ASSESSING SMALL GROUPS

BRANCH OF EXPERT SERVICE MEMBERS:

TIMOTHY LI
RYAN PETERSON
ANDY LEU
BRET PHILLIPS
DAVID GARBER

NILG 2020 VIRTUAL CONFERENCE
WEBINAR SERIES

Thank YOU SPONSORS FOR YOUR SUPPORT

affirmity

AMERICA'S JOB exchange

berkshire

biddle opportunity. for all.

DCI CONSULTING

GA GAUCHER ASSOCIATES Affirmative Action Professionals

HUDSONMANN

localjobnetwork

LOCKHEED MARTIN

MERCK INVENTING FOR LIFE

NORTHROP GRUMMAN

PEPSICO

resolution economics LLC

taapestry by DirectEmployers

vizient

Dr. Robert (Bob) LaJeunesse

Dr. Robert (Bob) LaJeunesse is Acting Director of Enforcement for Office of Federal Contract Compliance. Bob oversees the OFCCP's statistical and economic analysis program and serves as the agency's expert technical advisor in the development and resolution of systemic discrimination cases. Before joining OFCCP, Bob was a labor economist at the Equal Employment Opportunity Commission (EEOC), working in the Office of General Council as a consulting and testifying expert on Title VII cases. He also served as an Economist for the U.S. Treasury and the AFL-CIO. Prior to joining federal service, Dr. LaJeunesse was an assistant professor of economics at the State University of New York (New Paltz) and a senior lecturer at the University of Newcastle in Australia. Bob's scholarly publications are primarily in the field of labor economics, including a book on the socioeconomic and ecological virtues of work time regulation. Bob holds a Ph.D. in Economics from Colorado State University, and he also completed two years of service in the Peace Corps in Liepaja, Latvia.

Bogong Li

Bogong Li is the Statistician with the DOL Office of Federal Contract Compliance Programs. He helps develop statistical methods, conduct routine compensation, hiring, and selection data analysis to ensure federal contractor compliance with labor regulation and laws. He also helped the agency design and implement a new contractor scheduling procedure, among many other methodological developments. Bogong has been recognized by National Director of OFCCP for outstanding defense of statistical analysis results during conciliation and settlement process with a number of cases for compensation and hiring discrimination charges.

Previously he worked at the DOL Bureau of Labor Statistics as a Mathematical Statistician, serving as expert in statistical modelling, survey sampling methods and disclosure limitation techniques. He later joined Bank of America serving as a VP Economist responsible for developing corporate economic forecasts and internal data analytics. He received his Ph.D. in Statistics from University of California at Davis, with research interests in model-assisted survey design and robust statistical methods.

Aggregation and the identification of "Smallness"

- Some employee groups contain so few employees, *n*, such that it is impossible to apply the simple classical formulas that are only valid asymptotically as $n \to \infty$.
- Examples of small employee groups in statistical analysis:
  - Pay Analysis Groups (PAG)
  - Race, gender groups within PAGs
  - Job applicant pool for a specific position of a collective of positions
  - Multiple hiring requisitions of a single job
- For the same effect size, small-group statistics do not provide as reliable direct estimates as larger samples.

- Statistical professionals mostly rely on univariate small-group statistics to handle outcome comparisons. Methods include *non-parametric methods*, *nearest-neighbor matching*, *Cochran-Mantel-Haenszel (CMH)* or *Bayesian* methods to combine multiple groups.
- Though more advanced modelling techniques may be applied, it will require more auxiliary data and broader assumption of underlying structure. These approaches are also constrained by turn-around time and agency resources.

Ryan Peterson

Ryan Peterson is the labor economist for the Southeast Region of the U.S. Department of Labor's Office of Federal Contract Compliance Programs (OFCCP). Prior to his work with the Department of Labor, Ryan was in the Transfer Pricing Practice in the Internal Revenue Service, where he worked on large and complex audits of multinational corporations. A member of OFCCP's Branch of Expert Services, Ryan assists field staff in their EEO analyses and provides ongoing training to staff and management. Ryan's doctoral work was in labor economics and public finance at the University of Texas-Austin

Statistical (and Statistic-Adjacent) Tests

- Most basic level: matched pairs and simple cohort “tests”
  - Look for sets of individuals who are nearly identical besides protected class and check for similarly (equivalent hire rates, equivalent pay)
  - Look for individuals where members of one protected class have equal or better characteristics (e.g., equal or more training and tenure), assess whether outcomes are at least as good.
- Very simple test, very simple problem
  - “Failing the test” is a very bad sign, but passing it may convey little to no information
  - Matched pairs may not even exist

Some Other Examples

- Fisher’s Exact Test
  - Probably best known
  - Most typical use when comparing hire rates between two protected classes when there are fewer than 30 comparators
    - May use for promotions test as well
- Issue: “exact” here tells us we can use the test with small samples, not that it gives a definitive answer

Some Other Examples (2)

- Rank Sum tests such as Mann-Whitney U
  - Non-parametric
    - In principle this can work whenever we can rank outcomes (e.g., pay or length of time until promoted) for two sub-samples (e.g., male v female)
  - Mechanics
    - Rank everyone in the sample (lowest 1, second lowest 2, etc.)
    - Split into two sub-samples (e.g., male v female)
    - Sum ranks of the two sub-samples (very different totals suggesting a disparity)
  - Issue: we lose a bit of information when we summarize, for example, pay data, in terms of ranks.
    - Absence of a finding not necessarily dispositive

Andy Leu

Andy Leu is a senior statistician working under the Agency of Office of Contract Compliance Programs (OFCCP) in the United States Department of Labor. He manages the statistical application of compliance assessments, which includes planning, evaluating, reviewing, and partaking in the recommendations of compensation, hiring, and promotion analysis supported by the Agency's Pacific offices. Andy earned his Ph.D. in Applied Statistics, with an emphasize on Operation Research from the University of Northern Colorado. He became certified in Production and Inventory management from American APICS, where he also earned the black belt in Six Sigma. Prior to his work at OFCCP, he began teaching assembly language programming as an assist professor at Taiwan's National Chin-Yi Institute of Technology. He then worked for Gateway Computer and IOMEGA. Here he gained experience working in the arena of quality management, the customer call center, and the marketing research. He then worked with ANSWERTHINK Consulting company as a business forecasting implementation manager before he moved to the Federal Government as a Statistician.

K-Nearest Neighbor (KNN) Method to predict the Pay

The **KNN** algorithm **uses** ”features/attributes similarity” to **estimate** the target values. For example, use compensation factors to estimate pay.

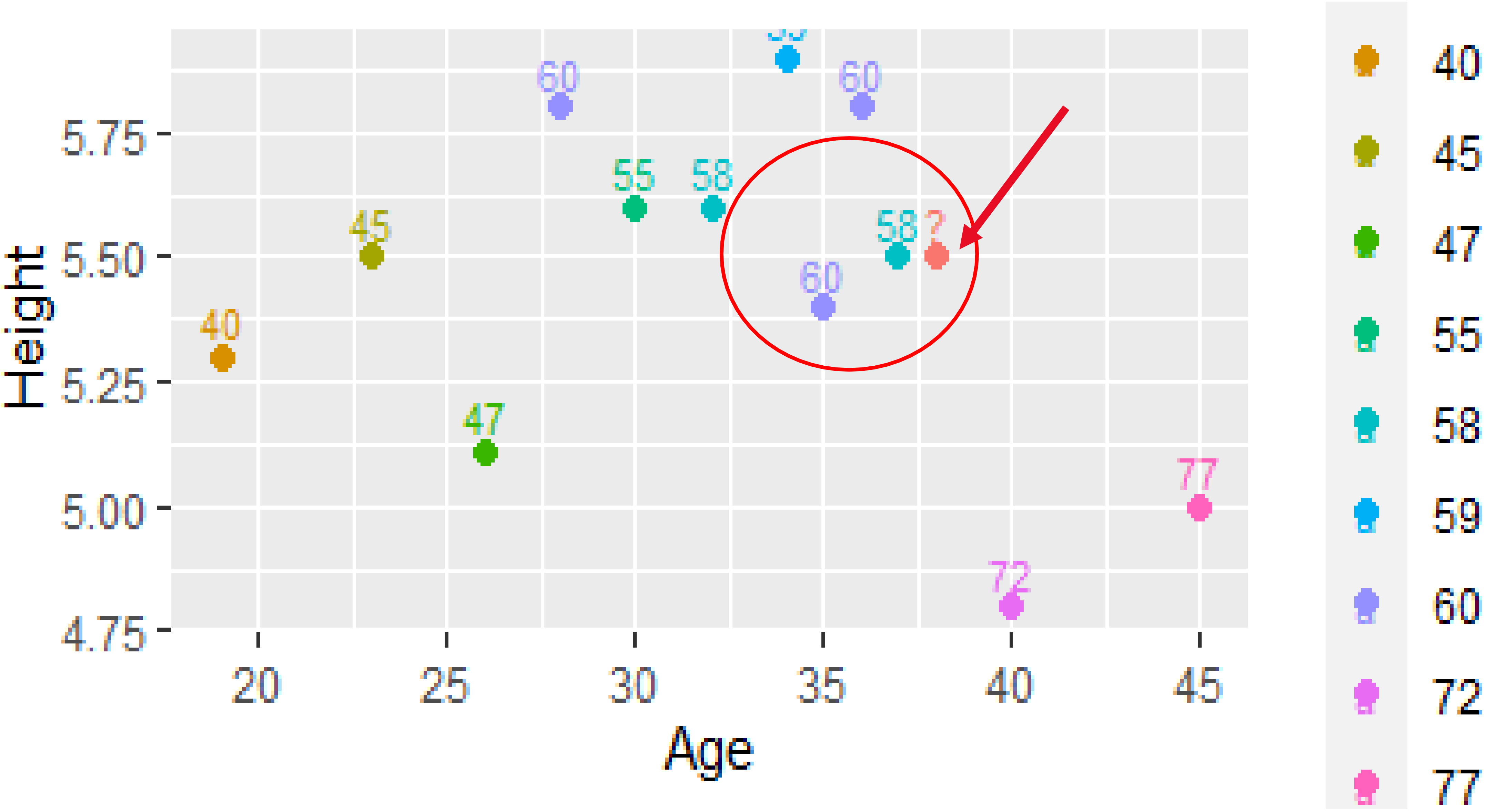
K-nearest neighbors is a nonparametric method. Based on the given data to find the best K neighbor - a **tuning parameter**, k which is decided by cross validation through evaluation by Root Mean Square Error (RMSE).

