# Datainsamling

Min grupp bestod av mig, David Norman, Ian Gutierrez, Filip Holmberg, Magdalena Wallner och Erik Fredrixon.

Gruppen delade upp datainsamlingen genom att välja ett bilmärke var, där alla skulle bidra med ca 50 observationer. Sedan loggade vi datan i en gemensam Excel-fil, säkerställde att vi kunde importera och använda datan i R och sedan diskuterade vi lite allmänt kring datan.

Jag tycker gruppen tog sig an uppgiften på ett väldigt bra sätt, vi strukturerade ett workflow så gott vi kunde, lyssnade på varandras idéer och hjälptes åt ifall någon fastnade. Vi höll också kontakten via en Teams-grupp under processen. Jag tror inte jag har något jag vill förbättra, alla var aktiva, lyssnade och villiga att arbeta. Olika personer är såklart mer vana vid olika ämnen, så vi kunde hjälpas åt när problem uppstod.

Mina styrkor och vad jag kan förbättra; känner att jag är duktig på problemlösning och att involvera alla som vill vara involverade. Jag är också väldigt avslappnad och lätt att arbeta med. Ibland kan jag ha svårt att förstå hur olika ämnen hänger ihop, vilket kan begränsa min kapacitet. Jag kan också vara ganska passiv och obekväm att prata i grupp. Det här är områden jag aktivt försöker förbättra.

På det stora hela skulle jag nog inte ändra något, jag är väldigt nöjd med våra resultat.

# Teoretiska frågor

1. En QQ-plot är ett sätt att kontrollera normaliteten i datan genom att titta på hur de observerade värdena förhåller sig till de teoretiska värdena för en normal fördelning. En relativt rak linje indikerar normalfördelning.
2. Prediktioner syftar på att förutspå någonting, exempelvis hur ett hus eller en bil ska prissättas baserat på tidigare observationer av sålda hus/bilar. Statistisk inferens å andra sidan handlar om att förstå och kvantifiera sambandet mellan de olika variablerna för att se hur de påverkar priset.

Exempel prediktion: baserat på variabler från alla andra hus vi sålt, hur mycket bör vi sälja det här nya huset för?

Exempel inferens: hur mycket kvadratmeter eller antal rum i snitt påverkar huspriset (om det påverkar alls).

1. Prediktionsintervall är alltid bredare än konfidensintervall, då prediktionsintervallet alltid tar med en slumpmässig felterm i ekvationen, medan konfidensintervall endast utgår från medelvärdet.
2. Betaparametrarna tolkas som koefficienter, det vill säga b1 visar hur mycket x1 ökar per enhet, givet att alla andra variabler är konstanta. B0 är interceptet, alltså där kurvan skär y-axeln.
3. Jag skulle säga att det stämmer, då BIC inkluderar en felterm som straffar komplexa modeller i högre grad, detta ger en sorts ”simulering” av variansen som kan finnas i test-data. Det kan dock fortfarande vara bra att använda validering/test data också.
4. P = predictors, k = antal predictors som väljs. Så *p choose k* betyder att den testar alla möjliga kombinationer av *k* över den totala mängden *p* (första modellen är tom)*.* Sedan använder den ett förvalt mått (Cp, BIC, AIC etc.) för att välja den kombination av predictors som har bäst poäng. Notera att när *p* blir större så växer antalet kombinationer exponentiellt, vilket kan bli prestandamässigt tungt.
5. Att modellerna bara är en förenkling av verkligheten, ingen modell kommer vara helt perfekt, men vissa är tillräckligt bra att man kan dra slutsatser ifrån dem.

# Självutvärdering

1. Utmaningar du haft under arbetet samt hur du hanterat dem.

* Den största utmaningen ligger i att förstå själva ekvationerna som utförs när man kör funktioner i R, det känns väldigt svårt att få ett grepp om exakt hur de fungerar. Men jag googlar, söker efter simpla förklaringar och går igenom ekvationer steg för steg tills jag känner att jag har en bra förståelse.
* Också att förstå uppbyggnaden i R till exempel datatyper, levels, vilka funktioner som går att använda med olika typer av data. Det här är dock relativt lätt att ta reda på.

1. Vilket betyg du anser att du skall ha och varför.

* Tycker jag nått god förståelse för ämnet och förhoppningsvis syns det i min rapport där jag tror mig fått med en hel del små detaljer som lätt kan missas. VG.

1. Något du vill lyfta fram till Antonio?

* -

# Appendix A

# Källförteckning