Final Review

Jacob M. Montgomery

Quantitative Political Methodology

Final Review

Last Call

- It's over. This is the end.

Last Call

- It's over. This is the end.
- Let's make sure all groups have their posters in

Last Call

- It's over. This is the end.
- Let's make sure all groups have their posters in
- **Peer evaluations** due at time of exam

Last Call

- It's over. This is the end.
- Let's make sure all groups have their posters in
- **Peer evaluations** due at time of exam
- Practice exam/review materials will be posted to website

Last Call

- It's over. This is the end.
- Let's make sure all groups have their posters in
- **Peer evaluations** due at time of exam
- Practice exam/review materials will be posted to website
- Office hours for me and grad students will continue.

Last Call

- It's over. This is the end.
- Let's make sure all groups have their posters in
- **Peer evaluations** due at time of exam
- Practice exam/review materials will be posted to website
- Office hours for me and grad students will continue.
- I will have additional office hours before the exam. Check Facebook.

Last Call

- It's over. This is the end.
- Let's make sure all groups have their posters in
- **Peer evaluations** due at time of exam
- Practice exam/review materials will be posted to website
- Office hours for me and grad students will continue.
- I will have additional office hours before the exam. Check Facebook.
- Joe and Cara are done. (Applause)!

Last Call

- It's over. This is the end.
- Let's make sure all groups have their posters in
- **Peer evaluations** due at time of exam
- Practice exam/review materials will be posted to website
- Office hours for me and grad students will continue.
- I will have additional office hours before the exam. Check Facebook.
- Joe and Cara are done. (Applause)!
- Bring your printed out posters to lab.

Last Call

- It's over. This is the end.
- Let's make sure all groups have their posters in
- **Peer evaluations** due at time of exam
- Practice exam/review materials will be posted to website
- Office hours for me and grad students will continue.
- I will have additional office hours before the exam. Check Facebook.
- Joe and Cara are done. (Applause)!
- Bring your printed out posters to lab.
- Poster session

Course evaluations

- I read all comments.

Course evaluations

- I read all comments.
- Used to evaluate me professionally and especially the graduate students.

Course evaluations

- I read all comments.
- Used to evaluate me professionally and especially the graduate students.
- 1% Extra credit for completing.

Exam checklist

- Peer evaluation form
- Pencils/erasers
- One sheet of paper
- Calculator with memory cleared

Overview of today

- Big picture overview of the class
- What test do I use here again?
- Questions?
- In-class review problems (answers will be posted to BB)

What is statistics?

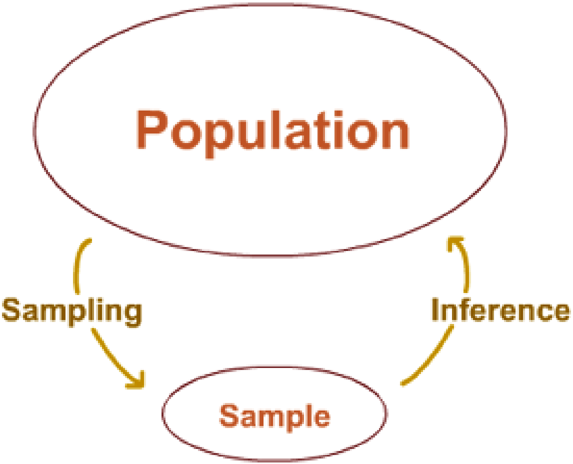
A body of methods for collecting and analyzing data.

What is statistics?

A body of methods for collecting and analyzing data.

Let's break that down.

- **Research design**: Gathering data that will allow us to answer research questions by testing empirical hypotheses.
- **Description**: Summarizing the *data*.
- **Inference**: Using data to make (probabilistic) statements about the *real world*. Testing our hypotheses.


Population
Sampling
Inference
Sample

(1) What kind of data are we gathering?

	Quantitative		**Qualitative (categorical)**	
	Continuous	*Discrete*	*Continuous*	*Discrete*
Interval				

(1) What kind of data are we gathering?

	Quantitative		**Qualitative (categorical)**	
	Continuous	*Discrete*	*Continuous*	*Discrete*
Interval	ex., Income			

(1) What kind of data are we gathering?

	Quantitative		Qualitative (categorical)	
	Continuous	*Discrete*	*Continuous*	*Discrete*
Interval	ex., Income	ex, Family size	NA	NA
Ordinal	NA			

(1) What kind of data are we gathering?

