Poli 30D Political Inquiry

Hypothesis Testing, χ^2 Distribution & Qualitative Methods

Shane Xinyang Xuan
ShaneXuan.com

November 22, 2016

Contact Information

Shane Xinyang Xuan
xxuan@ucsd.edu

We have someone to help you every day!

Professor Desposato	M	1330-1500 (Latin American Center)
Shane Xuan	Tu	1600-1800 (SSB332)
Cameron Sells	W	1000-1200 (SSB352)
Kelly Matush	Th	1500-1700 (SSB343)
Julia Clark	F	1200-1400 (SSB326)

Supplemental Materials

Our class oriented

ShaneXuan.com

UCLA SPSS starter kit

`www.ats.ucla.edu/stat/spss/sk/modules_sk.htm`

Princeton data analysis

`http://dss.princeton.edu/training/`

Quiz

You self-evaluated yourself during Week 5 in terms of how hard have you tried in this class. The average is below 60%. Hopefully you have tried harder during the second half of the quarter after the initial evaluation. Now, please re-evaluate yourself (on a scale of 0-10). Give you one score in your work ethics and explain.

Things to consider:

1. I went to Professor/TA's office hour very often
2. I participated in lectures/sections very often
3. I did not do (1)/(2) very often because I know the materials pretty well

Open-ended: Compared to my midterm evaluation, I have improved in the following way: (_____)

Announcement

I will have your grades (5% attendance, 2.5% quiz, 2.5% participation, 60% homework; that is a total of 70 points) ready by Dec 6. If you want to know your grade, you should come to see me during my office hours on Tuesday (12/6) from 4-6 pm. My access to email is really limited during the finals week, so please give me up to 48 hours to reply to your email. I have a few travel plans after the Thursday (12/8), so it will definitely take me 24-48 hours to get back to you.

Understand your grades

- 5% attendance
- 2.5% quiz
- 2.5% participation
- 60% homework
- 30% final

Understand your grades

- 5% attendance
- 2.5% quiz
- 2.5% participation
- 60% homework
- 30% final

You will know them by 12/6.

Understand your grades

- 5% attendance
- 2.5% quiz
- 2.5% participation
- 60% homework
- 30% final

You won't know it until the week after final.

Understand your grades

- 5% attendance
- 2.5% quiz
- 2.5% participation
- 60% homework
- 30% final

If I tell you that you have 60 points on 12/6, that means your highest possible score in this class is 90%, if no curve is given.

Understand your grades

- 5% attendance
- 2.5% quiz
- 2.5% participation
- 60% homework
- 30% final

If I tell you that you have 60 points on 12/6, that means your highest possible score in this class is 90%, if no curve is given.
If I tell you that you have 29 points on 12/6, that means you will fail this class even though you ace the final, if no curve is given.

Understand your grades

- 5% attendance
- 2.5% quiz
- 2.5% participation
- 60% homework
- 30% final

If I tell you that you have 60 points on 12/6, that means your highest possible score in this class is 90%, if no curve is given.
If I tell you that you have 29 points on 12/6, that means you will fail this class even though you ace the final, if no curve is given.
Most student will probably get around 55 points. That means if you try hard enough, say, get 25 out of the 30 points (around 83%) in the final, then you will get a B-.

