

General Instructions: You may use any calculator you like. You may have up to four 3x5 cards, but no other notes. Only the answer on the answer card will be graded.

Problems 1-20: Multiple choice. Each problem is worth 4 points.

1. The following table gives the number of times each digit occurs in the first 10,000 digits of the decimal expansion of e .

Digit	0	1	2	3	4	5	6	7	8	9
Frequency	974	989	1005	1008	982	992	1079	1008	994	969

Are the digits uniformly distributed? Perform an appropriate hypothesis test and find the P -value.

- (a) 1.0000
- (b) 0.8990
- (c) 0.5728
- (d) 0.5166
- (e) 0.4773
- (f) 0.0315
- (g) 0.0259
- (h) 0.0000

2. The following table shows who survived the sinking of the Titanic, based on their class of passage:

	Crew	1st/2nd	3rd
Alive	212	320	178
Dead	673	290	528

Were the chances of surviving the accident the same across all classes? Perform an appropriate hypothesis test and find the P -value.

- (a) 1.0000
 - (b) 0.3141
 - (c) 0.2245
 - (d) 0.1323
 - (e) 0.0660
 - (f) 0.0142
 - (g) 0.0030
 - (h) 0.0000
3. Assume that birthdays are uniformly distributed throughout the year, and that no one has a leap year birthday. What is the probability that at least one of a randomly chosen group of 105 people has a April 15th birthday?
- (a) 100%
 - (b) 75%
 - (c) 60%
 - (d) 50%
 - (e) 40%
 - (f) 25%
 - (g) 20%
 - (h) 0%

4. In a study of attitudes about higher education, you select 4 representative local colleges and universities, and perform a simple random sample of 50 men and 50 women from each. Which, if any, of the following techniques are you using?

I. Clustering

II. Stratification

III. Blocking

- (a) I, II, and III.
- (b) I and II only.
- (c) I and III only.
- (d) II and III only.
- (e) I only.
- (f) II only.
- (g) III only.
- (h) None of the above.

5. You pick a random number according to the Student t -distribution with 3 degrees of freedom. What is the expected value?

- (a) -3
- (b) -1.7
- (c) -1
- (d) 0
- (e) 1
- (f) 1.7
- (g) 3
- (h) ∞

6. Marilyn is flying from New York to Stockholm, with a connection in Amsterdam. The probability that his first flight leaves on time is 0.2. If the flight is on time, the probability that his luggage makes the connecting flight in Amsterdam is 0.95, but if the first flight is delayed, then the probability that his luggage makes the flight drops to 0.7. If Marilyn's luggage did not make it to Stockholm, what is the probability that his first flight was delayed?

- (a) 100%
- (b) 96%
- (c) 90%
- (d) 84%
- (e) 80%
- (f) 75%
- (g) 64%
- (h) 56%

7. Census data for St. Louis County shows that about 20.5% of the adult residents are black. If 80 people are called for jury duty, and only 10 of them are black, does this call into question the fairness of the jury selection system? Find the P -value from an appropriate 2-sided hypothesis test.

- (a) 1.0000
- (b) 0.3505
- (c) 0.1985
- (d) 0.1452
- (e) 0.0763
- (f) 0.0525
- (g) 0.0045
- (h) 0.0000

8. In an investigation on the environmental causes of disease, data was collected on the mortality rate (in deaths per 100,000) for males in 61 large towns in England and Wales. The concentration of calcium in the water (in ppm) in each town was also recorded, as well as whether the town was North or South of Derby (a town in the center of England). Read into R, the data looks like:

```

      mortality calcium Derby
1         1702         44 South
2         1309         59 South
3         1259        133 South
...
60        1828          8 North
61        1704         26 North

```

You perform the following linear regression:

```

Call: lm(formula = mortality ~ calcium + Derby)
Residuals:      Min       1Q   Median       3Q      Max
      -232.14   -83.33    2.95    84.53   307.89
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1696.3916    26.3191  64.455 < 2e-16 ***
calcium      -2.1601     0.4979  -4.338 5.81e-05 ***
DerbySouth  -158.8920    37.8733  -4.195 9.46e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 126.4 on 58 degrees of freedom
Multiple R-squared:  0.5618,    Adjusted R-squared:  0.5467
F-statistic: 37.18 on 2 and 58 DF,  p-value: 4.059e-11

```

(continued on the next page)

What does your model predict for the mean deaths per 100,000 in a town with 150,000 residents, 45 ppm calcium, in the far south of England?

