

## Part 2. R 통계분석 (데이터 분석 전문가 양성과정)

### 10

# 일원 분산분석과 이원 분산분석

경북대학교 배준현 교수  
(joonion@knu.ac.kr)



## 10. 일원 분산분석과 이원 분산분석

- 분산분석: *analysis of variance* (*ANOVA*)
  - 일원 분산분석: *one-way* ANOVA
  - 이원 분산분석: *two-way* ANOVA
  - 반복측정 분산분석: *repeated-measures* two-way ANOVA
  - 공분산 분석: *analysis of covariance* (*ANCOVA*)
  - 다변량 분산분석: *multivariate* analysis of variance (*MANOVA*)



## 10. 일원 분산분석과 이원 분산분석

### ■ 일원 분산분석: *one-way* ANOVA

- 집단을 구분하는 독립변수가 한 개인 경우에 모집단 간 평균의 동일성 검정
- 예제) 내장 데이터셋인 InsectSprays 데이터셋을 이용한 일원 분산분석

```
> str(InsectSprays)
```

```
'data.frame': 72 obs. of 2 variables:
```

```
$ count: num 10 7 20 14 14 12 10 23 17 20 ...
```

```
$ spray: Factor w/ 6 levels "A","B","C","D",...: 1 1 1 1 1 1 1 1 1 1 ...
```



## 10. 일원 분산분석과 이원 분산분석

- 집단별 요약통계량 구하기: 살충제별 평균, 분산, 표준편차

```
> tapply(InsectSprays$count, InsectSprays$spray, mean)
      A      B      C      D      E      F
14.500000 15.333333  2.083333  4.916667  3.500000 16.666667
```

```
> tapply(InsectSprays$count, InsectSprays$spray, sd)
      A      B      C      D      E      F
4.719399 4.271115 1.975225 2.503028 1.732051 6.213378
```

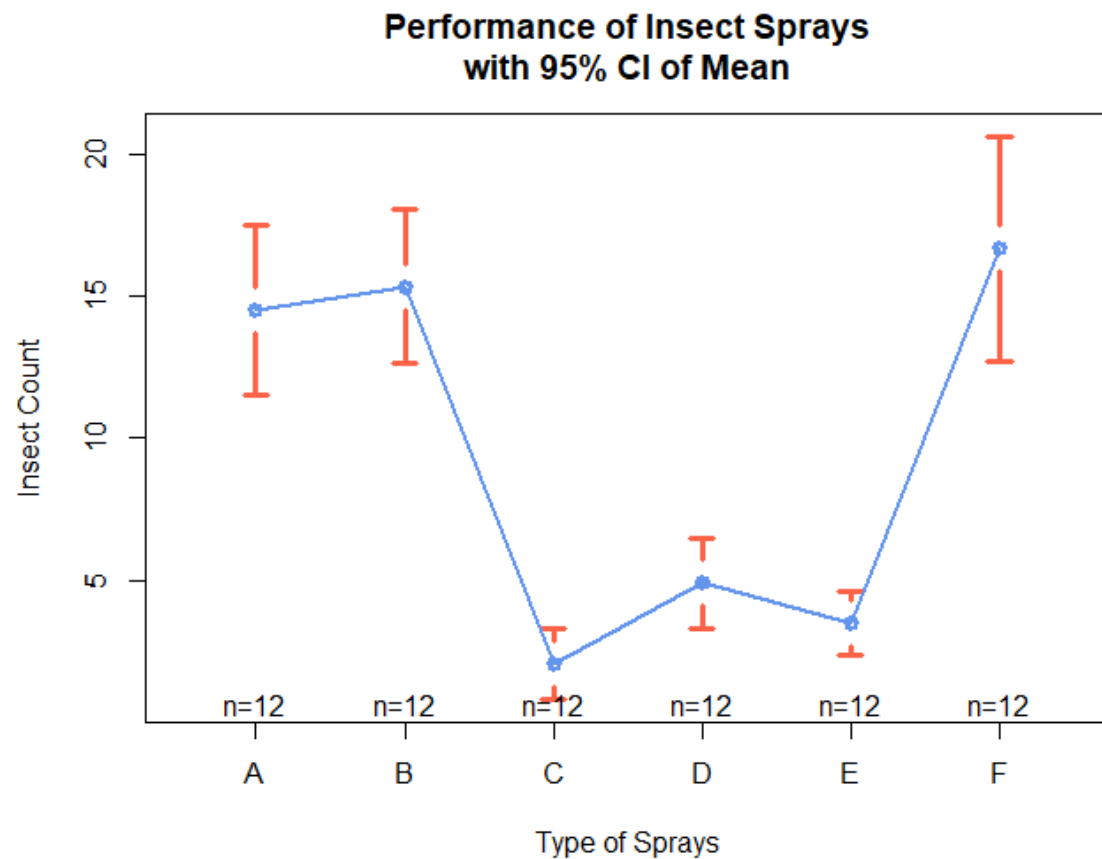
```
> tapply(InsectSprays$count, InsectSprays$spray, length)
      A      B      C      D      E      F
12 12 12 12 12 12
```



## 10. 일원 분산분석과 이원 분산분석

- 집단별 평균도표(mean plot) 그리기

```
library(gplots)
plotmeans(count ~ spray, data=InsectSprays,
main="Performance of Insect Sprays\nwith 95% CI of Mean",
  xlab="Type of Sprays",
  ylab="Insect Count",
  barcol="tomato",
  barwidth=3,
  col="cornflowerblue",
  lwd=2)
```

