STOR 151, SECTION 1: SPRING 2024
Final Exam, May 3, 2024

YOUR NAME (PRINT).....

PID.....

Please sign the **pledge**: "On my honor, I have neither given nor received unauthorized aid on this examination."

Sign.....

This is an open book exam. Course text, personal notes and calculator are permitted. You have 3 hours to complete the exam. Answers should be written on the exam. Personal computers and cellphones are not to be used in answering the questions but you may use a personal device for accessing course notes or for the calculator feature. If you have any queries about the meaning of a question, ask the instructor for advice.

SHOW ALL WORKING — the space allowed for each question should be sufficient for this. Even correct answers will not get full credit if it's not clear how they were obtained. Incorrect answers will gain substantial credit if the method of working is substantially correct.

Answer six of the eight questions. If you attempt more than six, all the answers will be graded but only the best six (complete questions) will count. Each question is worth a total of 20 points and the whole exam is worth 120 points (which will be rescaled to a maximum of 40 for grading purposes). Points for each individual part of a question are also given in square brackets.

Questions may be answered in any order. The order of questions on the exam has itself been randomized.

1. A survey was taken of a sample of heterosexual couples and their eye colors identified as one of blue, brown or green, with the following results:

		Female partner			
		Blue	Brown	Green	Total
Male partner	Blue	78	23	13	
	Brown	19	23	12	
	Green	11	9	16	
	Total				

   (a) Fill in the blanks of the above table to show the row, column and overall totals (no need to show working: just write your answers directly in the table). **[3 points.]**

(b) Calculate the proportions of blue, brown and green eyes among (i) the male partners, and (ii) the female partners. Based on these numbers, would you say the overall distribution of eye colors in the population is the same for males and females? (No need for a formal hypothesis test, but give rough statistical arguments to support your conclusion.) **[5 points.]**

(c) A more challenging hypothesis is whether men and women are more likely to be attracted to partners with the same eye color as theirs. Formulate this precisely as a statement about independence in a two-way table. Then, compute the χ^2 statistic and say whether the hypothesis is formally rejected at a significance level of 0.01. Based on this result, do you agree that "same colors attract"? **[12 points.]**

2. Burt is an enterprising businessman who moves to Chapel Hill and quickly hears a lot of his neighbors complaining about having their cars towed. He sees a business opportunity in this and offers the following insurance policy: for a premium of \$150 a year, Burt's company will pay all your towing fees up to \$500 for a single tow, or \$1,000 if you are towed more than once. He also estimates that in a given year, one driver in 8 gets towed exactly once and 1 driver in 15 gets towed more than once. For the purpose of this question, assume that these are the only possible payouts — 0 if the driver has no tows, \$500 if one tow, \$1,000 if more than one tow.

(a) Let X be Burt's profit in one year from a single customer (i.e. \$150 minus any payout). What are the mean and standard deviation of X? **[8 points.]**

(b) In his first year Burt gets 350 customers. What is the probability that Burt makes a profit at the end of the year? **[6 points.]**

(c) Assuming the terms of the policy remain unchanged, how many customers would he need to ensure that the probability of making a profit in a given year is at least 0.95? **[6 points.]**

3. The Tar Heel 10 is a 10-mile race held in Chapel Hill each April. A random sample of 100 runners was taken from all the finishers of the 2023 race, with the following results: mean 97.95 minutes, standard deviation 15.66 minutes. A second, independent, random sample of 100 runners was taken from all the finishers of the 2024 race, with mean 100.53 minutes, standard deviation 16.35 minutes. We also computed the mean and standard deviations of the *differences* between the times of the 100 runners selected from the 2023 race and those of the 100 runners selected from the 2024 race: these were $-2.58 (= 97.95 - 100.53)$ minutes and 20.69 minutes.

   (a) We would like to form estimates and confidence intervals for the differences in overall mean finishing times for the two races. Would this be an example of "Paired data" or "Difference of two means"? Briefly explain your reasoning. **[5 points.]**

(b) Based on the data given here, construct a 95% confidence interval for the difference in mean finishing times between the 2003 and 2004 races. State any assumptions you make. **[7 points.]**

(c) Based on the data given here, construct a hypothesis test whether the mean finishing times of the two races are the same. Be careful to state your null and alternative hypotheses, as well as any other assumptions you make. Assume a significance level of $\alpha = 0.05$. **[8 points.]**

4. According to a recent survey, the mean and standard deviation of total cholesterol level in US adults were 203.6 mg/dl and 40.7 mg/dl, respectively. A new drug is proposed whose manufacturers claim reduces total cholesterol by a mean of 25 mg/dl per individual. To test this claim, a group of researchers set up a clinical trial where n participants are randomly assigned to the new drug and another n to a placebo. Assume the standard deviation remains at 40.7 and the researchers use a two-sided significance level of $\alpha = 0.01$ to test the null hypothesis that the means for the new drug and placebo samples are the same.

