Sta 101
Fall 16

Dr. Mukherjee
Dec 19, 2016

Final (Version A)

Last Name: _____ **First Name:** _____

I hereby state that I have not communicated with or gained information in any way from my classmates during this exam, and that all work is my own.

Signature : _____

Any potential violation of Duke's policy on academic integrity will be reported to Undergraduate Conduct Board. All work on this exam must be your own.

1. You have 3 hours to complete the exam.
2. Show **all** your work on the open ended questions in order to get partial credit. No credit will be given for open ended questions where no work is shown, **even if the answer is correct.**
3. Mark the answers to the multiple choice and true/false questions by filling in the **bubbles** provided below. If you choose more than one answer, you will not receive any credit for that question. No partial credit will be given for multiple choice questions.
4. You are allowed a calculator, however you may not share a calculator with another student during the exam, one $8\frac{1}{2}$" $\times$ 11" sheet of notes ("cheat sheet") with writing on both sides, pen or a pencil, a dictionary, and to ask questions to me or the TA.
5. You are **not** allowed a cell phone, even if you intend to use it as a calculator or for checking the time, music device or headphones, notes (other than the one $8\frac{1}{2}$" $\times$11" sheet), books, or other resources, and to communicate with anyone other than myself during the exam.
6. Write clearly. Short answers are best!

Good luck!

11. (T) (F)	16. (T) (F)	21. (A) (B) (C)	26. (A) (B) (C) (D)	31. (A) (B) (C) (D)	36. (A) (B) (C) (D)
12. (T) (F)	17. (T) (F)	22. (A) (B) (C)	27. (A) (B) (C) (D)	32. (A) (B) (C) (D)	37. (A) (B) (C) (D)
13. (T) (F)	18. (T) (F)	23. (A) (B) (C)	28. (A) (B) (C) (D)	33. (A) (B) (C) (D)	38. (A) (B) (C) (D)
14. (T) (F)	19. (T) (F)	24. (A) (B) (C)	29. (A) (B) (C) (D)	34. (A) (B) (C) (D)	39. (A) (B) (C) (D)
15. (T) (F)	20. (T) (F)	25. (A) (B) (C)	30. (A) (B) (C) (D)	35. (A) (B) (C) (D)	40. (A) (B) (C) (D)

Question #1	/ 15	Match the following #6-10	/ 5
Question #2	/ 10	Fill in the blanks #11-15	/ 10
Question #3	/ 15	True / False #16-25	/ 10
Question #4	/ 10	Multiple Choice #26-35	/ 20
Question #5	/ 5	**TOTAL**	**/ 100**

1. ***Car window – skin cancer? - 15 points***

A recent study suggests that exposure to UV rays through the car window may increase the risk of skin cancer.[1] The study reviewed the records of all 1,047 skin cancer patients referred to the St. Louis University Cancer Center in 2004, some of which had melanoma (a type of skin cancer).

(a) **(8)** Of the 42 patients with melanoma, the cancer occurred on the left side of the body in 31 patients and on the right side in the other 11. Using $\alpha = 0.05$, test whether these data provide evidence that melanoma is found more often on the left side of the body. Clearly define your parameter(s) of interest, the sample statistic(s), state your hypotheses, check the conditions, calculate a p-value, give a conclusion, and interpret the results in context.

Parameter of interest: p = Proportion of all melanoma patients with cancer on the left side.
Sample statistic: $\hat{p} = 31 / 42 = 0.738$
Hypotheses: $H_0 : p = 0.5$; $H_A : p > 0.5$
Conditions:

1. *We do not know if the patients are randomly sampled, but they're likely less than 10% of the population of all melanoma patients. If this is a random sample, we can assume that the location of the melanoma of one patient is independent of another in the sample.*
2. *Expected number of successes and failures is $42 \times 0.5 = 21$, which is greater than 10.*

$$Z = \frac{0.738 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{42}}} = \frac{0.238}{0.0772} = 3.08$$

$$p - value = P(Z > 3.08) = 0.0010$$

Since p-value $< 5\%$, we reject H_0. These data provide convincing evidence that melanoma is found more often on the left side of the body.

[$Z = [3.08, 3.11]$ ***acceptable due to rounding.]***
[1 pt for parameter, 1 pt for statistic, 1 pt for hypotheses, 2 pts for conditions, 1 pt for correct SE, 1 pt for correct Z, 1 pt for correct p-value, 1 pt for correct interpretation in context.]

(b) **(1)** The same study also collected information on the gender of the patients. Of the patients with melanoma on the left 17 were male, and of the patients with melanoma on the right 5 were male. Using this information, fill in the cells of the following contingency table.

