Memorial University of Newfoundland

PhD in Statistics Comprehensive Examination

Exam 3: Design of Experiments and Survey Sampling   August 26, 2022

---

1. In an experiment to assess the effects of curing time (factor A) and type of mix (factor B) on the compressive strength of hardened cement cubes, three (3) different curing times were used in combination with four (4) different mixes, with three (3) observations obtained for each of the 12 curing time–mix combinations. The resulting sums of squares were computed to be $SSA = 30.8$, $SSB = 34.2$, $SSE = 97.4$, and $SST = 206$.

a. Write down an ANOVA type model and relevant model assumptions for the analysis of the data.

b. Construct an ANOVA table.

c. Test at level .05 the null hypothesis H_{0AB}: all $\gamma'_{ij}s = 0$ (no interaction of factors) against H_{1AB}: at least one $\gamma_{ij} \neq 0$.

d. Test at level .05 the null hypothesis H_{0A}: $\alpha_1 = \alpha_2 = \alpha_3 = 0$ (factor A main effects are absent) against H_{1A}: at least one $\alpha_i \neq 0$.

e. Test at level .05 the null hypothesis H_{0B}: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (factor B main effects are absent) against H_{1B}: at least one $\beta_i \neq 0$.

2. A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. There might be variability from one bolt to another. The chemist decides to use a completely randomized design (CRD). However, since her statistical knowledge is limited, she decided to ask your opinion.

a. Explain her why a randomized block design, with the bolts of cloth considered as blocks, would be a better design than a CRD for this study.

b. After your explanation she decides to apply a randomized complete block design (RCBD). She selected five bolts and applied all four chemicals in random order to each bolt. The resulting tensile strengths are given in the data table. Analyze the data from this experiment (use $\alpha = 0.10$) and draw appropriate conclusions. An R output for this experiment is given below.

c. Does the result of the analysis support your opinion explained in [a.]? (Use $\alpha = 0.05$).

d. Use Tukey's test to compare pairs of treatment means. (Use $\alpha = 0.05$). –See formula sheet as reference.

The Data Table

	Bolt (Block)						
Chemical	1	2	3	4	5	Totals	Averages
1	73	68	74	71	67	353	70.6
2	73	67	75	72	70	357	71.4
3	75	68	78	73	68	362	72.4
4	73	75	75	77	73	373	74.6
Totals	294	278	302	293	278	1445	
Averages	73.5	69.5	75.5	73.25	69.5	72.25	

The ANOVA Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
chem	3	44.95	14.98	3.610	0.04575 *
block	4	113.00	28.25		
Resid.	12	49.80	4.15		

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

3. A sample $y_1, y_2, \cdots, y_n$ is obtained from a finite population of size N. Assume that the inclusion probability of element i is π_i and the joint inclusion probability of elements i and j is π_{ij}. The Horvitz-Thompson estimator of the population total is

$$\widehat{\tau}_\pi = \sum_{i=1}^{n} \frac{y_i}{\pi_i},$$

where n is the sample size.

a. Prove that the Horvitz-Thompson estimator of the population total is design-unbiased. Hint: Let $Z_i = 1$ if the ith element is included in the sample.

b. Derive the variance of $\widehat{\tau}_\pi$.

c Find a design-unbiased estimator of this variance.

4. A population contains N units. A simple random sample of size n is selected, denoted $y_1, y_2, \cdots, y_n$. A subsample (simple random sample, second stage sample) of size n_1 is then selected from the first stage sample with $n_2 = n - n_1$ units remaining . Let $\bar{y}$ be the sample mean of the first stage sample, $\bar{y}_1$ the sample mean of the second stage sample, and $\bar{y}_2$ be the mean of the remaining n_2 units.

a. Find the expectation of $\bar{y}_1$ and $\bar{y}_2$

b. Find the variance of $\bar{y}_1 - \bar{y}_2$

Formula sheet

1. R functions related to F-distribution: $pf(x, d_1, d_2) = P(X \leq x)$ where $X \sim F(d_1, d_2)$
   $qf(\alpha, d_1, d_2) = q$ such that $P(X \leq q) = \alpha$.
2. R functions for Student's t distribution: $pt(x, d) = P(X \leq x)$, where $X \sim t(d)$
   $qt(\alpha, d) = q$ such that $P(X \leq q) = \alpha$.
3. R function for χ^2 distribution: $qchisq(\alpha, d) = q$ such that $P(\chi^2 \leq q) = \alpha$.
4. R function for standard normal distribution: $pnorm(x) = P(Z \leq x)$.
5. Values related to the above distributions:
   χ^2 distribution: qchisq(0.05, c(2,4,6,8))= (0.1025866, 0.7107230, 1.6353829, 2.7326368)
   qchisq(0.95, c(2,4,6,8)) = (5.991465, 9.487729, 12.591587, 15.507313)

```
qt(0.025,c(7,16,17)) = ( -2.364624 -2.119905 -2.109816),

qt(0.05,c(7,16,17)) = (-1.894579 -1.745884 -1.739607)

F distribution:
 > qf(.05, c(3,4,5), 12)
[1] 0.1143558 0.1691552 0.2137801
> qf(.05, c(3,4,5), 16)
[1] 0.1150445 0.1711122 0.2172136
> qf(.05, c(3,4,5), 20)
[1] 0.1154709 0.1723383 0.2193881
> qf(.95, c(3,4,5), 12)
[1] 3.490295 3.259167 3.105875
> qf(.95, c(3,4,5), 16)
[1] 3.238872 3.006917 2.852409
> qf(.95, c(3,4,5), 20)
[1] 3.098391 2.866081 2.710890
```

Tukey's test statistic comparing treatment means

$$T_\alpha = q_\alpha(a, f)\sqrt{\frac{MSE}{b}},$$

where a is the number of treatments and b the number of blocks and f is the degree of freedom of the MSE. Tukey's test statistic comparing block means

$$T_\alpha = q_\alpha(b, f)\sqrt{\frac{MSE}{a}}.$$

A few values of the $q_{0.05}(\cdot, \cdot)$: $q_{0.05}(4, (12, 16, 20)) = (4.20, 4.05, 3.96)$
$q_{0.05}(5, (12, 16, 20)) = (4.51, 4.33, 4.23)$