92%

Q1)

Blood Pressure (BP) → dependent variable

↳ not categorical, continuous

Clinician → independent variable

Age → independent variable

Scared → independent variable

MATLAB CODE:

```
>> BP = [80; 92; 89; 79; 32; 10; 78; 74; 86; 27; 82; 104; 110; 76;
   54; 98; 52; 100; 75; 120; 82; 90; 89; 76; 150; 78; 90; 84;
   115; 120];

>> Clinician = ['A','A','A','A','B','B','B','B','B','B','A','A','A','A','B','B','B'
   'B','B','B','A','A','A','A','C','C','C','C','C','C'];

>> Age = [25; 45; 32; 78; 100; 90; 45; 12; 74; 13; 86; 45; 23; 86;
   12; 86; 33; 25; 11; 90; 45; 30; 54; 13; 75; 23; 98; 55;
   12; 89];

>> Scared = [0;0;0;0;0;0;0;0;0;0;0;0;0;0;1;1;1;1;
   1;1;1;1;1;1;1;1;1;1;1;1;]

>> T = table (BP, Clinician, Age, Scared);

>> LM= fitlm (T, 'BP ~ Clinician + Age + Scared);

>> LM.ModelCriterion
```

A)

→ AIC: 281.93
BIC: 288.94

⇒ smaller AIC/BIC → better model

```
Linear regression model:
    BP ~ 1 + Clinician + Age + Scared

Estimated Coefficients:
                   Estimate      SE        tStat       pValue
                   ________    ______    _______    __________

    (Intercept)      79.041    10.875      7.268    1.2907e-07
    Clinician_B     -21.854    10.108    -2.1621      0.040377
    Clinician_C      6.3109    14.386    0.43869       0.66466
    Age            0.052699   0.15028    0.35068       0.72877
    Scared_1         17.723    10.503     1.6874       0.10396
```

↳ there is a statistically significant difference btwn Blood Pressure (BP) and Clinician B.

↳ there is NOT a

↳ There is NOT a statistical difference btwn Blood Pressure (BP) and Clinician C, Age, or Scared_1

B) >> LM2 = fitlm(T, 'BP ~ Age + Scared')

>> LM2.ModelCriterion → AIC: 285.03
BIC: 289.24

```
Linear regression model:
    BP ~ 1 + Age + Scared

Estimated Coefficients:
```

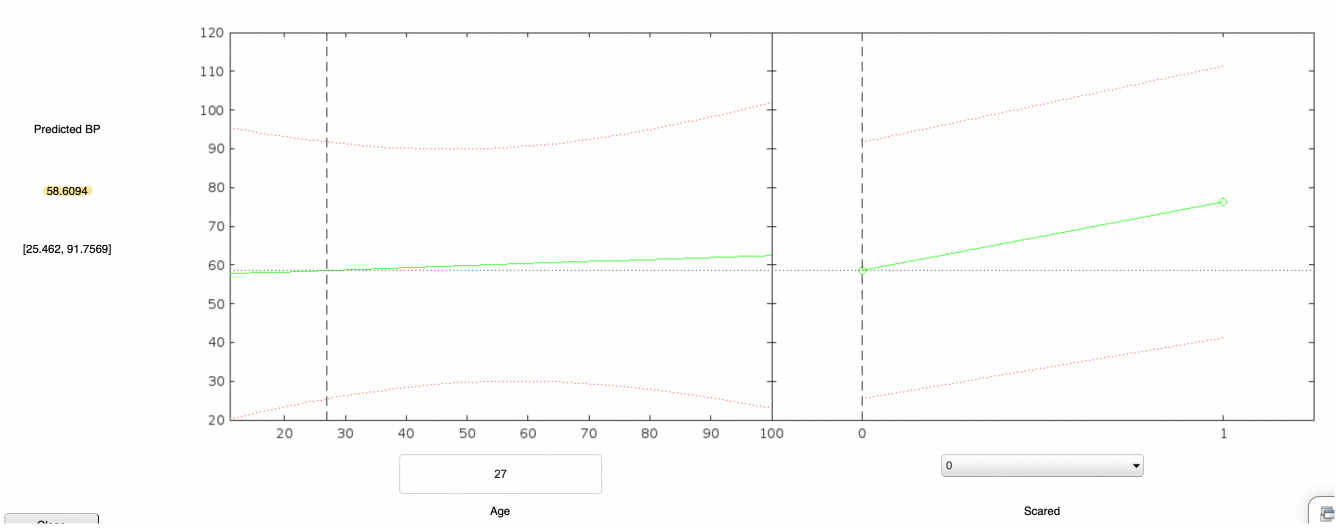
	Estimate	SE	tStat	pValue
(Intercept)	67.676	10.689	6.3313	8.8805e-07
Age	0.075544	0.16001	0.47213	0.64063
Scared_1	23.203	9.7503	2.3797	0.024653

There is a statistically significant difference btwn Blood Pressure (BP) and Scared_1.