		Location of melanoma		
		Left	Right	Total
Gender	Male	*17*	*5*	*22*
of patient	Female	*14*	*6*	*20*
	Total	*31*	*11*	*42*

[0.25 pts for each cell in the table, no points for cells on the margins]

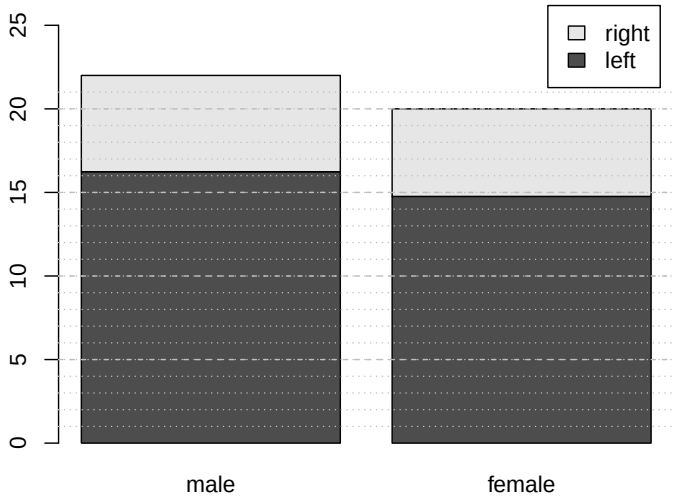
[1] Susan T. Butler, Scott W. Fosko, Increased prevalence of left-sided skin cancers, *Journal of the American Academy of Dermatology*, Volume 63, Issue 6, December 2010.

(c) **(3)** Assume there is **no association** between location of melanoma and gender. By drawing on the bar chart below, indicate what a segmented bar chart would look like under this assumption. Be sure to indicate which regions represent melanoma on the left and right sides, and show your calculations for calculating the heights of those regions.

Expected number of left in each category:

$22 * 0.738 = 16.236$

$20 * 0.738 = 14.76$



[1 pt each for each expected count, 1 pt for correct shading and labeling on the figure]

(d) **(2)** Which of the following method(s) can be used to test for the relationship between gender of patient and location of melanoma? Circle **all** that apply. Only select tests where conditions are **met**.

I. ANOVA

II. Chi-squared goodness of fit

III. F-test

IV. Multiple linear regression

V. ***Randomization test***

VI. Simple linear regression

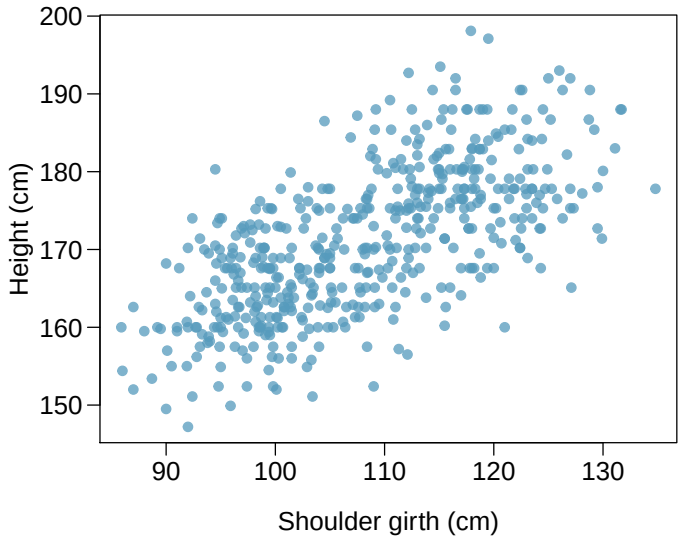
VII. T-test

VIII. Z-test

[0.25 for each part]

2. ***Body measurements - 10 points***

Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender for 507 physically active individuals. The scatterplot below shows the relationship between height and shoulder girth (over deltoid muscles), both measured in centimeters.



The mean shoulder girth is 107.20 cm with a standard deviation of 10.37 cm. The mean height is 171.14 cm with a standard deviation of 9.41 cm. The correlation between height and shoulder girth is 0.67.

(a) **(4)** Calculate the slope b_1 and intercept b_0 and interpret in the context of the data.

$b_1 = 10.37$. For each centimeter increase in shoulder girth, we would expect height to increase on average by 0.608 centimeters.
$b_0 = 171.14 - 0.608 \times 107.20 = 105.96$. People who have a shoulder girth of 0 cm are expected to be on average 105.79 cm tall. Here, the y-intercept serves only to adjust the height of the line and is meaningless by itself.

(b) **(1)** Fill in the blanks: The regression line can be written as

$$\underline{\widehat{height}} = \underline{105.96} + \underline{0.608} \times \underline{shoulder\ girth}$$

(c) **(2)** Calculate R^2 of the regression line for predicting height from shoulder girth, and interpret it in the context of the application.

$R^2 = 0.67^2 = 0.45$. Approximately 45% of the variation in heights is accounted for by the model, i.e. explained by shoulder girth.

(d) **(1)** A randomly selected student from your class has a shoulder girth of 100 cm. Predict the height of this student using the model

$\widehat{height} = 105.96 + 0.608 \times shouldergirth = 105.96 + 0.608 \times 100 \approx 167cm$

(e) **(1)** The student from part (d) is 160 cm tall. Calculate the residual, and explain what this residual means.