Understand your grades

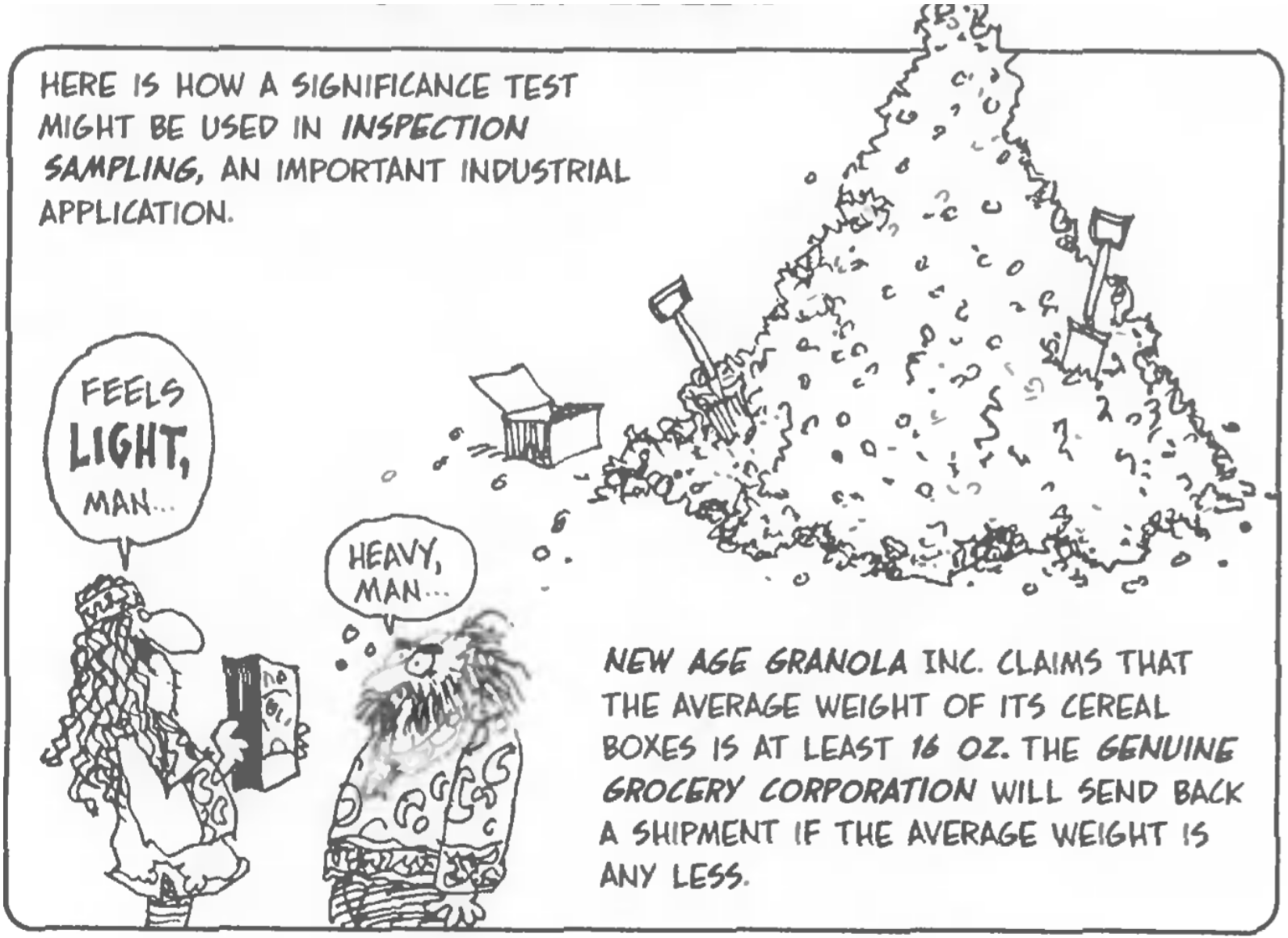
What is the point of trying?

You never know if a curve will be given. Say you do not try your best, and get 88% eventually. If we curve everyone up by 1%, then your friend who gets 89% initially will get an A-. However, because you have not tried your best, even with the curve, you will still keep your B+. I don't know if a curve will be given or if you will be the person in the aforementioned scenario – but please try your best so that you won't regret.

Research Design

IV/DV	Nominal/Ordinal	Interval/Ratio
Nominal/Ordinal	Crosstab & Barplot	Summary Table & Barplot
Interval/Ratio	Recode & Crosstab	Scatterplot & Regression

Hypothesis testing: z-test



Hypothesis testing: z-test

FIRST, THEY CHOOSE THEIR HYPOTHESES.

