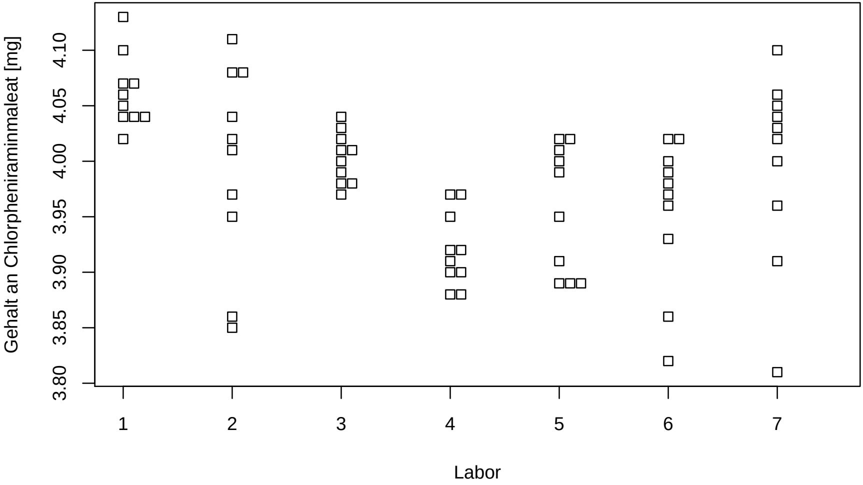
Ein Beispiel

7 verschiedene Labors haben jeweils 10 Messungen des Chlorpheniraminmaleat-Gehalts von Medikamentenproben vorgenommen.

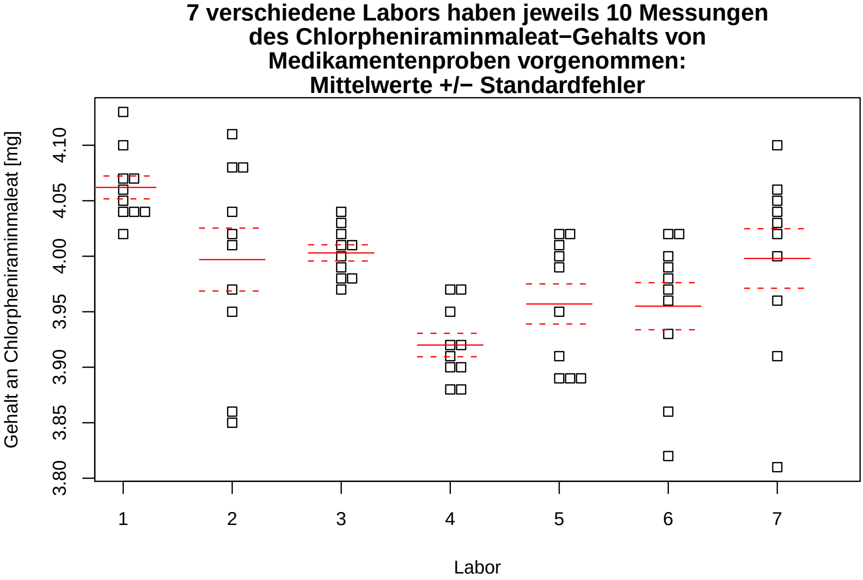
Die Daten liegen in der Datei `chlorpheniraminmaleat.txt` als Tabelle vor:

```
  Gehalt Labor
1 4.13 1
2 4.07 1
3 4.04 1
4 4.07 1
5 4.05 1
6 4.04 1
7 4.02 1
8 4.06 1
9 4.1 1
10 4.04 1
11 3.86 2
12 3.85 2
13 4.08 2
14 4.11 2
15 4.08 2
16 4.01 2
17 4.02 2
18 4.04 2
...
```

7 verschiedene Labors haben jeweils 10 Messungen des Chlorpheniraminmaleat−Gehalts von Medikamentenproben vorgenommen:



Daten aus R.D. Kirchhoefer, Semiautomated method for the analysis of chlorpheniramine maleate tablets: collaborative study, *J. Assoc. Off. Anal. Chem.* 62(6):1197-1201 (1979),
zitiert nach John A. Rice, *Mathematical statistics and data analysis*, 2nd ed., Wadsworth, 1995



Daten aus R.D. Kirchhoefer, Semiautomated method for the analysis of chlorpheniramine maleate tablets: collaborative study, *J. Assoc. Off. Anal. Chem.* 62(6):1197-1201 (1979),
zitiert nach John A. Rice, *Mathematical statistics and data analysis*, 2nd ed., Wadsworth, 1995

