Chapter 4 Review Sheet with Answers

1. Write down the type of data and for the following hypothesis tests: Two-population proportion test, Goodness of Fit, Categorical Relationship Test, Two-population mean test, ANOVA and Correlation.

2. Write down the assumptions, null and alternative hypothesis and the test statistic for the following hypothesis tests: Two-population proportion test, Two-population mean test, Goodness of Fit, Categorical Association Test, ANOVA, and the Correlation Test.

3. Fill out the following table to interpret the given test statistics.

Test Statistic	Critical Value	Tail	Does the sample data significantly disagree with H_0?	Explain why.
F = 2.174	3.823	Right		
T = −2.556	± 1.96	Two		
χ^2 = 16.87	9.977	Right		
F = 5.339	2.742	Right		
T = 1.349	± 2.576	Two		
χ^2 = 1.883	7.187	Right		

4. Fill out the following table to interpret the given P-value. Assume the data passed the conditions.

P-value	P-value %	Significance Level	Does the sample data significantly disagree with H_0?	Could be Sampling Variability or Unlikely?	Reject H_0 or fail to reject?
0.238		5%			
0.0003		1%			
5.7 x 10^{-6}		10%			
0.441		5%			
0.138		1%			
0		10%			

5. Complete the table by writing the conclusions for the following. Assume the data passed the assumptions.

P-value	Sig Level	Claim	Conclusion
0.238	5%	H_0	
0.0003	1%	H_A	
5.7 x 10^{-6}	10%	H_0	
0.441	5%	H_A	
0.138	1%	H_0	
0	10%	H_A	

6. If we want to see if two quantitative variables are related, what hypothesis test should we use?

7. If we want to see if two categorical variables with multiple options are related, what hypothesis test should we use?

8. If we want to see if categorical and quantitative variables are related, what hypothesis test should we use?

9. Suppose we want to see if the amount of money in peoples' checking accounts is related to city they live in. What hypothesis test should we use? Explain why.

10. Suppose we want to see if the percentage of people that own an Android phone is the same in nine different cities. What hypothesis test should we use? Explain why.

11. Suppose we want to see if the amount of rainfall in areas across Europe is related to the number of fires in those areas. What hypothesis test should we use? Explain why.

12. Suppose we want to see if a person's type of health insurance is related to their education level. What hypothesis test should we use? Explain why.

13. Suppose if we want to see if the population percentage of republicans that drink alcohol is the higher than the population percentage of democrats that drink alcohol.

14. Suppose we want to see if the population average number of movies owned on any streaming platform by people in Los Angeles is lower than the population average number of DVD's and VHS Taped movies owned by people in Los Angeles.

15. An orthopedic surgeon that specializes in knee injuries is wondering if the proportion of knee injuries is the same for the various sports. (This would indicate that the percent of knee injuries is not related to the sport being played.) The surgeon randomly selected knee injuries and made a note of what sport they occurred in. Assume that individual knee injuries were independent of each other.

Statement 1: $At\ least\ one\ is\ \neq\ (Type\ of\ sport\ is\ related\ to\ getting\ a\ knee\ injury)$

Statement 2: $\pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi_5\ (Type\ of\ sport\ is\ not\ related\ to\ getting\ a\ knee\ injury)$

Football	Baseball	Basketball	Soccer	Hockey	Tennis
23	8	14	31	19	5

Chi-Square Goodness-of-Fit Test:
Input: C4 observed counts
Expected frequency = 16.6667

Category	Observed Frequency	Expected Frequency	Contribution to X^2
0	23.0	16.6667	2.4067
1	8.0	16.6667	4.5067
2	14.0	16.6667	0.4267
3	31.0	16.6667	12.3267
4	19.0	16.6667	0.3267
5	5.0	16.6667	8.1667

N	Number of Categories	DOF	Significance	Critical Value	Test statistics	p-Value
100.0	6	5	0.01	15.0863	28.16	$3.3869 \cdot 10^{-5}$

a) What type of data is this?

b) Which statement is the null hypothesis?

c) Which statement is the alternative hypothesis?

d) Which statement is the claim?

e) Check the assumption (condition) that the expected counts (expected frequencies) must all be larger than 5. Does this data pass that assumption?

f) Does the Chi-square test statistic fall in the right tail determined by the critical value?

g) Does the sample data significantly disagree with the null hypothesis? Why or why not?

h) Convert the P-value into a percentage.

i) Convert the significance level into a percentage.

j) Is the P-value lower or higher than the significance level?

k) Could the sample data have occurred by sampling variability or is that unlikely? Why or why not?

l) Should we reject the null hypothesis or fail to reject the null hypothesis?

p) Write the conclusion addressing claim, evidence and assumptions.

