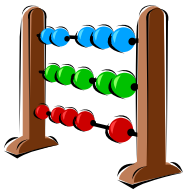
Understanding Statistics



Reasons for Analyzing Data

- Describe data
- Determine if two or more groups differ on some variable
  - Adverse impact analysis
- Determine if two or more variables are related
  - Reliability and validity of selection methods
  - Determining if merit variables explain group salary difference
- Reduce data

Types of Data

- Nominal
  - categories
    - race
    - hair color
- Ordinal
  - rank order
    - baseball standings
    - waiting list placements
    - Categories containing salary intervals (e.g., $10-$15, $16-$20)
- Interval
  - equality of intervals
    - performance ratings
    - temperature
- Ratio
  - true zero
  - equality of ratios
    - salary
    - height

Let's Practice

Variable	Type of Scale
Birth place	
Weight	
Class standing in high school	
Salary grade	
# Days absent from work	
Grade point average	

The Concept of Significance

- Interoccular Significance
- Statistical Significance
- Practical Significance

Significance Levels

- Indicate the probability that results occurred by chance
  - Indicates confidence in interpreting data
  - Does not indicate practical significance
- Standard is .05, but others can be used
  - Type I error: Concludes there is a difference when in fact there is none
  - Type II error: Concludes there is no difference when there is one

Statistical Significance

- When deviating from the .05 level, consider
  - the common sense of your finding
  - previous research
  - the quality of your data
  - the cost of being wrong
- Probability level is influenced by
  - sample size
  - differences between groups
  - within group variability

Significance Levels in Journal Articles

The job satisfaction level of female employees (M=4.21) was significantly higher than that of male employees (M=3.50), t (60) = 2.39, p < .02.

	Academy Score	Commendations
Cognitive ability	.43**	.03
Education	.28**	.24*

* p < .05, ** p < .01, *** p < .001

Statistics That Describe Data

Raw Data Are Not Usually Meaningful

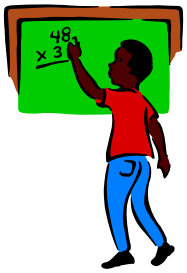
Applicant	IQ
Caffey	98
Doherty	104
Yokas	110
Boscarelli	93
Sullivan	121
Parker	114
Davis	99

Statistics That Describe Data

- Sample Size
  - overall (N)
  - subgroups (n)
- Frequencies
- Ranks
- Central Tendency
  - mean (statistical average)
  - median (midpoint)
  - mode (most common)
- Dispersion
  - range
  - variance & standard deviation

Measures of Central Tendency

- Mean
- Median
- Mode



The Mean

Applicant	IQ
Boscarelli	93
Caffey	98
Davis	98
Doherty	104
Yokas	110
Parker	114
Sullivan	121
Sum	738
N	7
Mean	105.4

The Median

- Median is the point at which 50% of your data fall above and 50% fall below
- Odd number of scores, the median is the middle score
- Even number of scores, the median is the average of the two middle scores

93 98 98 100 104 110 114 121

102

The Median

Applicant	IQ
Boscarelli	93
Caffey	98
Davis	98
Doherty	104
Yokas	110
Parker	114
Sullivan	121

The Mode
The Most Frequently Occurring Score

Applicant	IQ
Boscarelli	93
Caffey	98
Davis	98
Doherty	104
Yokas	110
Parker	114
Sullivan	121

Which Measure Should I Use?
Theoretical Considerations

- Mode
  - What is the most probable event
- Mean
  - Infer total resources
  - Predict future resources
  - Use in comparative analyses
- Median
  - Describe the "typical" person or situation

Which Measure Should I Use?
Statistical Considerations

- Mode
  - nominal data (categories)
- Mean
  - interval data
  - ratio data
- Median
  - ordinal (ranked) data
  - interval or ratio data if
    - outliers
    - skewed distribution

	Supervisor	
	Newhart	Brothers
	17	17
	18	18
	19	19
	20	20
	21	21
	22	22
	23	51
Mean	20	24
Median	20	20

If data are skewed, but you want to use the mean

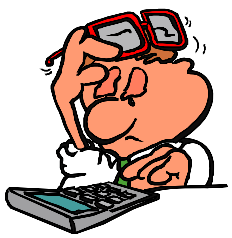
- Transform the data
  - Square-root (slight, affects the tails)
  - Logarithmic (moderate skew, affects center)
  - Inverse (large skew)
- Use a mean other than the arithmetic mean
  - M-estimators give less weight to values further from the center

