**일원분산분석 :3그룹 이상의 평균이 모두 같은지 검정**

F-test : F값이 작으면 관찰된 평균 간의 차이가 우연히 발생할 수 있다고 생각하고, 크면 그룹 간에 유의한 차이가 있다고 본다. F=그룹간 변동량/그룹 내 변동량

**다중비교:** 어떤쌍의 차이로 귀무가설이 기각되었는지 조사

* **Dunnett :하나의 대조군(reference)를 나머지 비교군(treatment)들과 비교할 때 사용**
* **Tukey :가능한 모든 조합의 쌍을 조사**

data("PlantGrowth")

weight group

1 4.17 ctrl

2 5.58 ctrl

11 4.81 trt1

12 4.17 trt1

13 4.41 trt1

28 6.15 trt2

29 5.80 trt2

30 5.26 trt2

ctrl대조군, trt1, trt2실험군

> out=lm(weight~group,data=PlantGrowth)

> out

Call:

lm(formula = weight ~ group, data = PlantGrowth)

Coefficients:

(Intercept) grouptrt1 grouptrt2

5.032 -0.371 0.494

*-> Trt1의평균4.661과 레퍼런스인 ctrl의평균5.032의 차이인 -0.3701*

**anova(out) #분산분석표**

Analysis of Variance Table

Response: weight

Df Sum Sq Mean Sq F value Pr(>F)

group 2 3.7663 1.8832 4.8461 0.01591 \*

Residuals 27 10.4921 0.3886

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

*F=4.8461 p값=0.01591<0.05로 세 그룹의 평균들이 모두 같다는 귀무가설을 기각한다. 그럼 어느 쌍에 차이가 있는지* ***다중비교****로 좀 더 조사!!*

1) 대조군인 ctrl과 2개의 비교군(trt1,trt2)를 비교하기 위해 Dunnett의 방법이용

library(multcomp)

dunnett=**glht**(out, **linfct=mcp(group="Dunnett")**)

#분산분석의 결과물을 out으로 저장한후 glht()를 사용한 다중비교함

summary(dunnett)

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Dunnett Contrasts

Fit: lm(formula = weight ~ group, data = PlantGrowth)

Linear Hypotheses:

Estimate Std. Error t value Pr(>|t|)

trt1 - ctrl == 0 -0.3710 0.2788 -1.331 0.323

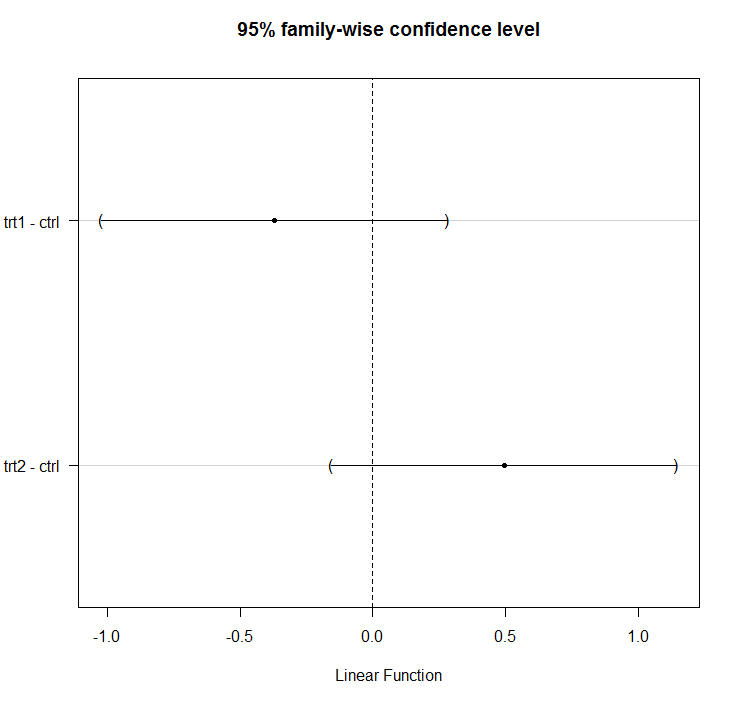
trt2 - ctrl == 0 0.4940 0.2788 1.772 0.153

(Adjusted p values reported -- single-step method)

->둘다 유의하지 않음

plot(dunnett)

