit: $\bar{x} + C_{\alpha,k}s$ One sided C value
$(1-\alpha)100\%$ one-sided lower bound interval	Lower limit: $(n-1)s^2/\chi^2_{\alpha,n-1}$ Upper limit: $+\infty$	Lower limit: $\bar{x} - t_{\alpha,n-1} * s * \sqrt{1+\frac{1}{n}}$ Upper limit: $+\infty$	Lower limit: $\bar{x} - C_{\alpha,k}s$ Upper limit: $+\infty$ One sided C value

Example 2(a) Escape from emergency

A sample of 25 offshore workers at an oil drilling company participate in the study on escape time during emergency. The mean and standard deviation of the escape time is 370.69 and 24.36. Assume, escape time is normally distributed.

What is the two-sided interval which contains standard deviation of the entire population with 95% probability?

(19.02, 33.89)

Example 2(b) Escape from emergency

A sample of 25 offshore workers at an oil drilling company participate in the study on escape time during emergency. The mean and standard deviation of the escape time is 370.69 and 24.36. Assume, escape time is normally distributed.

Suppose another worker is selected at random. What is the upper limit K such that the escape time of this worker is lower than K with 95% probability?

370.69+1.71*24.36*1.02 = 413.18

What is the interval which contains escape time of this worker with 99% probability?

(301.19, 440.19)

Example 2(c) Escape from emergency

A sample of 25 offshore workers at an oil drilling company participate in the study on escape time during emergency. The mean and standard deviation of the escape time is 370.69 and 24.36. Assume, escape time is normally distributed.

What is the upper limit K such that 90% of the workers will escape in time less than K with 99% probability?

370.69+2.129*24.36 = 422.55

Learning outcomes- Hypothesis Testing

- **Formulate the null hypothesis and alternative hypothesis** asked in the problem
- Determine **Type I and Type II error** given sample data
- Use the **correct formula for the determining whether to reject or not reject** a null hypothesis
  - By formulating the rejection region
  - By using p-values
- **Interpret the meaning** of results from hypothesis test

