AP Statistics CED 8.1 Daily Video 1

muouuuny	Statisti	cs – A	reiviy	Nesun	s one	xpecte	ed?				
What Will We	e Learn?										
How can we	determi	ne if ob	served	counts	in cate	egorical	data a	re cons	sistent	with ex	pected
counts due to random variation?											
More Catego	rical Da	ta									
In Unit 6, you	learnec	l about	inferer	nce pro	cedure	s for ca	tegorio	al data:	that w	ere cla	ssified in
terms of		ar	nd		·	But wha	at is a c	ategor	ical var	iable is	5
recorded with	า				cate	gories	? In Un	it 8, we	e will in	vestiga	te
inference pro	cedures	for the	e distrik	oution c	of	(catego	rical var	riable a	nd for	the
relationship b	etween		_ categ	gorical	variable	es.					
Fair Die?				-		Result	1	2 3	4 5	6 7	8 9 10
While playing	a role-	olaying	game	with so	me	Freque	ncy 7	11 7	11 9	8 12	10 13 12
friends, you r	otice th	at the ´	10-side	d die h	as beei	n rolling	g consis	stently			1.11
high numbers	s for a w	hile. Is	it your	imagin	ation, c	or could	, I the di	e be			
weighted? Yo	ou decid	e to ro	ll the d	ie 100 t	imes a	nd recc	ord the		1 2	3 4 5 6	7 8 9 10
results. Are th	nese res	ults cor	nsistent	with w	hat we	would	expect	in	1 2	Die Ro	II
random varia	tion?						I				
Expected Co	unts										
In order to ar	iswer ou	ır auest	ion ab	out whe	ether th	ne die is	s weiah	ited, we	e must	first co	nsider
how many of	each da	ita valu	e we w	ould		1	to see	if the di	e is		. If the
die is fair (un	weighte	d). we v	vould e	expect t	o see e	each of	the 10				
	re	epreser	nted in	100 rol	ls:			= ^	10 each		
Posulte	1	2	3	100101	5	6	7	8	0	10	Total
Observed	7	11	7	- 11	9	8	12	10	1.3	12	100
Expected	10	10	10	10	10	10	10	10	10	10	100
How can we determine if the observed counts are a good fit to what we expected based on random variation?											
Observed 5	xpected	ł									
	If we look at the deviations of the values from the values										
If we look at t	he devi	ations o	of the _			value	es from	the			_ values,
If we look at twe can see w	the devi here the	ations o ere are	of the _ large d	liscrepa	ncies.	value If the di	es from e were	the fair, we	e would	d b	_ values,
If we look at t we can see w these differen	the devi here the nces to k	ations o ere are pe close	of the _ large d e to	liscrepa	ncies.	value If the di	es from e were	the fair, we	e would	d b	_ values,
If we look at twe can see w these differen	the devi here the nces to k	ations o ere are pe close	of the _ large d e to	liscrepa 	ncies.	value If the di	es from e were	the fair, we	e would	b	_ values,
If we look at these different	the devi here the nces to k	ations o ere are be close	of the _ large d e to 3	liscrepa · 4	ncies.	value If the di	es from e were 7	the fair, we	e would	d	_ values,
If we look at these different these different conserved	the devi here the nces to k	ations o ere are be close 2 11	of the _ large d e to <u>3</u> 7	liscrepa · 1	ncies. 5 9	value If the di 6 8	es from e were 7 12	the fair, we <u>8</u> 10	e would 9 13	10 12	_ values,
If we look at these different these different conserved – Expected	the devi here the nces to k 1 7 10	ations of ere are be close 2 11 10	of the _ large d e to <u>3</u> 7 10	liscrepa · 11 10	ncies. 5 9 10	value If the di 6 8 10	es from e were 7 12 10	the fair, we 8 10 10	e would 9 13 10	10 12 10	_ values, Total 100 100
If we look at t we can see w these differen Results Observed Expected Obs – Exp	he devi here the nces to k 1 7 10	ations of ere are be close 2 11 10	of the _ large d e to 3 7 10	iscrepa · 11 10	ncies. 5 9 10	value If the di 8 10	es from e were 7 12 10	the fair, we 8 10 10	e would 9 13 10	10 12 10	_ values,
If we look at the we can see we can see we these different the	he devi here the nces to k 1 7 10 a proble	ations of ere are be close 11 10 m: If w	of the _ large d e to 3 7 10 e try to	liscrepa 4 11 10 summa	ncies. 1 9 10 arize th	value If the di 8 10 ese diff	es from e were <u>7</u> 12 10 ference	the fair, we 10 10 es by ta	e would 9 13 10 king th	10 12 10 e sum,	_ values, Total 100 100 this

Observed – E	xpected	Ł									
Let's try a diff	^f erent ap	oproacł	n. What	: about	taking	the					_ of the
differences to	o keep tł	ne value	es			? Tho	ough at	first thi	s seem	s like a	ı wise
approach, the	Э	_ will b	e much			_ even i	if the di	screpa	ncies a	re simi	lar.
Results	1	2	3	4	5	6	7	8	9	10	Total
Observed	7	11	7	11	9	8	12	10	13	12	100
Expected	10	10	10	10	10	10	10	10	10	10	100
Obs – Exp	-3	1	-3	1	-1	-2	2	0	3	2	0
Observed - Expected											
Another method is to . The values will again be .											
but this proce	edures a	similar	proble	em as u	sing the	e absol	ute valı	les bec	ause w	vith a la	irger
			the sun	n will b	e much	larger.					
Results	1	2	3	4	5	6	7	8	9	10	Total
Observed	7	11	7	11	9	8	12	10	13	12	100
Expected	10	10	10	10	10	10	10	10	10	10	100
Obs – Exp	-3	1	-3	1	-1	-2	2	0	3	2	0
(Obs – Exp) ²											
The benefit?	This give	es		diffe	rences	more v	veight i	n their			
to the sum al	lowing ι	is to		whe	en varia	tion m	ay not k	be due			•
Relativity Mat	tters!			. .							
Taking into a	ccount t	he	0	f the sa	imple, v	we can	divide	each of	f the		
differences b	y their _				·						
Results	1	2	3	4	5	6	7	8	9	10	Total
Observed	7	11	7	11	9	8	12	10	13	12	100
Expected	10	10	10	10	10	10	10	10	10	10	100
Obs - Exp	-3	1	-3	1	-1	-2	Z 4	0	3	2	0
$(Obs - Exp)^2$ $(Obs - Exp)^2$	7	1	7	1	1	4	4	0	7	4	42
Exp											
This results in	I		_ that b	etter re	epresen	it the _			O [.]	f the di	fferences
contributed b	y the			values			to	what is	S		·
The Chi-Squa	re Statis	stic									
This final sum	of thes	e	is	what v	ve call t	he				_ or	·
Back to the o	riginal q	luestior	nAre	these r	esults c	onsiste	nt with	what w	ve wou	ld	
in random va	riation?										
Simulating th	e χ^2 Sta	tistic		!	http://ww	w.rossm	anchanc	e.com/a	oplets/20)21/gof/(<u>GOF.html</u>
We will be us	ing the	online s	statistic	al appl	et Anal	yzing C	Dne-way	y Table	s to inv	vestigat	te if the
observed cou	ints are	a		t	o what	we exp	pected	based o	on ranc	lom va	riation.
What Should	We Tak	e Away	?								
How can we d	determir	ne if ob	served	counts	in cate	egorical	data a	re cons	istent v	with ex	pected
counts due to	randor	n variat	ion? W	e can u	ise the	chi-squ	iare sta	tistic to	measu	ure the	distance
between the			_ and _			_ count	s relativ	/e to			counts.
											1

Name_____



AP Statistics CED 8.2 Daily Video 1 (Skill 3.C)

Setting up a Chi-Squared goodness of Fit Test

	area good		55 01110		50			
What Will We Learn?								
What does the chi-square	e statistic mea	asur	e?				•	
How do the degrees of fr	reedom attec	t the	e shape of t	the	chi-square dis	stribution	s?	
Predatory Lending			<i>.</i> .					
Payday loans and title loa	ans are examp	oles	of predato	ory I	lending where	borrows	are left pay	ing interest
rates of 100% or more in	some cases,	mak	king it impo	ssi	ble to pay off	the loans	. Often, the	se
borrowers have few optic	ons due to lov	N Cr	edit scores	or	financial hards	ship. The	"predatory	" label of
these tactics comes from	. Pattern of ta	arge	ting people	e w	ho are low-inc	ome, eld	erly, or who	have little
formal education. Predat	ory lenders a	rgue	e that they	pro	vide a service	and that	people from	n all
backgrounds use their se	rvices, but is	that	true?					
Predatory Lending								
A random sample of 40 p	predatory lend	ding	y businesse	s w	ith Dallas, TX	addresse	s S	
was selected. Their appro	oximate locat	ions	are plotte	d a	s black points	on the m	ap.	
Do these types of busine	sses tend to l	be le	ocated prin	nari	ily in lower inc	ome regi	ons,	
or are they found propor	tionally in reg	gion	s of all inco	me	e levels?			
Observed Counts					Median Ho	usehold Inco	me Numbe	r of Predatory
Household incomes were	e divided into	thre	ee categori	es	\$0 to loss	than \$50.00	Lendi	ng Business
and the number of preda	itory lending	busi	inesses fou	nd	\$50,000 to le	es than \$100	000	17
in census tracts categoriz	ed by the thr	ee i	ncome		\$100.00	0 and above	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	3
brackets was determined								-
Expected Counts						12%	Percent of Dallas, TX Household	is
We will use the assumption	on that the nu	umb	per of busin	ess	es found in	10% 8% 6%	45.2%	25.6%
census tracks correspond	ling to the thr	ree i	income bra	cke	ets is	65		
to :	the number o	of hc	ouseholds in	h th	ne three	or on one same and a same and a same same	and an a star and a star a sta	an and an and an and an and and and and
income brackets.						I & & & & &	భ్ భ్ భ్ భ్ భ్ భ్ Median Household Income	a and a start and a start
Observed & Expected Co	ounts		Median		Number of	Expec	ed Number of	Difference
Then we find the			Household		Predatory Lendin Business	g Preda	atory Lending	(Obs – Exp)
between the observed ar	nd expected		\$0 to less that	an	20		431103303	
counts. Recall that if we f	ind the sum o	of	\$50,000		20			
these differences, we get	t	_!	\$50,000 to le than \$100.00	ss)0	17			
(Note: percentages come	e from the		\$100,000 an	d	3			
graph on the previous sli	de.		above		5			
The Chi-Square	Median		Number of	Ex	pected Number of	Difference	Squared	$(Obs - Exp)^2$
Statistic	Household	Lor	Predatory ding Business	Р	Predatory Lending	(Obs – Exp) Difference	<u>(003 – Exp/</u> Exp
We need something	\$0 to less than	Lei	20		Dusiliesses		(Obs – Exp)	
. so we	\$50,000							
, 30 the	\$50,000 to less than \$100 000		17					
differences to keep all	\$100,000 and		3					
the values	above							
Novt u	ve will compa	nro +	ha siza of t	he	squared differ	ences		
Next, v		ii e l			Squared unler			
	·							







AP Statistics CED 8.2 Daily Video 2 (Skill 1.F)