The residual can be calculated as $e_i = y_i - \hat{y}_i = 160 - 166 = -6cm$. A negative residual means that the model overestimated this student?s height.

(f) **(1)** A one year old has a shoulder girth of 56 cm. Would it be appropriate to use this linear model to predict the height of this child?

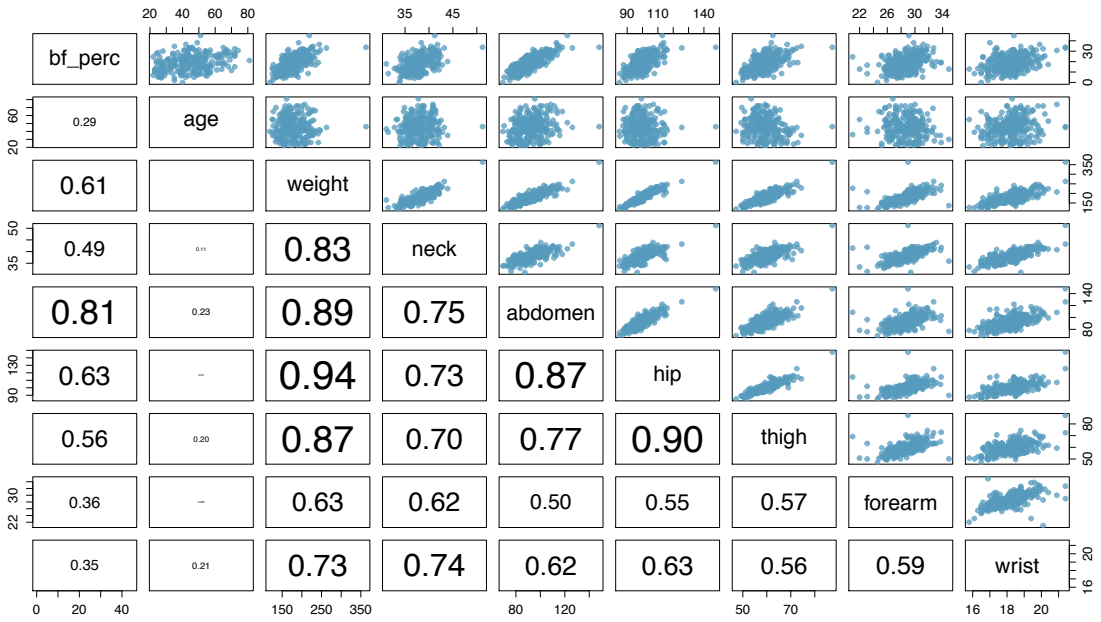
No. Predicting the height of a child would require extrapolation. The data used to create the model is for people with approximately 90 to 130 cm shoulder girth. The linear model may no longer hold outside the range of the data.

3. ***Body fat percentage - 15 points***

Body fat percentage can be complicated to estimate, while variables such age, height, weight, and measurements of various body parts are easy to measure. Based on data[2] on body fat percentage and other various easy to obtain measurements, we develop a model to predict body fat percentage based on the following variables:

- age (years)
- weight (pounds)
- neck circumference (cm)
- abdomen circumference (cm)
- hip circumference (cm)
- thigh circumference (cm)
- forearm circumference (cm)
- wrist circumference (cm)

The plot below shows the relationship between each of these variables and body fat percentage (the response variable) as well as the correlation coefficients between these variables:



And the following are the model outputs associated with this analysis:

Regression Summary	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-20.062	10.847	-1.850	0.066
age	0.059	0.028	2.078	0.039
weight	-0.084	0.037	-2.277	0.024
neck	-0.432	0.208	-2.077	0.039
abdomen	0.877	0.067	13.170	0.000
hip	-0.186	0.128	-1.454	0.147
thigh	0.286	0.119	2.397	0.017
forearm	0.483	0.173	2.797	0.006
wrist	-1.405	0.472	-2.978	0.003

ANOVA	Df	Sum Sq	Mean Sq	F value	Pr(>F)
age	1	1260.93	1260.93	80.21	0.0000
weight	1	5738.41	5738.41	365.04	0.0000
neck	1	153.37	153.37	9.76	0.0020
abdomen	1	3758.51	3758.51	239.09	0.0000
hip	1	6.42	6.42	0.41	0.5234
thigh	1	122.04	122.04	7.76	0.0058
forearm	1	79.91	79.91	5.08	0.0251
wrist	1	139.46	139.46	8.87	0.0032
Residuals	243	3819.99	15.72		
Total	251	15079.02			

Residual standard error: 3.965 on 243 degrees of freedom
F-statistic: 89.53 on 8 and 243 DF, p-value: $< 2.2e\text{-}16$

[2] Penrose, K., Nelson, A., and Fisher, A. (1985), *Generalized Body Composition Prediction Equation for Men Using Simple Measurement Techniques*, Medicine and Science in Sports and Exercise, 7(2), 189.

