Public Policy 529
Midterm Exam

Student ID number (8-digits): _____

List of Formulas

$$\bar{y} = \frac{\sum y_i}{n} \qquad s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} \qquad Z = \frac{y - \mu_y}{\sigma}$$

$$IQR = Q_3 - Q_1$$

$$SS = \sum (y_i - \bar{y})^2 \qquad s = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} \qquad \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

Probability

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} \qquad P(A \text{ and } B) = P(A) \times P(B|A)$$

$$P(\sim A) = 1 - P(A)$$

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \qquad P(A \text{ and } B) = P(A) \times P(B)$$

Confidence Intervals and Significance Tests

$$Z \text{ or } t = \frac{\bar{y} - \mu_0}{\hat{\sigma}_{\bar{y}}} \qquad \hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

$$Z = \frac{\hat{\pi} - \pi_0}{\hat{\sigma}_{\pi_0}} \qquad \text{c.i.} = \bar{y} \pm t \cdot \hat{\sigma}_{\bar{y}}$$

$$\hat{\sigma}_{\bar{y}} = \frac{s}{\sqrt{n}} \qquad \text{c.i.} = \bar{y} \pm Z \cdot \hat{\sigma}_{\bar{y}}$$

$$\hat{\sigma}_{\pi_0} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} \qquad \text{c.i.} = \hat{\pi} \pm Z \cdot \hat{\sigma}_{\hat{\pi}}$$

1. Suppose that, in the population, the number of hours people spend each month listening to news has a normal distribution with a mean of 10 and a standard deviation of 4.

   (a) What percentage of people listen to 12 or more hours?

   (b) What percentage of people listen to 11 or fewer hours?

   (c) What percentage of people are in the range of 6 to 12 hours ?

2. According to the 2012 American National Election Study, 43% of respondents report that they performed volunteer work during the past 12 months. The remaining 57% say that they did not perform volunteer work. The sample size of this survey is 5,507. Construct a 95% confidence interval around the estimated proportion of Americans who did volunteer work *and* interpret what this confidence interval tells us about the "true" population proportion.

3. In your own words, explain why the Central Limit Theorem is so important for significance testing.

4. A survey asked respondents whether they favor or oppose limits on imports of goods from other countries. The survey also collected information on political party identification. The joint probability distribution of these two variables is presented in the table below.

Favor/Oppose	Democratic	Independent	Republican	Total
Favor	.24	.21	.15	.60
Oppose	.16	.14	.10	.40
Total	.40	.35	.25	1.00

(a) What is P(Favor and Independent)?

(b) What is P(Oppose | Independent)?

(c) Are party identification and opinions about limits on imports independent? Demonstrate mathematically.

5. Suppose $\alpha = .01$ and you have 50 degrees of freedom.

(a) What is the critical value of the t-statistic needed to reject H_0 in a two-sided test?

(b) What is the corresponding critical value of Z?

(c) If your t-statistic were 2.403, what would be the associated p-value?

6. Use the Stata output below to answer the following questions. The data are about countries.

```
. sum oilprod, detail

        oil production, thousands of barrels per day
-------------------------------------------------------------
      Percentiles      Smallest
 1%            0             0
 5%            0             0
10%            0             0       Obs                 207
25%            0             0       Sum of Wgt.         207

50%        .9791                     Mean           409.7014
                        Largest       Std. Dev.      1333.729
75%        96.27          4172
90%         1023          9056       Variance        1778833
95%         2472          9764       Skewness       5.309319
99%         9056         10120       Kurtosis       35.10556
```

(a) What is the measurement level of this variable? Which measures of central tendency and dispersion are appropriate?

(b) Find the interquartile range. Explain the result in sentence form.

(c) After inspecting the summary output, including the measures of central tendency and dispersion, how would you describe the distribution of this variable?

(d) Suppose you were going to perform a significance test in which the null hypothesis is that the population mean is 502.5. Choose an appropriate test statistic and write out the formula to calculate it, plugging in the appropriate numbers from the table above. You do not need to calculate the statistic.

7. A statistics professor has taught a particular course 20 times. The mean number of students is 25 with a standard deviation of 7.5. When asked about the "typical enrollment" in this course, the professor naturally answers in the form of a 90% confidence interval.

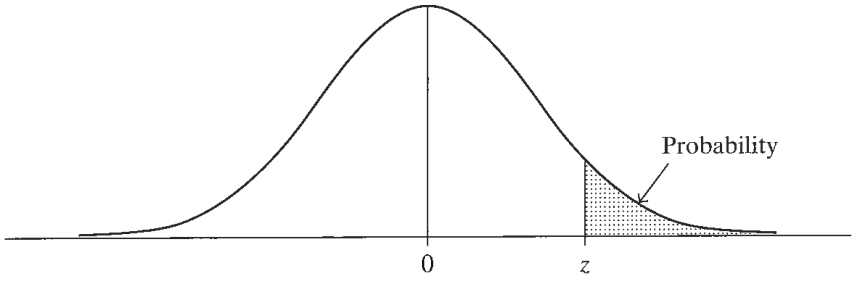
   (a) In thinking about this scenario, is the t-distribution or the Z-distribution the better approximation for the sampling distribution of the sample mean? Explain.

   (b) Find the 90% confidence interval for the mean.

8. A math teacher decides to experiment with a new approach. Her class has 16 students. At the end of the year, the mean standardized test score for her students was 328.4 and the standard deviation was 8. This compares to the baseline expected mean score of 324. Perform a test of statistical significance at $\alpha = .05$ on whether the new approach made a difference. How does the p-value compare to α?

Space for Work

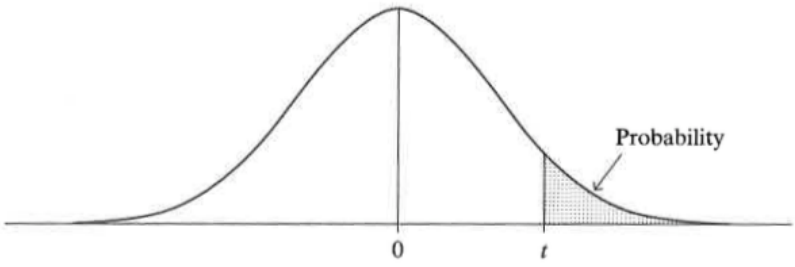
TABLE A: Normal curve tail probabilities. Standard normal probability in right-hand tail (for negative values of z, probabilities are found by symmetry)



	Second Decimal Place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000233									
4.0	.0000317									
4.5	.00000340									
5.0	.000000287									

Source: R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).

TABLE B: t Distribution Critical Values



	Confidence Level					
	80%	90%	95%	98%	99%	99.8%
	Right-Tail Probability					
df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.656	318.289
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.611
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.091

Source: "Table of Percentage Points of the t-Distribution." Computed by Maxine Merrington, Biometrika, 32 (1941): 300. Reproduced by permission of the Biometrika trustees.