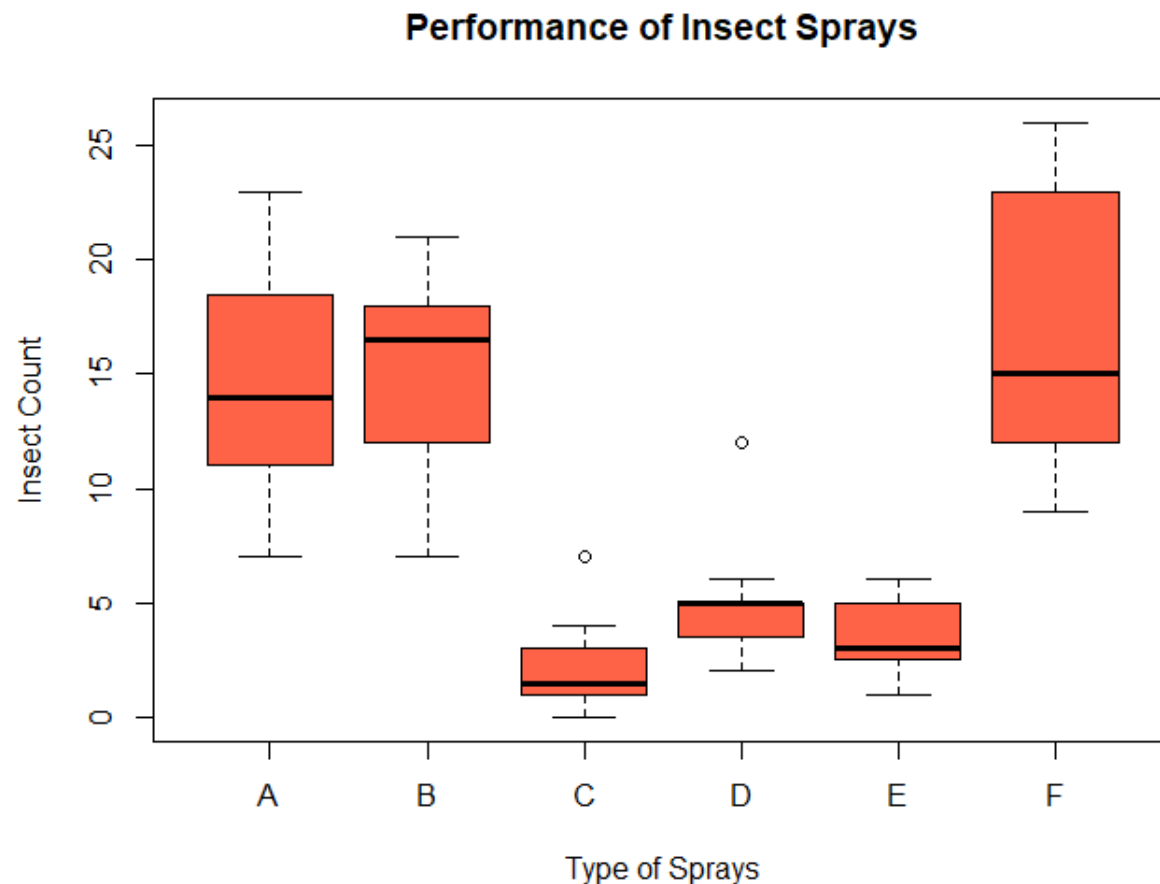




## 10. 일원 분산분석과 이원 분산분석

- 집단별 상자도표(box plot) 그리기

```
boxplot(count ~ spray, data=InsectSprays, col="tomato",  
        xlab="Type of Sprays", ylab="Insect Count",  
        main="Performance of Insect Sprays")
```





## 10. 일원 분산분석과 이원 분산분석

- 일원 분산분석: 살충제 간 살충효과의 평균이 동일한지 통계적으로 검정

```
> sprays.aov <- aov(count ~ spray, data=InsectSprays)
```

```
> sprays.aov
```

Call:

```
aov(formula = count ~ spray, data = InsectSprays)
```

Terms:

	spray
Sum of Squares	2668.833
Deg. of Freedom	5

	Residuals
Sum of Squares	1015.167
Deg. of Freedom	66

Residual standard error: 3.921902

Estimated effects may be unbalanced



## 10. 일원 분산분석과 이원 분산분석

```
> summary(sprays.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
spray	5	2669	533.8	34.7	<2e-16	***
Residuals	66	1015	15.4			

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1





## 10. 일원 분산분석과 이원 분산분석

- 분산분석은 귀무가설만 기각: 모집단평균이 모두 동일하다는 주장을 기각
- 개별 집단 간 평균의 차이 확인: 어느 집단과 어느 집단이 서로 달라서 이러한 결과가 나왔는가?

```
> model.tables(sprays.aov, type="mean")
```

Tables of means

Grand mean

9.5

spray

spray

	A	B	C	D	E	F
	14.500	15.333	2.083	4.917	3.500	16.667

```
> model.tables(sprays.aov, type="effects")
```

Tables of effects

spray

spray

	A	B	C	D	E	F
	5.000	5.833	-7.417	-4.583	-6.000	7.167



## 10. 일원 분산분석과 이원 분산분석

- 사후검정(post-hoc test): 두 살충제 간의 살충효과가 통계적으로 유의한지 검정
- 다중비교를 통한 Tukey's HSD(honest significant diff.) 검정