$H_0: \mu = 16$ OZ.

$H_a: \mu < 16$ OZ.

REJECTING THE NULL HYPOTHESIS MEANS REFUSING THE GRANOLA.

Hypothesis testing: z-test

THEY PULL OUT A SIMPLE RANDOM SAMPLE OF 49 BOXES, WEIGH EACH ONE, AND DETERMINE THE SAMPLE'S SUMMARY STATISTICS:

$\bar{x} = 15.90$ OZ.

$s = .35$ OZ.

A LITTLE LIGHT—BUT SIGNIFICANTLY SO?

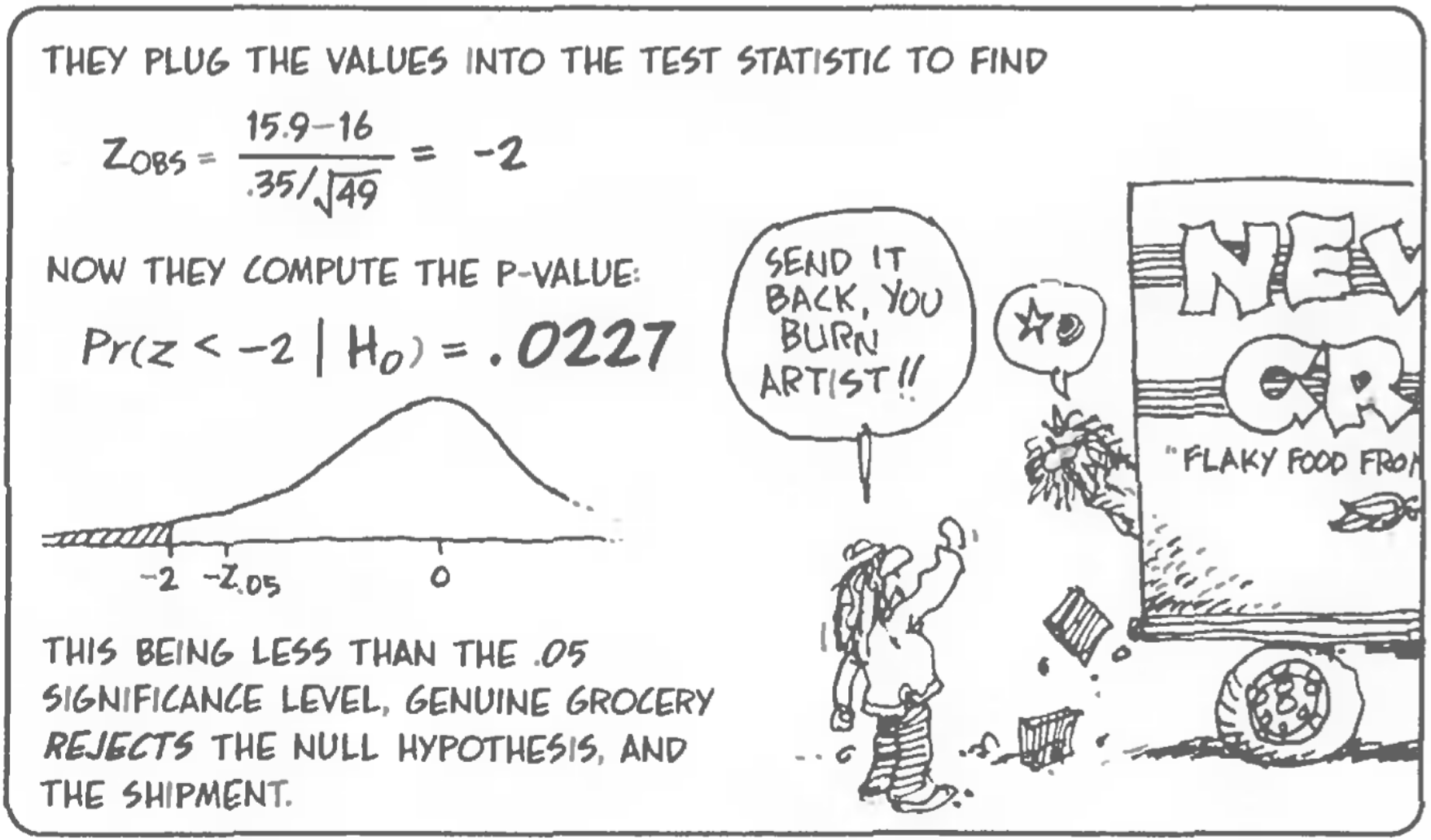
Hypothesis testing: z-test

NEXT, THEY CHOOSE A TEST STATISTIC. BY NOW, IT SHOULD BE PRETTY MUCH A KNEE-JERK REACTION TO KNOW THAT THE SAMPLE SPREAD FROM THE MEAN IS

$$\frac{\bar{X}-\mu_0}{SE(\bar{X})} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$$

WHERE s IS THE SAMPLE STANDARD DEVIATION. UNDER THE NULL HYPOTHESIS, THIS APPROXIMATES THE STANDARD NORMAL WHEN THE SAMPLE IS LARGE, BY THE CENTRAL LIMIT THEOREM.

Hypothesis testing: z-test



Hypothesis testing: z-test

For one-sample proportion test, the standard error is

$$\sqrt{\frac{p_0(1-p_0)}{n}}$$

So the z-score will be

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Hypothesis testing: t-test

If the absolute value of the t-value is greater than the critical value, you reject the null hypothesis. If the absolute value of the t-value is less than the critical value, you fail to reject the null hypothesis.

Hypothesis testing: t-test

WE RETURN TO CHAMELEON MOTORS, AND ITS 10 M.P.H. CRASH TEST. THE *RIGHTEOUS INSURANCE COMPANY* WILL INSURE AN AUTO ONLY IF THE MEAN REPAIR COST AFTER A 10 M.P.H. COLLISION IS LESS THAN $1000. THE COMPANY USES A STANDARD $\alpha = .05$ AS ITS SIGNIFICANCE LEVEL. SO...

