Introduction to the Course

Sergio I. Garcia-Rios

Government 3990: Statistics in the Social Science

General info

Teaching team

- Professor: Sergio Garcia-Rios
  `garcia.rios@cornell.edu`

Required materials

- OpenIntro Statistics, 3rd Edition:
  `http://openintro.org/os`
  - The textbook is freely available online. You're encuoraged to read on screen but you can print it out. If you prefer a paperback version you can buy it at the cost of printing (around $10) on Amazon
- (optional) Calculator (just something that can do square roots)

Puzzle Solving with Data - GOVT 3990

garciarios.github.io/govt_3990

Course structure

Learning units and course outline

- Pictures and summaries of data
    - Unit 1 - Intro to data: Observational studies & non-causal inference, principles of experimental design & causal inference, exploratory data analysis, introduction to simulation-based statistical inference.
- Mathematics behind statistics
    - Unit 2 - Probability & distributions: Basics of probability and chance processes, Bayesian perspective in statistical inference, the normal and binomial distributions.
- Statistical inference
    - Unit 3 - Framework for inference: CLT, sampling distributions, and introduction to theoretical inference.
    - Midterm 1
    - Unit 4 - Statistical inference for numerical variables
    - Unit 5 - Statistical inference for categorical variables
    - Midterm 2
- Modeling
    - Unit 6 - Simple linear regression: Bivariate correlation and causality, introduction to modeling.
    - Unit 7 - Multiple linear regression: More advanced modeling with multiple predictors.
    - Final Exam

Course structure

- Set of learning objectives and required and suggested readings, videos, etc. for each unit.
- Prior to beginning the unit, watch the videos and/or complete the readings and familiarize yourselves with the learning objectives.
- Begin a new unit with a readiness assessment: individual, then team.
- Class time: split between lecture, discussion/application, and lab.
- Complement your learning with problem sets.
- Wrap up a unit with a performance assessment.

Course structure

- Set of learning objectives and required and suggested readings, videos, etc. for each unit.
- Prior to beginning the unit, watch the videos and/or complete the readings and familiarize yourselves with the learning objectives.
- Begin a new unit with a readiness assessment: individual, then team.
- Class time: split between lecture, discussion/application, and lab.
- Complement your learning with problem sets.
- Wrap up a unit with a performance assessment.

Course structure

- Set of learning objectives and required and suggested readings, videos, etc. for each unit.
- Prior to beginning the unit, watch the videos and/or complete the readings and familiarize yourselves with the learning objectives.
- Begin a new unit with a readiness assessment: individual, then team.
- Class time: split between lecture, discussion/application, and lab.
- Complement your learning with problem sets.
- Wrap up a unit with a performance assessment.

Course structure

- Set of learning objectives and required and suggested readings, videos, etc. for each unit.
- Prior to beginning the unit, watch the videos and/or complete the readings and familiarize yourselves with the learning objectives.
- Begin a new unit with a readiness assessment: individual, then team.
- Class time: split between lecture, discussion/application, and lab.
- Complement your learning with problem sets.
- Wrap up a unit with a performance assessment.

Course structure

- Set of learning objectives and required and suggested readings, videos, etc. for each unit.
- Prior to beginning the unit, watch the videos and/or complete the readings and familiarize yourselves with the learning objectives.
- Begin a new unit with a readiness assessment: individual, then team.
- Class time: split between lecture, discussion/application, and lab.
- Complement your learning with problem sets.
- Wrap up a unit with a performance assessment.

Course structure

- Set of learning objectives and required and suggested readings, videos, etc. for each unit.
- Prior to beginning the unit, watch the videos and/or complete the readings and familiarize yourselves with the learning objectives.
- Begin a new unit with a readiness assessment: individual, then team.
- Class time: split between lecture, discussion/application, and lab.
- Complement your learning with problem sets.
- Wrap up a unit with a performance assessment.