Applicable for continuous target (KNN regression) and categorical target variable (KNN Classification).

TARGET IS CONTINUOUS -REGRESSION

Height	Age	Weight
5	45	77
5.11	26	47
5.6	30	55
5.9	34	59
4.8	40	72
5.8	36	60
5.3	19	40
5.8	28	60
5.5	23	45
5.6	32	58
5.4	35	60
5.5	37	58
5.5	38	?

TARGET IS CONTINUOUS -REGRESSION



Bret Phillips

Bret Phillips is Statistician for the Southwest and Rocky Mountain Region of the U.S. Department of Labor's Office of Federal Contract Compliance Programs (OFCCP). With over ten years of experience, Bret conducts complex statistical analyses to support field staff investigations of equal employment opportunity in the federal contractor workforce. Bret has received a number of awards for his work at OFCCP, most recently for developing tools to facilitate investigation of compensation discrimination. Prior to working at OFCCP, Bret worked as a statistical analyst for the State of Georgia and a program evaluator for the State of Texas. Bret holds M.A. and Ph.D. degrees in Applied Experimental Psychology from Southern Illinois University at Carbondale, and a B.S. in Psychology from Illinois State University.

Multiattribute Evaluation of Pay in Small Cohorts

Bret T. Phillips, Ph.D.

Branch of Expert Services, OFCCP

Cohort Analysis

- A nonstatistical technique where individual employees are compared to determine whether pay disparity exists by protected status.
- Classically, this is done by following comparators over time (“cohort” is a term from longitudinal data analysis).
- However, the EEO analyst is often presented with cross-sectional (single point in time or “snapshot”) data.

Techniques for Cross-Sectional Cohort Analysis

- Nooren & Biddle (2010)* suggest placing the employee data in a spreadsheet and sorting multiple ways to detect underpaid individuals.
- I teach COs to use PivotTables to drill-down on problematic PAGs, sort the new spreadsheet by descending pay, visually inspect, and highlight problematic comparators using color codes.

*Nooren, P. M., & Biddle, D. A. (2010). *Compensation analysis: A practitioner's guide to identifying and addressing compensation disparities*. Folsom, CA: Author.

Problems

- Trying to compare people using multiple factors simultaneously is hard work.
- Analyst will tend to focus only on the most direct comparator sets; narrows the study to an Equal Pay Act-style analysis
  - Title VII does not require direct comparators for an inference of discrimination (*Washington County v. Gunther*; *Lenzi v. Systemax*)
- No expectation for what anyone’s pay should be, so no way to say who is overpaid/underpaid

Multiattribute Utility Analysis (MAUT)

- Borrowed from decision theory*, MAUT is a nonstatistical method that produces a weighted-sum utility score to compare people across multiple attributes (or factors) simultaneously.
- Breaks analysis down into simpler components, which makes it easier on the analyst.

*Edwards, W., & Newman, J. R. (1982). *Multiattribute evaluation*. Thousand Oaks, CA: Sage.

MAUT, Preference & Pay

- MAUT is typically used to scale alternatives (e.g., policy options) by preferability.
- How is pay related to preference?
  - One way to think of pay is that it is the employer’s expressed preference for each of its employees.
  - Human capital factors are then objective measures of employee preferability.
  - Pay equity is then a question of the extent to which employer preference corresponds to employee preferability.

Step 1. Gather Data

- Requires a set of attributes along which each employee can be at least ranked.
  - Tenure (years of service, perhaps split into time in job vs. other time in company)
  - Years of Education
  - Prior Experience
  - Performance
  - Geographic Differential Zones
  - Etc.
- Spreadsheet of each employee by ID, race, gender, pay, and the above attributes

MAUT Cohort Example 20200626 BP - Excel

J1 <-- yellow columns calculated by OFCCP

ID Number	Gender	Race	Hire Date	Effective Date of Current Job Code	Employee Pay Rate	Total Years of Service	Years in Current Job	Years in Other Jobs	<-- yellow columns calculated by OFCCP
Employee 1	Male	White	2/22/79	4/9/90	$55,458	31.94	20.82	11.13	
Employee 2	Male	Hispanic	5/24/75	9/12/94	$52,936	35.69	16.39	19.30	
Employee 3	Male	Hispanic	6/24/90	8/31/03	$50,882	20.61	7.42	13.19	
Employee 4	Male	Hispanic	12/15/88	8/19/99	$48,724	22.13	11.46	10.67	
Employee 5	Male	Hispanic	11/18/81	2/29/04	$46,488	29.20	6.92	22.28	
Employee 6	Male	Hispanic	12/22/80	12/11/92	$45,604	30.11	18.14	11.97	
Employee 7	Male	Hispanic	1/2/78	9/6/00	$44,928	33.08	10.40	22.68	
Employee 8	Male	Hispanic	10/17/80	4/9/90	$43,108	30.29	20.82	9.48	
Employee 9	Male	Hispanic	4/8/85	1/1/02	$42,484	25.82	9.08	16.73	
Employee 10	Female	Hispanic	3/29/90	7/14/96	$41,210	20.85	14.55	6.29	
Employee 11	Male	Hispanic	6/3/94	2/29/04	$39,702	16.67	6.92	9.74	
Employee 12	Male	Hispanic	8/9/04	10/17/10	$37,102	6.48	0.29	6.19	
Employee 13	Female	Hispanic	12/9/02	10/17/10	$36,504	8.15	0.29	7.85	

Original Employee Data | Utilities | Equal Weighting | MAUT Scores | Predicted Pay | Remedy Th ...

Step 2a. Rank Employees by Attribute

- Rank employees along each attribute.
  - Ranks need to be ascending order, where 1 = lowest to *N* = highest (*N* = number of employees).
  - If having someone else do the ranking, it may be easier for them to rank 1 = highest to *N* = lowest. This is fine, but then analyst needs to reverse code the ranks.
- Do not worry about pay, race, or gender at this point! Those will be dealt with later.

Step 2b. Calculate Utilities for Rank Data

- Convert ranks to utilities using the following formula

$$u_{ij} = \frac{X_{ij} - 1}{N - 1}$$

Where X_{ij} = rank for employee *i* on attribute *j, and N* = number of employees

- Each utility ranges from 0 to 1

Step 3. Calculate Utilities for Interval Data

- Rank-based utilities may be too crude when the data are on an interval scale of measurement (e.g., years of service). Use alternate formula:

$$u_{ij} = \frac{x_{ij} - min_j}{max_j - min_j}$$

- Where x_{ij} = actual value (not rank!) for employee *i* on attribute *j*, max_j is the maximum observed value among employees on attribute *j*, and min_j is the minimum observed value among employees on attribute *j*.
- Rank- and interval-based utilities may be used within the same MAUT.

MAUT Cohort Example 20200626 BP - Excel

ID Number	Gender	Race	Hire Date	Effective Date of Current Job Code	Employee Pay Rate	Total Years of Service	Years in Current Job	Years in Other Jobs	Utility (Yrs Curr Job)	Utility (Yrs Other Jobs)
Employee 1	Male	White	2/22/79	4/9/90	$55,458	31.94	20.82	11.13	1.00	0.30
Employee 2	Male	Hispanic	5/24/75	9/12/94	$52,936	35.69	16.39	19.30	0.78	0.80
Employee 3	Male	Hispanic	6/24/90	8/31/03	$50,882	20.61	7.42	13.19	0.35	0.42
Employee 4	Male	Hispanic	12/15/88	8/19/99	$48,724	22.13	11.46	10.67	0.54	0.27
Employee 5	Male	Hispanic	11/18/81	2/29/04	$46,488	29.20	6.92	22.28	0.32	0.98
Employee 6	Male	Hispanic	12/22/80	12/11/92	$45,604	30.11	18.14	11.97	0.87	0.35
Employee 7	Male	Hispanic	1/2/78	9/6/00	$44,928	33.08	10.40	22.68	0.49	1.00
Employee 8	Male	Hispanic	10/17/80	4/9/90	$43,108	30.29	20.82	9.48	1.00	0.20
Employee 9	Male	Hispanic	4/8/85	1/1/02	$42,484	25.82	9.08	16.73	0.43	0.64
Employee 10	Female	Hispanic	3/29/90	7/14/96	$41,210	20.85	14.55	6.29	0.69	0.01
Employee 11	Male	Hispanic	6/3/94	2/29/04	$39,702	16.67	6.92	9.74	0.32	0.22
Employee 12	Male	Hispanic	8/9/04	10/17/10	$37,102	6.48	0.29	6.19	0.00	0.00
Employee 13	Female	Hispanic	12/9/02	10/17/10	$36,504	8.15	0.29	7.85	0.00	0.10

Original Employee Data | Utilities | Equal Weighting | MAUT Scores | Predicted Pay | Remedy Th ...

Step 4. Calculate Weights

- Attribute weights range from 0 to 1, and reflect the proportion of pay believed to be influenced by a particular attribute
- Equal weighting: $w_j = 1/j$, where j = # attributes.
- Other weighting systems are possible (rank sum, rank reciprocal, etc.).

