STAT 516 fa 2026 exam 02

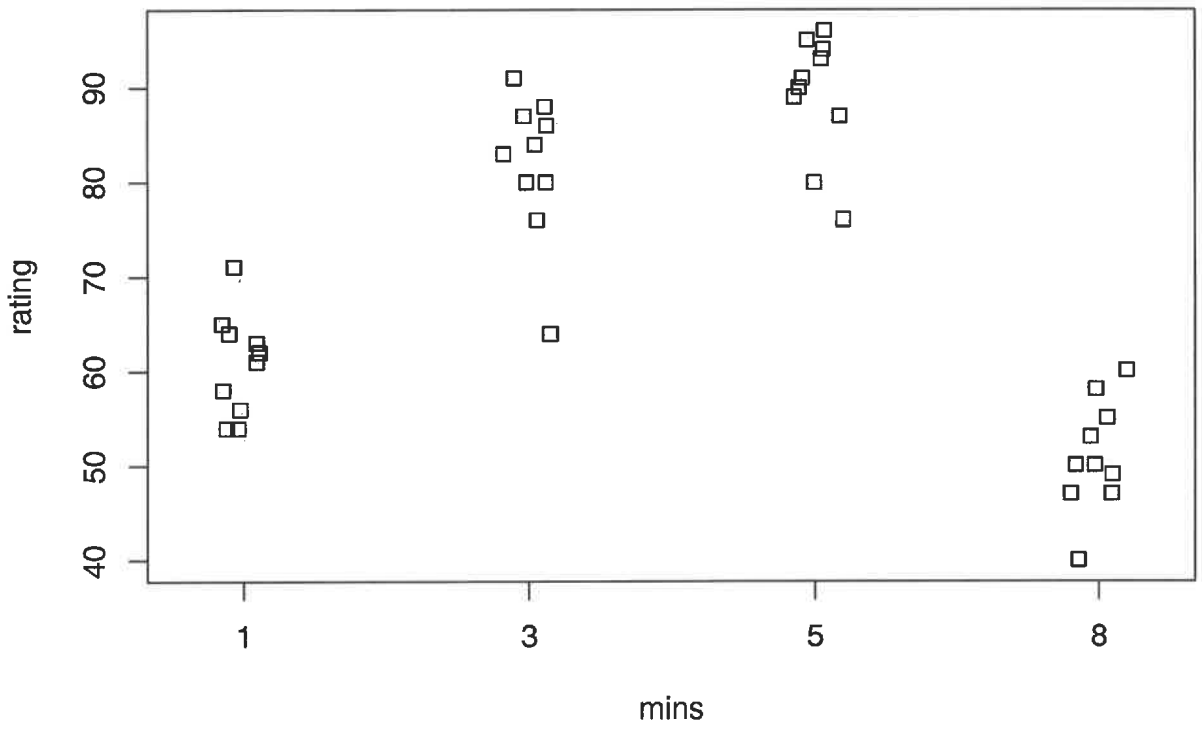
75 minutes, two pages of handwritten notes allowed, no calculators

1.

A chef wishes to learn how long one should sauté asparagus. He prepares trays of asparagus sautéed for one, three, five, and eight minutes and to each of forty guests at a catering event he gives a serving from one of the four trays such that ten guests are randomly assigned to be served from each tray. Each guest was then asked to rate the taste of the asparagus on a scale of 1 to 100.

Some R code and output based on the guests' ratings are below:

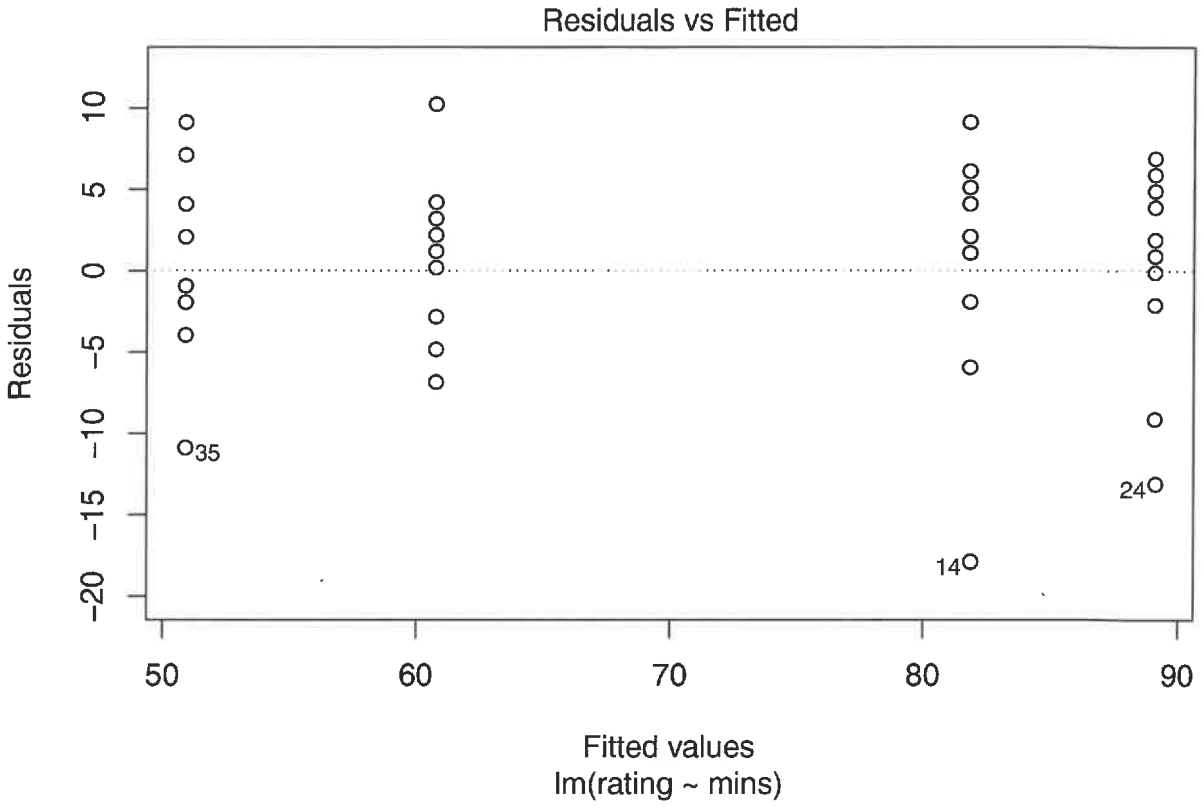
```
stripchart(rating ~ mins, vertical = T, method = "jitter", xlab = "mins")
```



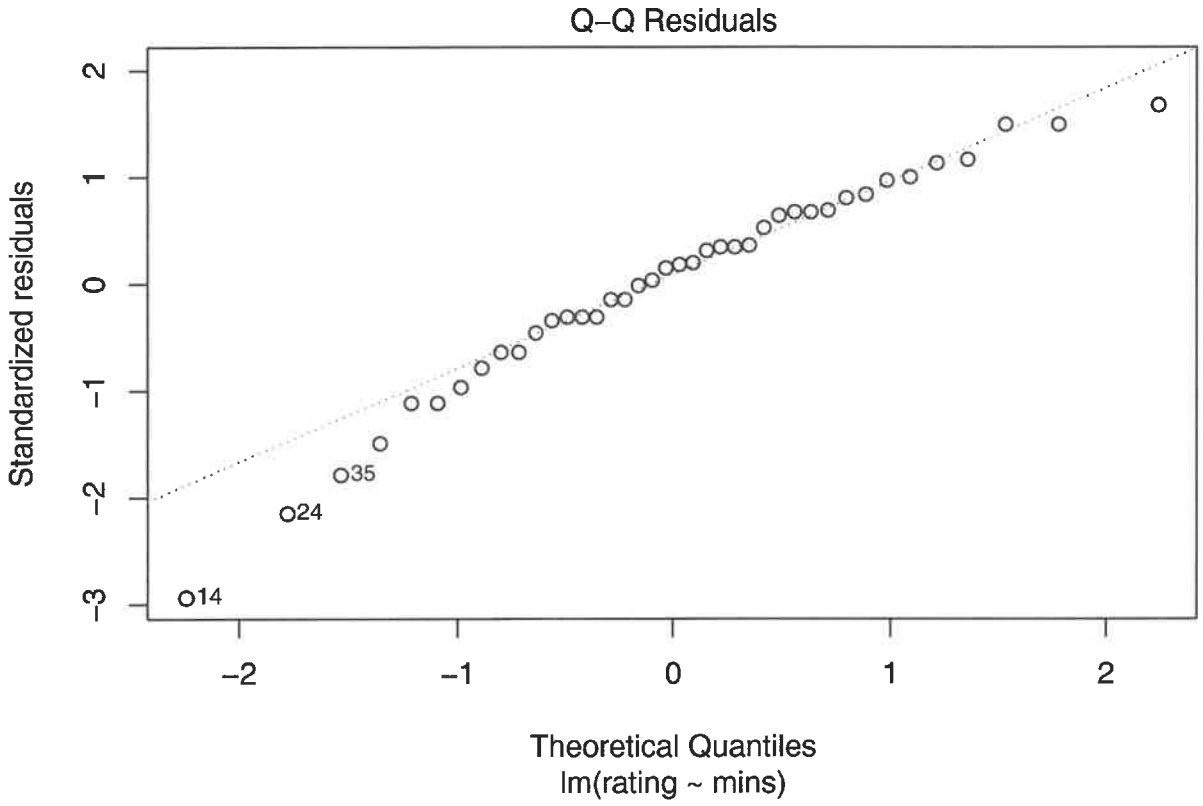

```
lm_out <- lm(rating ~ mins)
lm_out
```

```
Call:
lm(formula = rating ~ mins)

Coefficients:
(Intercept)        mins3        mins5        mins8
       60.8         21.1         28.3         -9.9
```

```
plot(lm_out,which = 1,add.smooth= F)
```