16. A forest ranger is looking into incidents of rabies among the animals. He thinks that the type of animal is related to whether or not they have rabies. This data was collected from one random sample of animals. From each animal, the type of animal was noted as well as their rabies status. Assume individual animals were independent of each other.

Statement 1: $Type\ of\ animal\ is\ not\ related\ to\ having\ rabies\ or\ not.$

Statement 2: $Type\ of\ animal\ is\ related\ to\ having\ rabies\ or\ not.$

	Squirrels	Chipmunks	Raccoons
Has Rabies	17	8	7
Does not have Rabies	21	22	20

Chi-Square Test: Contingency Table:

	Squirrels	**Chipmunks**	**Raccoons**	**Total**
Rabies	17.0 (12.8) [1.38]	8.0 (10.11) [0.44]	7.0 (9.09) [0.48]	32.0
No Rabies	21.0 (25.2) [0.70]	22.0 (19.89) [0.22]	20.0 (17.91) [0.25]	63.0
Total	38.0	30.0	27.0	95.0

(expected frequency), [test statistic contribution]

Significance Level	DOF	X^2	Critical value	p-Value
0.05	2	3.4670	5.9915	0.1767

a) What type of data is this?

b) Which statement is the null hypothesis?

c) Which statement is the alternative hypothesis?

d) Which statement is the claim?

e) Check the assumption (condition) that the expected counts (expected frequencies) must all be larger than 5. Does this data pass that assumption?

f) Does the Chi-square test statistic fall in the right tail determined by the critical value?

g) Does the sample data significantly disagree with the null hypothesis? Why or why not?

h) Convert the P-value into a percentage.

i) Convert the significance level into a percentage.

j) Is the P-value lower or higher than the significance level?

k) Could the sample data have occurred by sampling variability or is that unlikely? Why or why not?

l) Should we reject the null hypothesis or fail to reject the null hypothesis?

p) Write the conclusion addressing claim, evidence and assumptions.

17. The following printout and randomized simulations were made from StatKey. Assume the data was random and individual house prices were independent of each other. Use the printouts to answer the following questions. We think the state is related to the price of a home, but want some evidence to back up that statement. Use a significance level of 0.05.

Statement 1: $At\ least\ one\ is\ \neq\ (State\ is\ related\ to\ price\ of\ a\ home.)$

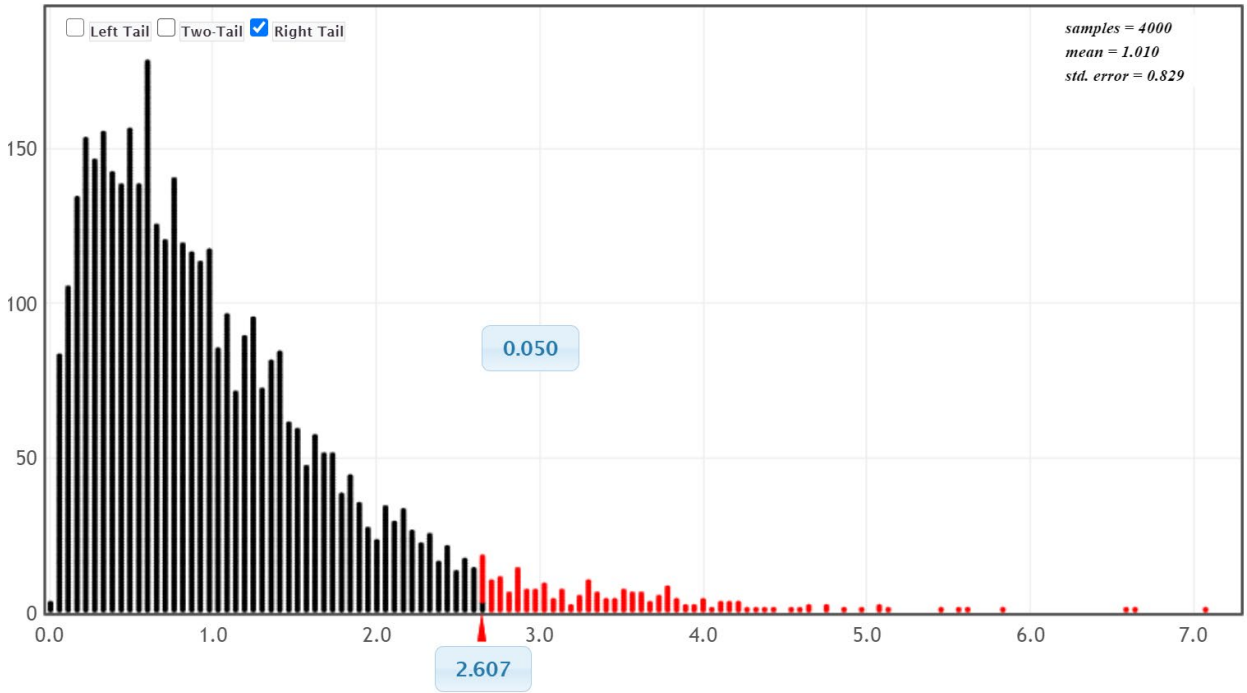
Statement 2: $\mu_1 = \mu_2 = \mu_3 = \mu_4\ (State\ is\ not\ related\ to\ price\ of\ a\ home.)$