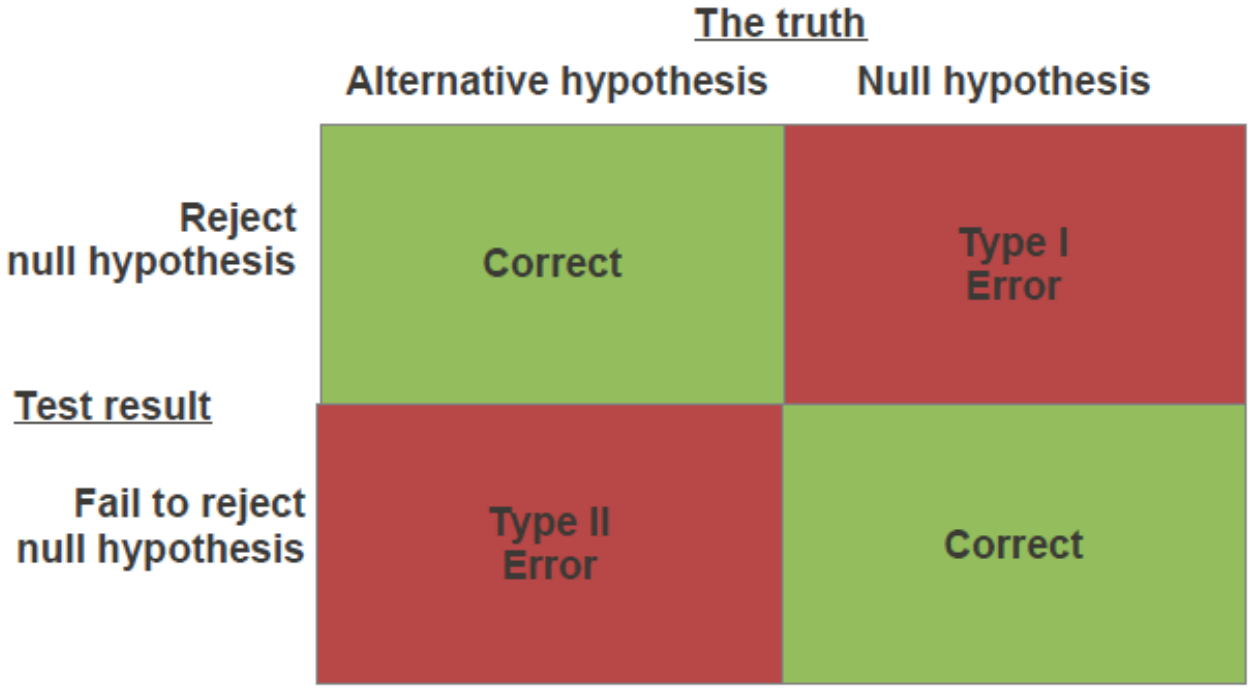
Review part 1

- Null hypothesis and alternate hypothesis
  - Null hypothesis is what you initially assume to be true (or what you will conclude if the evidence is ambiguous)
  - Alternative hypothesis is what you will believe only in the face of conclusive evidence (or what you want to test for)

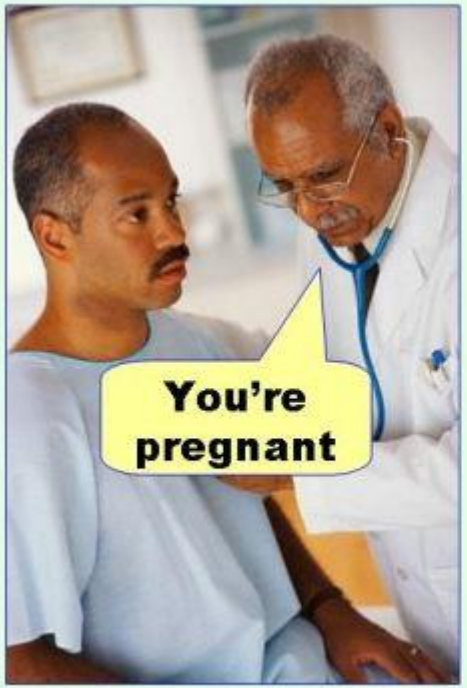
- The null and alternate hypothesis are statements about _____ parameter

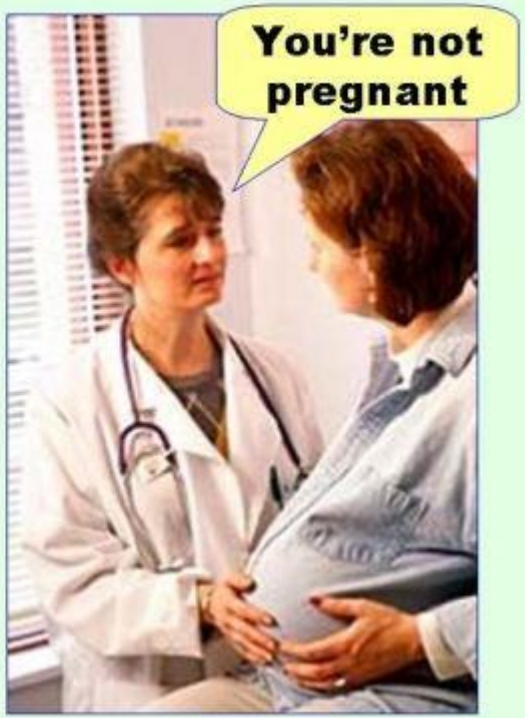
- Three types of alternate hypothesis

Review part 1



Type of error?



