Final for Statistics 503
Statistical Methods In Biology - Fall 2000
Material Covered: Chapters 1–10 of Workbook and text
11th December

This is a 2 hour final, worth 25% and marked out of 25 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 quiz. A calculator may also be used. No other aids are permitted.

Name (please print): ____________________________ . **ID Number:** ______.

last first

---

1. [2 points] One hundred and twenty (120) pea plants are selected at random and the number of pea pods produced per plant is measured (observed). From this group, an average number of pea pods per plant is computed. Match the columns: *All* of the items in the first column will be used up in the matching procedure; however, one item in the second column will be left unmatched.

statistical terms	pea pods example
(a) observation	**(a)** average number of pea pods per plant for 120 pea plants
(b) variable	**(b)** all pea plants
(c) parameter	**(c)** number of pea pods per plant for all pea plants
(d) statistical population	**(d)** number of pea pods for a pea plant
(e) sample	**(e)** average number of pods per plant for all pea plants
(f) statistic	**(f)** 120
(g) sample size	**(g)** number of pea pods per plant for 120 pea plants
	(h) number of pea pods for a particular pea plant

terms	(a)	(b)	(c)	(d)	(e)	(f)	(g)
pea pod example							

2. A tumor is cancerous 16% of the time; seven tumors are inspected. Assume this problem obeys the conditions of a binomial experiment.

(a) [1 point] The chance that at least 5 of the tumors are cancerous is
(circle one) **0.0017** / **0.1721** / **0.3334** / **0.6734** / **0.9983**

(b) [1 point] The variance in the number of tumors is
(circle one) **0.27** / **0.94** / **1.12** / **2.67** / **2.88**

(c) [1 point] A Poisson approximation in this example is *not* advisable because (circle one)

(i) $n \neq 100$ and $n\pi \leq 20$

(ii) $n > 100$ and $n\pi \leq 10$

(iii) $n < 100$ and $n\pi \leq 20$

(iv) $n \geq 100$ and $n\pi \leq 10$

(v) $n \leq 100$ and $n\pi \leq 10$

---

3. Consider the χ^2 distribution.

(a) [1 point] For this distribution with 12 degrees of freedom,

$P\left(\chi^2 \leq 3.24\right) =$ ______________________________.

(b) [1 point] The 37th percentile for this distribution with 9 degrees of freedom

is ______________________________.

(c) [1 point] A χ^2 distribution with 5 degrees of freedom is (circle one)

(i) more skewed left than

(ii) as skewed left as

(iii) as symmetric as

(iv) as right skewed as

(v) more right skewed than

a χ^2 distribution with 25 degrees of freedom.

4. The average survival time of leukemia patients in the midwest is assumed to be 17 months from time of diagnosis. The Cancer Research Society (CRS), however, claims the average survival time to be longer than this. The average survival time of a random sample of size $n = 15$ patients is $\bar{y} = 18.5$ months and the standard deviation in survival time is $s = 5.5$ months. Does this data support the CRS's claim at $\alpha = 0.05$? Assume normality.

(a) [1 point] A test of the CRS claim involves using the (circle best one)

- **(i)** normal distribution.
- **(ii)** t distribution with 14 degrees of freedom.
- **(iii)** χ^2 distribution with 14 degrees of freedom.
- **(iv)** F distribution with (14,15) degrees of freedom.
- **(v)** binomial distribution.

(b) [1 point] The p–value

is ____.

(c) [1 point] A 95% lower confidence interval for μ

is ____.

---

5. The observed data of the incidence of colon cancer in parents and their children from a random sample of 329 families in a midwestern city is given in the table below.

	children have colon cancer	children do not colon cancer	
parents have colon cancer	18	12	30
parents do not have colon cancer	22	277	299
	40	289	329

(a) [1 point] The test of whether or not those children who have colon cancer is *dependent* on whether or not the parents have colon cancer at $\alpha = 0.05$ has an observed χ^2 test statistic value of (circle closest one)
56.7 / 63.3 / 68.2 / 70.0 / 70.7

(b) [1 point] The test of whether there is a larger proportion of children with colon cancer of parents who had colon cancer than of children with colon cancer of parents who did not have colon cancer at $\alpha = 0.05$ has an observed z test statistic value of (circle closest one)
2.38 / 3.33 / 5.80 / 6.79 / 8.41

6. In a controlled randomized experiment, the effect of different levels of (bad) fat and salt on the average volume (in mm^3) of cancer tumor nodules in mice is investigated. For instance, the average volume of cancerous tumor nodules found in the first three mice, fed a diet with a low dose of (bad) fat and a low dose of salt, are 4.2, 3.1 and 2.9 mm^3, respectively.