There is NOT a statistical difference btwn Blood Pressure (BP) and Age.

C) To predict BP, use the first mode and run the plot-slice function

>> plotSlice(LM) ⇒ predicted BP is 58.6



27 Age

Not Scared (0)

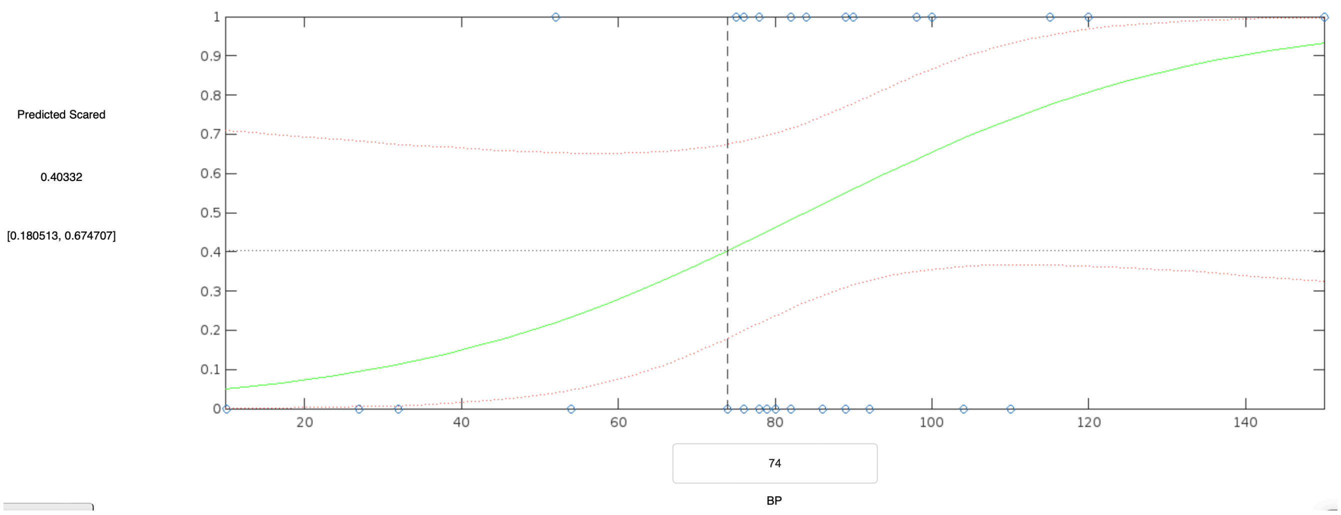
D) NEW MODEL

↳ dependent variable : Scared (categorical variable)

↳ independent variable : BP

Use logistic regression since your dependent variable is categorical

```
>> T4 = table(Scared, BP)
>> Loglm = fitglm(T4, 'Scared ~ BP', 'Distribution', 'binomial')
>> plotSlice(Loglm)
```



You have a 40.3% of being scared if your diastolic blood pressure is 74.

(Q2)

Since we are using a parametric test, we want to use a two-sample T-test since we have samples from two different groups + we only have two columns of data.

A) THOUGHT PROCESS:

i) What is the probability of picking a 35 (Class A) + that it's greater than any number in Class B? $1/6 \times 0 = 0$; 35 is NOT greater than any number in Class B

ii) What is the probability of picking a 67 (Class A) + that it's greater than any number in Class B? $1/6 \times 1/6 = 1/36$; 67 is greater than 46

iii) What is the probability of picking a 88 (Class A) + that it's greater than any number in Class B? $1/6 \times 4/6 = 4/36$; 88 is greater than 46, 67, 78, 78

iv) What is the probability of picking a 92 (Class A) + that it's greater than any number in Class B? $1/6 \times 4/6 = 4/36$; 92 is greater than 46, 78, 67, 78

v) What is the probability of picking a 100 (Class A) + that it's greater than any number in Class B? $1/6 \times 5/6 = 5/36$; 100 is greater than 46, 78, 67, 78, 99

vi) What is the probability of picking a 55 (Class A) + that it's greater than any number in Class B? $1/6 \times 1/6 = 1/36$ 55 is greater than 46

