Math 326 Section 01
Professor Carty
110 points possible

Final Exam
May 9, 2022

Name: Matthew Wilder

1. No hats or dark sunglasses. All hats are to be removed.
2. All book bags are to be closed and placed in a way that makes them inaccessible. Do not reach into your bag for anything during the exam. If you need extra pencils, pull them out now.
3. You are allowed a one-page / two-sided handwritten sheet of formulas. Notes, theorems, lemmas, etc. are not allowed on the formula sheet. Please put your name on your formula sheet and hand it in with the exam.
4. Be sure to print your proper name clearly.
5. Calculators are optional, but if you choose to use one you may ONLY use the battery or solarpowered Texas Instruments BA35 model calculator, the BA II Plus, the BA II Plus Professional, the TI30Xa or TI30X II (IIS solar or IIB battery), or TI-30X MultiView (XS Solar or XB Battery).
6. All electronic devices, including cell phones and other wearable devices, must be silenced or powered off and stored out of sight for the entirety of the exam.
7. If you have a question, raise your hand and I will come to you. Once you stand up, you are done with the exam. If you have to use the facilities, do so now. You will not be permitted to leave the room and return during the exam.
8. Every exam is worth a total of **110 points**. Including the cover sheet, each exam has 12 pages.
9. At 11:00, you will be instructed to put down your writing utensil. Anyone seen continuing to write after this announcement will have their exam marked and lose all points on the page they are writing on. At this time, you will use your phone to take PDFs of the pages and upload them into Canvas as a single PDF document (the same way you do with homework.) Solutions must be uploaded by 11:10.
10. If you finish early, quietly and respectfully perform the preceding tasks. You may leave early.
11. You will hand in the paper copy of the exam on your way out of the classroom.
12. You have hundred-twenty minutes to complete the exam. I hope you do well.

one?

1. (10 points) Five recent rainfalls at Vinegar Lake, NY, have been found to have the following pH values:

$$4.2, 4.5, 5.1, 3.9, \text{and } 4.3.$$

(a) Find a 95% confidence interval for μ, the mean pH level in all the rainfalls at Vinegar Lake.

$\hat{\mu} = 4.4$

$s^2 = \frac{1}{4} \sum (\mu_i - \bar{\mu})^2$

$s^2 = 0.2$

$s \approx 0.4472$

$\frac{s}{\sqrt{n}} = \frac{0.4472}{\sqrt{5}} = 0.2$

$t_{0.025}(4) = 2.776$

$2.776 \cdot \frac{s}{\sqrt{n}} \approx 0.5552$

$\boxed{4.4 \pm 0.56}$

$(3.8448, \; 4.9552)$

(b) Find a 95% confidence interval for σ, the true standard deviation of pH of rainfalls at Vinegar Lake.

$\alpha = 0.05$

$$\frac{4 \cdot 0.2}{\underset{\substack{\uparrow \\ 11.1433}}{\chi^2_{\alpha/2}(n-1)}}, \quad \frac{4 \cdot 0.2}{\underset{\substack{\uparrow \\ 0.484419}}{\chi^2_{1-\alpha/2}(n-1)}}$$

$\boxed{(0.0718, \; 1.65146)}$

2. (20 points) Suppose that we have a random sample $Y_1, Y_2, \ldots, Y_n$ from a distribution with the density function

$$f(y) = 2y/\theta^2 \text{ on support } 0 \le y \le \theta$$

(a) Show that $\hat{\theta} = 2\bar{Y}$ is a biased estimator for θ.

$$E[\hat{\theta}] = E[2\bar{Y}] = 2E[\bar{Y}] = \frac{4}{3}\theta$$

$$\text{Bias}(\hat{\theta}) = \frac{4}{3}\theta - \frac{2}{3}\theta = \frac{2}{3}\theta \neq 0 \quad \therefore \text{ biased}$$

$$E[Y] = \int_0^\theta 2y^2/\theta^2 \, dy = \frac{2}{\theta^2}\int_0^\theta y^2 dy = \frac{2}{\theta^2} \cdot \frac{y^3}{3}\Big|_0^\theta = \frac{2}{\theta^2} \cdot \left[\frac{\theta^3}{3} - 0\right]$$

$$= \frac{2}{3}\theta$$

(b) Determine the mean square error of $\hat{\theta}$.

