

Tellende prosjekt i ISTx1001 høst 2024

Industriell statistikk

Generell informasjon

I prosjektdelen av ISTx1001 Statistikk (industriell statistikk) har vi fokus på to hovedtemaer; *forsøksplanlegging og statistisk kvalitetsstyring*.

- Dette er oppgaveteksten til den tellende prosjektoppgaven. Besvarelsen teller 30% av karakteren i emnet.
- Det er 3-6 studenter i hver gruppe. Gruppene dere skal levere prosjektet i er de samme gruppene som ble etablert i forprosjektet.
- Prosjektet leveres gruppevis i Inspira.
- Alle studenter i samme gruppe får samme bokstavkarakter. Merk NTNUs regelverk for klage på karakter: Ved klage på karakterfastsettingen av gruppearbeid, der det gis en felles karakter, klager du individuelt. En eventuell endring etter klage vil kun gjelde for den/de gruppe medlemmene som har klaget.
- Informasjon om prosjektmodulen finnes i Blackboard, sammen med alt kursmateriellet.
- I forbindelse med oppgave 1 (2-nivå faktorielt forsøk) skal dere skrive en rapport på ca 6-7 sider, og maksimalt 10 sider (dette inkluderer tabeller og figurer).
- I forbindelse med oppgave 2 og 3 skal dere skrive en kort redegjørelse/refleksjon.
- De ulike oppgavene (1, 2, 3) er vektet ulikt (70%, 20% og 10%). Karakteren settes med prosentvurderingsmetoden.

Dere skal levere **en pdf-fil**. Dere er selv ansvarlige for at alle svar er med (rapport fra oppgave 1 samt svar på oppgave 2 og 3). Fila skal være ryddig og oversiktlig.

Frist for innlevering er 18. november klokka 12:00 i Inspira.

Det kan ikke gis utsettelse på innleveringsfristen.

Oppgave 1 - Forsøksplanlegging

Merknader

- Rapporten fra oppgave 1 skal være på **maksimalt 10 sider**. Dette inkluderer tabeller og figurer.
- Det legges vekt på at rapporten skal være sammenhengende og forståelig.
- I JupyterHuben ligger flere notatbøker som kan brukes for å analysere og visualisere data. Disse er også tilgjengelige for nedlasting fra <https://github.com/imfdrift/istx100y2024>
- Det *anbefales sterkt* at problemstillingen dere vil undersøke avklares med lokal faglærer før dere begynner med forsøkene.

Oppgavebeskrivelse

Målet med prosjektet er å planlegge, gjennomføre, analysere og diskutere et 2-nivå faktorielt forsøk. Dere står selv fritt til å velge problemstilling og eksperiment. Det kan være et laboratorieeksperiment eller et problem fra dagliglivet. Her er stegene gruppa skal gjennomføre:

1. Bestem dere for en problemstilling som dere kan studere med et 2^k faktorielt forsøk. Hva er responsvariabelen og hva er mulige forklaringsvariabler? Hva er egnede nivåer for forklaringsvariablene? Merk at responsvariabelen Y må være en kontinuerlig stokastisk variabel.
2. Design et forsøk med 16, 24 eller 32 målinger. Bruk enten 3 eller 4 faktorer. Dersom dere bruker 3 faktorer, må dere ha minst ett gjentak.
3. Planlegg innsamling av data.
4. Samle inn data.
5. Analyser data.
6. Skriv en rapport.

Rapportmal

Innledning

- Forklar kort hvorfor dere har valgt problemstillingen og hva målet deres er. Har dere forhåndskunnskap som kan være relevant?
- Diskuter valg av responsvariabel (en kontinuerlig variabel) og valg av forklaringsvariabler (3 eller 4). Forventer dere samspillseffekter mellom noen av disse forklaringsvariablene? Hvorfor/hvorfor ikke?

Planlegging av forsøket

- Hvordan vil dere måle responsen? Drøft kort usikkerheten forbundet med målinger av den valgte responsvariabelen deres.
- For hver forklaringsvariabel (faktor), hvilke to nivå vil dere velge? Begrunn kort hvorfor dere har valgt akkurat disse nivåene. Hvordan vil dere kontrollere at faktorene er på

nivåene dere har valgt?

- Oppgi informasjon om forsøksdesign, gjentak, randomisering og eventuell blokking. Hvordan skal forsøket i praksis gjennomføres for å sikre uavhengige forsøk under tilstrekkelig like omstendigheter? Presenter forsøksdesignet i en tabell.

Gjennomføring av forsøket

- Presenter målingene dere har gjort i en tabell.
- Beskriv gjennomføringen, gikk alt etter planen? Skjedde det noe underveis som kan påvirke analysen og validiteten av forsøket?

Analyse av målingene

- Presenter resultater fra multippel regresjonsanalyse (gjennomført i Python), samt utregninger av hovedeffekter og interaksjonseffekter.
- Kommenter på signifikans og effektstørrelser.
- Presenter hovedeffektsplott, interaksjonsplott (samspillsplott) og paretoplott. Alle figurer må ha en forklarende figurtekst.
- Sjekk og kommenter normalantagelsen med et QQ-plott. Diskuter implikasjoner for den statistiske analysen.

