MTF Scenario

Setup

In [117]:
```
# imports
import pandas as pd
import matplotlib.pyplot  as plt
import matplotlib as mpl
import math
import numpy as np
import seaborn as sns
```

In [200]:
```
# read indemnity data
data_indemnified = pd.read_excel("Scenarios\scenario_1.xlsx",
                                 sheet_name = 0,
                                 header = 0,
                                 # update column names
                                 names = ["year", "month_ID", "month_name",
                                        "county", "crop", "cause",
                                        "num_indem", "acres", "liability",
                                        "premium", "subsidy", "indemnity",
                                        "loss_ratio"],
                                 converters ={
                                    # make all crop names lowercase
                                    'crop': lambda x: x.lower()
                                 })

# read policy data
data_policies = pd.read_excel("Scenarios\scenario_1.xlsx",

                              sheet_name = 1,
                              header = 0,
                              # update colmn names
                              names = ["year", "crop", "county",
                                      "num_sold", "num_indem",
                                      "acres", "liabilities", "premium",
                                       "subsidy", "indemnity", "loss_ratio"],
                                converters ={
                                    # make all crop names lowercase
                                    'crop': lambda x: x.lower()
                                })

# read temperature data
data_temp = pd.read_excel("Scenarios\scenario_1.xlsx",
                          sheet_name = 2,
                          header = 0,
                         names = ["year", "month_ID", "month_name",
                                 "value", "anomaly"])
```

Question 1

Determine the loss (indemnity) per policy sold for almonds and for grapes in 2018?

In [54]:
```
# get almond policy in 2018
almond_2018 = data_policies.loc[(data_policies["year"] == 2018) & (data_polici
es["crop"] == "almonds")].iloc[0]

# find average indemnity
print("Almond loss per policy: ", (almond_2018.indemnity / almond_2018.num_sol
d))

# get grape policy in 2018
grapes_2018 = data_policies.loc[(data_policies["year"] == 2018) & (data_polici
es["crop"] == "grapes")].iloc[0]

# find average indemnity
print("Grapes loss per policy: ", (grapes_2018.indemnity / grapes_2018.num_sol
d))
```

```
Almond loss per policy:  21171.47646219686
Grapes loss per policy:  3040.013779527559
```

Question 2

Which crop had the higher annual loss per acre of all the policies sold in 2018?

In [60]:
```
print("Almond loss per acre: ", (almond_2018.indemnity / almond_2018.acres))
print("Grapes loss per acre: ", (grapes_2018.indemnity / grapes_2018.acres))
```

```
Almond loss per acre:  99.67028871144302
Grapes loss per acre:  28.02389874336524
Almonds have have a higher loss per acre.
```

Almonds have have a higher loss per acre.

Question 3

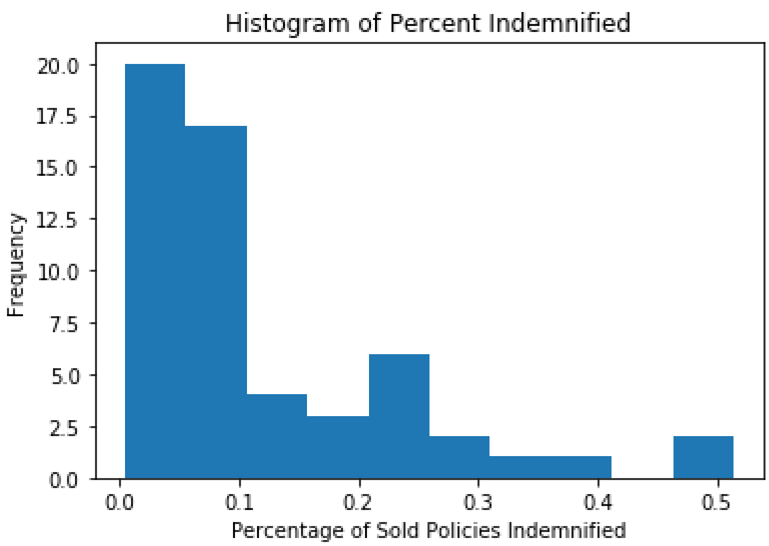
Which value, Loss per Policy Sold, or Loss per Acre, can give Omega Insurance a better understanding of the overall value of providing insurance policies to farmers in Acme County? Explain why.