```
> sprays.compare <- TukeyHSD(sprays.aov)
```

```
> sprays.compare
```

```
Tukey multiple comparisons of means  
95% family-wise confidence level
```

```
Fit: aov(formula = count ~ spray, data = InsectSprays)
```

```
$spray
```

	diff	lwr	upr	p adj
B-A	0.8333333	-3.866075	5.532742	0.9951810
C-A	-12.4166667	-17.116075	-7.717258	0.0000000
D-A	-9.5833333	-14.282742	-4.883925	0.0000014
E-A	-11.0000000	-15.699409	-6.300591	0.0000000



## 10. 일원 분산분석과 이원 분산분석

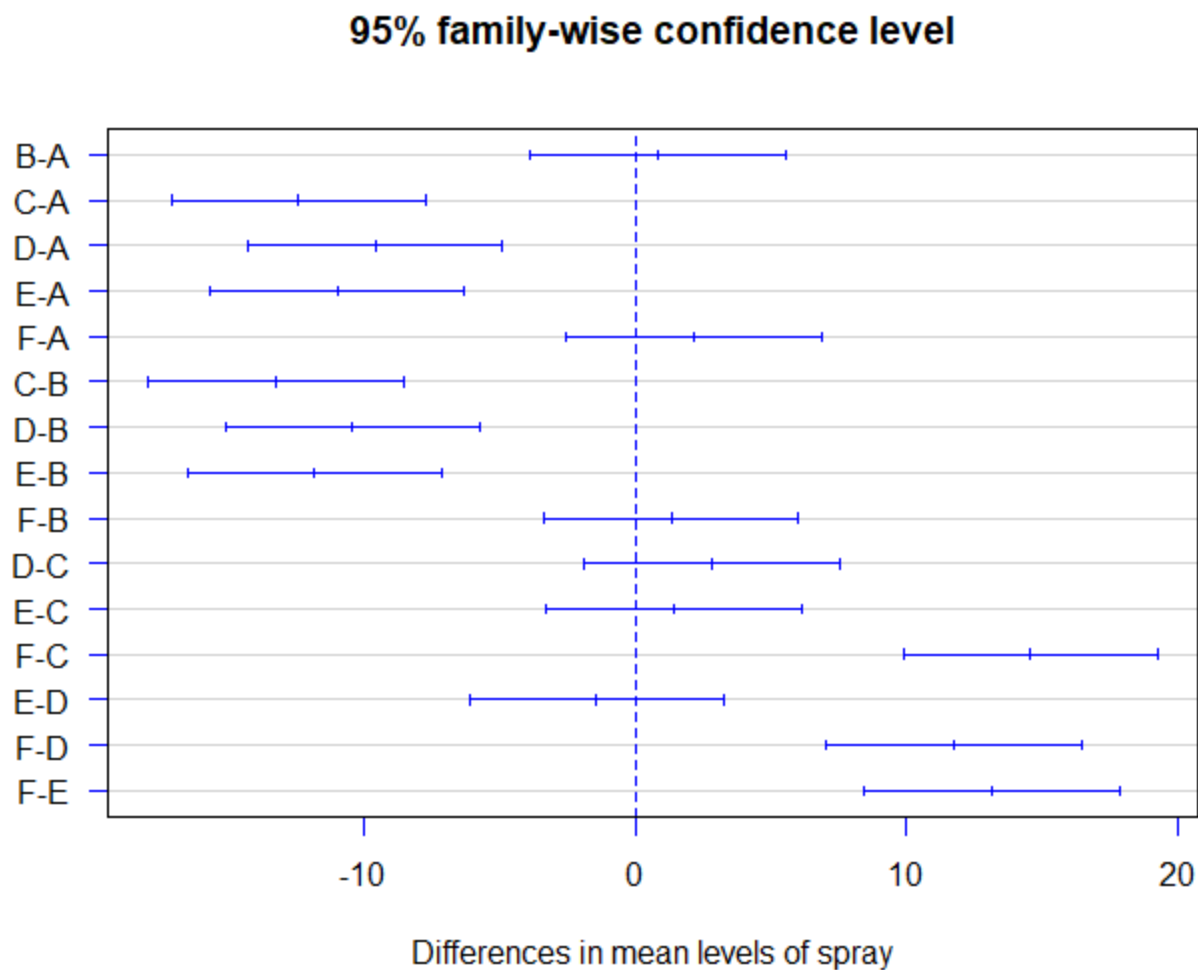
F-A	2.1666667	-2.532742	6.866075	0.7542147
C-B	-13.2500000	-17.949409	-8.550591	0.0000000
D-B	-10.4166667	-15.116075	-5.717258	0.0000002
E-B	-11.8333333	-16.532742	-7.133925	0.0000000
F-B	1.3333333	-3.366075	6.032742	0.9603075
D-C	2.8333333	-1.866075	7.532742	0.4920707
E-C	1.4166667	-3.282742	6.116075	0.9488669
F-C	14.5833333	9.883925	19.282742	0.0000000
E-D	-1.4166667	-6.116075	3.282742	0.9488669
F-D	11.7500000	7.050591	16.449409	0.0000000
F-E	13.1666667	8.467258	17.866075	0.0000000

```
> sprays.compare$spray['D-C',]  
      diff      lwr      upr      p adj  
2.8333333 -1.8660752  7.5327418 0.4920707
```



## 10. 일원 분산분석과 이원 분산분석

```
plot(TukeyHSD(sprays.aov), col="blue", las=1)
```





## 10. 일원 분산분석과 이원 분산분석

- Tukey HSD 다중비교 결과를 `glht()` 함수로 확인