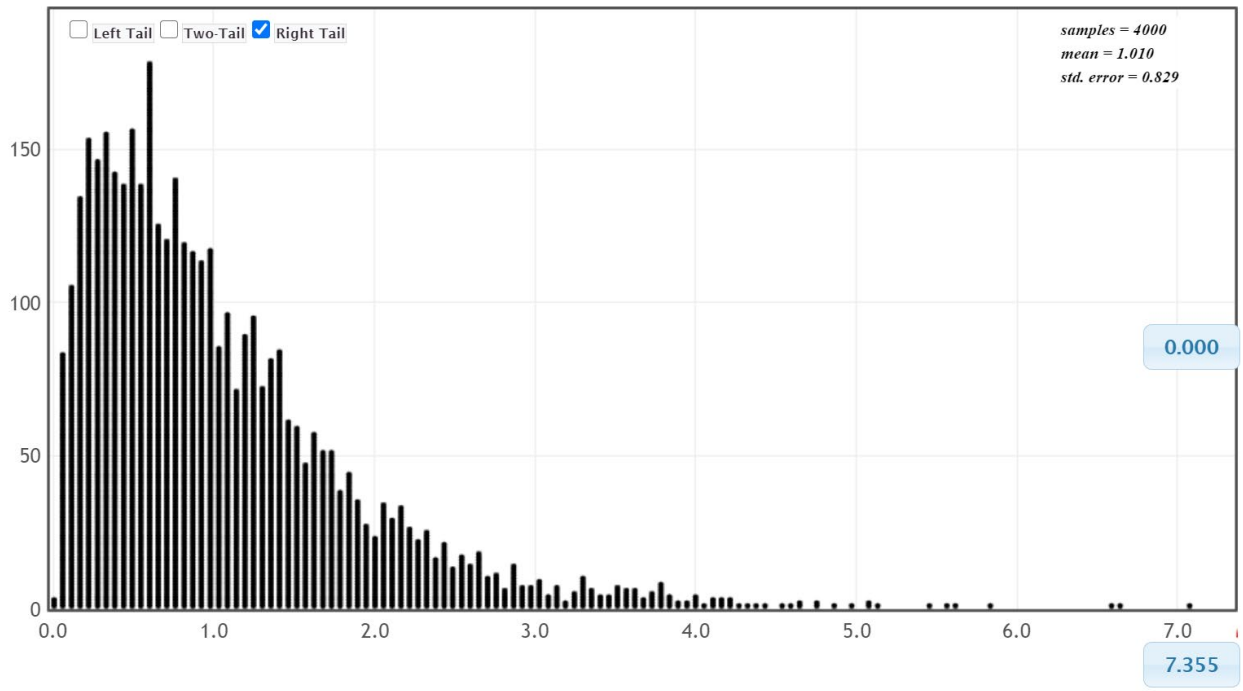
Original Sample

ANOVA Table

$n = 120, F = 7.355$

Statistics	CA	NJ	NY	PA	Overall
Sample Size	30	30	30	30	120
Mean	535.4	328.5	365.3	265.6	373.7
Standard Deviation	269.2	158.0	317.8	137.1	251.0





a) What type of data is this?

b) Which statement is the null hypothesis?

