

Projekt B

Simon Jorstedt, Erik Krook & Sepehr Zolfeghari

2019-10-14

Inledning

Svenska Träforskningsinstitutet utför två gånger årligen tillverkningskontroller av golvspånskivor producerade i Sverige. Vi skall fokusera på egenskapen böjhållfasthet¹. De tillverkare som uppfyller kontrollens krav kan märka sina spånskivor med ett kontrollmärke. Det är därför viktigt att eventuellt fusk i tillverknin-gen upptäcks så att kontrollens integritet bevaras. Syftet med rapporten är att undersöka variationen i datamaterialet, och rapportera eventuella oegentligheter till Svenska Träforskningsinstitutet

Sammanfattning

Vi inleder med att studera datamaterialet och diskutera de olika variansorsakerna i *Data*. Därefter följer *Modelldiskussion* där vi använder kunskaperna om datamaterialet för att resonera oss fram till en lämplig matematisk modell. Modellen har tvåsidig indelning och består av blandade systematiska och slumpmässiga faktorer, samt en hierarkisk slumpstruktur. I *Skattning och Analys* genomförs sedan en analys av modellens lämplighet där det konstateras att modellantagandena om normalfördelade slumpvariabler är lämpliga. Under *Rekommendationer* lämnas några rekommendationer till Svenska Träforskningsinstitutet om uppföljning av vissa tillverkare.

Data

Datamaterialet ingår i 1975 års kontrollmätningar och utgörs av nio tillverkare, vid två tillfällen (vår och höst). Vid varje tillverkare och tillfälle har sex skivor valts ut (detta antas slumpmässigt), och från varje skiva har tre prover tagits för att mäta böjhållfastheten. Proverna antas ha tagits slumpmässigt ur skivorna.

Datamaterialet utgörs således av 108 individuella spånskivor, och 324 individuella mätningar. För att möjliggöra variansanalys formateras datamängden till en tre gånger så lång dataframe, där varje skiva förekommer tre gånger. En gång för varje mätning. För att illustrera datamaterialet ritas vi i figur 1 upp en boxplot över böjhållfastheten för de olika tillverkarna.

Table 1: Ett övre utdrag ur datamaterialet.

Maker	Autspr	Board	bhf
1	1	1	20.62
1	1	2	22.35
1	1	3	21.30

¹Böjhållfastheten mäts i två riktningar. I datamängden ingår endast en av dessa riktningar, men det är inte angivet vilken denna är.

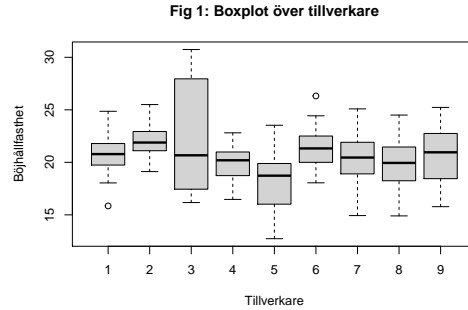


Figure 1: Boxplot över böjhållfastheten för de olika tillverkarna.

Variationen i data kan ha flera olika orsaker, som vi kan sammanfatta enligt följande.

- 1) Det verkar finnas skillnader mellan tillverkare.
- 2) Årstiden kan påverka, till exempel genom luftfuktighet, temperatur osv.
- 3) De individuella arbetarna eller maskinerna hos varje tillverkare kan ha olika arbetssätt och inställningar.
- 4) Mätinstrumenten som används vid datainsamlingen kan vara okalibrerade och har begränsad precision. Det kan även finnas systematiska fel i hur de tre mätningarna i en skiva tas.
- 5) Det kan råda samspel mellan tillverkare och årstid, till exempel om en tillverkare har vidtagit åtgärder för att skydda spånsnivorna från höstregn, medan andra inte har det.

Modelldiskussion

Om det antas att de individuella tillverkarna inte ändrat sina instruktioner till arbetare och maskiner under året då datamaterialet insamlats, så kan tillverkare betraktas som en systematisk faktor. Den kan då representeras av en additiv radeffekt α_i , $i = 1, \dots, 9$. Den är då tänkt att representera den systematiska variation som förekommer mellan två tillverkare.

