UCLA-ALIGNED COURSE FIELD NOTES

ECON 103

Introduction to Econometrics

Univariate regression, least squares, inference, robust errors, serial correlation, and R

How these notes are written. I wrote each chapter the way I wish the material had been explained the first time: the real idea, the point where I usually get stuck, the simplest mental model that still stays honest, and a worked decision instead of a polished answer appearing from nowhere. Every chapter closes with practice and a fully written solution. If a sentence sounds impressive but does not help me solve or explain something, it does not belong here.

Daniel Mastick

Curriculum map, working notes, practice, and solutions

Course map

Week	Core thread	What should be solid by the end
1	**Economic questions, data, and identification**	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
2	**Simple linear regression and least squares**	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
3	**OLS assumptions and the Gauss-Markov theorem**	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
4	**Sampling distributions and inference**	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
5	**Functional form and interpretation**	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
6	**Heteroskedasticity and robust inference**	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
7	**Time order, serial correlation, and trends**	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
8	**Reproducible R workflow and honest reporting**	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
9	**Integration studio**	Combine several chapters in one case, preserve their assumptions, and reconcile the result across representations.
10	**Cumulative review**	Retrieve the full course map from memory and solve mixed problems without chapter labels.

Scope anchor: https://catalog.registrar.ucla.edu/course/2025/ECON103?year=2025. The sequence is aligned to the official catalog description and expanded into a practical ten-week study plan.

How I would use this packet

Treat each numbered section as one chapter. Read the formal explanation, pause at the student interpretation, and see whether the easy version still says the same thing. Rework the example without looking, then cover the solution in the chapter practice box. The cumulative set at the end is deliberately mixed: knowledge becomes durable when I have to choose the tool instead of being told which chapter I am in.

Contents

CHAPTER 1 Economic questions, data, and identification

The idea

Econometrics connects a question to a population, variables, data-generating process, and estimand. Prediction, description, and causal inference require different assumptions; observational association is not automatically a treatment effect.

Rebuild the chapter from first principles

The claim I need to own is this: Econometrics connects a question to a population, variables, data-generating process, and estimand. Prediction, description, and causal inference require different assumptions; observational association is not automatically a treatment effect. I do not count that as learned just because the sentence looks familiar. I should be able to name the objects, state what changes and what stays fixed, recover the rule from the definitions, and explain why the result is allowed.

The move

My working sequence

1. **Translate.** Rewrite the prompt in the language of economic questions, data, and identification. List the known quantities, the unknown, the units or types, and any hidden constraint.
2. **Choose.** State the theorem, model, algorithm, accounting rule, or physical law before substituting values. This is where I make the assumption visible.
3. **Execute.** Work one justified step at a time and keep intermediate state readable enough that I can audit it later.
4. **Verify.** Check a boundary case, sign, unit, invariant, order of magnitude, or alternative derivation. Then translate the result back into an ordinary sentence.

Details I would refuse to skip

- **Definitions:** I should be able to define every technical noun in the chapter without using that same noun in the definition.
- **Assumptions:** I write the condition that makes the method legal beside the method itself. A correct procedure under the wrong assumptions is still a wrong answer.
- **Representation:** I move freely among words, equations or code, a diagram, and a small numerical example. If one representation disagrees with another, that disagreement is useful evidence.
- **Failure modes:** I keep the nearest counterexample in mind: the boundary where the rule changes, the invariant breaks, or the model stops matching reality.
- **Communication:** My final answer names what the result means and what it does not establish. I do not let notation do the explanatory work for me.

Sanity check

Closed-note check. Can I teach economic questions, data, and identification on a blank page, derive or reconstruct its main rule, work a fresh example, and diagnose a deliberately broken solution? If I can only recognize the finished work, I am not done.

How this chapter connects

I read this chapter as the bridge from the course foundations to simple linear regression and least squares. The earlier material supplies the language and constraints; this chapter adds a new reasoning move; the later material asks me to use that move without being told its name. That connection matters because exams and real projects mix chapters on purpose.

The full working model

- **Core claim.** Econometrics connects a question to a population, variables, data-generating process, and estimand.
- **What moves.** Prediction, description, and causal inference require different assumptions; observational association is not automatically a treatment effect.
- **What keeps the answer honest.** The model becomes useful only when I can apply economic questions, data, and identification to an unfamiliar case and defend the assumptions.
- **Mastery standard.** I can define the objects, rebuild the rule, work a fresh case, compare alternatives, and diagnose a broken solution without relying on recognition.

How I actually think about it

I write the estimand in words before opening R, because clean code cannot rescue a vague question.

The easy version

Econometrics connects a question to a population, variables, data-generating process, and estimand.

Worked example

Question. A regression finds that students who attend tutoring score higher. Is tutoring's causal effect identified?
Walkthrough. Not without assumptions or design. Motivation, prior achievement, scheduling, and selection may affect both attendance and scores; random assignment or a credible quasi-experiment is needed.

Common miss

The fastest way to get this chapter wrong is to use the visible procedure without naming its assumption, units, time period, reference group, or boundary. I put that condition beside the work and verify the result independently.