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias(\hat{\theta})^2$$

$$= var(\hat{\theta}) + \frac{4}{9}\theta^2$$

$Var(2\bar{Y})$

$$Var(2\bar{Y}) = 4Var(\bar{Y})$$

$$= 4\left[E[X^2] - E^2[X]\right]$$

$$= 4\left[\frac{\theta^2}{2} - \frac{4}{9}\theta^2\right]$$

$$2\theta^2 - \frac{16}{9}\theta^2$$

$18/9$

$$\frac{2}{9}\theta^2$$

$$E[Y^2] = \frac{2}{\theta^2} \cdot \frac{y^4}{4}\Big|_0^\theta = \frac{\theta^2}{2}$$

$$MSE(\hat{\theta}) = \frac{2}{9}\theta^2 + \frac{4}{9}\theta^2 = \frac{6}{9}\theta^2$$

$$\boxed{= \frac{2}{3}\theta^2}$$

(c) Find the method of moments estimator for θ.

$$\mu_1' = m_1$$

$$\frac{2}{3}\theta = \frac{1}{n}\sum Y_i$$

$$\frac{2}{3}\theta = \bar{Y}$$

$$\boxed{\theta_{mom} = \frac{3}{2}\bar{Y}}$$

(d) Show that the method of moments estimator for θ is consistent.

need: $\lim_{n\to\infty} Var(\theta_{mom}) = 0$

$$Var(\theta_{mom}) = \frac{9}{4} Var(\bar{Y})$$

$$= \frac{9}{4} Var\left(\frac{1}{n}\sum Y_i\right)$$

$$= \frac{9}{4n^2} Var\left(\sum Y_i\right)$$

$$= \frac{9}{4n^2} \underbrace{\sum_{i=1}^{n} Var(Y_i)}_{\text{finite in (b)}}$$

$$\lim_{n\to\infty}\left(\frac{9}{4n^2}\sum_{i=1}^{n} Var(Y_i)\right) = 0$$

$\therefore$ consistent

3. (15 points) Consider an iid sample $Y_1, Y_2, \ldots, Y_n$ from the distribution with density function:

$$f(y) = \frac{2}{\sqrt{\pi}}\theta^{3/2}y^2 \exp[-\theta y^2], y > 0$$

and θ is an unknown positive parameter.

(a) State and simplify the likelihood function.

$$L(\theta) = \prod f(y_i \mid \theta) = \left(\frac{2\theta^{3/2}}{\sqrt{\pi}}\right)^n y^{2n} \exp\left[-\theta \sum y_i^2\right]$$

$$g(u,\theta) = \left(\frac{2\theta^{3/2}}{\sqrt{\pi}}\right)^n \exp[-\theta u]$$

$$u := \sum y_i^2$$

$$h(\vec{y}) = y^{2n}$$

(b) Determine a sufficient statistic for θ.

$$u := \sum y_i^2$$

(c) Find the maximum likelihood estimator for θ.

$$\frac{d\ln(g)}{} = 0$$

$$\ln(g) = \ln\left(\left(\frac{2\theta^{3/2}}{\sqrt{\pi}}\right)^n \exp[-\theta u]\right)$$

$$= n\ln\left(\frac{2\theta^{3/2}}{\sqrt{\pi}}\right) + \ln(\exp(-\theta u))$$

$$= n\ln(\quad) - \theta u$$

note:

$$n\ln(\quad)$$

$$= n\left[\ln\left(\frac{2}{\sqrt{\pi}}\right) + \ln(\theta^{3/2})\right]$$

$$= n\left[\_\_\_\_ + \frac{3}{2}\ln\theta\right]$$

$$\frac{d\ln(g)}{d\theta} = \frac{d}{d\theta}\left[\frac{3}{2}\ln\theta - \theta u\right]$$

$$\frac{3}{2}\cdot\frac{1}{\theta} - u = 0 \qquad \frac{3}{2\theta} = u$$

$$3 = 2\theta u$$

$$\boxed{\theta_{MLE} = \frac{3}{2u}}$$

4. (15 points) A new medicine "Numbeze" has been developed as a hangover cure. To test its effectiveness, 16 seniors participating in the Senior Bar Crawl were randomly divided into two groups of 8 each. One group took Numbeze and the other were administered a placebo (fake drug). Each bar crawler quantified their overall relief on a 0 to 10 scale an hour after taking the medicine upon waking up the day after. The results were as follows:

$n_1 = n_2 = 8$

$n = 16$

Y_1	Numbeze:	$\bar{y} = 5.2$	$S = 3.3$	$n = 8$
Y_2	control:	$\bar{y} = 3.0$	$S = 2.3$	$n = 8$

(a) Formulate and complete a statistical test of whether Numbeze relieves hangover aches better than the placebo. Use the table to find bounds on the resultant p-value.

Difference in means.

