

# Quiz #7

Mar. 2, 2017

20 pts

{This is the same style of data we did in HW#7.}

A researcher wants to compare two "Diet" methods (A & B) using ten randomly selected people (i.e., five per each diet).

Each person's weight changes were measured four times (i.e., 3, 6, 9 and 12 weeks after starting diet). Data and relevant R printout are shown below.

```
> data1 <- read.csv("U:\\STAT510\\sampleDiet.csv")
> attach(data1)
> data1
```

	Diet	Person	Week	WtChange
1	A	1	3	-0.230
2	A	1	6	-1.700
3	A	1	9	-3.700
4	A	1	12	-4.680
5	A	2	3	-0.690
6	A	2	6	-1.590
7	A	2	9	-3.800
8	A	2	12	-4.530
9	A	3	3	0.690
10	A	3	6	-1.640
11	A	3	9	-3.730
12	A	3	12	-4.650
13	A	4	3	-0.180
14	A	4	6	-1.690
15	A	4	9	-3.630
16	A	4	12	-3.450
17	A	5	3	-0.081
18	A	5	6	-1.700
19	A	5	9	-3.800
20	A	5	12	-4.500
21	B	6	3	-0.700
22	B	6	6	-1.660
23	B	6	9	-3.200
24	B	6	12	-4.500
25	B	7	3	-0.580
26	B	7	6	-1.400
27	B	7	9	-3.300
28	B	7	12	-4.300
29	B	8	3	-0.600
30	B	8	6	-1.500
31	B	8	9	-3.600
32	B	8	12	-4.101
33	B	9	3	-0.499
34	B	9	6	-1.398
35	B	9	9	-4.201
36	B	9	12	-4.801
37	B	10	3	0.102
38	B	10	6	-1.698
39	B	10	9	-3.708
40	B	10	12	-3.901

```
> library( )
```

```
> model1 <- lmer( )
> summary(model1)
```

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Person	(Intercept)	0.0000	0.0000	
Person.1	(Intercept)	11.2333	3.3516	
	Week	0.2250	0.4744	-1.00
Residual		0.2064	0.4544	

Number of obs: 40, groups: Person, 10

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-2.25317	0.10243	-21.997
DietB	-0.03013	0.14486	-0.208

#1. In order to fit a model using "lmer" (linear mixed effect regression in R), which library do you have to load? Fill the blank.

> library( lme4 )

#2. What's wrong with fitting the following model? Explain briefly. Do NOT write a book.

> model19 <- lm(WtChange~Diet+Week)

you can not use a simple "lm" model to analyze this data because it can not analyze week as being nested within diet. ~~or more~~ -2

#3. Fill the blank, i.e., write appropriate R codes to fit a model accommodating the following description:

- WtChange = dependent variable
- Diet = fixed-effect term
- Person = random-effect term; random-effect affects each person the same way
- Week & Person = random-effect term; random-effect affects each person the same way for each week

> model11 <- lmer(Wtchange~Diet+(1|Person)+(1|Week))

#4. Look at the "summary" printout. Consider the t-value printed with "DietB". State the conclusion in plain terms. Reject  $H_0$ , or do not reject  $H_0$  == no points. Write in totally plain terms.

Diet does not significantly effect Wtchange ( $2 < t < 2$ ) ( $t=0.208$ )

#5. Fill the blanks, i.e., find numerical values for the following terms.

• Variance due to "Person"? 11.2333

• Variance due to "Week"? 0.2250

• Variance due to "pure error" (i.e., residual)?

0.2064

#6. (Continued from #5.)

If you had to redo this experiment, you'd need to collect more of \_\_\_\_\_? Just circle your choice.

people

or

weeks

# Quiz #7

Mar. 2, 2017

(20 pts)

{This is the same style of data we did in HW#7.}

A researcher wants to compare two "Diet" methods (A & B) using ten randomly selected people (i.e., five per each diet). Each person's weight changes were measured four times (i.e., 3, 6, 9 and 12 weeks after starting diet). Data and relevant R printout are shown below.

```
> data1 <- read.csv("U:\\STAT510\\sampleDiet.csv")
> attach(data1)
> data1
   Diet Person Week WtChange
 1     A      1    3   -0.230
 2     A      1    6   -1.700
 3     A      1    9   -3.700
 4     A      1   12  -4.680
 5     A      2    3   -0.690
 6     A      2    6   -1.590
 7     A      2    9   -3.800
 8     A      2   12  -4.530
 9     A      3    3    0.690
10    A      3    6   -1.640
11    A      3    9   -3.730
12    A      3   12  -4.650
13    A      4    3   -0.180
14    A      4    6   -1.690
15    A      4    9   -3.630
16    A      4   12  -3.450
17    A      5    3   -0.081
18    A      5    6   -1.700
19    A      5    9   -3.800
20    A      5   12  -4.500
21    B      6    3   -0.700
22    B      6    6   -1.660
23    B      6    9   -3.200
24    B      6   12  -4.500
25
```