Setting Up a Chi-Square Goodness of Fit Test						
What Will We Learn?						
How do we state the null hypothesis for a chi-square good	ness-of-fit test?					
How do we state the alternative hypothesis for the chi-squ	are goodness-of-fit test?					
Predatory Lending						
Payday loans and title loans are examples of predatory len	ding where borrows are left paying interest					
rates of 100% or more in some cases, making it impossible	e to pay off the loans. Often, these					
borrowers have few options due to low credit scores or financial hardship. The "predatory" label of						
these tactics comes from. Pattern of targeting people who are low-income, elderly, or who have little						
formal education. Predatory lenders argue that they provide a service and that people from all						
backgrounds use their services, but is that true?						
Predatory Lending						
A random sample of 40 predatory lending businesses with	Dallas, TX addresses					
was selected. Their approximate locations are plotted as b	plack points on the map.					
Do these types of businesses tend to be located primarily	in lower income regions,					
or are they found proportionally in regions of all income le	vels?					
Observed Counts	Median Household Income Number of Predatory					
Household incomes were divided into three categories	\$0 to less than \$50,000 20					
and the number of predatory lending businesses found	\$50,000 to less than \$100,000 17					
in census tracts categorized by the three income	\$100,000 and above 3					
brackets was determined.						
Hypothesized Proportions	Percent of Dallas, TX Households					
We will use the assumption that the number of businesses	tound in 45.2% 25.6%					
census tracks corresponding to the three income brackets is						
census tracks corresponding to the three income brackets	is					
to the number of households in the	is three					
to the number of households in the income brackets.	is three					
to the number of households in the income brackets. Null Hypothesis	is three					
to the number of households in the income brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne	is three ess-of-fit test, we can write it in					
to the number of households in the income brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne or	is three ess-of-fit test, we can write it in					
to the number of households in the income brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne or In words: H ₀ : The distribution of	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the					
	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the pome brackets.					
to the number of households in the tincome brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne or In words: H_0 : The distribution of of households in the specified income In symbols: H_0 : $p_1 =, p_2 =, p_3 = $	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets.					
to the number of households in the fince income brackets income brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne or In words: H_0 : The distribution of of households in the specified income In symbols: H_0 : $p_1 = p_2 = p_3 = $	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets. of predatory lending businesses found in					
to the number of households in the fince income brackets income brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne or In words: H_0 : The distribution of of households in the specified income In symbols: H_0 : $p_1 = p_2 = p_3 = \ where p_1, p_2 and p_3 represent theregions of Dallas, TX where household incomes are$	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets. of predatory lending businesses found in					
	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets. _ of predatory lending businesses found in , respectively.					
to the number of households in the fince income brackets income brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne or In words: H_0 : The distribution of of households in the specified income In symbols: H_0 : $p_1 = p_2 = p_3 = \ where p_1, p_2 and p_3 represent theregions of Dallas, TX where household incomes are and,Alternative Hypothesis$	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets. of predatory lending businesses found in , respectively.					
to the number of households in the tincome brackets income brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne or In words: H_0 : The distribution of of households in the specified income In symbols: H_0 : $p_1 = p_2 = p_3 = p_3 = p_3 = p_3 = p_4$ regions of Dallas, TX where household incomes are, and, Alternative Hypothesis Recall that we need a form of in the	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets. _ of predatory lending businesses found in , respectively. alternative hypothesis. We have					
	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets. of predatory lending businesses found in , respectively. alternative hypothesis. We have m categories that cover possible					
to the number of households in the trice income brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne or In words: H_0 : The distribution of	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets. of predatory lending businesses found in , respectively. alternative hypothesis. We have m categories that cover possible from its hypothesized proportion,					
to the number of households in the tincome brackets to the number of households in the tincome brackets. Null Hypothesis When we write the null hypothesis for a chi-square goodne or In words: H_0 : The distribution of of households in the specified income In symbols: H_0 : $p_1 = p_2 = p_3 = p_3 = p_3 = p_3 = p_4$ regions of Dallas, TX where household incomes are and, Alternative Hypothesis Recall that we need a form of in the proportions to consider. Since these proportions come fro outcomes, their sum will be If will a	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets. of predatory lending businesses found in respectively. alternative hypothesis. We have m categories that cover possible from its hypothesized proportion, also be different since all the proportions					
	is three ess-of-fit test, we can write it in in Dallas, TX is the same as the ome brackets. _ of predatory lending businesses found in , respectively. alternative hypothesis. We have m categories that cover possible from its hypothesized proportion, also be different since all the proportions					



Name
Alternative Hypothesis
You should state the alternative hypothesis in a way that suggests the proportions in
the null hypothesis are different. $\times H_a: p_1 \neq .452, p_2 \neq .292, p_3 \neq .256$
We are testing for, so given convincing evidence that at
proportion is different from what is stated in the null hypothesis is to show the
data do not Likewise, you should state the alternative hypothesis with
directional like < or > that we used for significance testing in previous units.
Remember: State the in!!
Summing Up Hypotheses
For hypotheses about a distribution of proportions for a categorical variable:
The null hypothesis is a statement of where the proportions are all equal
to specified values.
The alternative hypothesis states that on proportion is not as specified in
the null hypothesis.
Never refer to proportions (such as) in the hypotheses!
Remember to define any or you use.
Battleship Quadrants
Are certain quadrants on the Battleship game field preferred by players as strategic locations for their ships? A
random sample of 100 people who enjoy the game Battleship were surveyed. They were asked to place their
five ships on a Battleship game field as if they were going to begin playing against a worthy opponent. The
quadrant of the game field with the greatest number of spaced occupied by ships was recorded. Ties were
broken by identifying the greatest number of ships found in or near a quadrant. Quadrant 1 was defined as the
northwest corner (containing space A1) and subsequent quadrants were identified in clockwise order.
Battleship Quadrants Quadrant 1 2 3 4 4 6 7 8 9 10 A
The observed counts for each quadrant are found in Observed 16 22 33 29
the table below. Is there convincing evidence that
players have a preference for certain quadrants of the Battleship game field? State
the hypotheses for a chi-square goodness-of-fit test.
Null Hypothesis
We are not given a specified distribution to assume for the proportion of players that prefer each of
the four quadrants. Instead, if players have for where to place their
ships, we would to see all the quadrants represented.
H ₀ : The of Battleship quadrants preferred by players is the same across
quadrants, OR H_0 : $p_1 = p_2 = p_3 = p_4 = 0.25$ where p_1 , p_2 , p_3 , and p_4
are the of players that place the most ships in Battleship quadrant 1, 2, 3, and 4.
Alternative Hypothesis (Remember to write this in WORDS!)
H _a : The distribution of Battleship quadrants preferred by players is not the same across
quadrants OR H_a : of the proportions is not as specified in the hypothesis.
What Should We Take Away?
How do we state the null hypothesis for a chi-square goodness-of-fit test? The null hypothesis states
that the proportions for the categories in a categorical variable are
to specified values.
How do we state the alternative hypothesis for the chi-square goodness-of-fit test? In,
state that at of the proportions is not as specified in the hypothesis.



AP Statistics CED 8.2 Daily Video 3 (Skill 4.C)

What Will We Learn?						
How do we identify an appropriate significance test procee	dure for a distribution of	proportions for				
one categorical variable?						
How do we check the conditions for performing a significance test for a distribution of proportions						
for one categorical variable?						
Predatory Lending						
Payday loans and title loans are examples of predatory len	ding where borrows are l	eft paying interest				
rates of 100% or more in some cases, making it impossible	to pay off the loans. Oft	en, these				
borrowers have few options due to low credit scores or fin	ancial hardship. The "pre	datory" label of				
these tactics comes from. Pattern of targeting people who	are low-income, elderly,	or who have little				
formal education. Predatory lenders argue that they provid	le a service and that peop	ple from all				
backgrounds use their services, but is that true?						
Predatory Lending						
A random sample of 40 predatory lending businesses with	Dallas, TX addresses					
was selected. Their approximate locations are plotted as b	lack points on the map.					
Do these types of businesses tend to be located primarily	in lower income regions					
or are they found proportionally in regions of all income le	vels?					
Observed Counts	Median Household Income	Number of Predatory				
Household incomes were divided into three categories	\$0 to less than \$50 000	20				
and the number of predatory lending businesses found	\$50,000 to less than \$100,000	17				
in census tracts categorized by the three income	\$100.000 and above	3				
brackets was determined.		-				
Hypothesized Proportions	12% Percent o	of Dallas, TX Households				
We will use the assumption that the number of businesses	found in	25.6%				
census tracks corresponding to the three income brackets	is ¹⁵					
to the number of households in the t	hree	St and a start and a start and a start a s				
income brackets.	್ ಸ್ಟೇರ್ ವರ್ಷ್ ಪ್ರದೇಶ್ ಪ್ರದೇಶ್ Mediai	o ^{be} ge ^{ge} ge ^{ge} g ^{igte} g ^{igte} g ^{igter} g ^{igter} g ^{igter} g ^{igter}				
Hypotheses						
$H_0: p_1 = _, p_2 = _, p_3 = _$						
where p_1 , p_2 and p_3 represent the	of predatory lending bu	sinesses found in				
regions of Dallas, TX where household incomes are	, , , ,	,				
and ,	respectively.					
H _a : of the proportions is	as specified in the	hypothesis.				
Identifying the Procedure	I					
What type of data were collected? Business locations were	e recorded and	across				
three of variable: median hou	isehold income bracket.	0.0.000				
How many groups? We have sample	e of predatory ler	ndina husinesses				
What are we asked to do? Determine if there is		that				
predatory landing businesses are	aions of					
	gioris OI					
Chi Sauara Gaadaaaa a	f Eit Tost					



Checking the Conditions To check for independence 1. The data should come from a
To check for independence 1. The data should come from a sample OR a experiment. 2. When sampling, the sample should be less than or equal to of the respective population. The chi-square goodness-of-fit test becomes with more observations so should be greater than Checking the Conditions (Be sure to ✓ your conditions!) 1. A predatory businesses was selected. 2. We will assume that lending businesses is less then or equal to of lending businesses in Dallas, TX. To verify that the expected counts are all, we need the proportions from the hypothesis: $p_1 =, p_2 =, p_3 = Expected Counts Median Number of Predatory We find the expected counts by multiplying the Median Number of Predatory $
1. The data should come from a
 2. When sampling, the sample should be less than or equal to of the respective population. The chi-square goodness-of-fit test becomes with more observations so should be greater than Checking the Conditions (Be sure to ✓ your conditions!) 1. A predatory businesses was selected. 2. We will assume that lending businesses is less then or equal to of lending businesses in Dallas, TX. To verify that the expected counts are all, we need the proportions from the hypothesis: p₁ =, p₂ =, p₃ = Expected Counts We find the expected counts by multiplying the
of the respective population. The chi-square goodness-of-fit test becomes
The chi-square goodness-of-fit test becomes with more observations so should be greater than Checking the Conditions (Be sure to \checkmark your conditions!) 1. A predatory businesses was selected. 2. We will assume that lending businesses is less then or equal to of lending businesses in Dallas, TX. To verify that the expected counts are all, we need the proportions from the hypothesis: $p_1 = p_2 = p_3 = \$ Expected Counts We find the expected counts by multiplying the Heatory Lending
Ime chi-square goodness-of-int test becomes
So
Checking the Conditions (Be sure to \checkmark your conditions!) 1. A predatory businesses was selected. 2. We will assume that lending businesses is less then or equal to of lending businesses in Dallas, TX. To verify that the expected counts are all, we need the proportions from the hypothesis: $p_1 = p_2 = p_3 = \ Expected Counts Median Number of Predatory We find the expected counts by multiplying the Median Number of Predatory $
1. A predatory businesses was selected. 2. We will assume that lending businesses is less then or equal to of lending businesses in Dallas, TX. To verify that the expected counts are all, we need the proportions from the hypothesis: $p_1 = p_2 = p_3 = \ Expected Counts Median Number of Predatory We find the expected counts by multiplying the Median Number of Predatory $
2. We will assume that lending businesses is less then or equal to of lending businesses in Dallas, TX. To verify that the expected counts are all, we need the proportions from the hypothesis: $p_1 =, p_2 =, p_3 = Expected Counts Median Number of Predatory Expected Number of Predatory Lending $
of lending businesses in Dallas, TX. To verify that the expected counts are all, we need the proportions from the hypothesis: $p_1 = p_2 = p_3 = \ Expected Counts Median Number of Predatory Expected Number of Predatory Lending We find the expected counts by multiplying the Median Number of Predatory Predatory Lending $
To verify that the expected counts are all, we need the proportions from the
To verify that the expected counts are all, we need the proportions from the
Expected Counts Median Number of Expected Number of We find the expected counts by multiplying the Household Predatory Predatory
We find the expected counts by multiplying the Household Predatory Predatory Lending
proportions of each category from the null hypothesis
with the sample size. (Complete table!)
\$50,000 to less 17
3 All expected \$100,000
above
Battleship Quadrants
Are certain guadrants on the Battleship game field preferred by players as strategic locations for thei
ships? A random sample of 100 people who enjoy the game Battleship were surveyed. They were
asked to place their five ships on a Battleship game field as if they were going to begin playing
against a worthy opponent. The guadrant of the game field with the greatest number of spaced
against a worthy opponent. The quadrant of the game field with the greatest number of spaced
is an peak of supervision of supervision of the souther particular to the souther of supervision
In or near a quadrant. Quadrant I was defined as the northwest corner (containing space AI) and
subsequent quadrants were identified in clockwise order.
Battleship Quadrants Quadrant 1 2 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
The observed counts for each quadrant are found in Observed 16 22 33 29 C COURT COUR
the table below. Is there convincing evidence that
players have a preference for certain quadrants of the Battleship game field? Identify
the procedure and verify that the conditions for inference have been met.
Hypotheses
H_0 : $p_1 = p_2 = p_3 = p_4 = $; where p_1 , p_2 , p_3 , and p_4 are the of players
that place the most ships in Battleship quadrant 1, 2, 3, and 4.
H: of the proportions is as specified in the hypothesis
Identifying the Procedure
What time of data ware callected? Decide ware recorded as
what type of data were collected? People were recorded as counts across
of: which of the four quadrants had the most ships.
How many groups? We have sample of people who enjoy playing Battleship.
What are we asked to do? Determine if there is
that people prefer to place their ships in certain quadrants of the Battleship game field.
Chi-Square Goodness-of-Fit Test



	Na	ame_					
Checking the Conditions (Be sure to ✓ your conditions!)							
1. A of	who enjoy	/ play	ing th	ne gar	ne Ba	attleship was	
selected.			-	-			
2. There are at least = people w	vho enjoy p	olayin	g the	game	e Batt	leship.	
		2	0	U		·	
Battleship Quadrants (Be sure to ✓ your conditions!)						
To verify that the expected counts are all			_, we	need	the	A A A A A A A A A A A A A A A A A A A	
proportions stated in the hypothesis.	Quadrant	1	2	3	4	C Quadrant 1 Quadrant 2 D	
$H_0: p_1 = p_2 = p_3 = p_4 = $	Observed	16	22	33	29		
3. All are	Expected	25	25	25	25	H Quadrant 4 Quadrant 3	
						© Brisget Matamoros	
What Should We Take Away?							
How do we identify an appropriate significance test	procedure	e for a	n distr	ibutic	n of p	proportions for	
one categorical variable?							
Use a							
How do we check the conditions for performing a significance test for a distribution of proportions							
for one categorical variable?							
1. The data should come from a	or a					experiment.	
2. When samples, the sa	ample shou	ıld be)				
of the respective population.							
3. All should be			 •				



AP Statistics CED 8.3 Daily Video 1 (Skill 3.E)

Carrying Out a Chi-Square Goodness-of-Fit Test

What Will We Learn?