Project

Objective: Give you independent applied research experience using real data and statistical methods.

- Proposal and presentation: due mid-semester
- Final Presentation: last week of semester

Exams

There will be two midterms.

- See course info for dates and times of the exams.
- Exam dates cannot be changed and no make-up exams will be given.
- Calculator + cheat sheet allowed

Office Hours

Visit `garciarios.youcanbook.me` if you need to schedule office hours

- At least 24 hrs in advance
- Please choose only one slot, if you need more time, email me first
- If you schedule office hours, please do show up

Students with disabilities

Students with disabilities who believe they may need accommodations in this class are encouraged to contact the Student with Disability Services Office telephone 607.254.4545; **`e-mailsds_cu@cornell.edu`** as soon as possible to better ensure that such accommodations can be made.

Academic Dishonesty

Any form of academic dishonesty will result in an immediate 0 on the given assignment and will be reported to the Office of Academic Integrity. Additional penalties may also be assessed if deemed appropriate. If you have any questions about whether something is or is not allowed, ask me beforehand.

Tips for success

Tips for success

- Complete the reading before a new unit begins, and then review again after the unit is over.
- Be an active participant during lectures and labs.
- Ask questions - during class or office hours, or by email. Ask me, and your classmates.
- Do the problem sets - start early and make sure you attempt and understand all questions.
- Start your project early and and allow adequate time to complete them.
- Give yourself plenty of time time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

Tips for success

- Complete the reading before a new unit begins, and then review again after the unit is over.
- Be an active participant during lectures and labs.
- Ask questions - during class or office hours, or by email. Ask me, and your classmates.
- Do the problem sets - start early and make sure you attempt and understand all questions.
- Start your project early and and allow adequate time to complete them.
- Give yourself plenty of time time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

Tips for success

- Complete the reading before a new unit begins, and then review again after the unit is over.
- Be an active participant during lectures and labs.
- Ask questions - during class or office hours, or by email. Ask me, and your classmates.
- Do the problem sets - start early and make sure you attempt and understand all questions.
- Start your project early and and allow adequate time to complete them.
- Give yourself plenty of time time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

Tips for success

- Complete the reading before a new unit begins, and then review again after the unit is over.
- Be an active participant during lectures and labs.
- Ask questions - during class or office hours, or by email. Ask me, and your classmates.
- Do the problem sets - start early and make sure you attempt and understand all questions.
- Start your project early and and allow adequate time to complete them.
- Give yourself plenty of time time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

Tips for success

- Complete the reading before a new unit begins, and then review again after the unit is over.
- Be an active participant during lectures and labs.
- Ask questions - during class or office hours, or by email. Ask me, and your classmates.
- Do the problem sets - start early and make sure you attempt and understand all questions.
- Start your project early and and allow adequate time to complete them.
- Give yourself plenty of time time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

Tips for success

- Complete the reading before a new unit begins, and then review again after the unit is over.
- Be an active participant during lectures and labs.
- Ask questions - during class or office hours, or by email. Ask me, and your classmates.
- Do the problem sets - start early and make sure you attempt and understand all questions.
- Start your project early and and allow adequate time to complete them.
- Give yourself plenty of time time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

Tips for success

- Complete the reading before a new unit begins, and then review again after the unit is over.
- Be an active participant during lectures and labs.
- Ask questions - during class or office hours, or by email. Ask me, and your classmates.
- Do the problem sets - start early and make sure you attempt and understand all questions.
- Start your project early and and allow adequate time to complete them.
- Give yourself plenty of time time to prepare a good cheat sheet for exams. This requires going through the material and taking the time to review the concepts that you're not comfortable with.
- Do not procrastinate - don't let a unit go by with unanswered questions as it will just make the following unit's material even more difficult to follow.