MAUT Cohort Example 20200626 BP - Excel

	A	B	C	D	E	F	G	H	I	J	K
1									Weights -->	0.50	0.50
2	ID Number	Gender	Race	Hire Date	Effective Date of Current Job Code	Employee Pay Rate	Total Years of Service	Years in Current Job	Years in Other Jobs	Utility (Yrs Curr Job)	Utility (Yrs Other Jobs)
3	Employee 1	Male	White	2/22/79	4/9/90	$55,458	31.94	20.82	11.13	1.00	0.30
4	Employee 2	Male	Hispanic	5/24/75	9/12/94	$52,936	35.69	16.39	19.30	0.78	0.80
5	Employee 3	Male	Hispanic	6/24/90	8/31/03	$50,882	20.61	7.42	13.19	0.35	0.42
6	Employee 4	Male	Hispanic	12/15/88	8/19/99	$48,724	22.13	11.46	10.67	0.54	0.27
7	Employee 5	Male	Hispanic	11/18/81	2/29/04	$46,488	29.20	6.92	22.28	0.32	0.98
8	Employee 6	Male	Hispanic	12/22/80	12/11/92	$45,604	30.11	18.14	11.97	0.87	0.35
9	Employee 7	Male	Hispanic	1/2/78	9/6/00	$44,928	33.08	10.40	22.68	0.49	1.00
10	Employee 8	Male	Hispanic	10/17/80	4/9/90	$43,108	30.29	20.82	9.48	1.00	0.20
11	Employee 9	Male	Hispanic	4/8/85	1/1/02	$42,484	25.82	9.08	16.73	0.43	0.64
12	Employee 10	Female	Hispanic	3/29/90	7/14/96	$41,210	20.85	14.55	6.29	0.69	0.01
13	Employee 11	Male	Hispanic	6/3/94	2/29/04	$39,702	16.67	6.92	9.74	0.32	0.22
14	Employee 12	Male	Hispanic	8/9/04	10/17/10	$37,102	6.48	0.29	6.19	0.00	0.00
15	Employee 13	Female	Hispanic	12/9/02	10/17/10	$36,504	8.15	0.29	7.85	0.00	0.10

Original Employee Data | Utilities | **Equal Weighting** | MAUT Scores | Predicted Pay | Remedy Th ...

Step 5. Calculate MAUT Scores

- Calculate final MAUT scores for each employee using the formula.

$$MAUT_i = \sum_{1}^{j} w_j u_{ij}$$

- MAUT scores range from 0 to 1, with 0 being an employee ranked lowest on every attribute and 1 being an employee ranked highest on every attribute.

MAUT Cohort Example 20200626 BP - Excel

								Weights -->	0.50	0.50	
ID Number	Gender	Race	Hire Date	Effective Date of Current Job Code	Employee Pay Rate	Total Years of Service	Years in Current Job	Years in Other Jobs	Utility (Yrs Curr Job)	Utility (Yrs Other Jobs)	MAUT
Employee 1	Male	White	2/22/79	4/9/90	$55,458	31.94	20.82	11.13	1.00	0.30	0.65
Employee 2	Male	Hispanic	5/24/75	9/12/94	$52,936	35.69	16.39	19.30	0.78	0.80	0.79
Employee 3	Male	Hispanic	6/24/90	8/31/03	$50,882	20.61	7.42	13.19	0.35	0.42	0.39
Employee 4	Male	Hispanic	12/15/88	8/19/99	$48,724	22.13	11.46	10.67	0.54	0.27	0.41
Employee 5	Male	Hispanic	11/18/81	2/29/04	$46,488	29.20	6.92	22.28	0.32	0.98	0.65
Employee 6	Male	Hispanic	12/22/80	12/11/92	$45,604	30.11	18.14	11.97	0.87	0.35	0.61
Employee 7	Male	Hispanic	1/2/78	9/6/00	$44,928	33.08	10.40	22.68	0.49	1.00	0.75
Employee 8	Male	Hispanic	10/17/80	4/9/90	$43,108	30.29	20.82	9.48	1.00	0.20	0.60
Employee 9	Male	Hispanic	4/8/85	1/1/02	$42,484	25.82	9.08	16.73	0.43	0.64	0.53
Employee 10	Female	Hispanic	3/29/90	7/14/96	$41,210	20.85	14.55	6.29	0.69	0.01	0.35
Employee 11	Male	Hispanic	6/3/94	2/29/04	$39,702	16.67	6.92	9.74	0.32	0.22	0.27
Employee 12	Male	Hispanic	8/9/04	10/17/10	$37,102	6.48	0.29	6.19	0.00	0.00	0.00
Employee 13	Female	Hispanic	12/9/02	10/17/10	$36,504	8.15	0.29	7.85	0.00	0.10	0.05

Original Employee Data | Utilities | Equal Weighting | MAUT Scores | Predicted Pay | Remedy Th ...

Step 6. Compute Predicted Pay & Residuals

- Use MAUT scores to compute predicted pay

  $$\widehat{pay_i} = MAUT_i(pay_{max} - pay_{min}) + pay_{min}$$

  Where $MAUT_i$ = MAUT score for employee i, pay_{max} = highest observed pay in cohort, and pay_{min} = lowest observed pay in cohort.

- Residuals are predicted pay minus actual pay for each employee in the cohort.
  - Divide residual by actual pay and multiply by 100 to convert to percentage overpaid/underpaid.

MAUT Cohort Example 20200626 BP - Excel

ID Number	Gender	Race	Hire Date	Effective Date of Current Job Code	Employee Pay Rate	MAUT	Predicted Pay	Residual	% Overpay/ Underpay	
Employee 1	Male	White	2/22/79	4/9/90	$55,458	0.65	$48,819.54	$6,638.46	11.97%	
Employee 2	Male	Hispanic	5/24/75	9/12/94	$52,936	0.79	$51,475.16	$1,460.84	2.76%	
Employee 3	Male	Hispanic	6/24/90	8/31/03	$50,882	0.39	$43,817.96	$7,064.04	13.88%	
Employee 4	Male	Hispanic	12/15/88	8/19/99	$48,724	0.41	$44,237.36	$4,486.64	9.21%	
Employee 5	Male	Hispanic	11/18/81	2/29/04	$46,488	0.65	$48,814.92	-$2,326.92	-5.01%	**<-- 8 of 12 Hispanics underpaid**
Employee 6	Male	Hispanic	12/22/80	12/11/92	$45,604	0.61	$48,068.97	-$2,464.97	-5.41%	
Employee 7	Male	Hispanic	1/2/78	9/6/00	$44,928	0.75	$50,649.96	-$5,721.96	-12.74%	
Employee 8	Male	Hispanic	10/17/80	4/9/90	$43,108	0.60	$47,870.74	-$4,762.74	-11.05%	
Employee 9	Male	Hispanic	4/8/85	1/1/02	$42,484	0.53	$46,624.58	-$4,140.58	-9.75%	
Employee 10	Female	Hispanic	3/29/90	7/14/96	$41,210	0.35	$43,149.71	-$1,939.71	-4.71%	**<-- 1 of 2 females underpaid**
Employee 11	Male	Hispanic	6/3/94	2/29/04	$39,702	0.27	$41,608.43	-$1,906.43	-4.80%	
Employee 12	Male	Hispanic	8/9/04	10/17/10	$37,102	0.00	$36,504.00	$598.00	1.61%	
Employee 13	Female	Hispanic	12/9/02	10/17/10	$36,504	0.05	$37,462.24	-$958.24	-2.63%	

Original Employee Data | Utilities | Equal Weighting | MAUT Scores | Predicted Pay | Remedy Th ...

Step 7. Examine Underpaid Employees

- Set a practical significance criterion for % underpaid.
  - OFCCP has no specific guidance here (Threshold has varied with scheduling letter changes.)
  - Proactive self-audit: employer must decide what % threshold they are comfortable with (risk tolerance).
  - No clear legal consensus on practical significance threshold
- Examine all employees who meet the criterion.
  - Do these employees disproportionately come from a particular race or gender?

Step 8. Allocate Remedy

- If liability is accepted based on Step 7, allocate remedy to underpaid class members
- How far to remedy?
  - OFCCP position: make whole remedy for class members.
  - Proactive self-audit: remedy up to the practical significance threshold previously identified.
- In a similar way, must provide for prospective pay adjustments for class members. Management will need to commit to these as well.

MAUT Cohort Example 20200626 BP - Excel

	A	B	C	D	E	F	N	O	P	Q	R
2	ID Number	Gender	Race	Hire Date	Effective Date of Current Job Code	Employee Pay Rate	Residual	% Overpay/ Underpay	To 3.5% Threshold	Pay Adjustment Amount	New Pay Rate
3	Employee 1	Male	White	2/22/79	4/9/90	$55,458	$6,638.46	11.97%			
4	Employee 2	Male	Hispanic	5/24/75	9/12/94	$52,936	$1,460.84	2.76%			
5	Employee 3	Male	Hispanic	6/24/90	8/31/03	$50,882	$7,064.04	13.88%			
6	Employee 4	Male	Hispanic	12/15/88	8/19/99	$48,724	$4,486.64	9.21%			
7	Employee 5	Male	Hispanic	11/18/81	2/29/04	$46,488	**-$2,326.92**	**-5.01%**	**-1.51%**	**$699.84**	**$47,187.84**
8	Employee 6	Male	Hispanic	12/22/80	12/11/92	$45,604	**-$2,464.97**	**-5.41%**	**-1.91%**	**$868.83**	**$46,472.83**
9	Employee 7	Male	Hispanic	1/2/78	9/6/00	$44,928	**-$5,721.96**	**-12.74%**	**-9.24%**	**$4,149.48**	**$49,077.48**
10	Employee 8	Male	Hispanic	10/17/80	4/9/90	$43,108	**-$4,762.74**	**-11.05%**	**-7.55%**	**$3,253.96**	**$46,361.96**
11	Employee 9	Male	Hispanic	4/8/85	1/1/02	$42,484	**-$4,140.58**	**-9.75%**	**-6.25%**	**$2,653.64**	**$45,137.64**
12	Employee 10	Female	Hispanic	3/29/90	7/14/96	$41,210	**-$1,939.71**	**-4.71%**	**-1.21%**	**$497.36**	**$41,707.36**
13	Employee 11	Male	Hispanic	6/3/94	2/29/04	$39,702	**-$1,906.43**	**-4.80%**	**-1.30%**	**$516.86**	**$40,218.86**
14	Employee 12	Male	Hispanic	8/9/04	10/17/10	$37,102	$598.00	1.61%			
15	Employee 13	Female	Hispanic	12/9/02	10/17/10	$36,504	-$958.24	-2.63%			

Utilities | Equal Weighting | MAUT Scores | Predicted Pay | Remedy Threshold 3.5% | Rank Sum

Sensitivity Analysis

- Note that MAUT involves making decisions:
  - Who is in the cohort?
  - What attributes to include?
  - How should attributes be weighted?
- Perform MAUT under various scenarios and see how results are impacted – this will bring any issues regarding the calculation to a head.
- Management should sign off on the final MAUT calculation used.

