Max Shi
Professor Li
MA 331
October 18, 2019
I pledge my honor that I have abided by the Stevens Honor System.

9.37.a. Table

Stratum	Allowed Claims	Not Allowed Claims	Total
Small	51	6	57
Medium	12	5	17
Large	4	1	5
Total	67	12	79

b. Percentage of not allowed claims
Small – 6/57 = 10.5%
Medium – 5/17 = 29.4%
Large – 1/5 = 20%

c. We do this because the expected value for the large and not allowed claims is too small. It would be 5*12/79 which is about 0.75, which is too small.

d. Null hypothesis – There is no association between claim size and whether or not the claim is allowed.

e.

Stratum (Observed)	Allowed Claims	Not Allowed Claims	Total
Small	51	6	57
Medium + Large	16	6	22
Total	67	12	79

Stratum (Expected)	Allowed Claims	Not Allowed Claims	Total
Small	48.341	8.658	57
Medium + Large	18.658	3.342	22
Total	67	12	79

$$X^2 = \frac{(51-48.341)^2}{48.341} + \frac{(6-8.658)^2}{8.658} + \frac{(16-18.658)^2}{18.658} + \frac{(6-3.342)^2}{3.342}$$
$$X^2 = 0.146 + 0.816 + 0.379 + 2.11 = 3.455$$
$$df = (2-1)(2-1) = 1$$
$$P(X_1^2 > 3.455) = 0.063$$

9.38.a. Estimated Unallowed Claims

Strata	Claims in Strata	Sample Proportion	Estimated Unallowed Claims
Small	3342	10.5%	351.79
Medium	246	29.4%	72.32
Large	58	20%	11.6

b. Margins of Error:
Small –

$$SE_p = \sqrt{\frac{0.105(1-0.105)}{57}}$$
$$SE_p = 0.0406$$
$$SE = \sqrt{\frac{0.105(1-0.105)}{3342}}$$
$$SE = 0.0053\ for\ population$$

For confidence level C, the confidence interval for the estimation is:

$$\left(351.79 - (0.0053 * 3342 * z_C^*), 351.79 + (0.0053 * 3342 * z_C^*)\right)$$
$$\left(351.79 - (17.72 * z_C^*), 351.79 + (17.72 * z_C^*)\right)$$

Where z^*_C is the z value for which the area between z^*_C and $-z^*_C$ under the normal curve is C.

Medium –

$$SE = \sqrt{\frac{0.294(1-0.294)}{246}}$$
$$SE = 0.0290$$

For confidence level C, the confidence interval for the estimation is:

$$\left(72.32 - (0.0290 * 246 * z_C^*), 72.32 + (0.0290 * 246 * z_C^*)\right)$$
$$\left(72.32 - (7.134 * z_C^*), 72.32 + (7.134 * z_C^*)\right)$$

Where z^*_C is the z value for which the area between z^*_C and $-z^*_C$ under the normal curve is C.

Large –

$$SE = \sqrt{\frac{0.2(1-0.2)}{58}}$$
$$SE = 0.05252$$

For confidence level C, the confidence interval for the estimation is:

$$\left(11.6 - (0.0525 * 58 * z_C^*), 11.6 + (0.0525 * 58 * z_C^*)\right)$$
$$\left(11.6 - (3.045 * z_C^*), 11.6 + (3.045 * z_C^*)\right)$$

Where z^*_C is the z value for which the area between z^*_C and $-z^*_C$ under the normal curve is C.

9.50.

	x <= -0.6	-0.6 < x <= -0.1	-0.1 < x <= 0.1	0.1 < x <= 0.6	0.6 < x	Total
Count	139	102	41	78	140	500
Probability	0.274	0.186	0.0797	0.186	0.274	1
Expected Count	137.13	92.96	39.83	92.96	137.13	500

$$X^2 = \frac{(139-137.13)^2}{137.13} + \frac{(102-92.96)^2}{92.96} + \frac{(41-39.83)^2}{39.83} + \frac{(78-92.96)^2}{92.96} + \frac{(140-137.13)^2}{137.13}$$

$X^2 = 0.0255 + 0.879 + 0.0344 + 2.407 + 0.0600 = 3.4059$
$df = 5 - 1 = 4$
$P(X_4^2 > 3.4059) = 1 - 0.508 = 0.492$

There is about a 50% chance that these numbers were generated from a standard normal distribution.

	x <= -0.8	-0.8 < x <= -0.2	-0.2 < x <= 0.2	0.2 < x <= 0.8	0.8 < x	Total
Count	110	104	97	88	101	500
Probability	0.211	0.208	0.159	0.208	0.211	1
Expected Count	105.5	104	79.5	104	105.5	500

$$X^2 = \frac{(110 - 105.5)^2}{105.5} + \frac{(104 - 104)^2}{104} + \frac{(97 - 79.5)^2}{79.5} + \frac{(88 - 104)^2}{104} + \frac{(101 - 105.5)^2}{105.5}$$
$$X^2 = 0.1919 + 0 + 3.8522 + 2.4615 + 0.1919 = 6.698$$
$$P(X_4^2 > 6.698) = 1 - 0.847 = 0.153$$

There is a 15.3% chance that this data was generated by a normal distribution.