To do

To do

- Download or purchase the textbook
  - Download: **`http://openintro.org/os`**
  - Purchase: **`http://openintro.org/os/amazon`**
- Go to the course website, read the syllabus, let me know if you have any questions.
- Get started with your labs
  - Lab 0 is due Monday, Jan 30

An Example on the importance of (Good) Data Analysis and Presentation

The *Challenger* launch decision

- In 1986, the Challenger space shuttle exploded moments after liftoff
- Decision to launch one other most scrutinized in history
- Failure of O-rings in the solid-fuel rocket boosters blamed for explosion
- Could this failure have been foreseen?

Morton-Thiokol engineers made this table & worried about launching below 53 degrees (Why?)

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

Morton-Thiokol engineers made this table & worried about launching below 53 degrees (Why?)

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

O-ring would erode or have “blow-by” (2 ways to fail) in cold temp

Morton-Thiokol engineers made this table & worried about launching below 53 degrees (Why?)

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

Failed to convince administrators there was a danger

Morton-Thiokol engineers made this table & worried about launching below 53 degrees (Why?)

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

(Counter-argument: "damages at low and high temps")

The *Challenger* launch decision

Morton-Thiokol engineers made this table & worried about launching below 53 degrees (Why?)

Flights with O-ring damage	
Flt Number	Temp (F)
2	70
41b	57
41c	63
41d	70
51c	53
61a	79
61c	58

Are there problems with this presentation? with the use of data?

Engineers did not consider successes, only failures; selection on the dependent variable

Engineers did not consider successes, only failures; selection on the dependent variable

All flights, chronological order			
Damage?	Temp (F)	Damage?	Temp (F)
No	66	No	78
Yes	70	No	67
No	69	Yes	53
No	68	No	67
No	67	No	75
No	72	No	70
No	73	No	81
No	70	No	76
Yes	57	Yes	79
Yes	63	No	76
Yes	70	Yes	58

Other problems?

The *Challenger* launch decision

Engineers did not consider successes, only failures; selection on the dependent variable

All flights, chronological order			
Damage?	Temp (F)	Damage?	Temp (F)
No	66	No	78
Yes	70	No	67
No	69	Yes	53
No	68	No	67
No	67	No	75
No	72	No	70
No	73	No	81
No	70	No	76
Yes	57	Yes	79
Yes	63	No	76
Yes	70	Yes	58

Other problems? Why sort by launch number?

The *Challenger* launch decision

O-ring damage pre-Challenger, by temperature at launch			
Damage?	Temp (F)	Damage?	Temp (F)
Yes	53	**Yes**	70
Yes	57	No	70
Yes	58	No	70
Yes	63	No	72
No	66	No	73
No	67	No	75
No	67	No	76
No	67	No	76
No	68	No	78
No	69	**Yes**	79
Yes	70	No	81

The *Challenger* launch decision

O-ring damage pre-Challenger, by temperature at launch

Damage?	Temp (F)	Damage?	Temp (F)
Yes	53	Yes	70
Yes	57	No	70
Yes	58	No	70
Yes	63	No	72
No	66	No	73
No	67	No	75
No	67	No	76
No	67	No	76
No	68	No	78
No	69	Yes	79
Yes	70	No	81

The evidence begins to speak for itself.

The *Challenger* launch decision

O-ring damage pre-Challenger, by temperature at launch

Damage?	Temp (F)	Damage?	Temp (F)
Yes	53	Yes	70
Yes	57	No	70
Yes	58	No	70
Yes	63	No	72
No	66	No	73
No	67	No	75
No	67	No	76
No	67	No	76
No	68	No	78
No	69	Yes	79
Yes	70	No	81

The evidence begins to speak for itself.

What if engineers had made this table before the launch?

Why didn't NASA make the right decision?

Why didn't NASA make the right decision?

Many answers in the literature:
bureaucratic politics; group think; bounded rationality, etc

The *Challenger* launch decision

Why didn't NASA make the right decision?

