Solutions to Final STAT 421, Fall 2007

Fritz Scholz

With points as indicated in ( ) for each question. Total = 200

1. (8) Two treatments A and B were randomly assigned to 6 subjects (3 subjects to each treatment) with the following responses: $1, 3, 6$ and $5, 7, 9$ for treatments A and B, respectively. Based on the randomization reference distribution find the p-value of the two-sided test of no treatment effect when rejecting for large values of $|\bar{X}_B - \bar{X}_A|$. (Hint: you don't need to compute the full randomization reference distribution!)

There are $\binom{6}{3} = 6 \cdot 5 \cdot 4/(1 \cdot 2 \cdot 3) = 20$ possible splits of the pooled ordered sample $1, 3, 5, 6, 7, 9$. Of these splits the largest value for $|\bar{X}_B - \bar{X}_A|$ is $|9/3 - 22/3| = 13/3$ and it occurs for the split $1, 3, 5$ and $6, 7, 9$ (and the reverse). The next largest value of $|\bar{X}_B - \bar{X}_A|$, i.e., $|10/3 - 21/3| = 11/3$, is obtained for the split $1, 3, 6$ and $5, 7, 9$ (and the reverse) (interchanging 5 and 6 in the previous most extreme split). Thus the p-value for the above observed sample split $1, 3, 6$ and $5, 7, 9$, i.e., the chance of observing a value $\geq 11/3$, is $4/20 = 1/5 = .2$, not very significant.

2. (8) Which of the following vectors $\mathbf{c} = (c_1, \ldots, c_t)$ would define proper contrasts in a 1-way ANOVA with $t = 6$ treatment levels?

Circle the correct answer Y or N for yes (proper contrast) or no, respectively.

a) `c1=c(0,0,1,-.5,0,.5)` (Y or N), **N**

b) `c2=c(-1,0,1,-.5,0,.5)` (Y or N), **Y**

c) `c3=c(-1/6,0,-1/6,0,-1/6,1/2)` (Y or N), **Y**

d) `c4=c(0,-1/6,1/3,-1/6,-1/3,0)` (Y or N), **N** .

3. (8) The testing of which hypothesis is addressed by the modified Levine test and how does it link up with a 1-way ANOVA?

The Levene test is used to test the homogeneity of variances over several groups of data (as assumed in the typical ANOVA situation). It treats the absolute deviations of the observations from their respective group medians as new observations and performs an ANOVA for equality of means (expressing roughly equal variability) of these new groups of observations.

4. (4) If $\texttt{x} = \texttt{1} : \texttt{5}$ what are the responses to the following two R commands?

```
> x <= 3
```

```
> x[-c(1,3)]
```

`TRUE TRUE TRUE FALSE FALSE` and `2 4 5`, respectively.

5. (4) What is the response to the following R command?

```
> c(8,6,4,4)/c(2,2,1,2)
```

`4 3 4 2`

6. (6) What is the output of the following R function:

```
mystery.function=function(n){
out=NULL
for(i in 1:n){
   out=c(out,rep(i,i))
}
out
}
```

when you call `mystery.function(3)`? `1 2 2 3 3 3`

7. (6) If a random variable X has mean 0 and variance 25 what is the expectation of $(X-1)^2$, i.e., $E((X-1)^2) =$?

$E((X-1)^2) = E(X^2-2X+1) = E(X^2)-2E(X)+1 = E(X^2)+1 = \text{var}(X)+1 = 26.$

8. (8) Assume that for three samples of respective sizes $n_1 = 4$, $n_2 = 4$ and $n_3 = 8$ you are given the pooled standard deviation estimate $s = 2$. What is the standard error (SE) of the contrast estimate $\hat{C} = \frac{1}{2}\bar{X}_{1\bullet} + \frac{1}{2}\bar{X}_{2\bullet} - \bar{X}_{3\bullet}$? Also state which contrast is being estimated.

$$SE = s \times \sqrt{\frac{(1/2)^2}{4} + \frac{(1/2)^2}{4} + \frac{(-1)^2}{8}} = s \times \sqrt{\frac{1}{16} + \frac{1}{16} + \frac{1}{8}} = \frac{s}{2} = 1$$

and the contrast being estimated is: $C = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3$.