MAUT Cohort Example 20200626 BP - Excel

ID Number	Gender	Race	Years in Current Job	Years in Other Jobs	Utility (Yrs Curr Job)	Utility (Yrs Other Jobs)	MAUT	Predicted Pay	Residual	% Overpay/ Underpay	
				Weights -->	0.67	0.33					
Employee 1	Male	White	20.82	11.13	1.00	0.30	0.77	$51,076.61	$4,381.39	7.90%	
Employee 2	Male	Hispanic	16.39	19.30	0.78	0.80	0.79	$51,439.18	$1,496.82	2.83%	
Employee 3	Male	Hispanic	7.42	13.19	0.35	0.42	0.37	$43,569.89	$7,312.11	14.37%	
Employee 4	Male	Hispanic	11.46	10.67	0.54	0.27	0.45	$45,113.04	$3,610.96	7.41%	
Employee 5	Male	Hispanic	6.92	22.28	0.32	0.98	0.54	$46,711.42	-$223.42	-0.48%	<-- no longer underpaid
Employee 6	Male	Hispanic	18.14	11.97	0.87	0.35	0.70	$49,741.31	-$4,137.31	**-9.07%**	
Employee 7	Male	Hispanic	10.40	22.68	0.49	1.00	0.66	$49,015.23	-$4,087.23	**-9.10%**	
Employee 8	Male	Hispanic	20.82	9.48	1.00	0.20	0.74	$50,450.41	-$7,342.41	**-17.03%**	
Employee 9	Male	Hispanic	9.08	16.73	0.43	0.64	0.50	$45,944.10	-$3,460.10	**-8.14%**	
Employee 10	Female	Hispanic	14.55	6.29	0.69	0.01	0.47	$45,367.52	-$4,157.52	**-10.09%**	
Employee 11	Male	Hispanic	6.92	9.74	0.32	0.22	0.29	$41,955.14	-$2,253.14	**-5.68%**	
Employee 12	Male	Hispanic	0.29	6.19	0.00	0.00	0.00	$36,504.00	$598.00	1.61%	
Employee 13	Female	Hispanic	0.29	7.85	0.00	0.10	0.03	$37,136.44	-$632.44	-1.73%	

Equal Weighting | MAUT Scores | Predicted Pay | Remedy Threshold 3.5% | Rank Sum Scenario

MAUT Cohort Example 20200626 BP - Excel

ID Number	Gender	Race	Years in Current Job	Years in Other Jobs	Utility (Yrs Curr Job)	Utility (Yrs Other Jobs)	MAUT	Predicted Pay	Residual	% Overpay/ Underpay	
				Weights -->	0.33	0.67					
Employee 1	Male	White	20.82	11.13	1.00	0.30	0.53	$46,562.46	$8,895.54	16.04%	
Employee 2	Male	Hispanic	16.39	19.30	0.78	0.80	0.79	$51,511.13	$1,424.87	2.69%	
Employee 3	Male	Hispanic	7.42	13.19	0.35	0.42	0.40	$44,066.02	$6,815.98	13.40%	
Employee 4	Male	Hispanic	11.46	10.67	0.54	0.27	0.36	$43,361.68	$5,362.32	11.01%	
Employee 5	Male	Hispanic	6.92	22.28	0.32	0.98	0.76	$50,918.42	-$4,430.42	**-9.53%**	
Employee 6	Male	Hispanic	18.14	11.97	0.87	0.35	0.52	$46,396.63	-$792.63	-1.74%	**<-- no longer underpaid**
Employee 7	Male	Hispanic	10.40	22.68	0.49	1.00	0.83	$52,284.70	-$7,356.70	**-16.37%**	
Employee 8	Male	Hispanic	20.82	9.48	1.00	0.20	0.46	$45,291.07	-$2,183.07	**-5.06%**	
Employee 9	Male	Hispanic	9.08	16.73	0.43	0.64	0.57	$47,305.07	-$4,821.07	**-11.35%**	
Employee 10	Female	Hispanic	14.55	6.29	0.69	0.01	0.23	$40,931.89	$278.11	0.67%	**<-- no longer underpaid**
Employee 11	Male	Hispanic	6.92	9.74	0.32	0.22	0.25	$41,261.73	-$1,559.73	**-3.93%**	
Employee 12	Male	Hispanic	0.29	6.19	0.00	0.00	0.00	$36,504.00	$598.00	1.61%	
Employee 13	Female	Hispanic	0.29	7.85	0.00	0.10	0.07	$37,788.04	-$1,284.04	**-3.52%**	**<-- now underpaid**

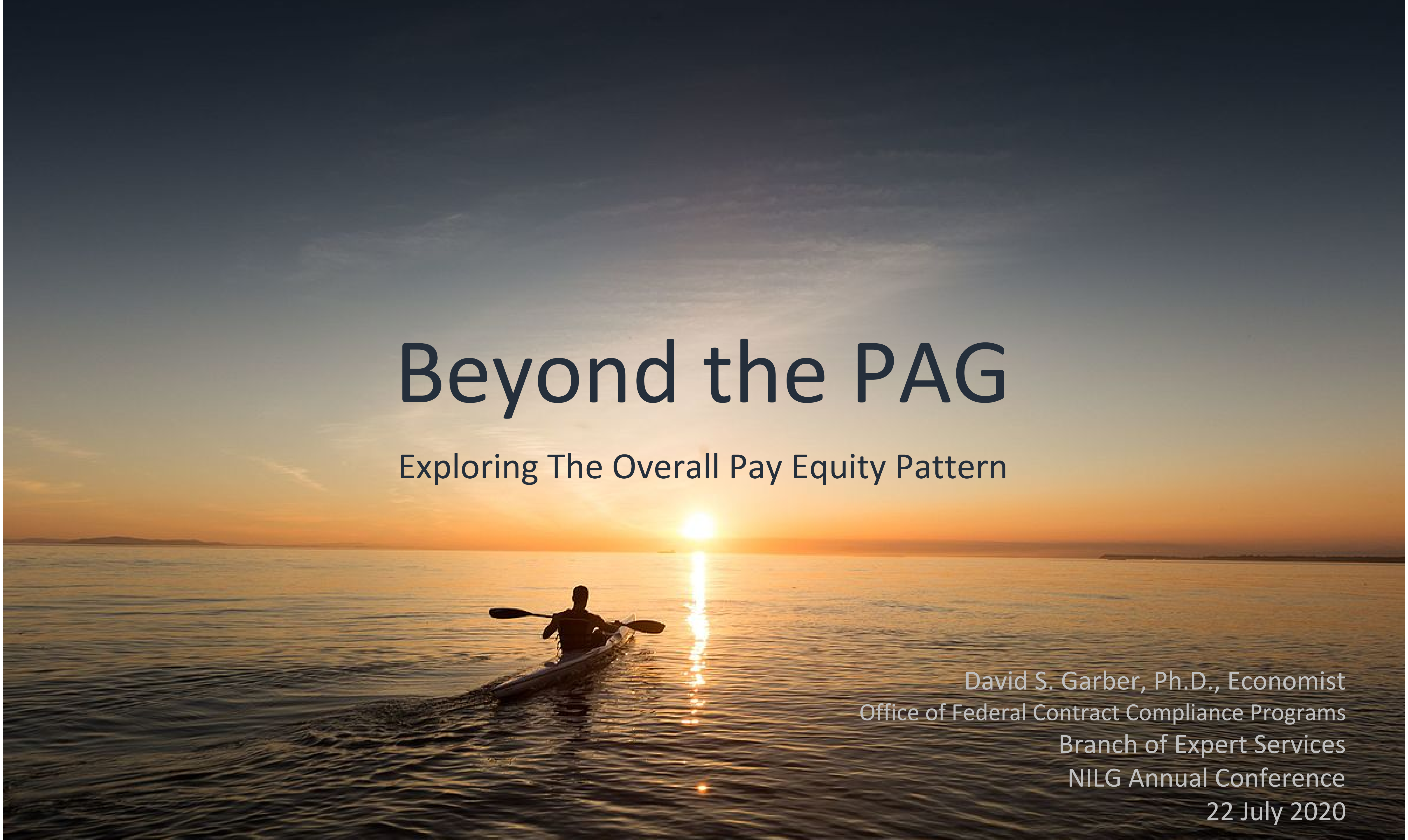
Predicted Pay | Remedy Threshold 3.5% | Rank Sum Scenario | Rank Sum Scenario (2)

Getting More Sophisticated

- Inferential statistics can be incorporated into MAUT as long as there are at least 2 employees in each group.
  - e.g., T-test (gender) or one-way ANOVA (race) on MAUT residuals.
  - Or use nonparametric equivalents (e.g., Mann-Whitney, Kruskal-Wallis).
- These tests will all be fairly underpowered, and may give an employer a false sense of security.
  - OFCCP does not need to show 2 SD if other evidence supports a finding of discrimination. Neither do private plaintiffs.

David Garber

David Garber is an economist with the U.S. Department of Labor's Office of Contract Compliance Programs (OFCCP) where he oversees the statistical component of compliance assessments undertaken by the Agency's Midwest offices. A member of the Agency's Branch of Expert Services, Dr. Garber is one of the Agency's thought leaders on the application of statistical methods to labor compliance evaluation. Prior to his work at OFCCP, David served as Economist with the U.S. Agency for International Development's (USAID) Office of Economic Growth. Having served as an economist for over 10 years, in both research and governmental capacities, in the U.S. and internationally, one of Dr. Garber's primary aims is to turn complex economic and statistical theory into assessment methods that are both accessible and useful to the non-academic practitioner. David earned his Ph.D. in Applied Economics from the University of Wisconsin - Madison and his B.A. from the College of William and Mary.



Beyond the PAG

Exploring The Overall Pay Equity Pattern

David S. Garber, Ph.D., Economist
Office of Federal Contract Compliance Programs
Branch of Expert Services

NILG Annual Conference
22 July 2020

Motivation

- Point #7 of NILG/OFCCP Roundtable – Systemic Discrimination
  - concerned by “isolated indicators” of pay disparity that are not reflective of systemic issues.
  - suggestion for a “minimum indicator prevalence threshold” to show consistency in findings.

