UNIVERSITY OF SOUTH CAROLINA

STAT 205, SPRING 2017

Midterm Exam

INSTRUCTOR: HAIGANG LIU

MARCH 2, 2017

•

GROUND RULES

No cheatsheet is allowed.

Calculator is allowed in this exam, but smartphone cannot be used as calculator.

No communication of any type with others is allowed. You must do it yourself.

Use number 2 pencil to fill in your scantron. Don't use a pen.

You need to fill in the I.D. number in the scantron form (start from the left).

Your I.D. Number can be found in the Page 3.

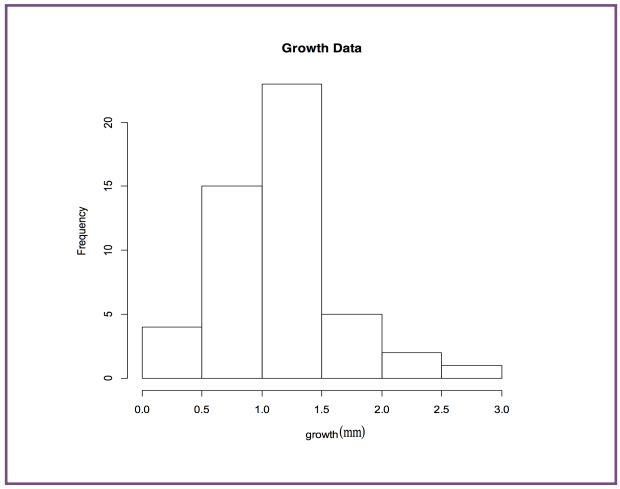
First Name	Last Name	I.D.	First Name	Last Name	I.D.	First Name	Last Name	I.D.
SOPHIA	Abdun	1	Claire	Hann	32	KATHERINE	Palmer	62
MATTHEW	Absher	2	CHAD	Hardin	33	Anna	Patrick	63
KATHRYN	Allgor	3	CALLIE	Hartsell	34	YURLEIDY	Piedrahita	64
JORDAN	Ard	4	KATIE	Heiken	35	CAMILLE	Pierner	65
KRISTEN	Barber	5	JOHN	Holler	36	WILLIAM	Portas	66
ASIA	Barnes	6	EMMA	Holtzclaw	37	LAUREN	Pruden	67
ALICIA	Barry	7	BRANDON	Houston	38	ERICA	Rader	68
BENJAMIN	Blanton	8	SAAD	Iftikhar	39	BRIANNA	Ray	69
BREANNA	Boone	9	ASHTON	Irvin	40	LAUREN	Reince	70
Margaret	Booth	10	DONTE	Jackson	41	MEREDITH	Riggs	71
EMILY	Boyd	11	MAEGAN	Jewson	42	JACK	Rogan	72
Morgan	Brett	12	Alexandra	Johnson	43	MEGHAN	Root	73
ELIZABETH	Broome	13	NOAH	Karch	44	MERRY	Rudinger	74
MICHELLE	Brown	14	SAVANNA	Kelly	45	VAIDA	Shelley	75
MYESHA	Butler	15	RONIT	Kulkarni	46	NICOLE	Smith	76
JAMES	Clark	16	SAMUEL	Lambert	47	REG	Taylor	77
SARAH	Colombo	17	SYDNEY	Lovelace	48	ADDISON	Testoff	78
AMANDA	Corrado	18	Adriana	Martinez Coronado	49	BRANDON	Todd	79
LOGAN	Crane	19	MATTHEW	McCrosson	50	ANDREAS	Tsakanias	80
KELSEY	Dillon	20	LAUREN	Medlin	51	RAINE	Valcich	81
JILLIAN	Faircloth	21	NATHAN	Meek	52	JOSHUA	Veloso	82
JOSHUA	Farris	22	HANNAH	Moore	53	EMMA	Wagner	83
RINESHA	Finklea	23	CASEY	Morrison	54	TAYLOR	Weidner	84
CASEY	Fissel	24	DANIEL	Nelson	55	ARIANA	Wilchenski	85
KAYLAN	Frame	25	COLIN	O'Connor	56	MACKENZIE	Williams	86
MATTHEW	Franklin	26	SIOBHAN	O'Dell	57	MATTHEW	Willis	87
KRYSTYN	Gainey	27	HALEIGH	O'Donegan	58	VICTORIA	Wills	88
AMALIA	Gelinas	28	CLAIRE	O'Loughlin	59	MICHEAL	Wood	89
MADISON	Genal	29	CASSIDY	Onley	60	EMILY	Wood	90
CALLUM	Gill	30	ALLISON	Osborne	61	CRISTINA	Young	91
BRELAND	Green	31				KATHRYN	Zabinski	92