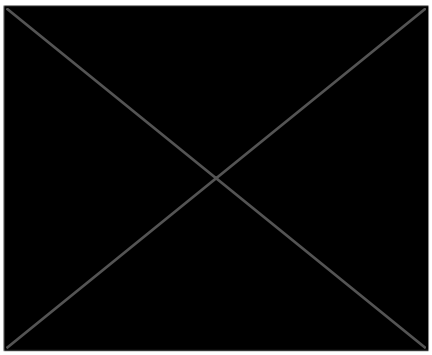


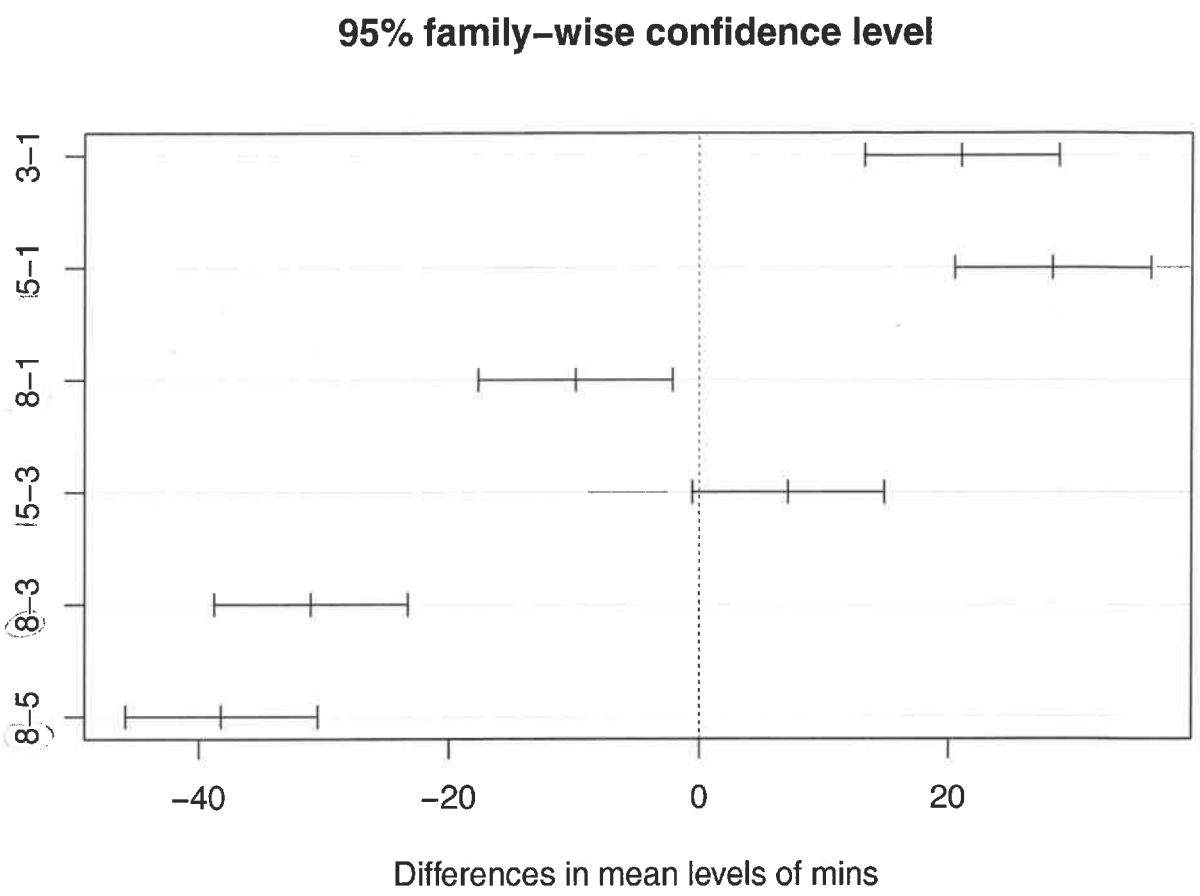

```
plot(lm_out,which = 2)
```




```
anova(lm_out)

Analysis of Variance Table

Response: rating
          Df Sum Sq Mean Sq F value Pr(>F)
mins       3   9627    3209     155 <2e-16 ***
Residuals ??36 747      21
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

plot(TukeyHSD(aov(lm_out)))
```