Vår- och höstmätningarna har valts ut, och skulle kunna betraktas som proxys för första och andra halvan av året. Det vi vill jämföra är alltså två lägen, eller positioner av året. Årstidvariationen betraktas därför som en systematisk faktor, som vi låter representeras av en additiv kolumneffekt β_j , $j = 1, 2$. Vi konstaterar även att systematiska förändringar av årstidernas egenskaper kan bortses ifrån eftersom tidsspannet för datamängden bara utgör ett år.

Att mätinstrument och kontrollanter bidrar med mätfel är ofrånkomligt, men studien har ett stort omfång och betydelse för Svenska Träforskningsinstitutet. Vi kan därför anta att detta fel inte är systematiskt, och att det kan beskrivas av en vanlig normalfördelad slumpterm ϵ_{ijkm} för varje mätning.

Råder det samspel mellan någon av faktorerna skiva, årstid eller tillverkare? För att undersöka detta ritar vi upp samspelsplottar i figur 2.

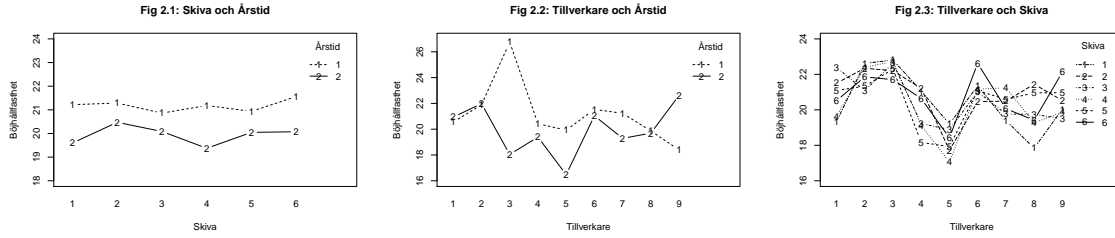


Figure 2: Samspelsplottar för interaktionerna mellan årstid, tillverkare och skiva.

I figur 2.2 framgår det att det råder tydligt samspel mellan tillverkare och årstid. Det är därmed lämpligt att vår modell tar hänsyn till detta, vilket görs med en samspelsfaktor γ_{ij} , $i = 1, \dots, 9$, $j = 1, 2$.

I figur 2.1 och figur 2.3 ser vi inget noterbart samspel.

Vi vill även argumentera för att det råder en hierarkisk slumpindelning. Det råder variation mellan en tillverkares väntevärde för böjhållfastheten och en individuell skiva, detta representerar vi med slumpvariablerna $\delta_{ijk} \sim N(0, \sigma_\delta^2)$. Variationen för mätningar av en och samma skiva representerar vi med slumpvariablerna $\epsilon_{ijkm} \sim N(0, \sigma_\epsilon^2)$. Alla δ_{ijk} och ϵ_{ijkm} antas sinsemellan oberoende.

Modellen beskriver tvåsidig indelning, har blandade faktorer, en hierarkisk slumpindelning och kan formuleras matematiskt enligt

$$Y_{ijkm} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \delta_{ijk} + \epsilon_{ijkm}$$

där $i = 1, \dots, 9$ representerar tillverkare, $j = 1, 2$ representerar årstid, $k = 1, \dots, 6$ anger en specifik spån-skiva, och $m = 1, 2, 3$ anger en specifik mätning. För att uppnå entydighet låter vi parametrerna uppfylla restriktionerna $\sum_{i=1}^9 \alpha_i = \sum_{j=1}^2 \beta_j = \sum_{i=1}^9 \gamma_i = \sum_{j=1}^2 \gamma_j = 0$.