## Quiz #7 (Keys)

```
#1 library(lme4)
#2 It treats all 40 cases are independent data, i.e., like 40 people. But this data is from 10 people and
each person produced four cases. That is, it's WRONG to fit a model as if all 40 are independent data.
#3 lmer(WtChange~Diet+(1|Person)+(Week|Person))
#4 There is NOT a significant mean difference between Diet A & B.
#5 11.2333, 0.225, 0.2064
#6 people (because there is so much more variation with people, you'd need more data)
```

```
39   B    10    9   -3.708
40   B    10   12  -3.901
```

> library( )

> model1 <- lmer( )
> summary(model1)

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Person	(Intercept)	0.0000	0.0000	
Person.1	(Intercept)	11.2333	3.3516	
	Week	0.2250	0.4744	-1.00
Residual		0.2064	0.4544	

Number of obs: 40, groups: Person, 10

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-2.25317	0.10243	-21.997
DietB	-0.03013	0.14486	-0.208

#1. In order to fit a model using "lmer" (linear mixed effect regression in R), which library do you have to load? Fill the blank.

> library( lme4 )

#2. What's wrong with fitting the following model? Explain briefly. Do NOT write a book.

> model19 <- lm(WtChange~Diet+Week)

*you can not use a simple "lm" model to analyze this data because it can not analyze week as being nested within diet*

*Say simply about it*

#3. Fill the blank, i.e., write appropriate R codes to fit a model accommodating the following description:

- WtChange = dependent variable
- Diet = fixed-effect term
- Person = random-effect term; random-effect affects each person the same way
- Week & Person = random-effect term; random-effect affects each person the same way for each week

#5. Fill the blanks, i.e., find numerical values for the following terms.

• Variance due to "Person"? 11.2333

• Variance due to "Week"? 0.2250

• Variance due to "pure error" (i.e., residual)? 0.2064

#6. (Continued from #5.)

If you had to redo this experiment, you'd need to collect more of \_\_\_\_\_? Just circle your choice.

people or weeks

# Quiz #8

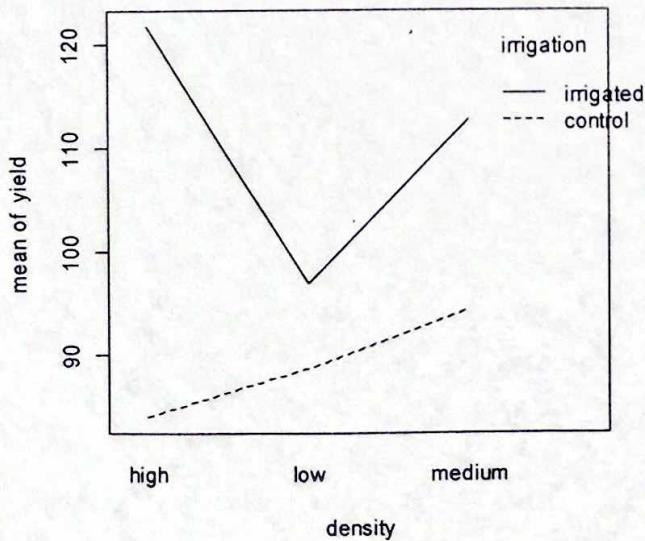
Mar. 10, 2017

20  
of

{This is the data from a split-plot design used in class.} This data is from an experiment on crop yield with three treatments: irrigation (irrigated or control), sowing density (low, medium or high), and fertilizer (N, P or NP). The largest plots were the four whole fields (block), each of which was split in half, and irrigation was applied at random to one half of the field. Each irrigation plot was split into three, and one of three different seed-sowing densities was used at random. Finally, each density plot was divided into three, and one of three fertilizer was used at random. Data and relevant R printout are shown below.

```
> attach(data1)
> head(data1)
  yield block irrigation density fertilizer
1   90     A    control    low        N
2   95     A    control    low        P
3  107     A    control    low      NP
4   92     A    control  medium        N
5   89     A    control  medium        P
6   92     A    control  medium      NP
```

> interaction.plot(\_\_\_\_\_, \_\_\_\_\_)



```
> library(lme4)
> modell1 <- lmer(yield ~ _____)
> summary(modell1)
```

Random effects:  
<<< Omitted for this Quiz >>>

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	85.2639	4.7556	17.929
irrigated	21.4444	4.4538	4.815
denlow	-10.0417	5.4548	-1.841
denmedium	0.8333	5.4548	0.153
ferNP	12.7500	2.8873	4.416
ferP	7.6667	2.8873	2.655

#1. This is called a **split-plot** design (or experiment). The reason why it's called so is some treatment variables are applied differently than other treatment variables.