Solved chapter practice

Practice 1. A regression finds that students who attend tutoring score higher. Is tutoring's causal effect identified?
Solution. Not without assumptions or design. Motivation, prior achievement, scheduling, and selection may affect both attendance and scores; random assignment or a credible quasi-experiment is needed.

Practice 2. Explain economic questions, data, and identification to a classmate using one definition, one concrete example, one assumption, and one failure mode.
Solution. Begin with the core claim above, use the worked case as the concrete example, state the condition highlighted by the warning, and choose a failure where that condition does not hold. A complete explanation connects all four pieces instead of listing them.

Mastery practice A. Reconstruct economic questions, data, and identification from a blank page. Include the central definition, the governing rule, one required assumption, and one boundary case.
Solution. Start from the blue idea box, but rewrite it in my own words. Identify the objects and state, name the rule before using it, copy the relevant assumption from the details audit, and choose a boundary where that assumption becomes equality or fails. A complete solution explains why each part belongs; a list of vocabulary is not enough.

Mastery practice B. A classmate gets a polished-looking answer to a economic questions, data, and identification problem but never states a model or checks the result. Write the shortest useful review of that solution.
Solution. I would ask four concrete questions: What exactly are the inputs and desired output? Which rule licenses the main step? Which assumption could invalidate it? Which independent check supports the conclusion? If the answer cannot survive those four questions, I treat the numerical or verbal result as unverified.

CHAPTER 2 Simple linear regression and least squares

The idea

The population regression separates conditional mean from disturbance. OLS chooses intercept and slope to minimize squared residuals, producing orthogonality conditions and a fitted line through sample means.

Rebuild the chapter from first principles

The claim I need to own is this: The population regression separates conditional mean from disturbance. OLS chooses intercept and slope to minimize squared residuals, producing orthogonality conditions and a fitted line through sample means. I do not count that as learned just because the sentence looks familiar. I should be able to name the objects, state what changes and what stays fixed, recover the rule from the definitions, and explain why the result is allowed.

The move

My working sequence

1. **Translate.** Rewrite the prompt in the language of simple linear regression and least squares. List the known quantities, the unknown, the units or types, and any hidden constraint.
2. **Choose.** State the theorem, model, algorithm, accounting rule, or physical law before substituting values. This is where I make the assumption visible.
3. **Execute.** Work one justified step at a time and keep intermediate state readable enough that I can audit it later.
4. **Verify.** Check a boundary case, sign, unit, invariant, order of magnitude, or alternative derivation. Then translate the result back into an ordinary sentence.

Details I would refuse to skip

- **Definitions:** I should be able to define every technical noun in the chapter without using that same noun in the definition.
- **Assumptions:** I write the condition that makes the method legal beside the method itself. A correct procedure under the wrong assumptions is still a wrong answer.
- **Representation:** I move freely among words, equations or code, a diagram, and a small numerical example. If one representation disagrees with another, that disagreement is useful evidence.
- **Failure modes:** I keep the nearest counterexample in mind: the boundary where the rule changes, the invariant breaks, or the model stops matching reality.
- **Communication:** My final answer names what the result means and what it does not establish. I do not let notation do the explanatory work for me.

Sanity check

Closed-note check. Can I teach simple linear regression and least squares on a blank page, derive or reconstruct its main rule, work a fresh example, and diagnose a deliberately broken solution? If I can only recognize the finished work, I am not done.

How this chapter connects

I read this chapter as the bridge from economic questions, data, and identification to ols assumptions and the gauss-markov theorem. The earlier material supplies the language and constraints; this chapter adds a new reasoning move; the later material asks me to use that move without being told its name. That connection matters because exams and real projects mix chapters on purpose.

The full working model

- **Core claim.** The population regression separates conditional mean from disturbance.
- **What moves.** OLS chooses intercept and slope to minimize squared residuals, producing orthogonality conditions and a fitted line through sample means.

- **What keeps the answer honest.** The model becomes useful only when I can apply simple linear regression and least squares to an unfamiliar case and defend the assumptions.
- **Mastery standard.** I can define the objects, rebuild the rule, work a fresh case, compare alternatives, and diagnose a broken solution without relying on recognition.

How I actually think about it

I see OLS as a projection with a specific loss function, not a machine that discovers truth.

The easy version

The population regression separates conditional mean from disturbance.

Worked example

Question. If covariance of x and y is 12 and variance of x is 4, what is the simple-regression slope?
Walkthrough. The slope is covariance divided by variance, or 3. The intercept is the y mean minus 3 times the x mean.

Common miss

The fastest way to get this chapter wrong is to use the visible procedure without naming its assumption, units, time period, reference group, or boundary. I put that condition beside the work and verify the result independently.

Solved chapter practice

Practice 1. If covariance of x and y is 12 and variance of x is 4, what is the simple-regression slope?
Solution. The slope is covariance divided by variance, or 3. The intercept is the y mean minus 3 times the x mean.

Practice 2. Explain simple linear regression and least squares to a classmate using one definition, one concrete example, one assumption, and one failure mode.
Solution. Begin with the core claim above, use the worked case as the concrete example, state the condition highlighted by the warning, and choose a failure where that condition does not hold. A complete explanation connects all four pieces instead of listing them.