Loss per acre gives a better understanding because the value of an insurance policy scales with the size of the farm that it covers.

Question 4

Create a histogram of the percentage of policies indemnified each year (the percentage of the policies sold that had a loss). Describe the shape of the histogram and explain logically why you think it is this shape.

In [66]:
```
# get percentage column
percentages = data_policies['num_indem'] / data_policies['num_sold']

# plot histogram
plt.hist(percentages)
plt.title("Histogram of Percent Indemnified")
plt.xlabel("Percentage of Sold Policies Indemnified")
plt.ylabel("Frequency")
plt.show()
```



The data appear to be skewed right, which indicates that for the majority of years, a small percentage (<10%) of policies were indemnified. This is expected, as we would expect that usually, most farms would not experience a loss. However, there are outliers, such as the year were almost half of all farms were indemnified. This indicates that during some years, an event may impact many farms at once, such as a flood or drought.

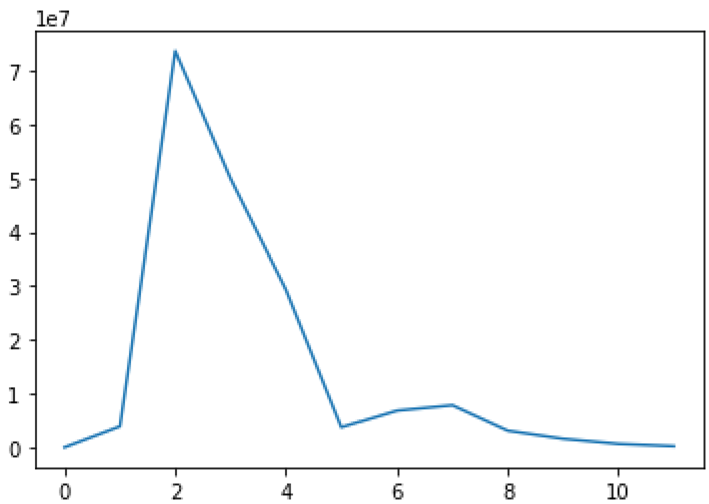
Question 5

Determine which month has the highest average agricultural loss in Acme county? Explain why you think this is the case.

In [76]:

```
month_losses = []

# iterate through months
months = range(0, 12)
for month in months:

    # gather month data
    data_bymonth = data_indemnified[data_indemnified["month_ID"] == month]

    # calculate loss
    total_loss = data_bymonth["indemnity"].sum()

    # collect loss
    month_losses += [total_loss]

    print("Month #", month, " Loss: ", total_loss, sep="")

plt.plot(month_losses)
plt.show()
```

```
Month #0 Loss: 389.4
Month #1 Loss: 3888358.4
Month #2 Loss: 73632651.98
Month #3 Loss: 49884767.870000005
Month #4 Loss: 29294589.84
Month #5 Loss: 3733096.9699999997
Month #6 Loss: 6833176.6899999995
Month #7 Loss: 7832170.91
Month #8 Loss: 3085194.1999999997
Month #9 Loss: 1583608.23
Month #10 Loss: 632557.36
Month #11 Loss: 206967.13
```



March is the season with the greatest average agrilcultural losses. This may be because March marks the beginning of planting season and thus is when plants are most suceptible to different forms of damage.

Question 6

Determine which cause of loss for agricultural claims was the largest in Acme county from 1991 to 2018?

In [82]:
```
cause_losses = []

# find all causes
causes = data_indemnified["cause"].unique()

# iterate through causes
for cause in causes:

    # gather month data
    data_bycause = data_indemnified[data_indemnified["cause"] == cause]

    # calculate loss
    total_loss = data_bycause["indemnity"].sum()

    # collect loss
    cause_losses += [[cause, total_loss]]

    print(cause, "Losses:", total_loss)
```

```
Hail Losses: 15383456.21
Cold Wet Weather Losses: 17237942.86
Excess Moisture Losses: 64527586.629999995
Heat Losses: 36033267.45
Wind/Excess Wind Losses: 5705580.949999999
Cold Winter Losses: 919792.4
Frost Losses: 20754693.48
Flood Losses: 10994.9
Freeze Losses: 16605613.889999999
Insects Losses: 248135.9
Failure of Irrigation Supply Losses: 6343898.7
Failure of Irrigation Equipment Losses: 162249.4
```

Excess moisture was the largest cuase of loss for those claims.