c) Which statement is the alternative hypothesis?

d) Which statement is the claim?

e) Check the assumption (condition) that all of the sample sizes are 30 or above. Does this data pass that assumption?

f) Check the assumption (condition) that the standard deviations are close. That is to say that no standard deviation is more than twice as large as any other. Does this data pass that assumption?

g) What was the F-test statistic?

h) What was the critical value?

i) Does the F-test statistic fall in the right tail determined by the critical value?

j) Does the sample data significantly disagree with the null hypothesis? Why or why not?

k) What was the P-value? Convert the P-value into a percentage.

l) Convert the significance level into a percentage.

m) Is the P-value lower or higher than the significance level?

n) Could the sample data have occurred by sampling variability or is that unlikely? Why or why not?

o) Should we reject the null hypothesis or fail to reject the null hypothesis?

p) Write the conclusion addressing claim, evidence and assumptions.

Answers

1. Write down the type of data and for the following hypothesis tests: Two-population proportion test, Goodness of Fit, Categorical Relationship Test, Two-population mean test, ANOVA and Correlation.

Two-population proportion: Categorical data in order to compare a proportion/percentage from two groups.

Goodness of Fit test: Categorical data in order to compare a proportion/percentage from 3 or more groups.

Categorical Association test: Seeing if a categorical data set is related to a different categorical data set.

Two-population mean: Comparing the same quantitative variable from two groups.

ANOVA: Seeing if a quantitative data set is related to a categorical data set.

Correlation: Seeing if two different quantitative data sets (usually with different units) have linear relationship.

2. Write down the assumptions, null and alternative hypothesis and the test statistic for the following hypothesis tests: Two-population proportion test, Two-population mean test, Goodness of Fit, Categorical Association Test, and ANOVA.

Two-population proportion test:

Assumptions: Random Sample data, Independent individuals, At least 10 success and 10 failures in both groups.

Test Statistic: Z

Ho: $\pi_1 = \pi_2$
Ha: $\pi_1 > \pi_2$ (Right tailed test)

Ho: $\pi_1 = \pi_2$
Ha: $\pi_1 < \pi_2$ (Left tailed test)

Ho: $\pi_1 = \pi_2$
Ha: $\pi_1 \neq \pi_2$ (Two-tailed test)

Two-population mean test:

Assumptions: Random Sample data, Independent individuals,
both samples normal shape or sample size at least 30.

Test Stat: T

Ho: $\mu_1 = \mu_2$
Ha: $\mu_1 > \mu_2$ (Right tailed test)

Ho: $\mu_1 = \mu_2$
Ha: $\mu_1 < \mu_2$ (Left tailed test)

Ho: $\mu_1 = \mu_2$
Ha: $\mu_1 \neq \mu_2$ (Two-tailed test)

Goodness of fit test (3 or more populations, uses a Chi-square right tailed test)

Assumptions: Random Sample data, Independent individuals, All expected counts at least 5.

Test Stat: Chi-square

Ho: $\pi_1 = \pi_2 = \pi_3 = \pi_4$
Ha: at least one ≠

Categorical Association Test: (two different categorical data sets, uses a Chi-square right tailed test)

Assumptions: Random Sample data, Independent individuals, All expected counts at least 5.

Test stat: Chi-square

Ho: Categorical variables are not related.
Ha: Categorical variables are related.

ANOVA: Uses a right tail F test

Assumptions: Random Sample data, Independent individuals, all sample sizes at least 30, no standard deviation more than twice as large as any other.

Test statistic: F

Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (Categorical Variable is not related to quantitative variable.)
Ha: at least one ≠ (Categorical Variable is related to quantitative variable.)

Correlation Test: Can be right tailed, left tailed or two tailed.

Assumptions: Random ordered pair sample data, Independent observations, number of ordered pairs at least 30, scatterplot and r show some linear correlation, no influential outliers in scatterplot, residual plot vs x value evenly spread out, histogram of residuals normal, histogram of residuals centered at zero.