(a) **(2)** Which of the following is **not** supported by information provided in the model outputs above?

i. The sample size is 252.

ii. All else held constant, for each additional cm the forearm circumference is higher, body fat percentage is expected to be higher by 0.483 percentage points.

iii. All else held constant, people with wider hips tend to have lower body fat percentages.

iv. ***Wrist circumference is the most significant predictor of body fat percentage since the slope associated with this variable has the highest magnitude.***

v. The F-test for the significance of the model overall suggests that at least one of the slope coefficients is significantly different than 0.

(b) **(2)** What is the adjusted R^2 for this model? Choose the closest answer.

i. 0.2286

ii. 0.2617

iii. 0.7467

iv. 0.7548

v. ***0.7383***

(c) **(2)** If we were to remove one variable from the model, which of the following values would **not** change in the ANOVA table? Check **all** that apply.

☐ $Df_{Residuals}$: 243

☐ Df_{Total} : 251

☐ $SumSq_{Residuals}$: 3819.99

☐ $SumSq_{Total}$: 15079.02

(d) **(3)** Investigate if these data provide convincing evidence that age and body fat percentage are significantly **positively** related. Answer the following questions for your investigation.

- Provide the p-value for the above hypothesis testing: *(1)* 0.0195

- Based on the above hypothesis test, what is your decision? (Circle one below, assume $\alpha = 0.05$) *(1)*

Reject H_0 Fail to Reject H_0

- Can you conclude from the p-value above that age and body fat percentage are significantly positively related? (Circle one below) *(1)*

Yes. No.

(1) Yes, the p-value for testing for a positive correlation between age and body fat percentage is 0.039 / 2 = 0.0195.
(1) Since the p-value is small we reject the null hypothesis of no relationship.

(e) **(4)** Construct a 95% confidence interval for the slope of abdomen circumference and interpret it in context of the data.

(1) $T^*_{243} = 1.97$
(1) $0.877 \pm 1.97 * 0.067$ = *(0.745, 1.009)*
(2) All else held constant, for each additional cm the abdomen circumference is higher, body fat percentage is expected to be higher by 0.745 to 1.009 percentage points.

[1 for T*, 1 for CI, 1 for interpretation
Common mistakes:
If missing "all else held constant" -1,
If sounding causal -1,
If missing "expected" and/or "on average" -1]

(f) **(2)** Upon a closer look at the model, your friend questions *"The coefficient for* `weight` *is negative? and significant? That doesn't make any sense! How can this be possible?"*. How would you respond?

(1) Weight and some of the other explanatory variables are highly correlated (such as neck, abdomen, hip, thigh, etc. as seen from the plot), which introduces collinearity into the model.
(1) This results in unreliable slope estimates.

[Common mistakes:
Term "collinearity" is not used -1
It's not clear that student understands impact of collinearity (unreliable slope estimates) -1]

4. ***Barking deer - 10 points***

Microhabitat factors associated with forage and bed sites of barking deer in Hainan Island, China were examined from 2001 to 2002. In this region woods make up 4.8% of the land, cultivated grass plot makes up 14.7%, and deciduous forests makes up 39.6%. Of the 426 sites where the deer forage, 4 were categorized as woods, 16 as cultivated grassplot, and 61 as deciduous forests. The table below summarizes these data.

Woods	Cultivated grassplot	Deciduous forests	Other	Total
4	16	61	345	426

(a) **(1)** Write the hypotheses for testing if barking deer prefer to forage in certain habitats over others.
H_0: Distribution of foraging preference follows distribution of available land type. H_A: Distribution of foraging preference follows distribution of available land type.

(b) **(1)** What type of test can we use to answer this research question?
Use a chi-squared goodness of fit test.

(c) **(2)** Calculate the expected counts for Woods, Cultivated grassplot, Deciduous forests and Other.
Expected counts can be calculated as follows: $E_{Wood} = 426 \times 0.048 = 20$ $E_{Cultivatedgrassplot} = 426 \times 0.147 = 63$ $E_{Deciduousforests} = 426 \times 0.396 = 169$ $E_{Other} = 426 \times (1 - (0.048 + 0.147 + 0.396)) = 174$
These are all above 5.

(d) **(1.5)** Check if the assumptions and conditions required for this test are satisfied.

1. Independence: We are told / ***not told*** (Circle one) if these plots are sampled randomly.
2. Sample size: All ***expected*** / observed counts (Circle one) are greater than _____.

(e) **(2.5)** Investigate if these data provide convincing evidence that barking deer prefer to forage in certain habitats over others. Given that Chi-squared test statistic value is 284.93 for the hypothesis test in part (a), calculate the following:

- What is the contribution of "Deciduous forests" in the Chi-squared test statistic? (Show your work)
  $(61 - 169)^2/169 = 69$ ***(1)***

- degrees of freedom: *3* ***(0.5)***

- p – value: $p\text{–}value = P(\chi^2 > 284.93) < 0.001$ ***(1)***

(f) **(2)** Based on the p–value above, what is the decision and conclusion of the test, in the context of the data? ($\alpha = 0.05$)

- Decision: ***Reject*** H_0 Fail to Reject H_0 (Circle one)

- Conclusion:

Since the p-value is less than 5%, we reject H_0. The data provide strong evidence that barking deer prefer to forage in certain habitats over others.

