Oppgave 1

En berømt kornprodusent hevder at det gjennomsnittlige sukkerinnholdet i en frokostblandingseske som han seller, er 0.2 gram per 1 gram frokostblanding. Et tilfeldig utvalg av n frokostblandingsesker velges ut og sukkerinnholdet registreres. Vi kan generelt anta at alle målinger, $X_1, X_2, \ldots X_n$ er uavhengige og normalfordelte med forventningsverdi μ, som representerer sukkerinnholdet i kornene, og standardavvik $\sigma = 0.03$.

En vil utføre en hypotesetest med nullhypothese $H_0 : \mu = 0.2$ mot den alternative hypotesen $H_1 : \mu > 0.2$, og forkaste nullhypotesen på signifikansnivå 0.05.

a) Foreslå hvordan en slik hypotesetest basert på gjennomsnittet $\bar{X}$ av målingene kan utføres. Hva blir konklusjonen hvis det tas $n = 10$ prøver og gjennomsnittet av målingene er 0.22?

b) Finn sannsynligheten for at nullhypotesen blir forkastet hvis det tas $n = 10$ prøver og sukkerinnholdet er $\mu = 0.22$.

c) Anta at sukkerinnholdet er $\mu = 0.22$. Beregn hvor mange prøver, n, det må tas for at sannsynligheten for at nullhypotesen forkastes skal være større enn 0.8.

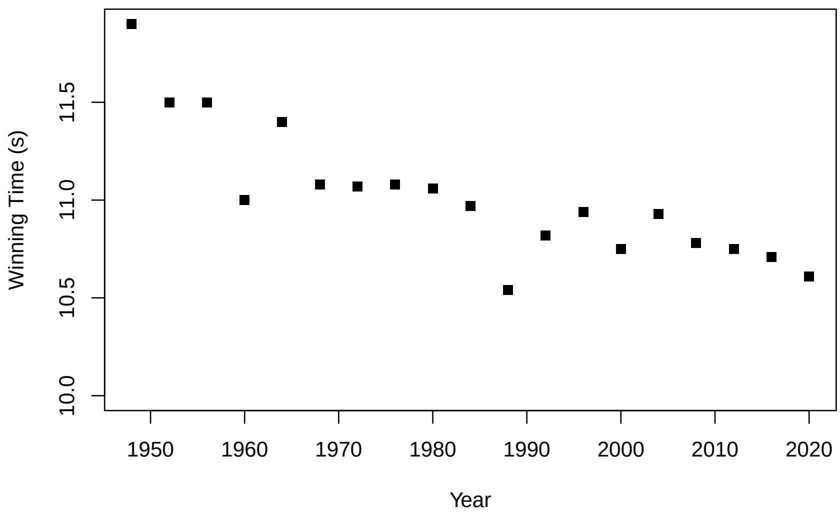
d) Et nytt spørsmål oppstår om fargen på frokostblandingen påvirker salget. For å teste dette selges blandingen i fire forskjelligfargede bokser. Antall bokser av hver farge solgt i løpet av den første uken var:

Blå = 31, Gul = 29, Rød = 23, Grønn = 29.

Bruk et 5% signifikansnivå og test nullhypotesen om at antall solgte bokser for hver av disse fire fargene er det samme.

Oppgave 2

Følgende figur viser olympiske vinnertider i sekunder for kvinner på 100-meter for de olympiske årene fra 1948 til 2020:



En linear regresjonsmodell, $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ med $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, ble tilpasset vinnertidene. Her er y_i vinnertiden i sekunder og x_i angir det olympisk året. Vi har totalt 19 observasjonspar slik at $i = 1, \ldots, 19$. Følgende parameterestimater ble beregnet:

$$\hat{\beta}_0 = 36.78, \quad \hat{\beta}_1 = -0.013, \quad \hat{\sigma}^2 = 0.036,$$

Videre, $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 9120$ (det vil si 0.036/9120 er et estimat av variansen til estimatoren for β_1).

a) Hva er endringen i forventet vinnertid fra den ene olympiaden til det neste (olympiske leker finner kun sted hvert 4. år)? Hva ville ifølge modellen vært vinnertiden i år 0?

b) Kan vi være sikre på at vinnertidene blir raskere over tid? Foreslå og utfør en hypotesetest for å løse dette spørsmålet. Når du gjør dette, definerer du nullhypotesen og den alternative hypotesen, beregner teststatistikken og bestemmer det kritiske området for et signifikansnivå på 5%. Hva konkluderer du med?