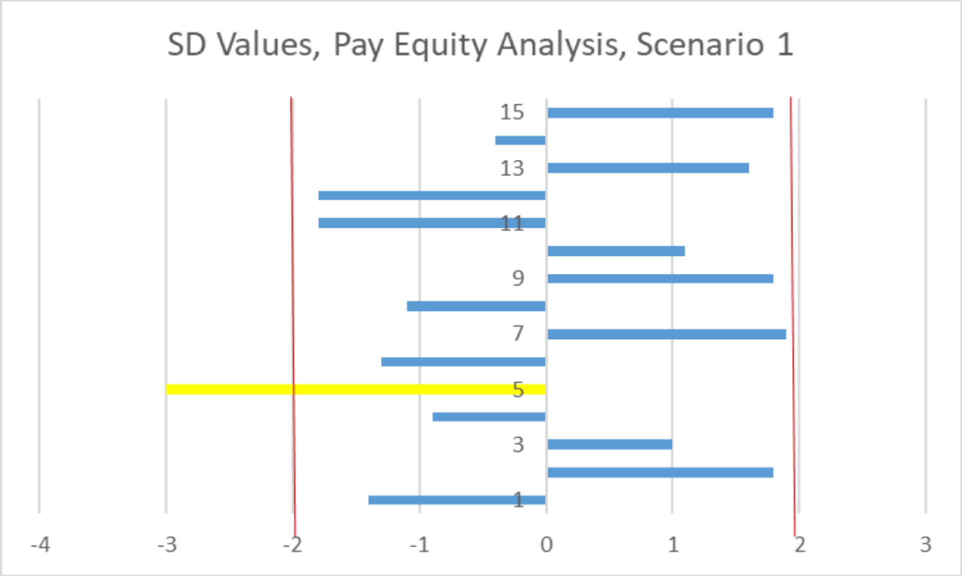
- OFCCP also strives to assess systemic, rather than isolated, issues.
  - Lack of statistical significance for any single PAG, particularly when due to small group size, should not be interpreted as pay parity within that group or overall.

The Problem, scenario 1 – “Mixed Signals”

Results, Pay Equity Regression Analysis, Contractor XyZ, Inc.

PAG	N	C_F-C_M (%)	SD
1	133	-2%	-1.4
2	106	+1%	+1.8
3	90	+2%	+1
4	79	-2%	-0.9
5	89	-4%	-3
6	79	-3%	-1.3
7	138	+1%	+1.9
8	13	-5%	-1.1
9	64	+3%	+1.8
10	85	+2%	+1.1
11	124	-1%	-1.8
12	144	-1%	-1.8
13	111	+1%	+1.6
14	2	-20%	-0.4
15	34	+2%	+1.8

Does this look like systemic bias?



Imperfect Answer:

Yes, because there’s a statistically significant pay disparity of 4% impacting females in PAG No. 5.

More Thorough Answer:

There may be something undesirable occurring in PAG No. 5, but it doesn’t seem to be part of a larger pattern.

Ronald Fisher, 1932, Statistics for Research Workers, p.103

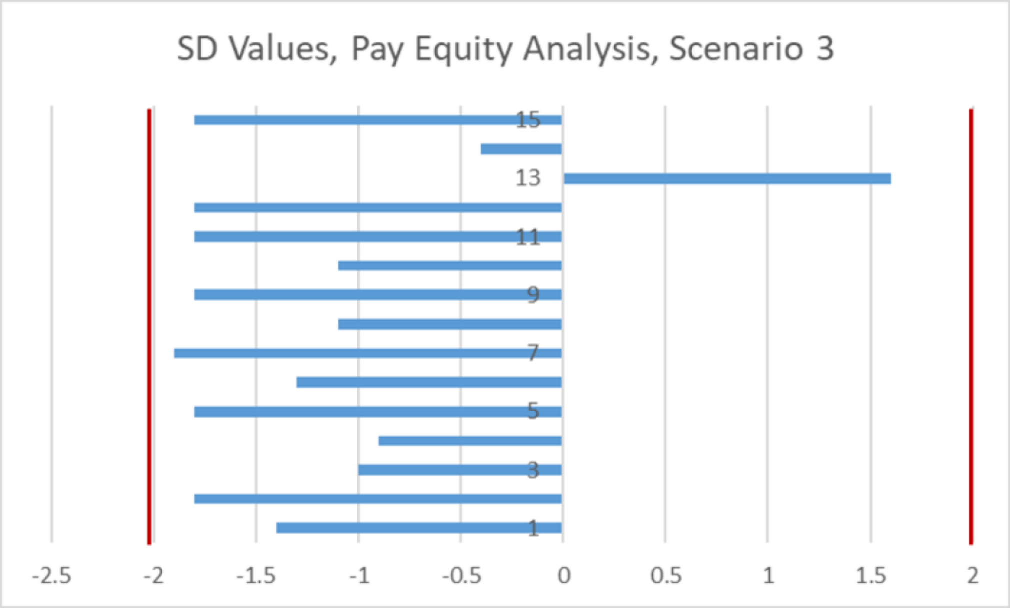
When a number of quite independent tests of significance have been made, it sometimes happens that although few or none can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are on the whole lower than would often have been obtained by chance. It is sometimes desired, taking account only of these probabilities, and not of the detailed composition of the data from which they are derived, which may be of very different kinds, to obtain a single test of the significance of the aggregate, based on the product of the probabilities individually observed.

Fisher, R. 1932. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh.

The Problem, scenario 2 – "The Subtle Pervasive Signal"

PAG	N	C_F-C_M (%)	SD
1	133	-2%	-1.4
2	106	-1%	-1.8
3	90	-2%	-1
4	79	-2%	-0.9
5	89	-2%	-1.8
6	79	-3%	-1.3
7	138	-1%	-1.9
8	13	-5%	-1.1
9	64	-3%	-1.8
10	85	-2%	-1.1
11	124	-1%	-1.8
12	144	-1%	-1.8
13	111	+1%	+1.6
14	2	-20%	-0.4
15	34	-2%	-1.8

Does this look like systemic bias?



Imperfect Answer:

No. The pay system seems to be equitable.

More Thorough Answer:

Not sure, but it doesn't look good:

1. Why do females earn less than their male counterparts in 14 of 15 PAGS?
2. Also, some of these PAGs contain too few employees for credible regression analysis.

The Problem,

Scenario 1: Isolated (s… …indicator not part of
larger pa…

Scenario 2: Lack of ar… …nt bias indicator
incorrect…

Generally, assessing the “within,” neglecting the “across”.

The Solution

- The current PAG framework, including regression analysis, “ignores the forest for the trees.”

- A comprehensive assessment of systemic pay equity in an employer’s workforce requires both trees and forest assessments.

- The SSEG v. sample size debate can be at least partially addressed by this more comprehensive approach.

- A variety of “summary” tests, some very simple to execute, are readily available to fill in the blanks of a comprehensive analysis.

An Omnibus Test, Demonstration

STEP 1:

Simulate a simple compensation database representing an employer that discriminates on the basis of gender.

STEP 2:

Show that regression analysis, in certain situations, is unable to detect the discrimination.

STEP 3:

Introduce a formal summary, or “combination,” test that does detect the overall bias pattern.

- Tippot (1931)
- **Fisher’s Combination Test (1932) – biased towards smallest p-values**
- Pearson’s Test (1934) – biased towards largest p-values
- **Z-transform Test (Stouffer, 1949) – reduces bias towards extreme p-values**
- Weighted Z-transform Test (Liptak, 1958) – idea to give different weights according to power of test
- Brown (1975) – combining non-independent tests
- Simes (1986) – combining non-independent tests, under certain conditions
- Kost & McDermott (2002) – combining non-independent tests
- Harmonic p-value test (Wilson, 2019) – combining non-independent tests

Simulating a Discriminating Employer, The Basics

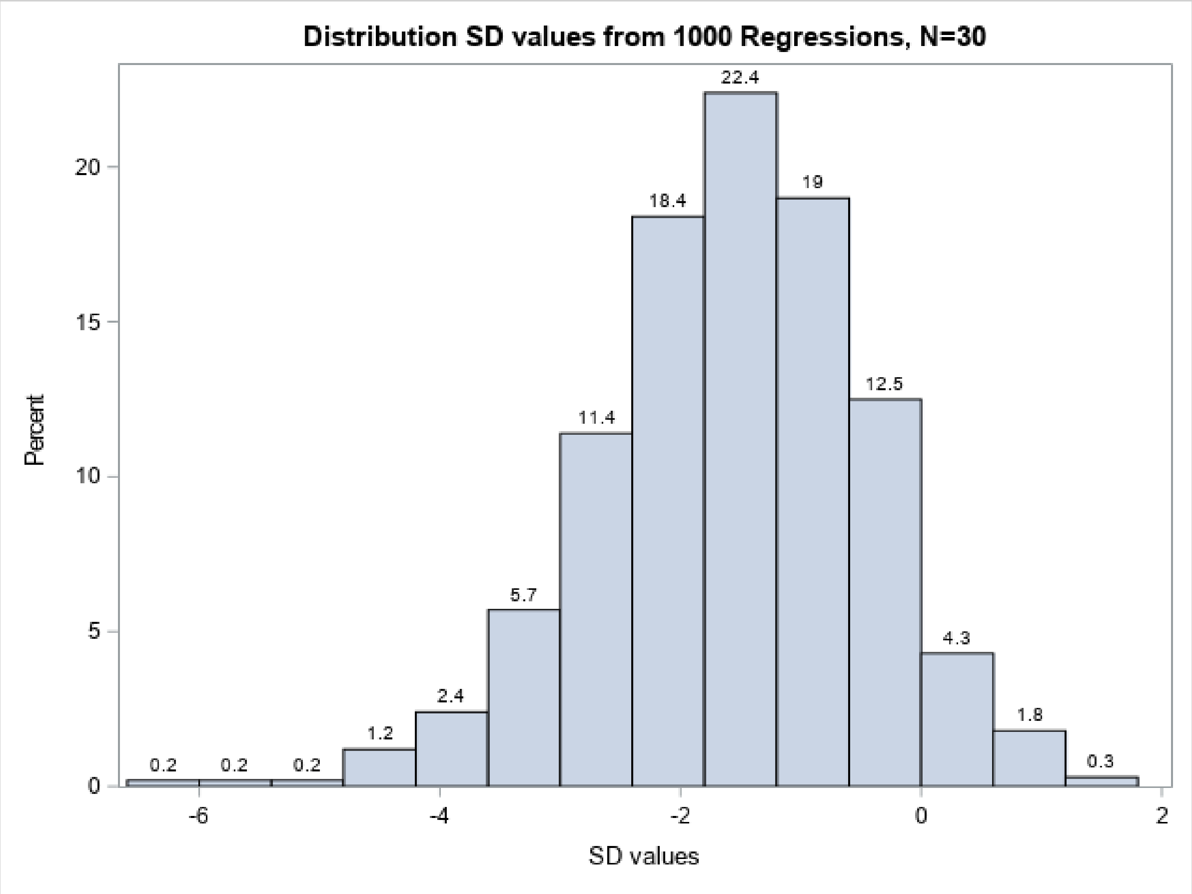
Biased Against Female Lawyers, Inc. (BAFLInc)