Many answers in the literature:
bureaucratic politics; group think; bounded rationality, etc

But Edward Tufte thinks it may have been a matter of presentation & modeling:

Why didn't NASA make the right decision?

Many answers in the literature:
bureaucratic politics; group think; bounded rationality, etc

But Edward Tufte thinks it may have been a matter of presentation & modeling:

- Never made the right tables or graphics
- Selected only failure data
- Never considered a simple statistical model

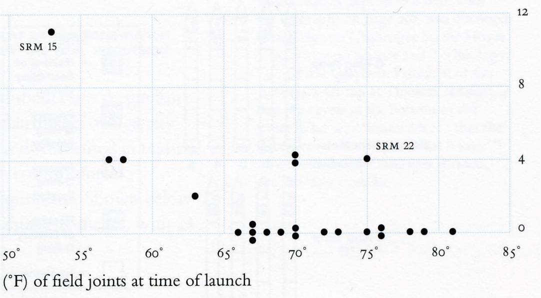
What do you think? How would you approach the data?

How about a scatterplot? Better for seeing relationships than a table.

Vertical axis is an O-ring damage index (due to Tufte, who made the plot)

How about a scatterplot? Better for seeing relationships than a table.

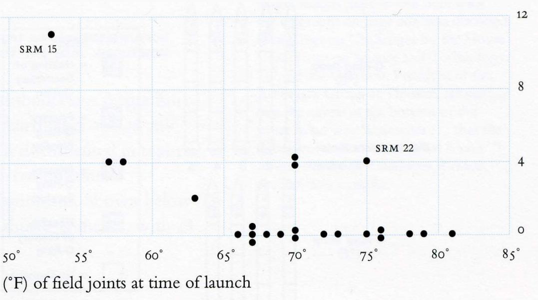
Vertical axis is an O-ring damage index (due to Tufte, who made the plot)



Suspicious.

How about a scatterplot? Better for seeing relationships than a table.

Vertical axis is an O-ring damage index (due to Tufte, who made the plot)

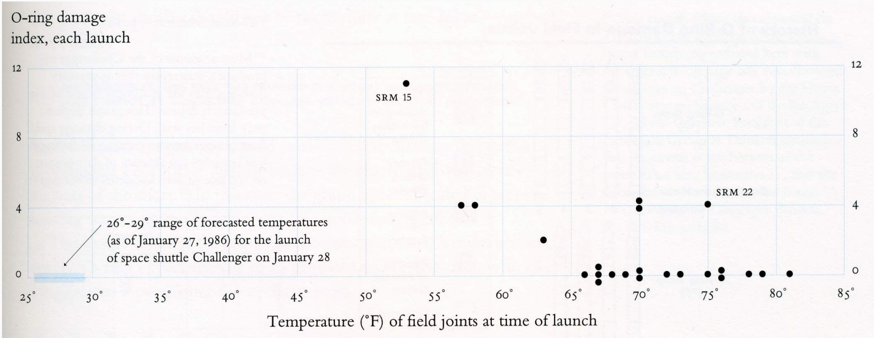


Suspicious. What was the forecast temperature for launch?

What was the forecast temperature for launch?

What was the forecast temperature for launch? 26 to 29 °F!

What was the forecast temperature for launch? 26 to 29 °F!



The shuttle was launched in unprecedented cold

Imagine you are the analyst making the launch recommendation.

You've made the scatterplot above. What would you add to it?

Put another way, what do you is the first question you expect from your boss?

Imagine you are the analyst making the launch recommendation.

You've made the scatterplot above. What would you add to it?

Put another way, what do you is the first question you expect from your boss?

"What's the chance of failure at 26 °F?"

Imagine you are the analyst making the launch recommendation.

You've made the scatterplot above. What would you add to it?

Put another way, what do you is the first question you expect from your boss?

"What's the chance of failure at 26 °F?"

The scatterplot suggests the answer is "high", but that's vague

Imagine you are the analyst making the launch recommendation.