How do we calculate a test statistic for a chi-square goodness-of-fit test?

How do we calculate a *p*-value for a chi-square goodness-of-fit test?

Predatory Lending Recap

Payday loans and title loans are examples of predatory lending where borrows are left paying interest rates of 100% or more in some cases, making it impossible to pay off the loans. Often, these borrowers have few options due to low credit scores or financial hardship. The "predatory" label of these tactics comes from. Pattern of targeting people who are low-income, elderly, or who have little formal education. Predatory lenders argue that they provide a service and that people from all backgrounds use their services, but is that true?

Predatory Lending Recap

A random sample of 40 predatory lending businesses with Dallas, TX addresses was selected. Their approximate locations are plotted as black points on the map. Do these types of businesses tend to be located primarily in lower income regions, or are they found proportionally in regions of all income levels?



Observed Counts

Household incomes were divided into three categories and the number of predatory lending businesses found in census tracts categorized by the three income brackets was determined.

Median Household Income	Number of Predatory Lending Business
\$0 to less than \$50,000	20
\$50,000 to less than \$100,000	17
\$100,000 and above	3

Hypotheses and Conditions

Recall from the previous video that the hypotheses for this chi-square goodness-of-fit test are:

*H*₀: $p_1 =$ _____, $p_2 =$ _____, $p_3 =$ _____; where p_1 , p_2 and p_3 represent the

	of predatory lending businesses found in regions of Dallas, ⁻	TX where household
incomes are _		and

_____, respectively.

H _a :	is not	as s	specified	in the
------------------	--------	------	-----------	--------

Additionally, the conditions for inference have _____

Our next step is to proceed with the mechanics of the test:

calculating the _____ and the _

Calculating the Test Statistic (Complete the table below as you watch the video.)

In the previous video	, we found the expected co	ounts by	the proportions of each
category from the	hypothesis by the _		Next we need to calculate
the	between the	and	counts. To keep the

values ______ as well as give ______ to extreme deviations, we will

the						
differences. Taking into	Median Household	Number of Predatory	Expected Number of Predatory Lending	Difference (<u>Obs</u> – Exp)	Squared Difference	(Obs – Exp) ² Exp
account the of	Income	Lending Business	Businesses		(Obs – Exp)	Слр
	\$0 to less than	20				
the sample, we will find	\$50,000					
the of the	\$50,000 to less	17				
	than \$100,000					
differences	\$100,000 and	3				
and the	above					



I	Name				
Calculating the Test Statistic	$(Obs - Exp)^2$				
Lastly, we will find the of these ratios.	Ехр				
	0.20389				
$\nu^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{(\text{Observed} - \text{Expected})^2} =$	2.42315				
λ Δ Expected					
	5.11891				
Calculating the <i>p</i> -value					
We can use technology to calculate the <i>p</i> -value. Remember :	df = number of categories - 1				
NORMAL FLORT AUTO REAL RADIAN HP NORMAL FLORT AUTO REAL RADIAN HP NORMAL FLORT AUTO REAL RADIAN HP 11 12 13 14 12 12 11.60	NORHAL FLOAT AUTO REAL RADIAN HP NORHAL FLOAT AUTO REAL RADIAN HP x2=7.74595074 NORHAL FLOAT AUTO REAL RADIAN HP x0=0.0207964003 If=2 CNTRB={0.2038938053 2.42				
Calculating the <i>p</i> -value					
You can use the χ^2 cdf function on the calculator					
Image: state Morrial FLOAT AUTO REAL RADIAN HP Morrial FLOAT AUTO REAL RADIAN HP Morrial FLOAT AUTO REAL RADIAN HP DISTE DRAM X2 cdf(7,746,1000.2) X2 cdf(7,746,1000.2) Morrial FLOAT AUTO REAL RADIAN HP Morrial FLOAT AUTO REAL RADIAN HP 1: normal Lcdf(3: invNorm(4: invT(5: tcdf(7: X2 cdf(7: X2 cdf(94 Prdf(94 Prdf(Morrial FLOAT AUTO REAL RADIAN HP Morrial FLOAT AUTO REAL RADIAN HP	χ^2 cdf (lowerbound, upperbound, df) Alternatively, you could use Table C. (If this is your choice, follow the video.)				
Why the Upper Tail? As we move further from zero, this implies the					
Why the Upper Tail? As we move the second se	further from zero, this implies the				
Why the Upper Tail? As we move for the observed and A chi-square statistic of means the observed and	further from zero, this implies the between the observed and expected counts are				
Why the Upper Tail? As we move for the observed and expected counts are	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail? As we move for the observed and expected counts are (A perfect fit!). Image: Count of the observed and expected counts are Battleship Quadrants Image: Count of the observed and expected counts are	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail? As we move for the square statistic of means the observed and expected counts are (A perfect fit!). As we move for the square statistic of means the observed and Battleship Quadrants As we move for the square statistic of As we move for the square statistic of A chi-square statistic of Image: square statistic of Image: square statistic of Image: square statistic of Battleship Quadrants Image: square statistic of square statistic on the Battleship game field preferred Image: square statistic of squ	further from zero, this implies the between the observed and expected counts are which means the is 				
Why the Upper Tail? As we move for the square statistic of means the observed and expected counts are (A perfect fit!). As we move for the square statistic of means the observed and Battleship Quadrants As we move for the square statistic of As we move for the square statistic of Battleship Quadrants Are certain quadrants on the Battleship game field preferred ships? A random sample of 100 people who enjoy the game	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail? As we move for the square statistic of means the observed and expected counts are As we move for the square statistic of (A perfect fit!). Image: square statistic of Image: square statistic of Battleship Quadrants Image: square statistic of statis statistic of statistic of statistic of stati	further from zero, this implies the between the observed and expected counts are which means the is I by players as strategic locations for their Battleship were surveyed. They were f they were going to begin playing against				
Why the Upper Tail? As we move for the square statistic of means the observed and expected counts are As we move for the square statistic of (A perfect fit!). Image: State statistic of Image: State statistic of Battleship Quadrants Image: State stat	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail? As we move for the second s	further from zero, this implies the between the observed and expected counts are which means the is I by players as strategic locations for their Battleship were surveyed. They were f they were going to begin playing against greatest number of spaced occupied by atest number of ships found in or near a				
Why the Upper Tail? As we move for the served and expected counts are (A perfect fit!). Image: Comparison of the served and expected counts are Battleship Quadrants Image: Comparison of the served and expected counts are Battleship Quadrants Image: Comparison of the served and expected counts are Battleship Quadrants Image: Comparison of the served and expected counts are Battleship Quadrants Image: Comparison of the served and expected counts are Battleship Quadrants Image: Comparison of the served and expected counts are Battleship Quadrants Image: Comparison of the served and expected counts are Battleship Quadrants Image: Comparison of the served and expected counts are	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail? As we move for the second s	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail? As we move for the served and expected counts are	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail? As we move for the second s	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail? As we move for the second counts are A chi-square statistic of means the observed and expected counts are	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail?As we move for the second counts are the observed and expected counts are the observed and expected counts are the observed and expected counts are the observed and the second count of the second count	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail?As we move for the spectral counts areA chi-square statistic of means the observed and expected counts are(A perfect fit!)Battleship QuadrantsAre certain quadrants on the Battleship game field preferred ships? A random sample of 100 people who enjoy the game asked to place their five ships on a Battleship game field as if a worthy opponent. The quadrant of the game field with the ships was recorded. Ties were broken by identifying the great quadrant. Quadrant 1 was defined as the northwest corner (or quadrants were identified in clockwise order.Battleship QuadrantsThe observed counts for each quadrant are found in the table video and use the space below to calculate the test-statisticQuadrant 1 2 3 4 29 $\chi^2 = \Sigma \frac{(Observed)}{16}$	further from zero, this implies the between the observed and expected counts are which means the is				
Why the Upper Tail? As we move for a statistic of means the observed and expected counts are (A perfect fit!). As we move for a statistic of means the observed and Battleship Quadrants Are certain quadrants on the Battleship game field preferred ships? A random sample of 100 people who enjoy the game asked to place their five ships on a Battleship game field as if a worthy opponent. The quadrant of the game field with the ships was recorded. Ties were broken by identifying the great quadrant. Quadrant 1 was defined as the northwest corner (a quadrants were identified in clockwise order. Battleship Quadrants The observed counts for each quadrant are found in the table video and use the space below to calculate the test-statistic Quadrant 1 2 3 4 2 3 2 2 Quadrant 1 2 3 4 2 3 2 2	further from zero, this implies the between the observed and expected counts are which means the is				



Multiple-Choice Example

Which of the following gives the correct test statistic and *p*-value for a chi-square goodness-of-fit test to determine if there is convincing evidence that the players have a preference for certain quadrants of the Battleship game field? (Eliminate choices as you watch the video!)

(A)
$$\chi^2 = \frac{(25-16)}{16} + \frac{(25-22)}{22} + \frac{(25-33)}{33} + \frac{(25-29)}{29}$$
; *p*-value = 0.957
(B) $\chi^2 = \frac{(16-25)}{25} + \frac{(22-25)}{25} + \frac{(33-25)}{25} + \frac{(29-25)}{25}$; *p*-value = 1.0
(C) $\chi^2 = \frac{(16-25)^2}{25} + \frac{(22-25)^2}{25} + \frac{(33-25)^2}{25} + \frac{(29-25)^2}{25}$; *p*-value = 0.147
(D) $\chi^2 = \frac{(16-25)^2}{25} + \frac{(22-25)^2}{25} + \frac{(33-25)^2}{25} + \frac{(29-25)^2}{25}$; *p*-value = 0.079
(E) $\chi^2 = \frac{(16-25)^2}{100} + \frac{(22-25)^2}{100} + \frac{(33-25)^2}{100} + \frac{(29-25)^2}{100}$; *p*-value = 0.637
Multiple-Choice Example

$$\frac{115}{25} + \frac{125}{25} + \frac{122}{25} + \frac{123}{25} + \frac{129}{25} + \frac{129}{2$$



	Name						
AP Statistics CED 8.3 Daily Video 2							
Carrying Out a Chi-Square Goodness-of-Fit Tes	st						
What Will We Learn?							
How do we interpret the <i>p</i> -value for a chi-square goodness	s-of-fit test?						
How so we state a conclusion for a chi-square goodness-of	f-fit test?						
Predatory Lending Recap							
Payday loans and title loans are examples of predatory len-	ding where borrows are l	eft paying interest					
rates of 100% or more in some cases, making it impossible	to pay off the loans. Oft	en, these					
borrowers have few options due to low credit scores or fina	ancial hardship. The "pre	datory" label of					
these tactics comes from. Pattern of targeting people who	are low-income, elderly,	or who have little					
formal education. Predatory lenders argue that they provide a service and that people from all							
backgrounds use their services, but is that true?							
Predatory Lending Recap							
A random sample of 40 predatory lending businesses with	Dallas, TX addresses						
was selected. Their approximate locations are plotted as b	lack points on the map.						
Do these types of businesses tend to be located primarily i	in lower income regions,						
or are they found proportionally in regions of all income le	vels?						
Observed Counts	Median Household Income	Number of Predatory					
Household incomes were divided into three categories	¢0 to loss their ¢50,000						
and the number of predatory lending businesses found	\$0 to less than \$50,000	20					
in census tracts categorized by the three income							
brackets was determined.							
Predatory Lending Recap							
Hypotheses: H ₀ : p ₁ = .452, p ₂ = .292, p ₃ = .256; where p ₁ ,	p_2 and p_3 represent the p	proportions of					
predatory lending businesses found in regions of Dallas, T	X where household incor	nes are \$0 to less					
than \$50,000, \$50,000 to less than \$100,000 and \$100,000) and above, respectively	·					
U. At least one of the preparience is not as an aritical in th	a null hypothesis						

 H_a : At least one of the proportions is not as specified in the null hypothesis.