```
> library(multcomp)
> tuk.hsd <- glht(model=sprays.aov, linfct=mcp(spray="Tukey"))
```

General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

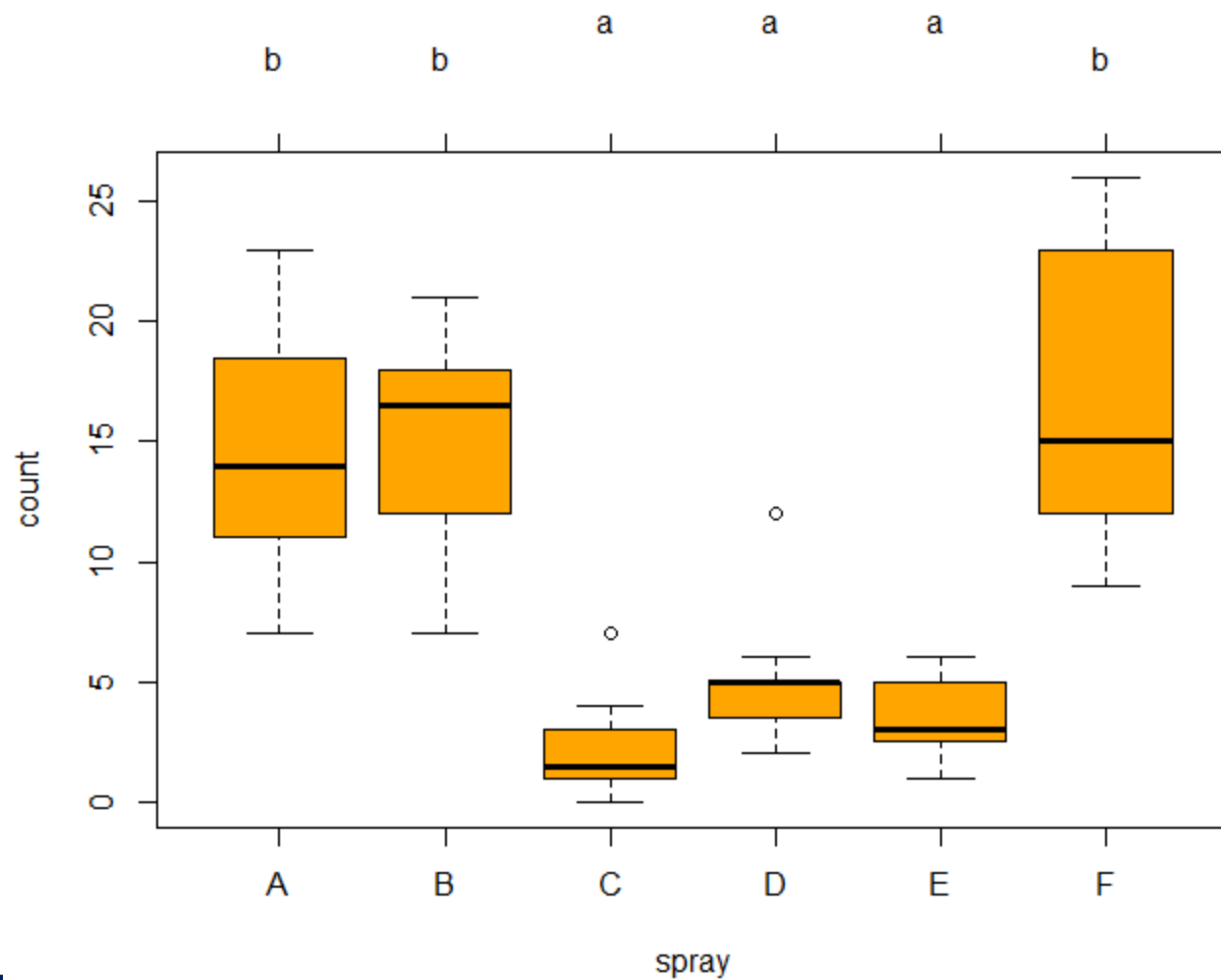
Linear Hypotheses:

	Estimate
B - A == 0	0.8333
C - A == 0	-12.4167
D - A == 0	-9.5833
E - A == 0	-11.0000
F - A == 0	2.1667
C - B == 0	-13.2500
D - B == 0	-10.4167
E - B == 0	-11.8333
F - B == 0	1.3333
D - C == 0	2.8333



## 10. 일원 분산분석과 이원 분산분석

```
plot(cld(tuk.hsd, level=0.05), col="orange")
```





## 10. 일원 분산분석과 이원 분산분석

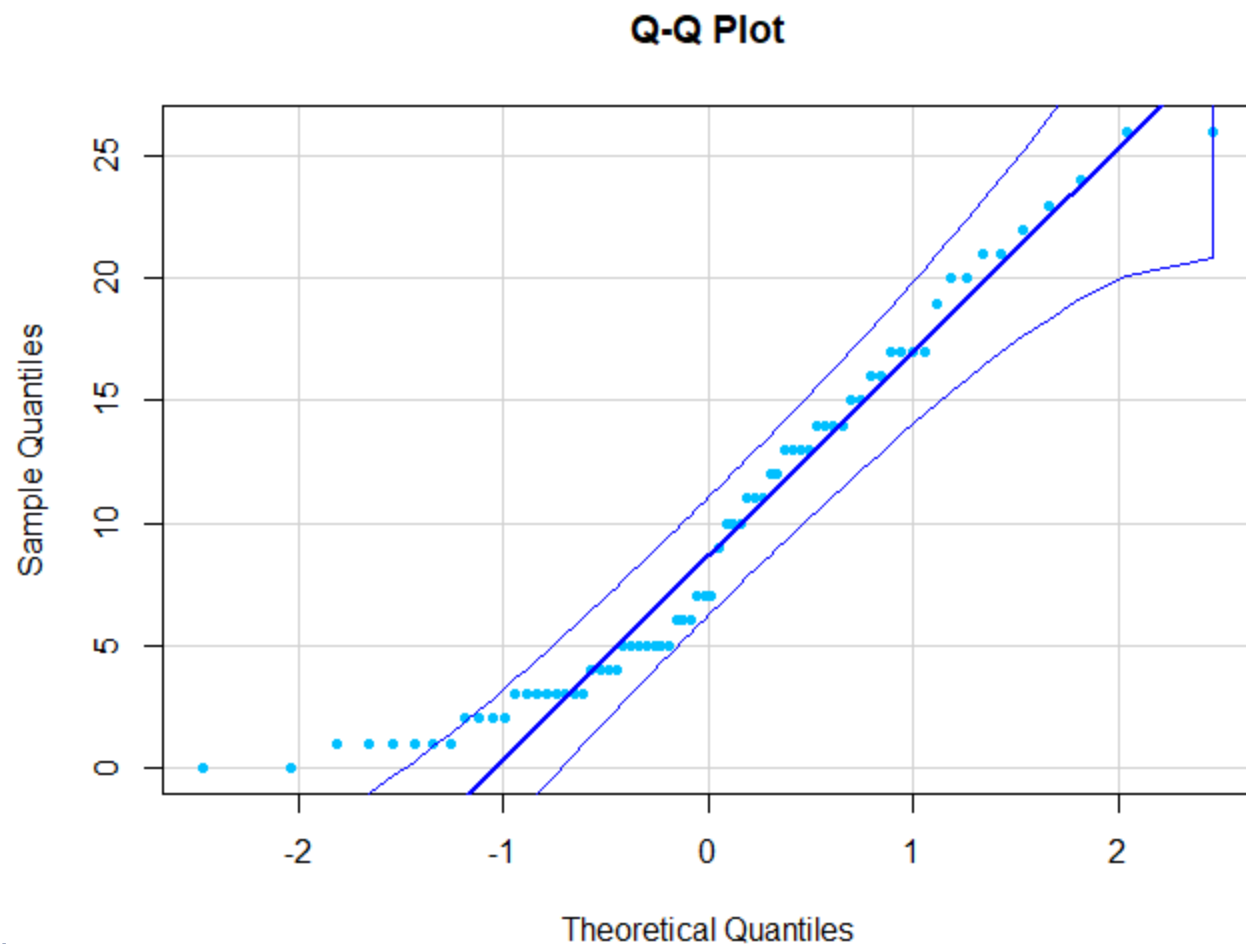
### ■ 분산분석의 가정과 진단

- 분산분석을 위해서는 관측값은 일반적으로 다음 조건을 충족해야 함
  - 정규성(*normality*): 종속변수는 정규분포를 따른다.
  - 등분산성(*equality of variance*): 각 집단의 분포는 모두 동일한 분포를 가진다.



## 10. 일원 분산분석과 이원 분산분석

