BST 140.652 Final Exam

Notes:

- You may use your one 8.5 by 11 formula sheet.
- Please use only the basic mathematical functions on your calculator.
- Show your work on all questions. Simple "yes" or "no" answers will be graded as if blank.
- Please be neat and write legibly. Use the back of the pages if necessary.
- Good luck!

_____

signature and **printed name**

1. Standard decks of playing cards have 13 cards of each suit (diamonds, clubs, hearts, spades) for a total of 52 cards. Someone gives you a 52 card deck. You draw a card, record its suit, replace the card, shuffle the deck and repeat that process 200 times, obtaining the following table:

$\diamondsuit$	$\clubsuit$	$\heartsuit$	$\spadesuit$
46	54	49	51

Does the distribution of suits appear to be standard? Perform the relevant hypothesis test, **stating your hypothesis defining any notation** and reporting a relevant P-value. Interpret your results.

2. A sample of 4 people resulted in the following contingency table comparing diabetes and a specific biomarker.

	Diabetes	
	Yes	No
Biomarker present	2	0
Biomarker absent	0	2

   a. Calculate the hypergeometric probability for every table satisfying the margins of this table.
   b. Calculate the Chi-squared statistic for every table satisfying the margins of this table.
   c. Report a P-value for an exact small sample test of independence between biomarker status and diabetes using the Chi-squared statistic as the test statistic.

3. A case-control study of esophageal cancer was performed. Daily alcohol consumption was ascertained ($80+$ gm = high, $0-79$ gm = low). The data was stratified by 3 age groups.

	Alcohol			Alcohol			Alcohol	
	H	L		H	L		H	L
case	8	5	case	25	21	case	50	61
control	52	164	control	29	138	control	27	208
	Age 35-44			Age 45-54			Age 55-64	

Assuming a constant odds ratio across age-strata, test to see if the odds ratio is 1. State your hypotheses **defining any notation that you use** report a relevant P-value, interpret your results.

4. The Biostatistics and Epidemilogy departments are running a 10K road race. There are three pairs of runners. In each case, a runner from Biostat was matched to a runner from Epi of the same age, gender and degree of running experience. For the pair with the largest difference in their finishing times, the winner was from Epi; for the pair with the second largest difference in their times, the winner was was from Biostat and for the final pair, the winner was from Epi.

   a. Give the exact small sample distribution of an appropriate statistic.

   b. State relevant hypotheses and perform the relevant test using your answer to part a.

5. Two computer based methods for diagnostic imaging are being studied. Ten images were processed with both methods, resulting in a $+$ or $-$ diagnosis for each method and image. The data are as follows

	Method A	
Method B	+	-
+	5	4
-	1	0

Does one method appear to have a different probability of giving a positive diagnosis than the other? State the relevant hypothesis **defining any notation that you use**, perform the test and interpret your results. Report a P-value.

6. Consider the following 2x2 table obtained from an experiment randomizing 200 people to a treatment and 200 to a placebo. Numbers of subjects with an improvement in symptoms were then recorded.

	Improvement	No improvement	Total
Treatment	41	159	200
Placebo	111	89	200

Let π_T be the population proportion of improved cases among the treated and π_P be the population proportion of improved cases for the placebo. Consider testing the hypothesis $H_0 : 2\pi_T = \pi_P$.

a. What are reasonable estimates of π_T and π_P under H_0?

b. Calculate the relevant expected counts.

c. Perform a Chi-squared test (hint the test has 1 df) and report a P-value.

Upper tail chi-square probabilities for 1 df.

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	1.000	0.752	0.655	0.584	0.527	0.480	0.439	0.403	0.371	0.343
1	0.317	0.294	0.273	0.254	0.237	0.221	0.206	0.192	0.180	0.168
2	0.157	0.147	0.138	0.129	0.121	0.114	0.107	0.100	0.094	0.089
3	0.083	0.078	0.074	0.069	0.065	0.061	0.058	0.054	0.051	0.048
4	0.046	0.043	0.040	0.038	0.036	0.034	0.032	0.030	0.028	0.027
5	0.025	0.024	0.023	0.021	0.020	0.019	0.018	0.017	0.016	0.015
6	0.014	0.014	0.013	0.012	0.011	0.011	0.010	0.010	0.009	0.009

Upper tail chi-square probabilities for 2 df.

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	1.000	0.951	0.905	0.861	0.819	0.779	0.741	0.705	0.670	0.638
1	0.607	0.577	0.549	0.522	0.497	0.472	0.449	0.427	0.407	0.387
2	0.368	0.350	0.333	0.317	0.301	0.287	0.273	0.259	0.247	0.235
3	0.223	0.212	0.202	0.192	0.183	0.174	0.165	0.157	0.150	0.142
4	0.135	0.129	0.122	0.116	0.111	0.105	0.100	0.095	0.091	0.086
5	0.082	0.078	0.074	0.071	0.067	0.064	0.061	0.058	0.055	0.052
6	0.050	0.047	0.045	0.043	0.041	0.039	0.037	0.035	0.033	0.032
7	0.030	0.029	0.027	0.026	0.025	0.024	0.022	0.021	0.020	0.019
8	0.018	0.017	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.012

Upper tail chi-square probabilities for 3 df.

	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
0	1.000	0.992	0.978	0.960	0.940	0.919	0.896	0.873	0.849	0.825
1	0.801	0.777	0.753	0.729	0.706	0.682	0.659	0.637	0.615	0.593
2	0.572	0.552	0.532	0.513	0.494	0.475	0.457	0.440	0.423	0.407
3	0.392	0.376	0.362	0.348	0.334	0.321	0.308	0.296	0.284	0.272
4	0.261	0.251	0.241	0.231	0.221	0.212	0.204	0.195	0.187	0.179
5	0.172	0.165	0.158	0.151	0.145	0.139	0.133	0.127	0.122	0.117
6	0.112	0.107	0.102	0.098	0.094	0.090	0.086	0.082	0.079	0.075
7	0.072	0.069	0.066	0.063	0.060	0.058	0.055	0.053	0.050	0.048
8	0.046	0.044	0.042	0.040	0.038	0.037	0.035	0.034	0.032	0.031
9	0.029	0.028	0.027	0.026	0.024	0.023	0.022	0.021	0.020	0.019
10	0.019	0.018	0.017	0.016	0.015	0.015	0.014	0.013	0.013	0.012
11	0.012	0.011	0.011	0.010	0.010	0.009	0.009	0.008	0.008	0.008