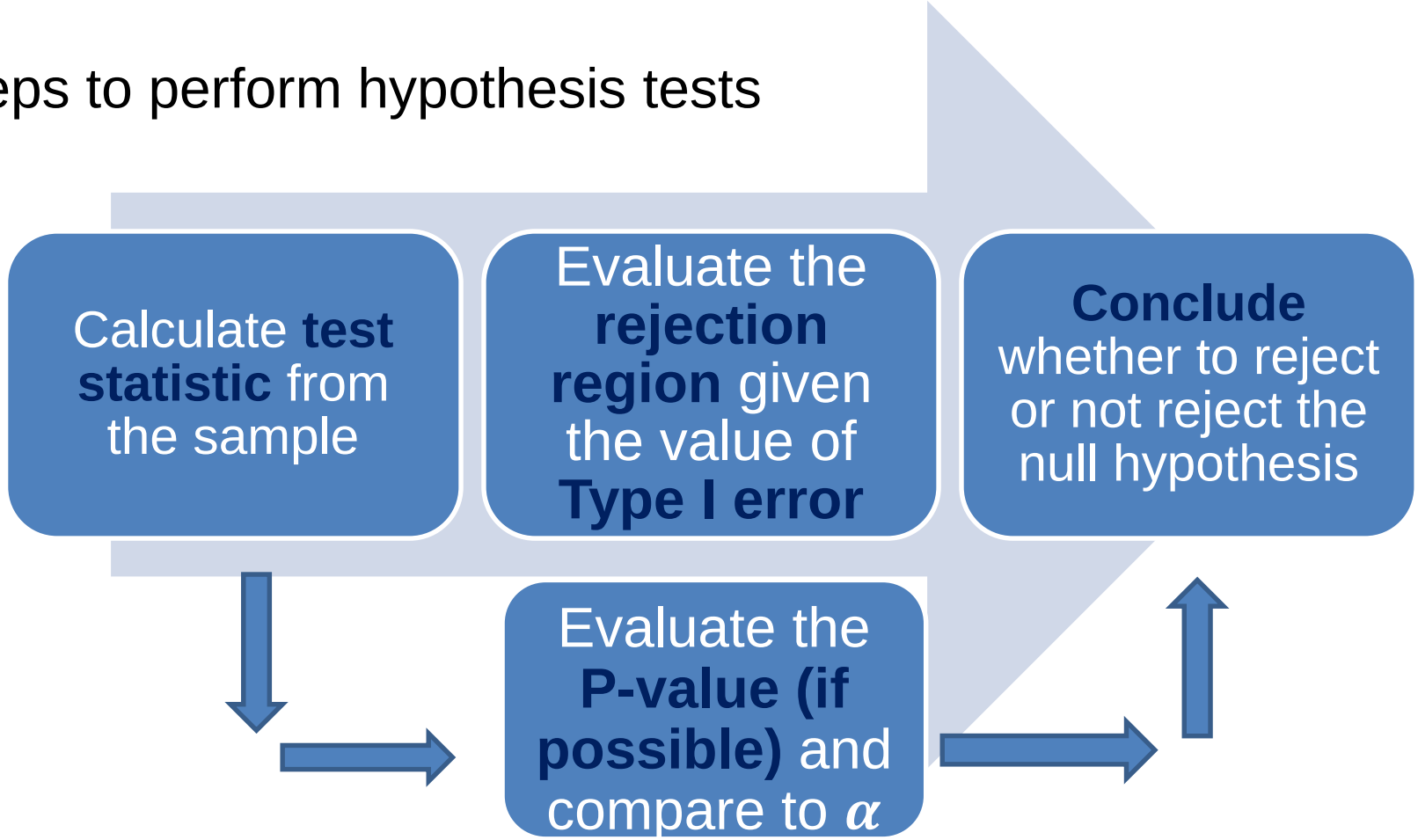


Remember, Type I error is false positive, and Type II error is false negative.

Example 1: Determine null and alternate hypothesis

- Let p be the proportion of people cured by a drug; the old drug cures 25%. Perform hypothesis test to check if the new drug is more effective.
  - $H_0: p = 0.25;\ H_a: p > 0.25$
- Proportion of students enrolled in clubs at UT is typically 70%. Perform hypothesis test to see if the proportion has changed in 2017.
  - $H_0: p = 0.70;\ H_a: p \neq 0.70$
- Let μ denote the true average radioactive level. The value 5 units is considered dividing line between safe and unsafe water. Perform hypothesis test to check if the radioactive level at a site is safe.
  - $H_0: \mu = 5;\ H_a: \mu < 5$

Steps to perform hypothesis tests



Test statistic

Population Parameter	Sample mean
Null hypothesis	
Test statistic	
Case 1	$n > 40$
Case 2	$n < 40$ AND population distribution is _____
Otherwise	Consult a knowledgeable statistician

Population Parameter	Sample proportion
Null hypothesis	$p = p_0$
Test statistic	
Case 1	$np_0 > 10$ AND $n(1 - p_0) > 10$
Otherwise	Consult a knowledgeable statistician

Review: Evaluating a hypothesis using rejection regions

Hypothesis test type	Rejection region Case 1	Rejection region Case 2
$H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$		
$H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$		
$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$		

If test statistic lies in the rejection region, then reject the null hypothesis

Example 2: Find rejection region

The drying time of certain type of paint is known to be normally distributed with mean 75 min and std dev. 9 min. Chemists have proposed a new additive designed to decrease the average drying time. It is believed that the drying times with this additive will remain normally distributed with $\sigma = 9$. You conduct an experiment using 25 samples and find the sample mean to be 70 min.