Mastery practice A. Reconstruct simple linear regression and least squares from a blank page. Include the central definition, the governing rule, one required assumption, and one boundary case.
Solution. Start from the blue idea box, but rewrite it in my own words. Identify the objects and state, name the rule before using it, copy the relevant assumption from the details audit, and choose a boundary where that assumption becomes equality or fails. A complete solution explains why each part belongs; a list of vocabulary is not enough.

Mastery practice B. A classmate gets a polished-looking answer to a simple linear regression and least squares problem but never states a model or checks the result. Write the shortest useful review of that solution.

Solution. I would ask four concrete questions: What exactly are the inputs and desired output? Which rule licenses the main step? Which assumption could invalidate it? Which independent check supports the conclusion? If the answer cannot survive those four questions, I treat the numerical or verbal result as unverified.

CHAPTER 3 OLS assumptions and the Gauss-Markov theorem

The idea

Linearity in parameters, random sampling, variation in x, zero conditional mean, and homoskedasticity support standard results. Under Gauss-Markov conditions, OLS is the best linear unbiased estimator, not necessarily best among every conceivable estimator.

Rebuild the chapter from first principles

The claim I need to own is this: Linearity in parameters, random sampling, variation in x, zero conditional mean, and homoskedasticity support standard results. Under Gauss-Markov conditions, OLS is the best linear unbiased estimator, not necessarily best among every conceivable estimator. I do not count that as learned just because the sentence looks familiar. I should be able to name the objects, state what changes and what stays fixed, recover the rule from the definitions, and explain why the result is allowed.

The move

My working sequence

1. **Translate.** Rewrite the prompt in the language of ols assumptions and the gauss-markov theorem. List the known quantities, the unknown, the units or types, and any hidden constraint.
2. **Choose.** State the theorem, model, algorithm, accounting rule, or physical law before substituting values. This is where I make the assumption visible.
3. **Execute.** Work one justified step at a time and keep intermediate state readable enough that I can audit it later.
4. **Verify.** Check a boundary case, sign, unit, invariant, order of magnitude, or alternative derivation. Then translate the result back into an ordinary sentence.

Details I would refuse to skip

- **Definitions:** I should be able to define every technical noun in the chapter without using that same noun in the definition.
- **Assumptions:** I write the condition that makes the method legal beside the method itself. A correct procedure under the wrong assumptions is still a wrong answer.
- **Representation:** I move freely among words, equations or code, a diagram, and a small numerical example. If one representation disagrees with another, that disagreement is useful evidence.

- **Failure modes:** I keep the nearest counterexample in mind: the boundary where the rule changes, the invariant breaks, or the model stops matching reality.
- **Communication:** My final answer names what the result means and what it does not establish. I do not let notation do the explanatory work for me.

Sanity check

Closed-note check. Can I teach ols assumptions and the gauss-markov theorem on a blank page, derive or reconstruct its main rule, work a fresh example, and diagnose a deliberately broken solution? If I can only recognize the finished work, I am not done.

How this chapter connects

I read this chapter as the bridge from simple linear regression and least squares to sampling distributions and inference. The earlier material supplies the language and constraints; this chapter adds a new reasoning move; the later material asks me to use that move without being told its name. That connection matters because exams and real projects mix chapters on purpose.

The full working model

- **Core claim.** Linearity in parameters, random sampling, variation in x, zero conditional mean, and homoskedasticity support standard results.
- **What moves.** Under Gauss-Markov conditions, OLS is the best linear unbiased estimator, not necessarily best among every conceivable estimator.
- **What keeps the answer honest.** The model becomes useful only when I can apply ols assumptions and the gauss-markov theorem to an unfamiliar case and defend the assumptions.
- **Mastery standard.** I can define the objects, rebuild the rule, work a fresh case, compare alternatives, and diagnose a broken solution without relying on recognition.

How I actually think about it

My shorthand for this chapter is: my shorthand for this chapter is: the zero conditional mean assumption carries the causal weight; the algebra is the easy part.

The easy version

Linearity in parameters, random sampling, variation in x, zero conditional mean, and homoskedasticity support standard results.

Worked example

Question. What does omitted-variable bias require?
Walkthrough. The omitted factor must affect the outcome and be correlated with the included regressor. The bias direction follows the product of those relationships.

Common miss

The fastest way to get this chapter wrong is to use the visible procedure without naming its assumption, units, time period, reference group, or boundary. I put that condition beside the work and verify the result independently.

Solved chapter practice

Practice 1. What does omitted-variable bias require?
Solution. The omitted factor must affect the outcome and be correlated with the included regressor. The bias direction follows the product of those relationships.

Practice 2. Explain ols assumptions and the gauss-markov theorem to a classmate using one definition, one concrete example, one assumption, and one failure mode.
Solution. Begin with the core claim above, use the worked case as the concrete example, state the condition highlighted by the warning, and choose a failure where that condition does not hold. A complete explanation connects all four pieces instead of listing them.