   (a) Suppose $n = 50$. What is the power of this test when the true effect size (reduction in total cholesterol) is 25 mg/dl? **[8 points.]**

(b) How large should n be to achieve a power of 80%? Assume all other conditions remain the same. **[12 points.]**

5. A group of students agree to participate in a survey about attitudes to immigration. Specifically, the students self-identify as Republicans (30%), Democrats (50%) or Independents (20%), and they are asked whether they support the immigration policy of one of the presidential candidates. Among the Republicans, 80% support the policy and 20% do not. Among the Democrats, 10% support the policy and 90% do not. Among the Independents, 40% support the policy and 60% do not. Now suppose we select one person random from all those who responded to the survey.

   (a) What is the probability that this person is a Democrat who supports the policy? **[4 points.]**

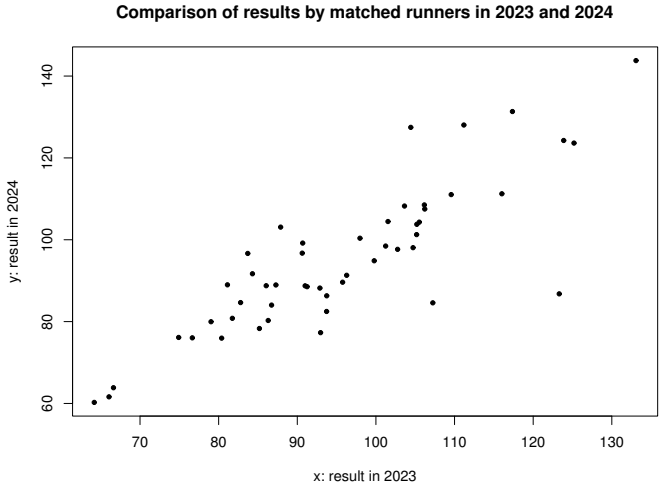
   (b) What is the probability that this person is an Independent who opposes the policy? **[4 points.]**

   (c) What is the unconditional probability that this person supports the policy? **[5 points.]**

   (d) Given that the selected person supports the policy, what is the probability that he/she is (i) a Republican, (ii) a Democrat, (iii) an Independent? **[7 points.]**

6. This is another question about the Tar Heel 10 but it is entirely separate from Question 3: it is not necessary to answer that question in order to tackle this one.

   I collected data for the finishing times of the Tar Heel 10 race in 2023 and 2024. I was successful in identifying 817 runners who ran the race both years. From those, I selected a sample of 50 runners and compared their results directly. In the following, x_i will denote the time in minutes of the ith runner in the sample in 2023, and y_i will denote the time in minutes of the same runner in 2024.



   (a) The figure shows a scatterplot of y_i against x_i for the sample of runners who ran both races. Based on this plot, which of the following do you think represents the correlation, to four decimal places? (No explanation is needed: just circle which one you think is the correct answer.)

   (i) 0.9347 (ii) 0.8445 (iii) 0.6002 (iv) –0.1141 **[5 points.]**

   (b) We would like to predict the value of y_i given x_i. To do this, it is convenient to have a prediction formula of the form

$$\hat{y} \quad = \quad b_0 + b_1 x$$

   You are given the following information: $\bar{x} = 95.618$, $\bar{y} = 94.961$, $s_x = 15.398$, $s_y = 17.802$. Using the correlation you selected in part (a), estimate b_0 and b_1. **[7 points.]**

(c) Jennie ran the 2023 Tar Heel 10 in a time of 80 minutes and 30 seconds (80.5 minutes). Using the regression equation, what is her predicted time for 2024? **[3 points.]**

(d) When this model is refitted using linear regression software, the standard error of the intercept is 8.65. We might naturally expect that the fitted straight line passes through the origin, i.e. that $\hat{y} = 0$ if $x = 0$. Are the data consistent with that conclusion? (Hint: the statement that the straight line passes through the origin is equivalent to $b_0 = 0$.) **[5 points.]**

7. A state has 10 universities, 25 four-year colleges and 50 community colleges, each of which offer multiple sections of an introductory statistics class each year. Researchers want to conduct a survey of students taking introductory statistics classes in the state.