Beachte: Die Labore sind sind mit Zahlen nummeriert. Damit R das nicht als numerische Werte sondern als Nummern der Labore auffasst, müssen wir die Variable “Labor” in einen sog. Factor umwandeln:

```
> chlor <- read.table("chlorpheniraminmaleat.txt")
> str(chlor)
'data.frame': 70 obs. of  2 variables:
 $ Gehalt: num  4.13 4.07 4.04 4.07 4.05 4.04 4.02 4.06 4.1
 $ Labor : int  1 1 1 1 1 1 1 1 1 1 ...
> chlor$Labor <- as.factor(chlor$Labor)
> str(chlor)
'data.frame': 70 obs. of  2 variables:
 $ Gehalt: num  4.13 4.07 4.04 4.07 4.05 4.04 4.02 4.06 4.1
 $ Labor : Factor w/ 7 levels "1","2","3","4",..: 1 1 1 1 1
```

Nun können wir die Varianzanalyse durchführen:

```
> chlor.aov <- aov(Gehalt~Labor,data=chlor)
> summary(chlor.aov)
           Df  Sum Sq  Mean Sq F value    Pr(>F)
Labor       6 0.12474 0.020789  5.6601 9.453e-05 ***
Residuals  63 0.23140 0.003673
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.
```

In vorigem Beispiel:

Sei μ_i der (uns unbekannte, wahre Populations-)Mittelwert der Messungen aus Labor i, für $i = 1, \dots, 7$.

In vorigem Beispiel:

Sei μ_i der (uns unbekannte, wahre Populations-)Mittelwert der Messungen aus Labor *i*, für $i = 1, \dots, 7$.

Die Varianzanalyse zeigte, dass es signifikante Unterschiede zwischen den Laboren gibt.

In vorigem Beispiel:

Sei μ_i der (uns unbekannte, wahre Populations-)Mittelwert der Messungen aus Labor i, für $i = 1, \dots, 7$.

Die Varianzanalyse zeigte, dass es signifikante Unterschiede zwischen den Laboren gibt.

Aber welche Labore unterscheiden sich signifikant?

In vorigem Beispiel:

Sei μ_i der (uns unbekannte, wahre Populations-)Mittelwert der Messungen aus Labor i, für $i = 1, \dots, 7$.

Die Varianzanalyse zeigte, dass es signifikante Unterschiede zwischen den Laboren gibt.

Aber welche Labore unterscheiden sich signifikant?

Wir könnten dazu für jedes Paar i, j von Labors jeweils einen (zwei-Stichproben-)t-Test durchführen, um die Nullhypothese

$$H_{0,(i,j)} : \mu_i = \mu_j$$

(zu einem vorgegebenen Signifikanzniveau α, sagen wir $\alpha = 5\%$) zu testen.

Welche Labore unterscheiden sich signifikant?

Wert der *t*-Statistik aus paarweisen Vergleichen mittels *t*-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Welche Labore unterscheiden sich signifikant?

Wert der t-Statistik aus paarweisen Vergleichen mittels t-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Das 97,5%-Quantil der t-Verteilung mit 18 Freiheitsgraden ist 2.101

Welche Labore unterscheiden sich signifikant?

Wert der t-Statistik aus paarweisen Vergleichen mittels t-Tests:

[(zweiseitiger) zwei-Stichproben t-Tests mit Annahme gleicher Varianzen]

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Das 97,5%-Quantil der t-Verteilung mit 18 Freiheitsgraden ist 2.101, also würde für die rot markierten Paare (jeweils für sich betrachtet) ein t-Test $H_{0,(i,j)}$ zum Signifikanzniveau 5% ablehnen.

Welche Labore unterscheiden sich signifikant?

Alternative Darstellung:

p-Werte aus paarweisen Vergleichen mittels t-Tests:

Erinnerung:
p-Wert = W'keit (unter der Nullhypothese) einen mindestens so extremen Wert der t-Statistik wie den beobachteten zu erhalten

[Hier: $2(1 - F_{Student(18)}(|t|))$, mit $F_{Student(18)}$ Verteilungsfunktion der Student-Verteilung mit 18 Freiheitsgraden]

Welche Labore unterscheiden sich signifikant?

