



Unuttered Questions of Statistical Programmers

PhUSE 2014, London

Aparajita Dey, Cytel

Disclaimer

Any comments or statements made herein solely those of the author and do not necessarily represent those of the company.

I am grateful to my colleagues for allowing me to use their photos in this presentation. The names are changed for privacy.

Concept and Flow

- Inspiration: Communication between Statisticians and Programmers
- Common questions asked by programmers – “What”, “How”, “When”
- Remains unuttered – “**Why**”



Case Study 1

Case Study 1: Random Effects

Tannu Mishra
SAS Programmer

Case Study 1: Random Effects

All confidence intervals constructed for pharmacokinetic parameters will be based on the least-squares means and variance components arising from a linear mixed effects model with **treatment and study period as a fixed effect and with subject as a random effect**.



Pharmacokinetic Parameter	Study Drug			Co-administration of Two Other Marketed Drugs			Study Drug / Co-administration	
	N	GM	90% CI	N	GM	90% CI	GMR	90% CI
AUC _{0-∞} [‡] (nM.hr)	xx	xx	(xx, xx)	xx	xx	(xx, xx)	xx.xx	(xx.xx, xx.xx)
C _{max} [‡] (nM)	xx	xx	(xx, xx)	xx	xx	(xx, xx)	xx.xx	(xx.xx, xx.xx)
[‡] Back-transformed least squares mean and confidence interval from linear mixed effects model with treatment and study period included as fixed effect and subject included as random effect ; performed on natural log-transformed values;								
GMR = Geometric least squares mean ratio, GM = Geometric Least-Squares Mean, CI = Confidence Interval								

