**Theory Questions:**

1. Kolla på följande video: https://www.youtube.com/watch?v=X9\_ISJ0YpGw&t=290s , beskriv kortfattat vad en Quantile-Quantile (QQ) plot är?

Ans: QQ plot is used to see that set of observations are normally distributed or skewed. If QQ plot result in approximately straight line, then set of observations are approximately normally distributed else the line changes its shape based on the data.

1. Din kollega Karin frågar dig följande: ”Jag har hört att i Maskininlärning så är fokus på prediktioner medan man i statistisk regressionsanalys kan göra såväl prediktioner som statistisk inferens. Vad menas med det, kan du ge några exempel?” Vad svarar du Karin?

Ans: Normally Machine learning models focuses more on predictions; how accurate it can give the output for unseen data. But it does not explain which variables are more influential for accuracy of model. Statistical regression concentrates on prediction and also check which variables are influencing for its accuracy.

For example, from the used Volvo cars data, it is clear that Linear regression model predicts unseen data and also explains which variables are influential for accuracy.

1. Vad är skillnaden på ”konfidensintervall” och ”prediktionsintervall” för predikterade värden?

Ans: Confidence interval gives us the range where the group of inputs with similar variables fall with certain percent of confidence whereas prediction interval gives us a range where a single input with certain variables will fit with certain percent of confidence.

1. Den multipla linjära regressionsmodellen kan skrivas som:

Y= 𝛽0 + 𝛽1𝑥1 + 𝛽1𝑥2+ ...+ 𝛽𝑝𝑥𝑝 +𝜀 . Hur tolkas beta parametrarna?

Ans: Here in the given equation 𝛽0 is the intercept when all the independent variables x1….xp are 0. Remaining beta parameters are the measures how independent values affects target variable.

Eg: In used volvo cars selling price is the target variable (Y)and if we consider 3 variables age(x1), Horsepower(x2), Fuel type(x3).

𝛽1 = Negative value, car price decreases by its age

𝛽2 = Horsepower value may not influence much

𝛽3 = Depends of type(Bensin,Diesel,El) beta value changes. So based on these beta parameters target variable value will be predicted.

1. Din kollega Nils frågar dig följande: ”Stämmer det att man i statistisk regressionsmodellering inte behöver använda träning, validering och test set om man nyttjar mått såsom BIC? Vad är logiken bakom detta?” Vad svarar du Hassan?

Ans: No, BIC is used for model selection and its not used for evaluating model performance on unseen data. When we use BIC to select a model, we have to split data into train, test and validation. This allows model to train on different dataset and evaluate on validation set and test data to predict unseen data.

1. Förklara algoritmen nedan för ”Best subset selection”?

Ans: step1: M0 denote null model, in used Volvo cars data if we don’t use any predictors and fit the model with only basic features of a car.

Step2: Every step increases the number of combinations of predictors. In our example, first consider only one predictor (age, year, type, Mileage).

a)In the next step try to take 2 combinations at a time(age and year, type and mileage, age and type)

b)In the next step try taking 3 combinations at a time.

After that pick best models from the models and (Mk).

Step3: Select best model based on selection criteria(Validation Error, Cp (AIC), BIC, or adjusted R2 or the cross validation).

1. Ett citat från statistikern George Box är: “All models are wrong, some are useful.” Förklara vad som menas med det citatet.

Ans: All the models which we fit are not completely accurate, but we use them to predict in this real world for solving complex problems.

**Självutvärdering:**

1. Vad tycker du har varit roligast i kunskapskontrollen?`

Ans: It was good to learn in depth about statistical inference.

1. Hur har du hanterat utmaningar? Vilka lärdomar tar du med dig till framtida kurser?

Ans: Practising more on code and analysis.

1. Vilket betyg anser du att du ska ha och varför?

Ans: I am not 100 percent perfect on statistical analysis, I am still learning to analyse data in a precise way. I followed complete course and gave my 100 percent. I cant evaluate my grade this time 😊