1
20
18
34
24
96

Biostatistics Midterm 2 2012 ~~116~~ 96 pts

I. Give (1) the NAME of the sampling distribution, (2) the null and alternate hypotheses using symbols where appropriate, and (3) the test statistics (formulas) for the following tests: 20 pts

1. To test whether each mean differs from each other mean following a significant F in Anova

Tukey (q) 1

$$q = \frac{\bar{X}_j - \bar{X}_i}{SE}$$ 2

here SE is not $S_{\bar{X}_A - \bar{X}_B}$

$H_0: \mu_i = \mu_j$

$H_1: \mu_i \neq \mu_j$ 1

it is based on MSerror

2. To test two groups for heteroscedasticity

F

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

you should use σ^2 in hypotheses not σ

$$F = \frac{s_1^2}{s_2^2}$$

3. To test for the random factor in a random block Anova

F

Ho: no variation among blocks, blocking not effective

H1: variation among blocks; blocking effective

$$F = \frac{MS_{blocks}}{MS_{error}}$$

if give wrong one (Fixed) − 2

4. According to the fossil record the average extinction rate for mammals is .005 species per year. In the last century 26 mammals have become extinct. Is this number unusually high?

Poisson

H_o: $\mu \leq .005/yr$
H_1: $\mu > .005/yr$ or $\mu > .005(100) = .5$

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$26 = x$

5. To test whether caffeine decreases blood calcium given two independent groups

t

H_o: $\mu_c \geq \mu_{nc}$

H_1: $\mu_c < \mu_{nc}$

$$t = \frac{\bar{X}_c - \bar{X}_{nc}}{S_{\bar{X}_c - \bar{X}_{nc}}}$$

II. For each of the following state the type of test. Assume that the original data is available for analysis and that a transformation will be performed if necessary. Some tests may be used more than once and some not at all. 18 pts

1 sample Z(X)	Unpaired t	Single classification Anova
1 sample Z($\bar{X}$)	Paired t	Random block Anova
1 sample t($\bar{X}$)	2 sample F-test	Nested (hierarchical) Anova
1 sample Binomial, Poisson		Factorial (two way with rep.)

1. A study was performed to measure the sublethal effects of acute nitrite toxicity on the fathead minnow in the laboratory. There were four artificial streams- two were with levels found in unpolluted conditions, 0.004mg/l nitrite, and two with 20mg/l nitrite- each stream contained twenty fish. One parameter measured was the rate of respiration (measured as opercular movements per minute) of each of the 80 fish.

   Nested   2 different streams in each Tr

2. In study of methods of increasing egg production in an endangered species of lizard investigators supplemented the natural diet of 10 lizards and left 10 lizards as a control. They then counted the number of eggs produced by each lizard (20 data points) to determine whether supplementary feeding increases egg production.   assume transformed, 1-way ANOVA but indicate $p \div 2$

   Independent (unpaired t)

3. A study was performed to test the hypothesis that melanic squirrels are predominate in dense forests and grey squirrels in open areas. Ten locations were used and in each location a forest and a field were trapped and the percentage of melanic squirrels in each were measured. (This gives a total of 20 percentages.)

   Paired t (paired by area) (assume transformed) 1-way ANOVA but $p \div 2$

4. In an effort to make a trap for the Colorado potato beetle that would not trap honeybees, researchers experimented with sticky traps of four different colors. They used six traps of each color and counted the number of insects stuck on each trap. The beetles and honeybees were tested separately. The data are thus 24 observations for each species.

   Factoral   2-way with replication (n=2)

5. A study was performed to compare the effects of two different diets on the protein content of the dorsal root ganglion in rats. Twenty rats were placed on each diet and the protein content was measured at 3 times for each rat for a total number of 120 data points.

   Nested   20 different rats on each diet

6. The mean baseline level of LDL of diabetic patients is known to be 127.6 mg/dl. Two hundred such patients on a particular antidiabetic had a LDL level of 109.2 mg/dl with a standard deviation of 25. Does the drug lower LDL?