Set up the Null and Alternate Hypothesis

H_0: $\mu = 75$; H_a: $\mu < 75$

Perform a hypothesis test for a significance level = 0.01. Interpret the result

Case 1 since n<25 AND distribution is normal AND population variance is known

Test statistic: $\frac{70-75}{\frac{9}{\sqrt{25}}} = -2.777$

Rejection region: $(-\infty, -z_{0.01}) = (-\infty, -2.33)$

Since $-2.777 \in (-\infty, -2.33)$, reject null hypothesis

Reject null hypothesis, i.e. new additive is effective in decreasing the drying time

Review: Evaluating a hypothesis using p-values

Hypothesis test type	P-values Case 1
$H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$	
$H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$	
$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$	

- Let $\bar{x}$ be the sample parameter and $z = (\bar{x} - \mu)\sqrt{n}/\sigma$ be the z-statistic
- If level of significance is **higher than p-value** reject; else accept

Example 2: Find rejection region

The drying time of certain type of paint is known to be normally distributed with mean 75 min and std dev. 9 min. Chemists have proposed a new additive designed to decrease the average drying time. It is believed that the drying times with this additive will remain normally distributed with $\sigma = 9$. You conduct an experiment using 25 samples and find the sample mean to be 70 min.

Find P-value for the given experiment

The P-value for the given hypothesis test is given by $\varphi(-2.77) = 0.0028$

Perform a hypothesis test for a significance level = 0.01 using this p-value

Since 0.01>0.0028, so reject the null hypothesis

Example 2: Find rejection region

The drying time of certain type of paint is known to be normally distributed with mean 75 min and std dev. 9 min. Chemists have proposed a new additive designed to decrease the average drying time. It is believed that the drying times with this additive will remain normally distributed with $\sigma = 9$. You conduct an experiment using 25 samples and find the sample mean to be 70 min.

Estimate the Type I error

0.01

Example 2: Find rejection region

The drying time of certain type of paint is known to be normally distributed with mean 75 min and std dev. 9 min. Chemists have proposed a new additive designed to decrease the average drying time. It is believed that the drying times with this additive will remain normally distributed with $\sigma = 9$. You conduct an experiment using 25 samples and find the sample mean to be 70 min.

Estimate the Type II error when the true mean is actually 72

H_0: $\mu = 75$; H_a: $\mu < 75$

Step 1: Find rejection region on z-critical value (we use z-critical value since we know the TRUE population variance; Case 1)

$$(-\infty, -z_\alpha) = (-\infty, -z_{0.01}) = (-\infty, -2.33)$$

Step 2: Convert the rejection region to a rejection region on sample mean

$$(-\infty, -2.33 * 1.8 + 75) = (-\infty, 70.806\ )$$

(contd…)

Example 2: Find rejection region

The drying time of certain type of paint is known to be normally distributed with mean 75 min and std dev. 9 min. Chemists have proposed a new additive designed to decrease the average drying time. It is believed that the drying times with this additive will remain normally distributed with $\sigma = 9$. You conduct an experiment using 25 samples and find the sample mean to be 70 min.

Step 3: find the error:

P(Type II error) = P(not rejecting null hypothesis when it not true)

We fail to reject null hypothesis when $(\bar{x} > 70.806)$

P(Type II error) = $P(\bar{x} > 72.039 \mid \mu = 72)$

Recall, $\bar{x} = \mu + Z * \sigma/\sqrt{n}$

We know, $\sigma = 9, n = 25, \mu = 72$

$$P(\bar{x} > 70.806) = P\left(Z > (70.806 - 72) * \frac{5}{9}\right) = P(Z > -0.6633) = 0.746$$