9. (8) For the 2-sample Student-t based confidence interval for $\Delta = \mu_Y - \mu_X$ show how you would convert a 95% confidence interval $[A, B] = [.5, 3.5]$ into a 99% confidence interval, if I give you the value for the ratio $t_{.995,m+n-2}/t_{.975,m+n-2} = \texttt{qt}(.995, \texttt{m}+\texttt{n}-2)/\texttt{qt}(.975, \texttt{m}+\texttt{n}-2) = 1.35$. Would this change the conclusion when testing the hypothesis $H_0 : \Delta = 0$ at the two levels $\alpha = .05$ and $\alpha = .01$? Justify your answer.

The 95% t-based interval is

$$\bar{Y} - \bar{X} \pm t_{.975,m+n-2}\; s\sqrt{\frac{1}{m} + \frac{1}{n}} = \frac{A+B}{2} \pm \frac{B-A}{2}$$

The 99% t-based interval is

$$\bar{Y} - \bar{X} \pm t_{.995,m+n-2}\; s\sqrt{\frac{1}{m} + \frac{1}{n}} = \frac{A+B}{2} \pm \frac{t_{.995,m+n-2}}{t_{.975,m+n-2}} \frac{B-A}{2} = 2 \pm 1.35 \times 1.5 = 2 \pm 2.025.$$

Since the first interval does not contain zero we would reject H_0 at level $\alpha = .05$, but at level $\alpha = .01$ we would not reject H_0 since the 99% interval contains zero.

10. (12) Give the definitions of the noncentral t, the noncentral χ^2, and the noncentral F, respectively. State these definitions in terms of functions of more basic random variables (name the random variables together with the appropriate parameters and relationships to each other, and give the relevant function).

Let $Z \sim \mathcal{N}(\delta, 1)$ and $V \sim \chi^2_f$ (central chi-square with f degrees of freedom) be independent, then $Z/\sqrt{V/f}$ is said to have a noncentral t-distribution with f degrees of freedom and noncentrality parameter δ.

Let $Z_i \sim \mathcal{N}(\mu_i, 1),\ i = 1, \ldots, f$ be independent, then $Z_1^2 + \ldots + Z_f^2$ is said to have a noncentral χ^2 distribution with f degrees of freedom and noncentrality parameter $\lambda = \mu_1^2 + \ldots + \mu_f^2$.

Let $V_1 \sim$ noncentral χ^2 distribution with f_1 degrees of freedom and noncentrality parameter λ and $V_2 \sim$ central χ^2 distribution with f_2 degrees of freedom, with V_1 and V_2 independent, then $(V_1/f_1)/(V_2/f_1)$ is said to have a noncentral F distribution with f_1 (numerator) and f_2 (denominator) degrees of freedom and with noncentrality parameter λ.

11. (8) In a 1-way ANOVA, comparing means for k groups at significance level $\alpha = .05$, what happens to the type II error probability when

a) the common σ for all k groups becomes small?

b) when the first $k-1$ group means are the same, say $= \mu_0$, and the k^{th} mean μ_k approaches μ_0? (Think of the noncentrality parameter).

a) The power function $\beta(\lambda)$ is a continuous and monotone increasing function of the noncentrality parameter $\lambda = \sum n_i(\mu_i - \bar{\mu})^2/\sigma^2 \nearrow$ as $\sigma \searrow$. Hence the type II error $1-\beta(\lambda) \searrow$ as $\sigma \searrow$ for any $\lambda > 0$.
b) As $\mu_k \to \mu_0$ all μ_i approach the same value μ_0, thus $\lambda \to 0$ and thus $1-\beta(\lambda) \to 1-\beta(0) = 1-\alpha = .95$.

