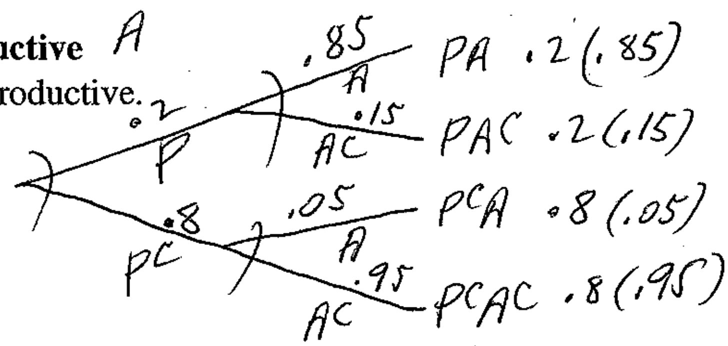
STT351-002
Final Exam

1-2. TREE.

20% of fields are productive. P
85% of productive fields **appear productive** A
5% of non-productive fields appear productive.
Make a tree.



1. P(field appears productive)

$$P(A) = P(PA) + P(P^C A)$$
$$= .2(.85) + .8(.05)$$

2. P(field is productive | field appears productive)

$$= P(PA)/P(A)$$
$$= .2(.85)/(.2(.85) + .8(.05))$$

3-4. CI and TEST for MEAN μ. A sample of n = 6 prescription eyeglass lenses is drawn from a process **under statistical control**. Each of these **six** is subjected to measurements which determine an overall score x = "conformity to prescription." The sample mean = 2.2 and the sample sd s = 2.8.

CENTRAL .95

3. Determine the 95% CI for μ. $\bar{x} \pm t_{DF5, .95}\, s/\sqrt{n}$

DF = 6 − 1 = 5

2.571

DF
5 $\boxed{2.571}$

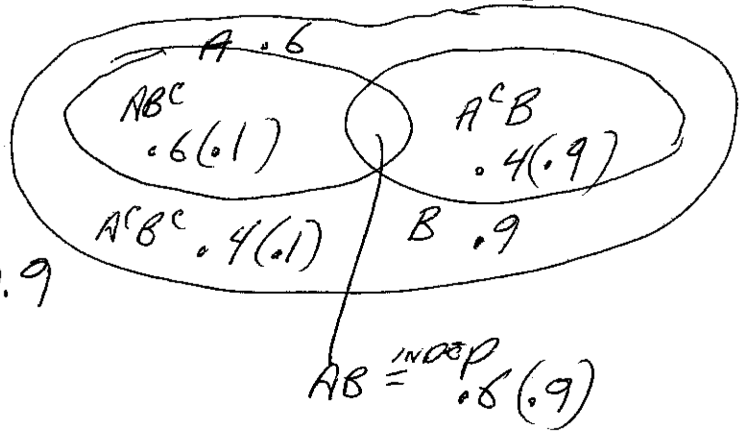
4. Determine the final sample size $n_{\mathbf{FINAL}}$ required for 95% **hybrid CI**
$\overline{x}_{\text{FINAL}} \pm 0.04$

$$t_{INIT} \frac{s_{INIT}}{\sqrt{n_{FINAL}}} = 0.04$$

$$n_{FINAL} = \left(\frac{t_{INIT}\, s_{INIT}}{0.04}\right)^2 = \left(\frac{2.571\ 2.8}{0.04}\right)^2$$

5-6. Probability Rules. P(A) = 0.6, P(B) = 0.9, events A, B are **independent.**

5. Complete a Venn diagram.



6. P(B | ~~A~~) INDEP $= P(B) = 0.9$

7-8. Drawing balls. Draws will be made **without replacement** and with equal probability on those remaining from {R R R Y Y B B B B} (i.e. **3 R, 2 Y, and 4 B**}.

7. P(R4) = $\frac{1}{3}$ by what simple principle? ORDER OF DEAL DOES NOT MATTER

$P(R4) = P(R1) = 3/9 = \frac{1}{3}$

8. Use rules of probability to PROVE P(R2) = $\frac{1}{3}$ by breaking down event R2 according to what happens on draw one. **Cite the rules you use.**

$P(R2) = P(R1\,R2) + P(R1^c\,R2)$ TOTAL PROBY

$= P(R1)\,P(R2|R1) + P(R1^c)\,P(R2|R1^c)$ MULT

$= 3/9 \; 2/8 + 6/9 \; 3/8 = 24/(9.8) = 3/9 = 1/3$

9-10. Estimates. A sample of n = 100 is selected **without replacement** and with equal probability from a population of size 300. This sample has mean $\overline{x} = 2.1$ with s = 0.6.

9. Estimate the sd $\sigma_{\overline{x}}$ of sample mean $\overline{x}$ WITHOUT REPL

$$\sqrt{\frac{N-n}{N-1}}\frac{s}{\sqrt{n}} = \sqrt{\frac{300-100}{300-1}}\frac{0.6}{\sqrt{100}}$$

10. Estimate the margin of error for $\overline{x}$. 1.96 TIMES (9)

$$1.96\sqrt{\frac{300-100}{300-1}}\frac{0.6}{\sqrt{100}}$$

11-12. CI for SUM of means. A with-replacement sample of 70 parts from supplier A finds sample mean breaking strength = 560 with sample sd s = 29. Independently of this, a with-replacement sample of 90 parts from supplier B finds sample mean breaking strength 540 with sample sd s = 38.

