Risk of death due to breast cancer and its relation to malignant

- Group 8:
- Xiao Xia Bian: 1000724219
- Jad Dimachkieh: 1000374114
- Shuntaro Kino: 997095806
- Ruoxin Liu: 1000543735
- Xian Qi: 998248070
- Jiawei Tian: 1000026517

Introduction

- Breast cancer is the second most common type of cancer in North America. Mostly found among females. According to the National Cancer Institute, the proportion of male diagnosed with Breast cancer is only 1% of the total number of patients.
- Moreover, Breast cancer incidence rates are decreasing due to new advance treatment. However, the US still is expecting to record 40,450 deaths due to breast cancer this year .
- Reference:
- http://www.cancer.gov/types/breast
- http://www.breastcancer.org/symptoms/understand_bc/statistics

Questions of interest

- Questions of interest that might be asked are:
- 1) Is the presence of malignant associated with an increased risk of death due to Breast cancer?
- 2) Are older patients more likely to die than younger ones ?
- 3) Is there a relation between malignant and the age group patients belongs to ?

About Data

Three factor tables

Survived = Yes

Age \ Malignant	Yes	No
<50	51	77
50-69	38	51
70+	6	7

Survived = No

Age \ Malignant	Yes	No
<50	13	10
50-69	11	20
70+	3	3

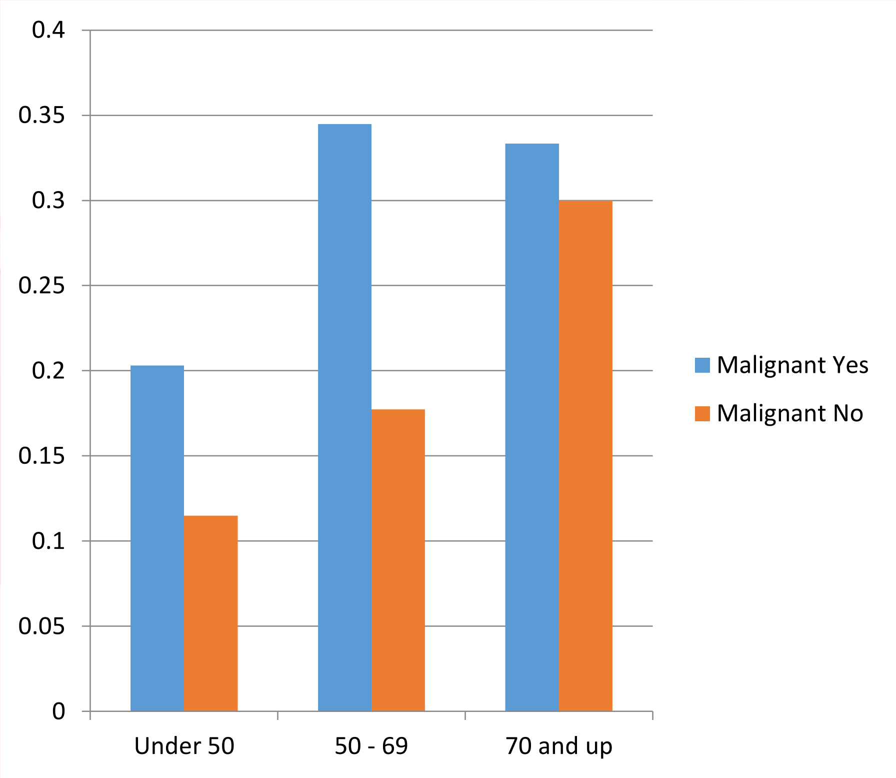
Total
151
120
19

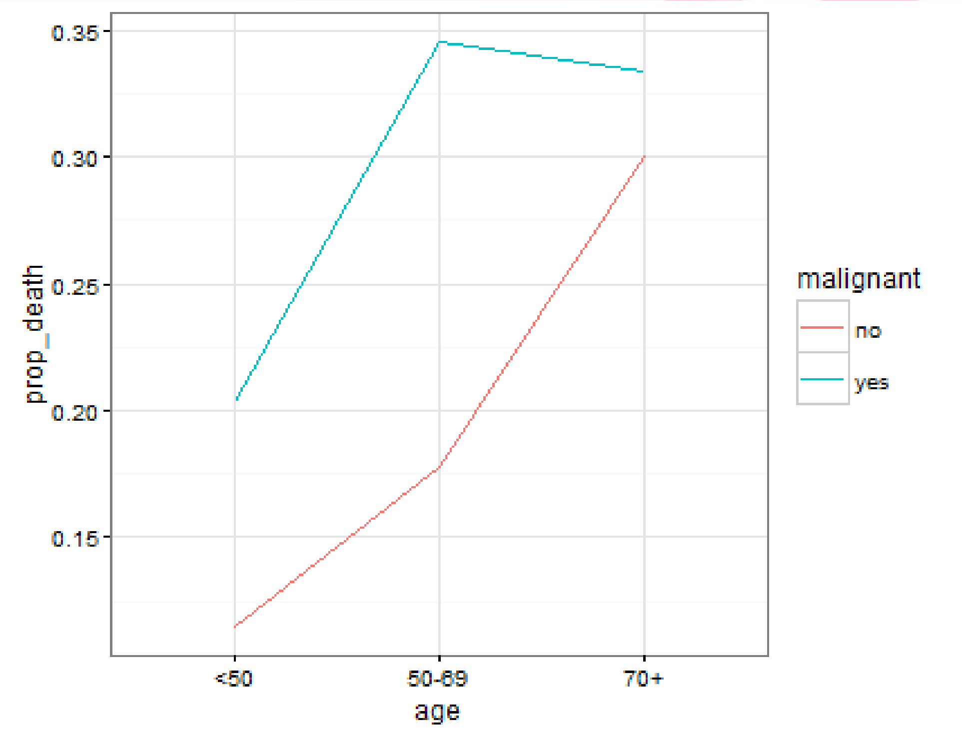
Total(Malignant:Yes) = 122
Total(Survived:Yes) = 230

Total(Malignant:No) = 168
Total(Survived:No) = 60

About Data

Plots





Methods used in the Analysis

1. Pearson Chi-Square Test of Independence
2. Multinomial Likelihood Ratio Test

   Analysis 1 and 2 both check whether there is relationship between 2 variables.

3. Binomial Sampling

   Fits data to a model that predicts odds (or probabilities) of death.

4. Poisson Regression

   Fits data to a model that predicts mean number of death within a period of time.