Procedures & Conditions: The conditions for a chi-square goodness-of-fit test have been verified	•
Mechanics: The test statistics is $\chi^2 = 7.746$ and the p-value = 0.0208	

Interpreting a *p*-value

Recall from the	e previous video that the <i>p</i> -value is calculated b	y finding the tail probability
of a chi-square	e distribution. What exactly does this describe?	From Units 6 and 7, we learned that
the	of obtaining a result as	the one in the study,
or	by chance alone,	the null hypothesis is

Interpreting a p-value

There is a of getting a chi-square statistic of j							just by the		
chance involved in the selection of the businesses, that the distribution of predatory l		Median Household Income	Number of Predatory Lending Businesses	Expected Number of Predatory Lending Businesses	Obs – Exp	(Obs – Exp) ²	$\frac{(Obs-Exp)^2}{Exp}$		
businesses in Dallas TX is the	lending	\$0 to less than \$50,000	20	18.08	1.92	3.6864	0.20389		
as the proportions of households	in the	\$50,000 to less than \$100,000	17	11.68	5.32	28.3024	2.42315		
specific		\$100,000 and above	3	10.24	-7.24	52.4176	5.11891		
specific	•						$\chi^2 = 7.746$		



Carrying Out a Chi-S

What Will We Learn?

Predatory Lending Recap

Predatory Lending Recap



		Nar	me					
Stating a Conclusion: A General Gu	uide							
When we state our conclusion, it ha	as main	parts:						
1. How does the <i>p</i> -value to our level of significance, α , and what must be								
make about H_0 ?								
2. What does this mean about H_a in	۱	?						
For small p -values \longrightarrow test statist	ics is	to occu	r by random	I		·		
Since the <i>p</i> -value of $___ \le \alpha$	z =, we _	Ho).					
There is	statistical e	vidence that [state H_{a} in _]			
For large p -values \longrightarrow test statist	ics is	to occu	r by random	I		·		
Since the <i>p</i> -value of > α	z =, we _		H_0 .					
There is	statistical e	vidence that [state H_{a} in _]			
Stating a Conclusion								
N	Since the <i>p</i> -val	ue of	is		_ α =	, we		
	th	е <i>H</i> ₀.						
	There co	nvincing						
$\chi^{2} = 7.746$	that the distrib	ution of pred	atory lendin	g busine	esses in E	Dallas, TX		
p-value = 0.0208	is	_ as the prop	ortions of ho	buseholo	ds in the	specified		
The problem did not state a significance level, so will use $\alpha = 0.05$.	income bracke	ts.						
Contributions								
Let's return to the original question	: Is there convi	ncing statistic	al evidence	that the	ese types	of		
businesses tend to be located prim	arily in lower in	businesses tend to be leasted primarily in lower income regions? We have determined that the						
businesses tend to be located primarily in lower income regions? We have determined that the								
proportions of predatory lending b	usinesses found	d in regions o	s? we nave f different ir	determi icome b	ned that rackets a	the re not		
proportions of predatory lending b the same as the proportions of hou	usinesses found seholds in the	d in regions o pecified inco	f different ir me bracket	determi icome b s.	ned that rackets a	the re not		
proportions of predatory lending by the same as the proportions of hou — Are the located prin	usinesses found seholds in the s marily in lower	d in regions o specified incc income regio	f different ir me bracket ns?	determi icome b s.	ned that rackets a	the re not		
proportions of predatory lending by the same as the proportions of hou Are the located print Contributions	usinesses found seholds in the marily in lower	d in regions o specified inco	f different ir me bracket ns?	determi icome b s.	ned that rackets a	the re not		
proportions of predatory lending be the same as the proportions of hou → Are the located prin Contributions By looking at the	usinesses found seholds in the marily in lower	in regions o specified inco income regio	f different ir ome bracket ns?	obs - Exp	ned that rackets a	the re not (Obs - Exp) ²		
proportions of predatory lending be the same as the proportions of hou → Are the located print Contributions By looking at the of each toward the chi-	usinesses found seholds in the marily in lower Median Household Income	in regions of specified inco income regio	s? vve nave f different ir ome bracket ns? Expected Number of Predatory Lending Businesses	oetermi ncome b s. Obs-Exp	ned that rackets a (Obs – Exp) ²	the re not (Obs - Exp) ² Exp		
proportions of predatory lending be the same as the proportions of hou → Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine	usinesses found seholds in the marily in lower Median Household Income \$0 to less the \$50,000	Number of Predatory Lending Businesses	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08	Obs - Exp	ned that rackets a (Obs – Exp) ² 3.6864	the re not (<u>Obs - Exp</u>) ² Exp 0.20389		
proportions of predatory lending by the same as the proportions of hou Are the located print Contributions By looking at the of each toward the chi- square statistic, we can determine where the	usinesses found seholds in the marily in lower Median Household Income \$00 to less the \$50,000 \$50,000 to less the s50,000 to	Number of Predatory Lending Businesses 1 20	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68	Obs - Exp 1.92 5.32	ned that rackets a (Obs – Exp) ² 3.6864 28.3024	the re not (<u>Obs - Exp</u>) ² Exp 0.20389 2.42315		
proportions of predatory lending be the same as the proportions of hou → Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i	usinesses found seholds in the marily in lower Median Househole so to less thi \$50,000 to le than \$100,00 \$100,000 ar above	Number of Predatory Lending Businesses	s? vve nave f different ir pme bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24	determi acome b s. 0bs – Exp 1.92 5.32 –7.24	ned that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176	the re not (Obs - Exp) ² Exp 0.20389 2.42315 5.11891		
proportions of predatory lending be the same as the proportions of hou Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i what we observed and what we	usinesses found seholds in the marily in lower Median Household Income \$00 to less th \$50,000 \$50,000 to le than \$100,00 \$100,000 ar above	In regions o specified income regio	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24	Obs - Exp 1.92 5.32 -7.24	ned that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$		
proportions of predatory lending be the same as the proportions of hou Are the located print Contributions By looking at the of each toward the chi- square statistic, we can determine where the ites in what we observed and what we expected. We can see that the	usinesses found seholds in the marily in lower Median Househole Income \$0 to less thi \$50,000 to le than \$100,00 \$100,000 ar above	Number of Predatory Lending Businesses 1 20 2 17 4 3	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 ad the large	Obs - Exp 1.92 5.32 -7.24 est contr	ned that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176 ribution.	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i what we observed and what we expected. We can see that the these data, we can infer that predat	usinesses found seholds in the marily in lower Median Household Income \$00 to less th \$50,000 \$50,000 to le than \$100,00 \$100,000 ar above	h regions o specified inco income regio Predatory Lending Businesses 1 20 3 17 4 3 bracket h sinesses do to	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 add the large	Obs - Exp 1.92 5.32 -7.24 est contr ocated ir	ned that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou → Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the what we observed and what we expected. We can see that the these data, we can infer that predat regions be	usinesses found seholds in the marily in lower Median Househole Income \$0 to less thi \$50,000 to le than \$100,00 \$100,000 ar above tory lending but cause they wer	h regions of specified income regions of specified income regions of the specified income region of the specified income region of the specified income region of the specified income regions of the specified income represented of the specified income regions of the specified income reg	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 ad the large	Obs - Exp 1.92 5.32 -7.24 est contro cated in	ned that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176 ribution.	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i what we observed and what we expected. We can see that the these data, we can infer that predat regions be in the highest income regions.	usinesses found seholds in the marily in lower Median Household Income \$50,000 \$50,000 to less th \$50,000 \$50,000 to less th \$50,000 \$50,000 to less th source than \$100,00 \$100,000 ar above	h regions of specified income regions of specified income regions of the specified income region of the sinesses do the represented of the specific terms of ter	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 add the large end to be lo	Obs - Exp 1.92 5.32 -7.24 est contr ocated ir	ned that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou → Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the what we observed and what we expected. We can see that the these data, we can infer that predat regions be in the highest income regions.	usinesses found seholds in the marily in lower Median Househole Income \$0 to less the \$50,000 to le than \$100,00 \$50,000 to le than \$100,00 \$100,000 ar above	In regions of specified income regions income regio Income regio <td>f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 add the large end to be lo</td> <td>Obs - Exp 1.92 5.32 -7.24 est contr ocated ir</td> <td>ned that rackets a (Obs - Exp)² 3.6864 28.3024 52.4176 ribution.</td> <td>the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From</td>	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 add the large end to be lo	Obs - Exp 1.92 5.32 -7.24 est contr ocated ir	ned that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176 ribution.	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou → Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i what we observed and what we expected. We can see that the these data, we can infer that predat regions be in the highest income regions. Multiple Choice Example	usinesses found seholds in the marily in lower Median Household Income \$00 to less th \$50,000 \$50,000 to le than \$100,00 \$100,000 ar above	h regions of specified income regions of specified income regions of the specified income region of the sinesses do the represented of the specific terms of	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 add the large end to be lo	Obs - Exp 1.92 5.32 -7.24 est contr ocated ir	ned that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou → Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i what we observed and what we expected. We can see that the these data, we can infer that predat regions beg in the highest income regions. Multiple Choice Example Are certain quadrants on the Battle	usinesses found seholds in the marily in lower Median Househole Income \$0 to less thi \$50,000 to le than \$100,00 \$50,000 to le than \$100,00 \$50,000 to le than \$100,00 \$50,000 ar above tory lending bu cause they wer	l in regions o specified inco income regio Predatory Lending Businesses n 20 50 17 d 3 bracket h sinesses do to e represented	f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 add the large end to be lo	obs - Exp 1.92 5.32 -7.24 est contr bcated in strategic	red that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176 ribution.	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i what we observed and what we expected. We can see that the these data, we can infer that predat regions be in the highest income regions. Multiple Choice Example Are certain quadrants on the Battle ships? A random sample of 100 peop	usinesses found seholds in the marily in lower Median Househol S0 to less thi \$50,000 to le than \$100,00 story lending bu cause they wer ship game field ople who enjoy	I preferred by the game Ba	 vve nave f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 ad the large end to be lood to be lo	Obs - Exp 1.92 5.32 -7.24 est contr ocated ir strategic re survey	ned that rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176 ribution.	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou → Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i what we observed and what we expected. We can see that the these data, we can infer that predat regions be in the highest income regions. Multiple Choice Example Are certain quadrants on the Battle ships? A random sample of 100 pee asked to place their five ships on a	usinesses found seholds in the marily in lower Median Household Income Sto less thi S50,000 to le than \$100,00 S50,000 to le than \$100,00 sto less thi S50,000 ar above tory lending bu cause they wer ship game field ople who enjoy Battleship game	I preferred by the game Ba e field as if th	<pre>s? vve nave f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 add the large end to be lo of players as ittleship wer ney were go</pre>	obs - Exp 1.92 5.32 -7.24 est contri ocated in strategic re survey ing to b	rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176 ribution.	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i what we observed and what we expected. We can see that the these data, we can infer that predat regions be in the highest income regions. Multiple Choice Example Are certain quadrants on the Battle ships? A random sample of 100 peo asked to place their five ships on a against a worthy opponent. The qu	usinesses found seholds in the marily in lower Median Household Income S0 to less thi \$50,000 or than \$100,00 \$100,000 ar above tory lending but cause they wer ship game field ople who enjoy Battleship game adrant of the g	I preferred by the game Ba e field as if th ame field witl	 vve nave f different ir pme bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 ad the large end to be lock y players as attleship were powere go the greate 	Obs - Exp 1.92 5.32 -7.24 est contro ocated ir strategic re survey ing to bo est numb	rackets a (Obs - Exp) ² 3.6864 28.3024 52.4176 ribution. The location yed. They egin play ber of spa	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From		
proportions of predatory lending be the same as the proportions of hou Are the located prin Contributions By looking at the of each toward the chi- square statistic, we can determine where the lies i what we observed and what we expected. We can see that the these data, we can infer that predat regions be in the highest income regions. Multiple Choice Example Are certain quadrants on the Battle ships? A random sample of 100 pee asked to place their five ships on a against a worthy opponent. The qu occupied by ships was recorded. Ti	usinesses found seholds in the marily in lower Median Household Income Sto less thi S50,000 to le than \$100,00 S50,000 to le than \$100,00 sto less thi S50,000 ar above story lending but cause they wer ship game field ople who enjoy Battleship gam adrant of the g	I preferred by the game Ba e field as if th ame field with by identifyin	<pre>s? vve nave f different ir ome bracket ns? Expected Number of Predatory Lending Businesses 18.08 11.68 10.24 add the large end to be lo</pre>	obs - Exp 1.92 5.32 -7.24 est contri ocated in cated in strategic re survey ing to bout est numb	(obs - Exp) ² 3.6864 28.3024 52.4176 ribution. They egin play ber of spa	the re not $\frac{(Obs - Exp)^2}{Exp}$ 0.20389 2.42315 5.11891 $\chi^2 = 7.746$ From the for their v were ing inced ips found		

subsequent quadrants were identified in clockwise order.