```
library(car)
qqPlot(InsectSprays$count, pch=20, col="deepskyblue", id=FALSE,
       main="Q-Q Plot", xlab="Theoretical Quantiles", ylab="Sample Quantiles")
```







## 10. 일원 분산분석과 이원 분산분석

- 샤피로-윌크 검정을 통해 정규성 충족 여부를 통계적으로 검정

```
> shapiro.test(InsectSprays$count)
```

Shapiro-Wilk normality test

```
data: InsectSprays$count  
W = 0.9216, p-value = 0.0002525
```

- 이상점의 존재여부를 통계적으로 검정

```
> outlierTest(sprays.aov)
```

No Studentized residuals with Bonferroni  $p < 0.05$

Largest  $|rstudent|$ :

	rstudent	unadjusted p-value	Bonferroni p
69	2.590966	0.011804	0.8499



## 10. 일원 분산분석과 이원 분산분석

- 레벤 검정이나 바틀렛 검정을 통해 집단 간 분산의 동일성 여부 검정

```
> leveneTest(count ~ spray, data=InsectSprays)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	5	3.8214	0.004223 **
	66		

---

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

```
> bartlett.test(count ~ spray, data=InsectSprays)
```

Bartlett test of homogeneity of variances

data: count by spray

Bartlett's K-squared = 25.96, df = 5, p-value = 9.085e-05



## 10. 일원 분산분석과 이원 분산분석

- 등분산의 가정을 충족하지 못할때: `oneway.test()` 함수를 이용한 일원분산분석

```
> oneway.test(count ~ spray, data=InsectSprays)
```

One-way analysis of means (not assuming equal variances)

data: count and spray

F = 36.065, num df = 5.000, denom df = 30.043, p-value = 7.999e-12

```
> oneway.test(count ~ spray, data=InsectSprays, var.equal=TRUE)
```

One-way analysis of means

data: count and spray

F = 34.702, num df = 5, denom df = 66, p-value < 2.2e-16



## 10. 일원 분산분석과 이원 분산분석

### ■ 이원 분산분석: *two-way* ANOVA

- 집단을 구분하는 독립변수가 두 개인 경우에 모집단 간 평균의 동일성 검정
- 두 개의 주 효과와 하나의 상호작용 효과를 검정
  - **주효과** 검정: 각 독립변수에 의해 만들어지는 집단 간 평균의 차이 검정
    - 두 독립변수가 각각 개별적으로 종속변수에 유의한 영향을 미치는가?
  - **상호작용효과** 검정: 두 독립변수의 조합에 의해 만들어지는 평균의 차이 검정
    - 두 독립변수의 조합이 종속변수와 유의한 영향관계를 갖는가?



## 10. 일원 분산분석과 이원 분산분석

- ToothGrowth 데이터셋을 이용한 이원 분산분석 수행

```
> str(ToothGrowth)
```

```
'data.frame':  60 obs. of  3 variables:
```

```
$ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
```

```
$ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
```

```
$ dose: Factor w/ 3 levels "low","med","high": 1 1 1 1 1 1 1 1 1 1 ...
```

```
> ToothGrowth$dose <- factor(ToothGrowth$dose,  
                             levels=c(0.5, 1.0, 2.0), labels=c("low", "med", "high"))
```



## 10. 일원 분산분석과 이원 분산분석

```
> ToothGrowth[seq(1, 60, 5),]
```

	len	supp	dose
1	4.2	VC	low
6	10.0	VC	low
11	16.5	VC	med
16	17.3	VC	med
21	23.6	VC	high
26	32.5	VC	high
31	15.2	OJ	low
36	10.0	OJ	low
41	19.7	OJ	med
46	25.2	OJ	med
51	25.5	OJ	high
56	30.9	OJ	high