Skattning och Analys

För att analysera vår models lämplighet låter vi R anpassa modellen och skatta parametrarna. Denna anpassning kan sedan användas för att skatta mätningarna i datamaterialet, och beräkna motsvarande residualer. Nedan följer utskriften för den anpassade modellen, och i figur 3 visas en residualplot samt en (ej standardiserad) qq-plot.

```
##
## Error: Board
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  5  15.33   3.067
##
## Error: Within
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Maker      8  451.4   56.43   14.75 < 2e-16 ***
## Autspr     1  123.3  123.28   32.23 3.23e-08 ***
## Maker:Autspr 8  875.4  109.42   28.60 < 2e-16 ***
## Residuals 301 1151.6    3.83
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

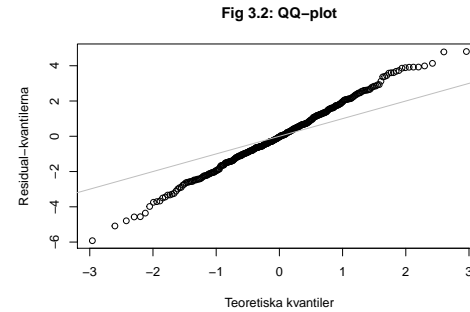
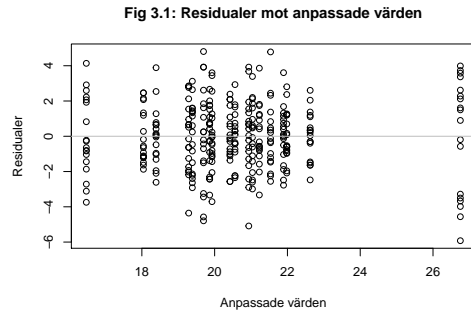


Figure 3: Residualplot och qq-plot för de beräknade residualerna. En horisontell linje $y=0$, samt en linje $y=x$ har ritats in i de respektive figurerna.

I modellutskriften framgår att samtliga faktorerna Tillverkare, Årstid och deras samspel är signifikanta på 99.9%-nivån vilket talar för modellens trovärdighet. I figur 3.1 tycks residualerna ligga ”jämnt” fördelade kring 0, det vill säga det råder homoskedesticitet. I figur 3.2 ligger punkterna ungefär längs en rak linje. Utifrån figur 3 kan vi alltså konstatera att residualerna är approximativt normalfördelade med samma varians. Modellantagandena är därmed uppfyllda. Att punkterna i fig 3.2 inte ligger längs linjen $y = x$ (och har en ”större” lutning) är inte konstigt. Det betyder bara att residualernas varians är större än 1. Mycket riktigt ger den anpassade modellen variansskattningarna $\hat{\sigma}_\delta^2 \approx 3.067$, samt $\hat{\sigma}_\varepsilon^2 \approx 3.826$. Variationen inom skivor tycks alltså vara större än variationen mellan skivor.

I Resultat och diskussion skall vi studera den anpassade modellens parameterskattningar.

Resultat och diskussion

Den anpassade modellen beskriver datamaterialet bra, varför det är rimligt att gå vidare med att studera parameterskattningarna.

```
## Tables of effects
##
##   Maker
##   Maker
##      1      2      3      4      5      6      7      8      9
## 0.1906 1.3798 1.8395 -0.6535 -2.3571 0.7298 -0.3002 -0.7788 -0.0502
##
##   Autspr
##   Autspr
##      1      2
## 0.6169 -0.6169
##
##   Maker:Autspr
##      Autspr
##   Maker 1      2
##      1 -0.806 0.806
##      2 -0.661 0.661
##      3  3.737 -3.737
##      4 -0.100 0.100
##      5  1.110 -1.110
##      6 -0.368 0.368
```

##	7	0.354	-0.354
##	8	-0.532	0.532
##	9	-2.734	2.734

I utskriften ser vi att tillverkare 5 har den lägsta parametern, med klar marginal. Den tillverkare med högst parametervärde är tillverkare 3. Det utläses också att böjhållfastheten tenderar att vara högre på våren än på hösten. Från samspelsparametrarna utläses att skivorna från tillverkare 3 varierar väldigt mycket mellan vår och höst. Den variationen ses också hos tillverkare 9, men är något mindre än för tillverkare 3.

Rekommendationer

Först av allt bör det noteras att det i uppgiftsbeskrivningen inte angetts någon absolut gräns för acceptabla böjhållfasthetsvärden. Med det sagt kan vi däremot uttala oss om de tillverkare som aviker från de övriga, speciellt de som producerar spånskivor med lägre böjhållfasthet. Det antas också vara intressant att vissa tillverkares spånskivor varierar mycket i böjhållfasthetsvärde mellan vår och höst.

Det rekommenderas därför att Svenska Träforskningsinstitutet undersöker tillverkare 3, 5 och 9 noggrannare.