

Using robust statistics in Oracle 10g to find people and departments that consistently submit proposals late

Ken Geis

University of California, Berkeley

Data Requirements


■ Sources

- Proposal Log and (maybe) Institute Proposal
 - Date proposal arrived at SPO
 - Deadline date

■ Large sample


- six month window = ~ 1700 proposals
 - balances timeliness and sample size


UCB Proposal Log


Add Proposal Log 

Proposal Number: Pending

Proposal Type: New

Lead Investigator:  Geis, Ken


Lead Unit:  OSRAC RAC Research Administration & Compliance

Sponsor:  005400 NSF National Science Foundation

Title: Study on timeliness of proposal submissions

Comments:

SPO RA: SPO - Howard, Deborah **Deadline:** Received By 04/01/2008



```

SELECT lead_unit, unit_name, COUNT(*),
       MEDIAN(TRUNC(deadline_date, 'DD') + 17/24 - create_timestamp) AS median_lead,
       AVG(CASE WHEN (TRUNC(deadline_date, 'DD') + 17/24 - create_timestamp) < 2
                THEN 1 ELSE 0 END) AS proportion_late,
       STATS_BINOMIAL_TEST(CASE WHEN TRUNC(deadline_date, 'DD') + 17/24 -
                        create_timestamp < 2
                        THEN 1 ELSE 0 END,
                        1, others.proportion, 'ONE_SIDED_PROB_OR_MORE') AS p
FROM osp$proposal_log proposal_log, osp$unit units,
     (SELECT unit_number,
            AVG(CASE WHEN (TRUNC(deadline_date, 'DD') + 17/24 - create_timestamp) < 2
                     THEN 1 ELSE 0 END) AS proportion
     FROM osp$unit units, osp$proposal_log proposal_log,
          ucb$administrators administrators
     WHERE units.unit_number != proposal_log.lead_unit
           AND proposal_type_code != 6
           AND (administrator_id IS NULL OR administrator_id NOT IN (7, 17, 18, 19, 20))
           AND create_timestamp BETWEEN ADD_MONTHS(SYSDATE - ?, -6) AND (SYSDATE - ?)
     GROUP BY unit_number) others
WHERE proposal_log.lead_unit = others.unit_number
     AND proposal_log.lead_unit = units.unit_number (+)
     AND proposal_type_code != 6
     AND (administrator_id IS NULL OR administrator_id NOT IN (7, 17, 18, 19, 20))
     AND create_timestamp BETWEEN ADD_MONTHS(SYSDATE - ?, -6) AND (SYSDATE - ?)
GROUP BY lead_unit, unit_name, others.proportion
HAVING STATS_BINOMIAL_TEST(CASE WHEN TRUNC(deadline_date, 'DD') + 17/24 -
                        create_timestamp < 2 THEN 1 ELSE 0 END,
                        1, others.proportion, 'ONE_SIDED_PROB_OR_MORE') < 0.05

```

This is a list of units whose proposals submitted to SPO between 9/20/2007 and 3/20/2008 have a significantly ($p < 0.05$) different (better or worse) proportion of late proposals than the rest of the campus. This is based on the 2-day lead requested by Vice Chancellor for Research Beth Burnside. Following this page are backup pages with proposal logs for the departments that are consistently late.

Name	Proposal Count	Median Lead	Proportion Late
ERSO Engineering Research Support Organization	212	3.06	29.2%
Helen Wills Neuroscience Institute	44	1.29	43.2%
The Institute for Quantitative Biomedical Research (QB3)	101	3.68	31.7%
CPAM Center for Pure & Applied Mathematics	27	1.34	40.7%
Bioengineering	7	0.66	57.1%
LHS Lawrence Hall of Science	24	3.77	8.3%
Energy & Resources Group	16	5.21	6.2%
Berkeley Natural History Museums	16	6.09	6.2%
ITS Institute of Transportation Studies	16	4.36	6.2%
ESPM Division of Organisms and Environment	34	10.17	8.8%
Nutritional Sciences & Toxicology	29	7.21	6.9%
MCB Barker	42	6.34	2.4%