Final for Statistics 301
Elementary Statistical Methods - Spring 2000
Material Covered: Chapters 1–11 of Workbook and text
For: 1st May

This is a 2 hour final, worth 28% and marked out of 28 points. The total possible points awarded for each question is given in square brackets at the beginning of each question. Anything that can fit on two sides of an $8\frac{1}{2}$ by 11 inch piece of paper may be used as a reference during this final. A calculator and appropriate statistical tables may also be used. No other aids are permitted.

Name (please print): ________________________ . **ID Number:** ____________.
last first

---

1. In a group of 6 disturbed males, the first episode of paranoia behavior occurred at the following ages.

22, 23, 23, 26, 26, 26

(a) [1] The histogram of this data is skewed (circle one) **left** / **right** and the mean is (circle one) **larger than** / **the same as** / **smaller than** the median.

(b) [1] The standard deviation, s, is (circle one) **larger than** / **the same as** / **smaller than** the range of the data.

(c) [1] In a group of disturbed females, the first episode of paranoia behavior occurred, on average, at age 29, where $s = 1.1$ years. The standard deviation in the age at which paranoid behavior started for the *combined* group of males and females is (circle one) **larger than** / **the same as** / **smaller than** 1.1 years.

---

2. In a random sample of 32 professors at PU/NC, 23 were found to have Type A personalities.

(a) [1] A 95% confidence interval of the percentage of professors with Type A personalities

is: ____________________________________.

(b) [1] If were interested in obtaining a confidence interval that is within 5% of the true (or actual, or population) percentage of professors at PU/NC who are Type A personalities, then we would have to take a random sample of size of at most $n \leq \frac{1}{4}\left(\frac{z_c}{E}\right)^2$ = (circle closest one) **135** / **189** / **283** / **335** / **384**.

3. Consider the data for artistic flair, y and level of insanity, x.

x	1	2	3	4	5
y	2	3	1	5	6

(a) [1] The regression equation for predicting artistic flair from level of insanity is given by (circle one)

(i) $y = x + 3$

(ii) $y = 3x + 0.4$

(iii) $y = 0.4x + 1$

(iv) $y = x + 0.4$

(v) $y = 0.4x + 3$

(b) [1] The predicted value of artistic flair for a level of insanity of $x = 3.5$ is (circle closest one) **2.4** / **4.4** / **3.9** / **6.5** / **10.9**

(c) [1] $S_e \approx$ (circle closest one) **0.56** / **1.32** / **1.55** / **2.31** / **3.27**

---

4. Suppose 2% of the 55,000 residences of Madeup town are criminally insane. What is the chance, in a random sample of 115 individuals from Madeup, that more than 3% of these individuals are criminally insane?

(a) [1] Calculate this *exactly*, using the binomial distribution. Circle closest one. **0.1224** / **0.1454** / **0.1992** / **0.2341** / **0.2745**.

(b) [1] Calculate this using the normal approximation to the binomial distribution. Do *not* use the continuity correction. Circle closest one. **0.1224** / **0.1454** / **0.1992** / **0.2218** / **0.2745**.

(c) [1] The normal approximation is *not* appropriate in this case

because ____.

5. Consider the following couple of questions which concern the t–distribution.

(a) [1] $P(t_{10} > 1.23) \approx$ (circle closest one) **0.123** / **0.131** / **0.197** / **0.222** / **0.259**.

(b) [1] The 67th percentile for the t_{15} distribution is (circle closest one) **0.17** / **0.44** / **0.62** / **1.03** / **1.95**.

(c) [1] The t–distribution is used in test problems where (circle *best* answer)

- **(i)** a small random sample size is taken, where it does not matter whether the distribution is normal or not (because of the central limit theorem).
- **(ii)** a big sample is taken, where it does not matter whether the distribution is normal or not (because of the central limit theorem).
- **(iii)** a small sample is taken, where it is known the distribution is normal.
- **(iv)** a small random sample size is taken, where it is known the distribution is normal.
- **(v)** a big random sample size is taken, where it is known the distribution is normal.

---

6. [3] Identify the following as either continuous or discrete variables and also as either qualitative or quantitative variables.

(a) number of telephones in a home:
(circle two) (yes, circle *two*!)
qualitative / **quantitative** / **continuous** / **discrete**

(b) height of trees:
(circle two) **qualitative** / **quantitative** / **continuous** / **discrete**

(c) drink size (medium, large, super size) at McDonalds:
(circle two) **qualitative** / **quantitative** / **continuous** / **discrete**

(d) names of trees in a back yard:
(circle two) **qualitative** / **quantitative** / **continuous** / **discrete**

(e) test date of exam:
(circle two) **qualitative** / **quantitative** / **continuous** / **discrete**

(f) temperature given by thermometer:
(circle two) **qualitative** / **quantitative** / **continuous** / **discrete**

7. The number of seizures, X, of a typical epileptic person in any given year is given by the following probability distribution.