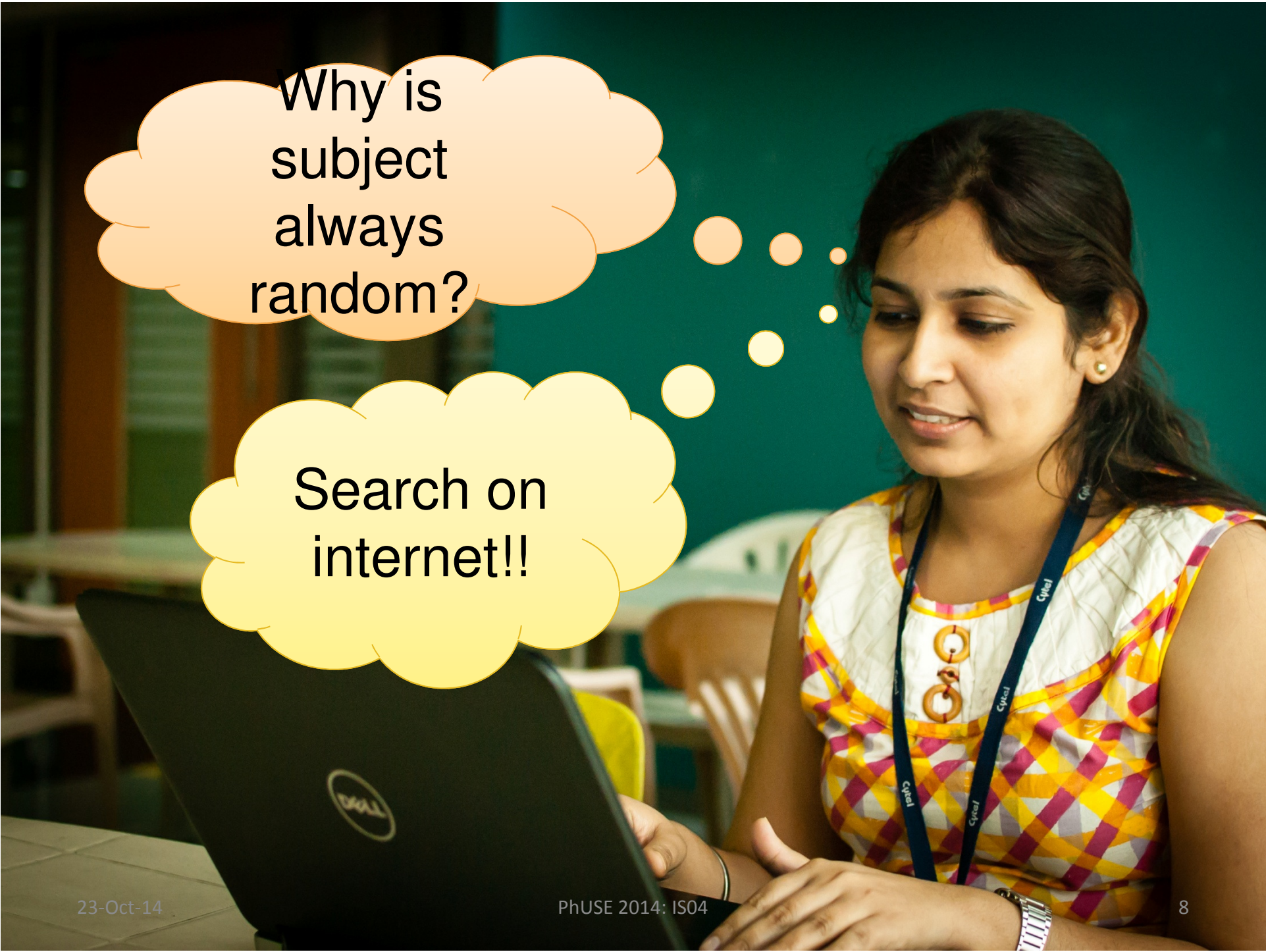
Case Study 1: Random Effects

Liver Function Tests will be analyzed with an analysis of covariance model. The dependent variable will be the log transformed LFT value. The model includes the **baseline measure, dose group, visit, and dose group by visit interaction. Subject will be included as a random effect.**



	Dose Group 1 (N = XX)	Dose Group 2 (N = XX)
Baseline Visit		
N	xx	xx
L.S. Mean	xx.x	xx.x
90% CI	(xx.x, xx.x)	(xx.x, xx.x)

Continued...

A woman with dark hair, wearing a yellow and pink patterned top and a blue lanyard, is sitting at a table and looking at a black Dell laptop. Two thought bubbles are overlaid on the image. The first bubble is orange and contains the text 'Why is subject always random?'. The second bubble is yellow and contains the text 'Search on internet!!'. The background is a blurred indoor setting with a teal wall and wooden chairs.

Why is
subject
always
random?

Search on
internet!!

Case Study 1: Random Effects

why subject is random effect?



... an effect is classified as a random effect when you want to make inferences on an entire population

... to be able to generalize the results to the so called population level, a Random Effects approach is necessary.

...data consists of a hierarchy of different populations whose differences relate to that hierarchy

biostatisticians use "fixed" and "random" effects to respectively refer to the population-average and subject-specific effects.

You visited this page on 20/9/14.

Did Tannu get the
answer to her
question?

No.

Case Study 1: Random Effects

Pharmacokinetic Parameter	Study Drug			Co-administration of Two Other Marketed Drugs			Study Drug / Co-administration	
	N	GM	90% CI	N	GM	90% CI	GMR	90% CI
$AUC_{0-\infty}^{\ddagger}$ (nM.hr)	22	8027	(7767, 8297)	20	7931	(7675, 8196)	1.01	(1.00, 1.03)
$C_{max}^{\ddagger}$ (nM)	21	895	(849, 945)	20	867	(822, 914)	1.03	(0.99, 1.08)
$\ddagger$ Back-transformed least squares mean and confidence interval from linear mixed effects model with treatment and study period included as fixed effect and subject included as random effect ; performed on natural log-transformed values;								
GMR = Geometric least squares mean ratio, GM = Geometric Least-Squares Mean, CI = Confidence Interval								

PK Parameter	Study Period	Treatment	Subject
$AUC_{0-\infty}$ C_{max}	1 2	- IP - Co-administration	1001 1002 ... 1025