Testing Dependency between Variables

Pearson Chi-Square Test of Independence

1. Age v.s. Malignancy

   P = 0.6080

2. Survival v.s. Age

   P = 0.04858

3. Survival v.s. Malignancy

   P = 0.009557

Multinomial Likelihood Ratio Test

1. Age v.s. Malignancy

   P = 0.6079

2. Survival v.s. Age

   P = 0.04942

3. Survival v.s. Malignancy

   P = 0.009638

Binomial Sampling

<u>Model 1: with interaction</u>

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,i}$$
$$+\beta_4(I_{50-69,i} \times I_{MalY,i}) + \beta_5(I_{70+,i} \times I_{MalY,i})$$

π_i = probablity of death for group i

$$I_{70+,i} = \begin{cases} 1, if\ i\ belongs\ to\ 70+ \\ 0, \quad otherwise \end{cases}$$

$$I_{50-69,i} = \begin{cases} 1, if\ i\ belongs\ to\ 50-69 \\ 0, \quad otherwise \end{cases}$$

$$I_{MalY,j} = \begin{cases} 1, if\ i\ belongs\ to\ Malignant: Yes \\ 0, \quad otherwise \end{cases}$$

Binomial Sampling

Testing Significance of Interaction in Model 1

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,i}$$
$$+\beta_4 (I_{50-69,i} \times I_{MalY,i}) + \beta_5 (I_{70+,i} \times I_{MalY,i})$$

Analysis of Deviance

Factor	P
Age	0.04942
Malignancy	0.01316
Age x Malignancy	0.7811

Binomial Sampling

Model 2: without interaction

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,i}$$

Coefficients	Estimate	P	exp(estimate)
β_0	-2.0730	1.68E-13	0.1258
β_1	0.6318	0.0424	1.8810
β_2	0.9282	0.0917	2.5300
β_3	0.7328	0.0141	2.0809

Binomial Sampling

Model 1 v.s. Model 2

AIC (Akaike's Information Criterion)

- Measures how well model fits the data
- Smaller the better

$\text{AIC}(\text{Model 1}) = 33.939 > 30.433 = \text{AIC}(\text{Model 2})$

Binomial Sampling

Estimates for Probablity of Death

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,i}$$

Age	Malignant No	Malignant Yes
<50	0.1117	0.2075
50-69	0.1914	0.3300
70+	0.2414	0.3984