	Name					
Battleship Quadrants	Quadrant 1 2 3 4 6 7 8 9 10					
The observed counts for each quadrant are found in	Observed 16 22 33 29					
the table below. A chi-square goodness-of-fit test to						
determine whether the proportions of players preferring	g each of the four quadrants					
are the same resulted in a <i>p</i> -value of 0.079.	© Bridget Materiours					
Multipe-Choice Example. (Eliminate answers as you wat	ch the video!)					
Which of the following is a correct interpretation of the p -value? \triangleleft	PAY ATTENTION TO WHAT THE QUESTION IS ASKING!					
(A) Since the <i>p</i> -value of 0.079 is greater than α = .05, we should reject the null hypothesis that the proportions are the same. There is convincing evidence that players prefer certain quadrants.						
(B) Three of the observed counts are not greater than 30, so the co interpretation of the <i>p</i> -value would not be valid.	nditions have not been met. An					
(C) Only 7.9% of those surveyed preferred quadrant 1. This is conv proportions of players preferring different quadrants is not the s	incing evidence that the ame.					
(D) There is a 0.079 probability that we would get results this extreme or more extreme, by chance alone, if the proportions of players preferring different quadrants is the same.						
(E) There is a 0.079 probability that we would get results this extremalone, assuming that players have preferences for different quartering.	ne or more extreme, by chance adrants.					
What Should We Take Away?						
How do we interpret the <i>p</i> -value for a chi-square goodr	ess-of-fit test?					
Assuming the proportions of a categorical v	variable as stated in the hypothesis					
are, there is a < <i>p</i> -value> of	getting a chi-square statistic as					
as the one in the study, I	by chance alone in sampling					
(or assignment).						
How so we state a conclusion for a chi-square goodness	s-of-fit test?					
Since the <i>p</i> -value of $\leq \alpha = $, we	H ₀ .					
L There is statistical e	evidence that [state H_a in]					
Since the <i>p</i> -value of > α =, we	H₀.					
L There is statistical e	evidence that [state <i>H</i> ª in]					



AP Statistics CED 8.3 Daily Video 3 (Skill 4.E)

Carrying Out a Chi-Square Goodness-of-Fit Test

What Will We Learn?

How do we perform a complete significance test for a distribution of proportions for one categorical variable?

2008 International Exam (The first section was omitted from the video, provided here for clarity!) "The department of parks and recreation of a certain city conducts summer programs for residents of its six districts. The summer programs include operating and maintaining community swimming pools in each of the districts as well as offering sports and recreational programs for school-age children, young adults, and older adults. The table above shows the proportion of households by district out of all the households that participated in the summer programs, based on annual data that were collected from simple random samples

each summer over a 10-year period, ending in the year 2000. The proportions are being used by the city for

planning purposes and for more efficiently targeting the introduction of future programs."

City leaders want to test if the proportions that are being used by the city are still valid. Data collected by a statistician from a simple

District	A	в	C	D	E	F
Proportion of Households	0.32	0.12	0.10	0.27	0.05	0.14
District	A	В	С	D	E	F
Number of Households	100	35	40	22	12	31

random sample this past summer indicated that the following **number** of households participated in each district.

(a) The statistician claims that the data for this past summer provide evidence that the proportions that are being used by the city are no longer valid. Give statistical evidence to justify the claim.(b) Which one of the six districts had the greatest change in participation since the year 2000? Use the information from part (a) to explain your choice.

Planning the Response (Highlight the key information above as you watch the video.)

First, this is a ______ test because we are asked to give ______ evidence for a claim. Let's identify the procedure, we have ______ sample and the variable of interest is a ______ variable with _____ categories. This tells us we will need to perform a

Part (a) – Significance test procedures generally require 4 parts:

- •
- _____
- · _____

Part (b) – We need the contributions!

Hypotheses & Parameters							
	District	A	В	С	D	E	F
Here are the current proportions being used by City	Proportion of Households	0.32	0.12	0.10	0.27	0.05	0.14
Leaders that are being challenged by the researchers.							
$H_0: p_A = _, p_B = _, p_C = _, p_D = _, p_D$	e =	_ p _F =		whe	ere p_A	throu	gh p _⊦
are theof households participating in the	e summer	progr	am fo	or		distri	ct.
H _a : of the proportions is different from	those sp	ecified	d in th	ne	ł	nypoth	nesis.
Procedures & Conditions (Be sure to ✓ your conditions!)							
Identification of procedure: We will use a chi-square							•
Conditions:							
The study uses a sample of households.							
 We must assume that households are less 	than or e	qual t	to	_ of a	ll hou	seholo	ds in
the city.		-					

📥 STATS MEDIC

Name										
If we multiply each proportion by the	District	Α	В		:	D	E		F	
of, you will get	Proportion of Households	0.32	0.12	0.1	10	0.27	0.0	5	0.14	
the expected values. You do need to	Expected									
state these on your AP Exam!	Number of Households									
3. All expected counts are										1
Mechanics (Use Technology!) (Note: Pa	ay close attentio	n to the	df: number	of categ	gories -	- 1 or 6 -	1 = 5)			
Enter Lists Select χ^2 GOF-Test MORHAL FLOAT AUTO REAL RADIAN MP NORMAL FLOAT AUTO REAL RADIAN MP NORMAL FLOAT AUTO REAL RADIAN MP 100 Frail 30 28.8 Control 100 Frail 100 Frail RADIAN MP Other State Stat	Assign Data HAL FLOAT AUTO REAL RAD CAGOFFICEL DSERVED:LL SPECTED:L2 F15 DLOT: BLUE alculate Draw		NORMAL FLOAT x2=47.48 p=4.5348 df=5 CNTRB={7	Calculate AUTO REAL RI 23007-Tes 300485 372594e-5 7.0083333	9 ADIAN MP 50 333 1.3	0 Υα #1 #34 χ ρ d	Resu to must re- ese value $^2 = 47.4$ -value = f = 5	ecord all s on the l8 4.53 x	three of AP Exam! 10 ⁻⁹ ≈ 0)
Since the <i>p</i> -value of approximately	IS		$\alpha = _$		_, we	e		_ the		-•
I here convincing	th	at the	e propoi	rtions	ot he	ouseh	olds p	partici	pating	J
in the for each c	district are _		the sar	me as	thos	se beir	ng use	e by tl	ne city	1.
Part (b) Which one of the six districts had the	greatest ch	ange i	in partic	ipatio	n sin	ce the	year 2	2000?	Use	_
the information from part (a) to explain your c	noice.		District Proportion of Households	A 0.32	B 0.12	C 0.10	D 0.27	E 0.05	F 0.14	1
the category that made the largest contribution	and then fill	na –	Expected Number of	76.8	28.8	24	64.8	12	33.6	1
		_	Households Number of	100	35	40	22	12	31	-
		((Obs – Exp) ² Exp	7.0083	1.3347	10.6667	28.2691	0	0.2012	1
Test Tip: List Operations You can calculate these contributions using Using the formula in Lists Image: Contribution of the formula in Lists Image: Contrib	technology HORMAL FLOAT AUTO RE NEINES OPS MATH 1:L1 2:L2 3:L3 4:L4 5:L5 6:L6 FCONTRB 8:RESID in ne chi-squar	/. stat AL RADIAN HP f	From the	stored I 1047 4010 Act 76.8 7.000 76.8 7.000 76.8 7.000 76.8 7.000 76.8 7.000 76.8 7.000 76.8 7.000 10.201	List @ .	2 nd List ¹⁰ <u>s</u> <u>s</u> <u>s</u> <u>s</u> <u>s</u> <u>s</u> <u>s</u> <u>s</u>	ItAL FLOAT AL 10 12.8 24.8 24.8 24.9 33.6 11)=7.0083 10=7.0083 LOAT AUTO RE 10=7.4800483 34872594 = (7.0083)	10 REAL RADZ 1.397 1.397 1.397 1.397 1.397 1.397 0.2012 0.20 0.2012 0.2012 0.20 0.2012 0.2012 0.20 0.2012 0.2012 0.20 0.2012 0.	AN 110 1 15 1 17 17 17 17 17 17 17 17 17 1	2
Scoring Guidelines										
Section 1: State a correct pair of	and c	heck	the					·		
Section 2: Show the correct mechanics inclu	uding the va	alue o	f the			/	, a	nd		_
Section 3: State a correct in _		_ usin	g the _			_ of th	e test			
Section 4: Identify and show supporting		foi	r the sel	lection	n of t	the co	rrect	distric	:t (D).	



m	е
	m

What Should We Take Away?							
How do we perform a complete significance test for a distribution of proportions for one categorical							
variable?							
Be sure to:							
Define any	_ and used.						
State the	Always specify value for the in each category.						
Identify the	you are using.						
Verify that the	for the procedure have been						
Calculate the	,, and						
Interpret the	in>						



Name

AP Statistics CED 8.4 Daily Video 1

Expected Counts in Two-Way Tables

What Will We Learn?					
How do we calculate the expected counts invo	lving two	-way tables o	of categor	rical data?	
A Look Forward					
In Topics 8.4 – 8.6 we will learn new typ	bes of chi-	square signif	icance te	sts.	
For both tests, we will be analyzing data in a				Variabl	e 1
			Category A	Category 1 Categor	y 2 Category 3
A two-way table data for	two	Variable 2	Category B		
variables.			Category C		© Luke Wilcox
Where Do You Go to School?					
A random sample of parents with school-aged				2010 Comple	2020 Sample
children was taken during 2019. A separate rar	ndom	Public		2019 Sample	2020 Sample
sample of parents with school-aged children w	as	Public Deixets (Deve els		200	105
taken in 2020. Parents were asked what type o	f school	Private/Paroch	ial/Charter	16	21
their children attended. Here are the results	1 301001	ноте		30	30
The variable of the represented in the	e column	s and the var	iable of		
is represented in the rows		s and the var			
Is represented in the rows.					
Expected Counts	to al	-			
Expected counts in a two-way table are calcula	ited	th	iere is no		
between the two categorical variables.					
I here are three steps to totaling a two-way			2019 Samp	ole 2020 Sam	ple Total
table:	Public		266	163	
> Find row totals	Private/	Parochial/Charter	16	21	
→ Find column totals	Home		38	30	
→ Find the table total	Total				
* Overall proportion that attended public scho * 80.34% of the 2019 sample should attend pu * expected count (public 2019) = ()()	ol = Iblic scho) = _	= ol, or (0.8034	(So)(320) = _	, assuming no	> relationship!)
General formula: expected	d count =	()())	
Expected Counts (Use the General	Formula	rom above!)			
Expected count (Private/Parochial/Charter, 201	9) =	=			
Expected count (Home, 2019) =					
Expected count (Public, 2020) =					
Expected count (Private/Parochial/Charter 202	20) =				
Expected count (Home $2020) =$					
			2010 5	ala 2020 S	
(Note: Make sure you include a key to			2019 Sam		
indicate the number in () are expected	Public Private (D	e chiel/Charter	266_(257.	1) 163	429
counte)	Home	ocniai/Charter	38	30	<u> </u>
counts.)	Total		320	214	534
	10001		220	2 17	



Are you Working?

A random sample of 2000 adults revealed the following data about education level and employment status.

is the categorical variable represented by rows.

_ is the categorical variable represented by columns.

Calculate the expected counts:

- Start by finding the row totals, column totals and the table total. (Add lines!)
- Use the formula to calculate expected counts for each cell. (Add these values to the table.)
- When possible, use subtraction to find the remaining expected counts.
- Don't forget the Key!

	No High School Diploma	High School Diploma, No College	High School Diploma, Some College
Employed	206	548	1186
Unemployed	14	22	24

What Should We Take Away?

How do we calculate the expected counts involving two-way tables of categorical data?

expected count = $\frac{(row total)(column total)}{table total}$

Once we have calculated enough of these expected counts, use the row totals and column totals and subtraction to find the others.