You’ve made the scatterplot above. What would you add to it?

Put another way, what do you is the first question you expect from your boss?

“What’s the chance of failure at 26 °F?”

The scatterplot suggests the answer is “high”, but that’s vague

But what if the next launch is at 58 °F? Or 67 °F?

Imagine you are the analyst making the launch recommendation.

You've made the scatterplot above. What would you add to it?

Put another way, what do you is the first question you expect from your boss?

"What's the chance of failure at 26 °F?"

The scatterplot suggests the answer is "high", but that's vague

But what if the next launch is at 58 °F? Or 67 °F?

Clearly, we want a more precise way to state the probability of failure

We need a *model*, and a way to convey that model to the public.

The *Challenger* launch decision

Model the probability of O-ring damage as a function of temperature

We can use a statistical tool called “logit” for this purpose

The model is nonlinear: $\Pr(\text{damage}) = (1 - \exp(-\beta_0 - \beta_1 \text{temperature}))^{-1}$

Model the probability of O-ring damage as a function of temperature

We can use a statistical tool called "logit" for this purpose

The model is nonlinear: $\Pr(\text{damage}) = (1 - \exp(-\beta_0 - \beta_1 \text{temperature}))^{-1}$

R gives us this lovely logit output...

Model the probability of O-ring damage as a function of temperature

We can use a statistical tool called "logit" for this purpose

The model is nonlinear: $\mathrm{Pr}(\mathrm{damage}) = (1 - \exp(-\beta_0 - \beta_1 \mathrm{temperature}))^{-1}$

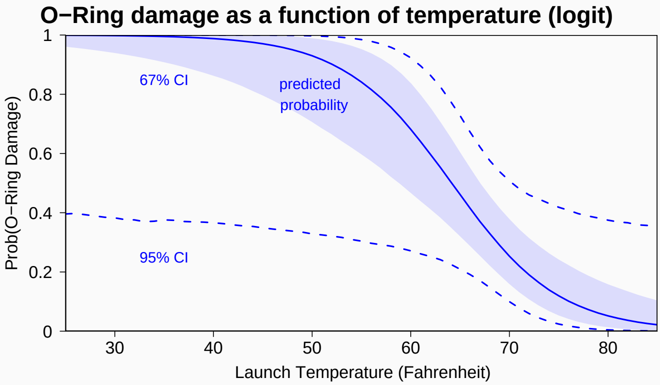
R gives us this lovely logit output...

Variable	est.	s.e.	*p*
Temperature (F)	-0.18	0.09	0.047
Constant	11.9	6.34	0.062
N	22		
log-likelihood	-10.9		

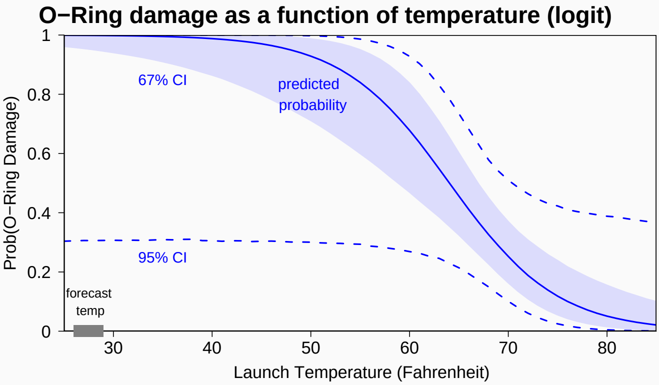
which most social scientists read as "a statistically significant negative relationship b/w temperature and probability of damage"

But that's pretty vague too.