## 10. 일원 분산분석과 이원 분산분석

- 각 집단별 요약통계량 계산: 각 집단의 크기와 집단별 평균 및 표준편차

```
> with(ToothGrowth, tapply(len, list(supp, dose), length))
```

	low	med	high
OJ	10	10	10
VC	10	10	10

```
> with(ToothGrowth, tapply(len, list(supp, dose), mean))
```

	low	med	high
OJ	13.23	22.70	26.06
VC	7.98	16.77	26.14

```
> with(ToothGrowth, tapply(len, list(supp, dose), sd))
```

	low	med	high
OJ	4.459709	3.910953	2.655058
VC	2.746634	2.515309	4.797731



## 10. 일원 분산분석과 이원 분산분석

- aov() 함수를 이용하여 이원분산분석 수행

```
> ToothGrowth.aov <- aov(len ~ supp * dose, data=ToothGrowth)
> ToothGrowth.aov <- aov(len ~ supp + dose + supp:dose, data=ToothGrowth)
```

```
> summary(ToothGrowth.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
supp	1	205.4	205.4	15.572	0.000231	***
dose	2	2426.4	1213.2	92.000	< 2e-16	***
supp:dose	2	108.3	54.2	4.107	0.021860	*
Residuals	54	712.1	13.2			

---

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





## 10. 일원 분산분석과 이원 분산분석

- `model.tables()` 함수를 이용하여 각 효과에 따른 요약통계량 계산

```
> model.tables(ToothGrowth.aov, type="means")
```

```
Tables of means
```

```
Grand mean
```

```
18.81333
```

```
supp
```

```
supp
```

```
0J
```

```
VC
```

```
20.663 16.963
```

```
dose
```

```
dose
```

```
low
```

```
med
```

```
high
```

```
10.605 19.735 26.100
```

```
supp:dose
```

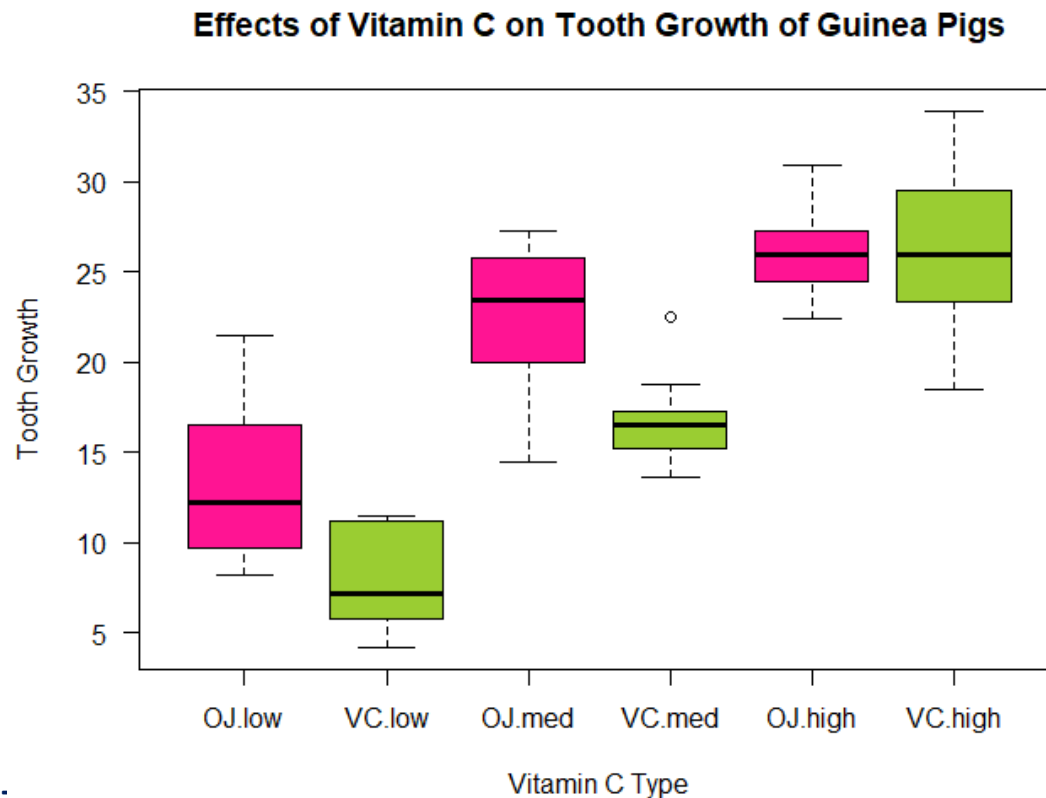
```
dose
```



## 10. 일원 분산분석과 이원 분산분석

- 상자도표로 비타민C 보충제와 투여량의 조합에 따른 이빨 성장의 정도를 파악

```
boxplot(len ~ supp * dose, data=ToothGrowth,  
        col=c("deeppink", "yellowgreen"), las=1,  
        xlab="Vitamin C Type", ylab="Tooth Growth",  
        main="Effects of Vitamin C on Tooth Growth of Guinea Pigs")
```

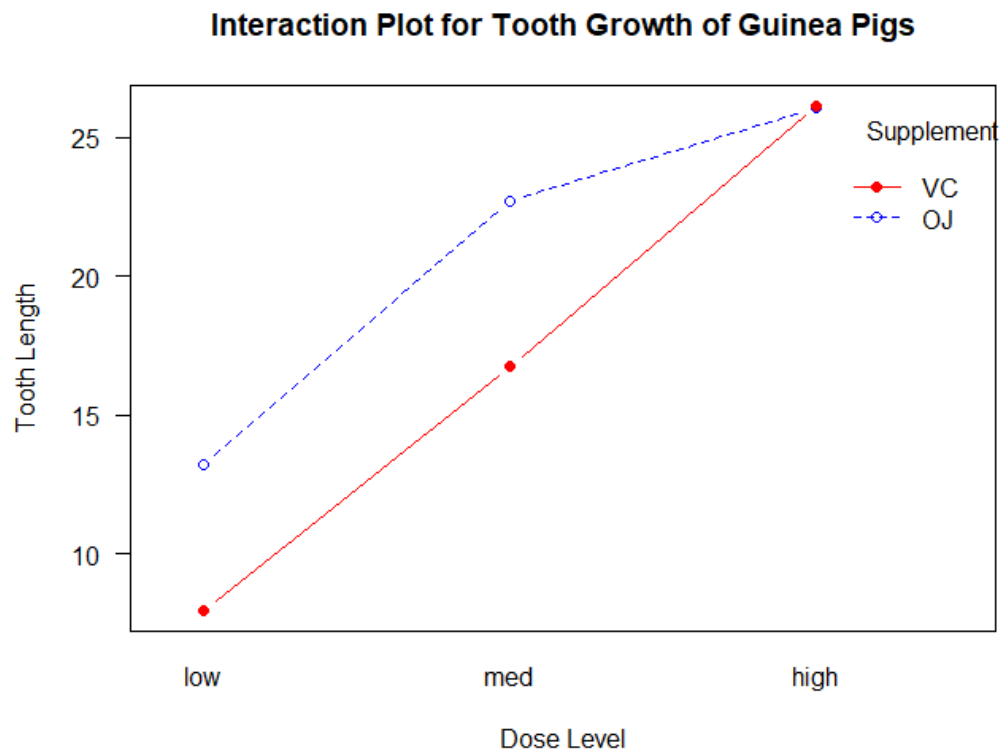




## 10. 일원 분산분석과 이원 분산분석

- 상호작용도표(interaction plot)으로 주효과와 상호작용효과 파악

