12.1 A HW

(A)

(1) PARAMETER: p_{nuts} = true population proportion of nuts

HYPOTHESIS: $H_0: p_{cashew} = .52$ $p_{almond} = .27$

$p_{macadamia} = .13$ $p_{brazil} = .08$

H_A: at least one of the p_i's is incorrect

(B)

NUT	# OBSERVED	% EXPECTED	# EXPECTED	$\frac{(O-E)^2}{E}$
Cashew	83	.52	78.0	0.3205
ALmond	29	.27	40.5	3.2654
Macadamia	20	.13	19.5	0.01282
Brazil	18	.08	12.0	3.0
Total	150	1.00	150	$\boxed{\Sigma x = 6.599}$

Take advantage of calculator

L1 L2 L3 = 150(L2) $L4 = \frac{(L-L3)^2}{L3}$

1VARSTAT L4

(3) $\chi^2 = .3205 + 3.2654 + .01282 + 3.0 = 6.599$

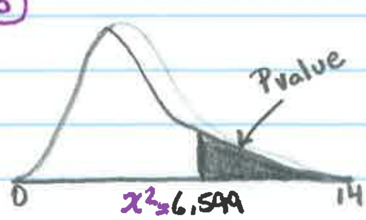
DO BY HAND - TIP USE LISTS IN CALC

$$\chi^2 = \frac{(83-78)^2}{78} + \frac{(29-40.5)^2}{40.5} + \frac{(20-19.5)^2}{19.5} + \frac{(18-12)^2}{12} = 6.599$$

(5) (a) The expected counts are all at least 5.

There are 4 categories - df = 3 for χ^2 distribution.

(b)



Chi-SQUARE distribution with 3 df

(c) $P(\chi^2 \geq 6.599) = .0858$ $> .05$

$\chi^2 cdf(6.599, E99, 3)$

Since th pvalue > .05, we fail to reject Ho. We do not have enough evidence to say the companies claim about the distribution of nuts is wrong.

11.1B

7

TREES IN FOREST	%	BIRDS OBSERVED	EXPECTED #	$\frac{(O-E)^2}{E}$
DOUGLAS FIRS	.54	70	84.24	2.4071
PINES	.40	79	62.40	4.416
OTHER TYPES	.06	7	9.36	0.595
	1.00	156	156	Σ 7.418 $= \chi^2$

USE L1, L2, L3 or do by hand

TEST : χ^2 GOODNESS OF FIT TEST FOR $\alpha = .05$

Hypothesis P_t = true proportion of trees in forest

H_0: $P_{Firs} = .54$ $P_{pines} = .40$ $P_{other} = .06$
H_A: At least one of the P_t's is incorrect

CONDITIONS
Random - a random sample was used
Independent - reasonable $156(10) = 1,560$ red breasted nut hatches
Large sample size - The expected counts in each category was greater than 5 (84.24, 62.4, 9.36)

MECHANICS $\chi^2 = 7.418$ df = 2
pvalue $\rightarrow P(\chi^2 \geq 7.418) = \chi^2cdf(7.418, E99, 2) = .0245$

Conclude: Since the pvalue (.0245) < .05, We Reject H_0, and conclude these birds prefer particular types of trees when they are searching for food.

$\boxed{11.1}$

$\boxed{\#9}$ THE DATA PROVIDED IS AMOUNT OF TIME (IN minutes).

THEREFORE χ^2 IS NOT APPROPRIATE BECAUSE TO DO A χ^2 TEST, THE DATA MUST BE OBSERVED COUNTS!!

$\boxed{\text{12.1B HW}}$

(11) P_{digit} = true proportion of Benford's law digit

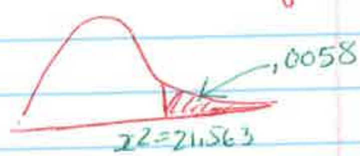
H_0: $p_1 = .301$ $\quad p_3 = .125 \quad p_5 = .079 \quad p_7 = .058 \quad p_9 = .046$

$p_2 = .176 \quad p_4 = .097 \quad p_6 = .067 \quad p_8 = .051$

H_A: at least one of the P_{digits} is incorrect

STATE TEST: CHI SQUARE (χ^2) Goodness of fit test

$\alpha = .05$



CONDITIONS

Random – random sample of 250 invoices

Independent – reasonable their are 10(250) = 2500 invoices at the company

Large sample size – The expected counts are at least 5:

Must Give all expected Counts and round 2 decimals → 75.25, 44, 31.25, 24.25, 19.75, 16.75, 14.50, 12.75, 11.5

MECHANICS:

$$\chi^2 = \sum \frac{(\text{observed} - \text{Expected})^2}{\text{Expected}} = \underbrace{\frac{(61-75.25)^2}{75.25} + \dots + \frac{(6-11.5)^2}{11.5}}_{\text{Can show 1st + last}}$$