   (a) Explain a method for collecting each of the following types of samples: (i) simple random sample, (ii) stratified sample, (iii) cluster sample, (iv) multistage sample. **[10 points.]**

(b) *Briefly* describe the advantages and disadvantages of each of the methods in (a). **[10 points.]**

8. A physician claims to be able to guess the sex of a baby as soon as the parents receive a positive pregnancy test. To test this claim, he is asked to guess the sex (male or female) of 30 babies. He is right in 19 of the 30 cases.

   (a) Based on this information, construct a 95% confidence interval for p, the proportion of times the physician correctly guesses the sex of the baby. **[8 points.]**

   (b) Does this study provide any evidence that the physician in fact has an ability beyond random guesswork to determine the sex of a baby based solely on a positive pregnancy test? Set this up as a formal hypothesis test, compute the p-value, and summarize your conclusions. **[6 points.]**

   (c) Suppose we want to conduct this experiment so that the margin of error of a 95% confidence interval for p is no more than 0.08. How large a sample size do we need? **[6 points.]**

SOLUTIONS AND COMMENTS

1. (a) See expanded table:

		Female partner			
		Blue	Brown	Green	Total
Male partner	Blue	78	23	13	114
	Brown	19	23	12	54
	Green	11	9	16	36
	Total	108	55	41	204

(b) The proportions of blue, brown and green among male partners are 0.558, 0.265, 0.176 and among female partners are 0.529, 0.270, 0.201 to three decimal places. A few calculations confirm that the differences are not statistically significant; for example, in the case of blue eyes (the biggest male-female discrepancy) the standard error of the difference is $\sqrt{\frac{0.558\times0.558}{204}+\frac{0.529\times0.529}{204}}=0.049$ which is bigger than the difference in proportions (0.029). (c) Here we proceed more formally by a χ^2 test. The null hypothesis is that the eye colors for the male and female partners are independent, and the alternative hypothesis is that they are not independent. We compute the expected counts under the assumption of independence. For example, the expected number of (Blue,Blue) combinations is $\frac{114\times108}{204}=60.35$. Calculating expected values for every cell in the table leads to

60.35	30.74	22.91
28.59	14.56	10.85
19.06	9.71	7.24

and $T=\sum\frac{(Obs-Exp)^2}{Exp}=33.7$. The *df* is $(3-1)\times(3-1)=4$ and the 0.001 upper tail value of the χ^2_4 distribution is 18.47. Therefore, we reject the hypothesis of independence with a p-value <0.001 (the actual p-value for $T=33.7$ is about 8×10^{-7} so it's a very clear-cut reject). Interpretation: all the diagonal entries of the "Observed" table are greater than the corresponding entries of the "Expected" table, and with one exception, the opposite is true for all the off-diagonal entries. Thus, the data support the hypothesis that men and women with the same eye color are more likely to be attracted to each other.

2. (a) X takes the values $x_1=-350$ with probability $p_1=\frac{1}{8}$, $x_2=-850$ with probability $p_2=\frac{1}{15}$, $x_3=+150$ with probability $p_3=1-\frac{1}{8}-\frac{1}{15}=\frac{97}{120}$. The mean of X is $\mu=x_1p_1+x_2p_2+x_3p_3=-43.75-56.6667+121.25000=20.8333$ and the variance is $\sigma^2\sum_{i=1}^3(x_i-\mu)^2p_i=17189.67+50556.71+13486.26=81232.64$ which leads to $\sigma=\sqrt{81232.64}$ which is about \$285 (these are nearly exact calculations; any reasonable numerical approximation will be accepted). (b) Taking the sum of n independent values of X, the net profit has mean $n\mu$ and standard deviation $\sqrt{n}\sigma$ which are about 7291.7 and 5332.1 when $n=350$. For a random variable Y with this mean and SD, the probability that $Y>0$ is the same as the probability that $z=\frac{Y-7281.7}{5332.1}>-\frac{7281.7}{5332.1}=-1.37$ to two decimal places, where z is standard normal. This is based on the fact that Y itself has an approximately normal distribution by the Central Limit Theorem. Using the normal table, this probability is $1-0.0853=0.9147$. (c) To make the one-sided probability up to 0.95, we replace $-\frac{7281.7}{5332.1}$ by $-\frac{\mu n}{\sigma\sqrt{n}}$ which we equate to –1.645 to get a left-tail probability of 0.05. Solving for $\sqrt{n}=\frac{1.645\times285}{20.8333}=22.5$ we deduce $n\approx22.5^2$ or 507 rounding to the next whole number.