Use the following information to answer questions 1 and 2.
For a group of patients in an age related medical study, the sample size was $n = 120$, the mean age of the group was $\bar{y} = 58$ and sample standard deviation was $s = 10$. Assume the age distribution is unimodal and bell shaped.

1. What is the sample variance of the sample data?
(a) 10 (b) 3.162 (c) 100 (d) cannot be determined

2. Approximately how many patients in the group will have ages between 38 and 78?
(a) 82 (b) 114 (c) 118 (d) 94

Use the following information to answer questions 3, 4 and 5.
The growth of a certain type of fungus in a favorable medium was observed and the 48 hour growth (diameter of the area in mm) data of 50 colonies is displayed in the following histogram.



3. Choose the most appropriate scenario for data:
(a) mean > median (b) mean < median (c) mean = median (d) cannot be determined

4. What is the observational unit for the study?
(a) a colony of fungus (b) the diameter of a colony of fungus
(c) hours of growth (d) type of fungus

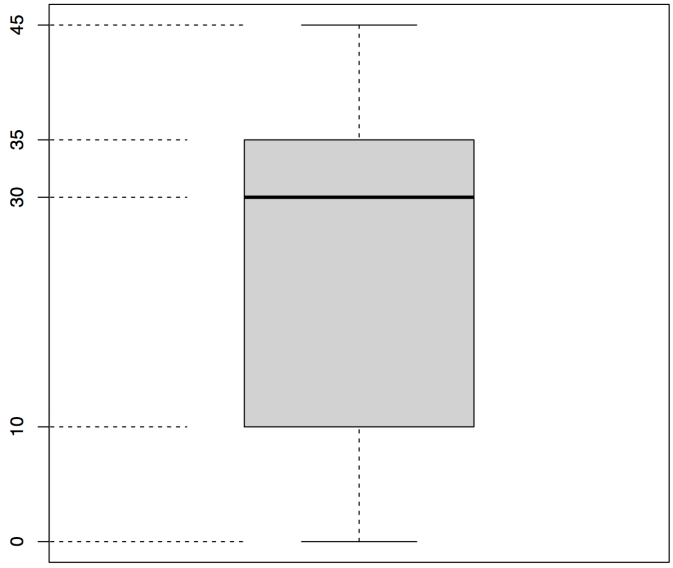
5. What is the study variable?
(a) a colony of fungus (b) the diameter of a colony of fungus
(c) hours of growth (d) type of fungus

6. A data set is consist of 4 observation, y_1, y_2, y_3 and y_4. If the mean $\bar{y} = 40$, then the value of the following sum is

$$(y_1 - 40) + (y_2 - 40) + (y_3 - 40) + (y_4 - 40)$$

(a) always positive (b) always negative (c) always 0 (d) cannot be determined

Use the following information to answer questions 7-11.
An experiment was carried out to see how long it takes toddlers aged 2–3 years to knock over a pile of blocks (in seconds). Use the box-plot for this data set, below, to answer the next five questions.



7. The interquartile range (IQR) for these data is
(a) 35 seconds (b) 25 seconds (c) 45 seconds (d) not computable from box plot

8. 75% of the observations are *less* than
(a) 35 seconds (b) 25 seconds (c) 45 seconds (d) not computable from box plot

9. 75% of the observations are *greater* than
(a) 0 seconds (b) 10 seconds (c) 30 seconds (d) not computable from box plot

10. The upper fence for these data is (*Hint*: the upper boundary to detect outliers)
(a) 45 seconds (b) -27.5 seconds (c) 25 seconds (d)72.5 seconds

11. The box plot suggests the shape of the data is:
(a) symmetric (b) left skewed (c) right skewed (d) Cannot be determined

Use the following information to answer questions 12-18.