Binomial Sampling

<u>Diagnostics</u>

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,i}$$

π_i = probablity of death for group i

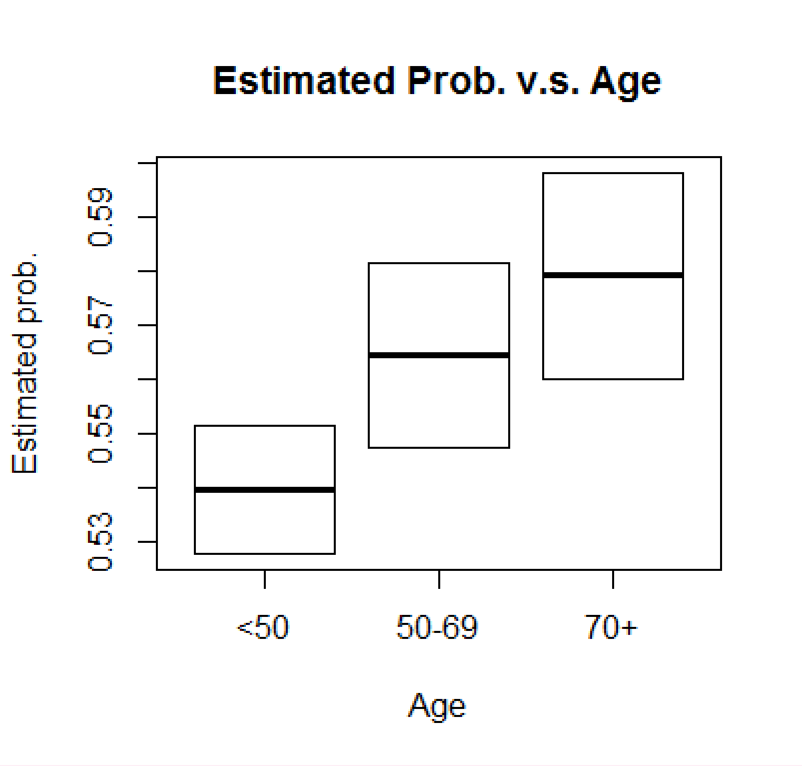
$$I_{70+,i} = \begin{cases} 1, if\ i\ belongs\ to\ 70+ \\ 0, \qquad otherwise \end{cases}$$

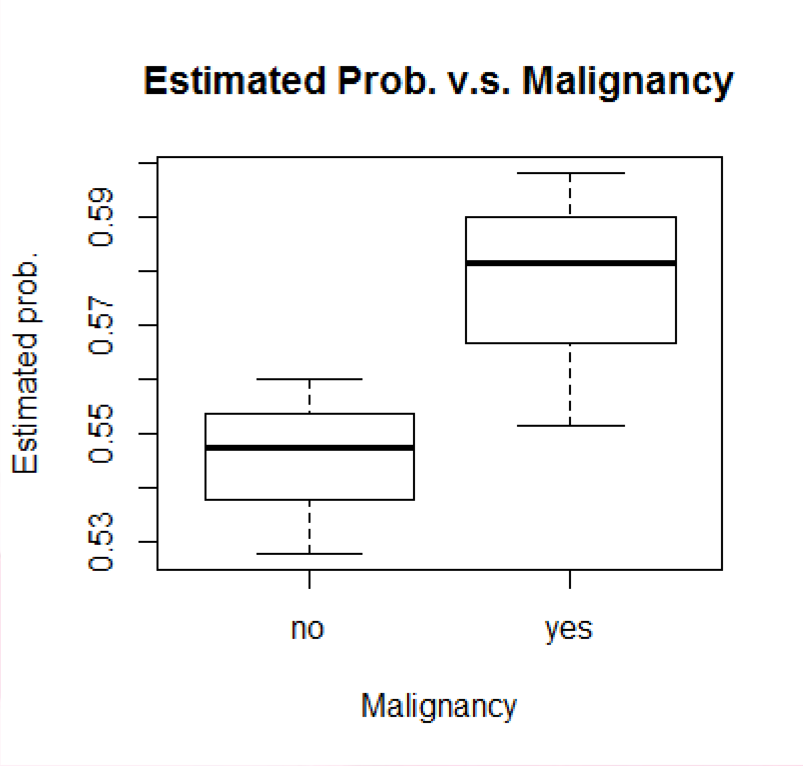
$$I_{50-69,i} = \begin{cases} 1, if\ i\ belongs\ to\ 50-69 \\ 0, \qquad otherwise \end{cases}$$

$$I_{MalY,j} = \begin{cases} 1, if\ i\ belongs\ to\ Malignant\colon Yes \\ 0, \qquad otherwise \end{cases}$$