For example, the units where "irrigation" was applied is very different from those where "fertilizer" was applied.

(3) (a) Which of the following two treatment variables are applied to "bigger" units?

(3) *Circle* your choice: irrigation or fertilizer ✓

(3) (b) Which of the following two treatment variables are applied to (many) more times?

(3) *Circle* your choice: density or fertilizer ✓

#2. Consider the "interaction" plot shown on the left.

(a) Write appropriate R codes to produce this plot.

(3) > interaction.plot(\_\_\_\_\_, ~~yield~~~~irrigation:density~~ - 3)

(b) According to the plot, is there a *significant* interaction between irrigation and density?

(3) *Circle* your choice: YES or NO ✓

#3. The following simplified names were used in the printout.

- irr = irrigation
- den = density
- fer = fertilizer

(4) (a) Suppose you want to fit an "lmer" model with the following three fixed-effect terms and a **random-effect** term to show "den" nested in "irr", which is again nested in "block".

Complete the following R codes.

> modell1 <-  
lmer(yield~irr+den+fer+~~(1|irr:den:block)~~  
~~(1|block:irr:den)~~)

(4) (b) {See the "summary" printout shown on the left} Among three levels of the "fertilizer" variable (i.e., N, P, and NP), write them in order of "yield". That is, write which "fertilizer" type produces the most, which is next, etc.

ferNP ← most ✓  
ferP  
fer N ← least

use data from a split-plot experiment from an experiment on **irrigation** (irrigated low, medium or high), a set plots were the four was split in half, and **irr** one half of the field. Each one of three c was used at random. Field into three, and one o random. Data and relevant

```
(data1)
lata1)
block irrigation dens]
A control ]
A control ]
A control medj
A control medj
A control medj
```

action.plot (\_\_\_\_\_)

## Quiz #8 (Keys)

#1 (a) irrigation, (b) fertilizer

#2 (a) `interaction.plot(density, irrigation, yield)`

#3 (a) `lmer(yield~irr+den+fer+(1|block:irr:den))`

(b) YES

(b)  $N < P < NP$

class.) three  
ng  
(or NP).  
, each  
ed at  
was  
ving  
y plot  
was  
hown

#1. This is a reason why applied. For example very different  
(a) Which one to "biggest  
Circle"  
(3)  
(3)

#2. Consider  
(a) Write appropriate  
>  
(3)  
(3)

(b) According to  
Circle  
(3)

on irrigated control

# Quiz #9

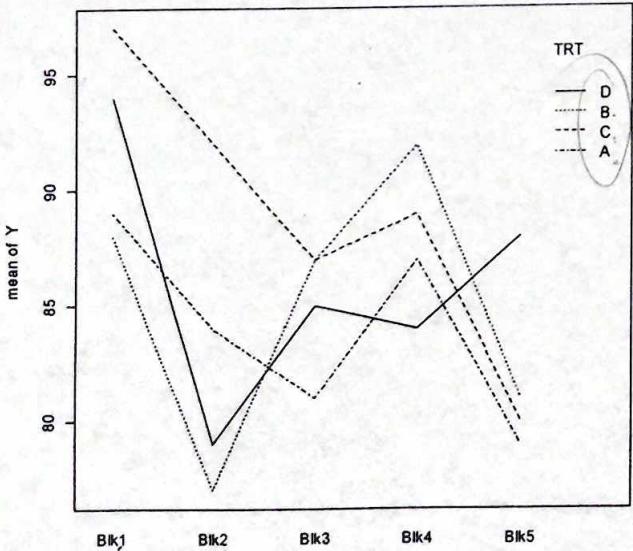
Mar. 24, 2017

20 pts

Suppose a researcher wanted to compare four different treatments (A, B, C and D). Five randomly selected blocks were used for the experiment and each block was divided into four small sections to accommodate each treatment.