X	0	2	4	6	8	10
P(X = x)	0.17	0.21	0.18	0.11	0.16	0.17

(a) [1] $P(X \leq 4) =$ ____.

(b) [1] The expected earnings

is $\mu =$ ____.

(c) [1] The standard deviation in earnings

is: ____.

---

8. In a study of annual family expenditures for mental health care, two populations were sampled at random with the following results:

$$\text{Population 1: } n_1 = 6, \quad \bar{x}_1 = \$351, \quad s_1 = 28$$
$$\text{Population 2: } n_2 = 5, \quad \bar{x}_2 = \$321, \quad s_2 = 3$$

We would like to test the hypothesis, at the 1% level of significance, that the difference in means is not zero (do *not* pool).

(a) [1] This is a test of the (circle one) **sample** / **population** difference in means. Since the sample size is small, we cannot rely on the central limit theorem and so must assume the data has been drawn from a (circle one) **normal distribution** / ***t*–distribution**.

(b) [2] Consider the following incomplete graphical representation of this test.

Complete this picture by determining the *numerical* values (the p–value may, for example, have a numerical value of 0.021) of the four items given in the picture above and then filling in the following table.

(a)	(b)	(c)	(d)

9. It is possible to make a mistake when using only the average strength of electrical activity in the thyroid gland of a human being to decide whether a patient has bipolar disorder or schizophrenia. It is assumed the patient has schizophrenia, unless compelling evidence arises to prove bipolar disorder.

(a) [1] In this case, the null hypothesis is (circle one)

- **(i)** the electrical activity in the thyroid gland is zero.
- **(ii)** the patient has bipolar disorder.
- **(iii)** the patient has schizophrenia.
- **(iv)** the patient is "normal", suffering from no mental disorders.
- **(v)** the average electrical activity indicates the patient has schizophrenia.

(b) [1] The type II error occurs when (circle none, one or more)

- **(i)** it is decided to reject the alternative when, in fact, the alternative is true.
- **(ii)** it is decided a patient has schizophrenia, when, in fact, the patient has bipolar disorder.
- **(iii)** mistakenly rejecting the alternative.
- **(iv)** mistakenly accepting the patient has schizophrenia.
- **(v)** the average electrical activity of a sampled group of patients indicates the patients do not have bipolar disorder, when, in fact, they do.

(c) [1] **True / False** In this case, decreasing the probability of a type I error, increases the probability of a type II error.

10. As part of a benefit for good mental health, a lottery is held where one hundred numbered marbles have been placed in a bowl, where each marble is either a \$100, \$200, \$300 or \$400 winner.

marble	01,...,21	22,...,45	46,...,75	76,...,99,00
wins	\$100	\$200	\$300	\$400

(a) [1] If a marble is taken out the bowl at random, the chance it is either a \$100, \$200, \$300 or \$400 winner is (circle one)

(i) Distribution A.

x	\$100	\$200	\$300	\$400
$P(X = x)$	0.21	0.25	0.29	0.24

(ii) Distribution B.

x	\$100	\$200	\$300	\$400
$P(X = x)$	0.21	0.24	0.30	0.25

(iii) Distribution C.

x	\$100	\$200	\$300	\$400
$P(X = x)$	0.21	0.26	0.31	0.25

(iv) Distribution D.

x	\$100	\$200	\$300	\$400
$P(X = x)$	0.21	0.25	0.30	0.24

(b) [1] Use the random numbers generator function of the TI–83 calculator, with seed "5". Generate $k = 10$ two–digit numbers, taken at random, between "00" and "99". Determine the average cost of these 5 pairs (where the first pair, for example, is the first two numbers generated by the calculator and the second pair is the third and fourth numbers generated and so on) and display these results in the following simulated distribution of $\bar{X}$, $n = 2$.

$\bar{x}$	\$100	\$150	\$200	\$250	\$300
proportion (out of 5)					

(1) (a) left, smaller; (b) smaller; (c) larger

(2) (a) $(0.56, 0.88)$ (b) **384**

(3) (a) (iv); (b) **3.9** (c) **1.55**

(4) (a) **0.1992** (b) **0.2218** (c) $np = 115(0.02) = 3.45 < 5$, meaning the Binomial distribution is not symmetric enough to be approximated by the symmetric Normal.

(5) (a) **0.123** (b) **0.44** (c) (iv)

(6) **quantitative discrete**; **quantitative continuous**; **qualitative** or **quantitative discrete**; **qualitative discrete**; **quantitative continuous**; **quantitative continuous**

(7) (a) **0.56** (b) **4.78** (c) **3.47**

(8) (a) **population**; **normal distribution** (b) -3.25, 0, 2.6066, 0.0466

(9) (a) (iii), (b) (i),(ii),(iii),(iv),(v), (c) **True**

(10) (a) (ii), (b) 0, $\frac{2}{5}$, $\frac{1}{5}$, $\frac{2}{5}$, 0