Alternative Darstellung:

p-Werte aus paarweisen Vergleichen mittels *t*-Tests:

[(zweiseitiger) zwei-Stichproben *t*-Tests mit Annahme gleicher Varianzen]

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	0.04506	0.00019	0.00000	0.00008	0.00025	0.03894
Lab2		0.84000	0.02033	0.24980	0.25103	0.97985
Lab3			0.00000	0.02982	0.04626	0.85929
Lab4				0.09398	0.15662	0.01446
Lab5					0.94356	0.22113
Lab6						0.22459

Erinnerung:

p-Wert = W'keit (unter der Nullhypothese) einen mindestens so extremen Wert der *t*-Statistik wie den beobachteten zu erhalten

[Hier: $2(1 - F_{Student(18)}(|t|))$, mit $F_{Student(18)}$ Verteilungsfunktion der Student-Verteilung mit 18 Freiheitsgraden]

Problem des multiplen Testens

Wir haben $7 \cdot 6 \cdot \frac{1}{2} = 21$ paarweise Vergleiche; auf dem 5%-Niveau zeigen einige davon Signifikanz an.

Problem des multiplen Testens

Wir haben $7 \cdot 6 \cdot \frac{1}{2} = 21$ paarweise Vergleiche; auf dem 5%-Niveau zeigen einige davon Signifikanz an.

Wenn die Nullhypothese(n) („alles nur Zufallsschwankungen") stimmt/en, verwirft man im Schnitt bei 5% der Tests die Nullhypothese zu Unrecht.

Problem des multiplen Testens

Wir haben $7 \cdot 6 \cdot \frac{1}{2} = 21$ paarweise Vergleiche; auf dem 5%-Niveau zeigen einige davon Signifikanz an.

Wenn die Nullhypothese(n) („alles nur Zufallsschwankungen") stimmt/en, verwirft man im Schnitt bei 5% der Tests die Nullhypothese zu Unrecht.

Testet man mehr als 20 mal und gelten jeweils die Nullhypothesen, wird man also im Schnitt mehr als eine Nullhypothese zu Unrecht verwerfen.

Problem des multiplen Testens

Wir haben $7 \cdot 6 \cdot \frac{1}{2} = 21$ paarweise Vergleiche; auf dem 5%-Niveau zeigen einige davon Signifikanz an.

Wenn die Nullhypothese(n) („alles nur Zufallsschwankungen“) stimmt/en, verwirft man im Schnitt bei 5% der Tests die Nullhypothese zu Unrecht.

Testet man mehr als 20 mal und gelten jeweils die Nullhypothesen, wird man also im Schnitt mehr als eine Nullhypothese zu Unrecht verwerfen.

Dieses Phänomen müssen wir bei multiplen Tests berücksichtigen

(und mit entsprechend angepassten Tests bzw. mit korrigierten *p*-Werten arbeiten).

Labor-Vergleichs-Beispiel mit Bonferroni-Korrektur

Wert der *t*-Statistik aus paarweisen Vergleichen mittels *t*-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Labor-Vergleichs-Beispiel mit Bonferroni-Korrektur

Wert der t-Statistik aus paarweisen Vergleichen mittels t-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Betrachte $\alpha = 5\%$. Hier $m = 21$, das $(1 - \frac{1}{2}\frac{\alpha}{m})$-Quantil
$(1 - \frac{1}{2}\frac{\alpha}{m} = 0.99881)$ der
t-Verteilung mit 18 Freiheitsgraden ist 3.532

Labor-Vergleichs-Beispiel mit Bonferroni-Korrektur

Wert der t-Statistik aus paarweisen Vergleichen mittels t-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Betrachte $\alpha = 5\%$. Hier $m = 21$, das $(1 - \frac{1}{2}\frac{\alpha}{m})$-Quantil $(1 - \frac{1}{2}\frac{\alpha}{m} = 0.99881)$ der t-Verteilung mit 18 Freiheitsgraden ist 3.532,

also können wir für die rot markierten Paare $H_{0,(i,j)}$ zum multiplen Signifikanzniveau 5% ablehnen.

Labor-Vergleichs-Beispiel mit Bonferroni-Korrektur

Alternativ: $21 \times$ (p-Wert aus paarweisem t-Test)

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	0.94626	0.00399	0.00000	0.00168	0.00525	0.8177
Lab2		17.64000	0.42693	5.24580	5.27163	20.576
Lab3			0.00000	0.62622	0.97146	18.045
Lab4				1.97358	3.28902	0.3036
Lab5					19.81476	4.6437
Lab6						4.7163

Für die rot markierten Paare ist der korrigierte p-Wert < 0.05.