- TRT = 4 different treatments (A, B, C, and D) (fixed effect)
- BLK = 5 blocks (Blk1, Blk2, ..., Blk5) (random effect)
- Y = agricultural yield (= dependent variable).

```
> attach(data1)
> data1
> data1
  TRT   BLK     Y
1   A   Blk1  89
2   B   Blk1  88
3   C   Blk1  97
4   D   Blk1  94
5   A   Blk2  84
6   B   Blk2  77
7   C   Blk2  92
8   D   Blk2  79
9   A   Blk3  81
10  B   Blk3  87
11  C   Blk3  87
12  D   Blk3  85
13  A   Blk4  87
14  B   Blk4  92
15  C   Blk4  89
16  D   Blk4  84
17  A   Blk5  79
18  B   Blk5  81
19  C   Blk5  80
20  D   Blk5  88
> interaction.plot(BLK, TRT, Y)
```



#1. Just by looking at the interaction plot, do you significant interaction between BLK & TRT?

Circle your answer:

YES, there seems a significant interaction ✓

NO, there doesn't seem a significant interaction

#2. (Refer to #1). In order to numerically test for an interaction, a researcher fit the following **model1** and obtained **anova** table. Fill the blanks for the df (degrees of freedom).

(4) > model1 <- lm(Y ~ BLK+TRT+BLK:TRT)

> anova(model1)

	SS	df	MS	F-value	p-value
BLK	264	4	omitted		
TRT	70	3	omitted		
BLK:TRT	226	12	omitted		
Residuals	0	0	omitted		
TOTAL	560	19			

#3. (Refer to #2). As a result, would it be possible or not possible to (numerically) test for the interaction between BLK and TRT?

Circle your answer:

YES, possible ✓

NO, not possible

#4. (a) Write an appropriate "lmer" model below.

(4) > library(lme4)  
> model2 <- lmer(Y ~ TRT + (1|BLK))  
> summary(model2)

Random effects:

Groups	Name	Variance	Std.Dev.
BLK	(Intercept)	11.79	3.434
Residual		18.83	4.340

Number of obs: 20, groups: BLK, 5

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	84.000	2.475	33.94
TRTB	1.000	2.745	0.36
TRTC	5.000	2.745	1.82
TRTD	2.000	2.745	0.73

(4) (b) What's the estimated mean of TRT=D? 80

(4) (c) Is TRT significant?

Circle your answer: YES or NO or don't know

# Quiz #9 (Keys)

#1 YES

#2 4, 3, 12, 0, 19

#3 NO

#4 (a) `lmer(Y ~ TRT + (1 | BLK))`

(b) 86

(c) N

1) (2) significant interaction between BLK & TRT?

Circle your answer:

YES, there seems a significant interaction

NO, there doesn't seem a significant interaction

#2. (Refer to #1). In order to numerically test for a researcher fit the following **model1** and obtain table. Fill the blanks for the df (degrees of freedom)

```
> model1 <- lm(Y ~ BLK+TRT+BLK:TRT)
> anova(model1)
```

	SS	df	MS	F-v.e
BLK	264	4	omitted	
TRT	70	3	omitted	
BLK:TRT	226	12	omitted	
Residuals	0	8	omitted	
TOTAL	560	19		

#3. (Refer to #2). As a result, would it be possible to (numerically) test for the interaction BLK and TRT?

Circle your answer:

YES, possible

NO, not possible

#4. (a) Write an appropriate "lmer" model below.  
> library(lme4) *(or lmer)*

# Quiz #10

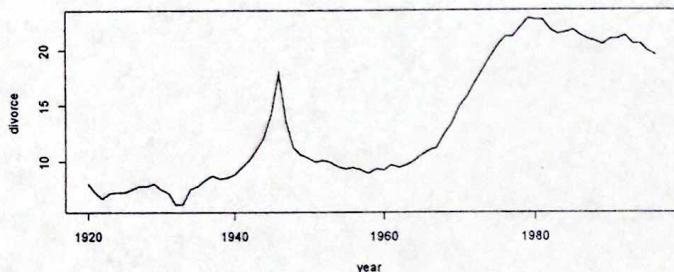
Apr. 7, 2017

20 pts

The "divusa" dataset (in "faraway" library) has 77 rows.

- (x) year = from 1920 to 1996 (i.e., during 77 years)
- (y) divorce = divorce per 1,000 women

```
> library(faraway)
> data(divusa)
> dim(divusa)
[1] 77 7
> attach(divusa)
> head(divusa[,1:2])
year divorce
1 1920 7.9
2 1921 7.3
3 1922 6.7
4 1923 7.3
5 1924 7.2
6 1925 7.2
> plot(divusa[,1:2], col=2)
```



- #1.** For a data like this is called time series data and all the y-points need to be connected by a line as shown on the left. Write R codes to connect all the data points. Fill the blank.