- (a) 2160
- (b) 1696
- (c) 1599
- (d) 1538
- (e) 1534
- (f) 1440
- (g) 562
- (h) 547

9. In the situation of Question 8, find the t^* corresponding with a 95% margin of error, together with the upper bound for a 95% confidence interval for the coefficient of the indicator variable.

- (a) $t^* = 2.002$, U.B. = -1
- (b) $t^* = 2.000$, U.B. = -1
- (c) $t^* = 1.960$, U.B. = -1
- (d) $t^* = 1.672$, U.B. = -1
- (e) $t^* = 2.002$, U.B. = -83
- (f) $t^* = 2.000$, U.B. = -83
- (g) $t^* = 1.960$, U.B. = -85
- (h) $t^* = 1.672$, U.B. = -96

10. The following table gives the fuel economy (in Miles Per Gallon) for several popular models of cars, together with the number of cylinders in the car's engine.

MPG	Cylinders	MPG	Cylinders	MPG	Cylinders
16.9	8	17	6	30	4
15.5	8	16.2	6	27.5	4
19.2	8	20.6	6	27.2	4

We found in Chapter 28 that the equal variance condition was not met for this data. We'd also previously seen that Gallons Per Mile was a good reexpression for fuel economy.

Is the mean Gallons Per Mile the same over all groups? Reexpress the data, then perform an appropriate ANOVA F -test and report the P -value.

- (a) 1.0000
- (b) 0.9999
- (c) 0.6945
- (d) 0.0038
- (e) 0.0031
- (f) 0.0007
- (g) 0.0001
- (h) 0.0000

11. Every year the students at Tukey Middle School take a physical fitness class, for which they do as many pushups as they can. Results for a random (stratified) selection are shown below, separated by gender.

Boys		17	27	31	17	25	32	28	20	29
Girls		24	20	18	14	5	22	20	18	16

Is there a gender difference in average number of pushup that students at Tukey Middle School can do? Find the P -value for an appropriate 2-sided t -test.

- (a) 1.0000
 - (b) 0.8846
 - (c) 0.0721
 - (d) 0.0196
 - (e) 0.0182
 - (f) 0.0111
 - (g) 0.0010
 - (h) 0.0000
12. Do freshmen gain weight on average during their first semester at Washington University in St. Louis? Perform an appropriate 2-sided test on the following weights (in pounds) of 7 randomly selected freshmen, and find the P -value.

August weight		171	110	134	115	104	120	142
December weight		168	111	136	119	106	124	148

- (a) 1.0000
- (b) 0.9999
- (c) 0.9958
- (d) 0.8529
- (e) 0.3688
- (f) 0.0797
- (g) 0.0500
- (h) 0.0000

13. Cuckoos lay their eggs in the nest of other (host) birds. The following table contains lengths (in mm) of eggs laid by cuckoos in robin and wagtail nests:

Robin	21.05	21.75	22.05	22.05	22.05	22.05	22.65	23.05	23.25	23.85
Wagtail	21.85	21.85	21.85	22.45	23.05	23.45	24.05	24.05	24.85	

Are the mean egg lengths different? Find the test statistic for this data from Tukey's quick test.

- (a) 0
 - (b) 1
 - (c) 2
 - (d) 3
 - (e) 4
 - (f) 5
 - (g) 6
 - (h) Tukey's test doesn't apply.
14. Perform a 2-sided 2-sample t -test on the data from Question 13 and find the P -value.
- (a) 1.00
 - (b) 0.64
 - (c) 0.32
 - (d) 0.16
 - (e) 0.10
 - (f) 0.08
 - (g) 0.05
 - (h) 0.00

15. In a recent poll of 500 likely voters on the upcoming Missouri Senate race, 45% of 260 women favored Republican candidate Roy Blunt, while 50% of 240 men favored him. (Blunt nonetheless polls as leading 48 to 42% on his Democratic opponent Robin Carnahan, due to third party candidates and undecided voters.)