```
interaction.plot(x.factor=ToothGrowth$dose, trace.factor=ToothGrowth$supp,  
  response=ToothGrowth$len, las=1, type="b",  
  pch=c(1, 19), col=c("blue", "red"), trace.label="Supplement",  
  xlab="Dose Level", ylab="Tooth Length",  
  main="Interaction Plot for Tooth Growth of Guinea Pigs")
```

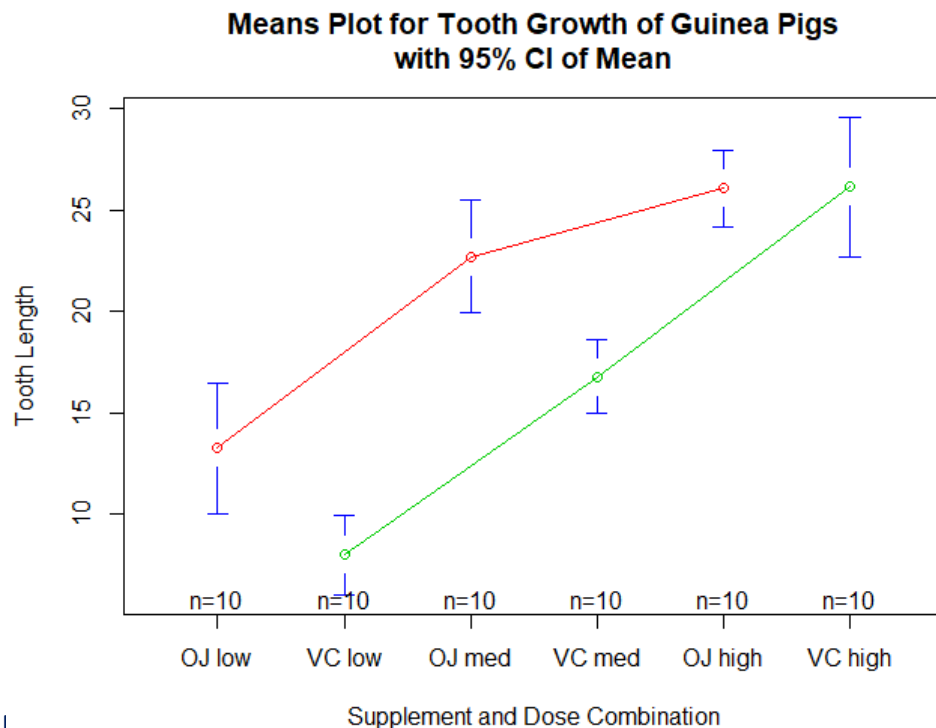




## 10. 일원 분산분석과 이원 분산분석

- 평균도표(means plot)로 독립변수의 조합별 평균, 신뢰구간, 표본크기를 함께 확인