Case Study 1: Random Effects



PK Parameter

- $AUC_{0-\infty}$
- C_{max}

Study Period

- 1
- 2

Treatment

- IP
- Co-administration

Subject

- 1001
- 1002 ...
- 1025

Case Study 2

Case Study 2: Deviation vs. Error

Bunty Jadhav
SAS Programmer

Case Study 2: Deviation Vs. Error

Example
Table Shell

Summary of Systolic BP mmHg (beats/minute) by Visit	Dose Group (N = xx)
--	------------------------

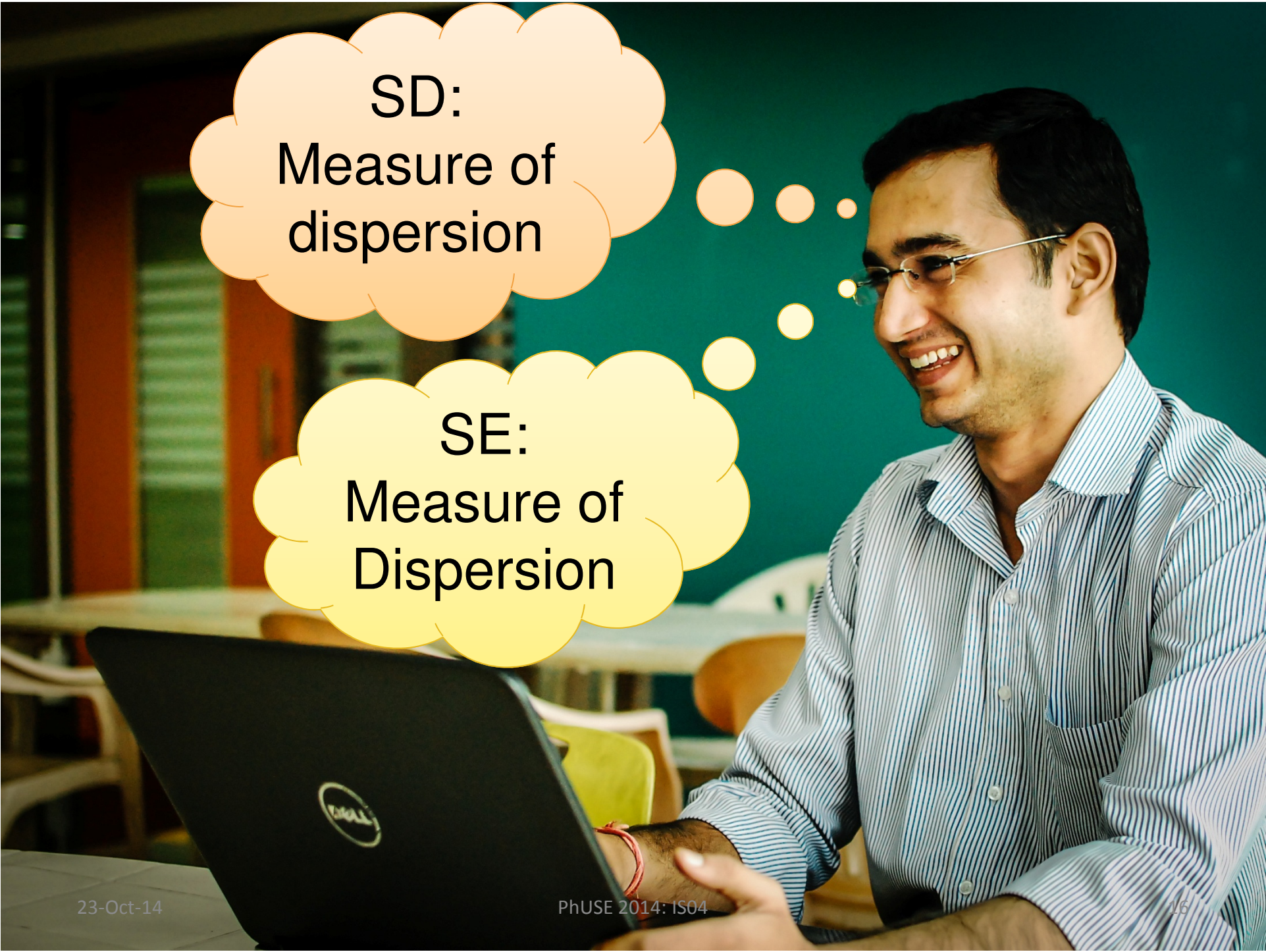
Baseline

n	xx
Mean	xx.x
SD	xx.x
SE	xx.x
Median	xx
Min, Max	xx, xx

Dose Group (N = xx)	
------------------------	--

Age (years)