AP Statistics CED 8.5 Daily Video 1 (Skill 1.F)

Setting up a Chi-Square Test for Homogeneity or Independence

		. 1	
How do we identify the appropriate significance test pro	ocedure for data in a tw	wo-way tab	le?
How do we state a null and alternative hypothesis for a	chi-square test for hon	nogeneity?	
How do we state a null and alternative hypothesis for a	chi-square test for inde	ependence	?
Identifying the Procedure			
There are types of significance tests for	da	ata in a two	-way table.
• Goal: to distributions of a categori	cal variable for	ро	pulations
Data collection: random sa	mples from	popu	ations (or
from groups in a randomized exp	eriment)		
\rightarrow chi-square test for	·		
Goal: to determine whether two categorical variations of the second	ables are		
Data collection: random sample from	n population	n	
\rightarrow chi-square test for			
The appropriate test for data in a two-way table	depends on		
Where Do You Go To School?			
A random sample of parents with school-aged			
children was taken during 2019. A separate random		2019 Sample	2020 Sample
comple of parents with school agod children was	Public	266	163
taken in 2020. Paranta ware saked what type of school	Private/Parochial/Charter	16	21
their children attended. Here are the regults	Home	38	30
their children attended. Here are the results.	a af a ab a al far tha two	waar whield	_
in we are interested in comparing the distribution of type	e of school for the two	year, which	1
	h.,		
Goal: The goal is to compare the distribution of	between	the two ye	ars
Data collection: independent random samples	s from differer	it populatio	ns
I herefore, the is a chi-square test for	·		
Null Hypothesis			
In a statistical test, the hypothesis is often a claim	ot	or	•
H_0 : There is in the distribution	on of		attended
by school-aged children from 2019 to 2020. (Notice tha	t there is no paramete	er to define	.)
Alternative Hypothesis			
In a statistical test, the hypothesis is t	he that w	e hope to s	upport with
from the data collected.			
H ₀ : There is in the distribution	on of		attended
by school-aged children from 2019 to 2020.			
H_{a} : There is in the distribution	on of		attended
by school-aged children from 2019 to 2020.			
Because the alternative hypothesis includes several prog	portions that could diff	fer in either	direction,
this test is said to be			· · · /



Are you Working?				
A random sample of 2000 adults revealed the following	data abou	t education	level and e	employment
status.		No High School Diploma	High School Diploma, No College	High School Diploma, Some College
If we are interested in deciding if education level and	Employed	206	548	1186
employment status are associated, what is the	Unemployed	14	22	24
appropriate significance test?				
Goal: To see if there is an betwee	en two cate	egorical varia	bles.	
Data collection: data collected from rar	ndom samp	ole from	_ populat	ion. Also,
note the work Therefore, this is a cl	ni-square te	est for		·
Null Hypothesis. (Remember, no difference or no chang	je!)			
H ₀ : There is between educ	ation level	and employ	ment statu	is for
OR				
H_0 : Education level and employment status are		fo	or all adult	ts.
Alternative Hypothesis. (Claim we hope to support with	evidence	from data co	llected.)	
H ₀ : There is between educati	on level an	d employme	nt status f	or
·				
H₄: There is between educati	on level an	d employme	nt status f	or
. OR		l J		
H_0 : Education level and employment status are			for all a	dults.
H_a : Education level and employment status are			for all a	dults.
What Should We Take Away?				
How do we identify the appropriate significance test pr	ocedure fo	r data in a tw	vo-way tab	ole?
categorical variable, populations -> chi-squ	ıare test fo	or	,	
categorical variables, population \rightarrow chi-squ	lare test fo			
How do we state a null and alternative hypothesis for a	chi-square	test for hom	oaeneitv?	
Ho: There is in the distributio	n of í]	across
populations or treatments.	- L		.	
H₄: There is in the distributio	n of [1	across
populations or treatments.	<u>.</u>		_	
How do we state a null and alternative hypothesis for a	chi-square	test for inde	pendence	?
H ₀ : There is between [categoric	al variable:] and [catego	rical varia	able] in a
given population.				
H _a : There is between [categorid	al variable:	l and [cateod	orical varia	ble] in a
given population.		5		•
OR				
H_0 : [categorical variable] and [categorical variable] are _		in a	a given po	pulation.
H_{a} : [categorical variable] and [categorical variable] are _		in a	a given po	pulation.



AP Statistics CED 8.5 Daily Video 2 (Skill 4.C)

Setting Up a Chi-Square Test for Homogeneity or Independence

How do we verify the conditions for performing a chi-sq	uare test for homogen	eity or inde	pendence?
Where Do You Go to School?		-	
A random sample of parents with school-aged		2010 Sampla	2020 Sampla
children was taken during 2019. A separate random	Public	2019 Sample .	163
sample of parents with school-aged children was	Private/Parochial/Charter	16	21
taken in 2020. Parents were asked what type of school	Home	38	30
their children attended. Here are the results.			
Researchers would like to perform a test to determine if	the distribution of sch	ool type for	all school-
aged children differed between the two years. Check th	e conditions for infere	nce.	
Where Do You Go to School?			
In a previous video, we stated the hypotheses:			
H ₀ : There is in the distribution of so	hool types attended b	y school-ag	ed children
from 2019 to 2020.			
H _a : There isin the distribution of se	chool types attended b	by school-ag	ged children
from 2019 to 2020.			
We also identified the procedure as a chi-square test for	r	·	
Checking the Conditions			
Remember that for procedures in Al	P Statistics you	verify tha	t the
for using that procedure are			
-	·		
- · ·	 		
In general, you should check for:			
In general, you should check for: • independence in the methods used to	 the data, and		
In general, you should check for: • independence in the methods used to • that the appropriate	the data, and has the correct sh	аре.	
In general, you should check for: • independence in the methods used to • that the appropriate	the data, and has the correct sh	ape.	
In general, you should check for: independence in the methods used to	the data, and has the correct sh	ape.	
In general, you should check for: independence in the methods used to	the data, and has the correct sh or independenc	ape. e.	
In general, you should check for: independence in the methods used to	the data, and has the correct sh eneity or independenc	ape. e.	
In general, you should check for: independence in the methods used to	the data, and has the correct sh eneity or independenc	ape. e.	
In general, you should check for: independence in the methods used to	the data, and has the correct sh eneity or independenc random sample or	ape. e. randomizeo	ł
In general, you should check for: independence in the methods used to	the data, and has the correct sh eneity or independenc random sample or random sam	ape. e. randomizeo ple	4
In general, you should check for: independence in the methods used to	the data, and has the correct sh eneity or independenc random sample or random sam	ape. e. randomized ple	ł
In general, you should check for:	the data, and has the correct sh eneity or independenc random sample or random sam	ape. e. randomized ple	to 10%
In general, you should check for:	the data, and has the correct sh eneity or independenc random sample or random sam	ape. e. randomized ple	d to 10%
In general, you should check for:	the data, and has the correct sh eneity or independenc random sample or random sam	ape. e. randomized ple	d to 10%
In general, you should check for:	the data, and has the correct sh eneity or independenc random sample or random sam ample size is bution is approximatel	ape. e. randomized ple	d to 10%
In general, you should check for:	the data, and has the correct sh eneity or independenc random sample or random sam ample size is bution is approximatel	ape. e. randomized ple	d to 10%
In general, you should check for:	the data, and has the correct sh eneity or independenc random sample or random sam ample size is bution is approximatel	ape. e. randomized ple	d to 10%



	Na	me		
Where Do You Go to School?				
A random sample of parents with school-aged			2019 Sample	e 2020 Sample
children was taken during 2019. A separate random	Public		266	163
sample of parents with school-aged children was	Private/	Parochial/Charter	16	21
taken in 2020. Parents were asked what type of scho	ool Home		38	30
their children attended. Here are the results.				
Researchers would like to perform a test to determin	ne if the dis	tribution of scl	nool type f	or all school-
aged children differed between the two years.				
Check the conditions for inference. (Be sure to \checkmark yo	ur conditio	ns!)		
1. There are random sam	ples (rando	om sample	s).
2. It is to believe parents	is			
to or all parents in 2019 a	nd		2019 Sample	2020 Sample
parents is to or all	Publi	c	266 (257.1)	163 (171.9)
parents in 2020.	Priva	te/Parochial/Charter	16 (22.2)	21 (14.8)
3. All expected counts are	Home	9	38 (40.7)	30 (27.3)
(In a previous video we calculated the expected cou	ints.)		(expected co	unts)
Are you Working?				
A random sample of 2000 adults revealed the follow	ving data al	pout educatior	n level and	employment
status. Researchers want to determine if there is		No High School	High School	High School
an association between education level and		Diploma	Diploma, No	Diploma,
employment status. Check the conditions for	Employed	206	College	Some College
inference.	Linomployed	14	22	24
	onempioyeu	14	22	24
Are you Working?				
In a previous video, we stated the hypotheses:				
H_0 : There is no association between education level	and emplo	yment status f	or all adult	S.
H_{a} : There is an association between education level	and emplo	yment status f	or all adults	5.
We also identified the procedure as a chi-square tes	st for			_•
Are you Working? (Check the conditions using infor	mation abo	ve and 🗸 your	conditions	s!)
1. The adults were select	ed.		High Scho	ool High school
2. It is to believe that 2000 adu	ılts is	No High Sc Diploma	hool Diploma, College	No Diploma, Some College
10% of all adults.	Empl	oyed 206 (213.	4) 548 (552.	9) 1186 (1173.7)
3. All expected counts are!	Unem	ployed 14 (6.6)	22 (17.1) 24 (36.3)
The conditions have all been		(expected c	ounts)	
What Should We Take Away?				
How do we verify the conditions for performing a ch	ni-square te	st for homoge	neity or inc	lependence?
1. Data should be collected using a	ran	dom sample o	r randomiz	ed
experiment () or a		random sar	nple	
()				
2. When sampling replacement, t	he sample s	ize is		to 10%
of the population size.				



AP Statistics CED 8.6 Daily Video 1 (Skill 3.E)

Carrying Out a Chi-Square Test for Homogeneity or Independence

What Will We Learn?

How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence?

How do we calculate a *p*-value for a chi-square test for homogeneity or independence?

Where Do You Go To School?

A random sample of parents with school-aged

children was taken during 2019. A separate random sample of parents with school-aged children was

taken in 2020. Parents were asked what type of school

	2019 Sample	2020 Sample
Public	266	163
Private/Parochial/Charter	16	21
Home	38	30

their children attended. Here are the results. Researchers would like to perform a test to determine if the distribution of school type for all school-aged children differed between the two years. Find the test statistic and *p*-value.

From Previous Videos

Calculating a Test Statistic

 H_0 : There is no difference in the distribution of school types attended by school-aged children from 2019 to 2020.

 H_a : There is a difference in the distribution of school types attended by school-aged children from 2019 to 2020.

We identified the procedure as a *chi-square test* for homogeneity.

We determined all three conditions are met.

(row total)(column total)		2019 Sample	2020 Sample	Total
$expected \ count = \frac{1}{table \ total}$	Public	266 (257.1)	163 (1/1.9)	429
(Observed Functed) ²	Private/Parochial/Charter	16 (22.2)	21 (14.8)	37
$\gamma^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{(\text{Observed} - \text{Expected})^2}$	Home	38 (40.7)	30 (27.3)	68
[∧] ∠ Expected	Total	320	214	534
		(expected	counts)	
$\chi^{2} = \frac{(266 -)^{2}}{(266 -)^{2}} + \frac{(16 -)^{2}}{(266 -)^{2}} + \frac{(38 -)^{2}}{(266 -)^{2}} + \frac{(166 -)^{2}}{(266 -$	$(33-)^2 + (21-)$	$^{2} + \frac{(30 - 1)}{(30 - 1)}$	_) ²	
$\chi^2 =+++$	+ ←	contributio	ons	
$\chi^2 =$				
Calculating a <i>p</i> -value				
The p-value is the of observing a	a chi-square statistic at		as large as t	he
one observed assuming the hypothesis ar	ud prohability model a		5	
For a chi aguara tost for homographity or independ			_• •	
For a chi-square test for homogeneity or independ	ence, the p-value is ca	iculated fro	m a chi-squ	are
distribution with degrees of freedom given by the f	ormula:			
df = (1)(1)				
		2019 Sample	e 2020 Sample	Total
df = ($)($	Public	266	163	429
di = ()()	Private/Parochial/Charte	r 16	21	37
	Home	38	30	68
df =	Total	320	214	534