12. (9) In a 1-way ANOVA with t treatments and n observations per treatment indicate your answer for a)-c) by True, False or Undecidable (without additional information) using T, F, U. Assume $\bar{Y}_{i\bullet} = \sum_{j=1}^{n} Y_{ij}/n$.

a) $\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i\bullet}) = 0$ **T**

b) $\sum_{i=1}^{t}(Y_{ij} - \bar{Y}_{i\bullet}) = 0$ **U**

c) $\sum_{j=1}^{n}\sum_{i=1}^{t}(Y_{ij} - \bar{Y}_{i\bullet}) = \sum_{i=1}^{t}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i\bullet}) = 0$ from a) **T**

13. (6) In a 1-way ANOVA with $t = 5$ treatments, if we add 10 to the population means for all t treatments, how does that affect the value of any given contrast? $C_{10} = \sum_i c_i(\mu_i + 10) = \sum_i c_i\mu_i + 10\sum_i c_i = \sum_i c_i\mu_i = C$, i.e., the contrast does not change since $\sum_i c_i = 0$ by definition of a contrast.

14. (18) The following table gives the responses Y_{ijk} from a 2 factor experiment with 3 observations per factor level combination. Give the 2×3 table of $\bar{Y}_{ij\bullet}$, $i = 1, 2$, $j = 1, 2, 3$ (leave answers in fractional form). Also give $\bar{Y}_{1\bullet\bullet}, \bar{Y}_{2\bullet\bullet}$, $\bar{Y}_{\bullet 1\bullet}, \bar{Y}_{\bullet 2\bullet}, \bar{Y}_{\bullet 3\bullet}$, and $\bar{Y}_{\bullet\bullet\bullet}$.

Data Table

		Factor 2		
	level	1	2	3
Factor 1	1	1, 4, 3	2, 2, 4	3, 1, 2
	2	6, 1, 2	3, 2, 1	4, 5, 2

Table of $Y_{ij\bullet}$

		Factor 2		
	level	1	2	3
Factor 1	1	8/3	8/3	6/3
	2	9/3	6/3	11/3

$\bar{Y}_{1\bullet\bullet} = 22/9$, $\bar{Y}_{2\bullet\bullet} = 26/9$, $\bar{Y}_{\bullet 1\bullet} = 17/6$, $\bar{Y}_{\bullet 2\bullet} = 14/6$, $\bar{Y}_{\bullet 3\bullet} = 17/6$, $\bar{Y}_{\bullet\bullet\bullet} = 48/18 = 8/3$.

15. (8) When comparing $k = 5$ independent normal samples with equal variances, we can compute confidence intervals for **how many pairwise** differences of means $\mu_i - \mu_j$, $i < j$? Using Bonferroni's method, how would you adjust the confidence levels of these individual confidence intervals so that they simultaneously cover their respective targets $\mu_i - \mu_j$ with probability $\geq .95$.

There are $\binom{5}{2} = 10$ possible paired differences. Thus the individual confidence levels need to be adjusted to $1 - \alpha/10 = 1 - .05/10 = .995$.

16. (6) Suppose in a 1-way ANOVA with t treatments we consider the full model $Y_{ij} = \mu_i + \epsilon_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, t$ with ϵ_{ij} being independent $\mathcal{N}(0, \sigma^2)$. What is the form of the reduced model?

The reduced model is: $Y_{ij} = \mu + \epsilon_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, t$ with ϵ_{ij} being independent $\mathcal{N}(0, \sigma^2)$.

17. (12) In a 2-way (2 factor) ANOVA rewrite the Sum of Squares Decomposition $SS_T = SS_A + SS_B + SS_{AB} + SS_E$ in terms of the responses Y_{ijk} using the dot-bar notation, e.g., $\bar{Y}_{i\cdot\cdot}$, etc.

$$\sum_{ijk}(Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2 = \sum_{ijk}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 + \sum_{ijk}(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 + \sum_{ijk}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2 + \sum_{ijk}(Y_{ijk} - \bar{Y}_{ij\cdot})^2$$

18. (10) The `Fmin.test` uses which test statistic T to test which hypothesis H_0? Which values of T would tend to speak against H_0 and explain why.

