Name ____

Fall 2024 **STAT 5870-1/A** **Exam II**

(80 points)

Instructions:

- Write your name on the top, but do not open the exam.
- You are allowed to use one 8.5" x 11" page of notes (front and back) and a calculator.
- A total of 5 pages with a front and back.
- For full/partial credit, show all your work.
- Please turn in extra pages.

(page intentionally left blank)

1. The Verbio company has an anaerobic digester in Nevada, IA that produces ethanol from corn stover (the post-harvest corn plant material). To understand daily production of ethanol, engineers measure the amount of ethanol produced from each of 15 digesters on one day in July.

   From these 15 digesters, they observe a sample mean of 4.7 m^3 and a sample standard deviation of 1.1 m^3. For the following questions, let y_i be the gallons of ethanol produced from digester i with $i = 1, \ldots, 15$ and assume $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ with prior $p(\mu, \sigma^2) \propto 1/\sigma^2$.

   (a) Identify the 4 assumptions indicated by the model $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$. (4 points)

   (b) What is the posterior expectation for the **mean** amount of ethanol (including units) produced by a digester on that day? (2 points)

   (c) What is the posterior expectation for the **variance** in the amount of ethanol (including units) produced by a digester on that day? (4 points)

   (d) Calculate a 95% credible interval for the **mean** ethanol produced (including units). (5 points)

   (e) Calculate the posterior probability that the **mean** amount of ethanol produced is greater than 5 m^3. (5 points)

2. The Iowa Environmental Council provides information to the public about the number of Iowa beaches that have *E. coli* (a bacteria) advisories, i.e. measured *E. coli* are above the state standard. On 6 Sep 6 2024, 14 of the 83 beaches in Iowa had *E. coli* advisories. For these data, assume a binomial distribution for the number of beaches with an advisory.

   (a) What are the two main assumptions in a binomial model? (2 points)

   (b) What is the maximum likelihood estimate for the probability of a beach having an advisory? (2 points)

   (c) Calculate a two-sided p-value for the null hypothesis that the probability of a beach advisory is 0.1. (5 points)

   (d) Construct an approximate, equal-tail 95% confidence interval for the probability of a beach having an advisory. (5 points)

   (e) Is it reasonable to use the formula that you used in constructing the confidence interval? Why or why not? (3 points)

3. For this question, please refer to the **R Output** on page 7.

(a) Let $Y_{i,g}$ be the ith observation from the gth group. Using statistical notation, write the model being fit to these data. (5 points)

(b) How many total observations do we have for this data set? (2 points)

(c) What is the null hypothesis for the p-value? (2 points)

(d) What is the alternative hypothesis for the p-value? (2 points)

(e) What is the estimated difference in the group means? (3 points)

(f) Construct an approximate, equal-tail 68% confidence interval for the difference in group means. (6 points)

4. High school robotics competitions are fought with 1 alliance of 3 robots competing against another alliance of 3 robots. After qualification rounds, the top robotics teams choose other teams to join an alliance that lasts for the whole playoffs. At the most recent competition, the Ames High School Team had to choose between Team A and Team B based on which team could score notes in an amplifier. During qualification, Ames scouted these two teams and found that Team A scored on 3 of 4 attempts while Team B scored on 14 of 20 attempts.

   (a) State the posterior distribution for the probability of scoring for **both** teams. (4 points)

   (b) Calculate the Bayes estimator for the probability of scoring for **Team B**. (2 points)

   (c) Write R code to construct an equal-tail 95% credible interval for the probability of scoring for **Team B**. (4 points)

   (d) Write R code to calculate the posterior probability that the probability of scoring for **Team B** is greater than 0.75. (4 points)

   (e) Write R code to estimate the posterior probability that the probability of scoring is larger for Team B than for Team A. (6 points)

R Output

(feel free to remove but I do need this page)

```
##
##  Two Sample t-test
##
## data:  y by group
## t = 3.3416, df = 20, p-value = 0.003251
## alternative hypothesis: true difference in means between group A and group B is not equ
## 95 percent confidence interval:
##  0.6371463 2.7541797
## sample estimates:
## mean in group A mean in group B
##       1.1972642      -0.4983988
```

(scratch paper)

Table 1: Cumulative distribution function, $P(Z < z)$, for standard normal

z	0	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

Table 2: T-critical values, i.e. $P(T_{df} < t_p) = p$, for a standard T distribution with df degrees of freedom.

df	$t_{0.75}$	$t_{0.8}$	$t_{0.85}$	$t_{0.9}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581
Inf	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576