	(bad) fat dosage →	low (L)	medium (M)	high (H)
salt dosage ↓	low (L)	4.2, 3.1, 2.9	4.4, 3.2, 3.3	5.9, 6.2, 5.5
	medium (M)	4.1, 3.7, 3.9	5.2, 4.5, 4.2	6.2, 6.7, 5.4
	high (H)	5.2, 4.1, 3.4	5.3, 4.1, 4.9	6.7, 7.5, 6.9

(a) [1 point] Subjecting three mice to each treatment is better than subjecting one mouse to each treatment because this helps (circle none, one or more answers)

(i) eliminate the confounder "salt".

(ii) eliminate the confounder "(bad) fat".

(iii) eliminate any confounders, whether external or internal.

(iv) decrease the variance in the mean response for each treatment.

(v) increase the reliability of statistical inference.

(b) [1 point] Twenty–seven mice, arranged from lightest in weight to heaviest in weight, are numbered 01, 02, ..., 27, respectively. How would these mice be assigned treatments so that the variable "salt" is confounded with the external variable "weight"? Complete the following table, where, for example, as shown below, mouse 01 is fed a low (bad) fat and low salt diet.

(bad) fat	L			M			H		
salt	L	M	H	L	M	H	L	M	H
mouse	01	___	___	___	___	___	___	___	___
mouse	___	___	___	___	___	___	___	___	___
mouse	___	___	___	___	___	___	___	___	27

(c) [1 point] Complete the following ANOVA table using the data above using "salt" factor as the treatment variable *and ignoring the "(bad) fat" factor.*

Source	Degrees of Freedom	Sum Of Squares	Mean Squares
Treatment (Salt)	___	___	___
Error	___	___	___
Total	___	___	

6 Continued. Continue to use the data on the average volume (in mm^3) of cancer tumor nodules in mice, given below again (for your convenience) to answer three more questions.

	(bad) fat dosage →	low (L)	medium (M)	high (H)
salt dosage ↓	low (L)	4.2, 3.1, 2.9	4.4, 3.2, 3.3	5.9, 6.2, 5.5
	medium (M)	4.1, 3.7, 3.9	5.2, 4.5, 4.2	6.2, 6.7, 5.4
	high (H)	5.2, 4.1, 3.4	5.3, 4.1, 4.9	6.7, 7.5, 6.9

(d) [1 point] Complete the following ANOVA table using the data above using "(bad) fat" factor as the treatment variable *and ignoring the "salt" factor.*

Source	Degrees of Freedom	Sum Of Squares	Mean Squares
Treatment (Fat)	______	______	______
Error	______	______	______
Total	______	______	

(e) [1 point] A q–q plot of the data, where "(bad) fat" is used as the treatment variable and the "salt" factor is ignored, indicates (circle one)
heavy tail / light tail / normality / left skew / right skew.

(f) [1 point] A $e \vee p$ plot for the data, where "(bad) fat" is used as the treatment variable and the "salt" factor is ignored, indicates variance is related to the mean μ, in the following way (circle one)
$\boldsymbol{\sigma^2 = k\mu(1-\mu)}$ / $\boldsymbol{\sigma^2 = k\mu}$ / $\boldsymbol{\sigma^2 = k}$ / $\boldsymbol{\sigma^2 = k\mu^2}$ / $\boldsymbol{\sigma^2 = k\mu^{-2}}$

---

7. [1 point] Consider the following table of wheat yield (kilograms) versus amount of water (liters per square meter).

amount of water, x	0.13	0.54	0.73	1.11	1.32	1.54	1.78	2.31	2.54	2.88
wheat yield, y	7.1	7.0	7.5	8.8	9.1	9.4	10.0	9.2	9.0	8.5

The residual at $x = 1.11$ is ______________________________.