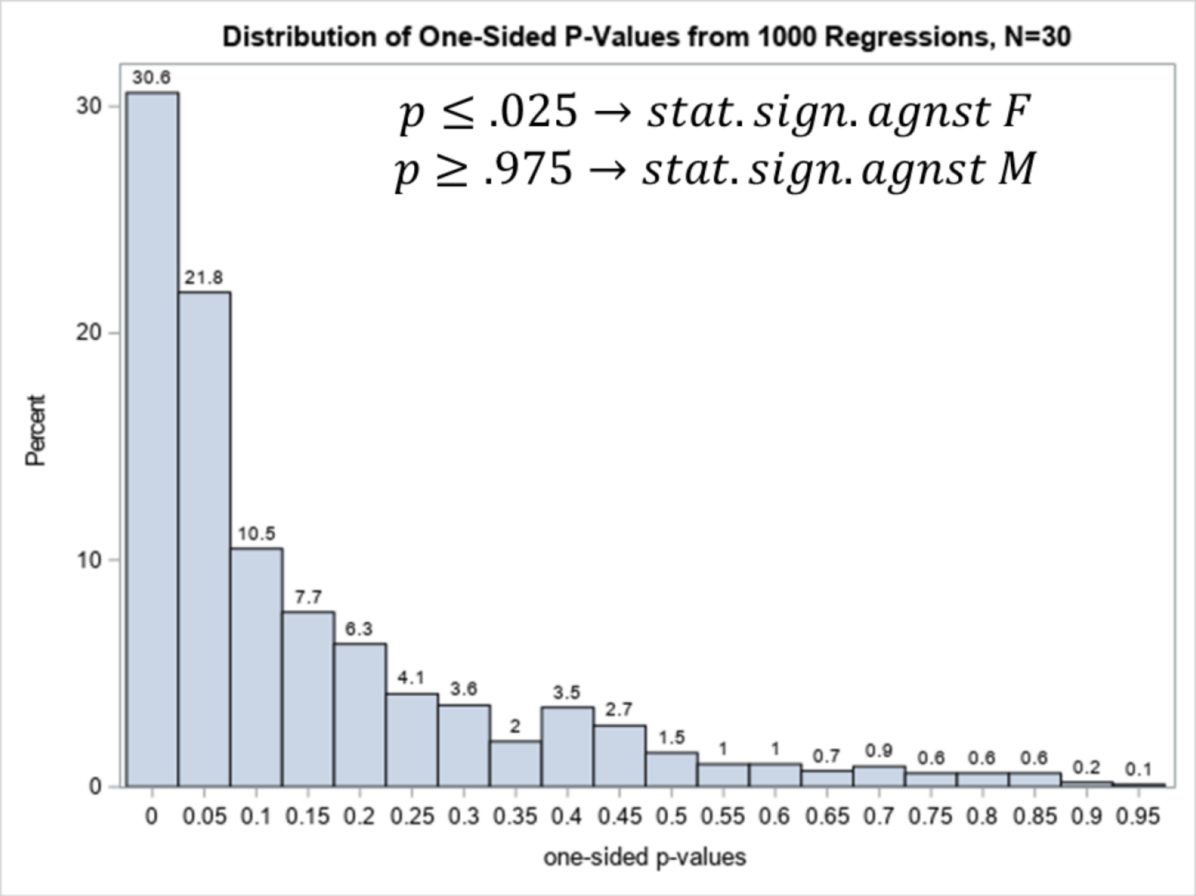
- For purposes of pay equity review, defines its analysis groups such that each contains 30 employees holding identical jobs.
- Employer calculates an employee's salary using a clear formula, although there is a very tiny bit of miscalculation (error).
  - For male employees: $C_M \cong 102500 + 2000t + 15000r$
  - For female employees: $C_F \cong 100000 + 1950t +$ 7500r r
  - Formulas show bias in starting pay (102500 v. 1000 7500r n returns from tenure (t) (2000 per year v. 1950 per year), and equal returns to performance (r).
- Each employee's tenure and performance rating is randomly generated (further details not necessary for this discussion).
- Gender imbalance -- each group is comprised of 1/3 female, 2/3 male.
- Independently generate 1000 groups (ie., PAGs) of 30 observations (employees).

Gastwirth, J.L., Bura, E. and Miao, W. (2011). Some important statistical issues courts should consider in their assessment of statistical analyses submitted in class certification motions: implications for Dukes v. Wal-Mart. Law, Probability and Risk, 10, 225-263.

Assessing BAFLinc.'s Pay System, Findings

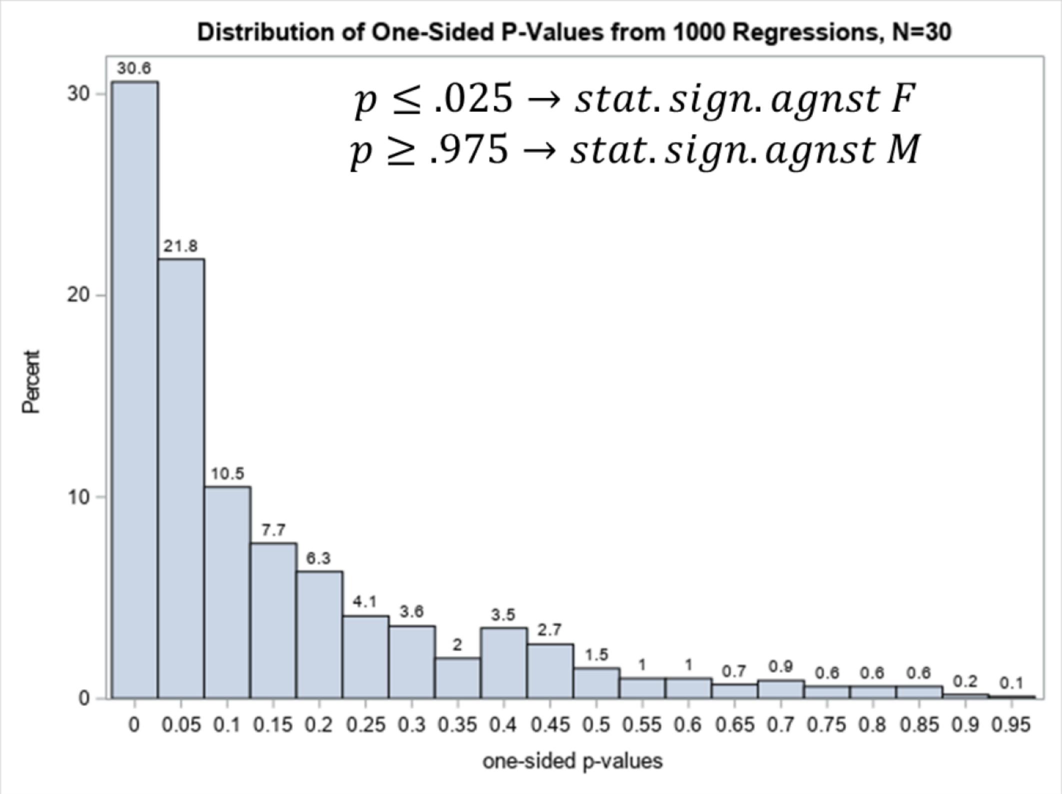
- Statistically significant pay disparity affecting females found in only 306 of 1000 groups (30.6%).
- But, also……pay disparity (not statistically significant) affecting females found in 630 additional groups.
- And…...pay disparity (not statistically significant) against male employees found in 64 groups.
- What's going on here?? We know the "employer" is discriminating, but regression is only detecting it in 31% of PAGs.