N-a





(a) Give the mean rating assigned to servings of asparagus which were sautéed for eight minutes.

60.8 − 9.9 = 50.9

$$\begin{array}{r} {}^{9} \\ 60.^{1}8 \\ -\ 9.9 \\ \hline 50.9 \end{array}$$

(b) Fill in the degrees of freedom value where the question marks ?? appear in the ANOVA table.

N − a 40 − 4 = 36

(c) Insofar as it is possible, rank the sauté times from worst to best according to their ratings using the output of Tukey's pairwise comparisons of means.

1. 5 min
2. 3 min
3. 1 min
4. 8 min

In direct comparisons, CI's comparing 8 min to the others is below or at 0. As with 1 min, direct comparisons to 3 min and 5 min show 1 well below those. The comparison with 1 and 8 is slightly below 0, favoring 1 as the better option. There is room for debate on whether 3 or 5 is better since 0 does fall in the CI.

(d) Explain the purpose of the residuals versus fitted values plot and give an interpretation of this plot for these data.

The residuals versus fitted graph shows how much variation is in each treatment group so we can compare how much each group varys. To meet the ANOVA model assumptions, we want each treatment group to have similar residuals / variation. Based on this graph, I would say this assumption is met with some slight larger residuals for the 3min treatment group.

(e) Explain the purpose of the normal quantile-quantile plot and give an interpretation of this plot for these data.

The normal Q-Q plot shows use if the data is normally distributed. If the data is mostly on the dotted line, we can say the data is normal and it meets the normality assumption of the ANOVA model. This data appears to be normally distributed with slight deviations on the tails. But because of the size of the data set, we can say it is normal.

(f) Carefully explain the meaning of "95% family-wise confidence level", which appears in the title of the plot of the Tukey confidence intervals.

This means the confidence intervals take into consideration that multiple comparisons were made (we used Tukey's HSD as opposed to Dunnett's). This means the CI's are larger than if we only did a direct comparison or chose a baseline.

2.

The chef now wishes to investigate the effects of salt and pepper on the taste of sautéed asparagus. He prepares for a catering event eight trays of asparagus, adding a light amount of salt to two trays, a medium amount to two trays, and a heavy amount to two trays, leaving two trays unsalted. Then, for each of these pairs of trays, he adds pepper to one tray and leaves the other without pepper.

- 8 trays
HS . P
HS NP

At the event he assigns guests to receive asparagus from one of the eight trays such that the guests are assigned to the trays at random. He then obtains from each guest a rating on the scale of 1 to 100 of the taste of the asparagus.