Test Statistic: T

H_0: ρ = 0 or $\beta_1 = 0$ (Two Quantitative variables do not have a linear relationship)

H_A: ρ ≠ 0 or $\beta_1 \neq 0$ (Two Quantitative variables have a linear relationship) (Two-tailed test)

H_A: ρ > 0 or $\beta_1 > 0$ (Two Quantitative variables have a positive (direct) linear relationship) (Right-tailed test)

H_A: ρ < 0 or $\beta_1 < 0$ (Two Quantitative variables have a negative (inverse) linear relationship) (Left-tailed test)

3. Fill out the following table to interpret the given test statistics.

Test Statistic	Critical Value	Tail	Does the sample data significantly disagree with H_0?	Explain why.
F = 2.174	3.823	Right	No. Sample data does not significantly disagree.	Test stat not in tail.
T = −2.556	± 1.96	Two	Yes. Sample data does significantly disagree.	Test stat in tail.
χ^2 = 16.87	9.977	Right	Yes. Sample data does significantly disagree.	Test stat in tail.
F = 5.339	2.742	Right	Yes. Sample data does significantly disagree.	Test stat in tail.
T = 1.349	± 2.576	Two	No. Sample data does not significantly disagree.	Test stat not in tail.
χ^2 = 1.883	7.187	Right	No. Sample data does not significantly disagree.	Test stat not in tail.

4. Fill out the following table to interpret the given P-value.

P-value	P-value %	Significance Level	Does the sample data significantly disagree with H_0?	Could be Sampling Variability or Unlikely?	Reject H_0 or fail to reject?
0.238	23.8%	5%	No. Not significantly disagree	Could be	Fail to reject Ho
0.0003	0.03%	1%	Yes. Significantly disagree	Unlikely	Reject Ho
5.7×10^{-6}	0.00057%	10%	Yes. Significantly disagree	Unlikely	Reject Ho
0.441	44.1%	5%	No. Not significantly disagree	Could be	Fail to reject Ho
0.138	13.8%	1%	No. Not significantly disagree	Could be	Fail to reject Ho
0	0%	10%	Yes. Significantly disagree	Unlikely	Reject Ho

5. Complete the table by writing the conclusions for the following. Assume the data passed the assumptions.

P-value	Sig Level	Claim	Conclusion
0.238	5%	H_0	There is not significant evidence to reject the claim.
0.0003	1%	H_A	The is significant evidence to support the claim.
5.7×10^{-6}	10%	H_0	The is significant evidence to reject the claim.
0.441	5%	H_A	There is not significant evidence to support the claim.
0.138	1%	H_0	There is not significant evidence to reject the claim.
0	10%	H_A	The is significant evidence to support the claim.

6. If we want to see if two quantitative variables are related, what hypothesis test should we use?

Correlation and Regression

7. If we want to see if two categorical variables with multiple options are related, what hypothesis test should we use?

Categorical Association Test

8. If we want to see if categorical and quantitative variables are related, what hypothesis test should we use?

ANOVA Test

9. Suppose we want to see if the amount of money in peoples' checking accounts is related to city they live in. What hypothesis test should we use? Explain why.

ANOVA test since money is quantitative and city is categorical.

10. Suppose we want to see if the percentage of people that own an Android phone is the same in nine different cities. What hypothesis test should we use? Explain why.

Goodness of fit since we are comparing a single percentage in multiple groups.

11. Suppose we want to see if the amount of rainfall in areas across Europe is related to the number of fires in those areas. What hypothesis test should we use? Explain why.

Correlation and Regression since both variables are quantitative.

12. Suppose we want to see if a person's type of health insurance is related to their education level. What hypothesis test should we use? Explain why.

Categorical Association Test since we have two different categorical variables.

13. Suppose if we want to see if the population percentage of republicans that drink alcohol is the higher than the population percentage of democrats that drink alcohol.

Two-population proportion Z-test since we are comparing the same percentage (proportion) between two groups.

14. Suppose we want to see if the population average number of movies owned on any streaming platform by people in Los Angeles is lower than the population average number of DVD's and VHS Taped movies owned by people in Los Angeles.

Two population mean T-test since we are comparing the same quantitative average between two groups.

15. An orthopedic surgeon that specializes in knee injuries is wondering if the proportion of knee injuries is the same for the various sports. (This would indicate that the percent of knee injuries is not related to the sport being played.) The surgeon randomly selected knee injuries and made a note of what sport they occurred in. Assume that individual knee injuries were independent of each other.

