STT 351 **Exam 2**

1. Four pumps are pulled from a process under statistical control. The four pumps average 3.2 gallons per minute with sample s.d.= $s_x = 0.7$.

a. Write the formula for a 99% confidence interval for μ_x.

$\bar{x} \pm 3.182 \frac{s_x}{\sqrt{n}}$

.95
3 3.182

b. Numerically evaluate (a) for the information given but do not reduce.

$3.2 \pm 3.182 \frac{0.7}{\sqrt{4}}$

c. What is your numerical estimate of the population sd σ_x from the information given? Do not reduce.

est is $s_x = 0.7$

d. What is your numerical estimate of the sd of the sample mean (i.e. your estimate of $\sigma_{\bar{x}}$) based on the information given? Don't reduce.

est is $\frac{s_x}{\sqrt{n}} = \frac{0.7}{\sqrt{4}}$

e. What $\boxed{\text{performance claim}}$ is made for a 99% confidence interval?

~~APPROX~~ 99% OF SAMPLES OF $n=4$ YIELD A 99% t CONF INTERVAL(1) CONTAINING μ_x.

t IS "EXACT"

f. For large n, what happens to the width of a CI if n is replaced by 4n?

$\frac{1}{\sqrt{4n}} = \frac{1}{2\sqrt{n}}$ SO WIDTH OF CI IS HALF AS LARGE.

g. What happens to $\frac{\bar{x} - \mu_x}{s_x / \sqrt{n}}$ if scores x_i are replaced by $(3\, x_i - 6)$?

UNCHANGED!

eg $\overline{3x_i} - \mu_{3x} = (\bar{x} - \mu_x)3$
$s_{3x} = 3 s_x$

RATIO IS SAME AS FOR x

LIKEWISE $\overline{x+c} - \mu_{x+c} = \bar{x} - \mu_x$
$s_{x+c} = s_x$

RATIO UNCHANGED

2. Sixty with-replacement samples are selected from parts plated with process x. Independently of these, 100 samples are selected from parts plated with process y. The following data apply to a measurement of corrosion resistance

	sample mean	sample sd	sample size
x	24.6	3.7	60
y	23.8	3.9	100

a. Numerically determine, but do not reduce, your large-n estimate of the margin of error for $\overline{x} - \overline{y}$.

$$1.96 \sqrt{s_x^2/n_x + s_y^2/n_y} = 1.96\sqrt{\frac{3.7^2}{60} + \frac{3.9^2}{100}}$$

b. Same as (a) except the samples are without relacement and the population sizes are 800 and 600 for x and y respectively. use FPC's

$$1.96\sqrt{\frac{3.7^2}{60}\,\frac{800-60}{800-1} + \frac{3.9^2}{100}\,\frac{600-100}{600-1}}$$

3. It is desired to estimate the mean of $\boxed{y = \text{burst pressure}}$ for a type of plastic inflatable already in use by the public. A random sample of 50 is selected. The sample is effectively with replacement since the population size N is so large relative to 50. It is anticipated that another measurement x = time in use may have a bearing on this, so both x and y scores are measured for each of the 50. We find sample s.d. are

$s_x = 17.8$ months $\qquad s_y = 20$ psi

If the sample correlation of x with y is r = 0.8 determine the numerical value of the margin of error of the regression based estimator of μ_y. Do not reduce.

$$1.96\sqrt{1-r^2}\,\frac{s_y}{\sqrt{n}}$$

$$1.96\sqrt{1-.8^2}\,\frac{20}{\sqrt{50}}$$

4. A random sample of 100 lead seals is selected from production. Of these are 29 found to be defective.

a. Give the formula for a 95% z-based CI for the population fraction of defective seals. $\hat{p} \pm 1.96 \sqrt{\hat{p}(1-\hat{p})}/\sqrt{n}$

$\frac{29}{100} \pm 1.96 \sqrt{(29/100)(71/100)}/\sqrt{100}$

b. Numerically evaluate (a) but do not reduce.