Some R code and output based on the data are shown below:

```
lm_out <- lm(rating ~ salt + pepper + salt:pepper)

anova(lm_out)

Analysis of Variance Table

Response: rating
            Df Sum Sq Mean Sq F value  Pr(>F)
salt         3   6904    2301   37.80 3.0e-09 ***
pepper       1?  2485    2485   40.82 1.3e-06 ***
salt:pepper  3    637     212    3.49   0.031 *
Residuals   24   1461      61
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(a-1)

(a-1)(b-1)

a = 4

3(b-1) = 3

b-1 = 1

b = 2

ab(n-1)

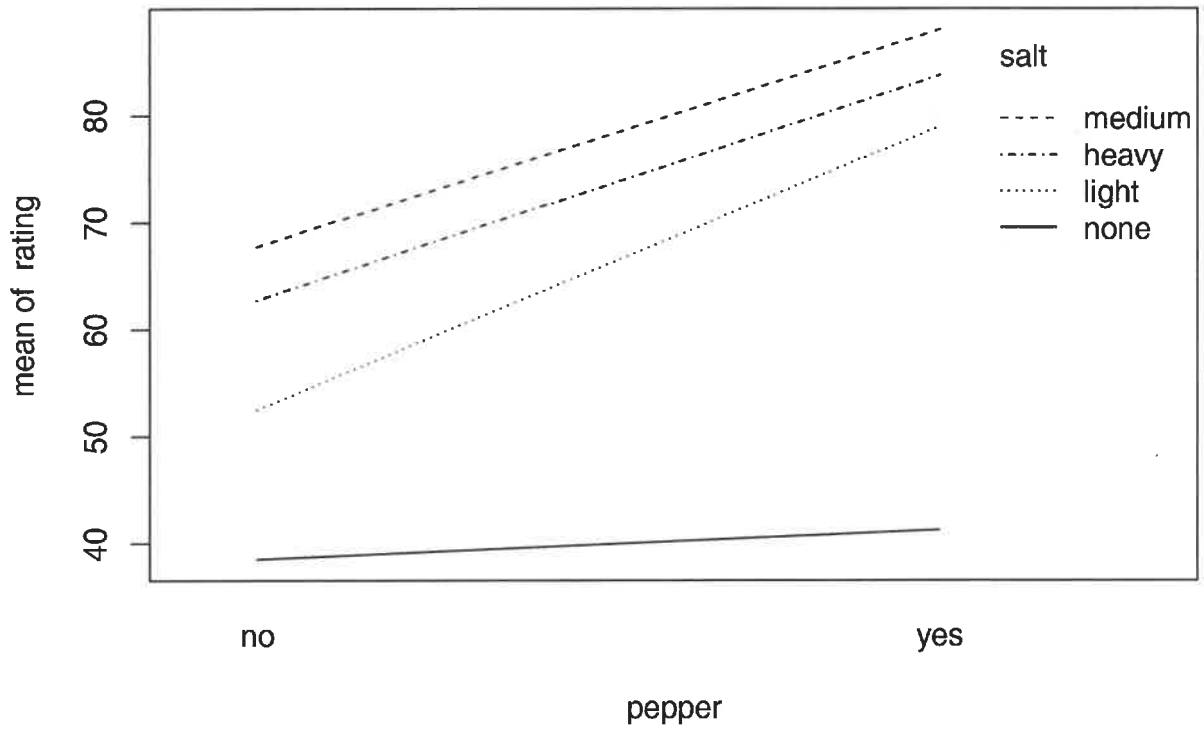
4·2(n-1) = 24

n-1 = 3

n = 4

4·8 = 32

```
interaction.plot(pepper, salt, rating)
```




```
library(DescTools)
salt_pepper <- paste(salt, pepper, sep = "_")
DunnettTest(rating ~ salt_pepper, control = "none_no")
```

```
  Dunnett's test for comparing several treatments with a control :
    95% family-wise confidence level

