Final exam prep 8 - 16 - 10 Try these. Not handed in. Will go over in class. Ask questions. Report errors.

1. Chi-square basic method (for classified independent samples having completely specified expected counts). Sales of sizes of particular athletic shirts have, from past experience, the probabilities

XS	S	M	L	XL	XXL
.02	**.04**	**.09**	**.31**	**.40**	**.14**

This season we have changed the fabric to one having a lighter, silky, breathable feel. Here are sales figures from our test run of the new shirts at an event where 1100 shirts were sold:

	XS	S	M	L	XL	XXL
observed sales	**30**	**67**	**106**	**319**	**430**	**148**
expected counts						

a. Fill in the above *expected counts* if 1100 sales follow the past model.

b. Chi-square statistic

c. df

e. P-value

f. If the past model continues to apply to present sales what is the probability of a P value as small or smaller than we have seen?

g. What, if anything, appears to have happened to size preferences due to the new fabric?

2. Another chi-square for classified independent samples having completely specified expected counts. This exercise will illustrate the fact that the basic method is not restricted to any particular shape for the table. A with-replacement random sample of **100** parts from production has been sorted below by weight of material, weight of scrap, and time of assembly. The usual cell probabilities (from past production) are given in parentheses and are used to determine the expected counts:

On-time parts

	over weight	not over weight
allowable scrap	3 (.04)	57 (.32)
excessive scrap	1 (.03)	4 (.06)

Late parts (includes machine down time)

	over weight	not over weight
allowable scrap	3 (.06)	19 (.30)
excessive scrap	3 (.05)	10 (.14)

The system has just returned to production following maintenance. We question whether there is evidence that the current sample differs materially from past experience. If so, we need to know if there seems to have been improvement.

a. Are all expected counts at least three (a rule of thumb sometimes used)?

b. Chi-square statistic

c. df

d. P-value (use your calculator and check against table entries)

e. Is P-value small enough to suggest to convince you that production is different from past experience? If so, what changes seem to have occurred and do these appear to be favorable or mixed?

Suppose the cell probabilities had been changed (in parentheses) as below.

On-time parts

	over weight	not over weight
allowable scrap	3 (.04)	57 (.33)
excessive scrap	1 (.02)	4 (.06)

Late parts (includes machine down time)

	over weight	not over weight
allowable scrap	3 (.06)	19 (.33)
excessive scrap	3 (.02)	10 (.14)

f. Merge cells, as makes most sense to you, bringing all expected counts up to at least 3.

g. Chi-square statistic after merge

h. df after merge

i. P-value after merge

j. Is P-value after merge small enough to convince you that production is different from past experience? If so, what changes seem to have occurred and do these appear to be favorable or mixed?

k. By comparison with your findings (e) has the need to merge interfered with any important conclusions you were able to make when no merge was needed?

3. A gene for flower color has the following visible outcomes

AA	**red flower**
aA or Aa	**pink flower**
aa	**white flower**

If there is random mating of flowers there will be a p in [0, 1] for which the population distribution takes the form

AA	aA or Aa	aa
p^2	2 p (1-p)	$(1-p)^2$

Suppose a random sample of 100 flowers finds

AA	aA or Aa	aa
41	38	21

a. Estimate p = $\frac{\text{\# letters } A \text{ in the population}}{\text{\# letters in the population}}$ by

$$\hat{p} = \frac{\text{\# letters } A \text{ in the sample of 100 flowers}}{\text{\# letters in 100 flowers}}$$

b. From (a) and the distribution determine the expected counts for AA, aA or Aa, aa.

c. Chi-square statistic

d. df

e. P-value

f. Does there seem to be strong evidence against what is expected in random mating?

Notice that in the above example there is no need for genetic analysis of the flowers in order to determine gene-types. Gene types are directly deduced from flower color. So we can directly get at the matter of whether the flowers are consistent with random mating.

4. Chi-square test of independence. A random sample of 113 customers of a large resort is classified according to number of nights and number of beds.

	1 night	2 nights	longer
1 bed	29	31	9
2 beds	12	19	5
more	0	6	2

a. Determine the marginal counts and from them the *expected counts* under the model that the number of beds is statistically independent of the number of nights.

	1 night	2 nights	longer	
1 bed				69
2 beds				36
more				8
	41	56	16	**113**

b. Chi-square statistic

c. df

d. P-value

e. Is there much reason to question independence of the number of beds from the nights stayed?