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{2.2}{S_p\sqrt{\frac{1}{4}}}$$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = \sqrt{\frac{7(3.3)^2 + 7(2.3)^2}{14}} = \sqrt{\frac{1}{2}\left((3.3)^2 + (2.3)^2\right)} \simeq 2.844$$

for $t(14)$

$$\Rightarrow t \approx 1.5471167$$

$$p = P(|T| > t) \in (0.050, 0.100)$$

(b) What do you conclude? What advice would you pass on the manufacturers of Numbeze?

We can conclude that the probability that there is a difference in effects of drug is between 90 and 95%.

I'd advise Numbeze that it should* be better, but to get a larger sample for $\alpha = 0.01$

(c) What underlying assumptions did you use in computing this hypothesis test?

The underlying distribution is nearly normal.

The underlying population μ and σ are the same.

Y_1 and Y_2 are independent

5. (10 points) Market researchers for Tony B Iced Tea want to test whether a majority of iced tea drinkers prefer the taste of their brand in comparison to a national brand. Let p denote the percentage of iced tea drinkers who prefer Tony B over the national brand. The researchers want to test $H_0 : p = 0.5$ versus $H_a : p > 0.5$. They visit the Illinois State Fair and solicit 100 people at random to take a taste test. The researchers decide the rejection region is $Y \geq 60$, where Y is the number of people out of 100 fairgoers who prefer Tony B.

(a) Find α, the probability of a Type I error.

$H_0: p = 0.5$

$H_a: p > 0.5$ $\quad n = 100$

$RR = \{Y \geq 60\}$

$\alpha = P\left(Z > \frac{0.6 - 0.5}{\sigma/\sqrt{n}}\right)$

$> \frac{0.1}{\sqrt{2}/10}$

$\sigma = \sqrt{\frac{q}{p^2}}$ ← binom var

$= \sqrt{\frac{0.5}{0.5^2}}$

$= \sqrt{\frac{1}{0.5}}$

$= \sqrt{2}$

$\alpha = P(Z > 0.7071)$

$\boxed{\approx 0.2389}$

(b) Find the power of the test if, in fact, 63% of the public prefers Tony B Iced Tea.

~~power(θ_0) = α~~

power(θ_a) = $1 - \beta(\alpha)$

$P\left(Z < \frac{0.6 - 0.63}{\sigma/\sqrt{n}}\right)$

$P(Z > 0.212)$

≈ 0.4168

$1 -$

$\boxed{= 0.5832}$

6. (25 points) Consider the following data on two independent variables x_1 and x_2 and one dependent variable y:

x_1	1	1	1	-1	-1	-1
x_2	-1	0	1	-1	0	1
y	4	3	1	7	5	2

We wish to fit the linear model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon.$$

(a) Determine the matrices $\mathbf{X}$, $\mathbf{X}^T$, $\mathbf{X}^T\mathbf{X}$, $\mathbf{X}^T\mathbf{Y}$, and $(\mathbf{X}^T\mathbf{X})^{-1}$.

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & 0 & 1 & -1 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 0 \\ 1 & -1 & 1 \end{bmatrix} = X \qquad Y = (4\ 3\ 1\ 7\ 5\ 2)^T$$

$$X^TX = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 4 \end{bmatrix} \qquad (X^TX)^{-1} = \begin{bmatrix} 1/6 & 0 & 0 \\ 0 & 1/6 & 0 \\ 0 & 0 & 1/4 \end{bmatrix}$$

$$X^TY = \begin{bmatrix} 22 \\ -6 \\ -8 \end{bmatrix}$$

(b) Find the model to the data.

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} 1/6 & & \\ & 1/6 & \\ & & 1/4 \end{bmatrix} \begin{bmatrix} 22 \\ -6 \\ -8 \end{bmatrix} = \begin{bmatrix} 11/3 \\ -1 \\ -2 \end{bmatrix}$$

$$\boxed{\hat{Y} = \frac{11}{3} - X_1 - 2X_2}$$

(c) Use matrices to calculate SSE and determine S.

$$SSE = \underbrace{Y^T Y}_{\Sigma y_i^2} - \hat{\beta}^T X^T Y$$

$$- \begin{bmatrix} \frac{11}{3} & -1 & -2 \end{bmatrix} \begin{bmatrix} 22 \\ -6 \\ -8 \end{bmatrix} = -\left(\frac{22 \cdot 11}{3} + 6 + 16\right)$$

$y_i^2 = 104$

$$\boxed{SSE = 1.\overline{3}}$$

$$S^2 = \frac{SSE}{n-3} = \frac{SSE}{3} = 0.\overline{4} \Rightarrow \boxed{S = \frac{2}{3}}$$

(d) Does the data present sufficient evidence to indicate that x_1 contributes information to the prediction of y? Use a 5% significance level.