Binomial Sampling

Plots





Poisson Regression

Three factor tables

Survived = Yes

Age \ Malignant	Yes	No
<50	51	77
50-69	38	51
70+	6	7

Survived = No

Age \ Malignant	Yes	No
<50	13	10
50-69	11	20
70+	3	3

Poisson Regression

Model 1: with 3 – way interaction

$$\log(\lambda_{i,j,k}) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,j} + \beta_4 I_{SurN,k}$$
$$+ \beta_5(I_{50-69,i} \times I_{SurN,k}) + \beta_6(I_{70+,i} \times I_{SurN,k}) + \beta_7(I_{MalY,j} \times I_{SurN,k})$$
$$+ \beta_8(I_{50-69,i} \times I_{MalY,j} \times I_{SurN,k}) + \beta_9(I_{70+,i} \times I_{MalY,j} \times I_{SurN,k})$$

$\lambda_{i,j,k} = expected\ counts\ in\ (i,j,k)th\ cell$

$$I_{50-69,i} = \begin{cases} 1, if\ i\ belongs\ to\ 50-69 \\ 0, \quad otherwise \end{cases}$$

$$I_{70+,i} = \begin{cases} 1, if\ i\ belongs\ to\ 70+ \\ 0, \quad otherwise \end{cases}$$

$$I_{MalY,j} = \begin{cases} 1, if\ j\ belongs\ to\ Malignant: Yes \\ 0, \quad otherwise \end{cases}$$

$$I_{SurN,k} = \begin{cases} 1, if\ k\ belongs\ to\ Survived: No \\ 0, \quad otherwise \end{cases}$$

Poisson Regression

Model 1: with 3 – way interaction

$$\log(\lambda_{i,j,k}) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,j} + \beta_4 I_{SurN,k}$$
$$+ \beta_5(I_{50-69,i} \times I_{SurN,k}) + \beta_6(I_{70+,i} \times I_{SurN,k}) + \beta_7(I_{MalY,j} \times I_{SurN,k})$$
$$+ \beta_8(I_{50-69,i} \times I_{MalY,j} \times I_{SurN,k}) + \beta_9(I_{70+,i} \times I_{MalY,j} \times I_{SurN,k})$$

Q: Are β_8 and β_9 statistically significant? (Does age affect the number of death/survival differently depending on whether the tumor is malignant or not?)

A) Yes!
B) No…
C) I don't know.

Poisson Regression

Model 1: with 3 – way interaction

$$\log(\lambda_{i,j,k}) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,j} + \beta_4 I_{SurN,k}$$
$$+ \beta_5(I_{50-69,i} \times I_{SurN,k}) + \beta_6(I_{70+,i} \times I_{SurN,k}) + \beta_7(I_{MalY,j} \times I_{SurN,k})$$
$$+ \beta_8(I_{50-69,i} \times I_{MalY,j} \times I_{SurN,k}) + \beta_9(I_{70+,i} \times I_{MalY,j} \times I_{SurN,k})$$

Testing Significance with Analysis of Deviance

Factor	P
Age x Malignancy	0.6080
Age x Survival	0.0494
Malignancy x Survival	0.0132
Age x Survival x Malignancy	0.7811

Poisson Regression

Model 2: without 3 – way interaction

$$\log(\lambda_{i,j,k}) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,j} + \beta_4 I_{SurN,k} + \beta_5(I_{50-69,i} \times I_{SurN,k}) + \beta_6(I_{70+,i} \times I_{SurN,k}) + \beta_7(I_{MalY,j} \times I_{SurN,k})$$

Coefficient	Estimate	P
β_0	4.3192	< 2e-16
β_1	-0.3634	0.27809
β_2	-2.2871	0.00338
β_3	-0.3514	0.12389