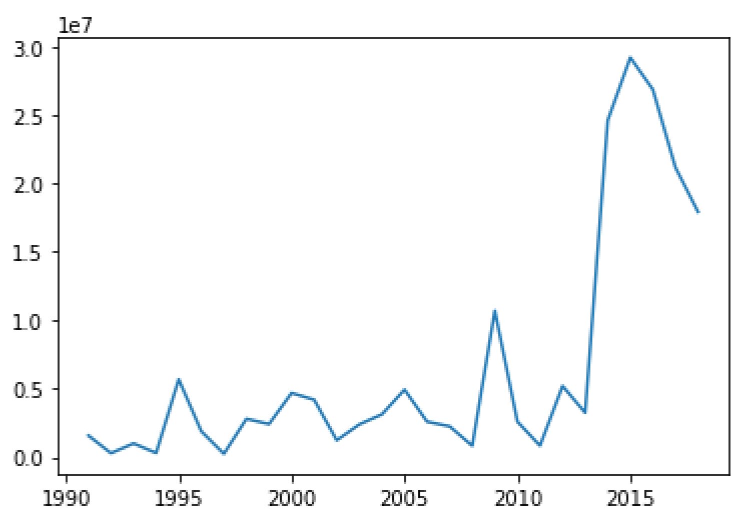
Question 7

What was the likelihood that the FCIC had a total annual loss of greater than $10,000,000 in Acme county? In which years were those losses?

In [90]:

```
yearly_losses = []
years_greater = []

# iterate years
years = range(1991, 2019)
for year in years:

    # subset year
    data_byyear = data_policies[data_policies["year"] == year]

    # get total losses
    total_losses = data_byyear["indemnity"].sum()
    yearly_losses += [total_losses]

    # check if greater than 10M
    if(total_losses > 10000000):
        years_greater += [year]
        print(year)

# calculate likelihood
print("Likelihood:", len(years_greater)/(2019-1991))

plt.plot(years, yearly_losses)
plt.show()
```

```
2009
2014
2015
2016
2017
2018
Likelihood: 0.21428571428571427
```



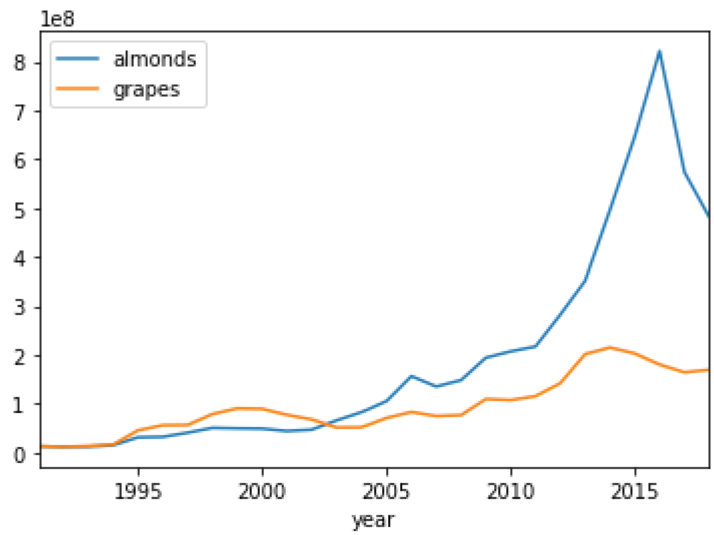
Question 8

What are some reasons that the FCIC high-loss years are concentrated in this most recent decade? Provide one or more reasons that the loss increased in these years.

There has been a large increase in the total amount of liabilities issued by the FCIC in the recent decade. This trend can be observed in the graph below.