**(4)**

```
> plot(divorce, type="l")
```

*what about "year"?* -3

**#2.** A researcher wants to fit an autoregressive model (i.e., y vs. lag1, lag12, and lag13 terms). To see how the "embed" command works, she first tried the following. Fill the blanks below, i.e., show what's "lag1" and "lag2" below.

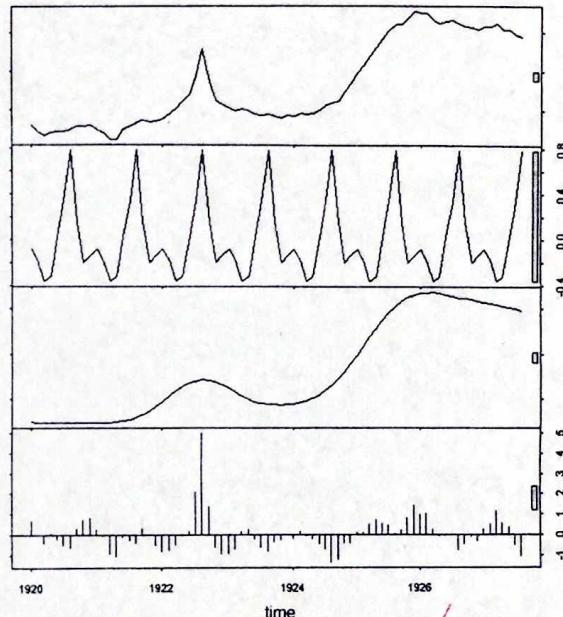
```
> lagdf <- embed(divorce, 3)
> lagdf <- data.frame(lagdf)
> colnames(lagdf) <- c("divorce", paste0("lag", 1:2))
> head(lagdf)
divorce      lag1      lag2
7.9        XXXXX      XXXXX
7.3        XXXXX      XXXXX
6.7        7.3       7.9
7.3        6.7       7.3
```

**(4)** <<< Omitted to save the world >>>

- #3.** There is a built in R plot shown below. To get this plot, (1) you first need to convert y-data into an R time-series format, and then (2) decompose the data by using "stl" (seasonal decomposition of time-series by loess) in R. This is shown below.

```
> modell1 <- ts(divorce, start=1920, freq=10)
> plot(stl(modell1, "periodic"))
```

- 1
- 2
- 3
- 4



*(Choose one for each of following question: 1, 2, 3, 4) (P)*

- Which plot is the trend? **3**
- Which one is plot of the raw data? **1**
- Which one is the seasonality plot? **2**
- Which plot is that of the remainder? **4**

- #4.** We have created a data frame "lagdf" with up to lag13 and found only "lag1" was significant. So the final model used "lag1" and log(y) was used (instead of y):

```
> model9 <- lm(log(divorce) ~ lag1, data=lagdf)
> summary(model9)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.54900	0.02852	54.32	<2e-16 ***
lag1	0.07241	0.00198	36.57	<2e-16 ***
---				

Predict y (i.e., "divorce") of next year when the current "divorce=19". Showing set up is good enough.

**(4)** divorce ~~19~~  $\log(19) \sim 0.07241X + 1.549$

-4

# Quiz #10 (Keys)

#1. divorce~year, type="l"

(OR .year, divorce, type="l")

#2. divorce    lag1                lag2  
    7.9           xxxxx               xxxxx  
    7.3           xxxxx               xxxxx  
    6.7           7.3                7.9  
    7.3           6.7                7.3

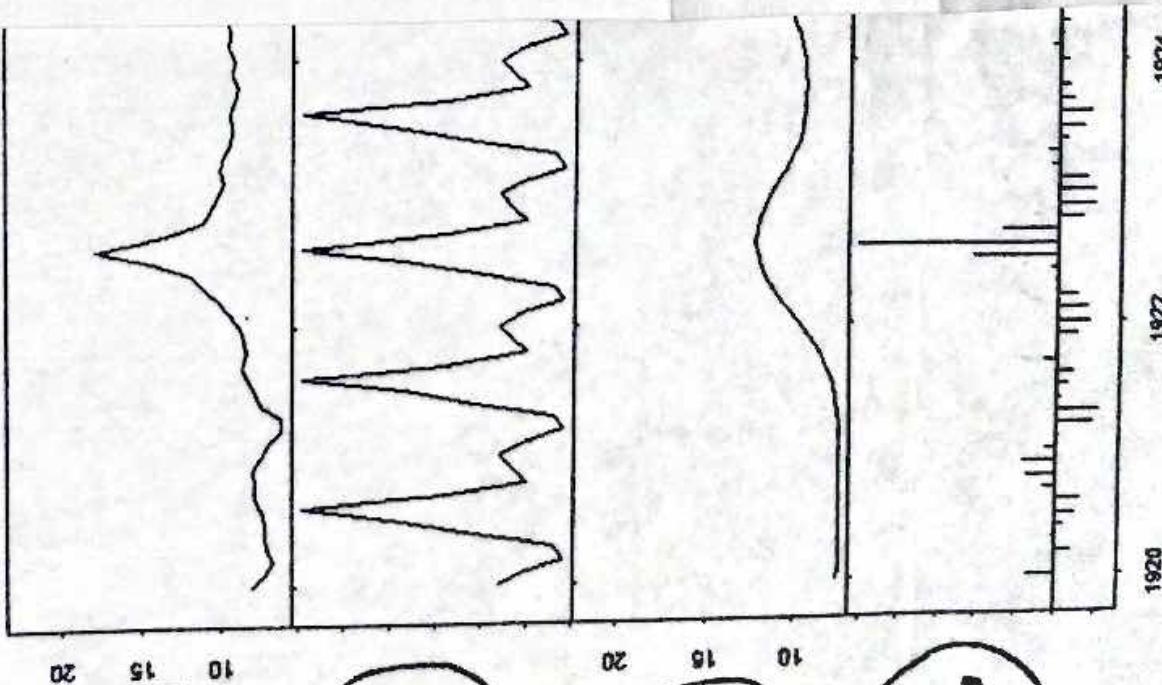
#3. 3, 1, 2, 4

#4.  $\log(y) = 1.549 + (0.07241 \times 19)$ , so  $y = e^{2.92479} = 18.6303$

the two variables used are:

There is a built in R plot showing:  
1) you first need to convert to  
format, and then (2) decompose  
seasonal decomposition of time  
series shown below.