Coefficient	Estimate	P
β_4	2.1000	8.98E-14
β_5	-0.6619	0.0315
β_6	0.9433	0.08235
β_7	0.7569	0.01045

Poisson Regression

Model 1 v.s. Model 2

AIC (Akaike's Information Criterion)

- Measures how well model fits the data
- Smaller the better

$\text{AIC(Model 1)} = 74.633 > 71.26 = \text{AIC(Model 2)}$

Poisson Regression

Interpreting Estimates of β′s

$$\log(\lambda_{i,j,k}) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,j} + \beta_4 I_{SurN,k}$$
$$+ \beta_5(I_{50-69,i} \times I_{SurN,k}) + \beta_6(I_{70+,i} \times I_{SurN,k}) + \beta_7(I_{MalY,j} \times I_{SurN,k})$$

- $\lambda_{50-69,\bullet,SurN} = \exp\{\beta_0 + \beta_1 + \beta_4 + \beta_5\}$
- $\lambda_{70+,\bullet,SurN} = \exp\{\beta_0 + \beta_2 + \beta_4 + \beta_6\}$
- $\lambda_{\bullet,MalY,SurN} = \exp\{\beta_0 + \beta_3 + \beta_4 + \beta_7\}$

$\exp\{\beta_1\}$: $\#(50-69) = \exp\{\beta_1\} \times \#(<50)$

$\exp\{\beta_2\}$: $\#(70+) = \exp\{\beta_2\} \times \#(<50)$

$\exp\{\beta_3\}$: $\#(MalignantYes) = \exp\{\beta_3\} \times \#(MalignantNo)$

$\exp\{\beta_4\}$: $\#(SurvivedNo) = \exp\{\beta_4\} \times \#(SurvivedYes)$

Poisson Regression

Interpreting Estimates of β′s

If data was balanced within each factor, then $\beta_1 = \beta_2 = \beta_3 = 0$

- $\lambda_{50-69,\bullet,SurY} = \exp\{\beta_0 + \beta_1 + \beta_4 + \beta_5\}$ → $\exp\{\beta_0 + \beta_4 + \beta_5\}$
- $\lambda_{70+,\bullet,SurY} = \exp\{\beta_0 + \beta_2 + \beta_4 + \beta_6\}$ → $\exp\{\beta_0 + \beta_4 + \beta_6\}$
- $\lambda_{\bullet,MalY,SurY} = \exp\{\beta_0 + \beta_3 + \beta_4 + \beta_7\}$ → $\exp\{\beta_0 + \beta_4 + \beta_7\}$

If N(<50) = N(50-69) = N(70+) and N(MalignantYes) = N(MalignantNo),

$\exp\{\beta_5\}$: #(50-69 and dead) = $\exp\{\beta_5\} \times$ #(< 50 and dead)

$\exp\{\beta_6\}$: #(70+ and dead) = $\exp\{\beta_6\} \times$ #(< 50 and dead)

$\exp\{\beta_7\}$: #(Malignant and dead = $\exp\{\beta_7\} \times$ #(Not malignant and dead)

Poisson Regression

Interpreting Estimates of $\lambda_{i,j,k}$

$$\log(\lambda_{i,j,k}) = \beta_0 + \beta_1 I_{50-69,i} + \beta_2 I_{70+,i} + \beta_3 I_{MalY,j} + \beta_4 I_{SurN,k}$$
$$+ \beta_5(I_{50-69,i} \times I_{SurN,k}) + \beta_6(I_{70+,i} \times I_{SurN,k}) + \beta_7(I_{MalY,j} \times I_{SurN,k})$$

Coefficients	Estimate	P	exp(estimate)	
β_5	0.6619	0.0315	1.9385	→ **i.**
β_6	0.9433	0.08235	2.5684	→ **ii.**
β_7	0.7569	0.01045	2.1317	→ **iii.**

If N(<50) = N(50-69) = N(70+) and N(MalignantYes) = N(MalignantNo),

i. #(50-69 and dead) = 1.9385 × #(< 50 and dead)

ii. #(70+ and dead) = 2.5684 × #(< 50 and dead)

iii. #(Malignant and dead) = 2.1317 × #(Not malignant and dead)

Binomial Sampling v.s. Poisson Regression

GOF (Goodness of Fit) Test

- Measures Prob(Model is appropriate given data)
- Bigger the better

$\text{GOF}(\text{Poisson Model 2}) = 0.9194 > 0.7811 = \text{GOF}(\text{Binom Model 2})$

Poisson Model is more appropriate!!!