In [108]:
```
# plot liabilities by crop group
liability_plot = data_policies.copy()
liability_plot.set_index('year', inplace = True)
liability_plot.groupby("crop")["liabilities"].plot(legend = True)
```

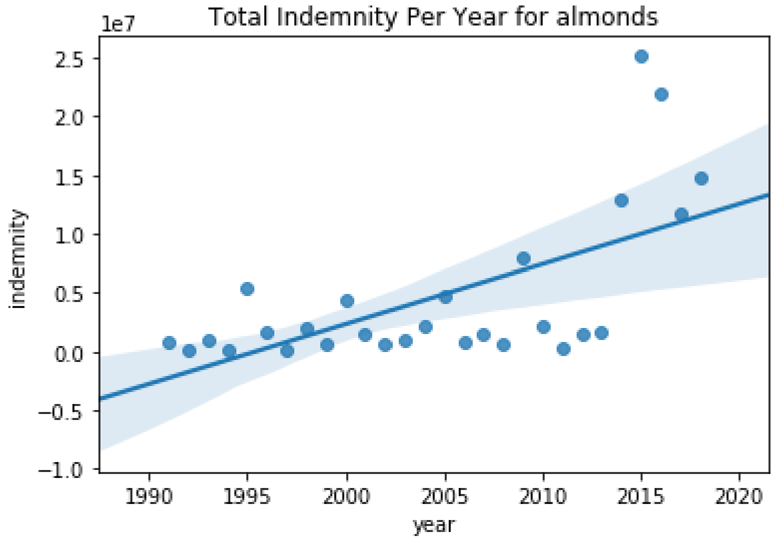
Out[108]:
```
crop
almonds     AxesSubplot(0.125,0.125;0.775x0.755)
grapes      AxesSubplot(0.125,0.125;0.775x0.755)
Name: liabilities, dtype: object
```

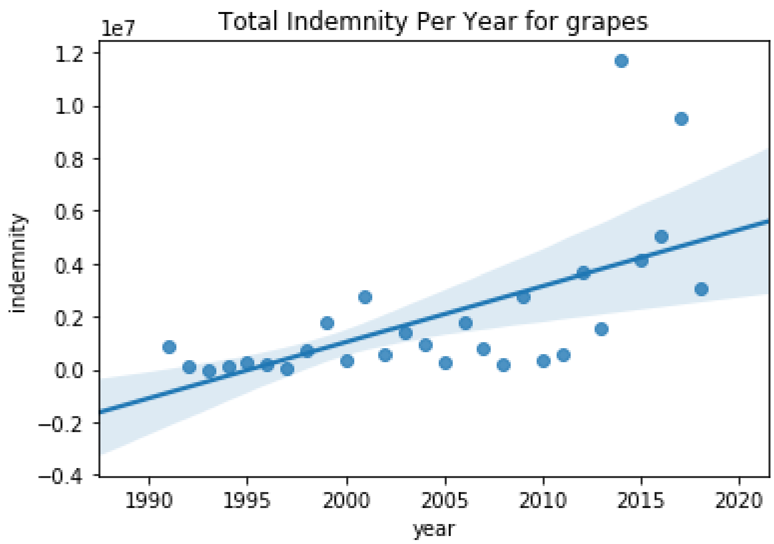


Question 9

Graph the total indemnity per year from 1991 to 2018. Create a linear regression trendline for almonds and for grapes.

In [184]:
```
# for each crop group
for key, grp in data_policies.groupby(['crop']):

    # plot graph with linreg
    sns.regplot(x='year',y='indemnity',label=key,data=grp, fit_reg=True)
    plt.title("Total Indemnity Per Year for "+key)
    plt.show()
```





Question 10

What percent of the total liability between 1991-2018 occurred in the last 5 years (2014-2018). What are some benefits or negative aspects of using just the past 5 years of data to project potential indemnities in the future 2019.

In [133]:
```
# get total liability
total_liability = data_policies['liabilities'].sum()

# get last five year liability
lastfive_liability = data_policies[data_policies['year'] >= 2014]['liabilitie
s'].sum()

# print percentage
print(lastfive_liability/total_liability)
```

```
0.494727260708278
```

Recent data can better reflect the policy and financial environment of the company which can lend greater credibiltiy to predictions. However, using only recent data may also overlook long-term trends in liability growth.

Question 11

In question #3, you determined there was a difference in the usefulness of information being provided in analyzing the annual loss per policy versus loss per acre. What difference in the usefulness of information do you see in analyzing the annual loss per liability ratio? What does the annual loss per liability tell you that the other ratios do not?