H_0: $\mu \geq \$1000$ MEAN COST IS TOO HIGH

H_a: $\mu < \$1000$ MEAN COST IS O.K.

THE TEST STATISTIC IS THE t DISTRIBUTION

$$t = \frac{\bar{X} - \mu_0}{SE(\bar{X})}$$

WHERE μ_0 IS THE HYPOTHETICAL MEAN OF $1000

Hypothesis testing: t-test

		α		
		.05	.025	.005
DEGREES OF FREEDOM	1	6.31	12.71	63.66
	2	2.92	4.30	9.92
	3	2.35	3.18	5.84
	4	2.13	2.78	4.60
	5	2.01	2.57	4.03

FROM THE TABLE OF CRITICAL t VALUES, WE SEE THAT $t_{.05} = 2.13$, SO WE DECIDE TO REJECT H_0 IF

$$t_{OBS} \leq -t_{.05} = -2.13$$

FROM CHAPTER 8, WE HAVE $\bar{x} = \$540$ AND $s = \$299$ FOR A SMALL, FIVE-CAR SAMPLE, SO WE FIND

$$t_{OBS} = \frac{540 - 1000}{299/\sqrt{5}}$$

$$= -3.44 < -t_{.05}$$

You reject the null hypothesis.

χ^2 Distribution

χ^2 is calculated by

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$f_e \sim$ expected value $\rightsquigarrow$ needs calculation

$f_o \sim$ observed value $\rightsquigarrow$ just observe

χ^2 Distribution

On December 18, 2010, the U.S. Senate voted on the question of whether to repeal the "don't ask, don't tell" policy regarding gays in the military. The relationship between party affiliation and vote is shown in the following table:

Repeal	Dem.	Rep.	Total
No	0	31	31
Yes	55	8	63
Total	55	39	94

Calculate the column percentage and interpret your results.