Mastery practice A. Reconstruct ols assumptions and the gauss-markov theorem from a blank page. Include the central definition, the governing rule, one required assumption, and one boundary case.
Solution. Start from the blue idea box, but rewrite it in my own words. Identify the objects and state, name the rule before using it, copy the relevant assumption from the details audit, and choose a boundary where that assumption becomes equality or fails. A complete solution explains why each part belongs; a list of vocabulary is not enough.

Mastery practice B. A classmate gets a polished-looking answer to a ols assumptions and the gauss-markov theorem problem but never states a model or checks the result. Write the shortest useful review of that solution.
Solution. I would ask four concrete questions: What exactly are the inputs and desired output? Which rule licenses the main step? Which assumption could invalidate it? Which independent check supports the conclusion? If the answer cannot survive those four questions, I treat the numerical or verbal result as unverified.

CHAPTER 4 Sampling distributions and inference

The idea

Standard errors measure estimator uncertainty; t statistics, confidence intervals, p-values, and tests compare data with a null model. Statistical significance, economic magnitude, and model validity are separate judgments.

Rebuild the chapter from first principles

The claim I need to own is this: Standard errors measure estimator uncertainty; t statistics, confidence intervals, p-values, and tests compare data with a null model. Statistical significance, economic magnitude, and model validity are separate judgments. I do not count that as learned just because the sentence looks familiar. I should be able to name the objects, state what changes and what stays fixed, recover the rule from the definitions, and explain why the result is allowed.

The move

My working sequence

1. **Translate.** Rewrite the prompt in the language of sampling distributions and inference. List the known quantities, the unknown, the units or types, and any hidden constraint.
2. **Choose.** State the theorem, model, algorithm, accounting rule, or physical law before substituting values. This is where I make the assumption visible.
3. **Execute.** Work one justified step at a time and keep intermediate state readable enough that I can audit it later.
4. **Verify.** Check a boundary case, sign, unit, invariant, order of magnitude, or alternative derivation. Then translate the result back into an ordinary sentence.

Details I would refuse to skip

- **Definitions:** I should be able to define every technical noun in the chapter without using that same noun in the definition.
- **Assumptions:** I write the condition that makes the method legal beside the method itself. A correct procedure under the wrong assumptions is still a wrong answer.
- **Representation:** I move freely among words, equations or code, a diagram, and a small numerical example. If one representation disagrees with another, that disagreement is useful evidence.
- **Failure modes:** I keep the nearest counterexample in mind: the boundary where the rule changes, the invariant breaks, or the model stops matching reality.
- **Communication:** My final answer names what the result means and what it does not establish. I do not let notation do the explanatory work for me.

Sanity check

Closed-note check. Can I teach sampling distributions and inference on a blank page, derive or reconstruct its main rule, work a fresh example, and diagnose a deliberately broken solution? If I can only recognize the finished work, I am not done.

How this chapter connects

I read this chapter as the bridge from ols assumptions and the gauss-markov theorem to functional form and interpretation. The earlier material supplies the language and constraints; this chapter adds a new reasoning move; the later material asks me to use that move without being told its name. That connection matters because exams and real projects mix chapters on purpose.

The full working model

- **Core claim.** Standard errors measure estimator uncertainty; t statistics, confidence intervals, p-values, and tests compare data with a null model.
- **What moves.** Statistical significance, economic magnitude, and model validity are separate judgments.
- **What keeps the answer honest.** The model becomes useful only when I can apply sampling distributions and inference to an unfamiliar case and defend the assumptions.
- **Mastery standard.** I can define the objects, rebuild the rule, work a fresh case, compare alternatives, and diagnose a broken solution without relying on recognition.

How I actually think about it

I report the estimate, units, uncertainty, and practical meaning before the p-value.

The easy version

Standard errors measure estimator uncertainty; t statistics, confidence intervals, p-values, and tests compare data with a null model.

Worked example

Question. An estimate is 2.4 with standard error 0.8. Test zero using a large-sample 5 percent rule.
Walkthrough. The t statistic is 3.0, exceeding about 1.96 in magnitude, so reject zero. A rough 95 percent interval is 2.4 plus or minus 1.57, or about 0.83 to 3.97.

Common miss

The fastest way to get this chapter wrong is to use the visible procedure without naming its assumption, units, time period, reference group, or boundary. I put that condition beside the work and verify the result independently.

Solved chapter practice

Practice 1. An estimate is 2.4 with standard error 0.8. Test zero using a large-sample 5 percent rule.
Solution. The t statistic is 3.0, exceeding about 1.96 in magnitude, so reject zero. A rough 95 percent interval is 2.4 plus or minus 1.57, or about 0.83 to 3.97.

Practice 2. Explain sampling distributions and inference to a classmate using one definition, one concrete example, one assumption, and one failure mode.
Solution. Begin with the core claim above, use the worked case as the concrete example, state the condition highlighted by the warning, and choose a failure where that condition does not hold. A complete explanation connects all four pieces instead of listing them.

Mastery practice A. Reconstruct sampling distributions and inference from a blank page. Include the central definition, the governing rule, one required assumption, and one boundary case.

