Homework2

Name: Zixin Ouyang

Exercise1

The FREQ Procedure

Table of treat by allergy			
treat	**allergy**		
Frequency Expected Row Pct	**N**	**Y**	**Total**
control	87 103.16 66.92	43 26.842 33.08	130
intervention	109 92.842 93.16	8 24.158 6.84	117
Total	196	51	247

a. The percent of allergic children in the control group is 33.08, while the percent of allergic children in the intervention group is 6.84.

Statistics for Table of treat by allergy

Statistic	DF	Value	Prob
Chi-Square	1	25.8765	<.0001
Likelihood Ratio Chi-Square	1	28.1791	<.0001
Continuity Adj. Chi-Square	1	24.2998	<.0001
Mantel-Haenszel Chi-Square	1	25.7717	<.0001
Phi Coefficient		-0.3237	
Contingency Coefficient		0.3079	
Cramer's V		-0.3237	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	87
Left-sided Pr <= F	<.0001
Right-sided Pr >= F	1.0000
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

b.

H_0: There is no association between treat and allergy.
H_A: There is association between treat and allergy.
We can use the chi-square tests in this case because the sample size is large, all expected cell counts are larger than 5. It is not appropriate to use the Mantel- Haenszel test since our categories are not ordinal. According to the results of Chi-Square test and Likelihood Ratio Chi-Square test, the p-values are less than 0.0001, so we reject the null hypothesis of no association. Therefore, we conclude that there is a statistically significant association between treat and allergy. Because the Phi Coefficient and the Cramer's V are -0.3237, the the association is negative and strong.

Column 1 Risk Estimates						
	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.6692	0.0413	0.5884	0.7501	0.5813	0.7492
Row 2	0.9316	0.0233	0.8859	0.9774	0.8697	0.9700
Total	0.7935	0.0258	0.7430	0.8440	0.7376	0.8422
Difference	-0.2624	0.0474	-0.3553	-0.1695		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.3308	0.0413	0.2499	0.4116	0.2508	0.4187
Row 2	0.0684	0.0233	0.0226	0.1141	0.0300	0.1303
Total	0.2065	0.0258	0.1560	0.2570	0.1578	0.2624
Difference	0.2624	0.0474	0.1695	0.3553		
Difference is (Row 1 - Row 2)						

Sample Size = 247

c.
P_C: the rate of having peanut allergy for control group
P_I: the rate of having peanut allergy for intervention group
H_0: $P_C=P_I$
H_A: $P_C>P_I$
$\hat{P}_C$=43/130=0.3308, $\hat{P}_I$=8/117=0.0684, $\hat{P}_C$-$\hat{P}_I$=0.3308-0.0684=0.2624
The risk difference is 0.2624 and 95% CI is 0.1695-0.3553.
Because the 95% CI does not contain zero, we reject the null hypothesis and conclude that the control group has a significantly higher rate of having peanut allergy than the intervention group.

Exercise2

a.

The FREQ Procedure

Table of Species by weight_cat				
Species	**weight_cat**			
Frequency Expected	**1-Light**	**2-Medium**	**3-Heavy**	**Total**
Bream	0 8.6545	13 12.982	21 12.364	34
Perch	20 14.255	17 21.382	19 20.364	56
Roach	8 5.0909	12 7.6364	0 7.2727	20
Total	28	42	40	110

Statistics for Table of Species by weight_cat

Statistic	DF	Value	Prob
Chi-Square	4	29.4210	<.0001
Likelihood Ratio Chi-Square	4	43.4811	<.0001
Mantel-Haenszel Chi-Square	1	23.6234	<.0001
Phi Coefficient		0.5172	
Contingency Coefficient		0.4594	
Cramer's V		0.3657	

Sample Size = 110

H_0: There is no association between species and weight_cat.
H_A: There is association between species and weight_cat.
We can use the chi-square tests in this case since all expected cell counts are larger than 5. It is not appropriate to use the Mantel- Haenszel test because the species are not ordinal. According to the results of Chi-Square test and Likelihood Ratio Chi-Square test, the p-values are less than 0.0001, so we reject the null hypothesis of no association. Therefore, we conclude that there is a statistically significant association between species and weight_cat. Because the Phi Coefficient and the Cramer's V are 0.5172 and 0.3657 respectively, the the association is positive and strong.

b.