χ^2 Distribution

This is what you should get:

Repeal	Dem.	Rep.	Total
No	0 (0%)	31 (79.49%)	31 (32.98%)
Yes	55 (100%)	8 (20.51%)	63 (67.02%)
Total	55 (100%)	39 (100%)	94 (100%)

Now, calculate χ^2 for this table.

χ^2 Distribution

This is what you should get:

Repeal	Dem.	Rep.	Total
No	0 (0%)	31 (79.49%)	31 (32.98%)
Yes	55 (100%)	8 (20.51%)	63 (67.02%)
Total	55 (100%)	39 (100%)	94 (100%)

Now, calculate χ^2 for this table.

Hint: To do this problem, the calculations are as follows. You need to calculate f_e and f_o for each cell, and sum up $\frac{(f_o - f_e)^2}{f_e}$. For example, f_e (Republican) for 'No' is

$$f_e(\text{Rep}) = 39 \times 0.3298 = 12.8622.$$

Similarly, f_e (Democrat) for 'Yes' is

$$f_e(\text{Dem}) = 55 \times 67.02\% \approx 37.$$

χ^2 Distribution

You need to have a table that calculate f_o, f_e, and $\frac{(f_o-f_e)^2}{f_e}$ for respondents who answer 'no' to the question:

	Dem.	Rep.
f_o	0	31
f_e	18.14	12.8622
$\frac{(f_o-f_e)^2}{f_e}$	18	25

And another table for those who answer 'yes':

	Dem.	Rep.
f_o	55	8
f_e	37	26
$\frac{(f_o-f_e)^2}{f_e}$	8.76	12.5

χ^2 Distribution

Then, you sum up $\frac{(f_o - f_e)^2}{f_e}$:

No		Dem.	Rep.
	f_o	0	31
	f_e	18.14	12.8622
	$\frac{(f_o - f_e)^2}{f_e}$	18	25

Yes		Dem.	Rep.
	f_o	55	8
	f_e	37	26
	$\frac{(f_o - f_e)^2}{f_e}$	8.76	12.5

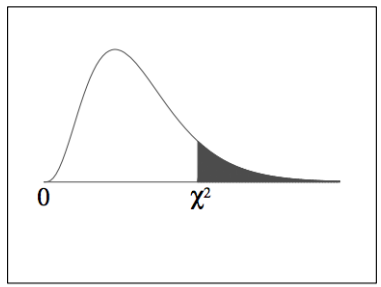
It follows that χ^2 is calculated by

$$\chi^2 = 18 + 25 + 8.76 + 12.5 = 64.25.$$

This is our test statistic. We will need to compare our test statistic to the critical value (to be discussed in the next 2 slides).

Read χ^2 table

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997

Interpret χ^2

- Degrees of freedom $=$ (#row – 1)(#col – 1)

Interpret χ^2

- Degrees of freedom $=$ (#row – 1)(#col – 1)
- The null hypothesis reigns over all the territory between 0 and the critical value of chi-square. We read the critical value from the table. In our case, the critical value is 3.841 because $\alpha = 0.05$ and $df = 1$.

Interpret χ^2

- Degrees of freedom $=$ (#row – 1)(#col – 1)
- The null hypothesis reigns over all the territory between 0 and the critical value of chi-square. We read the critical value from the table. In our case, the critical value is 3.841 because $\alpha = 0.05$ and $df = 1$.
- For any chi-square test statistic in this region ($0 < \chi^2 < 3.841$), the null hypothesis cannot be rejected.

Interpret χ^2

- Degrees of freedom $= (\#\text{row} - 1)(\#\text{col} - 1)$
- The null hypothesis reigns over all the territory between 0 and the critical value of chi-square. We read the critical value from the table. In our case, the critical value is 3.841 because $\alpha = 0.05$ and $df = 1$.
- For any chi-square test statistic in this region $(0 < \chi^2 < 3.841)$, the null hypothesis cannot be rejected.
- If our χ^2 statistic exceeds the critical value, then we should reject H_0.

