Sample: UNC ENT residency program applicants

# **Part 1**: female vs male difference in various metrics

Task 1: Check whether the t tests were done correctly.

Yes, but can be improved in two ways.

1. Alpha level: 66% false positive rate with 8 variables

* Bonferroni
* If *m* independent comparisons are performed, the [*family-wise error rate*](https://en.wikipedia.org/wiki/Family-wise_error_rate) (FWER), is given by
* {\displaystyle {\bar {\alpha }}=1-\left(1-\alpha \_{\{{\text{per comparison}}\}}\right)^{m}.}Text

  Description automatically generated
* **family-wise error rate** (**FWER**): the [probability](https://en.wikipedia.org/wiki/Probability) of making one or more false discoveries, or [type I errors](https://en.wikipedia.org/wiki/Type_I_and_type_II_errors) when performing [multiple hypotheses tests](https://en.wikipedia.org/wiki/Multiple_comparisons)

There are different ways to assure that the family-wise error rate is at most {\displaystyle {\bar {\alpha }}}. The most conservative method, which is free of dependence and distributional assumptions, is the [Bonferroni correction](https://en.wikipedia.org/wiki/Bonferroni_correction) {\displaystyle \alpha \_{\mathrm {\{per\ comparison\}} }={\alpha }/m}A picture containing text

Description automatically generated

Text

Description automatically generated

1. Correlation problem

* Logistic model: testing P+1 hypotheses
  + Q’s answered: Is any different? Which ones?

If (1) and (2) are the same 🡪 stronger conclusion

If different 🡪 further evaluation

Step 1: are M and F different (1 hypothesis)? F test

* Null: all are the same
* Alternative: at least one is different
* It’s incorrect to perform t test for each individual metric due to multiple testing, right?
* Should probably use False Discovery Rate. The input data need to have response and predictor variables. Theoretically, the response variable would be the match result (matched into the program or not), but I don’t have that data at this point. Is there anything I can do without a Y?

Graphical user interface, text, application, email

Description automatically generated

* 8 null hypotheses: Bonferroni (average Clout scores are the same) – upper bound to alpha, very restrictive
  + But Bonferroni does not take correlated variable into account
* Logistic regression with gender as Y: account for possible correlation among X’s
  + 8 null hypotheses
  + FDR function OR regular logistic function + Bonferroni

N < P: rank deficient matrix 🡪 can’t do modeling

If P is close to N: multiple testing correction is very important

P << N: as N increases 🡪 standard error decrease, standard error becomes more precise (p value will be much smaller if Ha is actually true)

The greater issue is correlation between predictors. Multiple testing isn’t as serious of an issue in this case.

\*Hypothesis

* Null hypothesis: there is no difference (in a given criteria) between male and female
* Alternate hypothesis: there is a difference

Objective metrics:

Step 1 Step 2 Number of Letters Peer Reviewed Journal Articles/Abstracts (Other than Published) Peer Reviewed Journal Articles/Abstracts Book Chapter (Peer Reviewed) Poster Presentation Oral Presentation Online Publication (Peer Reviewed) Non Peer Reviewed Online Publication Other Articles

Characteristics extracted from LOR/PS using a linguistic algorithm:

Standout Ability Grindstone Teaching Research Analytic Clout Authentic Tone

Other questions:

1. Effect size use?
2. Normality violation: Since n is large enough, it’s okay according to CLT.
3. PART 2 modeling:

Beta’s in logistic regression rely on Central Limit Theorem.

T-Test Assumptions

1. The scale of measurement. The scale of measurement applied to the data collected follows a **continuous** or ordinal scale, such as the scores for an IQ test.
2. **Simple random sample**: the data is collected from a **representative**, randomly selected portion of the total population.
3. The third assumption is the data, when plotted, results in a **normal distribution**, bell-shaped distribution curve. When a normal distribution is assumed, one can specify a level of probability (alpha level, level of significance, p) as a criterion for acceptance. In most cases, a 5% value can be assumed.
4. The fourth assumption is a **reasonably** **large sample** **size** is used. A larger sample size means the distribution of results should approach a normal bell-shaped curve.
5. The final assumption is homogeneity of variance. Homogeneous, or equal, variance exists when the standard deviations of samples are approximately equal.

# **Part 2**: substitute metric for step 1 score?

* Linear regression: looks like Step 2 score is the only predictor
* Possible transformations? Peer\_art\_abs, extra\_art\_abs, oral
* Possible to get combinations of variables to predict Step 1 score?
* +0.1 and log() to correct for variance OR sqrt
* Predicting step 1 score:
  + Regression (R=0.72)
  + Machine learning algorithm, use R^2 or MSE to evaluate performance

Calendar

Description automatically generated with medium confidence

* Y: step 1 score
* Transform all X variables except for step 2 score
* Exclude if have clear reasons to believe that it was erroneous (e.g. data reliability issue)
* Negative correlation
* +I(poster^2)

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated with medium confidenceA screenshot of a computer

Description automatically generated with low confidence