The following table cross-classifies 6549 subjects living in Massachusetts according to health risk (stressed or not stressed) and income (low, medium, or high). Use this table to answer the next seven questions.

	Income			
	Low	Medium	High	Total
Stressed	526	274	216	1,016
Not Stressed	1,954	1,680	1,899	5,533
Total	2,480	1,954	2,115	6,549

12. What is the probability that someone in this study is stressed?
(a) 0.298 (b) 0.925 (c) 0.102 (d) 0.155

13. Given that someone has high income, what is the probability that they are stressed, i.e. Pr{stressed|high income}
(a) 0.298 (b) 0.925 (c) 0.102 (d) 0.155

14. What is the probability of someone having low income *or* being not stressed?
(a) 0.298 (b) 0.925 (c) 0.102 (d) 0.155

15. What is the probability of someone having low income *and* being not stressed?
(a) 0.298 (b) 0.925 (c) 0.102 (d) 0.155

16. Given that someone is not stressed, what is the probability that they have high income?
(a) 0.323 (b) 0.213 (c) 0.343 (d) 0.845

17. In this study, income is what type of variable?
(a) numeric discrete (b) numeric continuous (c) categorical ordinal (d) categorical nominal

18. Is income level independent of being stressed?
(a) Yes (b) No (c) Cannot tell from the table (d) Both (a) and (b).

Use the following data to answer questions 19 and 20:

A rare type of HPV virus affects 1% of the population in a state. A new test is designed to detect the virus before an outward manifestation. If given to an infected person, the test shows positive result 97% of the time. If given to an uninfected person, the test shows positive (false) result 2% of the time. (Hint: use a probability tree).

19. If the test is given to a randomly chosen person in the population, what is the probability that it would be positive?
(a) 0.0295 (b) 0.3917 (c) 0.0198 (d) 0.0395

20. If the test is given to a randomly chosen person in the population and gives a positive result, what is the probability that the person is actually infected?
(a) 0.5411 (b) 0.9702 (c) 0.248 (d) 0.3288

21. A certain drug causes kidney damage in 1% of patients. Suppose the drug is to be tested on 50 patients. We want to compute the probability that one or more of the patients will experience kidney damage. The R-code we should use is:

(a) 1 - pnorm(1, 50, 0.01) (b) 1 - pbinom(1, 50, 0.01)
(c) pbinom(1, 50, 0.01) (d) 1 - dbinom(0, 50, 0.01)

22. Suppose we are trying to compare the average response of an experiment group to a control group. For which of the following cases a t-test will <u>not</u> be appropriate?

(a) We have 45 observations from the experiment group and 51 observations from the control group QQ-plot for both the samples are fairly close to straight lines.
(b) We have 15 observations from the experiment group and 10 observations from the control group and QQ-plot for both the samples are fairly close to straight lines.
(c) We have 12 observations from the experiment group and 23 observations from the control group and QQ-plot for both the samples shows clear curves away from straight lines.
(d) We have 123 observations from the experiment group and 213 observations from the control group and QQ-plot for both the samples shows clear curves away from straight lines.

<u>Use the following to answer questions 23 to 25</u>
As a part of study of the development of thymus glands, researchers weighed the glands of five chick embryos after 14 days of incubation. The thyme weights(mg) were as follows:

29.6 21.5 28.0 34.6 44.9

The mean is 31.7 and the standard deviation is 8.7.