11. Determine the 95% z-based CI for the difference $\mu_A - \mu_B$.

$$\bar{x} - \bar{y} +/- 1.96 \sqrt{\frac{\sigma_x^2}{n_x} + \sigma_y^2/n_y}$$

$$(560-540) \pm 1.96 \sqrt{\frac{29^2}{70} + \frac{38^2}{90}}$$

12. Determine the 95% z-based CI for the **SUM** $\mu_A + \mu_B$.

$$\bar{x} + \bar{y} \pm 1.96 \sqrt{\frac{\sigma_x^2}{n_x} + \sigma_y^2/n_y}$$

$$560+540 \pm 1.96 \sqrt{\frac{29^2}{70} + \frac{38^2}{90}}$$

13-14. Binomial. On average 15% of a population of vases will be damaged in shipment (i.e. p = 0.15). Recall the discrete probability density for binomial is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, ..., n.$$

FIX

CORRECTION

13. Determine p(2). Evaluate the binomial coefficient but do not otherwise reduce. $n=20$

$$p(2) = \binom{20}{2} .15^{2} .85^{20-2}$$

$$= \frac{20 \cdot 19}{2} .15^2 .85^{18}$$

14. Determine a 90% z-based CI for p if we find that out of a particular sample of 200 vases there are 25 damaged in shipment.

$$\hat{p} = \frac{25}{200}$$

$$\hat{p} \pm 1.645 \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

$$\frac{25}{200} \pm 1.645 \frac{\sqrt{\frac{25}{200} \frac{175}{200}}}{\sqrt{200}}$$

15-16. Poisson. We average around 3.7 work stoppages per day. If the distribution of the number of work stoppages is Poisson recall $p(x) = e^{-\mu} \frac{\mu^x}{x!}$, x = 0, 1, 2, ... ad inf.

15. In the z approximation of P(more than two stoppages tomorrow) we use a standard score z of 2.5 (continuity correction of "more than two"). Determine this z.

$$z = \frac{2.5 - \mu}{\sqrt{\mu}} = \frac{2.5 - 3.7}{\sqrt{3.7}}$$

16. Determine P(more than two stoppages tomorrow) directly from the discrete density but express it as a finite, not infinite, sum.

$$1 - p(0) - p(1) = 1 - e^{-3.7}\frac{3.7^0}{0!} - e^{-3.7}\frac{3.7^1}{1!}$$

17-20. Chi-square. A process has been producing LEDs on average in proportions 25% excellent 10% good 65% avg

A random sample of 200 LEDs finds

observed:	48 excellent	32 good	120 avg
expected:	50	20	130

eg $50 = (.25)(200)$

17. Fill in the expected counts above consistent with past performance.

18. Determine the contribution of category "good" to the chi-square statistic.

$$\frac{(O-E)^2}{E} = \frac{(32-20)^2}{20}$$

19. Determine the degrees of freedom of the chi-square.

DF = 2

RUBRIC $k-1 = 3-1 = 2$

GENERAL PRINCIPLE:

#FREE PARAMETERS FULL = 3-1 = 2

− #FREE " IN H_0 = 0

DIFF = 2

20. Determine p_{SIG} if the chi-square statistic is 11.6571.

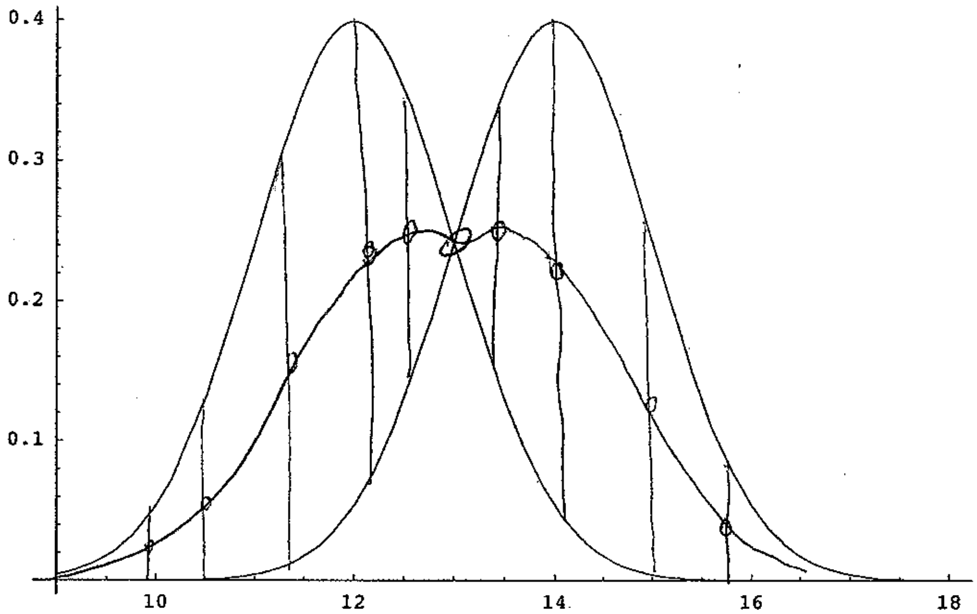
$$p_{SIG} = P(\chi^2_{2DF} > 11.6571)$$

BETWEEN .001 & .005.