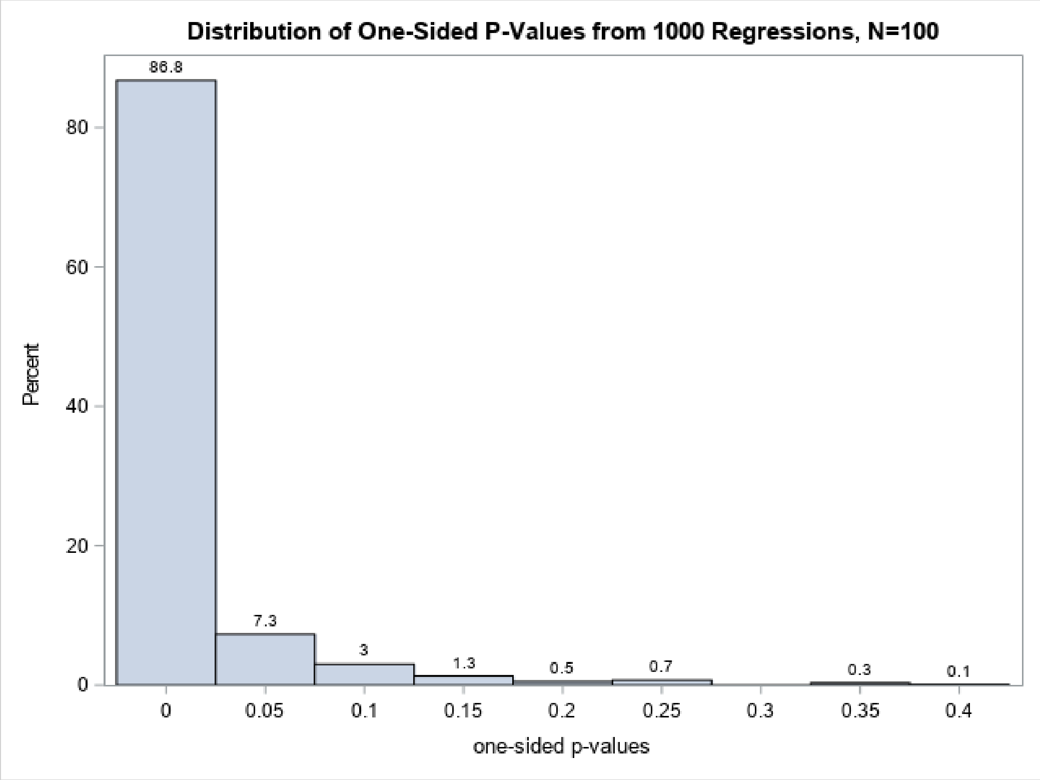




Detecting BAFLinc.'s Gender-biased Pay System

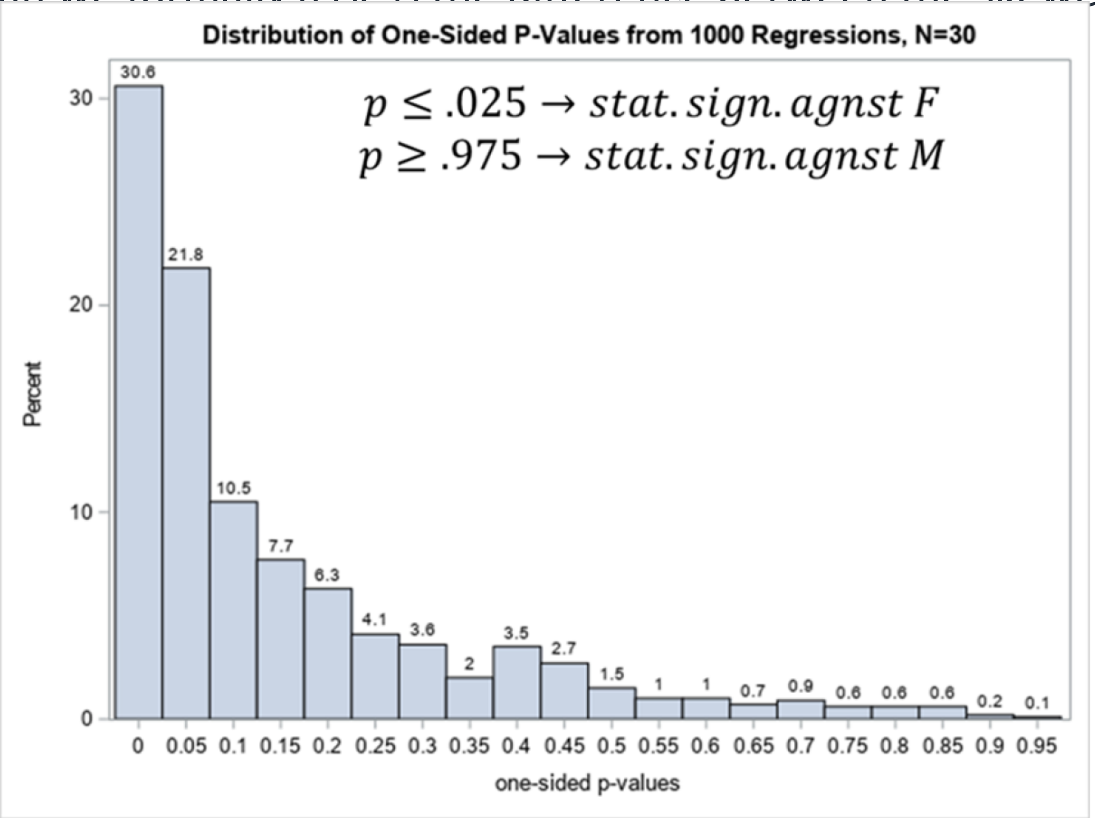
- PAGs, containing 30 employees, are small (also, there are 5 control variables – only 6 observations per variable).
- Size of Disparity is relatively small (2.3% on average).
- Gender imbalance.
- This all leads to Low Power of analysis – high rate of false negatives (69.1%)
- In this example, one may interpret the 30.6% as the Power of the regression analysis – the rate of detecting the true behavior.
- If we simply increase group size to 100, the Power increases from 30.6% to 86.8%.

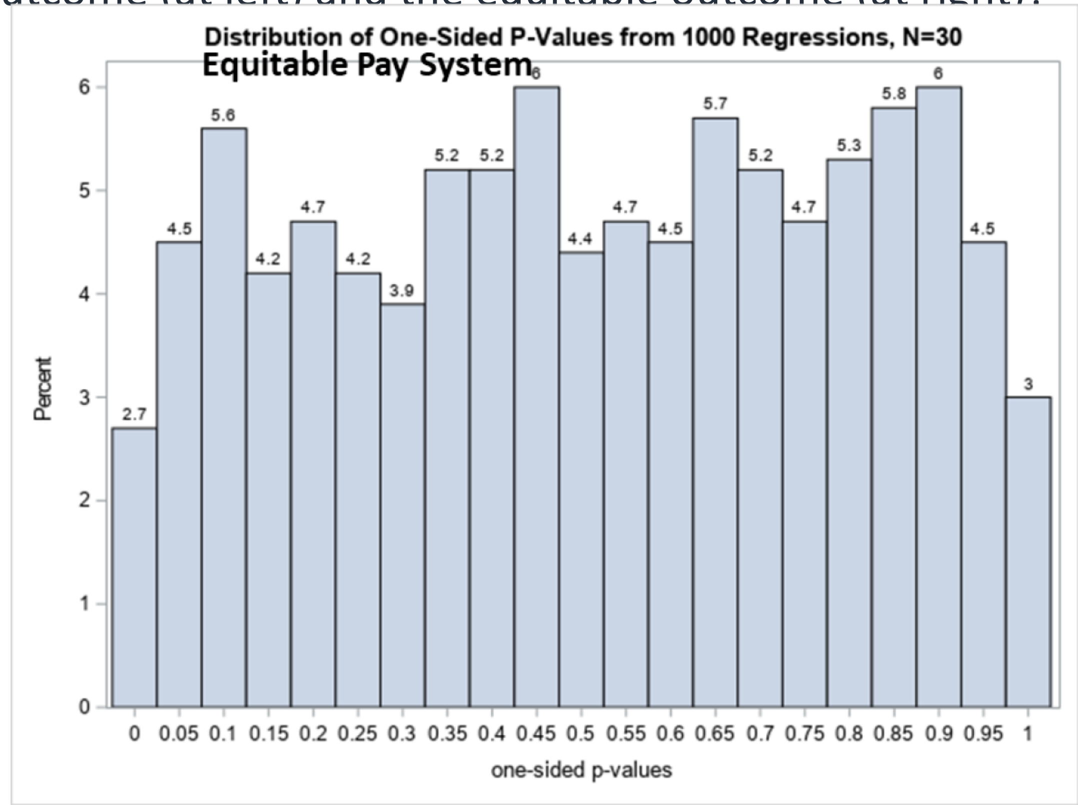




Introducing an Omnibus Test

- We know that BAFLinc.'s compensation system is biased against female employees.
- We can only detect this in 31% of groups using group-by-group regression analysis (see below left).
- We also see, however, that females are the disfavored group in 93.6% of groups (whether statistically significant or not).
- If the pay system were equitable, we would expect around 50% of groups to favor males, 50% to favor females (chart at right).
- Can we formally assess the difference between the skewed outcome (at left) and the equitable outcome (at right)?



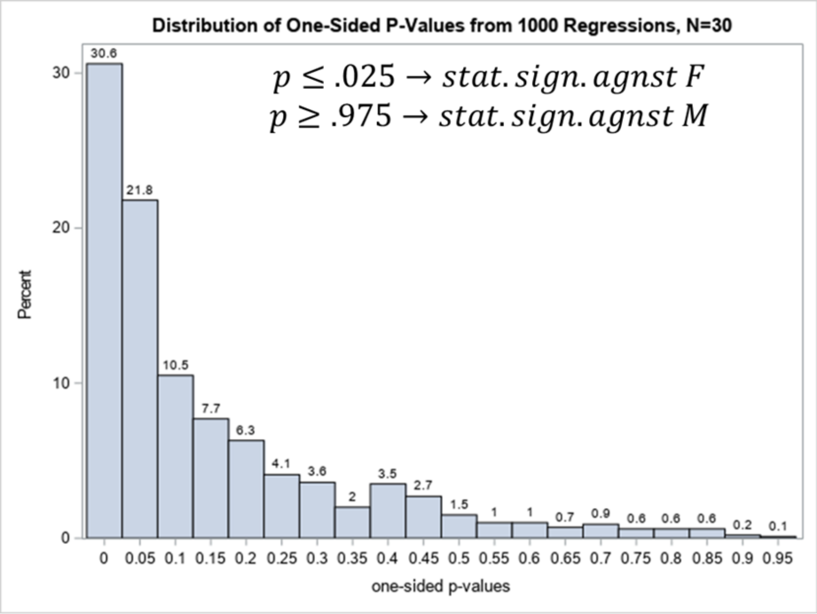


An Example of an Omnibus Test – Fisher's (Combined) Test

$$X^2_{2k} \sim -2\sum_{i=1}^{k} \ln(p_i)$$ (R. Fisher, 1932)

DISCLAIMER: While OFCCP is considering methods to combine individual tests to assess for systemic issues, the current example is simply an experiment and should not be interpreted as the Agency's prescribed method.

- The sum of the log transformation of one-side p-values across multiple strata, multiplied by negative 2, is distributed as a chi-squared random variable, with degrees of freedom equal to [2*(number of strata)].
- Otherwise said:
  - take the log of all p-values from individual regressions,
  - sum them and multiply by -2,
  - lookup value of this statistic in chi-squared "tables" to get a p-value for the whole.
- This p-value for the whole tells us if there is a systemic issue, there is a pattern across the workforce.