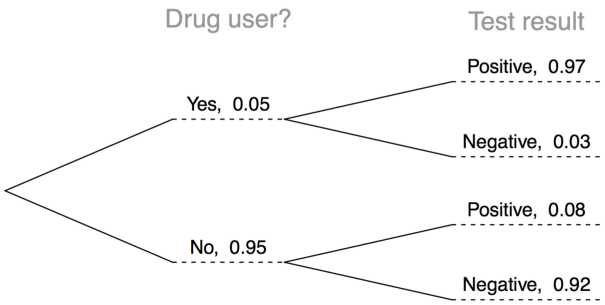
5. ***Drug testing - 5 pts***

Most companies drug test their employees before they start employment, and sometimes regularly during their employment as well. Suppose that a drug test for an illegal drugs is 97% accurate in the case of a user of that drug, and 92% accurate in the case of a non-user for that drug. Suppose also that 5% of the entire population uses that drug.

(a) **(1)** You are the hiring manager at a company that drug tests their employees. You have recently decided to hire a new employee. What is the prior probability that this employee is a user of this drug? (You may assume that this prospective employee is a randomly drawn person from the population.)

P(drug user) = 0.05

(b) **(2)** The prospective employee gets drug tested, and the test comes out to be positive. What is the posterior probability that they are actually a user for the drug? (Answer this question based on the tree)



$$
\begin{aligned}
&P(drug\ user \mid +) \\
&= \frac{P(drug\ user\ AND\ +)}{P(+)} \\
&= \frac{0.0485}{0.0485 + 0.076} \\
&\approx 0.39
\end{aligned}
$$

(c) **(1)** When the employee finds out that they tested positive, they refuse the test results, and say they would like to be tested again. What is the new prior probability you should use for this employee being a user of this drug? *P(drug user) = 0.39*

(d) **(1)** The employee tests positive again in the second test. Should the new probability of them actually being a user of this drug be higher or lower than what you calculated before, or the same?

Higher. Lower. (Circle one)

6. (5 pts) ***City council survey. Match the following***
A city council has requested a household survey be conducted in a suburban area of their city. The area is broken into many distinct and unique neighborhoods, some including large homes, some with only apartments, and others a diverse mixture of housing structures. Match the appropriate sampling methods.

   (1) Divide the city into neighborhoods, randomly sample 10 neighborhoods, and sample all households from those neighborhoods.

   (2) Sample the 200 households closest to the city council offices.

   (3) Randomly sample 50 households from the city.

   (4) Divide the city into neighborhoods, randomly sample 10 neighborhoods, and then randomly sample 20 households from those neighborhoods.

   (5) Divide the city into neighborhoods, and sample 20 households from each neighborhood.

   (a) Simple random sampling.

   (b) Stratified sampling.

   (c) Cluster sampling. .

   (d) Multi-stage sampling.

   (e) Convenience sampling.

Fill in the blanks - 2 points each (Total 10 points)

For questions 7 to 11, fill in the blanks. Write legibly.

7. **(2)** The difference between an observed and predicted response is

   called the _____. *Residual*

8. **(2)** All else equal, a 99% confidence interval is _____ than a 90% confidence interval. *wider*

9. **(2)** An extraneous variable that is related to the explanatory and response variables and that prevents

   us from deducing causal relationships based on observational studies is called a _____. *confounding variable*

10. **(2)** The probability of rejecting the null hypothesis when the null hypothesis is false is

    called the _____. *power of the test*

11. **(2)** The theorem that describes the sampling distribution of a sample statistic is called the _____. *central limit theorem*

True or False - 1 point each (Total 10 points)

Determine if questions 12 to 21 are true or false. Fill in the bubbles on the first page of the exam.

12. (1) (T / ***F***) When the null hypothesis that the means are all the same is rejected using ANOVA, we can then conclude that all the means are different from one another.

13. (1) (***T*** / F) As the degrees of freedom increases, the t distribution approaches the normal distribution.

14. (1) (T / ***F***) If subjects are randomly assigned to treatments, conclusions can be generalized to the population.

15. (1) (T / ***F***) You are going to collect income data from a right-skewed distribution of incomes of politicians. If you take a large enough sample, the sample mean and sample median will always have the same value.

16. (1) (T / ***F***) The margin of error of a confidence interval increases as the sample size increases.

17. (1) (T / ***F***) Least squares regression lines minimize the sum of residuals.

18. (1) (***T*** / F) The χ^2 statistic has a right skewed distribution and is always positive.

19. (1) (***T*** / F) Not all leverage points are influential.

20. (1) (***T*** / F) A correlation coefficient of -0.90 indicates a stronger linear relationship than a correlation coefficient of 0.5.

21. (1) (***T*** / F) The F-test assessing the overall significance of a multiple linear regression model may yield a significant p-value even if some of the variables included in the model are not significant predictors.