8. [1 point] Consider the following table below which compares the Fisher, Bonferroni and Scheffe critical contrast values (CCVs) at $\alpha = 0.05$ for three contrasts.

	contrast	$\lvert\hat{\theta}\rvert$	Fisher CCV	Bonferroni CCV	Scheffe CCV
1	$(-\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, 1)$	3.9	2.19	3.25	4.01
2	$(0, 1, -1, 0, 0)$	2.0	3.25	4.01	5.18
3	$(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, 0)$	4.1	2.73	3.25	3.39

Circle true or false.

(i) True / False The first contrast, contrast 1, is significant with respect to both the Fisher CCV and Bonferroni CCV, but not the Scheffe CCV.

(ii) True / False The second contrast, contrast 2, is significant with respect to all three CCVs.

(iii) True / False The third contrast, contrast 3, is significant with respect to all three CCVs.

(iv) True / False To say the second contrast, contrast 2, is significant implies $\mu_2 \neq \mu_3$.

---

9. Twelve roses are subjected to three fertilizers to determine the influence on length of bloom time (measured in days):

fertilizer 1	53	72	69	89	$\bar{y}_1 \approx 70.75$
fertilizer 2	47	62	55	56	$\bar{y}_2 \approx 55$
fertilizer 3	33	45	41	50	$\bar{y}_3 \approx 42.25$

Test if the mean rose bloom time for fertilizer 2 is larger than the average of the mean rose bloom time for fertilizer 1 and fertilizer 3 at $\alpha = 0.05$.

(a) [1 point] This test concerns the parameter (circle one)

- **(i)** $\theta = \mu_1 - \frac{1}{2}(\mu_2 + \mu_3)$
- **(ii)** $\theta = \mu_2 - \frac{1}{2}(\mu_1 + \mu_3)$
- **(iii)** $\theta = \mu_2 + \frac{1}{2}(\mu_1 + \mu_3)$
- **(iv)** $\theta = \mu_2 - \frac{1}{2}(\mu_2 - \mu_3)$
- **(v)** $\theta = \mu_3 - \frac{1}{2}(\mu_1 + \mu_2)$

(b) [1 point] $\hat{\theta} =$ ______________________________.

10. Consider the following table of salinity (milligrams) versus depth of water (meters).

depth of water, x	13	54	73	111	132	154	178	231	254	288
salinity, y	13	11.2	7.5	4.8	3.1	3.4	2.0	2.2	1.0	0.5

(a) [1 point] A Bonferroni 95% confidence interval (calculated simultaneously with seven other CIs) for the expected response at $x = 13$

is given by ____.

(b) [1 point] A Bonferroni 95% prediction interval (calculated simultaneously with seven other CIs) for the expected response at $x = 13$ for the mean of four (4) future observations

is given by ____.

(1) h, d, e, c, g, a, f

(2) (a) **0.0017**; (b) **0.94** (c) (v) $n \leq 100$ and $n\pi \leq 10$

(3) (a) 0.0064; (b) 7.068 (c) **more right skewed than**.

(4) (a) (ii) t distribution with 14 degrees of freedom; (b) 0.154; (c) (15.999, ∞).

(5) (a) **70.7**; (b) **8.41**.

(6) (a) (iv) decrease the variance in the mean response for each treatment,
(v) increase the reliability of statistical inference;
(b) 01, . . . , 09 in Salt L; 10, . . . , 18 Salt M; 19, . . . , 27 Salt H;
(c) 2, 4.927, 2.464; 24, 38.238, 1.593; 27, 43.165
(d) 2, 31.201, 15.600; 24, 11.964, 0.499; 26, 43.165
(e) **normality**
(f) $\boldsymbol{\sigma^2 = k}$

(7) 0.52

(8) **True**, **False**, **True**, **True**

(9) (a) (ii) $\theta = \mu_2 - \frac{1}{2}(\mu_1 + \mu_3)$; (b) $\mathbf{-1.50}$.

(10) (a) $\mathbf{(6.76, 14.88)}$; (b) $\mathbf{(5.51, 16.13)}$.