

UCLA-ALIGNED COURSE FIELD NOTES

STATS 100C

Linear Models

Matrix regression, projection, general hypotheses, diagnostics, selection, and principal-component regression

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	The linear model in matrix form	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
2	Least squares as orthogonal projection	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
3	Distribution theory and Gauss-Markov	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
4	General linear hypotheses and extra sums of squares	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
5	Residuals, leverage, and influence	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
6	Categorical predictors, interactions, and polynomial structure	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
7	Model selection and validation	Reconstruct the model, work an application, test its assumptions, and explain the result in ordinary language.
8	Principal-component regression and stabilized fitting	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/2024/stats100c>. 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

1	The linear model in matrix form	5
1.1	Rebuild the chapter from first principles	5
1.2	Details I would refuse to skip	5
1.3	How this chapter connects	6
1.4	The full working model	6
2	Least squares as orthogonal projection	7
2.1	Rebuild the chapter from first principles	7
2.2	Details I would refuse to skip	8
2.3	How this chapter connects	8
2.4	The full working model	8
3	Distribution theory and Gauss-Markov	10
3.1	Rebuild the chapter from first principles	10
3.2	Details I would refuse to skip	10
3.3	How this chapter connects	11
3.4	The full working model	11
4	General linear hypotheses and extra sums of squares	12
4.1	Rebuild the chapter from first principles	12
4.2	Details I would refuse to skip	13
4.3	How this chapter connects	13
4.4	The full working model	13
5	Residuals, leverage, and influence	15
5.1	Rebuild the chapter from first principles	15
5.2	Details I would refuse to skip	15
5.3	How this chapter connects	16
5.4	The full working model	16
6	Categorical predictors, interactions, and polynomial structure	17
6.1	Rebuild the chapter from first principles	17
6.2	Details I would refuse to skip	18
6.3	How this chapter connects	18

6.4	The full working model	18
7	Model selection and validation	20
7.1	Rebuild the chapter from first principles	20
7.2	Details I would refuse to skip	20
7.3	How this chapter connects	21
7.4	The full working model	21
8	Principal-component regression and stabilized fitting	22
8.1	Rebuild the chapter from first principles	22
8.2	Details I would refuse to skip	23
8.3	How this chapter connects	23
8.4	The full working model	24

CHAPTER 1 The linear model in matrix form

The idea

The model $y = X\beta + \text{error}$ describes a mean vector in the column space of X . Rank, identifiability, estimability, coding, and the geometry of projection determine what coefficients can be learned.

Rebuild the chapter from first principles

The claim I need to own is this: The model $y = X\beta + \text{error}$ describes a mean vector in the column space of X . Rank, identifiability, estimability, coding, and the geometry of projection determine what coefficients can be learned. 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 the linear model in matrix form. 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 the linear model in matrix form 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 least squares as orthogonal projection. 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 model $y = X\beta + \epsilon$ describes a mean vector in the column space of X .
- **What moves.** Rank, identifiability, estimability, coding, and the geometry of projection determine what coefficients can be learned.
- **What keeps the answer honest.** The model becomes useful only when I can apply the linear model in matrix form 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 check the design matrix before interpreting coefficients; the model is hiding there in plain sight.

The easy version

The model $y = X\beta + \epsilon$ describes a mean vector in the column space of X .

Worked example

Question. What fails if one column of X is an exact linear combination of others?

Walkthrough. X lacks full column rank, β is not uniquely identified, and $X^T X$ is singular. Reparameterize or remove redundant columns.

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 fails if one column of X is an exact linear combination of others?

Solution. X lacks full column rank, β is not uniquely identified, and X -transpose- X is singular. Reparameterize or remove redundant columns.

Practice 2. Explain the linear model in matrix form 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 the linear model in matrix form 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 the linear model in matrix form 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 Least squares as orthogonal projection

The idea

OLS projects y onto the column space of X . Normal equations make residuals orthogonal to every regressor; $\hat{y}$ and residual-maker matrices describe fitted values, leverage, and decomposition.

Rebuild the chapter from first principles

The claim I need to own is this: OLS projects y onto the column space of X . Normal equations make residuals orthogonal to every regressor; $\hat{y}$ and residual-maker matrices describe fitted values, leverage, and decomposition. 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 least squares as orthogonal projection. 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 least squares as orthogonal projection 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 linear model in matrix form to distribution theory and gauss-markov. 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.** OLS projects y onto the column space of X .
- **What moves.** Normal equations make residuals orthogonal to every regressor; hat and residual-maker matrices describe fitted values, leverage, and decomposition.
- **What keeps the answer honest.** The model becomes useful only when I can apply least squares as orthogonal projection 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 geometry makes the algebra believable: fitted plus residual are perpendicular pieces of the same data vector.

The easy version

OLS projects y onto the column space of X .

Worked example

Question. State the normal equations.

Walkthrough. X -transpose times $(y - X \text{ beta-hat}) = 0$. Equivalently, X -transpose residual equals zero.

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. State the normal equations.