*95%신뢰구간이 0을 포함하므로 유의수준 0.05에서 유의하지 않음을 알수 있음*



2)tukey 다중비교

**tukey=glht(out, linfct=mcp(group="Tukey"))**

summary(tukey)

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = weight ~ group, data = PlantGrowth)

Linear Hypotheses:

Estimate Std. Error t value Pr(>|t|)

trt1 - ctrl == 0 -0.3710 0.2788 -1.331 0.391

trt2 - ctrl == 0 0.4940 0.2788 1.772 0.198

trt2 - trt1 == 0 0.8650 0.2788 3.103 0.012 \*

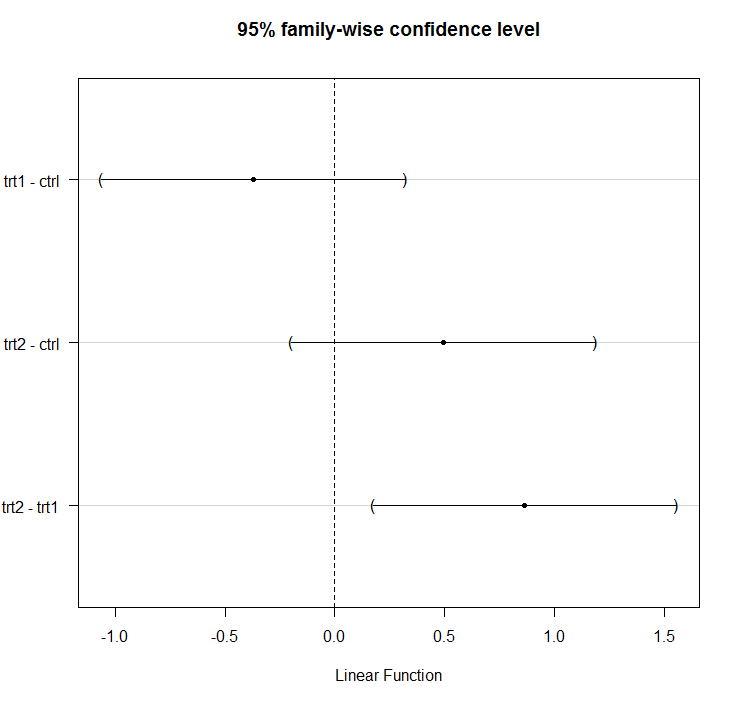
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- single-step method)

*Trt2와 trt1이 유의한 차이가 있다. 그래프로도 확인할 수 있다.*

plot(tukey)



**이원분산분석:두개의 그룹변수를 가짐. 두 그룹변수들의 효과뿐만아니라 두 그룹변수들이 서로 어떤 영향을 미치는지 교호작용을 볼 수 있다.**

warpbreaks

breaks wool tension

1 26 A L

2 30 A L

16 35 A M

17 30 A M

18 36 A M

19 36 A H

20 21 A H

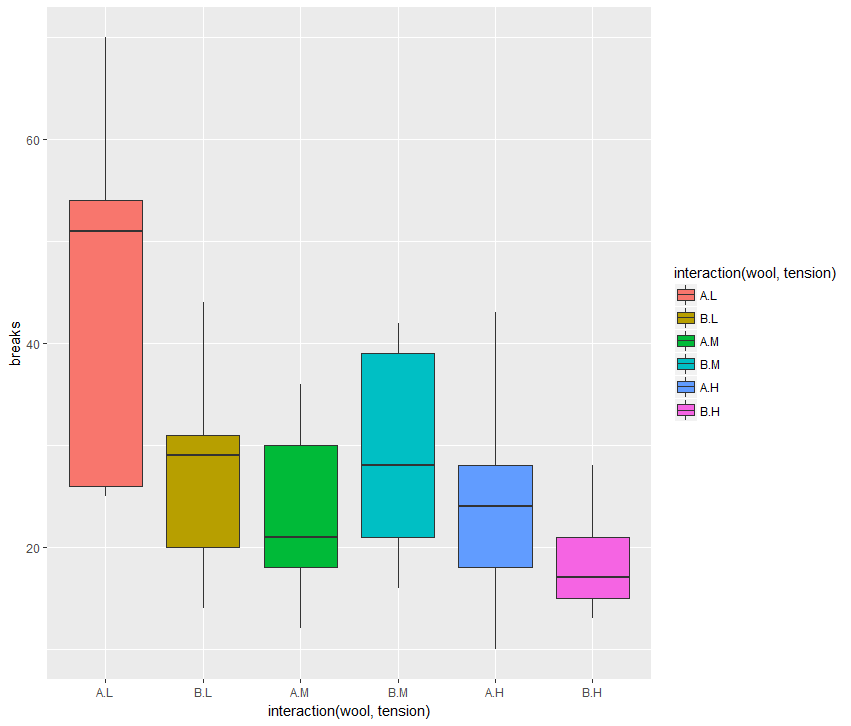
> levels(warpbreaks$tension)