Solution. Start from the blue idea box, but rewrite it in my own words. Identify the objects and state, name the rule before using it, copy the relevant assumption from the details audit, and choose a boundary where that assumption becomes equality or fails. A complete solution explains why each part belongs; a list of vocabulary is not enough.

Mastery practice B. A classmate gets a polished-looking answer to a sampling distributions and inference problem but never states a model or checks the result. Write the shortest useful review of that solution.

Solution. I would ask four concrete questions: What exactly are the inputs and desired output? Which rule licenses the main step? Which assumption could invalidate it? Which independent check supports the conclusion? If the answer cannot survive those four questions, I treat the numerical or verbal result as unverified.

CHAPTER 5 Functional form and interpretation

The idea

Logs, polynomials, interactions, indicator variables, and nonlinear transformations let marginal effects vary. Coefficient meaning depends on the scale and reference group; extrapolation beyond support is a model claim.

Rebuild the chapter from first principles

The claim I need to own is this: Logs, polynomials, interactions, indicator variables, and nonlinear transformations let marginal effects vary. Coefficient meaning depends on the scale and reference group; extrapolation beyond support is a model claim. I do not count that as learned just because the sentence looks familiar. I should be able to name the objects, state what changes and what stays fixed, recover the rule from the definitions, and explain why the result is allowed.

The move

My working sequence

1. **Translate.** Rewrite the prompt in the language of functional form and interpretation. List the known quantities, the unknown, the units or types, and any hidden constraint.
2. **Choose.** State the theorem, model, algorithm, accounting rule, or physical law before substituting values. This is where I make the assumption visible.
3. **Execute.** Work one justified step at a time and keep intermediate state readable enough that I can audit it later.
4. **Verify.** Check a boundary case, sign, unit, invariant, order of magnitude, or alternative derivation. Then translate the result back into an ordinary sentence.

Details I would refuse to skip

- **Definitions:** I should be able to define every technical noun in the chapter without using that same noun in the definition.
- **Assumptions:** I write the condition that makes the method legal beside the method itself. A correct procedure under the wrong assumptions is still a wrong answer.
- **Representation:** I move freely among words, equations or code, a diagram, and a small numerical example. If one representation disagrees with another, that disagreement is useful evidence.
- **Failure modes:** I keep the nearest counterexample in mind: the boundary where the rule changes, the invariant breaks, or the model stops matching reality.
- **Communication:** My final answer names what the result means and what it does not establish. I do not let notation do the explanatory work for me.

Sanity check

Closed-note check. Can I teach functional form and interpretation on a blank page, derive or reconstruct its main rule, work a fresh example, and diagnose a deliberately broken solution? If I can only recognize the finished work, I am not done.

How this chapter connects

I read this chapter as the bridge from sampling distributions and inference to heteroskedasticity and robust inference. The earlier material supplies the language and constraints; this chapter adds a new reasoning move; the later material asks me to use that move without being told its name. That connection matters because exams and real projects mix chapters on purpose.

The full working model

- **Core claim.** Logs, polynomials, interactions, indicator variables, and nonlinear transformations let marginal effects vary.
- **What moves.** Coefficient meaning depends on the scale and reference group; extrapolation beyond support is a model claim.
- **What keeps the answer honest.** The model becomes useful only when I can apply functional form and interpretation to an unfamiliar case and defend the assumptions.
- **Mastery standard.** I can define the objects, rebuild the rule, work a fresh case, compare alternatives, and diagnose a broken solution without relying on recognition.

How I actually think about it

I translate every coefficient into units before I decide whether it is large.

The easy version

Logs, polynomials, interactions, indicator variables, and nonlinear transformations let marginal effects vary.

Worked example

Question. In log wage equals a plus 0.08 education, interpret the education slope.
Walkthrough. One additional year of education is associated with approximately 8 percent higher wage, or exactly about 8.33 percent using exp(0.08)-1, holding modeled factors fixed.

Common miss

The fastest way to get this chapter wrong is to use the visible procedure without naming its assumption, units, time period, reference group, or boundary. I put that condition beside the work and verify the result independently.

Solved chapter practice

Practice 1. In log wage equals a plus 0.08 education, interpret the education slope.
Solution. One additional year of education is associated with approximately 8 percent higher wage, or exactly about 8.33 percent using exp(0.08)-1, holding modeled factors fixed.

Practice 2. Explain functional form and interpretation to a classmate using one definition, one concrete example, one assumption, and one failure mode.
Solution. Begin with the core claim above, use the worked case as the concrete example, state the condition highlighted by the warning, and choose a failure where that condition does not hold. A complete explanation connects all four pieces instead of listing them.

Mastery practice A. Reconstruct functional form and interpretation from a blank page. Include the central definition, the governing rule, one required assumption, and one boundary case.
Solution. Start from the blue idea box, but rewrite it in my own words. Identify the objects and state, name the rule before using it, copy the relevant assumption from the details audit, and choose a boundary where that assumption becomes equality or fails. A complete solution explains why each part belongs; a list of vocabulary is not enough.