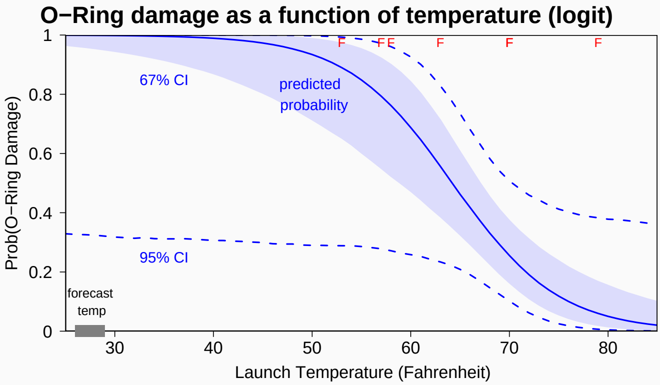
Is there a more persuasive/clear/useful way to present these results?



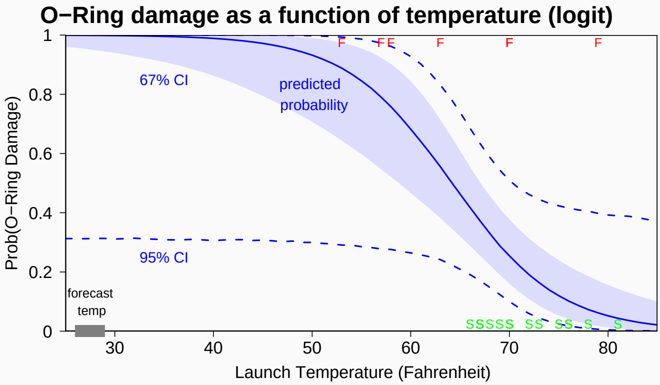
A picture clearly shows non-linear model predictions *and* uncertainty



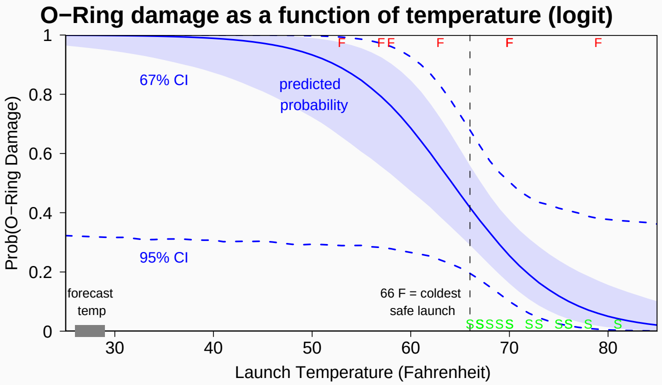
And gives a more precise sense of how foolhardy launching at 29 F is.



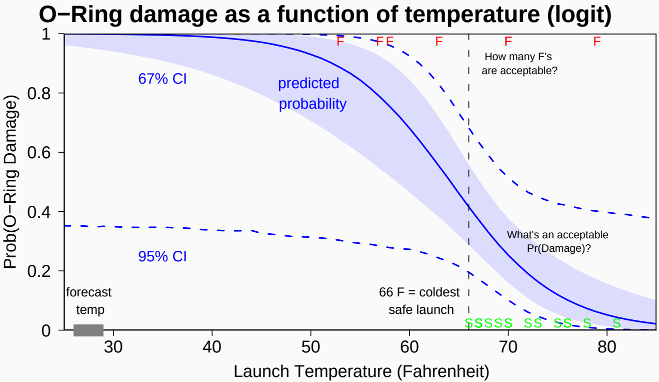
It's also good to show the data giving rise to the model.



Remembering that the Failures are only meaningful compared to Successes



Looking just at the data might show that launches under 66 F likely O-ring failures.



This inference is based on an unstated model.

In a hearing, Richard Feynmann dramatically showed O-rings lose resilence when cold by dropping one in his ice water.

Experiment cut thru weeks of technical gibberish concealing flaws in the O-ring

But it shouldn't have taken a Nobel laureate:
any scientist with a year of statistical training could have used the launch record to reach the same conclusion

And it would take no more than a single graphic to show the result