Analysis Summary

Methods used

1. Pearson Chi-Square Test of Independence
2. Multinomial Likelihood Ratio Test

   Analysis 1 and 2 both check whether there is relationship between 2 variables.
3. Binomial Sampling

   Fits data to a model that predicts odds (or probabilities) of death.
4. Poisson Regression

   Fits data to a model that predicts mean number of death within a period of time.

Analysis Summary

1. There is no evidence about interaction effect between age and malignancy of cancer on death/survival of breast cancer patients within 3 years.

   By analyses 1, 2, 3, 4.

2. There is some evidence that age is associated to death/survival of breast cancer patients within 3 years.

   By analyses 1, 2, 3, 4.

3. There is a strong evidence that malignancy of cancer is associated to death/survival of breast cancer patients within 3 years.

   By analyses 1, 2, 3, 4

Analysis Summary

4. There is some evidence that breast cancer patients in age 50 – 69 have higher risk of death within 3 years compared to patients under 50. (Odds of death is 1.88 times higher).

   By analyses 3.

5. There is not enough evidence that breast cancer patients who are 70 and over have higher risk of death within 3 years compared to patients under 50.

   By analyses 3.

6. There is a strong evidence that presence of malignant tumor is associated to higher risk of death within 3 years for breast cancer patients. (Odds of death is 2.08 times greater).

   By analyses 3.

Analysis Summary

7. There is some evidence that expected number of death within 3 years for breast cancer patients in age 50 – 69 is greater compared to patients under 50. (1.94 times greater).

   By analyses 4.

8. There is not enough evidence that expected number of death within 3 years for breast cancer patients who are 70 and over is greater compared to patients under 50.

   By analyses 4.

9. There is a strong evidence that presence of malignant tumor is associated to more expected number of death within 3 years for breast cancer patients. (2.13 times greater)

   By analyses 4.

Questions of Interest: Answers

1) <u>Which factors are associated to death/survival of breast cancer patients?</u>
   a) Age (under 50, 50-69), Malignancy of cancer
   b) No observed effect from age (70 and over)
   c) Effect of age does not depend on malignancy of cancer or vice versa.
2) <u>What effect does each factor have?</u>
   a) #(50-69 and dead) $= 1.94 \times$ #($<$ 50 and dead)
   b) #(70+ and dead) $= 2.57 \times$ #($<$ 50 and dead)
   c) #(Malignant and dead) $= 2.13 \times$ #(Not malignant and dead)
3) <u>Which types of patients have higher mortality rates?</u>
   a) Older patients are more likely to die than younger patients
   b) Patients with malignant tumor are more likely to die than patients without
   c) #(70+, malignant and dead) $= 5.48 \times$ #($<$ 50, not malignant and dead)

Notable Limitations

- Small sample size for age group 70 and up.
  - N(<50) = 151, N(50-69) = 120, N(70+) = 19
  - This limits us from having a statistically significant result on the relationship between age and death of cancer patients.

    Possible fix

  1. Make sure to collect enough samples for all age groups for similar research.
  2. Treat 50-69 and 70 and up as one group.

- Possible confounding variable
  - Higher age is usually associated with higher number of death due to causes other than breast cancer.
  - So we possibly overestimate the association between breast cancer and death.

    Possible fix

  1. Collect data for causes of death.
  2. Look at other studies and compare our results with the relationship between age and general cause of death.

Conclusion

I-clicker Question 3

Which model discussed best illustrate the difference of risk of death among all groups.

A)The odds of death
B)The probability of death
C)The expected number of death
D)None of the above

- We can then conclude from the result that the model that fits best this particular research is the comparison of death probability .

Conclusion for Question of Interest

- 1) Is the presence of malignant associated with an increased risk of death due to Breast cancer?
- Strong evidence shows that patients that suffer from malignant have higher risk to death due to breast cancer .
- 2) Are older patients more likely to die than younger ones ?
- According to the study there is enough evidence to confirm that older patients have on average twice the risk of death then a younger patient.
- 3) Is there a relation between malignant and the age group patients belongs to ?
- According to the tests there is no relation between the age group and the malignant status of a patient

- Finally it is safe to say that no matter what age group a patient belongs to he or she would have a bigger risk of death if they suffer from malignant.