```
modell1 <- ts(divorce, start=1921, end=1924, frequency=4)
plot(stl(modell1, "periodic"))
```



Choose one for each of following question

# Quiz #11

Apr. 14, 2017

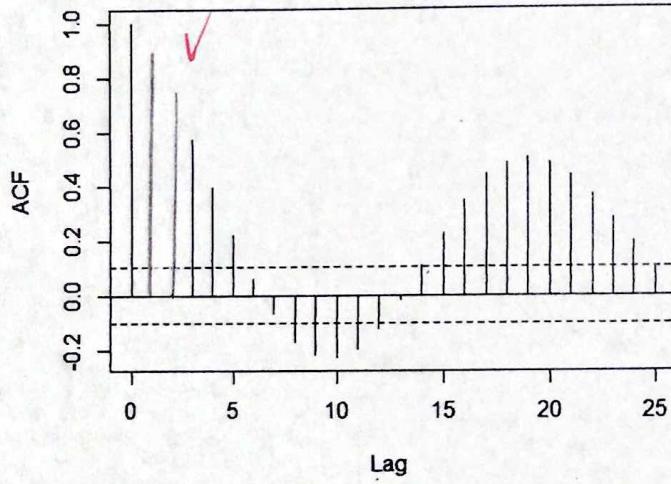
20 pts

17/21

This is from the sample dataset "blowfly" from class not technicians kept watching for almost 7 years (361 week is "flies" (i.e., number of flies).

```
> data1 <-  
read.table("U:\\STAT510\\Data\\blowfly.txt", header=T)  
> dim(data1)  
[1] 361 1  
> head(data1)  
  flies  
1 948  
2 942  
3 911  
4 858  
5 801  
6 676  
> attach(data1)  
> _____ (flies)
```

Series flies



- What's shown in the "acf" plot at `lag=1` is the plain *correlation* between the original data and `lag=1` data (i.e., data after `lag=1` is taken). Consider the following:

```
> cor(flies[-c(360:361)], flies[-c(1:2)])  
[1] 0.7502315  
> cor(flies[-c(359:361)], flies[-c(1:3)])  
[1] 0.581824  
> cor(flies[-361], flies[-1])  
[1] 0.8863121
```

- #1. I have erased two lines at `lag=1` and `lag=2` in the acf plot. Complete the plot, i.e., sketch appropriate lines at `lag=1` & `lag=2` in the acf plot shown above.

(4) ✓

- #2. What would be the answer to the following calculation?

```
> cor(flies, flies)
```

(2) 1 ✓

- #3. Write an appropriate R command to create an acf plot as shown above, i.e., fill the blank.

(2) > \_\_\_\_\_ acf (flies) ✓

- #4. Seen from the acf plot:

- Is the series cyclic? (Choose one: Yes, No, Don't know)
- If YES, what's the period? 19 (If NO, skip this question.) ✓

- #5. Suppose an ARIMA model with parameters  $(p=3, d=0, q=1)$  was fit as shown below.