Mastery practice B. A classmate gets a polished-looking answer to a functional form and interpretation problem but never states a model or checks the result. Write the shortest useful review of that solution.
Solution. I would ask four concrete questions: What exactly are the inputs and desired output? Which rule licenses the main step? Which assumption could invalidate it? Which independent check supports the conclusion? If the answer cannot survive those four questions, I treat the numerical or verbal result as unverified.

CHAPTER 6 Heteroskedasticity and robust inference

The idea

Heteroskedasticity means conditional error variance changes with regressors. OLS coefficients may remain unbiased under zero conditional mean, but conventional standard errors fail; heteroskedasticity-robust errors repair inference asymptotically.

Rebuild the chapter from first principles

The claim I need to own is this: Heteroskedasticity means conditional error variance changes with regressors. OLS coefficients may remain unbiased under zero conditional mean, but conventional standard errors fail; heteroskedasticity-robust errors repair inference asymptotically. I do not count that as learned just because the sentence looks familiar. I should be able to name the objects, state what changes and what stays fixed, recover the rule from the definitions, and explain why the result is allowed.

The move

My working sequence

1. **Translate.** Rewrite the prompt in the language of heteroskedasticity and robust inference. List the known quantities, the unknown, the units or types, and any hidden constraint.
2. **Choose.** State the theorem, model, algorithm, accounting rule, or physical law before substituting values. This is where I make the assumption visible.
3. **Execute.** Work one justified step at a time and keep intermediate state readable enough that I can audit it later.
4. **Verify.** Check a boundary case, sign, unit, invariant, order of magnitude, or alternative derivation. Then translate the result back into an ordinary sentence.

Details I would refuse to skip

- **Definitions:** I should be able to define every technical noun in the chapter without using that same noun in the definition.
- **Assumptions:** I write the condition that makes the method legal beside the method itself. A correct procedure under the wrong assumptions is still a wrong answer.
- **Representation:** I move freely among words, equations or code, a diagram, and a small numerical example. If one representation disagrees with another, that disagreement is useful evidence.
- **Failure modes:** I keep the nearest counterexample in mind: the boundary where the rule changes, the invariant breaks, or the model stops matching reality.
- **Communication:** My final answer names what the result means and what it does not establish. I do not let notation do the explanatory work for me.

Sanity check

Closed-note check. Can I teach heteroskedasticity and robust inference on a blank page, derive or reconstruct its main rule, work a fresh example, and diagnose a deliberately broken solution? If I can only recognize the finished work, I am not done.

How this chapter connects

I read this chapter as the bridge from functional form and interpretation to time order, serial correlation, and trends. The earlier material supplies the language and constraints; this chapter adds a new reasoning move; the later material asks me to use that move without being told its name. That connection matters because exams and real projects mix chapters on purpose.

The full working model

- **Core claim.** Heteroskedasticity means conditional error variance changes with regressors.
- **What moves.** OLS coefficients may remain unbiased under zero conditional mean, but conventional standard errors fail; heteroskedasticity-robust errors repair inference asymptotically.
- **What keeps the answer honest.** The model becomes useful only when I can apply heteroskedasticity and robust inference to an unfamiliar case and defend the assumptions.
- **Mastery standard.** I can define the objects, rebuild the rule, work a fresh case, compare alternatives, and diagnose a broken solution without relying on recognition.

How I actually think about it

I use robust standard errors by default for cross-sectional work unless a stronger variance model is justified.

The easy version

Heteroskedasticity means conditional error variance changes with regressors.

Worked example

Question. Does heteroskedasticity automatically bias the OLS slope?
Walkthrough. No. It primarily invalidates the usual variance formula. Bias comes from failures such as nonzero conditional mean, not unequal variance by itself.

Common miss

The fastest way to get this chapter wrong is to use the visible procedure without naming its assumption, units, time period, reference group, or boundary. I put that condition beside the work and verify the result independently.

Solved chapter practice

Practice 1. Does heteroskedasticity automatically bias the OLS slope?
Solution. No. It primarily invalidates the usual variance formula. Bias comes from failures such as nonzero conditional mean, not unequal variance by itself.

Practice 2. Explain heteroskedasticity and robust inference to a classmate using one definition, one concrete example, one assumption, and one failure mode.
Solution. Begin with the core claim above, use the worked case as the concrete example, state the condition highlighted by the warning, and choose a failure where that condition does not hold. A complete explanation connects all four pieces instead of listing them.

Mastery practice A. Reconstruct heteroskedasticity and robust inference from a blank page. Include the central definition, the governing rule, one required assumption, and one boundary case.
Solution. Start from the blue idea box, but rewrite it in my own words. Identify the objects and state, name the rule before using it, copy the relevant assumption from the details audit, and

choose a boundary where that assumption becomes equality or fails. A complete solution explains why each part belongs; a list of vocabulary is not enough.

Mastery practice B. A classmate gets a polished-looking answer to a heteroskedasticity and robust inference problem but never states a model or checks the result. Write the shortest useful review of that solution.
Solution. I would ask four concrete questions: What exactly are the inputs and desired output? Which rule licenses the main step? Which assumption could invalidate it? Which independent check supports the conclusion? If the answer cannot survive those four questions, I treat the numerical or verbal result as unverified.