3. (a) The two samples are independent so it's a two-sample test, not a paired test. (b) $\bar{y}-\bar{x}=100.53-97.95=2.58$ with a standard error $\sqrt{\frac{15.66^2}{100}+\frac{16.35^2}{100}}=2.2640$ and the corresponding

t^* value for a 95% confidence interval based on df=99 is 1.98 (table, page 415; we determined df by the $\min(n_x, n_y) - 1$ formula which gives 99, but we actually looked up df=100 because this is the nearest entry in the table. The exact df here is not too important but since the standard errors are estimated from the data, this is a situation where we should be using the t distribution rather than the normal distribution.) Therefore, the 95% confidence interval for the difference of means is $2.58 \pm 1.98 \times 2.264 = (-1.90, 7.06)$. (c) The pooled standard error for $\bar{y} - \bar{x}$ is $\sqrt{\frac{15.66^2}{100} + \frac{16.35^2}{100}} = 2.2641$ to 4 decimal places. For a test that the two means are equal, we compute the t statistic $t = \frac{100.53 - 97.95}{2.2641} = 1.14$ which does not appear to be statistically significant: again referring to the t distribution table with df=100, the rejection points for two-sided significance levels of 0.2 and 0.05 respectively are 1.29 and 1.98, so in each case, we *do not reject* the null hypothesis. Practical comment on this conclusion: in fact the actual mean race times (99.6 for 2023, 100.0 for 2024) were even closer than this sample suggests, but it was a genuine random sample and initially created the impression that there was a difference in the mean running times of the two races. However, when you compute the standard error by either method it becomes clear that the standard error is only slightly smaller than the observed difference of mean running times, so in practice, the conclusion is that there was no significant difference in the mean running times.

[Two comments. First, quite a few students got the answer to (a) right but still used the wrong SE in (b). Second, in class I made a bit of a fuss about computing a "pooled standard deviation" (which in this case comes to 16.01) before computing the standard error of $\bar{y} - \bar{x}$, but because the two sample sizes are equal, it actually makes no difference which way you do it.]

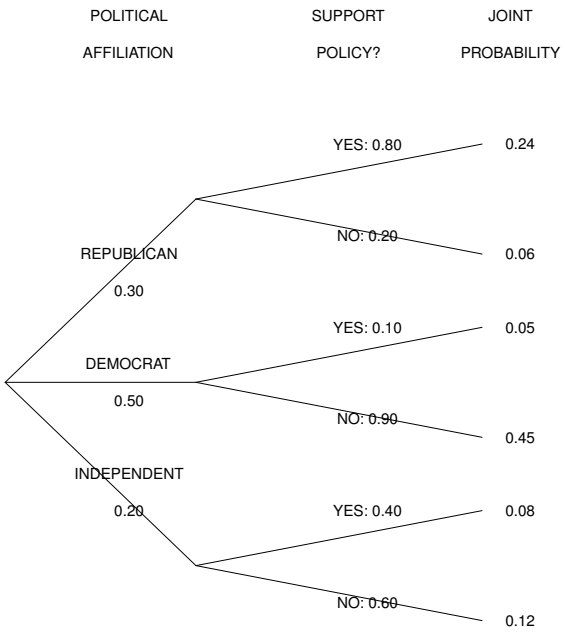
4. Let $\bar{x}$ and $\bar{y}$ be the means of the treatment and control groups respectively. The test will reject H_0 when $|\bar{x} - \bar{y}| > 2.58 \times 40.7 \times \sqrt{\frac{2}{n}} = 2.58 \times 8.14 = 21.0$ when $n = 50$. If the true mean of $\bar{x} - \bar{y}$ is –25, then the probability that this test rejects H_0 in the right direction is $\Pr\{\bar{x} - \bar{y} < -21.0\} = \Pr\left\{\frac{\bar{x} - \bar{y} + 25}{8.14} < \frac{4}{8.14}\right\} = \Pr\{z < 0.49\} = 0.6879$ or 0.69 to two decimal places, where z is a standard normal random variable. If we want to get this power up to 0.8, in the notation adopted in class, we set $z^* = 2.58$, $z^\dagger = 0.84$ corresponding to $\alpha = 0.01$, power 0.8 respectively, so $z^* + z^\dagger = 3.42$. The formula then becomes $25 = 3.42 \times SE = 3.42 \times 40.7 \times \sqrt{\frac{2}{n}}$ which solves to $n = 62$.