Statement 1: $At\ least\ one\ is\ \neq\ (Type\ of\ sport\ is\ related\ to\ getting\ a\ knee\ injury)$

Statement 2: $\pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi_5\ (Type\ of\ sport\ is\ not\ related\ to\ getting\ a\ knee\ injury)$

Football	Baseball	Basketball	Soccer	Hockey	Tennis
23	8	14	31	19	5

Chi-Square Goodness-of-Fit Test:
Input: C4 observed counts
Expected frequency = 16.6667

Category	Observed Frequency	Expected Frequency	Contribution to X^2
0	23.0	16.6667	2.4067
1	8.0	16.6667	4.5067
2	14.0	16.6667	0.4267
3	31.0	16.6667	12.3267
4	19.0	16.6667	0.3267
5	5.0	16.6667	8.1667

N	Number of Categories	DOF	Significance	Critical Value	Test statistics	p-Value
100.0	6	5	0.01	15.0863	28.16	$3.3869 \cdot 10^{-5}$

a) What type of data is this? Categorical Data (Checking single categorical variable in multiple groups)

b) Which statement is the null hypothesis? Statement 2 (= is always null)

c) Which statement is the alternative hypothesis? Statement 1

d) Which statement is the claim? Statement 2 Ho (claiming they are the same)

e) Check the assumption (condition) that the expected counts (expected frequencies) must all be larger than 5. Does this data pass that assumption? Yes. It passes the assumption since all expected counts are 16.6667 which are all greater than 5.

f) Does the Chi-square test statistic fall in the right tail determined by the critical value? Yes. Right tail starts at 15.0863. So the test stat of 28.16 would fall to the right of that on the number line.

g) Does the sample data significantly disagree with the null hypothesis? Why or why not? Yes. Sample data significantly disagrees with Ho since the test stat fell in the tail determined by critical value.

h) Convert the P-value into a percentage. 0.0033869%

i) Convert the significance level into a percentage. 1%

j) Is the P-value lower or higher than the significance level? P-value is lower than significance level.

k) Could the sample data have occurred by sampling variability or is that unlikely? Why or why not? Unlikely to be sampling variability since P-value is low.

l) Should we reject the null hypothesis or fail to reject the null hypothesis? Reject Ho since P-value is low.

p) Write the conclusion addressing claim, evidence and assumptions.

Low P-value, passed all assumptions, claim is Ho

Conclusion: There is significant evidence to reject the claim that the chance of getting a knee injury is the same in all of the sports listed.

16. A forest ranger is looking into incidents of rabies among the animals. He thinks that the type of animal is related to whether or not they have rabies. This data was collected from one random sample of animals. From each animal,

the type of animal was noted as well as their rabies status. Assume individual animals were independent of each other.

Statement 1: $Type\ of\ animal\ is\ not\ related\ to\ having\ rabies\ or\ not.$

Statement 2: $Type\ of\ animal\ is\ related\ to\ having\ rabies\ or\ not.$

	Squirrels	Chipmunks	Raccoons
Has Rabies	17	8	7
Does not have Rabies	21	22	20

Chi-Square Test: Contingency Table:

	Squirrels	**Chipmunks**	**Raccoons**	**Total**
Rabies	17.0 (12.8) [1.38]	8.0 (10.11) [0.44]	7.0 (9.09) [0.48]	32.0
No Rabies	21.0 (25.2) [0.70]	22.0 (19.89) [0.22]	20.0 (17.91) [0.25]	63.0
Total	38.0	30.0	27.0	95.0

(expected frequency), [test statistic contribution]

Significance Level	DOF	χ^2	Critical value	p-Value
0.05	2	3.4670	5.9915	0.1767

a) What type of data is this? Two different categorical data sets (rabies or not, type of animal)

b) Which statement is the null hypothesis? Statement 1 (not related is always null)

c) Which statement is the alternative hypothesis? Statement 2 (related is always alternative)

d) Which statement is the claim? Statement 2 (says they think they are related)

e) Check the assumption (condition) that the expected counts (expected frequencies) must all be larger than 5. Does this data pass that assumption? Yes. It passes the assumption. The expected counts are 12.8, 10.11, 9.09, 25.2, 19.89, and 17.91 which are all greater than 5.

f) Does the Chi-square test statistic fall in the right tail determined by the critical value?

No. The right tail starts at 5.9915. The test stat 3.467 would not fall to the right of that on the number line.

g) Does the sample data significantly disagree with the null hypothesis? Why or why not?

