A correlation between variables, however, does not automatically mean that the change in one variable is the cause of the change in the values of the other variable (i.e., Pearson or Spearman correlations.

Causality: Causation indicates that one event is the result of the occurrence of the other event; i.e. there is a causal relationship between the two events. This is also referred to as cause and effect.

Solve Endogeneity with 1. Instrumental variable 2. Experimental Design

An instrumental variable is correlated with dependent variable, but it affects Y only through its effects on independent variables

**Course high level summary**

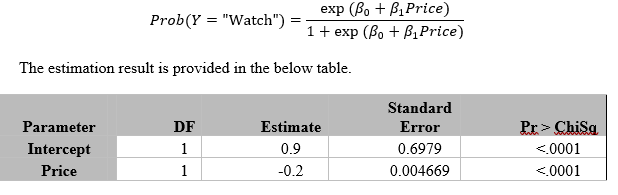
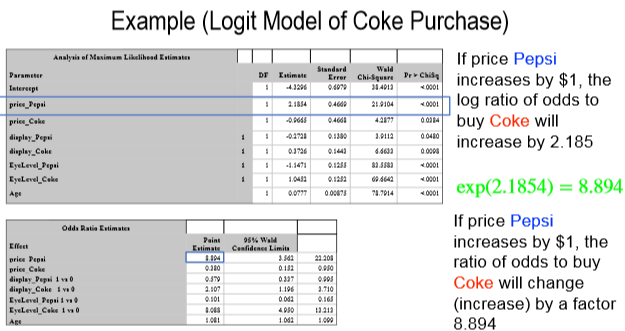
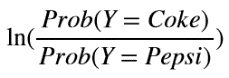
If dependent variable is continuous, the we use Regression

• If dependent variable is binary and discrete, then we use Logistic / Probit regression

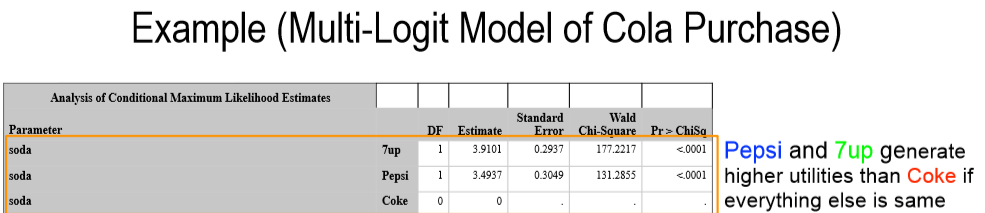
• If dependent variable is multinomial and discrete, then we use Multinomial Logit / Multinomial Probit

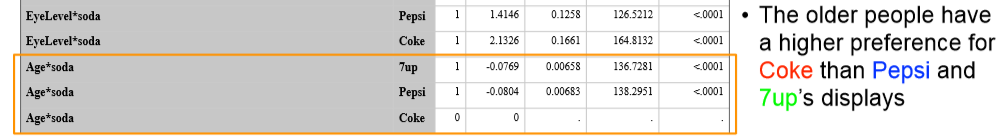
• If dependent variable is discrete and violates IIA, then we use Nested Logit / Multinomial Probit

• If dependent variable is binary and ordinal, then we use Ordered Logit

Not a good idea to use Linear on classification because Residuals are not normal, homoscedastic…. Predicting probabilities that are negative or larger than 1. So Linear still can be used on binary but not for multinomial Sensitivity = TP/(TP+FP) Spec = TN/(TN+FP)

If price goes up by $1, the ratio of odds of buying to not buying will decrease by a factor .





IF IIA ( INDEPENDENCE OF Irrelevant alternatives) is violated means it doesn’t hold. Multi-Logit suffers from IIA issue to capture asymmetric switching patterns, EVEN if the new alternative is more similar to one existing alternative than another. In this case we can use **Multinomial Probit:** We can assume the unobserved error terms are normally distributed. It is called **Multi-Probit model.** Or we can use **The Nested Logit Model:** We can have Muti-layer of decision process to model the decision-maker’s choice process.

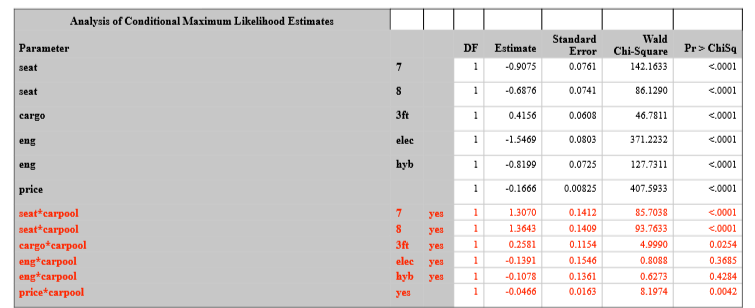
Conjoint: Technique for finding consumer preferences based on their ratings, rankings or choices of alternatives

Ratings-based conjoint: Consumers has to rate each alternative (e.g., on a scale of 1 – 100), OR allot 100 points between handful of alternatives

• Ranking-based conjoint: Consumers rank the alternatives in the order of preference

• Choice-based conjoint: Consumers choose their best alternative

• Incentivized conjoint: Participants are given a chance to win one of the products that they choose

We use multinomial logit model to do conjoint analysis.

For no carpool interpret is straight.

6carseats are preferred over 7&8. Cargo3ft is preferred over base.

Gas car engine which is base is preferred over Elec and hyb.

With Car pool is calculated

Is calculated by adding the respective estimate.

So car7 seats actual estimate is 1.3070 +(-0.9075) = 0.399 (respective to 6seater which is the base)

Carpool are more price sensitive as -0.1666+(-0.0466) = -0.23 so this will decrease sales more as compared to no carpoo

Limited dependent variable can be left censoring: insurance example where claims start at $3000 so its left censored

residuals are not equally distributed. Right Censoring: Y is capped out and doesn’t go above a certain point.

Censoring: All outcomes are recorded, but values may only be partially known because of endogenous reporting by people or organizations

Truncating: the data set does not include observations in the analysis that are beyond a boundary value.

We use Tobit Model for truncated or censored data. Especially left censored.

Two step regression can also solve the limited dependent variable. In first step we use Probit and then we use linear regression.