

Stat {4601, 6872} Final (90 minutes)

#1. (60 pt) A number of welders (total of 6 by 5) were asked to weld two pipes together using 5 different welding torches. Each welder was assigned a single torch and performed a single weld using the torch. Finished pipes were measured on a variety of quality factors, and rated from 1 to 10, where 10 represents a perfect weld. We are interested if different torch has a significant performance difference as measured by the quality factor. The data are as follows. (For computational convenience, the last three rows are striken through; do not use them in your computation)

Welder	Torch1	Torch2	Torch3	Torch4	Torch5
1	3.9	4.1	4.2	4.1	3.3
2	9.4	9.5	9.4	9.0	8.6
3	9.7	9.3	9.3	9.2	8.4
4	8.3	8.0	7.9	8.6	7.4
5	9.8	8.9	9.0	9.0	8.3
6	9.9	10.0	9.7	9.6	9.1

- To randomize properly, how should the experiment be performed?
- What's the relevant nonparametric test? Define the model formally/mathematically (using X_i , Y_i , μ , Δ etc) and state the necessary assumptions to apply the nonparametric test in a)
- Formally state the hypothesis (e.g. $H_0: \mu=0$ vs $H_1: \mu > 0$ where μ is the median ...). Compute the relevant nonparametric test statistic and compute the P-value.
- What's the procedure for selecting a "best" torch in the 5 from such results called?
- If the parametric and nonparametric tests give different answers, with a very small P-value (say 0.001) for the nonparametric test and a very large P-value (say 0.20) for the parametric test, what could be possible explanations of the reasons why? In such case, what would you do to clarify the issue? Finally, which test result would you choose?
- What test would you use if the above experiment was actually done for "Six welders with different expertise", in which case a welder performed five welds each time with a different torch? How would you randomize such experiment?

#2. (40 pt) A claim has been made that there are more people of blood type 'B' among CEOs than general (non-CEO) population. To confirm the conjecture, the following data has been collected:

	Blood Type	
	Non-B	B
CEO	57	36
Non-CEO	70	30

- State the assumptions and the null-hypothesis for a relevant test.
- For the alternative hypothesis: "There is a difference between the two proportions (of those with blood type B among CEOs and among non-CEOs)", compute the test statistic and the P-value.
- For the alternative hypothesis: "The proportion of those with blood type B is greater among CEOs than among non-CEOs", compute the test statistic and the P-value.
- If you sample more Non-CEOs (there are a lot more of them...), then how do you expect the above P-values to become? (It's convenient to assume the proportion of those with blood type B will stay the same (30%) among non-CEOs as you sample more, e.g. (700 and 300), (7,000 and 3,000) and so on...) Will they get smaller or larger? Justify your answer.

#3. (20 pt) Answer the following questions.

- What procedures can one use to check if using a parametric two-sample t-test is OK compared to nonparametric Wilcoxon rank sum test?
- What is data snooping?

Final Solution

#1. (50 pt) Two-way version:

- Make sure torches are used in random order for each welder.
- Friedman test. $X_{ijt} = \theta + \beta_i + \tau_j + e_{ijt}$ are all mutually independent; $e_{ijt} \sim \text{iid}$ continuous F
- $H_0: [\tau_j = \dots = \tau_k]$ vs $H_1: [\tau_j, \dots, \tau_k \text{ not all equal}]$
- Multiple comparison (or Jonckheere Terpstra test)
- Causes: Outliers; Departure from the normal assumption; small sample.

Solutions: Plotting. Formal investigation. Using the nonparametric result is recommended.

	2	3.5	5	3.5	1
	3.5	5	3.5	2	1
	5	3.5	3.5	2	1
R	10.5	12	12	7.5	3
R.	3.5	4	4	2.5	1

$$S = \frac{12n}{k(k+1)} \sum_{j=1}^k \left(R_{.j} - \frac{k+1}{2} \right)^2$$

$$S = 12 \cdot 3 / (5 \cdot 6) * ((3.5-3)^2 + (4-3)^2 + (4-3)^2 + (2.5-3)^2 + (1.5-3)^2)$$

$$= 5.7$$

```
friedman.test(read.table('welder.dat'), exact=TRUE)
```

```
1-pchisq(12.9, 4)
```

Use table A.22 $k=5$, $n=3$, to get $P > .117$ (Using approximate $\chi^2(4)$, the P-value is 0.11)

One-way version

- Make sure the welders-torch combination is random and the welding is done in a random order.
- Kruskal test. $X_{ij} = \tau_j + e_{ij}$, $i=1, \dots, n_i$, $j=1, \dots, k$ are all mutually independent; $e_{ij} \sim \text{continuous F}$
- $H_0: [\tau_j = \dots = \tau_k]$ vs $H_1: [\tau_j, \dots, \tau_k \text{ not all equal}]$

	2	3.5	5	3.5	1
	12.5	14	12.5	8	7
	15	10.5	10.5	9	6
R	29.5	28	28	20.5	14
R.	9.83	9.33	9.33	6.83	4.67

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k n_j \left(R_{.j} - \frac{N+1}{2} \right)^2$$

$$H = 12 / (15 \cdot 16) * 3 * ((9.83-8)^2 + (9.33-8)^2 + (9.33-8)^2 + (6.83-8)^2 + (4.67-8)^2)$$

$$= 2.90$$

```
kruskal.test(read.table('welder.dat'), exact=TRUE)
```

```
1-pchisq(2.9, 4)
```

Use table A.12 to get $P > .1008$ (Using approximate $\chi^2(4)$, the P-value = .574)

#2.

- Two independent bernoulli trials.
- $A=1.27$ $P=.2024$
- $A=1.27$, $P=.1012$
- P-values will get smaller. (more sample; larger power.)

$$D = 36/93 - 30/100$$

$$\text{Phat} = (36+30) / (93+100)$$

$$\text{SD} = \sqrt{\text{Phat} * (1-\text{Phat}) * (1/100 + 1/93)}$$

$$A = D / \text{SD}$$

$$A = 1.2745$$

$$2 * (1 - \text{pnorm}(A))$$

$$0.2024$$

$$1 - \text{pnorm}(A)$$

$$0.1012$$

```
kruskal.test(read.table('welder.dat'))
```

#3.

- Visualization; normal Q-Q plot; Shapiro-Wilk test.
- Perform many tests until one gets significant p-value.