23. A point estimate for the mean thymus gland weight of chick embryos is:
(a) 31.7 (b) 8.7 (c) $31.7/\sqrt{5}$ (d) $8.7/\sqrt{5}$

24. The *standard error* of the point estimate in the previous question is:
(a) 31.7 (b) 8.7 (c) $31.7/\sqrt{5}$ (d) $8.7/\sqrt{5}$

25. If we assume the thymus gland weights of chick embryos follow approximately a normal distribution, based on the empirical law an approximate 95% confidence interval, for mean thymus gland weight will be given by:

(a) $[(31.7 - 2\times 8.7),\ \ (31.7 + 2\times 8.7)]$ (b) $[(31.7 - 2\times \frac{8.7}{\sqrt{5}}),\ \ (31.7 + 2\times \frac{8.7}{\sqrt{5}})]$

(c) $[(8.7 - 2\times 31.7),\ \ (8.7 + 2\times 31.7)]$ (d) $[(8.7 - 2\times \frac{31.7}{\sqrt{5}}),\ \ (8.7 + 2\times \frac{31.7}{\sqrt{5}})]$

Use the following to answer questions 26 and 27.

Six healthy three-year old Suffolk sheep were injected with antibiotic Gentamicin, at a dose of 10mg/kg body weight. Their blood serum concentration (mg/ml) of Gentamicin 1.5 hours after injection were as follows:

33 26 34 31 23 25

The data in R gives the following output:

```
> concentration=c(33,26,34,31,23,25)
> t.test(concentration)

        One Sample t-test

data:  concentration
t = 15.2987, df = 5, p-value = 2.164e-05
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 23.84992 33.48341
sample estimates:
mean of x
 28.66667
```

26. A point estimate for the mean concentration of Gentamicin in the population of three year old Suffolk sheep after 1.5 hours of injection is

(a) 10 (b) 15.2987 (c) 2. 164e-05 (d) 28.66667

27. A 90% confidence interval estimate for the mean concentration of Gentamicin in the population of three year old Suffolk sheep after 1.5 hours of injection is

(a) [23.84992, 33.48341] (b) [24.89088 32.44246]

(c) [21.11125 36.22208] (d) [15.79589 41.53745]

28. The Wisconsin Fast Plant grows fast. Ancymidol (ancy) slows growth. In a study control (no ancy) and plants treated with ancy were measured (cm) after two weeks. We want to test if Ancymidol indeed reduces plant growth significantly. Based on the output from R (next page), what is the Satterthwaite-Welch degrees of freedom used for the t-test in this case?

(a) -2.1583 (b) 13.245 (c) 0.02491 (d) 15

```
> ancy=c(12.2,18.5,13.0,9.8,12.8,7.1,7.7,18.9)
> control=c(16.0,15.2,18.8,19.3,20.2,13.9,20.3,10.6)
> t.test(ancy, control, alternative="less")

        Welch Two Sample t-test

data:  ancy and control
t = -2.1583, df = 13.245, p-value = 0.02491
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -0.774485
sample estimates:
mean of x mean of y
  12.5000   16.7875
```

Use the following to answer questions 29 and 30

The serum cholesterol levels of a population of 12 to 14 year olds follow a normal distribution with mean 162 mg/dl and standard deviation 28 mg/dl. Here are some calculations done in R that are related to this problem. Use the appropriate ones to answer next two questions:

pnorm(152,162,28) = 0.3604924	**pnorm(152,162,$\frac{28}{3}$)** = 0.1419884	**pnorm(152,162,$\frac{28}{9}$)** = 0.0006538474
pnorm(172,162,28) = 0.6395076	**pnorm(172,162,$\frac{28}{3}$)** = 0.8580116	**pnorm(172,162,$\frac{28}{9}$)** = 0.9993462

29. What percentage of 12 to 14 years olds has serum cholesterol between 152 and 172 mg/dl?
(a) 0.9993462 (b) 0.9986923 (c) 0.7160232 (d) 0.2790151

30. If $\overline{Y}$ represents the mean cholesterol value of a random sample of nine 12 to 14 year from the population, what is $\Pr(152 \le \overline{Y} \le 172)$?
(a) 0.9993462 (b) 0.9986923 (c) 0.7160232 (d) 0.2790151

This page is left blank intentionally as scratch paper.

This page is left blank intentionally as scratch paper.

This page is left blank intentionally as scratch paper.