Oppgave 3

En studie undersøkte om hunder foretrekker klapping eller vokal ros. Forskere plasserte 14 hunder tilfeldig i to grupper med 7 i hver. I gruppe A ville eieren klappe hunden, mens i gruppe B ville eieren gi lokal ros. Responsvariabelen er tiden, i sekunder, som hunden samhandlet med eieren sin og angis som:

Gruppe A: X_i	114	203	217	254	256	284	296
Gruppe B: Y_j	4	7	24	25	48	71	294

med $i = 1, \dots, n_A$ og $j = 1, \dots, n_B$, med $n_A = n_B = 7$.

Vi tester

$$H_0 : f_X(x) = f_Y(x) \quad \text{for all } x \in (-\infty, +\infty)$$
$$H_1 : f_X(x) = f_Y(x - c) \quad \text{for all } x \in (-\infty, +\infty)$$

for en vilkårlig konstant $c \neq 0$.

a) Beregn teststatistikken og p-verdien for Wilcoxons to-utvalgstest (The Wilcoxon Rank Sum Test). Hva blir konklusjonen din?

Oppgave 4

Det siste året har det vært av interesse å følge med i hvordan inflasjonen har utviklet seg og i hvilke kategorier den har vært størst. I en undersøkelse innenfor dagligvarehandelen ble det spesielt sett på bakevarer, fersk frukt og grønnsaker, hermetisk frukt og grønnsaker, kjøtt og kategorien meieriprodukt. Innenfor hver av de 5 kategoriene ble 10 produkter plukket ut tilfeldig og prosentvis prisøkning i en 2 måneders periode for hvert produkt ble registrert. Du kan i resten av oppgaven anta følgende modell for den observerte prosentvise prisøkningen:

$$Y_{ij} = \mu_j + \epsilon_{ij}, \quad i = 1, 2, \ldots, 10; \quad j = 1, 2, \ldots, 5,$$

hvor $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ og uavhengige.

Fra de observerte dataene fikk en følgende verdier for total kvadratsum og kvadratsummen for kategorier:

$$\text{SSTOT} = \sum_{j=1}^{5}\sum_{i=1}^{10}(y_{ij} - \bar{y}_{..})^2 = 27.5$$
$$\text{SSTR} = \sum_{j=1}^{5} 10(\bar{y}_{.j} - \bar{y}_{..})^2 = 4.6$$

a) Konstruer den komplette variansanalysetabellen og foreta en test på nullhypotesen om at forventet prisstigning er den samme innenfor hver kategori. Hva blir konklusjonen med et signifikansnivå valgt til 0.05?

Gjennomsnittlig observert økning for hver kategori ble: 0.1% for bakevarer, 0.5% for fersk frukt og grønnsaker, 0.6% for hermetisk frukt og grønnsaker, 1% for kjøtt og 0.3% for meieriprodukt. .

b) Konstruer en kontrast for å teste hypotesen at den forventede prisøkningen for kjøtt er den samme som den gjennomsnittlige forventede prisøkningen for de 4 andre kategoriene. Utfør testen. Hva blir konklusjonen med et signifikansnivå valgt til 0.05?

Alle tall i variansanalysetabellen utregnet fra de observerte dataene kunne vært konstruert med informasjon om verdiene for $\bar{y}_{.j}, j = 1, 2, \ldots, 5$ og $\sum_{j=1}^{5}\sum_{i=1}^{10} y_{ij}^2$.

c) Vis at SSTOT kan utregnes dersom denne informasjonen er gitt. Hva blir tallet for $\sum_{j=1}^{5}\sum_{i=1}^{10} y_{ij}^2$ for disse dataene?