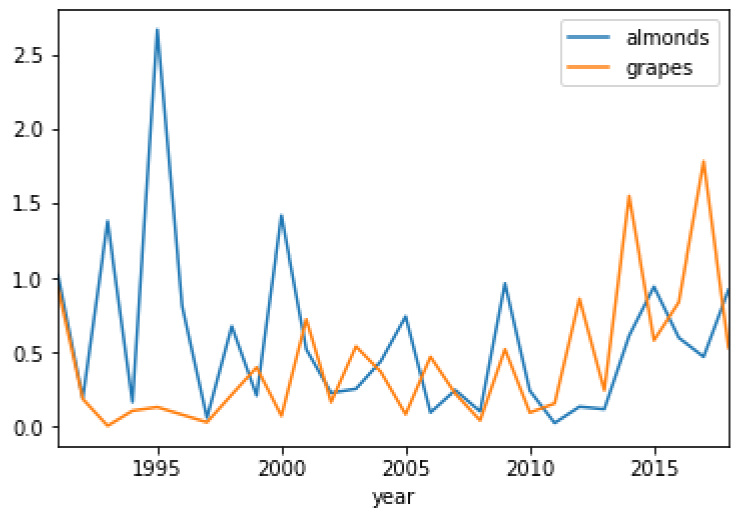
The loss per liability offers a unique calculation that can account for not only the acres covered, but also the value of those acres. This may offer even better insight, as land can have different per-acre value if it is employed for different purposes.

Question 12

Graph the annual loss (indemnity) to liability ratios for grapes and almonds from 1991 to 2018 (for all policies). Compare the variability in these ratios for the first 10 years (1991 to 2000) with the last ten years (2009-2018)? Is there a difference in variability (identified by the Standard Deviation) in the ratios between the first 10 years and last 10 years for either grapes or almonds? Using the data provided explain why it might be the case that some of the data has a higher variability than others.

In [135]:
```
# plot loss ratios by crop group
ratios_plot = data_policies.copy()
ratios_plot.set_index('year', inplace = True)
ratios_plot.groupby("crop")["loss_ratio"].plot(legend = True)
```

Out[135]:
```
crop
almonds    AxesSubplot(0.125,0.125;0.775x0.755)
grapes     AxesSubplot(0.125,0.125;0.775x0.755)
Name: loss_ratio, dtype: object
```



In [148]:
```
# for each crop group
for key, grp in data_policies.groupby(['crop']):

    # 1991-2000
    first_ten = grp[grp['year'] < 2001]

    # calculate SD
    ft_SD = np.std(first_ten['indemnity']/first_ten['liabilities'])

    # 2009-2018
    last_ten = grp[grp['year'] >= 2009]

    # calculate SD
    lt_SD = np.std(last_ten['indemnity']/last_ten['liabilities'])

    print("SD in first 10 for", key, "crop:", ft_SD)
    print("SD in last 10 for", key, "crop:", lt_SD)
```

```
SD in first 10 for almonds crop: 0.049807706127372606
SD in last 10 for almonds crop: 0.013734090949714539
SD in first 10 for grapes crop: 0.018473063937323345
SD in last 10 for grapes crop: 0.01784415943927369
```

There is a large difference in standard diaviation for almonds and a much smaller difference for grapes. One explanation is that there are a greater number of policies as time passes. By the law of large numbers, we can expect sampled values to better predict the actual value, thus decreasing the standard deviation.

Question 13

Between 1991 and 2018, what is the likelihood of having a positive monthly max temperature anomaly?

In [170]:
```
# count total anomalies
num_anoms = len(data_temp.index)

# count positive anomalies
num_posanoms = len(data_temp[data_temp['anomaly'] > 0])

prob_pos_anom = num_posanoms/num_anoms

# calculate proporiton
print("Likelihood of postiive anomaly: ", prob_pos_anom)
```

```
Likelihood of postiive anomaly:  0.6517857142857143
```

For an arbitrary month from 1991 to 2018, the likelihood that it will have a postiive anomaly is 65.18%.