No. The sample data does not significantly disagree with the null hypothesis since the test stat did not fall in the tail determined by the critical value.

h) Convert the P-value into a percentage. 17.67%

i) Convert the significance level into a percentage. 5%

j) Is the P-value lower or higher than the significance level? The P-value is higher than the significance level.

k) Could the sample data have occurred by sampling variability or is that unlikely? Why or why not? Since the P-value is high, the sample data could have occurred by sampling variability.

l) Should we reject the null hypothesis or fail to reject the null hypothesis? Fail to reject Ho since we have a high P-value.

p) Write the conclusion addressing claim, evidence and assumptions.

Met all assumptions, High P-value, Claim is Ha

There is significant evidence to support the claim that having rabies or not is related to the type of animal.

17. The following printout and randomized simulations were made from StatKey. Assume the data was random and individual house prices were independent of each other. Use the printouts to answer the following questions. We think the state is related to the price of a home, but want some evidence to back up that statement. Use a significance level of 0.05.

Statement 1: $At\ least\ one\ is\ \neq\ (State\ is\ related\ to\ price\ of\ a\ home.)$

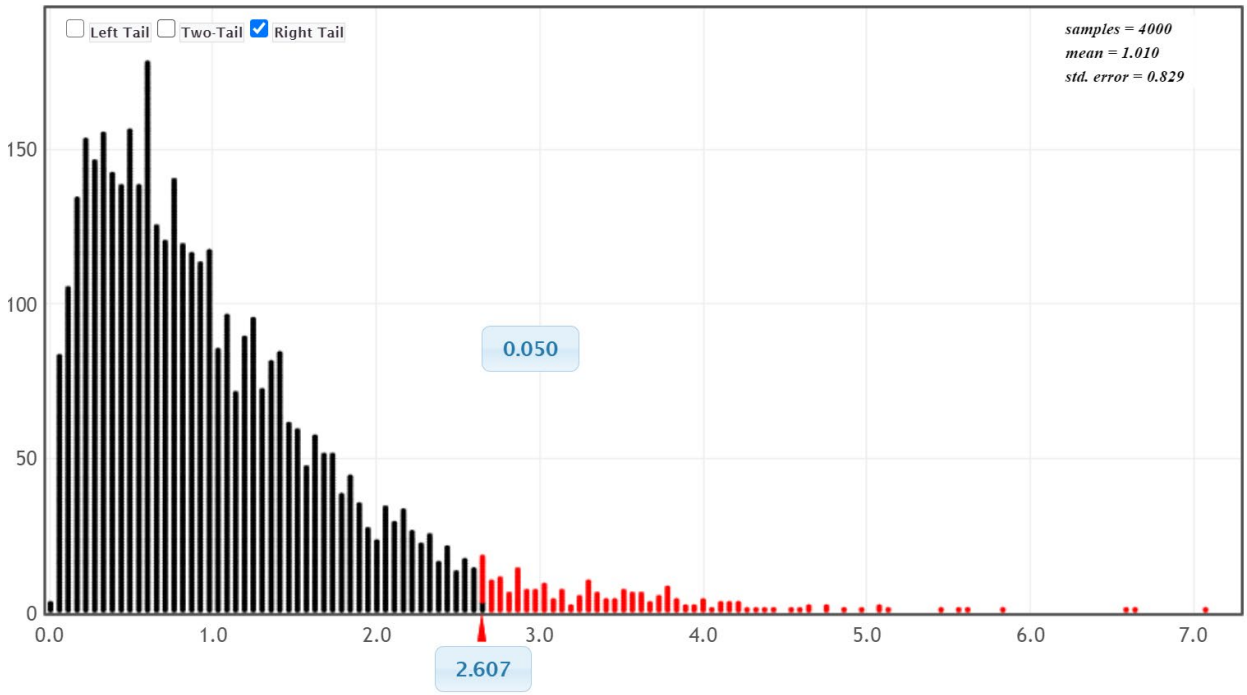
Statement 2: $\mu_1 = \mu_2 = \mu_3 = \mu_4\ (State\ is\ not\ related\ to\ price\ of\ a\ home.)$