Multiple choice (fill in bubbles on the cover page) - 2 points each (Total 30 points)
For questions 22 to 31, select the best answer, and fill in the bubbles on the first page of the exam.

22. **(2)** A random sample of 100 runners who completed the Cherry Blossom 10 mile run yielded an average completion time of 95 minutes. A 95% confidence interval calculated based on this sample is 92 minutes to 98 minutes. Which of the following is **false** based on this confidence interval?

    (a) We are 95% confident that the true average finishing time of all runners who completed the Cherry Blossom 10 mile run is between 92 minutes and 98 minutes.

    (b) ***95% of the time the true average finishing time of all runners who completed the Cherry Blossom 10 mile run is between 92 minutes and 98 minutes.***

    (c) Based on this 95% confidence interval, we would reject a null hypothesis stating that the true average finishing time of all runners who completed the Cherry Blossom 10 mile run is 90 minutes.

    (d) The margin of error of this confidence interval is 3 minutes.

23. **(2)** In a blood testing procedure, blood samples from 4 people are combined into one mixture. The mixture will only test negative if all the individual samples are negative. If the probability that an individual sample tests positive is 0.21, what is the probability that the mixture will test positive? Assume blood samples are independent of each other. Choose the closest answer.

    (a) 0.16

    (b) ***0.61***

    (c) 0.84

    (d) 0.998

    $(1-(1-0.21)^4) = 0.61$

24. **(2)** For post-hoc tests of the results of an ANOVA we use a corrected α. If we want an overall type 1 error rate of 5% what should the α be for the individual pairwise tests if $df_G = 6$.

    (a) 0.00179

    (b) ***0.00238***

    (c) 0.00333

    (d) 0.00714

    $k = 7; K = 7 * 6/2 = 21; 0.05/21 = 0.00238$

25. **(2)** An introductory stats professor hypothesizes that 30% of students learn best by reading the book and watching videos, 35% doing problem sets, and the rest in class. She surveys a random sample of 100 students asking them how they learn best, and wants to use these data to evaluate her hypothesis. Which method should she use?

    (a) ***χ^2 test of goodness of fit***

    (b) χ^2 test of independence

    (c) Z-test

    (d) ANOVA

26. **(2)** Consider a regression predicting weight (kg) from height (cm) for a sample of adult males. What are the units of measurement of the correlation, the intercept, and the slope?

    (a) correlation - cm, intercept - kg, slope - kg

    (b) correlation - cm×kg, intercept - cm, slope - cm/kg

    (c) ***correlation - no units, intercept - kg, slope - kg/cm***

    (d) correlation - no units, intercept - cm, slope - cm

27. **(2)** Hypotheses: $H_0 : p = 0.8; H_A : p > 0.8$. $n = 30$ and $\hat{p} = 0.6$. Suppose we wanted to use simulation-based methods. Which of the following is the correct set up for this hypothesis test? Red: success, blue: failure, $\hat{p}_{sim}$ = proportion of reds in simulated samples.

    (a) Place 60 red and 40 blue chips in a bag. Sample, <u>with</u> replacement, 30 chips and calculate the proportion of reds. Repeat this many times and calculate the proportion of simulations where $\hat{p}_{sim} \geq 0.8$.

    (b) ***Place 80 red and 20 blue chips in a bag. Sample, <u>with</u> replacement, 30 chips and calculate the proportion of reds. Repeat this many times and calculate the proportion of simulations where $\hat{p}_{sim} \geq 0.6$.***

    (c) Place 80 red and 20 blue chips in a bag. Sample, <u>without</u> replacement, 30 chips and calculate the proportion of reds. Repeat this many times and calculate the proportion of simulations where $\hat{p}_{sim} \geq 0.6$.

    (d) Place 80 red and 20 blue chips in a bag. Sample, <u>with</u> replacement, 100 chips and calculate the proportion of reds. Repeat this many times and calculate the proportion of simulations where $\hat{p}_{sim} \geq 0.6$.

28. **(2)** Which of the following is **<u>true</u>**?

    (a) A model selected using the adjusted R^2 backwards selection approach will only contain explanatory variables that are significant at the 5% level.

    (b) Adjusted R^2 may or may not be smaller than R-squared, depending on the sample size and the number of predictors in the model.

    (c) A parsimonious model is the model containing the highest possible number of predictors.

    (d) ***We cannot use R^2 as a criteria for model selection because it always increases when an additional variable is added to the model.***

29. **(2)** The *trimmed mean* is a statistical measure of central tendency, much like the mean and median. It involves the calculation of the mean after discarding given parts of the sample at the high and low ends. For example, the 5% trimmed mean is calculated after discarding the bottom and top 2.5% of the sample data. Based on this information, how does the robustness of the trimmed mean to outliers and extreme skew compare to that of the (regular) mean?