```
library(gplots)
plotmeans(len ~ interaction(supp, dose, sep=" "), data=ToothGrowth,
          connect=list(c(1,3,5), c(2,4,6)), col=c("red", "green3"),
          xlab="Supplement and Dose Combination", ylab="Tooth Length",
          main="Means Plot for Tooth Growth of Guinea Pigs\nwith 95% CI of Mean")
```

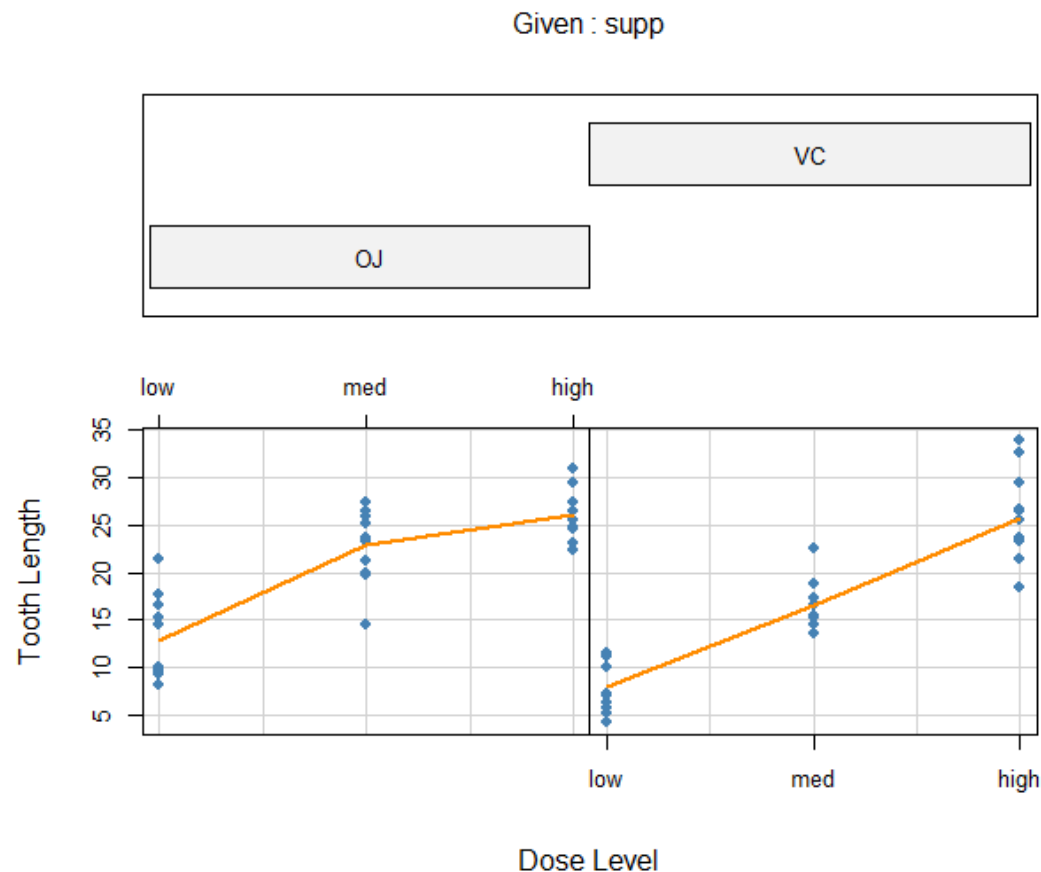




## 10. 일원 분산분석과 이원 분산분석

- 조건부도표(conditioning plot)로 집단 간의 영향의 차이를 파악

```
coplot(len ~ dose | supp, data=ToothGrowth,  
       col="steelblue", pch=19, panel=panel.smooth, lwd=2, col.smooth="darkorange",  
       xlab="Dose Level", ylab="Tooth Length")
```



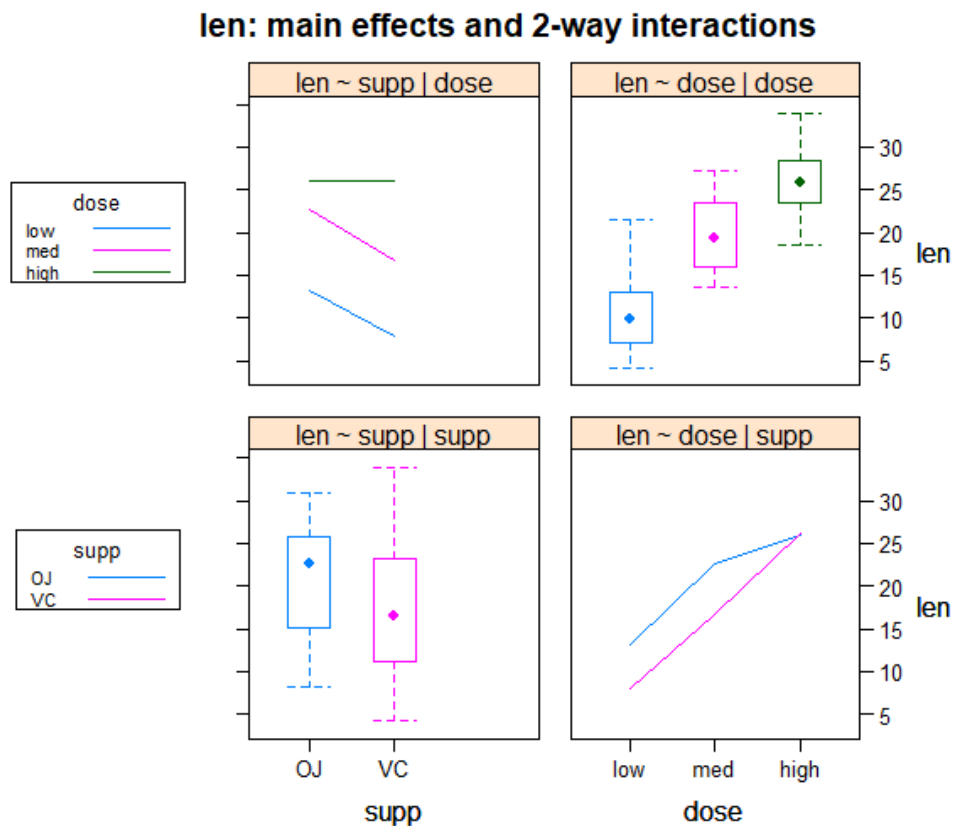


## 10. 일원 분산분석과 이원 분산분석

- `interaction2wt()` 함수로 주효과와 상호작용효과를 동시에 파악

```
library(HH)
```

```
interaction2wt(len ~ supp * dose, data=ToothGrowth)
```





## 10. 일원 분산분석과 이원 분산분석

- TukeyHSD() 함수를 이용한 사후분석을 통해 집단 간 차이에 대한 추가 검정 수행

```
> TukeyHSD(ToothGrowth.aov)
```

```
Tukey multiple comparisons of means  
95% family-wise confidence level
```

.....(생략)

```
> TukeyHSD(ToothGrowth.aov, which=c("dose"), conf.level=0.99)
```

```
Tukey multiple comparisons of means  
99% family-wise confidence level
```

```
Fit: aov(formula = len ~ supp + dose + supp:dose, data = ToothGrowth)
```

```
$dose
```

	diff	lwr	upr	p adj
med-low	9.130	5.637681	12.622319	0.0e+00
high-low	15.495	12.002681	18.987319	0.0e+00
high-med	6.365	2.872681	9.857319	2.7e-06

*Any Questions?*