Calculating the p-value $\chi^2 = 14.30$ df = ()() = ()() =				Ν	ame	e					
$\chi^{2} = 14.30$ $df = (\ (\) (\) = (\) (\) (\) = (\) (\) = (\) (\) (\) = (\) (\) (\) = (\) (\$	Calculating the <i>p</i> -value										
df = () () = () () = () () = () ()	$\chi^2 = 14.30$										
p-value = $P(\chi^2 \ge 14.30)$ There are two ways to calculate this <i>p</i> -value: 1. Using Table C 2. Using Table C 2. Using Table C we can only get an estimation 1. Identify the df 2. locate the chi-square statistic 3. find the interval for the <i>p</i> -value Here we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = \$	df = ()()	= ()() =	=				
There are two ways to calculate this <i>p</i> -value: 1. Using Table C 2. Using Technology Calculating <i>p</i> -value Using Table C Using Table C we can only get an estimation 1. Identify the df 2. locate the chi-square statistic 3. find the interval for the <i>p</i> -value Here we can determine that the <i>p</i> -value Using Technology Using technology we can find the eact <i>p</i> -value. $P(\chi^2 \ge 14.30) = _$ $\chi^2 = 14.30$ Here we can determine that the <i>p</i> -value $P(\chi^2 \ge 14.30) = _$ Using Technology Using technology we can find the eact <i>p</i> -value. $P(\chi^2 \ge 14.30) = _$ $\frac{\chi^2 = 14.30}{\chi^2 = 14.30}$ Here we can determine that the <i>p</i> -value $P(\chi^2 \ge 14.30) = _$ $\frac{\chi^2 = 14.30}{\chi^2 = 14.48 \text{ Witcox}}$ $\frac{\chi^2 = \sum (Observed - Expected)^2}{(Deerved - Expected)^2}$ How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $P(\chi^2 \ge _$ test statistic) $\chi^2 = \sum (Observed - Expected)^2$ $\chi^2 = \sum (Observed - Expected)^2$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge _$ test statistic) $df = (_ _](\chi = -](\chi = -](\chi = -])$	p-value = $P(\gamma^2 > 14.30)$	•		/ \		/					
There are two ways to calculate this p-value: 1. Using Table C 2. Using Technology Calculating p-value Using Table C Using Table C we can only get an estimation 1. Identify the df 2. locate the chi-square statistic 3. find the interval for the p-value Here we can determine that the p-value is between and Calculating p-value Using Technology Using technology we can find the exact p-value. $P(\chi^2 \ge 14.30) = __$. Calculating p-value Using Technology Using technology we can find the exact p-value. $P(\chi^2 \ge 14.30) = __$. What Should We Take Away? How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $\chi^2 = \sum_{n=1}^{\infty} \frac{(biserved - Expected)^2}{Expected}$ How do we calculate a p-value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge ___\{i=1}^{\infty} test statistic)$ df = (Ν									
There we can determine that the <i>p</i> -value $P_{x^{2} = 14.30$ $p_{x^{2} = 14.30$ $p_{x^{$	There are two ways to calculate this p-value:										
The Using Table C 2. Using Table C 2. Using Table C 2. Using Table C we can only get an estimation 1. Identify the df 2. Iocate the chi-square statistic 3. find the interval for the <i>p</i> -value Here we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = $ Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = $ What Should We Take Away? How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $\chi^2 = \sum_{n=1}^{\infty} \frac{(0bserved - Expected)^2}{Expected}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, $	1 Using Table C		$\overline{\ }$						p-val	ue	
$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}$	2 Using Tashaology										
$\chi^2 = 14.30$ Calculating p-value Using Table C Using Table C we can only get an estimation 1. Identify the df $\frac{1}{20}$ $\frac{15}{20}$ $\frac{10}{20}$ $\frac{11}{20}$ $\frac{10}{120}$ $\frac{10}{120$	2. Using rechnology								t	*	
Table C we can only get an estimation 1. Identify the df 1. Identify the df 2. Iocate the chi-square statistic 3 411 446 522 425 747 728 544 1028 748 1029 1029 1029 1029 1029 1029 1029 1029									ι χ²= 14	.30	
Table C we can only get an estimation 1. Identify the df 2. Identify the df											
Using Table C we can only get an estimation 1. Identify the df 2. locate the chi-square statistic 3. find the interval for the <i>p</i> -value Here we can determine that the <i>p</i> -value is between and Calculating <i>p</i>-value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = $ Calculating <i>p</i>-value Wilcox What Should We Take Away? How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $\chi^2 = \sum_{p = 1}^{(Observed - Expected)^2}$ How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $P(\chi^2 \ge 1, \chi^2) = \sum_{p = 1}^{(Observed - Expected)^2}$ How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $P(\chi^2 \ge 1, \chi^2) = \sum_{p = 1}^{(Observed - Expected)^2}$ How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $P(\chi^2 \ge 1, \chi^2) = \sum_{p = 1}^{(Observed - Expected)^2}$ How do we calculate an appropriate test for homogeneity or independence? $P(\chi^2 \ge 1, \chi^2) = \sum_{p = 1}^{(Observed - Expected)^2}$ How do we calculate an appropriate test for homogeneity or independence? $P(\chi^2 \ge 1, \chi^2) = \sum_{p = 1}^{(Observed - Expected)^2}$ How do we calculate an appropriate test for homogeneity or independence? $P(\chi^2 \ge 1, \chi^2) = \sum_{p = 1}^{(Observed - Expected)^2}$	Calculating p-value Using Table C	Table C χ^2	Critica	l Values	•						
Using Table C we can only get an estimation 1. Identify the df 2. locate the chi-square statistic 3. find the interval for the <i>p</i> -value Here we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = $ $\chi^2 = 14.30$ What Should We Take Away? How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge 1_{10}, 0_{10$						Tail prob	ability p	,			
estimation 1. Identify the df 2. locate the chi-square statistic 3. find the interval for the <i>p</i> -value there we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = _\$ Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = _\$ There we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = _\$ What Should We Take Away? How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge ___________________________________$	Using Table C we can only get an	df .25	.20	.15	.10	.05 .025	.02	.01	.005	.0025 .00	1 .0005
1. Identify the df 2. locate the chi-square statistic 3. find the interval for the <i>p</i> -value Here we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = $ $\chi^{2}=14.30$ Here we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = $ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=14.30$ $\chi^{2}=2$	estimation	1 1.32	1.64 3.22	2.07 3.79	2.71 4.61	3.84 5.02 5.99 7.38	5.41 7.82	6.63 9.21	7.88	9.14 10. 11.98 13.	83 12.12 82 15.20
2. locate the chi-square statistic 3. find the interval for the <i>p</i> -value $4 \ 539 \ 559 \ 647 \ 729 \ 649 \ 6114 \ 1148 \ 1138 \ 1138 \ 1148 \ 1048 \ 1047 \ 1057 \ 1059 \ 1077 \ 1059 \ 201 \ 2$	1. Identify the df	3 4.11	4.64	5.32	6.25	7.81 9.35	9.84	11.34	12.84	14.32 16	7 17 73
3. find the interval for the <i>p</i> -value Here we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = \$	2. locate the chi-square statistic	4 5.39 5 6.63	5.99 7.29	6.74 8.12	7.78 9.24	9.49 11.14 11.07 12.83	11.67 13.39	13.28 15.09	14.86 16.75	16.42 18 18.39 20	7 200 51 22.11
Here we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = \$ Important flash and the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = \$ Important flash and the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = \$ Important flash and the exact <i>p</i> -value is the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = \$ Important flash and the exact <i>p</i> -value is the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = \$ The exact flash and the exact <i>p</i> -value is the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = \$ How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $P(\chi^2 \ge \$	3. find the interval for the <i>p</i> -value										
Here we can determine that the <i>p</i> -value is between and Calculating <i>p</i> -value Using Technology Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) =$						$\chi^2 = 14.30$					
Calculating p-value Using Technology Using technology we can find the exact p-value. $P(\chi^2 \ge 14.30) = $ Implied FLOAT AUTO BLAL BADJAN HP	Here we can determine that the <i>p</i> -value is be	tween _		a	nd _						
Using technology we can find the exact <i>p</i> -value. $P(\chi^2 \ge 14.30) = $ Using technology we can find the exact <i>p</i>-value. $P(\chi^2 \ge 14.30) =$ Using technology we can find the exact <i>p</i>-value. $P(\chi^2 \ge 14.30) =$ Using technology we can find the exact <i>p</i>-value. P($\chi^2 \ge 14.30$) = Using technology we can find the exact <i>p</i>-value. P($\chi^2 \ge 14.30$) = Using technology we can find the exact <i>p</i>-value. P($\chi^2 \ge 14.30$) = Using technology we can find the exact <i>p</i>-value for a chi-square test for homogeneity or independence? P($\chi^2 \ge$ test statistic) df = ()() Use Table or χ^2cdf on the calculator.	Calculating <i>p</i> -value Using Technology										
$\frac{ VBEHAL FLOAT AUTO REAL RADZAM H?}{ I normalcdf(2: normalcdf(3: invNorm(4: invT(5: trdf(5: tr$	Using technology we can find the exact <i>p</i> -valu	ue. P(x	$^{2} \geq$	14.3	0) =						
What Should We Take Away? How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $x^{2} = \sum \frac{(Observed - Expected)^{2}}{Expected}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(x^{2} \ge \underline{\qquad} test statistic)$ of = ()() Use Table or x^{2}cdf on the calculator.	NORMAL FLOAT AUTO REAL RADIAN MP 👖 NORMAL FLOAT AUTO REAL	RADIAN MP	n	G	IORMAL	FLOAT AUTO REA	L RADIAN	N MP	n		
$\frac{1:\text{normalledf(}}{2:\text{normalledf(}} \\ \frac{2:\text{normalledf(}}{3:\text{invNorm(}} \\ \frac{4:\text{invT(}}{5:\text{trdf(}} \\ \frac{6:\text{trdf(}}{7:\text{trbdf(}} \\ \frac{6:\text{trdf(}}{9:\text{trbdf(}} \\ \frac{1}{2:\text{vectedf(}} \\ \frac{1}{2:\text{vected}} \\ \frac{1}{2:vected$	DISTR DRAW X2cdf			2	(²cdf	(14.30,100	00000	,2)			
$\frac{2 : i n \text{VNorm}(}{\text{SitPaf(}}) = \frac{1}{2} \text{Paste} $	1:normalpdf(2:normalcdf(upper:1000000					7.8	348640	0813	<u>e4</u>		
S: t-pdf(S: t-pdf(3:invNorm(4:invT(
$\frac{7:122\text{ bdf}(}{\text{PLFPdf}(\text{Constraints})} = \text{Cluke Wilcox} \qquad \text{@ Luke Wilcox} \qquad \text{& I Luke Wilcox} \qquad & I Luke Wilco$	5:tpdf(6:tcdf(
Purpoint © Luke Wilcox © Luke Wilcox What Should We Take Away? How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge ___\$	7:X ² Pdf(B H X ² Cdf(
$ \begin{array}{c} \hline \\ \hline $											
What Should We Take Away? How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge ____]$ test statistic) df = ()(\) Use Table or χ^2 cdf on the calculator.	© Luke Wilcox	© Luke Wile	сох				© Lu	ike Wi	lcox		
How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $\chi^{2} = \sum \frac{(\text{Observed} - \text{Expected})^{2}}{\text{Expected}}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^{2} \ge ___________________________________$	M/hat Chauld M/a Tales August										
How do we calculate an appropriate test statistic for a chi-square test for homogeneity or independence? $\chi^{2} = \sum \frac{(\text{Observed} - \text{Expected})^{2}}{\text{Expected}}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^{2} \ge ___________________________________$	vvnat Snould vve Take Away?					to at famil					
Independence? $\chi^{2} = \sum \frac{(\text{Observed} - \text{Expected})^{2}}{\text{Expected}}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^{2} \ge ___________________________________$	How do we calculate an appropriate test stati	stic for a	a chi	-squ	are	test for h	nomc	bgei	neity	/ or	
$x = \angle \qquad \text{Expected}$ How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge ___________________________________$	independence? (Observed – Expected)	ed) ²									
How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge ___________________________________$	x ⁻ ∠ Expected										
How do we calculate a <i>p</i> -value for a chi-square test for homogeneity or independence? $P(\chi^2 \ge ___________________________________$		-								_	
$P(\chi^2 \ge __\ \text{ test statistic})$ df = ()() Use Table or χ^2 cdf on the calculator.	How do we calculate a <i>p</i> -value for a chi-squar	re test fo	or ho	mog	gene	eity or inc	depe	ende	ence	?	
df = ()() Use Table or γ^2 cdf on the calculator.	$P(\chi^2 \geq$ test statistic)										
Use Table or χ^2 cdf on the calculator.	df = ()())									
	Use Table or χ^2 cdf on the calculator.										



AP Statistics CED 8.6 Daily Video 2 (Skill 4.B)

Carrying Out a Chi-Square Test for Homogeneity or Independence

What Will We Learn?			
How do we interpret the p-value in a chi-square test for	homogeneity or inde	pendence?	
How do we state a conclusion for a chi-square test for h	omogeneity or indepe	endence?	
Where Do You Go To School?			
A random sample of parents with school-aged			
children was taken during 2019 A separate random	Dublic	2019 Sample 20	20 Sample
sample of parents with school-aged children was	Public	266	163
taken in 2020. Parents were asked what type of school	Private/Parochial/Charter	16	21
their children attended. Here are the results. Researcher	Home	38 matest to det	30 tormino if
the distribution of school type for all school aged childre	s would like to period		Find the
test statistic and pivalue		le two years. I	ind the
From Province Videoc			
H: There is no difference in the distribution of school tu	nos attanded by sche	al agod childr	on from
r_{0} . There is no amerence in the distribution of school ty 2019 to 2020	pes allended by scho	or-aged childr	
UIT IU ZUZU.	or attended by sehes	and childre	n from
n_a . There is a difference in the distribution of school typ 2010 to 2020	es allended by schoo	-aged childrei	
We identified the procedure as a chi square test for her	nogeneity		
We determined all three conditions are rest	nogeneity.		
we determined all three conditions are met. $y^2 = 5.55$ df = 2 p value = 0.042			
$\chi^{-} = 5.55, \text{ di} = 2, p \text{-value} = 0.082$			
The pivolue measure how it is to get avide	and for U an	22 or	
the p-value measure now It is to get evidence	$\frac{1}{100} H_a dS \underline{\qquad}$		
Interpreting a pixelue			
Assuming a <i>p</i> -value there is a < <i>p</i> -value >	of aetting a 1	r^2 of coalculat	ed chi-
square> or greater by in the random	or getting a		
In context:	sumple(s) of random	ussignment.	
Assuming there is in the distribution of	of fo	r school-aged	children
from 2019 to 2020, there is a	$\mathbf{v}^2 =$	or	by
chance alone in the random samples	Jetting x –	OI	Oy
Stating a Conclusion			
Small p values	by random	alone	
Similar p-values $-$ test statistic is to occur	by random		
• Decause the p -value of $\leq u =,$ we There is convincing	= Teject 7 10. 5+ [1	
Large pivalues — Notest statistic is	by random	j. Dono	
Large p-values r less statistic is to occur	by fail to reject H		
• Decause the p-value of $\leq u - $, we There is not convincing	that [1	
Stating a Conducion	5 ulat []·	
No significance lovel was stated in the school time over			ic the
most	ipie, so we will use α :	–, wnicr	i is the
Peopuse the pivolue of			
because the <i>p</i> -value of, v	ve	·	and the state
I nere is that the	ere is a	In the di	Istribution
of school types for school-aged children from			
	-	STATS N	1EDIC