CHAPTER 7 Time order, serial correlation, and trends

The idea

Time-series observations bring dependence, trends, seasonality, persistence, and structural change. Serial correlation distorts standard errors and may signal omitted dynamics; spurious regression can arise from unrelated trending series.

Rebuild the chapter from first principles

The claim I need to own is this: Time-series observations bring dependence, trends, seasonality, persistence, and structural change. Serial correlation distorts standard errors and may signal omitted dynamics; spurious regression can arise from unrelated trending series. I do not count that as learned just because the sentence looks familiar. I should be able to name the objects, state what changes and what stays fixed, recover the rule from the definitions, and explain why the result is allowed.

The move

My working sequence

1. **Translate.** Rewrite the prompt in the language of time order, serial correlation, and trends. List the known quantities, the unknown, the units or types, and any hidden constraint.
2. **Choose.** State the theorem, model, algorithm, accounting rule, or physical law before substituting values. This is where I make the assumption visible.
3. **Execute.** Work one justified step at a time and keep intermediate state readable enough that I can audit it later.
4. **Verify.** Check a boundary case, sign, unit, invariant, order of magnitude, or alternative derivation. Then translate the result back into an ordinary sentence.

Details I would refuse to skip

- **Definitions:** I should be able to define every technical noun in the chapter without using that same noun in the definition.

- **Assumptions:** I write the condition that makes the method legal beside the method itself. A correct procedure under the wrong assumptions is still a wrong answer.
- **Representation:** I move freely among words, equations or code, a diagram, and a small numerical example. If one representation disagrees with another, that disagreement is useful evidence.
- **Failure modes:** I keep the nearest counterexample in mind: the boundary where the rule changes, the invariant breaks, or the model stops matching reality.
- **Communication:** My final answer names what the result means and what it does not establish. I do not let notation do the explanatory work for me.

Sanity check

Closed-note check. Can I teach time order, serial correlation, and trends on a blank page, derive or reconstruct its main rule, work a fresh example, and diagnose a deliberately broken solution? If I can only recognize the finished work, I am not done.

How this chapter connects

I read this chapter as the bridge from heteroskedasticity and robust inference to reproducible r workflow and honest reporting. The earlier material supplies the language and constraints; this chapter adds a new reasoning move; the later material asks me to use that move without being told its name. That connection matters because exams and real projects mix chapters on purpose.

The full working model

- **Core claim.** Time-series observations bring dependence, trends, seasonality, persistence, and structural change.
- **What moves.** Serial correlation distorts standard errors and may signal omitted dynamics; spurious regression can arise from unrelated trending series.
- **What keeps the answer honest.** The model becomes useful only when I can apply time order, serial correlation, and trends to an unfamiliar case and defend the assumptions.
- **Mastery standard.** I can define the objects, rebuild the rule, work a fresh case, compare alternatives, and diagnose a broken solution without relying on recognition.

How I actually think about it

My shorthand for this chapter is: my shorthand for this chapter is: two lines rising over time are not evidence that one causes the other.

The easy version

Time-series observations bring dependence, trends, seasonality, persistence, and structural change.

Worked example

Question. Why include a time trend in a regression of two trending variables?
Walkthrough. It can absorb a shared deterministic trend, but does not solve every nonstationarity problem. Residual diagnostics, differencing, and a defensible dynamic model remain necessary.

Common miss

The fastest way to get this chapter wrong is to use the visible procedure without naming its assumption, units, time period, reference group, or boundary. I put that condition beside the work and verify the result independently.

Solved chapter practice

Practice 1. Why include a time trend in a regression of two trending variables?
Solution. It can absorb a shared deterministic trend, but does not solve every nonstationarity problem. Residual diagnostics, differencing, and a defensible dynamic model remain necessary.

Practice 2. Explain time order, serial correlation, and trends to a classmate using one definition, one concrete example, one assumption, and one failure mode.
Solution. Begin with the core claim above, use the worked case as the concrete example, state the condition highlighted by the warning, and choose a failure where that condition does not hold. A complete explanation connects all four pieces instead of listing them.

Mastery practice A. Reconstruct time order, serial correlation, and trends from a blank page. Include the central definition, the governing rule, one required assumption, and one boundary case.
Solution. Start from the blue idea box, but rewrite it in my own words. Identify the objects and state, name the rule before using it, copy the relevant assumption from the details audit, and choose a boundary where that assumption becomes equality or fails. A complete solution explains why each part belongs; a list of vocabulary is not enough.

Mastery practice B. A classmate gets a polished-looking answer to a time order, serial correlation, and trends problem but never states a model or checks the result. Write the shortest useful review of that solution.
Solution. I would ask four concrete questions: What exactly are the inputs and desired output? Which rule licenses the main step? Which assumption could invalidate it? Which independent check supports the conclusion? If the answer cannot survive those four questions, I treat the numerical or verbal result as unverified.

CHAPTER 8 Reproducible R workflow and honest reporting

The idea

A sound workflow imports without overwriting raw data, validates types and joins, explores distributions, estimates declared models, checks residuals and sensitivity, produces reproducible tables and figures, and records limitations.