5. See tree diagram (next page). (a) 0.05 (b) 0.12 (c) 0.24+0.08+0.05=0.37 (d) $\frac{(0.24, 0.05, 0.08)}{0.37} = (0.649, 0.135, 0.216)$

6. (a) (ii). (b) $b_1 = \frac{0.8445 \times 17.802}{15.398} = 0.976$ and $b_0 = 94.961 - 0.976 \times 95.618 = 1.605$. (c) $1.605 + 0.976 \times 80.5 = 80.173$, or about 80 minutes, 10 seconds. [You may get slightly different answers because of rounding error.] d) To make a distinction between population values and their estimates, we rewrite the regression equation as $\hat{y} = \beta_0 + \beta_1 x$ where β_0, β_1 are the population values and b_0, b_1 are their estimates. We are formally testing $H_0 : \beta_0 = 0$ against the alternative $H_A : \beta_0 \neq 0$. Given that the standard error of b_0 is 8.65, the t statistic has the value $\frac{1.638}{8.65} = 0.19$. We could state without any further analysis that this result is clearly not significant, but a more formal analysis would note that the critical t^* corresponding to a two-tailed probability of 0.2 is 1.30 (from the table on page 415 of the text with df=48, i.e. $n - 2$ in this instance). The observed value is $0.19 < 1.30$ so the two-sided p-value is definitely

greater than 0.2, which is not a statistically significant result. Practical interpretation: there is no evidence against the stated hypothesis (that the straight line goes through the origin).

[Added comment. Quite a few students guessed wrong in (a) but I still gave credit for (b) and (c) if you got the rest of the calculations right. The solutions to (c) under each of the answers (i)–(iv) in (a) are 78.6, 80.2, 84.5, 97.0 all expressed in minutes to the nearest 0.1.]



Tree diagram for question 5.

7. (a) (i) SRS: get a list of all students who took a statistics class in any of the institutions, then use a random number generator to select a random subset of that list. (ii) stratified sample: same, but form separate SRSs for the universities, the four-year colleges and the community colleges. (iii) cluster sample: select a random subset of the universities, and a random subset of the four-year colleges, and a random subset of the community colleges, then survey all students who took a statistics class in those institutions. (iv) multi-stage sample: start the same way as in (iii), but within each selected institution, select a random sample of all students who took a statistics class in that institution. Some variants on this process are permitted. For instance, in (iii), you could just lump all 85 institutions together and select a random sample of those — that would be a valid definition of a cluster sample but would seem less natural in this context, where presumably you are interested in learning something about the differences among the three types of institutions. (b) (i) is the purest form of random sample, but it would be extremely tedious to have to compile a list of all students who took a statistics class in all the institutions. (ii) would be better for differentiating the three types of institution, but suffers from the same disadvantage that it's tedious to implement. (iii) is much quicker to identify the sample, but it may be difficult to contact all the students in the selected institutions, so that could also be tedious to implement. (iv) is similar to (iii)

but because we are not trying to get responses from all the students, should be simpler and quicker to implement. Probably (iv) is best for providing a statistically valid sample without excessive administrative costs. Again, there are other valid points you could make which would receive credit.

8. (a) $\hat{p} = \frac{19}{30} = 0.6333$. The 95% confidence interval is $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/30} = (0.461, 0.806)$. Note that this confidence interval includes 0.5 (the value at which we would conclude he was just guessing). *Conditions for the central limiting theorem*: so long as $\frac{1}{3} \leq p \leq \frac{2}{3}$ (where p is the true probability of a correct guess) and with $n = 30$, we have $np \geq 10$ and $n(1-p) \geq 10$ so the conditions for the central limit theorem are satisfied, justifying a confidence interval based on the normal distribution. (b) Test $H_0 : \ p = \frac{1}{2}$ against $H_A : \ p \neq \frac{1}{2}$. First compute the standard error as $SE = \sqrt{\frac{p_0(1-p_0)}{30}} = 0.0913$ when $p_0 = \frac{1}{2}$. Then $z = \frac{\hat{p}-p_0}{SE} = 1.46$, for which the two-sided p-value is $2 \times 0.0721 = 0.1442$. Since this is > 0.05, we *do not reject* the null hypothesis and conclude that the result is not significant. In plain English, there is no evidence that this physician is able to detect the sex of a baby better than random guesswork. [Note: it's not clear-cut whether you should use a one-sided or two-sided test here since presumably you're not interested in the possibility that the physician might guess right less than half the time. The conclusion using a one-sided test is the same so I also gave credit.] (c) The ME based on a sample of size n is bounded by $1.96 \times \sqrt{\frac{0.5 \times 0.5}{n}}$ so we set this equal to 0.08 to get $n = \left(\frac{1.96}{0.08}\right)^2 /4 = 150.06$ — should round up to the next whole number which is 151.