Solution. X -transpose times $(y - X \text{ beta-hat}) = 0$. Equivalently, X -transpose residual equals zero.

Practice 2. Explain least squares as orthogonal projection 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 least squares as orthogonal projection 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 least squares as orthogonal projection 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 Distribution theory and Gauss-Markov

The idea

Under zero-mean spherical errors OLS is best linear unbiased; normal errors add exact t and F distributions. Estimated variance, degrees of freedom, covariance of coefficients, and prediction uncertainty follow from projection geometry.

Rebuild the chapter from first principles

The claim I need to own is this: Under zero-mean spherical errors OLS is best linear unbiased; normal errors add exact t and F distributions. Estimated variance, degrees of freedom, covariance of coefficients, and prediction uncertainty follow from projection geometry. 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 distribution theory and gauss-markov. 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 distribution theory and gauss-markov 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 least squares as orthogonal projection to general linear hypotheses and extra sums of 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.** Under zero-mean spherical errors OLS is best linear unbiased; normal errors add exact t and F distributions.
- **What moves.** Estimated variance, degrees of freedom, covariance of coefficients, and prediction uncertainty follow from projection geometry.
- **What keeps the answer honest.** The model becomes useful only when I can apply distribution theory and gauss-markov 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: bLUE is a precise conditional claim, not a slogan that OLS is always best.

The easy version

Under zero-mean spherical errors OLS is best linear unbiased; normal errors add exact t and F distributions.

Worked example

Question. What is $\text{Var}(\hat{\beta})$ under homoskedastic independent errors?

Walkthrough. $\sigma^2 \text{ times } (X\text{-transpose}\cdot X)^{\text{inverse}}$, assuming full column rank.

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 is $\text{Var}(\hat{\beta})$ under homoskedastic independent errors?

Solution. $\sigma^2 (X^T X)^{-1}$, assuming full column rank.

Practice 2. Explain distribution theory and gauss-markov 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 distribution theory and gauss-markov 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 distribution theory and gauss-markov 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 General linear hypotheses and extra sums of squares

The idea

A general hypothesis $R\beta=r$ can test one or many coefficient restrictions. Full and reduced models produce extra sums of squares; t , F , confidence regions, and partial tests are different views of the same comparison.

Rebuild the chapter from first principles

The claim I need to own is this: A general hypothesis $R\beta=r$ can test one or many coefficient restrictions. Full and reduced models produce extra sums of squares; t , F , confidence regions, and partial tests are different views of the same comparison. 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 general linear hypotheses and extra sums of 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 general linear hypotheses and extra sums of 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 distribution theory and gauss-markov to residuals, leverage, and influence. 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 general hypothesis $R\beta=r$ can test one or many coefficient restrictions.
- **What moves.** Full and reduced models produce extra sums of squares; t , F , confidence regions, and partial tests are different views of the same comparison.

- **What keeps the answer honest.** The model becomes useful only when I can apply general linear hypotheses and extra sums of 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 write the restriction matrix before fitting models so the test matches the scientific question.

The easy version

A general hypothesis $R\beta=r$ can test one or many coefficient restrictions.

Worked example

Question. How does a one-restriction F test relate to a t test?

Walkthrough. For the same two-sided linear restriction, F equals t squared and yields the same p-value.

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. How does a one-restriction F test relate to a t test?

Solution. For the same two-sided linear restriction, F equals t squared and yields the same p-value.

Practice 2. Explain general linear hypotheses and extra sums of 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 general linear hypotheses and extra sums of 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 general linear hypotheses and extra sums of 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 5 Residuals, leverage, and influence

The idea

Residual diagnostics examine nonlinearity, unequal variance, dependence, outliers, leverage, and influence. Studentized residuals, hat values, Cook's distance, and case-deletion sensitivity reveal different failure modes.

Rebuild the chapter from first principles

The claim I need to own is this: Residual diagnostics examine nonlinearity, unequal variance, dependence, outliers, leverage, and influence. Studentized residuals, hat values, Cook's distance, and case-deletion sensitivity reveal different failure modes. 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 residuals, leverage, and influence. 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 residuals, leverage, and influence 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 general linear hypotheses and extra sums of squares to categorical predictors, interactions, and polynomial structure. 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.** Residual diagnostics examine nonlinearity, unequal variance, dependence, outliers, leverage, and influence.
- **What moves.** Studentized residuals, hat values, Cook's distance, and case-deletion sensitivity reveal different failure modes.
- **What keeps the answer honest.** The model becomes useful only when I can apply residuals, leverage, and influence 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

An unusual x value is leverage; an unusual y after fitting is residual. I need both to understand influence.

The easy version

Residual diagnostics examine nonlinearity, unequal variance, dependence, outliers, leverage, and influence.

Worked example

Question. Can a high-leverage point have low influence?

Walkthrough. Yes. If it lies close to the fitted pattern, deleting it may barely change coefficients despite unusual predictors.

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. Can a high-leverage point have low influence?

Solution. Yes. If it lies close to the fitted pattern, deleting it may barely change coefficients despite unusual predictors.