	Quantitative		**Qualitative (categorical)**	
	Continuous	*Discrete*	*Continuous*	*Discrete*
Interval	ex., Income	ex, Family size	NA	NA
Ordinal	NA	ex., Love stats.	NA	NA
Nominal	NA	NA	NA	

(1) What kind of data are we gathering?

	Quantitative		**Qualitative (categorical)**	
	Continuous	*Discrete*	*Continuous*	*Discrete*
Interval	ex., Income	ex, Family size	NA	NA
Ordinal	NA	ex., Love stats.	NA	NA
Nominal	NA	NA	NA	ex., Eye color

(2) Now that we have this data, how can we summarize it?

Sample mean

$$\bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

(2) Now that we have this data, how can we summarize it?

Sample mean

$$\bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n} = \frac{1}{n}\sum_{i=1}^{n} y_i \equiv \frac{1}{n}\sum y_i$$

Median

The observation that falls in the center of an ordered sample.

Mode

The most frequently occurring value

(2) Now that we have this data, how can we summarize it?

Range

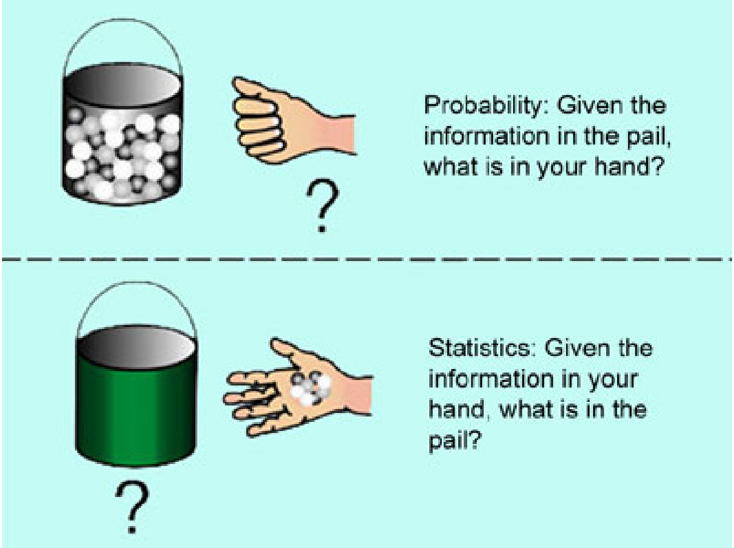
$$(max(y) - min(y)) \equiv |min(y) - max(y)|$$

Standard deviation

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

Interquartile range (IQR)

75^{th} *percentile* - 25^{th} *percentile*


Probability: Given the information in the pail, what is in your hand?
?
Statistics: Given the information in your hand, what is in the pail?
?

(3) Now that we have these statistics, how can we make inferences? Sampling distributions

Sampling distributions

> *A sampling distribution is the distribution of a* ***statistic*** *given repeated sampling.*

(3) Now that we have these statistics, how can we make inferences? Sampling distributions

Sampling distributions

> *A sampling distribution is the distribution of a* ***statistic*** *given repeated sampling.*

We use probability theory (especially the central limit theorem) to calculate the sampling distribution of various statistics.

- $\bar{y} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ (large n)

(3) Now that we have these statistics, how can we make inferences? Sampling distributions

Sampling distributions

> *A sampling distribution is the distribution of a* ***statistic*** *given repeated sampling.*

We use probability theory (especially the central limit theorem) to calculate the sampling distribution of various statistics.

- $\bar{y} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ (large n)
- $\frac{\bar{y}-\mu}{\hat{\sigma}_{\bar{y}}} \approx \frac{\bar{y}-\mu}{\sigma} = Z \sim N(0,1)$ (large n)

(3) Now that we have these statistics, how can we make inferences? Sampling distributions

Sampling distributions

> *A sampling distribution is the distribution of a* ***statistic*** *given repeated sampling.*

We use probability theory (especially the central limit theorem) to calculate the sampling distribution of various statistics.

- $\bar{y} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ (large n)
- $\frac{\bar{y}-\mu}{\hat{\sigma}_{\bar{y}}} \approx \frac{\bar{y}-\mu}{\sigma} = Z \sim N(0,1)$ (large n)
- $\frac{\bar{y}-\mu}{\hat{\sigma}_{\bar{y}}} = t \sim t_{df=n-1}$ (small n with normal y)

(3) Now that we have these statistics, how can we make inferences? Sampling distributions

Sampling distributions

> *A sampling distribution is the distribution of a* ***statistic*** *given repeated sampling.*

We use probability theory (especially the central limit theorem) to calculate the sampling distribution of various statistics.

