Lsn 9

Clark

September 16, 2019

Admin

Awhile back I was asked to consult on a project with DPE examining the surface that the sled pull was conducted on. They wanted to demonstrate that there was an effect due to the surface. They took 25 volunteers and had them do the sled pull on both grass and sand.

Their sources of variation diagram was:

The model this used is:

Their statistical question was:

The data was fit using:

```
ACFT=read.csv("ACFT.csv")
ACFT<-ACFT %>% mutate(Participantf=as.factor(Participant))
GrassSand=ACFT %>% filter(Surface %in%c("G","S", "S "))%>%droplevels()
levels(GrassSand$Surface)<-c("G","S","S")
contrasts(GrassSand$Surface)=contr.sum
lm.mod<-lm(Sled~Surface,data=GrassSand)
summary(lm.mod)
```

```
##
## Call:
## lm(formula = Sled ~ Surface, data = GrassSand)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

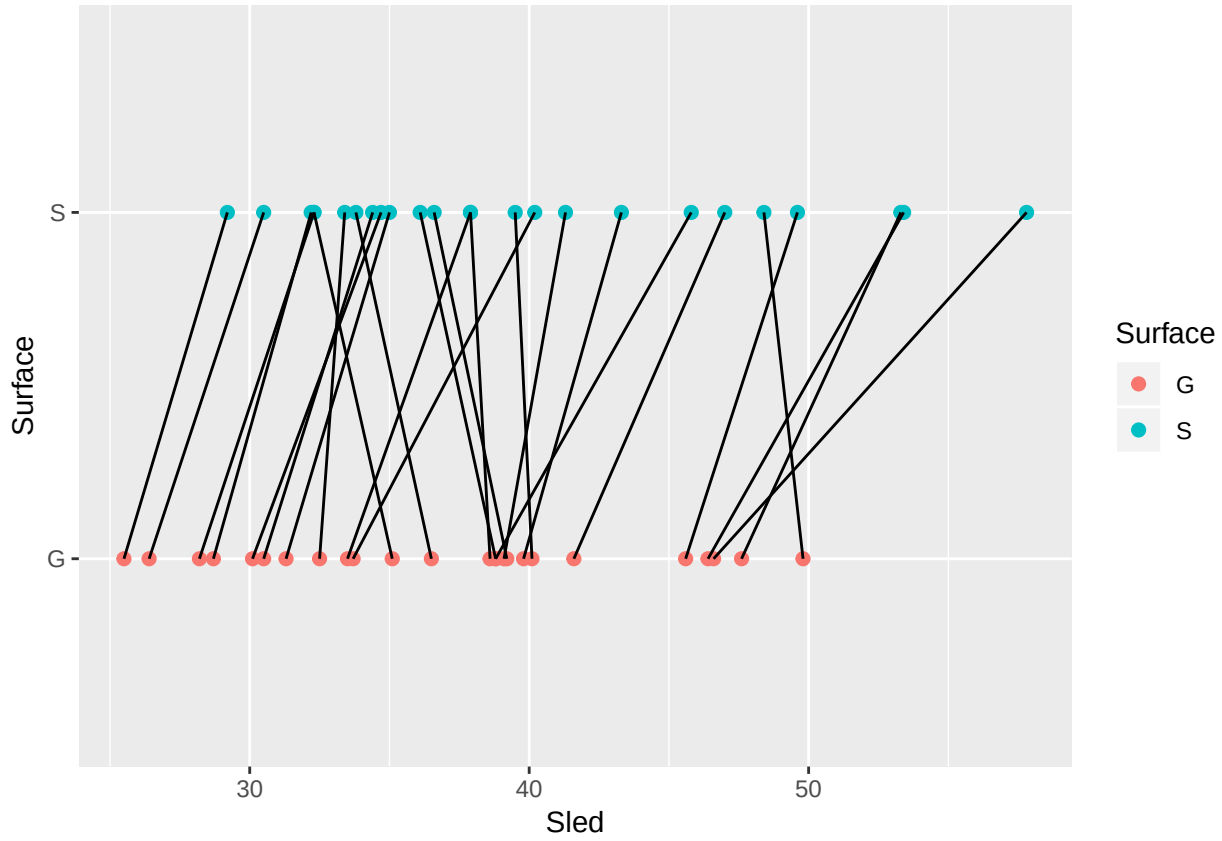
```
## -11.460  -5.942  -1.160   4.346  17.964
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   38.398      1.054  36.423   <2e-16 ***
## Surface1      -1.438      1.054  -1.364    0.179
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.454 on 48 degrees of freedom
## Multiple R-squared:  0.03732,    Adjusted R-squared:  0.01726
## F-statistic: 1.861 on 1 and 48 DF,  p-value: 0.1789
```

```
anova(lm.mod)
```

```
## Analysis of Variance Table
##
## Response: Sled
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Surface    1  103.39 103.392  1.8606 0.1789
## Residuals 48 2667.28  55.568
```

What is our conclusion?

```
GrassSand %>% ggplot(aes(x=Sled,y=Surface,group=Participantf,col=Surface))+
  geom_point(size=2)+geom_line(col="black")
```



What can we learn by looking at this plot?

Let's modify our sources of variation diagram, from our plot what is a major source of variability that is not accounted for?:

Let's write out a new statistical model:

Does our hypothesis we are testing change?

Now we have two ways to approach this statistical question, if we look at within subject, what happens if we take the difference of our two observations?

These differences are typically analyzed using a paired t-test. Now instead of our observations we are looking at our differences.

```
diff.dat<-GrassSand %>% group_by(Participantf)%>%summarize(diff=diff(Sled))%>%
  select(diff)
```

Then we can just do a standard one sample T-test to see if the difference is indeed zero.

```
n=nrow(diff.dat)
y.vals=diff.dat$diff
t.stat=mean(y.vals)/(sd(y.vals)/sqrt(n))
p.val=2*pt(-3.966,n-1)
p.val
```

```
## [1] 0.000574154
```

Which you can do in R using `t.test`. The second way to analyze the data (which will be more helpful for this class) is to continue in an ANOVA framework:

```
contrasts(GrassSand$Participantf)=contr.sum
lm.mod<-lm(Sled~Surface,data=GrassSand)
anova(lm.mod)
```

```
## Analysis of Variance Table
##
## Response: Sled
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Surface    1  103.39 103.392  1.8606 0.1789
## Residuals 48 2667.28  55.568
```

```
full.lm.mod<-lm(Sled~Surface+Participantf,data=GrassSand)
anova(full.lm.mod)
```

```
## Analysis of Variance Table
##
## Response: Sled
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## Surface       1  103.39 103.392  15.734 0.0005733 ***
## Participantf 24 2509.56 104.565  15.912 1.333e-09 ***
## Residuals    24  157.71   6.571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the ANOVA table we see that by adding a second factor (participants) what we actully have done is taken some of our unexplained variation (residuals) and now explained it through participant. This makes the F statistic for Surface bigger because we are no longer comparing 103.392 to 55.56, but rather 13.392 to 6.57. Note other things that happen:

We won't take the time to prove it in this class, but the $SSA \approx I \sum_{j=1}^{J} \hat{\alpha}_j^2$. Likewise $SSB \approx J \sum \sum_{i=1}^{I} \hat{\beta}_i^2$