Does Blunt have a gender gap? Perform a 2-sided 2-proportion z -test, and find the P -value.

- (a) 1.0000
- (b) 0.2632
- (c) 0.1692
- (d) 0.1195
- (e) 0.1069
- (f) 0.0460
- (g) 0.0009
- (h) 0.0000

16. You test out three different frisbee grips for your forehand throw, making 10 throws with each grip and measuring the distance that the frisbee travelled. After checking the conditions, you perform an ANOVA to test if the mean distance from all three grips is the same. If your F -statistic is 8.0, what is the P -value of your test?

- (a) 1.0000
- (b) 0.5341
- (c) 0.1170
- (d) 0.0156
- (e) 0.0019
- (f) 0.0076
- (g) 0.0005
- (h) 0.0000

17. In our ANOVA analysis of SUGARS with respect to factor(SHELF) in the breakfast cereals data set, the grand mean of SUGARS is 6.9 g, while the “treatment effects” from shelf 1, 2, and 3 are -2.1 g, 2.7 g, and -0.4 g respectively.

If Chocolate Frosted Sugar Bombs are a Shelf 2 cereal with 17 grams of sugar per serving, find their residual.

- (a) 17
- (b) 14.5
- (c) 14.3
- (d) 12.2
- (e) 10.5
- (f) 8.6
- (g) 7.4
- (h) 5.7

18. The following table of actuarial data indicates the estimated additional years of life for a black male of a given age in the United States.

Age	10	20	30	40	50	60	70	80	90
Years Left	59	50	41	32	24	17	12	7	4

Reexpress the Years Left appropriately, create a linear model for the reexpressed data, and find the absolute value of the t -score for a 2-sided regression slope t -test.

(Be sure to check that your reexpressed scatterplot is “straight enough”!)

- (a) 120
- (b) 95
- (c) 25
- (d) 17
- (e) 5
- (f) 2
- (g) 1
- (h) 0

19. From the following multi-factor ANOVA, identify the P -value of the interaction term.

Analysis of Variance Table

Response: Accuracy

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Distance	2	51.044	25.522	41.9769	2.023e-09	***
Hand	1	39.690	39.690	65.2796	5.101e-09	***
Distance:Hand	2	10.355	5.177	8.5156	0.001178	**
Residuals	30	18.240	0.608			

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

- (a) 41.9769
- (b) 25.522
- (c) 2.023e-09
- (d) 65.2796
- (e) 5.101e-09
- (f) 8.5156
- (g) 5.177
- (h) 0.001178

20. A consumer organization has reported test data for 50 car models, including fuel efficiency (in Miles Per Gallon) and weight (in thousands of pounds). We reexpress fuel efficiency in Gallons Per Mile, and use R to create the following regression model:

```
Call: lm(formula = 1/MPG ~ wt.1000)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.001661   0.003050   0.545   0.588
wt.1000      0.013774   0.001040  13.241  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.003726 on 48 degrees of freedom
Multiple R-squared:  0.7851,    Adjusted R-squared:  0.7806
F-statistic: 175.3 on 1 and 48 DF,  p-value: < 2.2e-16
> mean(fuel2)
      MPG wt.1000
25.0200  2.8878
```

Find the margin of error for a 95% confidence interval for the mean Gallons Per Mile consumption of all cars weighing 2500 pounds.

- (a) 0.0052
- (b) 0.0049
- (c) 0.0026
- (d) 0.0021
- (e) 0.0013
- (f) 0.0011
- (g) 0.0007
- (h) 0.0000

Problems 21-30: True/false. Each problem is worth 2 points.

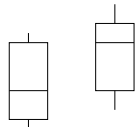
21. True/false: In Question 16, we conclude that the significant difference between mean distances of frisbee grips is likely caused by the different grips.

- (a) True
- (b) False

22. True/false: An archer performs an experiment, with 3 levels of the factor Distance (Near, Middle, Far). She shoots 10 arrows at a target for each of the 3 treatment conditions, and counts the number of bullseyes. To determine if there is a significant difference in the number of bullseyes, she should perform ANOVA.