```
> model301 <- arima(flies, order=c(3,0,1))  
> model301
```

Coefficients:

	ar1	ar2	ar3	ma1	intercept
ar1	1.51	-0.47	-0.14	-0.56	3489.73
s.e.	0.11	0.45	0.06	0.10	264.68

$\sigma^2 = 1147998$ , log likelihood = -3031.77, aic = 6075

The forecasting equation with AR (order=p), MA (order=q) is written as

$$y_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Using only significant terms of `model301`, write the estimated forecasting model.

(4)  $y_t = 3489.73 + 0.11(\text{ar1}) - 0.56(\text{ma1})$  ✓

- #6. Consider the following R printout from `model301`.

```
> Box.test(model301$residuals)  
Box-Pierce test
```

```
data: model301$residuals  
X-squared = 0.00044619, df = 1, p-value = 0.9831
```

- Is this a good news? (Choose one: Yes, No, Don't know)
- Write the conclusion in totally plain terms. Saying "Reject/Accept  $H_0$ " is NOT enough.)

It means there is not significant autocorrelation of the residuals. ✓

It means more than just autocorrelation. See Ans. key

# Quiz #12

Apr. 21, 2017

(20 pts) 16/24

#1. The pdf of a binomial random variable  $X$  (= total number of "successes") is  $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ , where  $x = 0, 1, 2, \dots, n$

Suppose you flip a biased coin with  $p(H) = 0.2$  five times, what's the probability of getting two heads? Show the complete set up, at least.

$$\binom{5}{2} = \frac{5 \times 4}{2 \times 1} = 10 \quad \text{Ans: } 1.2788 \quad (4)$$

$$\binom{5}{2} \cdot .2^2 (1-.2)^{5-2}$$

$$\times (0.04)(0.512)$$

$$(-4)$$

#2. In a survey of 1,000 people, 500 are "whites", 220 are "blacks" and 280 are "Hispanics." Also, 10% of "whites," 20% of "blacks" and 35% of "Hispanics" are smokers. Fill the blanks. Hint:  $280 \times 0.35 = 98$

W	B	H
500	220	280
10%	20%	35%

(1) (a)  $p(\text{white}) = \frac{500}{1000}$ ,  $p(\text{black}) = \frac{220}{1000}$ ,  $p(\text{hispanic}) = \frac{280}{1000}$

(1) (b)  $p(\text{smoker} | \text{white}) = \frac{50}{500} = 0.1$ ,  $p(\text{smoker} | \text{black}) = \frac{44}{220} = 0.2$ ,  $p(\text{smoker} | \text{hispanic}) = \frac{98}{280} = 0.35$

Now, find the following:

(3) (c) The probability of selecting a smoker, i.e.,  $p(\text{smoker}) = \frac{192}{1000}$

(3) (d) Suppose a randomly selected person is a smoker. What's the probability that the person is a "white"? i.e.,

$$\boxed{.260} \quad \begin{matrix} .10 & .5 & \frac{(.10)(.5)}{(.10)(.5) + (.20)(.22) + (.35)(.28)} \\ p(\text{white} | \text{smoker}) = P(\text{SMOKER} | \text{white}) \cdot P(\text{white}) \\ P(\text{SMOKER} | \text{white}) \cdot P(\text{white}) + P(\text{smoker} | \text{black}) \cdot P(\text{black}) + P(\text{smoker} | \text{Hispanic}) \cdot P(\text{Hispanic}) \end{matrix}$$

#3. Which of the following is a correct writing for the Bayes' formula? Choose one.

(a)  $P(B_i | A) = \frac{P(B_i | A) \cdot P(A)}{\sum_k P(A | B_k) \cdot P(B_k)}$

(c)  $P(B_i | A) = \frac{P(B_i | A) \cdot P(A)}{\sum_k P(B_k | A) \cdot P(A)}$

(b)  $P(B_i | A) = \frac{P(B_i | A) \cdot P(A)}{\sum_k P(A | B_k) \cdot P(B_k)}$

(d)  $P(B_i | A) = \frac{P(A | B_i) \cdot P(B_i)}{\sum_k P(A | B_k) \cdot P(B_k)}$

#4. The following is the famous principle of the Bayesian probability. Fill the blank.

Posterior (probability)  $\propto$  prior  $\times$  likelihood (4)

## Quiz #12 (Keys)

#1.  $f(2) = \binom{5}{2} 0.2^2 0.8^3 = 0.2048$

#2. (a)  $p(\text{white}) = 0.5, p(\text{black}) = 0.22, p(\text{hispanic}) = 0.28$

(b)  $p(\text{smoker} | \text{white}) = 0.1, p(\text{smoker} | \text{black}) = 0.2, p(\text{smoker} | \text{hispanic}) = 0.35$

(c)  $p(\text{smoker}) = P(S | W)P(W) + P(S | B)P(B) + P(S | H)P(H) = (0.5 \cdot 0.1) + (0.2 \cdot 0.22) + (0.35 \cdot 0.28) = 0.192$

(d)  $p(\text{white} | \text{smoker}) = \frac{P(S | W)P(W)}{P(S)} = \frac{0.5 \times 0.1}{0.192} = 0.2604$

#3. (d)

#4. Posterior  $\propto$  prior  $\times$  likelihood

# Quiz #13

Apr. 28, 2017

20 pts  
13)  $\Gamma(2)$

#1. The gamma function is defined as  $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$  and  $\Gamma(n) = (n-1)!$  when  $n$  is a positive integer. Find  $\Gamma(2)$

$$\Gamma(2) = (2-1)!$$

ans: 1 (2)

#2. The beta function is defined as  $B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$  and a key property is its relationship with the gamma function,

$$B(a, b) = \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)}$$

$$\frac{(2-1)! \cdot (1-1)!}{(3-1)!}$$

$$\frac{1 \cdot 1}{2}$$

ans: 1/2 ✓ (2)

#3. The pdf of a beta distribution with two parameters  $(\alpha, \beta)$  is  $f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}$ ,  $0 < x < 1$ .

(a) Write the pdf of a beta distribution with  $(\alpha=1, \beta=1)$ .

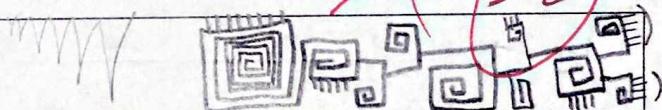
$$ans: f(x) = \frac{1}{B(1,1)} x^{1-1} (1-x)^{1-1} = 1 \quad (2)$$

(b) There is an even more popular name than beta distribution with  $(\alpha=1, \beta=1)$ . What would it be?

ans: posterior distribution (2)

#4. Consider the following R codes.