$\chi^2 = 21.563 \qquad df = 8$

pvalue = $P(\chi^2 \geq 21.563) = \chi^2cdf(21.563, E99, 8) = .0058$

CONCLUDE: Since the pvalue is less than .05, we reject H_0 and conclude that the invoices are inconsistent with Benford's Law

12.1B HW

#11 cont

(A) Follow up

Analysis:

Digit	Observed		Expected	χ^2
1	61		75.25	2.7
2	50		44.00	0.8
3	43	>	31.25	4.4 *
4	34	>	24.25	3.9 *
5	25		19.75	1.4
6	16		16.75	0.03
7	7	<	14.50	3.9 *
8	8		12.75	1.8
9	6		11.5	2.6

Reviewing χ^2 contribution - 3,4,7 have the largest contribution. Digits 3+4 have too many and Digit 7 has not enough.

(11B) Type I error: says that the company's invoices did not follow Benford's law (suggesting fraud) when in fact they were consistent with Benford's law.

Type II error: says that the invoices were consistents with Benford's law (suggesting fraud) when in fact they were not.

A Type I error would be more serious here. Alleging that the company had committed fraud when it had not

15 H_0: $p_i = 1/12 = .083$ for all astrological signs
H_A: At least 1 of the proportions is incorrect

Test: χ^2 Goodness of Fit $\alpha = .05$ $df = 12-1 = 11$

Conditions: Random: a random sample of 4,344 people
Independent: sampling without replacement
There are more than 43,440 people in the U.S.
Large sample: All the expected counts are greater than 5. The expected count is $1/12(4344) = 362.00$ for every sign

Calculations Show this work

$$\chi^2 = \frac{(321-362)^2}{362} + \cdots + \frac{(355-362)^2}{362} = 19.76$$

$$P(\chi^2 > 19.76) = .0487$$

To do calculations:
① L1 are Observed counts
② L2 - Expected counts = 362
③ STAT TESTS χ^2GOF-TEST
L1 L2 df=11
$\chi^2 = 19.76$
$p = .0487$
$df = 11$

Conclusion: Since the pvalue (.0487) is less than $\alpha = .05$, Reject H_0. and conclude that the 12 signs are not equally likely

Followup Analysis: Refer to χ^2GOF-TEST - CNTB values
4.6, .01, .07, .39, 1.2, 4.4, 2.4, 3.0, 2.6, .17, .54, .13
Aries Virgo

The largest contributions to the χ^2 statistic are Aries and Virgo. There are fewer Aries and more Virgo than expected

11.1 B HW

17 TEST: χ^2 goodness-of-fit test $\alpha = .05$

H_0: $p_{SMOOTH} = .75$ $p_{WRINKLED} = .25$

H_A: AT LEAST ONE OF THE p_i's IS INCORRECT.

CONDITIONS

Random and Independent Conditions were given

Large enough sample size -

The expected counts 417 and 139 are both greater than 5.

PEAS	%	OBS	EXPECTED	$\chi^2 = \frac{(O-E)^2}{E}$
SMOOTH	.75	423	417	.0863
WRINKLED	.25	133	139	.2589
	1.00	n = 556	556	$\boxed{\chi^2 = .3452}$

Mechanics TEST: χ^2 GOODNESS OF FIT

$\chi^2 = .352$ df = 1

PVALUE = $P(\chi^2 \geq .352) = \chi^2 cdf(.3452, E99, 1) = .5568$

Conclude:

Since the pvalue is very large and greater than .05, we fail to reject Ho, we do not have enough evidence to dispute Mendel's belief.

11.25 REVIEW QUESTION

IN THE NEXT CHAPTER WE WILL GO BACK TO LSRL AND DO REGRESSION INFERENCE

$\widehat{GPA} = \underbrace{3.42}_{\text{y intercept}} + \underbrace{0.024}_{\text{slope}}(\text{BOOKS\_READ})$ $\quad r^2 = .083$

A) The yintercept is 3.42. This means that we would predict an ENGLISH GRADE OF 3.42 FOR A STUDENT THAT HAD READ NO BOOKS.

B) STUDENT READ 17 BOOKS w/ ENGLISH GPA OF 2.85.

$\widehat{GPA} = 3.242 + .024(17) = 3.828$

RESIDUAL $= y - \hat{y} = 2.85 - 3.828 = \underline{\underline{-.9\cancel{78}}}$

C) $r^2 = .083$ ← STRENGTH OF MODEL $\quad r = \sqrt{.083} = .288$ ← STRENGTH OF ASSOCIATION

THE RELATIONSHIP BETWEEN GPA AND NUMBER OF BOOKS READ IS NOT VERY STRONG.

(1) THE VALUE OF r^2 is only .083 which means that only 8.3% of the variation in ENGLISH GRADES is ACCOUNTED FOR BY THE LINEAR RELATIONSHIP WITH NUMBER OF BOOKS READ

OR

(2) The correlation coefficient between GPA and number of books read is only .288 indicating a weak positive association