Estimate	t Value	Pr > |t|	trial	onesided	lnonesided	zvalue
-0.027203	-2.01	0.0558	1	0.02791	-3.5788	-1.91247
-0.011992	-0.66	0.5153	2	0.25765	-1.3561	-0.6506
-0.040477	-3.2	0.0037	3	0.00186	-6.2893	-2.90165
-0.021554	-1.52	0.1407	4	0.07033	-2.6545	-1.47331
-0.00617	-0.5	0.6205	5	0.31023	-1.1705	-0.49521
-0.025455	-1.73	0.0953	6	0.04765	-3.0439	-1.66809
-0.023115	-1.53	0.1397	7	0.06983	-2.6617	-1.47708
0.001114	0.08	0.9398	8	0.53011	-0.6347	0.07554
-0.019835	-1.76	0.0907	9	0.04537	-3.0928	-1.69146
-0.011453	-0.77	0.4506	10	0.2253	-1.4903	-0.7544
-0.008134	-0.58	0.566	11	0.28298	-1.2624	-0.574

Fisher Stat	6131.041
Fisher P	0E+00

Step 1: take natural log of 1-sided p-values from each regression **[=ln('pvalue')].**

Step 2: sum all values from step 1 and multiply by -2 **[=sum(g2:g1002)*-2].**

Step 3: use chi-sq table, look up Fisher's Statistic with degrees of freedom equal to 2*(number of PAGs) **[=chisq.dist.rt(h2,2000)]**.

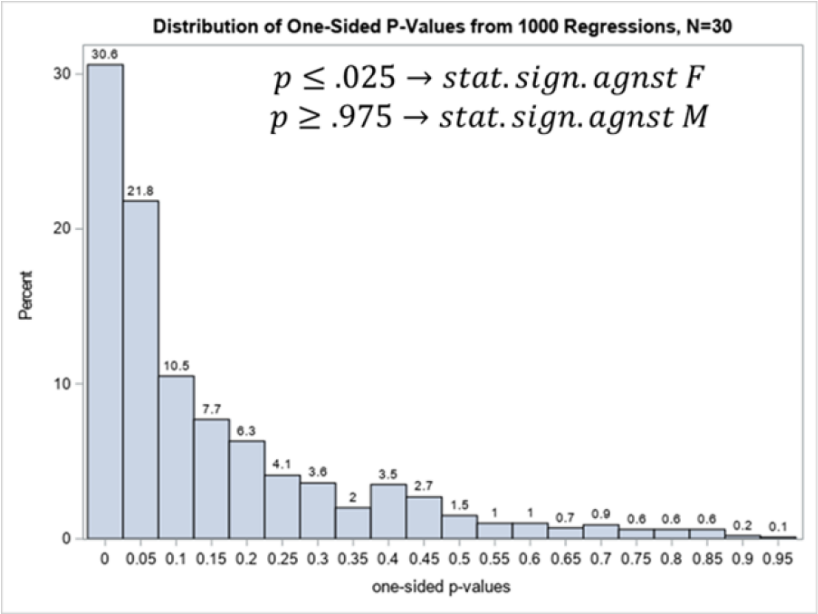
The resulting p-value, nearly zero, allows us to reject (with an extremely high level of confidence) the assertion that the outcomes and patterns we're seeing in this case are the result of a neutral pay process.

DISCLAIMER #2: The basic Fisher's Combination method requires that pay in each PAG be independently determined – while OK for this example, any prescribed will need to account for non-independence.

An Example of an Omnibus Test – Stouffer's (Combined) Test

DISCLAIMER: While OFCCP is considering methods to combine individual tests to assess for systemic issues, the current example is simply an experiment and should not be interpreted as the Agency's prescribed method.

- Otherwise said:
  - For each regression's p-value, convert to z-score,
  - sum them and divide by square root of number of groups (PAGs),
  - lookup value of this statistic in z "tables" to get a p-value for the whole.
- This p-value for the whole tells us if there is a systemic issue, there is a pattern across the workforce.



Estimate	t Value	Pr > |t|	trial	onesided	lnonesided	zvalue
-0.027203	-2.01	0.0558	1	0.02791	-3.5788	-1.91247
-0.011992	-0.66	0.5153	2	0.25765	-1.3561	-0.6506
-0.040477	-3.2	0.0037	3	0.00186	-6.2893	-2.90165
-0.021554	-1.52	0.1407	4	0.07033	-2.6545	-1.47331
-0.00617	-0.5	0.6205	5	0.31023	-1.1705	-0.49521
-0.025455	-1.73	0.0953	6	0.04765	-3.0439	-1.66809
-0.023115	-1.53	0.1397	7	0.06983	-2.6617	-1.47708
0.001114	0.08	0.9398	8	0.53011	-0.6347	0.07554
-0.019835	-1.76	0.0907	9	0.04537	-3.0928	-1.69146
-0.011453	-0.77	0.4506	10	0.2253	-1.4903	-0.7544

Stouffer Stat -46.5465853

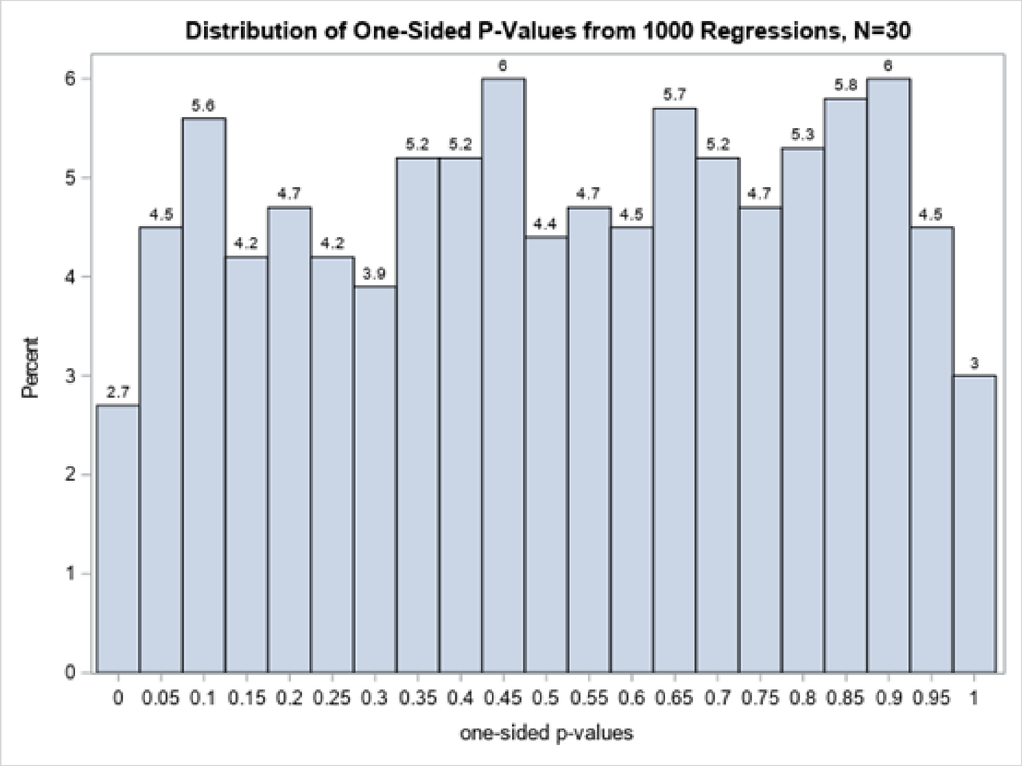
Stouffer P 0E+00

Step 1: convert 1-sided p-value to z-value
[=norm.s.inv('pvalue')]
Step 2: sum all values from step 1 and divide by sqrt(N)
[=sum(h2:h1002)/sqrt(1000)]
Step 3: use standard normal (Z) table, look up Stouffer's Statistic to find Stouffer's P value.
[=norm.s.dist('stouffers stat',"true")]

DISCLAIMER #2: The basic Stouffer's Combination method requires that pay in each PAG be independently determined – while OK for this example, any prescribed will need to account for non-independence.

Bonus Example – Mistaken Discovery

- In bias-free case, if PAGs are independent from each other, we expect any false positives to be equally split between those indicating bias for males & those for females.
- Otherwise said, the p-values should be distributed evenly on [0,1] interval.
- Use combined test to see if the p-distribution across PAGs (as shown at left) may come from an equitable system.



pagsize	sum	fishers	ztotal	p_fishers	p_stouffer
30	-978.421	1956.84	1.2872	0.75073	0.90099

Both Fishers Test P and Stouffers Test P, both being within (.025,.975), indicate an extremely low likelihood of bias pattern, whether adversely impacting females or males.

Main Takeaways

- An myopic analysis of individual pay analysis groups prevents comprehensive assessment of systemic pay issues.
    - Misinterpreting an isolated occurrence as a general pattern.
    - Misinterpreting smaller, but consistent, occurrences as pattern-free.

- The traditional approach to pay equity assessment, PAG-by-PAG analysis, is missing a critical piece – a summary, or combination, test -- for it to be considered systemic pay equity analysis.

- Simple tests, including Fisher's Combination Test (among many others, depending on the context), are easy to implement.

Thank you for attending!

SAVE THE

DATE!

We hope you'll join us at the NILG 2021 National Conference

August 1 – August 4, 2021
Omni® Nashville Hotel
Nashville, Tennessee

Win a Complimentary Registration to the NILG 2021 National Conference!

TO LEARN MORE VISIT:

https://www.nationalilg.org/2021-conference-giveaway/

NILG
2020 VIRTUAL CONFERENCE
WEBINAR SERIES

Thank YOU
SPONSORS
FOR YOUR SUPPORT

affirmity
AMERICA'S JOB exchange®
berkshire
biddle
opportunity. for all.
DCi CONSULTING
GA GAUCHER ASSOCIATES
Affirmative Action Professionals
HUDSONMANN
localjobnetwork
LOCKHEED MARTIN
MERCK
INVENTING FOR LIFE
NORTHROP GRUMMAN
PEPSICO
pepsi. FritoLay Tropicana QUAKER G
resolution economics LLC
taapestry
by DirectEmployers
vizient®

2020 VIRTUAL CONFERENCE
WEBINAR SERIES

Thank You

Join the NEW NILG LinkedIn Page to stay current on agency news, free NILG webinars and national conference updates:
https://www.linkedin.com/company/nilg

SHRM PDC # 20-KGTYH

HRCI Activity # 527152

"The use of this official seal confirms that this Activity has met HR Certification Institute's® (HRCI®) criteria for recertification credit pre -approval."

2020 VIRTUAL CONFERENCE
WEBINAR SERIES