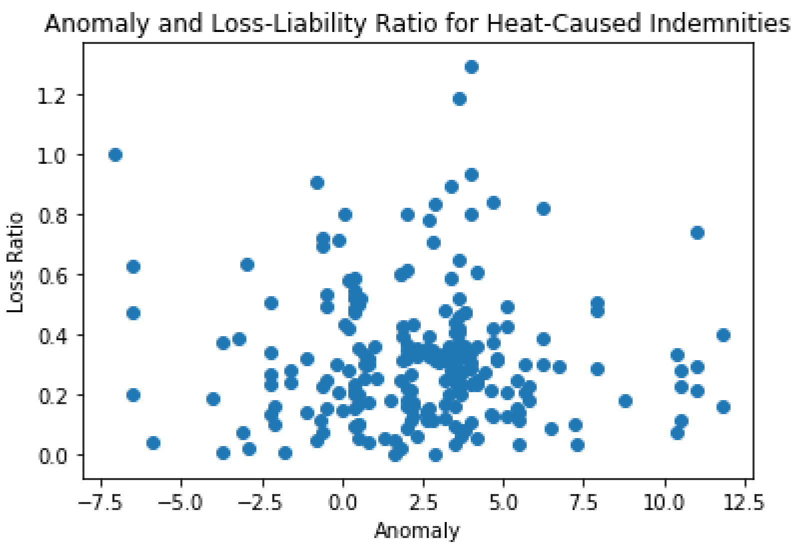
Question 14

Create a scatterplot of the relationship between the max temperature anomaly and the loss to liability ratio just for the losses due to “Heat”. Is there a strong correlation between these values? Should a linear regression be used on this data to predict future values? Explain why or why not.

In [215]:

```
# get heat data
data_heat = data_indemnified[data_indemnified['cause'] == 'Heat']
data_heat['date'] = data_heat['month_ID'].map(str)+ '-' +data_heat['year'].map
(str)
data_heat['date'] = pd.to_datetime(data_heat['date'], format='%m-%Y').dt.strft
ime('%m-%Y')

# get temp data
temp_plot = data_temp.copy()
temp_plot['date'] = temp_plot['month_ID'].map(str)+ '-'+temp_plot['year'].map(
str)
temp_plot['date'] = pd.to_datetime(temp_plot['date'], format='%m-%Y').dt.strft
ime('%m-%Y')


data_merged = pd.merge(data_heat, temp_plot, how='inner', on = 'date')

plt.scatter(x=data_merged['anomaly'], y=data_merged['indemnity']/data_merged[
'liability'])
plt.title("Anomaly and Loss-Liability Ratio for Heat-Caused Indemnities")
plt.xlabel("Anomaly")
plt.ylabel("Loss Ratio")
plt.show()
```

```
C:\Users\thepe\Anaconda3\lib\site-packages\ipykernel_launcher.py:3: SettingWi
thCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/st
able/indexing.html#indexing-view-versus-copy
  This is separate from the ipykernel package so we can avoid doing imports u
ntil
C:\Users\thepe\Anaconda3\lib\site-packages\ipykernel_launcher.py:4: SettingWi
thCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/st
able/indexing.html#indexing-view-versus-copy
  after removing the cwd from sys.path.
```



There is not a strong correlation between these values. Linear regression should not be used on the data because there the data does not meet the Straight Enough Condition.

Question 15

Using the monthly Max Temperature anomalies and indemnity values for losses due to heat, what is the likelihood of having an indemnity due to heat on grape crops if the anomaly is less than or equal to 0? What is the likelihood if the anomaly is positive? Use this information to calculate the overall likelihood of having a monthly loss due to heat for grapes?

In [168]:

```
# get neg anomalies
total_neg = data_temp[data_temp['anomaly'] <= 0]
indem_neg = data_merged[data_merged['anomaly'] <= 0]
prob_neg = len(indem_neg.index) / len(total_neg.index)

# get pos anomalies
total_pos = data_temp[data_temp['anomaly'] > 0]
indem_pos = data_merged[data_merged['anomaly'] > 0]
prob_pos = len(indem_pos.index) / len(total_pos.index)

# get total prob
prob_overall = len(data_merged.index) / len(data_temp.index)

print("Probability for negative anomaly:", prob_neg)
print("Probability for positive anomaly:", prob_pos)
print("Probability overall:", prob_overall)
```

```
Probability for negative anomaly: 0.3333333333333333
Probability for positive anomaly: 0.817351598173516
Probability overall: 0.6488095238095238
```

Question 16

How could this information help Omega Insurance better predict or plan for future indemnities due to heat?