Konklusjon og diskusjon

- Gjør en tolkning av verdien til effektene.
- Hvilke konklusjoner kan dere trekke fra forsøket, og hva kan dere eventuelt ikke konkludere om?
- Hva ville dere gjort annerledes om dere skulle vurdere problemstillingen på nytt? Diskuter svakheter med forsøket og gjennomføringen.
- Var det hensiktsmessig med et 2-nivå faktorielt forsøk, eller kunne dere tenke dere f.eks. flere nivåer per faktor? Hva med andre faktorer?

Eksempler på forsøk

Vi har satt sammen noen forslag som lar seg gjøre hjemmefra, men bruk gjerne det dere har tilgang på av laboratorier, utstyr, etc. for å svare på mer relevante problemstillinger.

- Baking av muffins. Responsvariabel: høyde på muffins, forklaringsvariabler: bakepulver, mel, smør, osv.
- Sagetid. Responsvariabel: tid for å sage gjennom en planke, forklaringsvariabel: sag, treverk, posisjon, osv.
- Koking av vann. Responsvariabel: Tid til fosskoking, forklaringsvariabler: temperatur, kjele, salt, induksjon, osv.
- Basketball. Responsvariabel: Snitt antall treff på 10 kast, forklaringsvariabler: posisjon, ball, teknikk, osv.

Oppgave 2 - Statistisk prosesskontroll

Målet med oppgaven er å bestemme kontrollgrenser i en prosesskontroll, samt visualisere og forklare hva dette innebærer ved hjelp av stokastiske simuleringer.

Tenk dere at dere skal lage et produkt basert på forsøket dere gjennomførte i Oppgave 1, f.eks muffins. Dere må selv velge nivåer for forklaringsvariablene som produktet skal produseres under. En stor produksjon skal settes i gang, og responsvariabelen fra Oppgave 1 skal brukes for å kontrollere (med stikkprøver) at prosessen er under kontroll.

2a) Kontrollgrenser

Finn en estimert forventningsverdi og et standardavvik på Y fra Oppgave 1 for den valgte faktorkombinasjonen. Dette setter dere som verdier på μ og σ for en prosess som er under kontroll. Bestem hvor mange (n) produkter dere vil kontrollere i hver stikkprøve.

Bestem kontrollgrenser for gjennomsnittet til en stikkprøve. Bruk Shewarts' anbefaling om 3-sigma-grenser.

Bestem kontrollgrenser for det empiriske standardavviket til en stikkprøve basert på kritiske verdier fra kjikvadratfordelingen med $\alpha/2 = 0.002$.

2b) Stokastisk simulering

Gjennomfør simuleringene under, hver med 50 simulerte stikkprøver. Bruk Python. For hver av simuleringene skal dere plote et kombinert $\bar{x}$ -s-diagram og *forklare hva vi ser i diagrammet*.

Simulering 1: Trekk n responsverdier y for hver stikkprøve. Disse trekkes fra en normalfordeling med forventning μ og standardavvik σ som spesifisert over.

Simulering 2: Nå skal μ gradvis økes ved å legge til 0.5σ etter hver tiende stikkprøve. Når går alarmen?

Presenter arbeidet deres (inkl. figurer) på maksimalt to sider.

Oppgave 3 - Relevans i studieprogram

I denne oppgaven skal dere reflektere rundt forsøksplanlegging og/eller prosesskontroll i ingeniørfaget. Dersom alle på gruppa er fra samme studieprogram velger dere et eksempel fra den studieretningen, ellers bruker dere et eksempel som er relevant for minst en av ingeniørretningene. Skriv omtrent en halv side med diskusjon for denne oppgaven.

Her er noen forslag til diskusjon:

- Skriv ned et eksempel på en problemstilling fra deres ingeniørretning som dere kunne tenke dere å studere. Vil et 2-nivå faktorielt forsøk være egnet her? Hvorfor / hvorfor ikke?
- Har dere allerede gjennomført forsøk i en lab i løpet av studiet der flere forklaringsvariabler (faktorer) kan ha påvirket resultatet? Skriv ned mulige faktorer og om dere kunne/burde ha gjort et 2-nivå faktorielt forsøk for å bestemme egnede nivåer.
- Kan dere tenke dere en produksjonsprosess relatert til deres ingeniørretning der prosesskontroll vil være relevant? Hva slags målinger burde gjøres med stikkprøver, og hva er kritiske avvik?

I denne oppgaven finnes det ingen riktige eller gale svar. Poenggiving gis basert på refleksjon og diskusjon, og forståelse for den statistiske teorien for forsøksplanlegging og/eller prosesskontroll. Dersom dere har skrevet en god refleksjon får dere ikke trekk i poeng for en 'negativ' konklusjon.