```
> x <- rnorm(1000, 100, 13)
> quantile(x, c(0.025, 0.975))
> pnorm(115, 100, 13)
> qnorm(0.9, 100, 13)
```



(a) Explain briefly what the first command does, i.e.  $rnorm(1000, 100, 13)$

ans: will give you random #'s from a Normal distribution (2)

(b) Explain briefly what the second command does, i.e.  $quantile(x, c(0.025, 0.975))$

ans: will give you the # values associated w/ a % range (2)

(c) (i) Explain briefly what the third command does, i.e.  $pnorm(115, 100, 13)$

ans: gives a percentile from a Normal(2) distribution (2)

(ii) Which of the following can be a possible answer to  $pnorm(115, 100, 13)$ ? Just circle one.

0.22, 0.55, 0.88, 0.99, 16, ✓ 86, 116, 136 (2)

(d) (i) Explain briefly what the fourth command does, i.e.  $qnorm(0.9, 100, 13)$

ans: gives a quantile from a normal(2) (2)

(ii) Which of the following can be a possible answer to  $qnorm(0.9, 100, 13)$ ? Just circle one.

0.22, 0.55, 0.88, 0.99, 16, 86, 116, ✓ 136 (2) dist.

## Quiz #13 (Keys)

#1.  $\Gamma(2) = 1! = 1$

#2.  $B(2, 1) = \frac{\Gamma(2) \cdot \Gamma(1)}{\Gamma(3)} = \frac{1}{2!} = \frac{1}{2}$

#3. (a)  $f(x) = \frac{1}{B(1, 1)} x^0 (1-x)^0 = \frac{\Gamma(2)}{\Gamma(1) \cdot \Gamma(1)} 1 = 1, 0 < x < 1$

(b) uniform distribution on  $(0, 1)$

#4.

- (a) generates 1,000 "normal" random numbers with mean 100 & SD 13.
- (b) finds the 95% c.i., i.e., finds 2.5<sup>th</sup> percentile and 97.5<sup>th</sup> percentile.
- (c) (i) finds the percentile of 115, i.e., what percentage is behind 115. (ii) 0.88
- (d) (i) finds the quantile of 90%, i.e., what number will be the 90<sup>th</sup> percentile. (ii) 116