Practice 2. Explain residuals, leverage, and influence 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 residuals, leverage, and influence 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 residuals, leverage, and influence 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 Categorical predictors, interactions, and polynomial structure

The idea

Indicator coding chooses reference comparisons; interactions make one effect depend on another variable; polynomial and basis expansions model curvature while remaining linear in coefficients.

Rebuild the chapter from first principles

The claim I need to own is this: Indicator coding chooses reference comparisons; interactions make one effect depend on another variable; polynomial and basis expansions model curvature while remaining linear in coefficients. 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 categorical predictors, interactions, and polynomial structure. 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 categorical predictors, interactions, and polynomial structure 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 residuals, leverage, and influence to model selection and validation. 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.** Indicator coding chooses reference comparisons; interactions make one effect depend on another variable; polynomial and basis expansions model curvature while remaining linear in coefficients.

- **What moves.** The model becomes useful only when I can apply categorical predictors, interactions, and polynomial structure to an unfamiliar case and defend the assumptions.
- **What keeps the answer honest.** The model becomes useful only when I can apply categorical predictors, interactions, and polynomial structure 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: a main effect inside an interaction is conditional at the other variable's zero or reference level.

The easy version

Indicator coding chooses reference comparisons; interactions make one effect depend on another variable; polynomial and basis expansions model curvature while remaining linear in coefficients.

Worked example

Question. In $y=b_0+b_1x+b_2g+b_3xg$, what is the x slope when $g=1$?

Walkthrough. The slope is b_1+b_3 . When $g=0$ it is b_1 , and b_3 is the slope difference.

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 $y=b_0+b_1x+b_2g+b_3xg$, what is the x slope when $g=1$?

Solution. The slope is b_1+b_3 . When $g=0$ it is b_1 , and b_3 is the slope difference.

Practice 2. Explain categorical predictors, interactions, and polynomial structure 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 categorical predictors, interactions, and polynomial structure 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 categorical predictors, interactions, and polynomial structure 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 Model selection and validation

The idea

Adjusted R-squared, information criteria, nested tests, stepwise methods, cross-validation, and subject knowledge choose among models under different goals. Data-driven selection changes uncertainty and can overfit.

Rebuild the chapter from first principles

The claim I need to own is this: Adjusted R-squared, information criteria, nested tests, stepwise methods, cross-validation, and subject knowledge choose among models under different goals. Data-driven selection changes uncertainty and can overfit. 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 model selection and validation. 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 model selection and validation 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 categorical predictors, interactions, and polynomial structure to principal-component regression and stabilized fitting. 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.** Adjusted R-squared, information criteria, nested tests, stepwise methods, cross-validation, and subject knowledge choose among models under different goals.
- **What moves.** Data-driven selection changes uncertainty and can overfit.
- **What keeps the answer honest.** The model becomes useful only when I can apply model selection and validation 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 decide whether the goal is explanation, prediction, or compression before selecting a model.

The easy version

Adjusted R-squared, information criteria, nested tests, stepwise methods, cross-validation, and subject knowledge choose among models under different goals.

Worked example

Question. Why can stepwise p-values be misleading?

Walkthrough. They ignore the adaptive search that selected variables, understating uncertainty and exaggerating evidence. Validate externally or account for selection.

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 can stepwise p-values be misleading?

Solution. They ignore the adaptive search that selected variables, understating uncertainty and exaggerating evidence. Validate externally or account for selection.

Practice 2. Explain model selection and validation 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 model selection and validation 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 model selection and validation 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 Principal-component regression and stabilized fitting

The idea

Correlated predictors create unstable coefficient directions. Principal-component regression rotates to orthogonal directions and discards low-variance components; ridge shrinks directions instead. Scaling and validation are essential.

Rebuild the chapter from first principles

The claim I need to own is this: Correlated predictors create unstable coefficient directions. Principal-component regression rotates to orthogonal directions and discards low-variance components; ridge shrinks directions instead. Scaling and validation are essential. 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 principal-component regression and stabilized fitting. 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 principal-component regression and stabilized fitting 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 model selection and validation 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.** Correlated predictors create unstable coefficient directions.
- **What moves.** Principal-component regression rotates to orthogonal directions and discards low-variance components; ridge shrinks directions instead.
- **What keeps the answer honest.** Scaling and validation are essential.
- **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: a low-variance predictor direction can still matter for y , so unsupervised compression is not automatically predictive.

The easy version

Correlated predictors create unstable coefficient directions.

Worked example

Question. Why standardize before PCA when variables use different units?

Walkthrough. Without scaling, high-variance units dominate the components. Standardization makes the analysis reflect correlation structure unless raw variance is substantively meaningful.

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 standardize before PCA when variables use different units?

Solution. Without scaling, high-variance units dominate the components. Standardization makes the analysis reflect correlation structure unless raw variance is substantively meaningful.

Practice 2. Explain principal-component regression and stabilized fitting 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 principal-component regression and stabilized fitting 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 principal-component regression and stabilized fitting 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.