If we know that positive anomalies are associated with greater losses due to heat, then we can use predictions of future positive anomalies to inform projections of future indemnities.

Question 17

If the likelihood of having a positive max temperature anomaly increased by 0.05 above what you calculated previously, what would happen to the likelihood of having a monthly loss due to heat for grapes?

In [172]:

```
# get new probs
new_pos_prob = prob_pos_anom + 0.05
new_neg_prob = 1-new_pos_prob

# calculate new likelihood
new_likelihood = new_pos_prob * prob_pos + new_neg_prob * prob_neg

# print results
print("The new likelihood would be", new_likelihood)
print("This is an increase of", new_likelihood - prob_overall)
```

```
The new likelihood would be 0.673010437051533
This is an increase of 0.024200913242009126
```

Question 18

To help measure the size or severity of a loss on crops, we can measure the indemnity per acre of policies that had a loss. In 2018, what was the indemnity per acre of the policies that had a loss (just those policies indemnified) for losses due to heat only?

In [183]:
```
# get 2018 heat losses
data_heat18 = data_heat[data_heat['year'] == 2018]

# get sums
total_indem = data_heat18['indemnity'].sum()
total_acres = data_heat18['acres'].sum()

# calculate average
print("Average indemnity per acre:", total_indem/total_acres)
```

```
Average indemnity per acre: 626.5293047746101
```

Question 19

Expected value is a way to calculate how much, over time one would expect to pay. It is the value of the payment times the likelihood of it happening. In 2020 the FCIC expects to insure 110,000 acres of grapes. What is the difference in the expected value of the indemnity due to heat if the likelihood of a positive Max Temperature Anomaly remained what it has been between 1991-2018, versus if the likelihood increased by 0.05 as you noted above?

In [193]:
```
# get monthly avg indemnity/acre for pos
avg_pos = ((indem_pos['indemnity']/indem_pos['acres']).sum()) / len(indem_pos.
index)

# get monthly avg indemnity/acre for neg
avg_neg = ((indem_neg['indemnity']/indem_neg['acres']).sum()) / len(indem_neg.
index)

# calculate expected values
normal_expected = 110000 * (prob_pos_anom * avg_pos + (1 - prob_pos_anom) * av
g_neg)
increased_expected = 110000 * (new_pos_prob * avg_pos + new_neg_prob * avg_neg
)

print("Normal likelihood:", normal_expected)
print("+5% likelihood:", increased_expected)
print("Difference:", increased_expected - normal_expected)
```

```
Normal likelihood: 51848179.96313401
+5% likelihood: 52529181.7241633
Difference: 681001.7610292882
```

Question 20

What are the premiums per acre for grapes and for almonds for all policies (not just indemnified) in 2018? Is the premium per acre different for all policies versus just the policies that are indemnified? Explain why this might be.

In [203]:
```
# get 2018 policies
data_indemnified2018 = data_indemnified[data_indemnified['year'] == 2018]
data_policies2018 = data_policies[data_policies['year'] == 2018]

# calculate premium/acre
print("Indemnified policies", data_indemnified2018['premium'].sum() / data_ind
emnified2018['acres'].sum())
print("All policies:", data_policies2018['premium'].sum() / data_policies2018[
'acres'].sum())
```

```
Indemnified policies 163.46123890652785
All policies: 84.35132256346529
```

The premium for indemnified policies is almost twice as high as the premium for all policies. This could be an attempt by Omega Insurance to recoup losses by charging higher for risky policies.

Question 21

In the data provided, there are two Loss Ratio columns, one in the Indemnified Policies by Month sheet and one in the All Policies Annual Summary sheet. Explain the difference between these two columns.

The loss-ratio column of the "by Month" sheet identifies the loss-ratio of only those policies which have been indemnified. Of course, these values will be high, as these are only for policies which have incurred a loss. The "All Policies" sheet includes the policies which have *not* been indemnified and finds the overall loss ratio. As expected, these values are usually much lower than the previous set of values.

Question 22

Using the loss ratio for all policies, not just the indemnified ones, how likely was it between 1991 and 2018 that for the grape farmer policies the FCIC lost more money to indemnity than it brought in from the premiums?