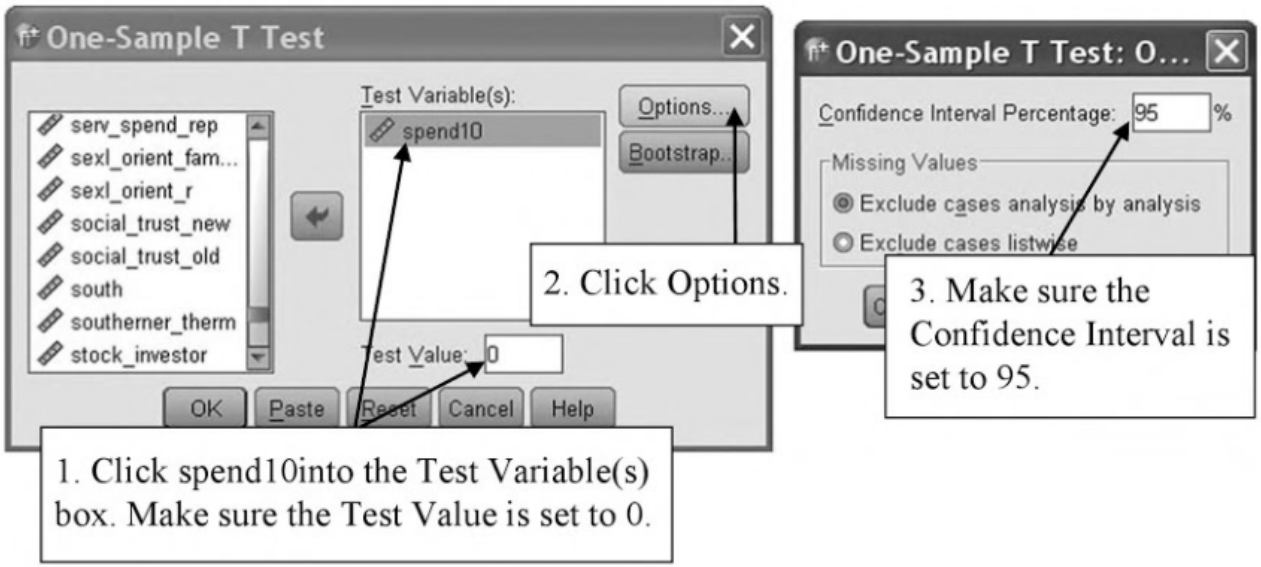
Interpret χ^2

- Degrees of freedom $= (\#\text{row} - 1)(\#\text{col} - 1)$
- The null hypothesis reigns over all the territory between 0 and the critical value of chi-square. We read the critical value from the table. In our case, the critical value is 3.841 because $\alpha = 0.05$ and $df = 1$.
- For any chi-square test statistic in this region $(0 < \chi^2 < 3.841)$, the null hypothesis cannot be rejected.
- If our χ^2 statistic exceeds the critical value, then we should reject H_0.
- In our case, since $64.25 > 3.841$, we can reject the null.

Interpret χ^2

- Degrees of freedom $= (\#\text{row} - 1)(\#\text{col} - 1)$
- The null hypothesis reigns over all the territory between 0 and the critical value of chi-square. We read the critical value from the table. In our case, the critical value is 3.841 because $\alpha = 0.05$ and $df = 1$.
- For any chi-square test statistic in this region ($0 < \chi^2 < 3.841$), the null hypothesis cannot be rejected.
- If our χ^2 statistic exceeds the critical value, then we should reject H_0.
- In our case, since $64.25 > 3.841$, we can reject the null.
- You have to calculate the test statistic, and read the critical value. The value that you read depends on both degree of freedom and α.