- $\bar{y} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ (large n)
- $\frac{\bar{y}-\mu}{\hat{\sigma}_{\bar{y}}} \approx \frac{\bar{y}-\mu}{\sigma} = Z \sim N(0,1)$ (large n)
- $\frac{\bar{y}-\mu}{\hat{\sigma}_{\bar{y}}} = t \sim t_{df=n-1}$ (small n with normal y)
- $\frac{\hat{\pi}-\pi}{\sigma_{\hat{\pi}}} = Z \sim N(0,1)$ (large n)

(3) Now that we have these statistics, how can we make inferences? Sampling distributions

Sampling distributions

> *A sampling distribution is the distribution of a* ***statistic*** *given repeated sampling.*

We use probability theory (especially the central limit theorem) to calculate the sampling distribution of various statistics.

- $\bar{y} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ (large n)
- $\frac{\bar{y}-\mu}{\hat{\sigma}_{\bar{y}}} \approx \frac{\bar{y}-\mu}{\sigma} = Z \sim N(0,1)$ (large n)
- $\frac{\bar{y}-\mu}{\hat{\sigma}_{\bar{y}}} = t \sim t_{df=n-1}$ (small n with normal y)
- $\frac{\hat{\pi}-\pi}{\sigma_{\hat{\pi}}} = Z \sim N(0,1)$ (large n)
- $\hat{\beta}$, $\hat{\alpha}$, F, χ^2, $\bar{y}_2 - \bar{y}_1$

(3) Now that we have these statistics, how can we make inferences? Confidence intervals

We want to say something useful about the unknown population parameter. One way to do this is to set up a confidence interval.

Point estimation

> *A sample statistic that gives a good guess about a population parameter.*

(3) Now that we have these statistics, how can we make inferences? Confidence intervals

We want to say something useful about the unknown population parameter. One way to do this is to set up a confidence interval.

Point estimation

> *A sample statistic that gives a good guess about a population parameter.*

Caonfidence intervals

A confidence interval for a population parameter is a range of numbers within which a parameter is believed to fall.

Caonfidence intervals

A confidence interval for a population parameter is a range of numbers within which a parameter is believed to fall.

Confidence coefficient

The probability that an interval would contain the parameter with repeated sampling.

This is calculated as: Point Estimate $\pm$ (Z or T) $\times$ Standard error

(4) Now that we have these statistics, how can we make inferences? Hypothesis tests

We calculate statistics with known sampling distributions to:

- Make systematic statements about population parameters for a single population.
- Make systematic statements about how population parameters are related to explanatory variabels

(5) Now that we have these statistics, how can we make inferences? Hypothesis tests

Five steps of hypothesis testing:

1. Make/state assumptions about our data
2. Formulate null and alternative hypotheses
3. Calculate the appropriate test statistic
4. Calculate the p-value
5. Draw a conclusion

P-value

A P-Value is a measure of surprise.

(5) Now that we have these statistics, how can we make inferences? Hypothesis tests

Five steps of hypothesis testing:

1. Make/state assumptions about our data
2. Formulate null and alternative hypotheses
3. Calculate the appropriate test statistic
4. Calculate the p-value
5. Draw a conclusion

P-value

> *A P-Value is a measure of surprise. We ask, "If the null hypothesis is true, how likely is it that we would observe a test-statistic this extreme* ***or more?"***

(6) Using the appropriate test for your data

Resp. Variable	Explanatory Variables	Measure	Statistical Test(s)	Sample Size
Interval	None (population estimate)	Mean	One-sample z-test (1)	large

(6) Using the appropriate test for your data

Resp. Variable	Explanatory Variables	Measure	Statistical Test(s)	Sample Size
Interval	None (population estimate)	Mean	One-sample z-test (1)	large
			One-sample t-test (2)	small

(6) Using the appropriate test for your data

Resp. Variable	**Explanatory Variables**	**Measure**	**Statistical Test(s)**	**Sample Size**
Interval	None (population estimate)	Mean	One-sample z-test (1)	large
			One-sample t-test (2)	small
Categorical	None (population estimate)	Proportion	One sample z-test (3)	large

(6) Using the appropriate test for your data

Resp. Variable	Explanatory Variables	Measure	Statistical Test(s)	Sample Size
Interval	None (population estimate)	Mean	One-sample z-test (1)	large
			One-sample t-test (2)	small
Categorical	None (population estimate)	Proportion	One sample z-test (3)	large
Interval	Categorical (2 Groups)	Mean	Two-sample z-test (4)	large