   $t(\bar{X})$ you do not have a σ

III. Answer briefly. ~~**40**~~ 34 **pts**

1. Compare the relative power of multiple t, Tukey and Dunnett tests when you have 3 groups. Show the relative alpha values either for the whole set or in each individual test. 8 pts

I didn't ask you to explain

multiple t > Dunnetts > Tukey 2 pts

for any 1 comparison $\alpha = 0.05$ $\alpha = \frac{0.05}{2}$ $\alpha = \frac{0.05}{3}$

2 pts 2 pts 2 pts

overwhole $\alpha = 0.14$ .05 .05

2. What is the predicted value of F_p if Ho is true? If it is false? Explain in terms of expected mean squares. 4pts

true $F_p = 1$ 1 pt

False $F_p > 1$ 1 pt

adds to variance

if 0 F=1

if >0 F>1

$$\frac{\text{Ems groups}}{\text{Ems error}} = \frac{\sigma^2 + \sigma^2_{gr}}{\sigma^2} = F$$

3. You have 2 groups. If you compare them with Anova how many tails will you use to look up the probabilities? Explain. 3pts

1 tail ~~1~~ pts

H_1: $F_p > 1$ ~~1~~ pts so H_1 is directional

4. If you test exactly the same hypothesis as in #3 by a t test instead of Anova, how many tails will it have? What is the numerical relationship between the two test statistics? 3pts

2 1½

$t^2 = F$ 1½

pg 18 pts

5. Show the calculation of the error sum of squares in a one way and a random block Anova (using other sums of squares). 2 pts

1 1way SS TOTAL - SSgroups

1 RB SS TOTAL - SSgroups(treat) - SSblocks

6. Based on the above demonstrate why the random block Anova is likely to be more powerful than a one way Anova. Why might it not be more powerful? 6pts

2 SS error smaller in RB $MS_{error} = \frac{SS\ error}{df_{error}}$

a smaller SS error likely means smaller MS_{error}

2 $F = \frac{MS_{gr(tr)}}{MS_{error}}$ ↓ denominator ↑ F

2 but lower df in RB means MS_{error} not smaller

7. Are the following likely to be suitable for analysis by parametric tests such as Anova and t tests? Explain why or why not including the assumptions of these tests and which if any are likely to be violated. If not suitable how could you correct them? 8pts

a. You wish to determine the effect of food addition and food restriction on lifespan in the wild of the endangered channel island foxes.

2 lifespan + skewed ∴ violates normality (maybe homoscedasticity + additivity)

2 try a log transformation maybe square root note it is not counted

b. You wish to compare the proportion of seeds germinating with two different concentrations of nitrate.

2 proportions binomial + likely to violate normality maybe homo + additivity

2 try arc sine √p

pg 16

IV. An experiment was performed to measure the effect of low levels of hydrocarbon pollution on species diversity in aquatic mesocosms. They used 3 mesocosms with and three without pollution. Four samples were taken in each mesocosm and the number of species in each sample determined (24 data points).

Source of variation	SS	Df	MS	F calc	Fcrit $F_{.05(1)}$	p
Treatment	15.44	1 1	15.44 1	3.328 1	7.71 1/2	.1<p<.25 1/2
Mesocosm	18.56	4 1	4.64 1	3.7591	2.93 1/2	.01<p<0.25 1/2
Error	22.22 1	18 1	1.23 1			
Total	56.22	23 1				

1. What kind of Anova is this? 1pt
   Nested

-4 for block ANOVA correctly done

2. Fill in the table above. (Assume that a transformation has been performed if necessary.) Show your work in the table. 12 pts

3. State the level of significance. 1pt $\alpha = 0.05$

4. State all sets of hypotheses using symbols where appropriate. 4 pts

   Ho: $\mu_{HC} = \mu_{none}$
   H1: $\mu_{HC} \neq \mu_{none}$
   Ho no variation among mesocosm within same treatment
   H1 variation among mesocosms ...

5. State the statistical conclusions for each set of hypotheses and verbalize. 6pts

   treatment Fcalc < Fcrit p > 0.05 Fail to reject
   no significant variation among mesocosms

   Mesocosms Fcalc > Fcrit p < 0.05 reject
   significant variation among mesocosms within treatment