$none_no
                       diff lwr.ci upr.ci    pval
heavy_no-none_no     24.25   8.73   39.8 0.00118 **
heavy_yes-none_no    45.25  29.73   60.8 4.9e-08 ***
light_no-none_no     14.00  -1.52   29.5 0.08899 .
light_yes-none_no    40.50  24.98   56.0 1.6e-06 ***
medium_no-none_no    29.25  13.73   44.8 0.00011 ***
medium_yes-none_no   49.50  33.98   65.0 3.0e-09 ***
none_yes-none_no      2.75 -12.77   18.3 0.99521
```

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

(a) What is the name of the experimental design the chef is using?

Two-way ANOVA ~~test~~

"factorial" design

(b) How many guests participated in the experiment?

$a=4$
$b=2$

$24+8=32$

residual df + $a \cdot b = 32$

(c) Fill in the degrees of freedom value where the question mark ? appears in the ANOVA table.

1

(d) State whether you believe there is an interaction effect between the salt and pepper factors. Explain your answer.

Yes, I have reason to believe there is an interaction between salt and pepper factors because the p-value for the interaction $.031 < .05$.

(e) Comment carefully, considering all the output shown, on whether adding pepper generally improves the average taste rating of sautéed asparagus.

I would say no, because while their is evidence to support this claim when pepper is added alongside salt, there isn't sufficient evidence to support this in the case of no salt + yes pepper.

(f) How many comparisons would be involved if one were to compare all pairs of treatment means in this experiment?

7

$8(8-1)/2$

$7 \times 8 = 56/2 = 28$

(g) In this experiment, each tray of asparagus was rated by the same number of guests. What term do we apply to an experimental design in which we assign the same number of experimental units to every set of experimental conditions?

Balanced

3.

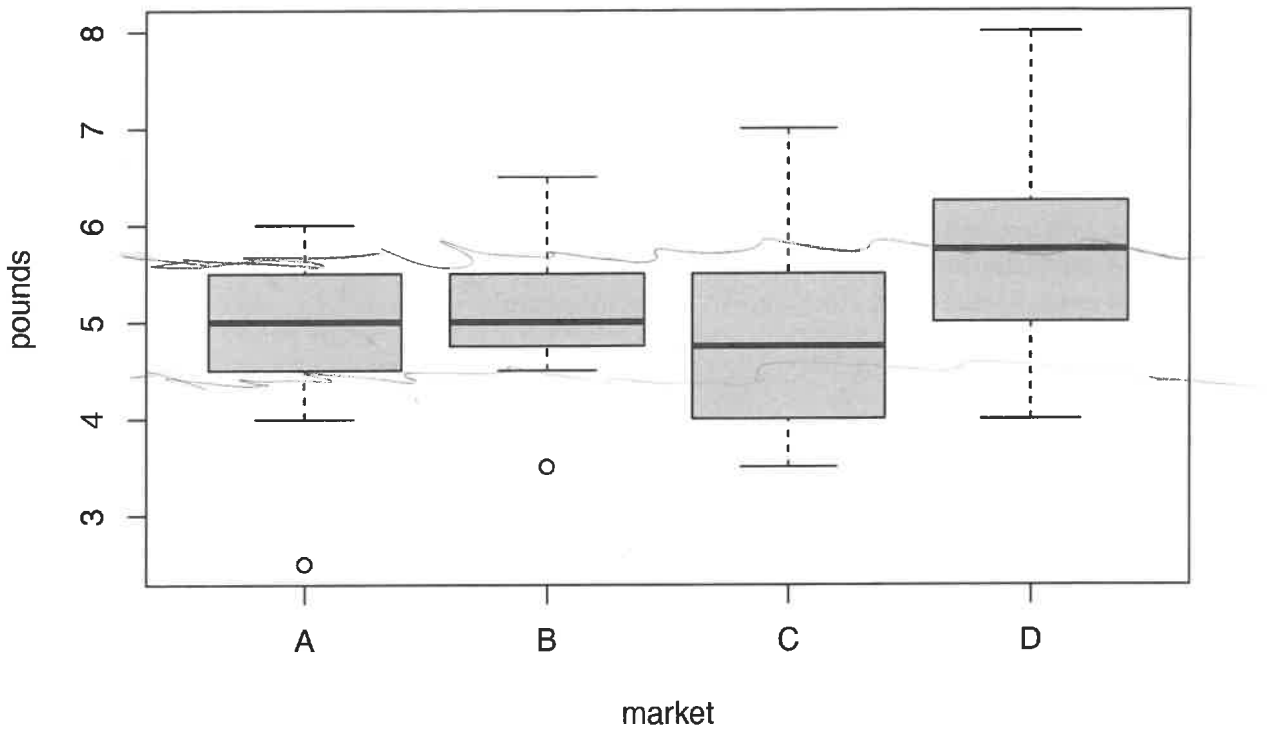
An asparagus grower who plans to sell asparagus at farmers markets around her state this season is interested in learning the number of pounds a typical customer will buy. After attending the first four markets of the season, she constructs a data set containing the number of pounds purchased by each of the first twenty customers at each market. She decides it will be appropriate to fit to her data the one-way random effects model

$$Y_{ij} = \mu + A_i + \varepsilon_{ij},$$

where Y_{ij} is the number of pounds purchased by customer j at market i, μ represents an overall mean, A_i is the random effect of market i, and ε_{ij} is an error term accounting for variability among customers. Here, $i = 1, \ldots, 4$ and $j = 1, \ldots, 20$, and the market random effects $A_1, \ldots, A_4$ are assumed to be independent $\mathcal{N}(0, \sigma_A^2)$ random variables and the error terms to be independent $\mathcal{N}(0, \sigma_\varepsilon^2)$ random variables, where the A_i and ε_{ij} are independent of each other.

Below is some R code and output based on the data:

```
boxplot(pounds ~ market, data = ppc)
```




```
anova(lm(pounds ~ market, data = ppc))
```

```
Analysis of Variance Table

Response: pounds
          Df Sum Sq Mean Sq F value Pr(>F)
market     3   10.2    3.41    3.86  0.013 *
Residuals 76   67.1    0.88
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lmer(pounds ~ 1 + (1|market), data = ppc)
```