[1] "H" "L" "M"

> warpbreaks$tension=factor(warpbreaks$tension, levels=c("L","M","H"))

**#tension의 레벨을 낮은순부터 높은순으로 바꿔준다**

ggplot(warpbreaks, aes(x=interaction(wool,tension), y=breaks, fill=interaction(wool,tension)))+geom\_boxplot()

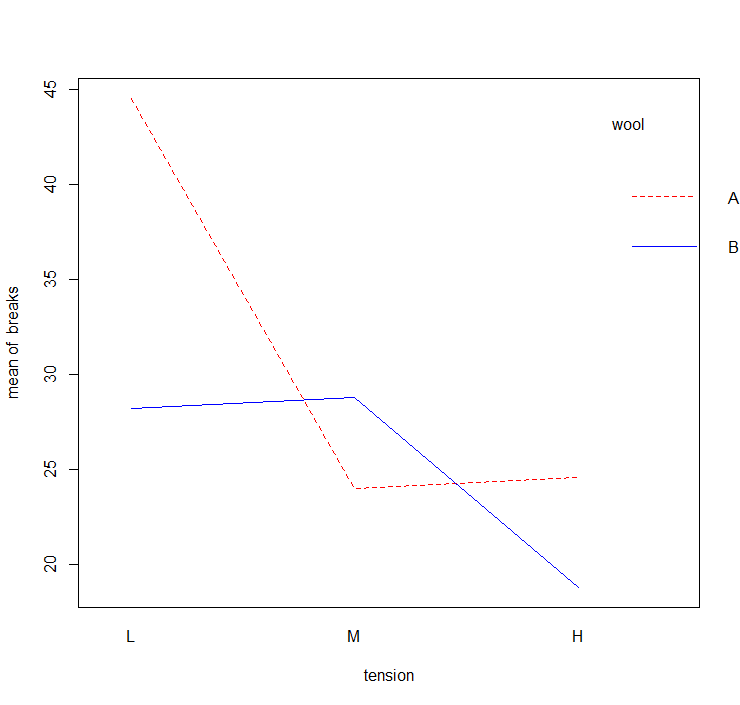


Wool,tension조합의 분포차이가 있다.

**#교호작용 interactionplot:두 그룹변수의 조합으로 y의 평균을 그래프에 넣어 두 그룹 변수가 서로 y의 평균에 영향을 주는지 알아보는 방법**

> interaction.plot(warpbreaks$tension, warpbreaks$wool, warpbreaks$breaks, col=c("red","blue"))

Interaction.plot(x-축에 올 그룹변수, 두번째그룹변수, y) 여기서 wool이 두종류이므로 두가지색을 지정해주었고, 종류가 셋이라면 세가지색을 지정해야,,

 *인터랙션이 없다면 두개의 선이 비슷한 거리를 유지하면서 평행에 가까울텐데, 두개의 선이 서로 만나므로 인터랙션이 존재함!*

> out=lm(breaks~wool+tension+**wool\*tension**, data=warpbreaks)

> anova(out)

Analysis of Variance Table

Response: breaks

Df Sum Sq Mean Sq F value Pr(>F)

wool 1 450.7 450.67 3.7653 0.0582130 .

tension 2 2034.3 1017.13 8.4980 0.0006926 \*\*\*

wool:tension 2 1002.8 501.39 4.1891 0.0210442 \*

Residuals 48 5745.1 119.69

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

인터랙션 텀인 wool\*tension을 넣어주어보면,*Wool은 유의하지는 않지만 0,.05에 근접하고 tension은 유의함. 교호작용도 유의하다*