SUM ALL PROBS SINCE "OR" PROBLEM: $15/36 = \boxed{5/12}$

↳ you need to "add" them together

B)

	PROBABILITY OF REMOVING ONE GRADE FROM CLASS A		BOTH CLASS A (REMOVED) AND CLASS B WILL INDIVIDUALLY HAVE MEANS STAT. DIFF. THAN 50	
REMOVE 35:	1/6	x	$\underset{(A)\ \times\ (B)}{(1 \times 1)}$	= 1/6
				+
REMOVE 67:	1/6	x	$\underset{(A)\ \times\ (B)}{(0 \times 1)}$	= 0
				+
REMOVE 88:	1/6	x	$\underset{(A)\ \times\ (B)}{(0 \times 1)}$	= 0
				+
REMOVE 92:	1/6	x	$\underset{(A)\ \times\ (B)}{(0 \times 1)}$	= 0
				+
REMOVE 100:	1/6	x	$\underset{(A)\ \times\ (B)}{(0 \times 1)}$	= 0
				+
REMOVE 55:	1/6	x	$\underset{(A)\ \times\ (B)}{(0 \times 1)}$	= 0
				= 1/6

MATLAB CODE:

```
>> ClassB = [47;78;67;99;78;104];
>> 50;
>> [h,p,ci,stats] = ttest(ClassB, 50)
```

→ use a 1-sample t-test b/c you are comparing a data set to a specific number

p-value: 0.02

$0.02 < 0.05$ → Statistically Significant → $\boxed{\text{always}}$

PROBABILITY: $\boxed{1/6}$

= 16.7%

The probability of Class B will be statistically different than 50 → 100% or 1

```
>> ClassA = [67; 88; 92; 100; 55];
>> 50;
>> [h, p, ci, stats] = ttest(ClassA, 50)
```

↱ use a 1-sample

p-value: 0.02

0.02 < 0.05 → Statistically Significant

```
>> ClassA = [35; 88; 92; 100; 55];
>> 50;
>> [h, p, ci, stats] = ttest(ClassA, 50)
```

↱ use a 1-sample

p-value: 0.13

0.13 > 0.05 → NOT Statistically Significant

```
>> ClassA = [35; 67; 92; 100; 55];
>> 50;
>> [h, p, ci, stats] = ttest(ClassA, 50)
```

↱ use a 1-sample

p-value: 0.17

0.17 < 0.05 → NOT Statistically Significant

```
>> ClassA = [35; 67; 88; 100; 55];
>> 50;
>> [h, p, ci, stats] = ttest(ClassA, 50)
```

↱ use a 1-sample

p-value: 0.18

0.17 < 0.05 → NOT Statistically Significant

```
>> ClassA = [35; 67; 88; 92; 55];
>> 50;
>> [h, p, ci, stats] = ttest(ClassA, 50)
```

↱ use a 1-sample

p-value: 0.18

0.18 < 0.05 → NOT Statistically Significant

```
>> ClassA = [35; 67; 88; 92; 100]
>> 50;
>> [h, p, ci, stats] = ttest(ClassA, 50)
```

↱ use a 1-sample

p-value: 0.09

0.09 < 0.05 → NOT Statistically Significant

Q3)

IV1 and IV2 → independent variables

DV → continuous, dependent variable

Since the dependent variable is <u>NOT</u> categorical, you want to use linear regression, <u>NOT</u> logistic

MATLAB CODE:

the plane has to shift - doing AIC / BIC w interactions + w/o interactions

```
>> IV1_I = [1;2;3;4;5;7;7;10];
>> IV2_I = [1;2.2;2.9;4.3;5.5;6.7;6.7;11];
>> DV_D = [1;9;30;67;122;160;319;1015];

>> T1 = Table(DV_D, IV1_I, IV2_I)
>> LM = fitlm(T1, 'DV_D ~ IV1_I + IV2_I')
```
multiple regression

p-value: 0.00699

```
>> T2 = Table(DV_D, IV1_I)
>> LM2 = fitlm(T2, 'DV_D ~ IV1_I')
```
when IV1 is the ONLY independent variable

p-value: 0.00681

```
>> T3 = Table(DV_D, IV2_I)
>> LM3 = fitlm(T3, 'DV_D ~ IV2_I')
```
when IV2 is the ONLY independent variable

p-value: 0.00268

OF INTEREST:

- the highest p-value - 0.007 - is with two independent (IV1 and IV2) and 1 dependent variable (DV)
- when IV1 or IV2 is the only independent variable, the IV2 model has a lower p-value - 0.00268 - than the IV1 model

MULTIDIMENSIONAL SPACE:

first model - multi-dimensional model - in 3D space
↳ since P_{fit} is 3 → higher p-value

second and third model - linear model - in 2D space
↳ since P_{fit} is 2 → lower p-value

BTWN 2nd + 3rd Model
- IV1: lower mean (4.875), lower standard deviation (2.24) ⇒ higher p-value compared to 3rd Model b/c lower STD + less spread
- IV2: higher mean (5.038), greater standard deviation (3.18)

Sums of squares around mean / Sums of squares around fit

$$F = \frac{SS(mean) - SS(fit) / (P_{fit} - P_{mean})}{SS(fit) / (n - P_{fit})}$$

Simple regression: $P_{fit} = 2$ → int + IV1 or IV2

multiple regression: $P_{fit} = 3$
↳ int + IV1 + IV2

Q4) Boulder → 1 independent variable

Fungi Color → dependent variable (3)

A) Assuming normality, we want to use an ANOVA b/c it compares 1 independent variable w 3 dependent variables.

MATLAB CODE:

```
>> Red = [30; 17; 33; 18; 25; 50; 40; 19];
>> Green = [35; 10; 18; 34; 20; 15; 20; 5];
>> Cream = [8; 5; 12; 2; 25; 11; 6; 14];
>> [p, tbl, stats] = anova1([Red, Green, Cream])
```

ANOVA TABLE:

- F Statistic: 6.98
- P value: 0.0047

$\alpha = 0.05$

Since $0.0047 < \alpha$, we can conclude that there is a significant difference between the quantity of red, green, and cream fungi at the beach

B) Not assuming normality → use Kruskal-Wallis Test

MATLAB CODE:

```
>> Red = [30; 17; 33; 18; 25; 50; 40; 19];
>> Green = [35; 10; 18; 34; 20; 15; 20; 5];
>> Cream = [8; 5; 12; 2; 25; 11; 6; 14];
>> [p2, tbl2, stats2] = kruskalwallis([Red, Green, Cream])
```

KRUSKAL-WALLIS ANOVA Table

- F Statistic: 9.57
- P value: 0.0083

$\alpha = 0.05$

Since $0.0083 < \alpha$, we can conclude that there is a significant difference between the quantity of red, green, and cream fungi at the beach

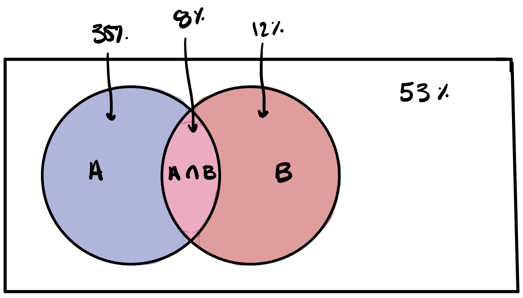
Boulder - paired t-tests

Friedman Test and 2-Way ANOVA

↳ comparing all the rows together

↳ Comparison + similarity btwn boulders

Q5)



Probability one has senile dementia if they don't have artherosclerosis?

$P(LA) : 0.65$

$P(B \mid LA) : \frac{0.04}{0.65} = 0.0615 \times 100 = \boxed{6.15\%}$

USING BAYES FORMULA

① $P(A)$: probability one has atherosclerosis → 0.35

$P(B)$: probability one has senile dementia → 0.12

$P(A \cap B)$: probability having both illness → 0.08

$P(1 - P(A) - P(B))$: probability of NOT having an illness → 0.53

② : BAYES

$P(B) = P(C)$: probability of senile dementia

$P(D)$: probability of not having artherosclerosis

$P(D \mid C)$: probability of not having artherosclerosis given senile dementia

$$P(C \mid D) = \frac{P(D \mid C) \times P(C)}{P(D)} = \frac{\frac{0.04}{\cancel{0.12}} \times \cancel{0.12}}{0.65} = 0.0615 = \boxed{6.15\%}$$