n	Xx
Mean	xx.x
SD	xx.x
Median	xx.x
Q1, Q3	xx.x, xx.x
Min, Max	xx, xx



SD:
Measure of
dispersion

SE:
Measure of
Dispersion

Case Study 2: Deviation Vs. Error

SD = Standard Deviation

SE = Standard Error
of Sample Mean

Definition

Spread of data

SD of Sample Mean

If measured in a sample, it estimates –

Spread of population
data

Accuracy of sample
mean as an estimate of
population mean

If no sampling string attached, it serves as –

Descriptive Statistics
measuring dispersion

No Significance

Case Study 3

Case Study 3: P-value

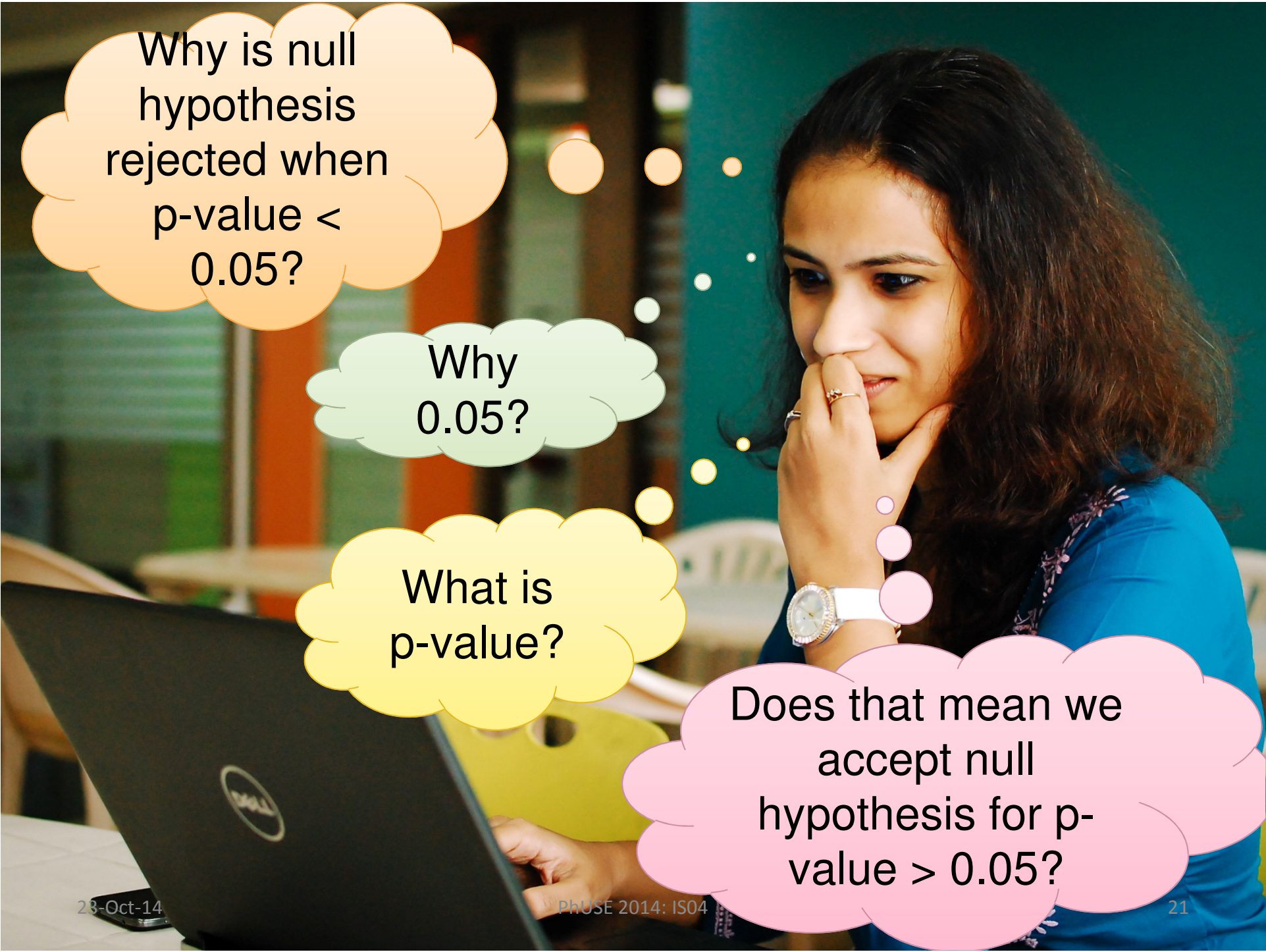
Kirti Inamdar
SAS Programmer

Case Study 3: P-value

Example
Table Shell

	objective responses	Odds ratio	95% Confidence Interval for Odds Ratio	P-value
Objective responses	XX			