```
Linear mixed model fit by REML ['lmerModLmerTest']
Formula: pounds ~ 1 + (1 | market)
   Data: ppc
REML criterion at convergence: 223
Random effects:
 Groups   Name        Std.Dev.
 market   (Intercept) 0.356
```

```
 Residual             0.940
Number of obs: 80, groups:  market, 4
Fixed Effects:
(Intercept)
       5.13
```

(a) Why might it make sense for the asparagus grower to regard the market effect as random rather than fixed?

we are "pulling" markets from a pool of markets assumed to be distributed normally, with mean zero and variance σ_A^2, and the effect of each market we pull comes from, or is governed by, this variance σ_A^2 from a larger pool of markets we pick from.

(b) State the null and alternate hypotheses for which the F value shown in the ANOVA table is the test statistic.

$H_0: \sigma_A^2 = 0$

$H_1: \sigma_A^2 > 0$

(c) From which F distribution is the p value corresponding to the F statistic computed?

$F_{a-1, a(n-1)}$ which in this case is $F_{3, 4(19)} = F_{3, 76}$

$a = 4$ $\quad 3$

$n = 20$ $\quad 19$

$\quad 4$

$\quad 76$

(d) Is there significant variation from market to market in the pounds purchased per customer? Explain your answer.

p-value for this is .013

At $\alpha = .05$, we reject H_0, meaning that the variance σ_A^2 is not zero, meaning there is variation from market to market.

(e) A 95% confidence interval for the overall mean μ based on these data is $[4.474, 5.788]$. Give an interpretation of this interval.

for # pounds purchased

we are 95% certain that the overall mean from all markets and all customers is between 4.474 and 5.788

(f) Give the value of the restricted maximum likelihood estimate of the standard deviation σ_A based on these data.

$\sigma_A = .356$

4.

The analytical chef from questions 1 and 2 wishes to run another experiment investigating the best length of time to sauté asparagus. This time, however, he has access to only eight people who can serve as taste testers. He decides, therefore, that each person will give a rating for all four lengths of sauté time. Thus he prepares four trays of asparagus sautéed for one, three, five, and eight minutes and assigns to each of his eight participants a random order in which to taste servings from the trays.

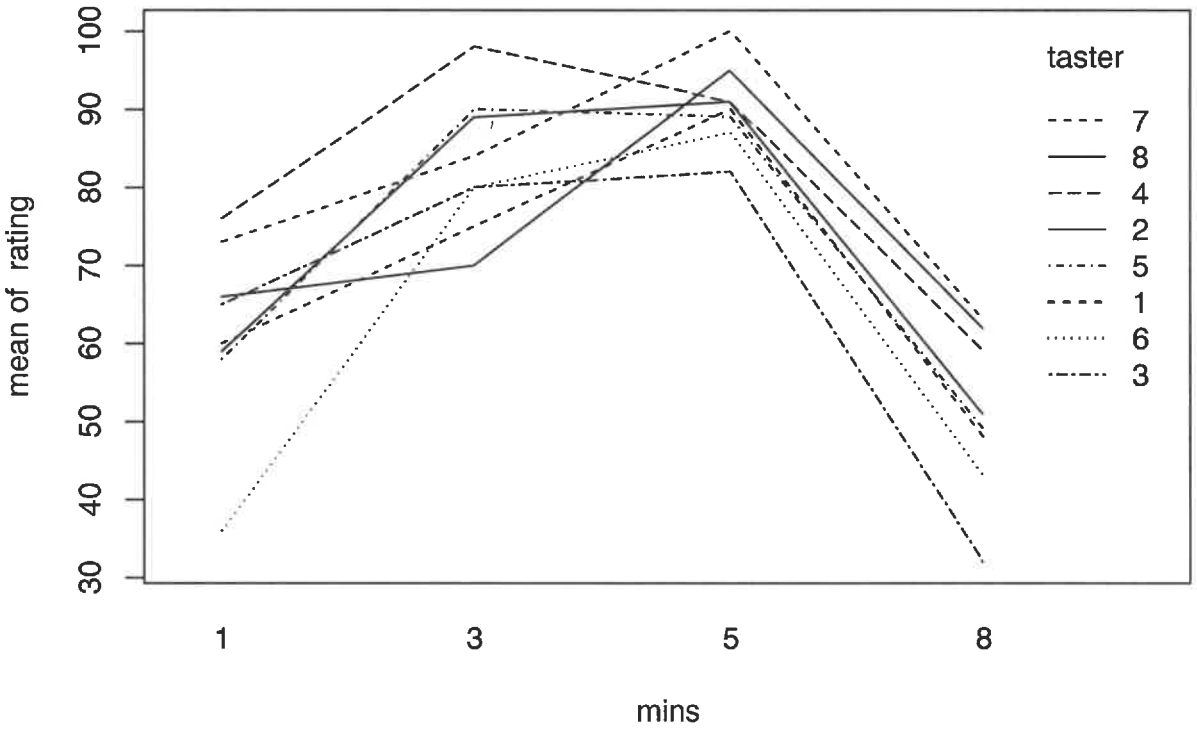
He regards the ratings as arising from the model given by

$$Y_{ij} = \mu + \tau_i + B_j + \varepsilon_{ij}$$

for $i = 1, \ldots, 4$ and $j = 1, \ldots, 8$, where μ is a baseline mean, the τ_i are sauté time effects, the B_j are independent $\mathcal{N}(0, \sigma_B^2)$ random variables representing tasters' tendencies to give higher or lower ratings regardless of sauté time, and the ε_{ij} are independent $\mathcal{N}(0, \sigma_\varepsilon^2)$ random variables. The B_j and the ε_{ij} are also assumed to be independent of each other.

Below is some R code and output based on his data:

```
interaction.plot(mins,taster,rating)
```




```
lm_out <- lm(rating ~ mins + taster)
anova(lm_out)
```

```
Analysis of Variance Table

Response: rating
          Df Sum Sq Mean Sq F value  Pr(>F)
mins       3   8214    2738   44.86 2.6e-09 ***
taster     7   1290     184    3.02   0.023 *
Residuals 21   1282      61
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lmer_out <- lmer(rating ~ mins + (1|taster))
lmer_out
```

```
Linear mixed model fit by REML ['lmerModLmerTest']
Formula: rating ~ mins + (1 | taster)
REML criterion at convergence: 211
Random effects:
```

```
 Groups   Name        Std.Dev.
 taster   (Intercept) 5.55
 Residual             7.81
Number of obs: 32, groups:  taster, 8
Fixed Effects:
(Intercept)        mins3        mins5        mins8
       61.6         21.6         29.0        -10.7
```

```
alpha <- 0.05/6 # adjust alpha for making six comparisons
ls_means(lmer_out, pairwise = T, level = 1 - alpha)
```

```
Least Squares Means table:

             Estimate Std. Error df t value   lower   upper Pr(>|t|)
mins1 - mins3  -21.625      3.906 21   -5.54 -33.000 -10.250  1.7e-05 ***
mins1 - mins5  -29.000      3.906 21   -7.42 -40.375 -17.625  2.7e-07 ***
mins1 - mins8   10.750      3.906 21    2.75  -0.625  22.125    0.012 *
mins3 - mins5   -7.375      3.906 21   -1.89 -18.750   4.000    0.073 .
mins3 - mins8   32.375      3.906 21    8.29  21.000  43.750  4.6e-08 ***
mins5 - mins8   39.750      3.906 21   10.18  28.375  51.125  1.4e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

  Confidence level: 99%
  Degrees of freedom method: Satterthwaite
```

(a) Give the name of the model the chef is assuming for his data.

Randomized complete block design ✓

(b) What would it mean if taster 1 had a negative value of B_1?

The tendency of taster #1 to give a lower rating regardless of sauté time. ✓

(c) Does it appear that there was significant taster to taster variability in the ratings?

The p-value measuring this is .023, so using $\alpha = .05$ we can probably reject H_0.

$H_0: \sigma_B^2 = 0$

Therefore, we conclude that there is significant variability from taster to taster

(d) Report the value of the restricted maximum likelihood estimate of σ_B.

$\sigma_B = 5.55$

(e) Rather than following Tukey's procedure for making all pairwise comparisons, in the construction of the six confidence intervals shown above, the significance level was set to 0.05/6 in order to ensure a familywise coverage probability of at least 0.95. What is the name of this adjustment strategy?

Bonferroni adjustment

(f) Based on the lower and upper bounds of the confidence intervals for comparing all pairs of means, give a careful summary of how the sauté times were rated by the tasters, pointing out where significant differences were found and where no significant differences were found. You may ignore the p values in the table, as these are not adjusted for multiple comparisons.

The difference between 1 and 8, and the difference between 3 and 5, had no significant difference

Significant differences were found for all other comparisons of means in the output.

It looks like 3 and 5 were both better than 1, and 3 and 5 were both better than 8 too.