Rebuild the chapter from first principles

The claim I need to own is this: A sound workflow imports without overwriting raw data, validates types and joins, explores distributions, estimates declared models, checks residuals and sensitivity, produces reproducible tables and figures, and records limitations. I do not count that as learned just because the sentence looks familiar. I should be able to name the objects, state what changes and what stays fixed, recover the rule from the definitions, and explain why the result is allowed.

The move

My working sequence

1. **Translate.** Rewrite the prompt in the language of reproducible r workflow and honest reporting. List the known quantities, the unknown, the units or types, and any hidden constraint.
2. **Choose.** State the theorem, model, algorithm, accounting rule, or physical law before substituting values. This is where I make the assumption visible.
3. **Execute.** Work one justified step at a time and keep intermediate state readable enough that I can audit it later.
4. **Verify.** Check a boundary case, sign, unit, invariant, order of magnitude, or alternative derivation. Then translate the result back into an ordinary sentence.

Details I would refuse to skip

- **Definitions:** I should be able to define every technical noun in the chapter without using that same noun in the definition.
- **Assumptions:** I write the condition that makes the method legal beside the method itself. A correct procedure under the wrong assumptions is still a wrong answer.
- **Representation:** I move freely among words, equations or code, a diagram, and a small numerical example. If one representation disagrees with another, that disagreement is useful evidence.
- **Failure modes:** I keep the nearest counterexample in mind: the boundary where the rule changes, the invariant breaks, or the model stops matching reality.
- **Communication:** My final answer names what the result means and what it does not establish. I do not let notation do the explanatory work for me.

Sanity check

Closed-note check. Can I teach reproducible r workflow and honest reporting on a blank page, derive or reconstruct its main rule, work a fresh example, and diagnose a deliberately broken solution? If I can only recognize the finished work, I am not done.

How this chapter connects

I read this chapter as the bridge from time order, serial correlation, and trends to the cumulative course model. The earlier material supplies the language and constraints; this chapter adds a new reasoning move; the later material asks me to use that move without being told its name. That connection matters because exams and real projects mix chapters on purpose.

The full working model

- **Core claim.** A sound workflow imports without overwriting raw data, validates types and joins, explores distributions, estimates declared models, checks residuals and sensitivity, produces reproducible tables and figures, and records limitations.
- **What moves.** The model becomes useful only when I can apply reproducible r workflow and honest reporting to an unfamiliar case and defend the assumptions.
- **What keeps the answer honest.** The model becomes useful only when I can apply reproducible r workflow and honest reporting to an unfamiliar case and defend the assumptions.
- **Mastery standard.** I can define the objects, rebuild the rule, work a fresh case, compare alternatives, and diagnose a broken solution without relying on recognition.

How I actually think about it

My shorthand for this chapter is: my analysis should run from a clean session and leave an audit trail another person can follow.

The easy version

A sound workflow imports without overwriting raw data, validates types and joins, explores distributions, estimates declared models, checks residuals and sensitivity, produces reproducible tables and figures, and records limitations.

Worked example

Question. What belongs in a minimal regression report?
Walkthrough. Question and estimand, data and sample construction, exact specification, coefficient units, appropriate standard errors, uncertainty, diagnostics, sensitivity checks, limitations, and reproducible code.

Common miss

The fastest way to get this chapter wrong is to use the visible procedure without naming its assumption, units, time period, reference group, or boundary. I put that condition beside the work and verify the result independently.

Solved chapter practice

Practice 1. What belongs in a minimal regression report?
Solution. Question and estimand, data and sample construction, exact specification, coefficient units, appropriate standard errors, uncertainty, diagnostics, sensitivity checks, limitations, and reproducible code.

Practice 2. Explain reproducible r workflow and honest reporting to a classmate using one definition, one concrete example, one assumption, and one failure mode.
Solution. Begin with the core claim above, use the worked case as the concrete example, state the condition highlighted by the warning, and choose a failure where that condition does not hold. A complete explanation connects all four pieces instead of listing them.

Mastery practice A. Reconstruct reproducible r workflow and honest reporting from a blank page. Include the central definition, the governing rule, one required assumption, and one boundary case.
Solution. Start from the blue idea box, but rewrite it in my own words. Identify the objects and state, name the rule before using it, copy the relevant assumption from the details audit, and choose a boundary where that assumption becomes equality or fails. A complete solution explains why each part belongs; a list of vocabulary is not enough.

Mastery practice B. A classmate gets a polished-looking answer to a reproducible r workflow and honest reporting problem but never states a model or checks the result. Write the shortest useful review of that solution.
Solution. I would ask four concrete questions: What exactly are the inputs and desired output? Which rule licenses the main step? Which assumption could invalidate it? Which independent check supports the conclusion? If the answer cannot survive those four questions, I treat the numerical or verbal result as unverified.

Cumulative study plan

I use this packet in three passes. First I reconstruct each model and work its examples without looking. Second I mix chapters so the tool is no longer named for me. Third I explain a complete case aloud, including assumptions, units, uncertainty, failure modes, and what evidence would change my conclusion.