$\alpha = 0.05$

$H_0: \beta_1 = 0$

$H_a: \beta_2 \neq 0$

$$t = \frac{\hat{\beta}_1 - 0}{\sqrt{Var(\hat{\beta}_1)}} = \frac{\hat{\beta}_1}{\frac{2}{3}\sqrt{C_{11}}} = \frac{-1}{\frac{2}{3}\sqrt{1/6}}$$

$$t = -3.67$$

RR: $|t| > t_{0.025}(3) = 3.182$

$|-3.67| > 3.182$

$\Longrightarrow$ Reject H_0

$\boxed{X_1 \text{ contributes useful information!}}$

(e) Find a 90% confidence interval for $E(Y)$ when $x_1 = 0$ and $x_2 = -1$.

$$E[Y] = \frac{11}{3} - 0 - 2(-1) \qquad a = (1, 0, -1)$$

$$\frac{11}{3} + 2 = \frac{11}{3} + \frac{6}{3} = \frac{17}{3}$$

$$t_{0.05}(3) = 2.353 \qquad s\sqrt{a^T(X^TX)^{-1}a}$$

$$\frac{2}{3}\sqrt{}$$

$$[1, 0, -1]\begin{bmatrix} 1/6 & & \\ & 1/6 & \\ & & 1/4 \end{bmatrix}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = [1 \ 0 \ -1]\begin{bmatrix} 1/6 \\ 0 \\ -1/4 \end{bmatrix} = \frac{1}{6} + \frac{1}{4}$$

$$= 0.41\overline{6}$$

$$2.353 \cdot \frac{2}{3}\sqrt{0.41\overline{6}}$$

$$\boxed{CI := \frac{17}{3} \pm 1.012569979}$$

$$CI = (4.654097, 6.67924)$$

7. (15 points) A study wanted to compare different supplement regimens to the time required to recover from the flu. The was double-blind in design. That is, the physicians evaluating recovery did not know what treatment an individual was receiving nor did the individuals themselves. The time to recovery and the treatments appear in the table below:

μ_1 placebo	μ_2 vitamin C	μ_3 vitamin Z
4	7	3
11	6	2
7	5	5

\hline

(a) This data set is small enough to do by hand. Determine $\hat{\mu}_i$'s, $\hat{\mu}_0$, SSA and SSW.

$$\hat{\mu}_0 = \frac{50}{9} = 5.\bar{5}$$
$$\hat{\mu}_1 = \frac{22}{3} = 7.\bar{3}$$
$$\hat{\mu}_2 = \frac{18}{3} = 6$$
$$\hat{\mu}_3 = \frac{10}{3} = 3.\overline{33}$$

$$SSA = \sum n_i (\hat{\mu}_i - \hat{\mu}_0)^2 = 3 \sum (\mu_i - \mu_0)^2 = 3\left[(7.\bar{3}-5.\bar{5})^2 + (6-5.\bar{5})^2 + (3.\bar{33}-5.\bar{5})^2\right]$$
$$\boxed{\approx 24.93} \text{ SSA}$$

$$SSW = \sum\sum (Y_{ij} - \mu_i)^2 = (4-7.33)^2 + (11-7.33)^2 + (7-7.33)^2$$
$$+ \underbrace{(7-6)^2}_{1} + \underbrace{(6-6)^2}_{0} + \underbrace{(5-6)^2}_{1}$$
$$+ (3-3.33)^2 + (2-3.33)^2 + (5-3.33)^2 = \boxed{31.\bar{3}} \text{ SSW}$$

(b) Fill out the ANOVA table for this experiment and use the tables to find bounds on the p-value.

		df	MS	Fobs	P
among	24.93	2	12.465	2.3869	$P(F(8,6) > 2.3869)$
within	$31.\bar{3}$	6	5.222		
total	$56.26\bar{3}$	8			

$F(8,6)_{0.1} = 2.98 = \alpha$

$\boxed{p > 0.1}$ (and ≤ 1)

(c) Have we provided sufficient evidence that some treatments provide for quicker recovery? Briefly explain.

No because p is above 0.1 no $\alpha \leq 0.1$ will reject the null hypothesis.

Aside because i don't believe b/c: $(Y_2 - Y_3)$ CI $= 2.\bar{6} \pm t_{0.05}(2)\left(\sqrt{\frac{S_1^2}{3} + \frac{S_2^2}{3}}\right)$ $\quad t_{0.05}(2) = 2.92$

$\sqrt{\frac{2}{3} + \frac{4.\bar{6}}{3}}$

2.6 ± 4.353

wow...

And $Y_1 - Y_3 = 4 \pm 8.7057$

But what does this mean? All contain all μ_i's in CI

So its reasonable that theres no difference?