(6) Using the appropriate test for your data

Resp. Variable	Explanatory Variables	Measure	Statistical Test(s)	Sample Size
Interval	None (population estimate)	Mean	One-sample z-test (1)	large
			One-sample t-test (2)	small
Categorical	None (population estimate)	Proportion	One sample z-test (3)	large
Interval	Categorical (2 Groups)	Mean	Two-sample z-test (4)	large
			Two-sample t-test (5)	small

(6) Using the appropriate test for your data

Resp. Variable	Explanatory Variables	Measure	Statistical Test(s)	Sample Size
Interval	None (population estimate)	Mean	One-sample z-test (1)	large
			One-sample t-test (2)	small
Categorical	None (population estimate)	Proportion	One sample z-test (3)	large
Interval	Categorical (2 Groups)	Mean	Two-sample z-test (4)	large
			Two-sample t-test (5)	small
Categorical	Categorical (2+ Groups)	Proportion	χ^2 (Contingency Tables) (6)	large*

(6) Using the appropriate test for your data

Resp. Variable	**Explanatory Variables**	**Measure**	**Statistical Test(s)**	**Sample Size**
Interval	None (population estimate)	Mean	One-sample z-test (1)	large
			One-sample t-test (2)	small
Categorical	None (population estimate)	Proportion	One sample z-test (3)	large
Interval	Categorical (2 Groups)	Mean	Two-sample z-test (4)	large
			Two-sample t-test (5)	small
Categorical	Categorical (2+ Groups)	Proportion	χ^2 (Contingency Tables) (6)	large*
Interval	Categorical (2+ Groups)	Mean	Regression (F-test) (7)	large

(6) Using the appropriate test for your data

Resp. Variable	Explanatory Variables	Measure	Statistical Test(s)	Sample Size
Interval	None (population estimate)	Mean	One-sample z-test (1)	large
			One-sample t-test (2)	small
Categorical	None (population estimate)	Proportion	One sample z-test (3)	large
Interval	Categorical (2 Groups)	Mean	Two-sample z-test (4)	large
			Two-sample t-test (5)	small
Categorical	Categorical (2+ Groups)	Proportion	χ^2 (Contingency Tables) (6)	large*
Interval	Categorical (2+ Groups)	Mean	Regression (F-test) (7)	large
Interval	Interval	Continuous	Regression (8, 9)	

(6) Using the appropriate test for your data

Resp. Variable	Explanatory Variables	Measure	Statistical Test(s)	Sample Size
Interval	None (population estimate)	Mean	One-sample z-test (1)	large
			One-sample t-test (2)	small
Categorical	None (population estimate)	Proportion	One sample z-test (3)	large
Interval	Categorical (2 Groups)	Mean	Two-sample z-test (4)	large
			Two-sample t-test (5)	small
Categorical	Categorical (2+ Groups)	Proportion	χ^2 (Contingency Tables) (6)	large*
Interval	Categorical (2+ Groups)	Mean	Regression (F-test) (7)	large
Interval	Interval	Continuous	Regression (8, 9)	

Test number	Test statistic	How is that distributed?
1	$\dfrac{\bar{y} - \mu_0}{S/\sqrt{n}}$	Standard normal (Z)

Test number	Test statistic	How is that distributed?
1	$\dfrac{\bar{y} - \mu_0}{S/\sqrt{n}}$	Standard normal (Z)
2	$\dfrac{\bar{y} - \mu_0}{S/\sqrt{n}}$	t-dist., $df = n - 1$

Test number	Test statistic	How is that distributed?
1	$\dfrac{\bar{y}-\mu_0}{S/\sqrt{n}}$	Standard normal (Z)
2	$\dfrac{\bar{y}-\mu_0}{S/\sqrt{n}}$	t-dist., $df = n-1$
3	$\dfrac{\hat{\pi}-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$	Standard normal (Z)

Test number	Test statistic	How is that distributed?
4	$\dfrac{(\bar{y}_2 - \bar{y}_1) - 0}{\sqrt{\frac{S_2^2}{n_2} + \frac{S_1^2}{n_1}}}$	Standard normal (Z)

Test number	Test statistic	How is that distributed?
4	$\dfrac{(\bar{y}_2 - \bar{y}_1) - 0}{\sqrt{\frac{S_2^2}{n_2} + \frac{S_1^2}{n_1}}}$	Standard normal (Z)
5	$\dfrac{(\bar{y}_2 - \bar{y}_1) - 0}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	t-dist, $df = n - 2$

