Results and conclusions

1. Fit appropriate regression models for fatigue (SSC-F) given weakness (SSC-W) and treatment group (IC). Select the regression model that is best supported by the data using the likelihood ratio test, AIC and BIC.

Observation-

For this model, we consider treatment group and/or weakness as predictors for the response variable, fatigue. We consider different combinations of predictors using Ordered Logistic Regression model as the response variable, Fatigue (SSC-F) has four naturally ordered levels (0 = absent → 3 = severe)

#	Model	Deviance	AIC	BIC
1	sscf ~ ic	3478.594	3486.594	2894.334
2	sscf ~ sscw	2851.118	2863.118	3507.405
3	sscf ~ ic + sscw	2840.904	2854.904(best)	2891.322(best)

Model 1 co-efficient interpretation:

```
Coefficients:
      Value Std. Error t value
ic1 -0.4311        0.1   -4.31
```

The ordinal logistic regression shows that IC has a **significant** (as |t| > 2) **negative effect** (as co-efficient is -ve) **on SSCF** (β = –0.4311, OR = 0.65). This means individuals with IC = 1 have 35% lower odds of being in higher SSCF categories.

Model 2 co-efficient interpretation:

```
Coefficients:
           Value Std. Error t value
sscw.L  3.13424     0.1421 22.0601
sscw.Q -0.02423     0.1155 -0.2098
sscw.C  0.33348     0.1009  3.3047
```

The ordered logistic regression shows a **strong positive linear association** between weakness (SSC-W) and fatigue (SSC-F). The coefficient for the linear trend (3.13) indicates that as weakness increases, the odds of being in a higher fatigue category increase dramatically (exp (β)≈ 23). The quadratic trend is not significant, suggesting no substantial curvature in this relationship. The cubic trend is significant but small, indicating minor deviations from linearity. Overall, weakness is a **very strong predictor** of fatigue.

Model 3 co-efficient interpretation:

```
Coefficients:
            Value Std. Error  t value
ic1    -0.344299     0.1078 -3.19422
sscw.L  3.130716     0.1423 22.00414
sscw.Q  0.004253     0.1160  0.03668
sscw.C  0.332470     0.1011  3.28965
```

For IC effect: Holding weakness constant, individuals in the **intervention group (IC = 1)** have **29% lower odds** of being in a higher fatigue category than those in the control group. As odd ratio = exp (-0.344) ~ 0.71. This effect is **significant** but much **smaller** than the effect of weakness.

For linear effect of weakness: For each one-level increase in weakness (0→1→2→3), the odds of being in a *higher fatigue category* increase by about **23 times**, controlling for treatment. This is still the **dominant predictor**.

Odd ratio = exp (3.131) ~ 22.88

Now let us compare Model 2 and Model 3 to determine if adding the treatment group variable as predictor is significant in building model.

We obtain the lowest scores for AIC and BIC for model 3 that indicates that Model 3 is the best in terms of fit and complexity of model.

We know in model 3 the dominant predictor is weakness. In this case is it significant to include IC as another predictor. Let us do a likelihood ratio test to check if adding IC or treatment group is significant.

LRT results-

#	Model comparison	Delta Deviance	Delta df	p-value
Model 2 v/s Model 3	sscf ~ sscw v/s sscf ~ ic + sscw	10.21397	1	0.001393811

From the LRT, we fail to accept null hypotheses (simpler model is better). Hence the best model we can fit is Model 3 (sscf ~ ic + sscw). Though weakness is dominant predictor, adding IC improves the model fit.

To conclude, weakness has a **major significant positive relation** with fatigue. Receiving treatment also **lower odds** of being in a higher fatigue category.

2. Fit appropriate regression models for weakness (SSC-W) given fatigue (SSC-F) and treatment group (IC).
Select the regression model that is best supported by the data using the likelihood ratio test, AIC and BIC.

Observation-

#	**Model**	**Deviance**	**AIC**	**BIC**
1	sscw ~ sscf	2943.778	2955.778(best)	2986.994(best)
2	sscw ~ ic	3574.942	3582.942	3603.753
3	sscw ~ ic + sscf	2943.612	2957.612	2994.031

Model 1 (sscw ~ sscf) already has the **lowest AIC (2955.778) and BIC (2986.994)**, so it is preferred based on both fit and complexity. So we will not perform LRT with any other model. Because we can see that adding ic with sscf increases both AIC and BIC values.

Interpretation of co-efficient for model 1:

```
Coefficients:
         Value Std. Error t value
sscf.L  3.1713     0.1537  20.630
sscf.Q -0.1347     0.1213  -1.111
sscf.C  0.2407     0.1026   2.346
```

The **relationship is strongly positive**: higher sscf levels are associated with much higher odds of being in a higher sscw category.

The **linear trend dominates**, while quadratic and cubic effects are minor, suggesting a mostly **monotonic increase**.