(a) ***trimmed mean is more robust than the mean***

(b) trimmed mean is less robust than the mean

(c) trimmed mean is equally as robust as the mean

(d) cannot tell from the information given

30. **(2)** Does meditation cure insomnia? Researchers randomly divided 400 people into two equal-sized groups. One group meditated daily for 30 minutes, the other group attended a 2-hour information session on insomnia. At the beginning of the study, the average difference between the number of minutes slept between the two groups was about 0. After the study, the average difference was about 32 minutes, and the meditation group had a higher average number of minutes slept. To test whether an average difference of 32 minutes could be attributed to chance, a statistics student decided to conduct a randomization test. She wrote the number of minutes slept by each subject in the study on an index card. She shuffled the cards together very well, and then dealt them into two equal-sized groups. Which of the following **best** describes the outcome?

(a) If meditation is effective, the average difference between the two stacks of cards will be more than 32 minutes.

(b) The average difference between the two stacks of cards will be about 32 minutes.

(c) ***The average difference between the two stacks of cards will be about 0 minutes.***

31. **(2)** Which of the following is **false**?

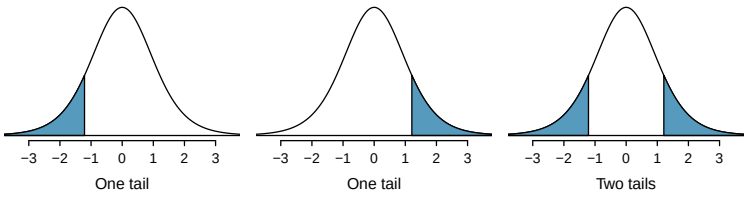
(a) When doing prediction based on a regression model, confidence intervals for a new observation are always narrower than prediction intervals for the same new observation.

(b) All else held constant, a model based on a higher sample size yields a narrower prediction interval.

(c) ***All else held constant, a model with a lower residual standard error yields a wider prediction interval.***

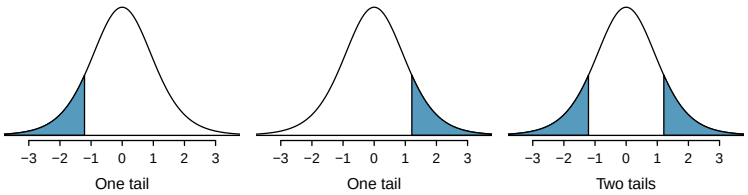
(d) Prediction intervals for new observations closer to the center of the observed data are narrower than those for new observations closer to the bounds of the observed data.

t distribution probability table



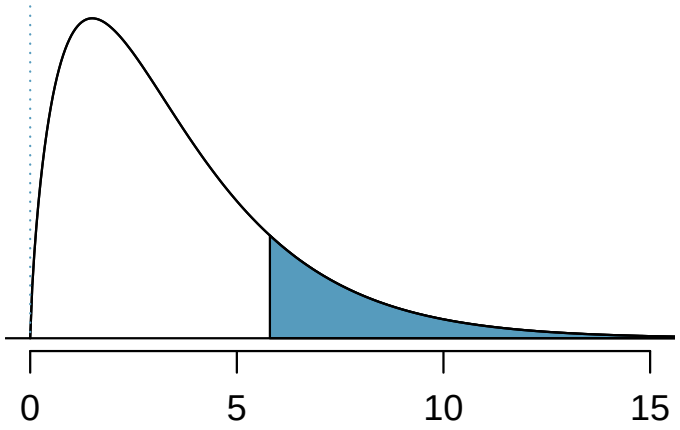
one tail		0.100	0.050	0.025	0.010	0.005
two tails		0.200	0.100	0.050	0.020	0.010
df	1	3.08	6.31	12.71	31.82	63.66
	2	1.89	2.92	4.30	6.96	9.92
	3	1.64	2.35	3.18	4.54	5.84
	4	1.53	2.13	2.78	3.75	4.60
	5	1.48	2.02	2.57	3.36	4.03
	6	1.44	1.94	2.45	3.14	3.71
	7	1.41	1.89	2.36	3.00	3.50
	8	1.40	1.86	2.31	2.90	3.36
	9	1.38	1.83	2.26	2.82	3.25
	10	1.37	1.81	2.23	2.76	3.17
	11	1.36	1.80	2.20	2.72	3.11
	12	1.36	1.78	2.18	2.68	3.05
	13	1.35	1.77	2.16	2.65	3.01
	14	1.35	1.76	2.14	2.62	2.98
	15	1.34	1.75	2.13	2.60	2.95
	16	1.34	1.75	2.12	2.58	2.92
	17	1.33	1.74	2.11	2.57	2.90
	18	1.33	1.73	2.10	2.55	2.88
	19	1.33	1.73	2.09	2.54	2.86
	20	1.33	1.72	2.09	2.53	2.85
	21	1.32	1.72	2.08	2.52	2.83
	22	1.32	1.72	2.07	2.51	2.82
	23	1.32	1.71	2.07	2.50	2.81
	24	1.32	1.71	2.06	2.49	2.80
	25	1.32	1.71	2.06	2.49	2.79
	26	1.31	1.71	2.06	2.48	2.78
	27	1.31	1.70	2.05	2.47	2.77
	28	1.31	1.70	2.05	2.47	2.76
	29	1.31	1.70	2.05	2.46	2.76
	30	1.31	1.70	2.04	2.46	2.75