The FREQ Procedure

Table of Species by weight_cat			
Species	**weight_cat**		
Frequency Row Pct	**1-Light**	**3-Heavy**	**Total**
Bream	0 0.00	21 100.00	21
Perch	20 51.28	19 48.72	39
Total	20	40	60

Statistics for Table of Species by weight_cat

Column 1 Risk Estimates						
	Risk	**ASE**	**(Asymptotic) 95% Confidence Limits**		**(Exact) 95% Confidence Limits**	
Row 1	0.0000	0.0000	0.0000	0.0000	0.0000	0.1611
Row 2	0.5128	0.0800	0.3559	0.6697	0.3478	0.6758
Total	0.3333	0.0609	0.2141	0.4526	0.2169	0.4669
Difference	-0.5128	0.0800	-0.6697	-0.3559		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	**ASE**	**(Asymptotic) 95% Confidence Limits**		**(Exact) 95% Confidence Limits**	
Row 1	1.0000	0.0000	1.0000	1.0000	0.8389	1.0000
Row 2	0.4872	0.0800	0.3303	0.6441	0.3242	0.6522
Total	0.6667	0.0609	0.5474	0.7859	0.5331	0.7831
Difference	0.5128	0.0800	0.3559	0.6697		
Difference is (Row 1 - Row 2)						

Sample Size = 60

P_B: the proportion in the heavy weight category for Bream group
P_P: the proportion in the heavy weight category for Perch group
H_0: $P_B=P_P$, H_A: $P_B>P_P$
$\hat{P}_B=21/21=1$, $\hat{P}_P=19/39=0.4872$, $\hat{P}_B-\hat{P}_P=1-0.4872=0.5128$
The risk difference is 0.5128 and 95% CI is 0.3559-0.6697.
Because the 95% CI does not contain zero, we reject the null hypothesis and conclude that bream has larger proportion in the heavy weight category than perch.

Exercise3

The ANOVA Procedure

Dependent Variable: lweight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	29.5255102	14.7627551	20.90	<.0001
Error	106	74.8669742	0.7062922		
Corrected Total	108	104.3924844			

R-Square	Coeff Var	Root MSE	lweight Mean
0.282832	14.85247	0.840412	5.658400

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Species	2	29.52551019	14.76275510	20.90	<.0001

a.
H_0: $\alpha_1 = \alpha_2 = \ldots \alpha_k = 0$, all group means are equal. (e.g. the term lweight is useless)
H_A: $\alpha_i \neq 0$ for at least one i. (e.g. the term lweight is useful)
According to the one-way ANOVA, the F-statistics is large (the p-value is less than 0.0001), so we reject the null hypothesis and conclude that lweight is a statistically significant predictor.
The assumption of one-way balance ANOVA is homogeneity of variance (variances are equal among all groups). The one-way ANOVA can be used in balanced or unbalanced cases.

	lweight		
	Mean	**Std**	**N**
Species			
Bream	6.38	0.36	34
Perch	5.46	1.09	56
Roach	4.95	0.53	19

Levene's Test for Homogeneity of lweight Variance ANOVA of Squared Deviations from Group Means					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Species	2	27.3257	13.6629	7.04	0.0013
Error	106	205.7	1.9407		

Welch's ANOVA for lweight			
Source	**DF**	**F Value**	**Pr > F**
Species	2.0000	62.74	<.0001
Error	49.0915		

b. We have unbalanced data, but the one-way ANOVA can be used in unbalanced cases. We use option hovtest in proc anova to check the assumption of homogeneity. According to the Levene's test for equal variance, the p-value is 0.0013 (less than 0.05). So we reject homogeneity and use Welch's Test. The p-value of Welch's test is less than 0.0001, so we reject the null hypothesis and conclude that group means are different for some groups and factor α is useful.

Tukey's Studentized Range (HSD) Test for lweight

Comparisons significant at the 0.05 level are indicated by ***.				
Species Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
Bream - Perch	0.9216	0.4873	1.3559	***
Bream - Roach	1.4317	0.8595	2.0039	***
Perch - Bream	-0.9216	-1.3559	-0.4873	***
Perch - Roach	0.5101	-0.0203	1.0405	
Roach - Bream	-1.4317	-2.0039	-0.8595	***
Roach - Perch	-0.5101	-1.0405	0.0203	

c. According to the indication of ***, we know species comparisons between Bream and Perch, and between Bream and Roach are significant. The mean log(weight) of Bream is 0.9216 larger than the mean log(weight) of Perch. The mean log(weight) of Bream is 1.4317 larger than the mean log(weight) of Roach. The mean log(weight) of Perch is 0.5101 larger than the mean log(weight) of Roach. The interval estimate for the difference of means of log (weight) between Bream and Perch is (0.4873, 1.3559). The interval estimate for the difference of means of log (weight) between Bream and Roach is (0.8595, 2.0039). The interval estimate for the difference of means of log (weight) between Perch and Roach is (-0.0203, 1.0405). We can see that if the interval estimate for the difference of means of log (weight) between two species does not contain zero, the comparison is significant, vice versa.