			Name_				
Are you Working?							
A random sample of 2000 adults revealed	ed the follo	owing dat	a about	education	level	and e	employment
status. Researchers want to determine if	there is a	n	ľ	No High School	High	School	High School
association between education level and				Dipionia	Co	llege	Some College
employment status. The p-value for a ch	ii-square te	est Empl	loyed	206	5	548	1186
for independence is 0.0008. Interpret th	e p-value	and Unen	nployed	14	:	22	24
make a conclusion at the $\alpha = 0.01$ signif	icance leve	el.					
	1					I. I.	
H_0 : There is no association between edu	ication lev	el and em	ployme	ent status to	r all a	adults	•
H _a : There is an association between edu	cation leve	el and em	ployme	nt status to	r all a	adults	
We identified the procedure as a chi-squ	uare test t	or indepe	ndence				
We determined all three conditions are	met. χ ² =	14.30, dt	= 2, p-	/alue = 0.00)08		
Here is what the <i>p</i> -value actually means:							
Assuming there is bet	tween edu	cational le	evel and	d employme	ent st	tatus f	or all adults,
there is a 0.0008 of g	etting a χ	² of	or		by c	chance	e alone in a
random sample of	·						
State A Conclusion							
Because the <i>p</i> -value of		, we		•			
There is	that	the		betw	een e	educa	tional level
and employment status for all adults.							
Follow-Up Analysis			High Schoo	High School			
	r	lo High School Diploma	Diploma, N College	o Diploma, Some College	Total		
The largest contribution to the chi-	Employed	206 (213.4)	548 (552.9)	1186 (1173.7)	1940	2 (($phserved - expected)^2$
square statistic is, because	Unemployed Total	14 (6.6) 220	22 (17.1) 570	24 (36.3) 1210	60 2000	$\chi^2 = \Sigma^{\underline{\frown}}$	expected
the number of	iotai (expected counts	s)	1210	2000		
unemployed adults with no high	$\chi^2 = \frac{(206 - 21)^2}{212}$	$(548 - 552)^{2} + (548 - 552)^{2}$	$\frac{(1186)^2}{(1186)^2}$ + $\frac{(1186)^2}{(1186)^2}$	$\frac{-1173.7)^2}{172.7} + \frac{(14-6.7)^2}{6.6}$	$\frac{(6)^2}{(2)} + \frac{(2)}{(2)}$	$\frac{2-17.1)^2}{17.1}$	$+\frac{(24-36.3)^2}{26.2}$
school diploma is much	$\chi^2 = 0.26 + 0.1$	04+ 0.13 + 8.30 +	1.40 + 4.17	← contributions		17.1	30.3
than	$x^2 = 14.30 \leftarrow$	chi-square statistic					
What Chauld M/a Taka Away?	χ = 14.00 €	on oqualo olanon					
what Should we Take Away?				· ·		2	
How do we interpret the <i>p</i> -value in a chi	I-square te	est for non	nogene	ity or indep	ende	ence	
Assuming, there is	a <p-value< td=""><td>9></td><td> OT</td><td>geπing a χ</td><td>/~ OT</td><td><caic< td=""><td>ulated chi-</td></caic<></td></p-value<>	9>	OT	geπing a χ	/~ OT	<caic< td=""><td>ulated chi-</td></caic<>	ulated chi-
square> or greater, by	_ in the ra	indom sar	npie(s)	or random	assig	Inmen	τ.
		(I			1		
How do we state a conclusion for a chi-s	square test	t for home	ogeneit	y or indepei	nden	ce?	
Because the <i>p</i> -value of	$\leq \alpha = $, we re	ject H ₀ .	-			
	evider	nce that [••••]•		
Because the <i>p</i> -value of	$\leq \alpha = $, we fai	il to reje	ect H₀.	-		
There is not convincing	ev	vidence th	at [].		



AP Statistics CED 8.6 Daily Video 3 (Skill 4.E)

Carrying Out a Chi-Square Test for Homogeneity or Independence

What Will We Learn?

How do we perform a complete chi-square test for homogeneity or independence?

2017 #5

The table and bar chart below summarize the age at diagnosis, in years, for a random sample of 207 mean and women currently being treated for schizophrenia.

		Age-Grou	up (years)		
	20 to 29	30 to 39	40 to 49	50 to 59	Total
Women	46	40	21	12	119
Men	53	23	9	3	88
Total	99	63	30	15	207



Do the data provide convincing statistical evidence of an association between age-group and gender in the diagnosis of schizophrenia?

In the wording of the question there is a clue to us that we need to do a full significance test. Highlight that wording. What type of test do we want here? We have a ______ and we know that we are trying to see if there are two ______ that are associated so we know that this is going to require a *chi-square test for* ______ 2017 #5 Hypotheses _____ and _____ are _____ (that is, they are not H₀: _____ _____ for the population of people ______ for schizophrenia. H_a: ______ and _____ are _____ (that is, they are

_____ for schizophrenia. associated) for the population of people _____

No significance level was stated, so we'll use $\alpha =$ _____.

2017 #5 Procedure and Conditions

Procedure: chi-square test for _____

Conditions: (Be sure to \checkmark the conditions!)

1. _____ men and women currently being treated for schizophrenia.

2. ______ is less that ______ of ______ currently being treated for schizophrenia.

3. All _____ counts are greater than ____.

table total

		Age-Gro	up (years)		
	20 to 29	30 to 39	40 to 49	50-59	Total
Women	46	40	21	12	119
Men	53	23	9	3	88
Total	99	63	30	15	207

Use the formula to calculate all of the expected counts.

All conditions have been _____.

expected count = $\frac{(row total)(column total)}{(row total)(column total)}$

Key: (expected counts)



						Ν	Vame	
2017 #	5 Test Stat	istic an	d <i>p</i> -valu	е				
		Ago-Gro						.2
	20 to 29	30 to 39	40 to 49	50 to 59	Total	$\chi^2 = \sum \frac{1}{2}$	observed – expected	
Wome	n 46 (56.91) 4	40 (36.22)	21 (17.25)	12 (8.62)	119		expected	
Men	53 (42.09)	23 (26.78)	9 (12.75)	3 (6.38)	88			
Total	99	63	30	15	207	Use the form	mula to calculate t	he χ^2 statistic.
	(expected	counts)						
$\chi^2 =$								
$\chi^2 =$								p-value
$\chi^2 =$		d	-		p-value	=		$\chi^2 = 10.884$
2017 #	5 Calculate	or	_				a —	
	elect Matrix	HP O	Ente	r Observe	d Data		ect χ^2 Test	Select Calculate
NOMES ME 1 [A] 1> 2:[B] 1> 3:[C] 4:[D] 5:[E] 6:[F] 7:[G] 8:[H]	ATH EDIT ×1 ×1	<u>" 0</u>	NORMAL FLOAT	2 ×4 21 12 9 3		EDIT CALC CONTRACTOR CALC CONTRACTOR CALC CONTRACTOR CALC CONTRACTOR CALC CONTRACTOR CALC CONTRACTOR CONTRAC	0 REAL KADIAN HP	NORHAL FLOAT AUTO REAL RADIAN HP
94[1]	©L	uke Wilcox	(R)(1,1)= 46		© Luke Wilcox	H:HNOVH(© Luke Wilcox	© Luke Wilcox
	Test Sta	tistic		Expecte	d Counts			
	<i>p</i> -value	and df		are nov	w in Matrix	В		
-	NORMAL FLOAT AUTO A X2-10.883837 p=0.01237088 df=3	EAL RADIAN M est 3 41 © Luke	U Milcox	DRHAL FLOAT AUT ATRIX(B) 2 55.913 36.217 42.087 26.783	TO REAL RADIAN HP 2 ×4 17.246 8.6232 12.754 6.3768	1) B Wilcox		
2017 #	5 Conclusi	on	THICOX					
Becaus	se the p-va	lue				we		
There					•	that there i	s an	 between
			and			for the r	population curr	ently being treated for
schizor	ohrenia.		0.10					
What S	Should We	Take A	wav?					
How de	o we perfo	rm a co	molete	chi-squa	re test fo	r homoaen	eity or indeper	idence?
Make s	sure to:		p.oto	0				
•	State the		and		hypo	theses		
•	Identify th	e				lse	is not is stated	d.
•	Identify th	e			vou are u	sina.	-	
•	Verify that	the		f	, or the pro	ocedure are	e met (with evi	dence!).
•	Calculate	the			and	the	•	•
•	Make a			base	d on the	p-value. (Y	ou do not need	d to the <i>p</i> -
	value unle	ss spec	ificallv a	sked.)		,		•••• P
			· · · · · · · · · · · · · · · · · · ·					



AP Statistics CED 8.7 Daily Video 1

Skill Focus – Inference Procedures for Categorical Data

What Will We Learn?			
How do we identify an appropriate chi-square tes	t for a set of c	ategorical data?	
Identifying the Procedure			
Goal: to a distribution of a categori	cal variable to	a dist	ribution
Data collection: random sample from _	popu	lation	
\rightarrow (1) chi-square test for	(_ sample, categ	orical variable)
Goal: to distributions of a categorica	l variable for _	por	oulations
Data collection: random sampl	es from	populations	(or from multiple
groups in a randomized experiment)			
ightarrow (2) chi-square test for	(samples, cate	gorical variable)
Goal: to determine whether categorical v	ariables are	·	
Data collection: random sample fro	m poj	pulation.	
ightarrow (3) chi-square test for	(sample, categ	gorical variables)
2008 #5			
A study was conducted to determine where moos	se are found ir	a region containing	a large burned
area. A map of the study area was partitioned into	o the following	g four habitat types.	
(1) Inside the burned are, not near the edge of the	e burned area		
(2) Inside the burned area, near the edge.			4
(3) Outside the burned area, near the edge, and			3
(4) Outside the burned area, not near the edge.			
The figure shows these four habitat types.		Note: Figure not dr	awn to scale.
2008 #5			
The proportion of total acreage in each of the			
habitat types was determined for the study	Habitat Type	Proportion of Total Acreage	Observed
area. Using an aerial survey, moose locations	1	0.340	25
were observed and classified into one of the	2	0.101	22
of the four habitat types. The results are given	3	0.104	30
in the table below. The researcher who are	4	0.455	40
conducting the study expect the number of	Total	1.000	117
moose observed in a habitat type to be proportic	nal to the amo	ount of acreage of the	at type of
habitat. Are the data consistent with this expectat	ion? Conduct	an appropriate statis	tical test to
support your conclusion.			
2008 #5			
What type of test is appropriate here?			
We need to know:			
Number of samples:			
Number of categorical variables:			
We want compare the distribution of habitat type	from the	to the	distribution
that it would be proportional to the amount of ac	reage.		
	-		
So, in this case we will use a $ ightarrow$ chi-square test for			



Name___

2016 #2

Product advertisers studied the effects of television ads on children's choices for two new snacks. The advertisers used two 30-second television ads in tan experiment. One ad was for a new sugary snack called Choco-Zuties, and the other was for a new healthy snack called Apple-Zuties. For the experiment, 75 children were randomly assigned to one of three groups, A, B, or C. Each child individually watched a 30-minute television program that was interrupted for 5 minutes of advertising. The advertising was the same for each group with the following exceptions.

- The advertising for group A included the Choco-Zuties ad but not the Apple-Zuties ad.
- The advertising for group B included the Apple-Zuties ad but not the Choco-Zuties ad.
- The advertising for group C included neither the Choco-Zuties ad not the Apple-Zuties ad.

2016 #2

After the program, the children were offered a choice between the two snacks. The table below summarizes their choices. Do the data provide

Group	Type of Ad	Number Who Chose Choco-Zuties	Number Who Chose Apple-Zuties
Α	Choco-Zuties only	21	4
В	Apple-Zuties only	13	12
С	Neither	22	3

convincing statistical evidence that there is an association between type of ad and children's choice of snack among all children similar to those who participated in the experiment?

2016 #2

What type of test is appropriate here? Since this is a randomized experiment, we need to know:

Number of groups: ____

Number of categorical variables: _____

So, in this case we will use a \rightarrow chi-square test for _

2013 #4

The Behavioral Risk Factor Surveillance System is an ongoing health survey system that tracks health

conditions and risk behaviors in the United States. In one of their studies, a random sample of 8,866 adults answered the question "Do you consume five or more servings of fruits and vegetables per day?" The data are summarized by response and by agegroup in the frequency table below. Do the data provide

convincing statistical evidence that there is an association

between age-group and whether or not a person consumes five or more servings of fruits and vegetables per day for adults in the United States?

2013 #4

What type of test is appropriate here? Since this is a randomized experiment, we need to know: Number of samples:

Number of categorical variables: _____

So, in this case we will use a \rightarrow chi-square test for ____

What Should We Take Away?

How do we identify an appropriate chi-square test for a set of categorical data?

____ sample, _____ categorical variable ightarrow (1) chi-square test for ______

____ samples, _____ categorical variable ightarrow (1) chi-square test for ___

____ sample, _____ categorical variables ightarrow (1) chi-square test for _

Age-Group (years)	Yes	No	Total
18–34	231	741	972
35-54	669	2,242	2,911
55 or older	1,291	3,692	4,983
Total	2,191	6,675	8,866