TABLE VII

RT TAIL	DF 2
.005	10.59
.001	13.81

21. Kernel density. Bell curves are placed at each of two points (see below). Plot the kernel density estimate. Take care to do it correctly (show five pts accurately).



22-24. Rules for E, Var, sd. Random variables X, Y are **independent** with

E X = 6 Var X = 4

E Y = 9 Var Y = 2

22. E (XY) $\overset{IND}{=} EX\,EY = 6(4)$

23. E (Y^2) (follows from Var Y = E (Y^2) - $(E\,Y)^2$) $= Var\,Y + (EY)^2$
$= 2 + 9^2$

24. sd (6X - 3Y + Y - ~~4~~) $= SD(6X - 2Y) = \sqrt{Var(6X-2Y)}$
$= \sqrt{36\,Var X + 4\,Var Y} = \sqrt{36(4) + 4(2)}$

25. Plot regression line. Parts are sampled with-replacement and scored (x, y) where

x = serial number of part   y = hardness.

The sample data are:

$\bar{x} = 1343$ $s_x = 433$ **n** = 200 pairs (x, y)

$\bar{y} = 12.7$ $s_y = 1.1$ **r** = 0.7

What is the value y for a point on the regression line with x = 1343 + 433 (i.e. one sample sd above the sample mean in the x-scale)?

$$\underline{\text{ANS}}.\ \bar{y} + r(1)\sigma_y = 12.7 + 0.7(1.1)$$

26. Proportionally stratified. A population of motors is stratified by supplier

20% A   10% B   70% C

A **stratified sample** of motors produces the following sample means by stratum

stratum	A	B	C
sample mean	2.4	2.7	2.0

Estimate the population mean μ from the above data.

$$\bar{x} = \sum_{i=1}^{3} w_i \bar{x}_i = .2(2.4) + .1(2.7) + .7(2.0)$$

27. Calculating SD. For the following discrete distribution calculate the standard deviation σ.

x	p(x)	$x\,p(x)$	$x^2 p(x)$
0	0.8	0	0
1	0.2	.2	.2
		$EX = .2$	$EX^2 = .2$

$$\sigma = \sqrt{\mathrm{Var}\,X} = \sqrt{EX^2 - (EX)^2} = \sqrt{.2 - .2^2} = \sqrt{.2 \cdot .8}$$

USUAL FORMULA $\sqrt{p(1-p)}$ FOR 0-1 SCORES

28-29. Multiple regression. A random sample of 400 of our products is selected from stores nationwide. Each is scored for

y = selling price

x1 = 1 if store is major retailer, 0 if not

x2 = quantity ordered by store

A multiple linear regression is fit to this data resulting in the fitted model

y = 44.75 - 7.80 x1 - 0.083 x2

28. Determine the average effect on price (according to the fitted model) occasioned by adding 500 to the order and switching from a major retailer to one that is not a major retailer. $\Delta y = 7.8$ (FROM $-7.80\,1$ TO $-7.80(0)$)

$- 0.083(500)$ (FROM $-0.083\,x_2$ TO $-0.083(x_2+500)$)

$\Delta y = 7.8 - (0.083)(500)$

29. Compare the 95% CI of μ based on $\bar{y} = 42.76$ with that based on the regression based estimator if

CORRECTION

sample multiple correlation is $\hat{R} = 0.6$,
regression based estimator works out to 37.80.

95% CI using $\bar{y} \pm 1.96 \frac{s_y}{\sqrt{n}} = 42.76 \pm 1.96 \frac{s_y}{\sqrt{400}}$

95% CI using regression based estimator $37.80 \pm \underbrace{\sqrt{1-\hat{R}^2}}_{\sqrt{1-0.6^2}\,(=.8)} \, 1.96 \frac{s_y}{\sqrt{400}}$

30. t–TEST. A process is in control. Each part produced is score x = finishing time. A sample of 12 will be used to monitor the process in a test of the null hypothesis

H_0: $\mu_x = 5$ (minutes) vs H_1: $\mu_x \neq 5$ with $\alpha = \mathbf{0.1}$ 2 SIDED $\alpha/2 = .05$

30. If the test statistic for a sample of 12 evaluates to t = 2.8 what action is taken by the test? Indicate your reasoning. $DF = n-1 = 12-1 = 11$.

$.05 = \alpha/2$

2-SIDED TEST

REJECT H_0 IF $|\text{TESTSTAT}| > t_{\text{CRITICAL}}$

i.e. IF $|2.8| > 1.796$

$\Rightarrow$ REJECT H_0

TABLE IV

DF	CUMULATIVE .95
11	1.796

OR TABLE VI $P_{SIG} = P(|T| > 2.8) = 2(.009)$

DF	
11	
2.8	.009