t distribution probability table

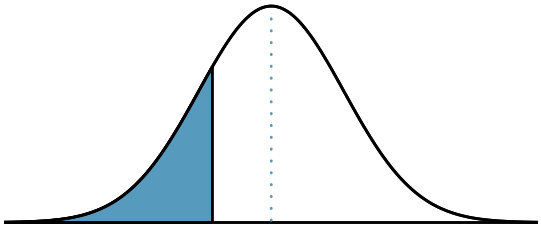


one tail		0.100	0.050	0.025	0.010	0.005
two tails		0.200	0.100	0.050	0.020	0.010
df	31	1.31	1.70	2.04	2.45	2.74
	32	1.31	1.69	2.04	2.45	2.74
	33	1.31	1.69	2.03	2.44	2.73
	34	1.31	1.69	2.03	2.44	2.73
	35	1.31	1.69	2.03	2.44	2.72
	36	1.31	1.69	2.03	2.43	2.72
	37	1.30	1.69	2.03	2.43	2.72
	38	1.30	1.69	2.02	2.43	2.71
	39	1.30	1.68	2.02	2.43	2.71
	40	1.30	1.68	2.02	2.42	2.70
	41	1.30	1.68	2.02	2.42	2.70
	42	1.30	1.68	2.02	2.42	2.70
	43	1.30	1.68	2.02	2.42	2.70
	44	1.30	1.68	2.02	2.41	2.69
	45	1.30	1.68	2.01	2.41	2.69
	46	1.30	1.68	2.01	2.41	2.69
	47	1.30	1.68	2.01	2.41	2.68
	48	1.30	1.68	2.01	2.41	2.68
	49	1.30	1.68	2.01	2.40	2.68
	50	1.30	1.68	2.01	2.40	2.68
	60	1.30	1.67	2.00	2.39	2.66
	70	1.29	1.67	1.99	2.38	2.65
	80	1.29	1.66	1.99	2.37	2.64
	90	1.29	1.66	1.99	2.37	2.63
	100	1.29	1.66	1.98	2.36	2.63
	150	1.29	1.66	1.98	2.35	2.61
	200	1.29	1.65	1.97	2.35	2.60
	300	1.28	1.65	1.97	2.34	2.59
	400	1.28	1.65	1.97	2.34	2.59
	500	1.28	1.65	1.96	2.33	2.59
	∞	1.28	1.65	1.96	2.33	2.58

Chi-square probability table



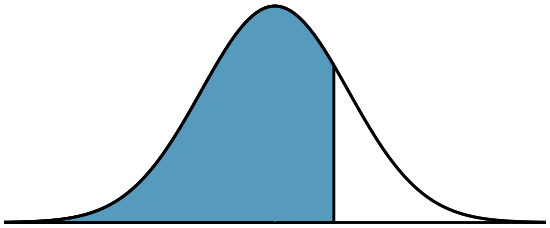
Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
	6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
	9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59
	11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26
	12	14.01	15.81	18.55	21.03	24.05	26.22	28.30	32.91
	13	15.12	16.98	19.81	22.36	25.47	27.69	29.82	34.53
	14	16.22	18.15	21.06	23.68	26.87	29.14	31.32	36.12
	15	17.32	19.31	22.31	25.00	28.26	30.58	32.80	37.70
	16	18.42	20.47	23.54	26.30	29.63	32.00	34.27	39.25
	17	19.51	21.61	24.77	27.59	31.00	33.41	35.72	40.79
	18	20.60	22.76	25.99	28.87	32.35	34.81	37.16	42.31
	19	21.69	23.90	27.20	30.14	33.69	36.19	38.58	43.82
	20	22.77	25.04	28.41	31.41	35.02	37.57	40.00	45.31
	25	28.17	30.68	34.38	37.65	41.57	44.31	46.93	52.62
	30	33.53	36.25	40.26	43.77	47.96	50.89	53.67	59.70
	40	44.16	47.27	51.81	55.76	60.44	63.69	66.77	73.40
	50	54.72	58.16	63.17	67.50	72.61	76.15	79.49	86.66



negative Z

Second decimal place of Z										
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Z
0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	−3.4
0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005	−3.3
0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	−3.2
0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010	−3.1
0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013	−3.0
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	−2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	−2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	−2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	−2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	−2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	−2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	−2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	−2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	−2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	−2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	−1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	−1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	−1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	−1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	−1.5
0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808	−1.4
0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968	−1.3
0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151	−1.2
0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357	−1.1
0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587	−1.0
0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841	−0.9
0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119	−0.8
0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420	−0.7
0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743	−0.6
0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085	−0.5
0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446	−0.4
0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821	−0.3
0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207	−0.2
0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602	−0.1
0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000	−0.0

*For $Z \leq -3.50$, the probability is less than or equal to 0.0002.



positive Z

	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

*For $Z \geq 3.50$, the probability is greater than or equal to 0.9998.