Hypothesis testing: SPSS



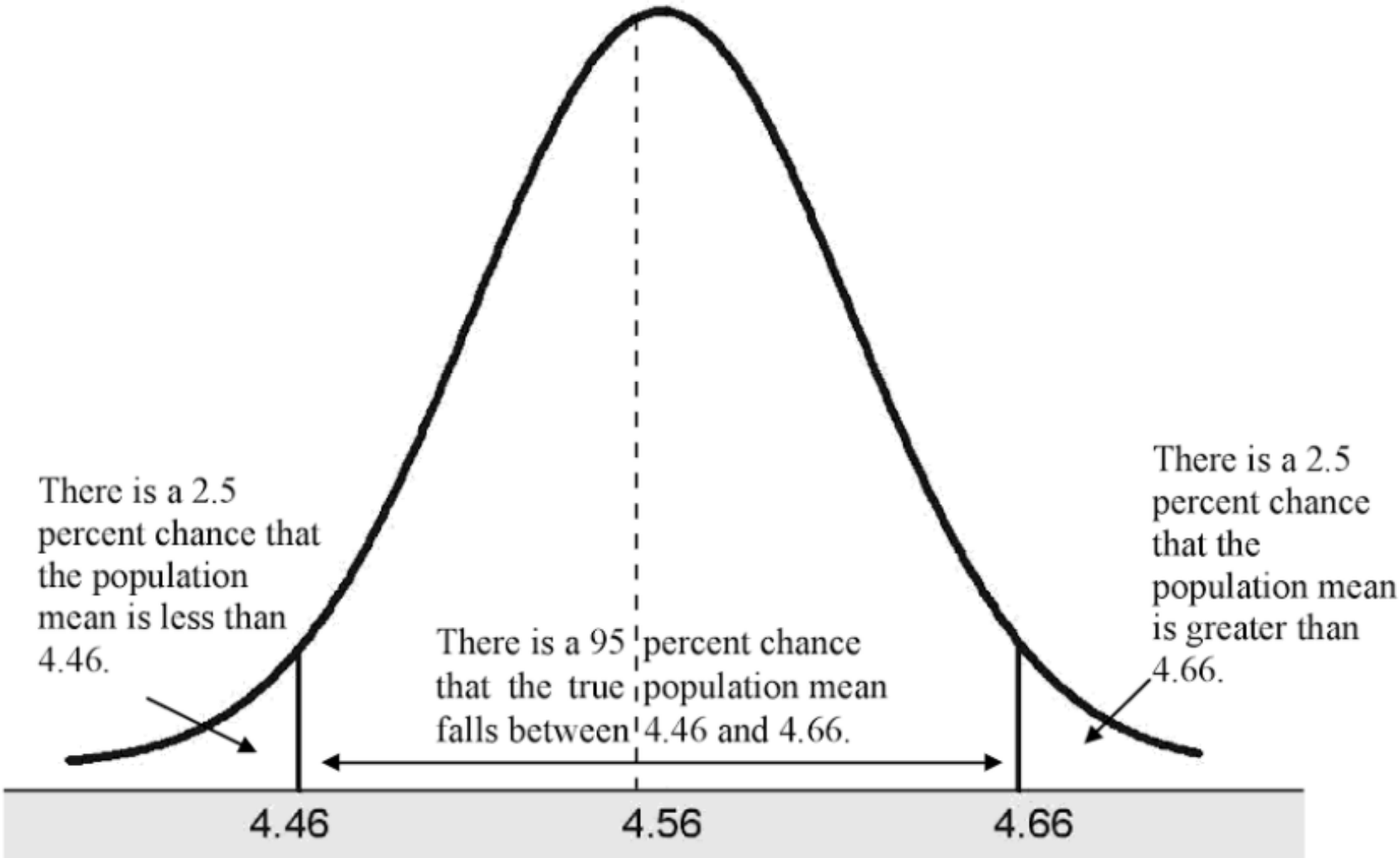
One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
spend10 Incr federal $:10 programs	2323	4.56	2.497	.052

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
spend10 Incr federal $:10 programs	88.014	2322	.000	4.561	4.46	4.66

Hypothesis testing: SPSS



Regression: SPSS

Coefficients[a]