3. Fit loglinear models for weakness (SSC-W), fatigue (SSC-F) and treatment group (IC). Select the loglinear model that is best supported by the data using G2.
Observation -

As suggested in the question, I dropped the time variable to form a contingency table of dimension – 4x2x4 as below-

	ic	0				1				Row total
ssc-f	**ssc-w**	0	1	2	3	0	1	2	3	
0		145	33	33	9	239	37	38	13	547
1		24	81	44	9	25	69	54	13	319
2		13	17	105	26	19	24	77	35	316
3		15	1	20	56	10	3	8	48	161
Column total		197	132	202	100	293	133	177	109	1343

We examine all possible hierarchical log-linear models that involve SSC-W,SSC-F and IC as in the table.

	Model type	Model	G^2	df	p-value	Fits the data well?
1	Complete Independence	[F][IC][W]	846.3169	24	0	No
2	One-var Independent of the other 2	[F][IC, W]	830.2331	21	0	No
3	One-var Independent	[IC][F, W]	38.37768	15	0.000793269	No
4	One-var Independent	[W][F,IC]	826.6999	21	0	No
5	**Conditional Independence**	**[F,IC][F,W]**	**18.76075**	**12**	**0.09446808**	**Yes**
6	Conditional Independence	[F,IC][IC,W]	810.6161	18	0	No
7	Conditional Independence	[F,W][IC,W]	22.29388	12	0.03435513	No
8	**Pairwise**	**[F,IC][F,W][IC,W]**	**10.0994**	**9**	**0.342498**	**Yes**
9	**Saturated**	**[F,IC,W]**	**0**	**0**	**1**	**Yes**

Only Models 5, 8 and 9 from Table have a p-value associated with the log-likelihood Ratio test statistic G^2 greater that 0.05 and fit the data well.

Summary of these models:

- **Model 5**: Conditional independence - simple, explains the main associations without overfitting.
- **Model 8**: Pairwise associations - slightly more complex, still fits well, includes all two-way relationships.
- **Model 9**: Saturated - perfect fit, but fully parameterized; usually used as benchmark.

We determine which of these three models is best supported by the data by performing two pairwise comparisons of the 3 hierarchical loglinear models that fits the data well.

	Comparison	G^2 difference	Df difference	p-value	Comment
1	Model 8 v/s Model 9	10.0994	9	0.657502	As p-value>0.5, accept null hypothesis and keep the simpler model 8.
2	Model 5 v/s Model 8	8.661349	3	0.9658507	As p-value>0.05, accept null hypothesis to keep the simpler model, model 5

From the above table, it is clear that model-5 **[F,IC][F,W]** best fits the data. This means that Fatigue is influenced by treatment and fatigue has direct effect on weakness. But once you know a participant's fatigue level, knowing treatment group gives **no extra information** about weakness.

Clinical interpretation

Fatigue acts as a **mediator variable**.
Treatment affects fatigue, and fatigue affects weakness.
But treatment does **not** directly affect weakness after accounting for fatigue.

4. Summarize your findings in the statistical analyses you perform in the first three problems. Is weakness related to fatigue? If so, in which way? Is the intervention effective in altering the levels of weakness or fatigue? Are the findings from your regression models consistent with those from your loglinear models?

Both the regression and loglinear models indicate a strong positive relationship between fatigue and weakness. Regression Model 1 shows that treatment reduces fatigue, while Regression Model 2 does not show a direct effect of treatment on weakness. The loglinear model confirms that treatment affects fatigue, which in turn indirectly influences weakness, with fatigue acting as a mediator between treatment and weakness. Regression models quantify these effects, whereas the loglinear model highlights the structural pathways of influence.

Regression Model 1: Fatigue as Response (SSCF ~ IC + SSCW)

- **Relationship with Weakness:**
  Weakness (SSC-W) is the **dominant predictor** of fatigue. Each one-level increase in weakness increases the odds of being in a higher fatigue category by ~23 times (OR $\approx$ 22.88).
- **Effect of Intervention (IC):**
  Receiving treatment (IC = 1) **reduces the odds of higher fatigue** by 29% (OR $\approx$ 0.71), even after controlling for weakness.
- **Interpretation:**
  Weakness strongly drives fatigue, but treatment also has a **significant, though smaller, effect** on reducing fatigue.

Regression Model 2: Weakness as Response (SSCW ~ SSCF)

- **Relationship with Fatigue:**
  Fatigue (SSC-F) is a strong positive predictor of weakness. Higher fatigue levels are associated with substantially higher odds of being in a higher weakness category.
- **Effect of Intervention (IC):**
  Adding IC does **not improve model fit**, indicating treatment does not directly affect weakness.
- **Interpretation:**
  Weakness is strongly linked to fatigue, but treatment's effect on weakness not seen.

Loglinear Model: Hierarchical Model [F,IC][F,W]

- Fatigue is influenced directly by treatment (IC -> F).
- Weakness is influenced directly by fatigue (F -> W).

- **No direct path from IC to weakness**, meaning the effect of intervention on weakness is mediated through fatigue.
-