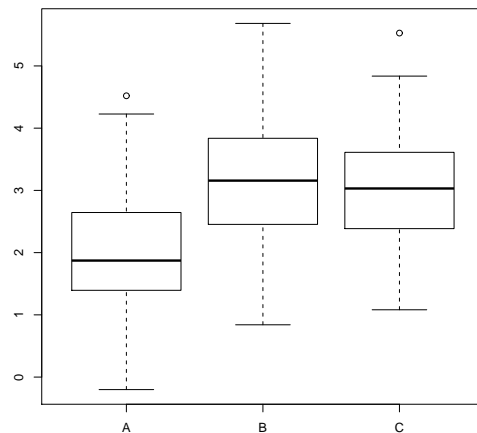
- (a) True
- (b) False

23. True/false: The following boxplot represents two samples of size 9. We conclude from Tukey's quick test that we can reject the null hypothesis (that the means are equal) at a significance level of 0.001.



- (a) True
- (b) False

24. True/false: You are interested in comparing the mean weight of Wash U students with that of SLU students. You take a random sample of 100 students from each university, and weigh each of them. The weights are skew and bimodal, with an outlier in each sample. It is appropriate to use a Wilcoxon test.
- (a) True
(b) False
25. True/false: In the situation of Question 24, it is appropriate to use a 2-sample t -test.
- (a) True
(b) False
26. True/false: You perform an ANOVA test on the data represented in the following boxplot



and find a P -value of 0.0035. We conclude that there is a statistically significant difference between group A and group C.

- (a) True
(b) False

27. True/false: Interested in whether pepperoni pizza has more fat than cheese pizza, you collect nutritional information for both pepperoni and cheese in 10 brands of frozen pizza. Assuming the conditions are met, you will use a 2-sample t -test to compare them.

(a) True

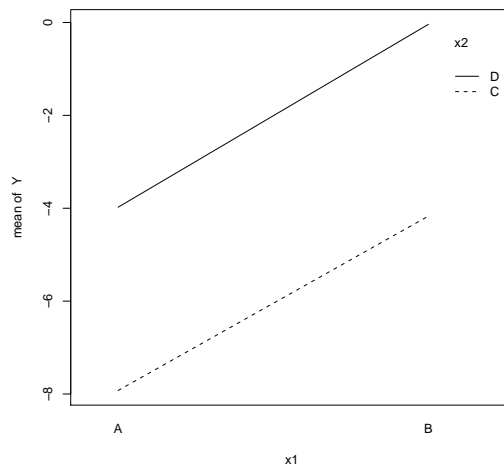
(b) False

28. True/false: Two random variables X and Y both have expected value 0 and standard deviation 1. It follows that $P(X > 0) = P(Y > 0)$.

(a) True

(b) False

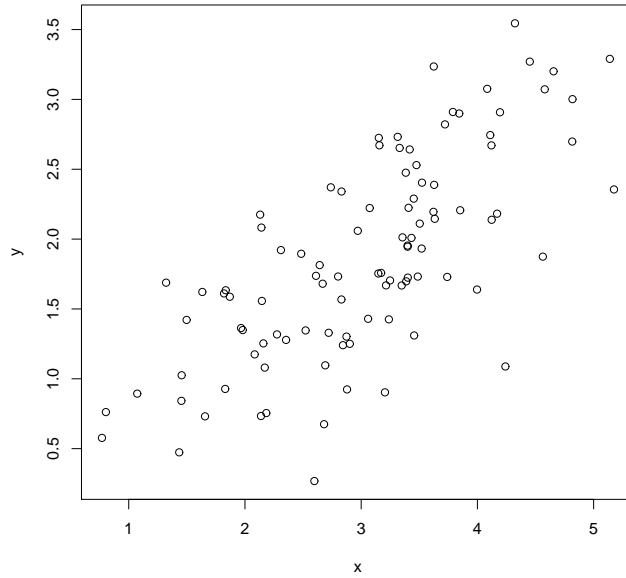
29. True/false: From the following interaction plot, it appears that the interaction term $x_1 * x_2$ will be significant in the multifactor ANOVA of Y over x_1 and x_2 .



(a) True

(b) False

30. True/false: The data represented in the following scatterplot obeys the “straight enough” condition, and it is appropriate to fit a linear regression to this data.



- (a) True
(b) False