5. A 95% bootstrap ci for the population mean, based on a dataset named "prices" is given by a call to the function

bootci[mean, prices, 10000, 0.95]

a. Give the call required to generate a $\boxed{\text{99 \%}}$ bootstrap ci for the population $\boxed{\text{median}}$ but with $\boxed{\text{twice the number of bootstrap replications}}$.

bootci[MEDIAN, PRICES, 20000, 0.99]

b. Is it correct to say that the bootstrap method effectively increases the sample size of the dataset? NO, OUR BOOTSTRAP REPLICATE SAMPLES ARE NOT NEW DATA, THEY MERELY SUBSTITUTE FOR ANALYTICAL CALCULATION.

6. Machine processes are scored for x = efficiency. It is desired to obtain a 95% ci for μ_x by the method of $\boxed{\text{proportional stratified sampling}}$ with strata corresponding to three levels of cycling rate.

stratum	**1**	**2**	**3**
stratum size	**1000**	**1500**	**800**

$N = 3300$

a. If a total sample size of 33 is used, how is this allocated among the three strata?

$n_1 = w_1 n = \frac{1000}{3300} 33 = 10$

$n_2 = 15$

$n_3 = 8$ / TOTAL 33

b. The stratum by stratum sample means are

stratum	1	2	3
sample mean	2.6	3.1	2.8

Determine the overall sample mean of all 66. Don't reduce.

$$\bar{x} = \Sigma_1^3 w_i \bar{x}_i = \frac{10}{33}(2.6) + \frac{15}{33}(3.1) + \frac{8}{33}(2.8)$$

c. The estimated s.d. of $\overline{X}$ (from this stratified sample) is given by $\sqrt{\Sigma_1^3 W_i^2 \frac{s_i^2}{n_i}}$. Give the numerical values of the weights W_i.

$w_1 = \frac{10}{33}$ $w_2 = \frac{15}{33}$ $w_3 = 8/33$

7. A maximum likelihood estimator selects the model giving the most probability to what has been seen (the data). By this way of thinking, which model best explains the event R1 G2?

model 1: select two with replacement from [R R G G G]

model 2: select two without replacement from [R G G]

Show your calculations.

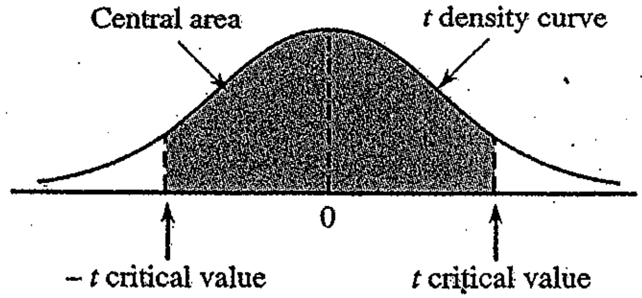
① $P(R1G2) = \frac{2}{5}\frac{3}{5} = \frac{6}{25}$

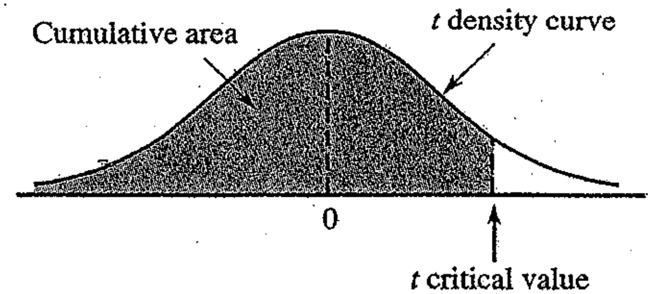
② $P(R1G2) = \frac{1}{3} 1 = \frac{1}{3}$

$\frac{1}{3} > \frac{6}{25}$ (i.e. $25 > 18$) so

MODEL ② IS THE MLE CHOICE AMONG THE TWO MODELS.

Table IV *t* critical values for confidence and prediction intervals





Central area = confidence/prediction level for two-sided interval:		80%	90%	95%	98%	99%	99.8%	99.9%
Cumulative area = confidence/prediction level for one-sided interval:		90%	95%	97.5%	99%	99.5%	99.9%	99.95%
Degrees of freedom	**1**	3.078	6.314	12.706	31.821	63.657	318.31	636.62
	2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
	3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
	4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
	5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
	6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
	7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
	8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
	9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
	10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
	11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
	12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
	13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
	14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
	15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
	16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
	17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
	18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
	19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
	20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
	21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
	22	1.321	1.717	2.074	2.508	2.819	3.505	3.792