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Test number	Test statistic	How is that distributed?
4	$\dfrac{(\bar{y_2} - \bar{y_1}) - 0}{\sqrt{\frac{S_2^2}{n_2} + \frac{S_1^2}{n_1}}}$	Standard normal (Z)
5	$\dfrac{(\bar{y_2} - \bar{y_1}) - 0}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	t-dist, $df = n - 2$
	$\hat{\sigma} = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$	
6	$\chi^2 = \sum \dfrac{(f_0 - f_e)^2}{fe}$	χ^2, $df = (rows - 1)(cols - 1)$

Test number	Test statistic	How is that distributed?
4	$\dfrac{(\bar{y}_2 - \bar{y}_1) - 0}{\sqrt{\frac{S_2^2}{n_2} + \frac{S_1^2}{n_1}}}$	Standard normal (Z)
5	$\dfrac{(\bar{y}_2 - \bar{y}_1) - 0}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	t-dist, $df = n - 2$
	$\hat{\sigma} = \sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$	
6	$\chi^2 = \sum \dfrac{(f_0 - f_e)^2}{fe}$	χ^2, $df = (rows - 1)(cols - 1)$
7	F-statistic	F, $df_1 = k, df_2 = n - (k + 1)$

Test number	Test statistic	How is that distributed?
8 (bivariate)	$\frac{\hat{\beta}}{\hat{\sigma}/\sqrt{(X_i - \bar{X})^2}}$	$t - dist$, $df = n - 2$
	$\hat{\beta} = \frac{\sum_{i=1}^{n} \left((X_i - \bar{X})(Y_i - \bar{Y}) \right)}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$	
	$\hat{\sigma} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}}$	

Test number	Test statistic	How is that distributed?
8 (bivariate)	$\dfrac{\hat{\beta}}{\hat{\sigma}/\sqrt{(X_i - \bar{X})^2}}$	$t - dist$, $df = n - 2$
	$\hat{\beta} = \dfrac{\sum_{i=1}^{n}\left((X_i - \bar{X})(Y_i - \bar{Y})\right)}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$	
	$\hat{\sigma} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n-2}}$	
	$\hat{\alpha} = \dfrac{\bar{Y} - \hat{\beta}\bar{X}}{\hat{\sigma}_{\hat{\alpha}}}$	$t - dist$, $df = n - 2$

Test number	Test statistic	How is that distributed?
8 (bivariate)	$\dfrac{\hat{\beta}}{\hat{\sigma}/\sqrt{(X_i - \bar{X})^2}}$	$t - dist,\ df = n - 2$
	$\hat{\beta} = \dfrac{\sum_{i=1}^{n} \left((X_i - \bar{X})(Y_i - \bar{Y}) \right)}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$	
	$\hat{\sigma} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}}$	
	$\hat{\alpha} = \dfrac{\bar{Y} - \hat{\beta}\bar{X}}{\hat{\sigma}_{\hat{\alpha}}}$	$t - dist,\ df = n - 2$
9 (multivariate)	$\dfrac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}$	$t - dist, df = n - (k + 1)$
	$\dfrac{\hat{\alpha}}{\hat{\sigma}_{\hat{\alpha}}}$	$t - dist,\ df = n - (k + 1)$

Using regression to establish causality

1. Experiments (Dummies)
2. Statistical control
3. Statistical control (Fixed effects dummies)
4. Difference-in-differences (Dummies with interactions)
5. Regression discontinuity
6. Instrumental variables

Parting thoughts

I hope you take away from this class:

- Quantitative analysis is within your grasp.

Parting thoughts

I hope you take away from this class:

- Quantitative analysis is within your grasp.
- Using data to answer questions can be hard, but using intuition is often just wrong.

Parting thoughts

I hope you take away from this class:

- Quantitative analysis is within your grasp.
- Using data to answer questions can be hard, but using intuition is often just wrong.
- Don't skip the tables.

Parting thoughts

I hope you take away from this class:

- Quantitative analysis is within your grasp.
- Using data to answer questions can be hard, but using intuition is often just wrong.
- Don't skip the tables.
- Be an informed and thoughtful consumer of ..
    1. political science research,

Parting thoughts

I hope you take away from this class:

- Quantitative analysis is within your grasp.
- Using data to answer questions can be hard, but using intuition is often just wrong.
- Don't skip the tables.
- Be an informed and thoughtful consumer of ..
  1. political science research,
  2. all research,

Parting thoughts

I hope you take away from this class:

- Quantitative analysis is within your grasp.
- Using data to answer questions can be hard, but using intuition is often just wrong.
- Don't skip the tables.
- Be an informed and thoughtful consumer of ..
    1. political science research,
    2. all research,
    3. polls,
    4. rhetoric,
    5. and common wisdom.