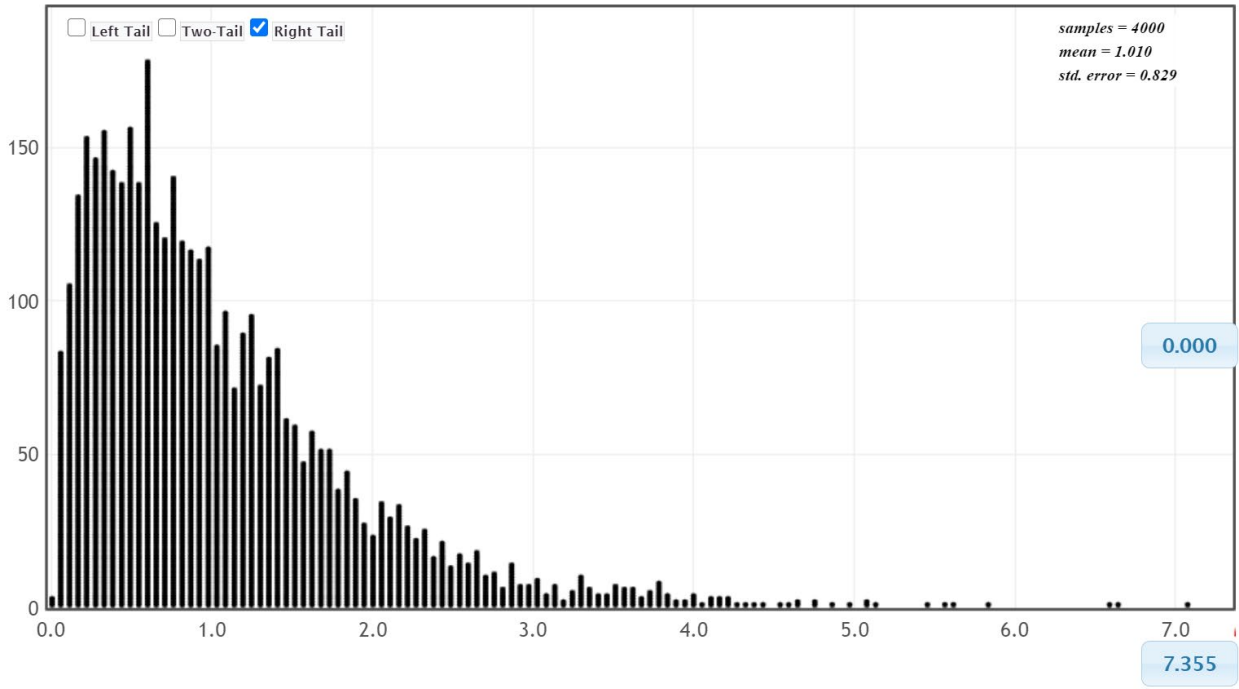
Original Sample

ANOVA Table

$n = 120, F = 7.355$

Statistics	CA	NJ	NY	PA	Overall
Sample Size	30	30	30	30	120
Mean	535.4	328.5	365.3	265.6	373.7
Standard Deviation	269.2	158.0	317.8	137.1	251.0





a) What type of data is this? State is categorical and Home price is quantitative.

b) Which statement is the null hypothesis? Statement 2 (=, not related is always null)

c) Which statement is the alternative hypothesis? Statement 1 (not equal, related always alternative)

d) Which statement is the claim? Statement 1 (they think they are related)

e) Check the assumption (condition) that all of the sample sizes are 30 or above. Does this data pass that assumption? Yes. It passes the assumption since all of the sample sizes were 30.

f) Check the assumption (condition) that the standard deviations are close. That is to say that no standard deviation is more than twice as large as any other. Does this data pass that assumption? No. It fails this assumption. The standard deviation of NY (317.8) is more than double that of PA or NJ.

g) What was the F-test statistic? 7.355

h) What was the critical value? 2.607 (corresponds to 5% sig level in tail)

i) Does the F-test statistic fall in the right tail determined by the critical value? Yes. The F-test stat 7.355 falls to the right of 2.607 on the number line.

j) Does the sample data significantly disagree with the null hypothesis? Why or why not? Yes. The sample data significantly disagrees with the null hypothesis since the F-test stat fell in the tail.

k) What was the P-value? Convert the P-value into a percentage. P-value = 0 = 0% (tail % corresponding to F-test stat)

l) Convert the significance level into a percentage. 5%

m) Is the P-value lower or higher than the significance level? P-value (0%) is lower than the 5% significance level.

n) Could the sample data have occurred by sampling variability or is that unlikely? Why or why not? Unlikely to have occurred by sampling variability since the P-value was low.

o) Should we reject the null hypothesis or fail to reject the null hypothesis? Reject Ho since P-value was low.

p) Write the conclusion addressing claim, evidence and assumptions.

Did not pass all assumption, Low P-value, Claim is Ha

Since the data did not pass all of the assumptions, there is not significant evidence to support the claim that the state is related to the home price.