IP : Marketed Drug		X.XX	(X.XX, X.XX)	0.XXXX
Covariate 1 Level 1: Level 0		X.XX	(X.XX, X.XX)	0.XXXX
Covariate 2 Level 1 : Level 0		X.XX	(X.XX, X.XX)	0.XXXX

	Marketed Drug (N = XXX)	IP (N = XXX)
Subjects with events - n(%)	XX (XX)	XX (XX)
Disease progression	XX (XX)	XX (XX)
Death without disease progression	XX (XX)	XX (XX)
Censored Subjects - n(%)	XX (XX)	XX (XX)
Stratified Cox proportional hazards model		
Hazard ratio		X.XXX
95% CI		X.XXX,X.XXX
P-value for treatment effect		0.XXXX



Why is null hypothesis rejected when $p\text{-value} < 0.05$?

Why 0.05?

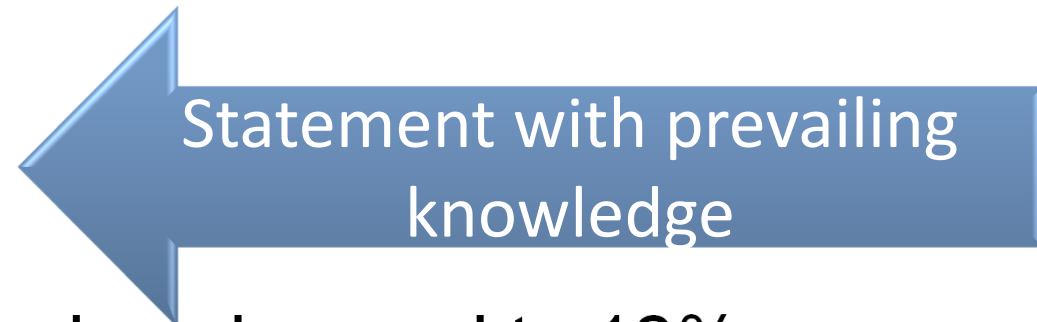
What is p-value?

Does that mean we accept null hypothesis for $p\text{-value} > 0.05$?

Case Study 3: P-value

Example:

Null Hypothesis:



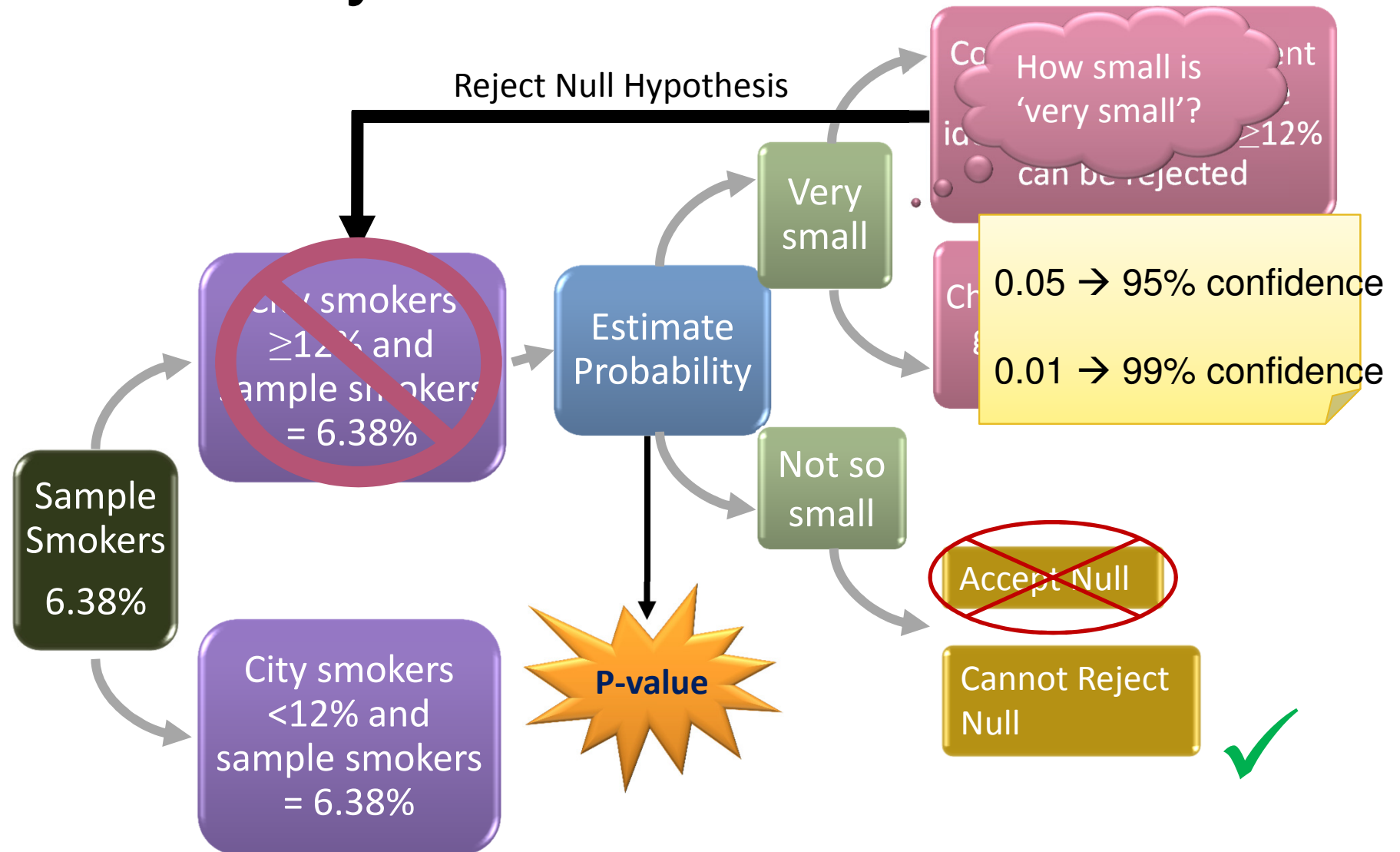
The percentage of smokers is equal to 12%, vs.

Alternative Hypothesis



the percentage of smokers is less than 12%

Case Study 3: P-value



Summary

Why is Subject always used as random effect in linear models?

Why p-value is not generally used in Safety tables?

Why reject null hypothesis for p-value less than 0.05?

between

What is degrees of freedom?

Why are t-test and paired t-test different?

Why are analyses?

Why are there two types of error bars—“mean +/- SE” and “Mean and CI”?

Summary

- ☑ Gaining skill in Statistics while concentrating on programming is extremely difficult time taking and not always feasible
- ☑ Cannot target to know everything but that does not stop us from starting the process
- ☑ No harm in asking questions – to Statisticians, to Programmers who are more experienced
- ☑ Build a Glossary – responsibility that comes with experience

Questions?



Thank you



Aparajita.Dey@Cytel.com