Biometry, Math 2520, Spring 2008
Final Examination Solutions

1. The following table of data is from an unpublished morphometric study of cottonwood by T.J. Crovello. Y_1 is fresh leaf width and Y_2 is dry leaf width, both in millimeters. Calculate r for this data and determine if there is significant correlation.

Y_1	Y_2
90	88
99	87
55	52
100	95
86	83
90	88
82	77
78	75
115	109
100	95
110	105
84	78
76	71
100	97

Here $\overline{Y}_1 = 90.357, \overline{Y}_2 = 85.71$, then we calculate $\sum y_1^2 = 2836$ and $\sum y_2^2 = 3085$, finally $\sum y_1 y_2 = 2919$, so that $r = \frac{2919}{\sqrt{2836 \cdot 3085}} = .9868$. To test if this is significant, with $H_0 : \rho = 0$, $t_s = r \cdot \sqrt{\frac{n-2}{1-r^2}} = .9868\sqrt{\frac{12}{1-.97377}} = 21.12$. This is much larger than $t_{.001[12]}$, so we reject the null hypothesis, accepting the alternative hypothesis that $\rho \neq 0$, that is, there is a significant correlation.

2. In this experiment, stem mothers of the aphid *Myzus persicae* were placed on one of three diets one day before giving birth to nymphs (data from Mittler and Dadd, 1966). Test whether the proportion of nymphs that developed into winged forms depends on the diet of the mother. Use a G-test or prove to me that you are trapped on a desert island without a way to calculate logarithms and use a chi-squared test.

Diet	Winged forms	Non-winged forms	row total
Synthetic	216	0	216
Cotyledon "sandwich"	211	19	260
Free cotyledon	27	48	75
col. total	454	67	521

We're clearly not on a desert island, so we'll do this with a G-test for independence. To calculate G, we take $f \ln f$ for all the numbers in the chart, letting $0 \ln 0 = 0$ even though it's an indeterminate form (well, in fact $\lim_{f \to 0^+} f \ln f = 0$, so this is the right thing to do [there are papers that suggest other numbers to use in this case, more or less saying that we ought to act as though f is a small fraction, but I digress even more than usual]) then give the row and column totals a negative sign, adding them all up, and multiplying by 2. This gives us $G = 170.674$, which is much much more than $\chi^2_{.001[2]}$, so we don't need to calculate G_{adj}, and just accept the alternative hypothesis that diet does affect type.

3. The mean length of developmental period (in days) for 3 strains of houseflies at 7 densities is given below. Perform a 2-way analysis of variance to determine if these flies differ in developmental period with density and among strains. You may assume that there is no interaction between strain and density. (Data from Sullivan and Sokal, 1963).

Density	OL strain	BELL strain	bwb strain	Row average
60	9.6	9.3	9.3	9.40
80	10.6	9.1	9.2	9.63
160	9.8	9.3	9.5	9.53
320	10.7	9.1	10.0	9.93
640	11.1	11.1	10.4	10.87
1280	10.9	11.8	10.8	11.17
2560	12.8	10.6	10.7	11.37
Column average	10.785	10.043	9.9857	$\overline{\overline{Y}} = 10.271$

Source	SS	df	MS	F_s	sig.
Strain	2.788857	2	1.3942857	4.18	5%
Density	12.5428	6	2.090476	6.14	1%
Interaction	4.033	12	.3358		
Total	19.44286	20			

Note that we use the MS for interaction as the denominator in our F_s, so that our F statistics are compared with critical values for 2,12 df and 6,12 df. In both cases we accept the alternative hypothesis that developmental period differs with density and among strains.

4. You have performed four different experiments to test if Komodo dragons have a larger lung capacity than Gila monsters. In all cases, the null hypothesis is that the capacities are the same and the alternative hypothesis is that the Komodo dragons have larger capacities. All of your experiments were inconclusive with P-values of .131, .072, .180, and .112. Combine these P-values to determine if you can reject the null hypothesis on the basis of all four experiments.

The statistic to calculate is $-2\sum \ln P$, and it is distributed as a chi-square with 8 degrees of freedom. $-2\,(\ln .131 + \ln .072 + \ln .180 + \ln .112) =$
$-2\,(-2.032558 - 2.631089 - 1.714798 - 2.189256) = 17.1354$, which gives a P-value between .01 and .05, so that we may reject the null hypothesis.

5. In a game, you roll three 6-sided dice, and count the number of sixes obtained. Compute expected binomial frequencies for 1000 repetitions of these rolls and determine if the following data fits such a binomial distribution.

I choose to use $\frac{1}{6}$ as p, thus having an extrinsic hypothesis, and thus can test my G-statistic for goodness of fit against a chi-square distribution with 3 df. If you choose to find $p = (342 + 2 \cdot 73 + 3 \cdot 5)/3000 = .167\bar{6}$, then you have lost a df, so should compare with 2 df.

Number of sixes	f	$\left(\binom{3}{Y}\right)$	p^Y	q^{3-Y}	$\hat{f}_{rel}$	$\hat{f}$
0	580	1	1	.5787	.5787	578.7
1	342	3	.1667	.6944	.3472	347.2
2	73	3	.02778	.8333	.30644	69.4
3	5	1	.00462963	1	.00462963	4.6

Then $G = 2\left(580 \ln \frac{580}{578.8} + 342 \ln \frac{342}{347.2} + 73 \ln \frac{73}{69.4} + 5 \ln \frac{5}{4.6}\right) = .2897$, which is less than $\chi^2_{.1[3]}$, so we accept the null hypothesis of a good fit.

6. The Mythbusters are testing if the humidity level at which baseballs are stored affects the distance that they travel when hit. (Episode 83, aired August 8, 2007) Three groups of 36 baseballs are stored at different humidities and then a machine is used to hit the balls from home plate. The distances are measured. Use analysis of variance to determine if the humidity affects distance.

Humidity	Average distance	Variance of distance
10%	135	240
50%	120	220
90%	110	225

First we find $\overline{\overline{Y}} = 121.\bar{6}$. Since we have variances of equally sized samples, we can simply average them to get MSE=$228.\bar{3}$. Our complete ANOVA table follows:

Source	SS	df	MS	F_s	sig.
AMONG	11400	2	5700	24.96	.1%
WITHIN	23975	105	$228.\bar{3}$		
TOTAL	35375	107			

So we take the alternative hypothesis that there are differences in groups based on humidity. This suggests that the Coors field humidor is effective, though correlation does not imply causation.