Model[b]		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B[d]	Std. Error[e]	Beta[f]	t[g]	Sig.[g]	Lower Bound[h]	Upper Bound[h]
1	(Constant)[c]	12.325	3.194		3.859	.000	6.027	18.624
	math score[c]	.389	.074	.368	5.252	.000	.243	.535
	female[c]	-2.010	1.023	-.101	-1.965	.051	-4.027	.007
	social studies score[c]	.050	.062	.054	.801	.424	-.073	.173
	reading score[c]	.335	.073	.347	4.607	.000	.192	.479

a. Dependent Variable: science score

$$Y = \beta_0 + \beta_1(math) + \beta_2(female) + \beta_3(soc) + \beta_4(read)$$

Regression: SPSS

Coefficients[a]

Model[b]		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B	
		B[d]	Std. Error[e]	Beta[f]	t[g]	Sig.[g]	Lower Bound[h]	Upper Bound[h]
1	(Constant)[c]	12.325	3.194		3.859	.000	6.027	18.624
	math score[c]	.389	.074	.368	5.252	.000	.243	.535
	female[c]	-2.010	1.023	-.101	-1.965	.051	-4.027	.007
	social studies score[c]	.050	.062	.054	.801	.424	-.073	.173
	reading score[c]	.335	.073	.347	4.607	.000	.192	.479

a. Dependent Variable: science score

$$Y = \beta_0 + \beta_1(math) + \beta_2(female) + \beta_3(soc) + \beta_4(read)$$

This is a multiple regression. Note that under the column of B, you can read coefficients of the independent variables. Thus, the model can be written as

$$Y = 12.325 + 0.389(math) - 2.01(female) + 0.05(soc) + 0.335(read)$$

Regression: SPSS

Coefficients[a]

Model[b]		Unstandardized Coefficients		Standardized Coefficients	t[g]	Sig.[g]	95% Confidence Interval for B	
		B[d]	Std. Error[e]	Beta[f]			Lower Bound[h]	Upper Bound[h]
1	(Constant)[c]	12.325	3.194		3.859	.000	6.027	18.624
	math score[c]	.389	.074	.368	5.252	.000	.243	.535
	female[c]	-2.010	1.023	-.101	-1.965	.051	-4.027	.007
	social studies score[c]	.050	.062	.054	.801	.424	-.073	.173
	reading score[c]	.335	.073	.347	4.607	.000	.192	.479

a. Dependent Variable: science score

$$Y = 12.325 + 0.389(math) - 2.01(female) + 0.05(soc) + 0.335(read)$$

Under the column of Sig, SPSS tells you the p-values for each coefficient. Using $\alpha = 0.5$, the coefficient of math is significantly different from 0 because p-value $0 < \alpha = 0.05$. The coefficient of female is not significantly different from 0 because $0.051 > \alpha = 0.05$. The coefficient of socst is not significantly different from 0. The coefficient of read is significantly different from 0.

For final exam:

- Concepts: how to get z-score, t-score, and χ^2-statistic; how to get p-value from z-score and how to get critical values from t and χ^2 statistic
- Hypothesis test: what is null/alternative hypothesis; when to reject the null (for z-test, t-test, and χ^2-test)
- Confidence interval: how to construct a confidence interval for mean; CI for proportion
- How to read SPSS outputs on (1) regression table, (2) z-test, (3) χ^2-test
- My office hour is Tuesday from 4-6 pm $\rightsquigarrow$ please use them! These hours are reserved for you!

Study hard!

- Teaching evaluation $\rightsquigarrow$ helps me to improve my teaching; helps my department to get feedback; help my boss to determine if he should fire me
- For the quantitative reasoning certificate at UCSD, you can take Poli 5D in the winter quarter $\rightsquigarrow$ I will work with Professor Molly Roberts to teach you really cool stuff (R, and Excel)
- My office hour is Tuesday from 4-6 pm $\rightsquigarrow$ please use them! These hours are reserved for you!