In [206]:
```
# find all loss ratios > 1
losing_years = data_policies[(data_policies['crop'] == "grapes") & (data_polic
ies['loss_ratio'] > 1)]

# calculate probability
print("Probability of loss:", len(losing_years.index) / (2018-1990))
```

```
Probability of loss: 0.07142857142857142
```

Question 23

Insurance companies, like all companies, have other expenses to be paid in order to maintain a healthy business. Explain how Omega Insurance might price the premium for their new crop insurance policies to make sure they can cover these other operating expenses.

Omega Insurance must price the premiums such that they can cover the cost of indemnities and operation while also retaining profit. Therefore, they may use expected indemnity and predicted operating expenses as a baseline for premiums.

Question 24

The Omega Insurance Co. CEO is considering providing crop insurance policies to farmers in Acme county, but only wants to insure one type of crop to start. The CEO also wants to minimize the risk for the company on how much they might have to pay due to loss from the policies. Which crop would you recommend Omega pursue, insurance policies for Almond farmers, or insurance policies for grape farmers? Explain why.

In [207]:
```
# get grape and almond data for the past five years
grape_policies = data_policies[(data_policies['year'] >= 2014) & (data_policie
s['crop'] == 'grapes')]
almond_policies = data_policies[(data_policies['year'] >= 2014) & (data_polici
es['crop'] == 'almonds')]

# get average loss ratios
grape_ratio = grape_policies['loss_ratio'].mean()
almond_ratio = almond_policies['loss_ratio'].mean()

print("Grape ratio:", grape_ratio)
print("Almond ratio:", almond_ratio)
```

```
Grape ratio: 1.0581102
Almond ratio: 0.7114368000000001
```

For the past five years, the grape policies appeal to have a much greater average loss ratio than the almond policies. This suggest that they may be a safer investment.

Question 25

Who do you think is subsidizing the premiums for crop insurance and why?

The government subsidizes the premiums for the farmers in order to not only aid in rural development but also influence the cost and supply of agricultural goods. That means that the farmers have more money to improve their farming efficiency and encourage more farmers to plant. This helps to stabilize the markets and regulates the economy by providing incentives for the farmers to plant. It also helps the insurance company keep premiums low.

Question 26

Taking the subsidy into account, how many years did the FCIC lose more from indemnities on grape farms than it took in from the portion of the premiums that were not covered by a subsidy?

In [212]:

```
# find years with loss without subsidies
real_loss = data_policies[(data_policies['crop'] == 'grapes') &
                          (data_policies['indemnity'] +data_policies['subsidy'
] > data_policies['premium'])]

# count
print("This was true in", len(real_loss.index), "years.")
```

```
This was true in 15 years.
```

Question 27

Besides what is provided in the spreadsheets, what other information would be helpful in projecting future crop losses and analyzing the potential for Omega Insurance to provide crop insurance policies in Acme county in the future?

1. Humidity - trends in humidity level can help predict losses due to excess moisture.
2. Percipitation - changes in the inches of percipitation annual can predict losses due to flooding and hail.
3. Insect Patterns - variation in the level of insects (due to, for example, migration patterns) can influence losses due to insects.

Question 28

If you were representing the government of Acme county and were faced with the knowledge that the likelihood of having a positive temperature anomaly were going to increase to 75% by 2030 what recommendations, incentives, or new policies could you make to help the farming community in your county?

```
In [213]: increase75_expected = 110000 * (.75 * avg_pos + .25 * avg_neg)
          print("Normal likelihood:", normal_expected)
          print("+75% likelihood:", increase75_expected)
          print("Difference:", increase75_expected - normal_expected)
```

```
Normal likelihood: 51848179.96313401
+75% likelihood: 53185861.99372726
Difference: 1337682.0305932462
```

Our team would recommend two policy changes.

First, we would increase subsidies for companies seeking to insure farms. If we assume that 110,000 acres would be covered in 2030, than we would expect an additional $1,337,682 in indemnities for insurance companies. As we found previously, subsidies play a large role in maintaining competitive policy rates for insurance companies.

Second, in order to hedge against the damages brought about by climate change, we believe that local governments like the Acme county government should incentivize farmers to adopt green practices. By using innovative farm equipment, better seeds, green energy and climate-smart practices, U.S. farmers and ranchers are producing mcan use less water, protect against erosion and conserve soil. Such incentives can be implemented via tax credits and loan gaurantees.