The test statistic is $T = \min(s_1^2, \ldots, s_t^2)/\max(s_1^2, \ldots, s_t^2) \leq 1$ and it is used to test the hypothesis $H_0 : \sigma_1^2 = \ldots = \sigma_t^2$.
Under H_0 we have $\tau = \min(\sigma_1^2, \ldots, \sigma_t^2)/\max(\sigma_1^2, \ldots, \sigma_t^2) = 1$. Since T is a reasonable estimate of $\tau \leq 1$ we can view small values of T as evidence against H_0.

19. (6) In the context of an experiment with two factors A and B with t_1 and t_2 levels, respectively, and with n replications of the response Y per factor level combination, what is the difference in performed analyses in the following two commands? `anova(lm(Y ~ A*B))` and `anova(lm(Y ~ A+B))`?

Under the first command we assume the full model and we test for main effects and interactions being zero, respectively. Under the second command we assume an additive model (no interactions). We can then only test for the two main effects being zero, respectively.

20. (10) In the full 2-factor ANOVA model $\mu_{ij} = \mu + a_i + b_j + c_{ij}$ (equal number $n > 1$ of replications $Y_{ijk}, k = 1, \ldots, n$ per factor level combination (i,j)) which type of constraints were used in the derivation of the least squares estimates? Give these least squares estimates of a_i, b_j, and c_{ij}.

The zero-sum constraints were used in deriving the following least squares estimates

$$\hat{a}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}\ , \quad \hat{b}_j = \bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdot\cdot\cdot}\ , \quad \text{and} \quad \hat{c}_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot}\ .$$

21. (10) In the full 2-factor ANOVA model $\mu_{ij} = \mu + a_i + b_j + c_{ij}$ ($i = 1, \ldots, t_1$, $j = 1, \ldots, t_2$, and with equal number $n > 1$ of replications per factor level combination) what is the null distribution of the F-test for testing $H_0 : b_1 = \ldots = b_{t_2}$? What is the null distribution of the F-test for testing $H_0 : c_{ij} = 0 \ i = 1, \ldots, t_1, \ j = 1, \ldots, t_2$?

For $H_0 : b_1 = \ldots = b_{t_2}$ the null distribution of the F-test is $F_{t_2-1, t_1 t_2 (n-1)}$ while for $H_0 : c_{ij} = 0 \ \forall i, j$ the null distribution of the F-test is $F_{(t_1-1)(t_2-1), t_1 t_2 (n-1)}$.

22. (10) In the reduced or additive ANOVA model $\mu_{ij} = \mu + a_i + b_j$ with just one observation Y_{ij} per factor level combination (i,j), what is the denominator of the F-test statistic for $H_0 : a_1 = \ldots = a_{t_1}$?

$$\text{The denominator is} \qquad \frac{\sum_{i=1}^{t_1} \sum_{j=1}^{t_2} (Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2}{(t_1 - 1)(t_2 - 1)}$$

23. (10) What is the mean of a noncentral chi-square random variable with f degrees of freedom and noncentrality parameter λ.

$$\text{With } Z_i \sim \mathcal{N}(0,1) \quad E(\sum_{i=1}^{f} (Z_i + \mu_i)^2) = \sum_{i=1}^{f} E(Z_i^2 + 2\mu_i Z_i + \mu_i^2) = \sum_{i=1}^{f} (1 + \mu_i^2) = f + \lambda.$$

24. (5) What kind of plot would `qqplot(c(1,3,5),c(4,6,8))` produce and what would it suggest as a diagnostic about the two samples of size 3 when comparing it to the main diagonal?

The plot would show 3 points appearing in a linear pattern, shifted 3 units up and parallel to the main diagonal. This would suggest that the second sample is shifted 3 units to the right of the first sample (which is exactly the case).