Measures of Dispersion

- Range
  - Minimum
  - Maximum
  - Spread
- Variance (s^2)

$$\frac{\sum(score - mean)2}{N - 1}$$

- Standard deviation (s)
  - Square root of the variance
  - 1 SD = 68% of scores
  - 2 SD = 95% of scores



Number of Days Absent

	Day Shift	Evening Shift	Night Shift
	4	2	3
	4	3	3
	4	4	4
	4	5	5
	4	6	5
Mean	4	4	4

Performance Ratings

	Supervisor	
	Tribbiani	Geller
	3	2
	3	2
	3	3
	3	3
	3	4
	3	4
Mean	3	3

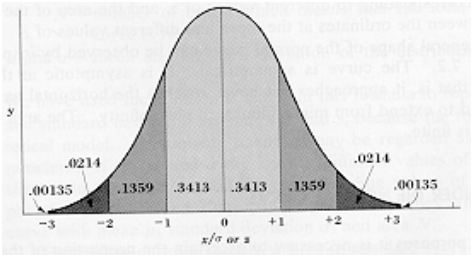
IQ Scores for Two Training Groups

Training Group	Mean IQ	SD	1 SD Range	2 SD Range
Morning	100	3	97 – 103	94 – 106
Afternoon	100	15	85 – 115	70 - 130

Salary Survey Example

- Salary Survey Data
  - Mean for police officer is $25,000
  - SD = $3,000
- Our Salary
  - $24,000

The Normal Curve

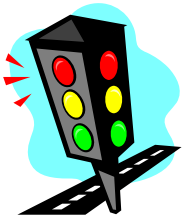


We know that an applicant's test score is one standard deviation above the mean

Standard Deviation	Cumulative %
- 3.0	0.14
- 2.0	2.28
- 1.5	6.68
- 1.0	15.87
- 0.5	30.85
0.0	50.00
+ 0.5	69.15
+ 1.0	84.13
+1.5	93.32
+2.0	97.72
+3.0	99.86

Caution About Inferences From Standard Deviations

- Inferences can be made only when
  - Data are normally distributed
  - Sample size is large
- If conditions are not met, using percentiles based on actual data is best



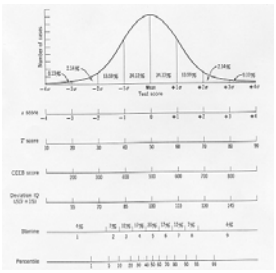
Employee	Day Shift	Night Shift
A	1	1
B	2	1
C	2	1
D	3	1
E	3	1
F	3	1
G	4	1
H	4	1
I	4	1
J	4	1
K	5	1
L	5	1
M	5	1
N	5	9
O	5	9
P	5	9
Q	6	9
R	6	9
S	6	9
T	6	9
U	7	9
V	7	9
W	7	9
X	8	9
Y	8	9
Z	9	9

Number of citations issued

	Day Shift	Night Shift
Mean	5.00	5.00
SD	2.00	4.08
1 SD Range	3 – 7	.92 – 9.08
2 SD Range	1 – 9	- 5.0 – 14.16

Measures of Comparison and Explanation

- Percent
- Percentile
  - Q1
  - Q2
  - Q3
- Standard Score (Z)
  - mean of zero
  - standard deviation of 1
- T-Score



Let's Think

- I am going on a cruise and want to know the ages of the passengers so I'll know if I fit in
- I am traveling to a new country and want to get information about the temperature so that I will know what to wear
- I am planning a trip to the Bahamas in September and want to know the odds that there will be a hurricane
- I want to get information on the frequency of employee absenteeism in my company
- I want to know if the salaries of men and women are different

Using Descriptive Statistics to Ensure Data Integrity

- Reasons for Errors
  - Inaccurate source data
  - Copied incorrectly from source data
  - Input error
    - misread
    - keystroke error
    - conversion error
  - Input statement error
- Methods to Check
  - Proofread raw data
  - "Sure thing" analysis that didn't work
  - Use descriptive statistics to
    - check for values outside the possible range
    - check for values that don't make sense

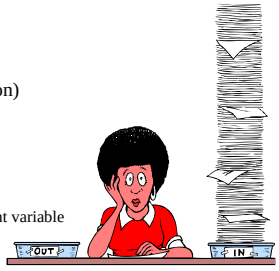
Statistics That Test Differences Between Groups

Examples

- Do average salaries differ between men and women (salary equity analysis)?
- Are female employees found more often in administrative positions and male employees in executive positions (glass ceiling analysis)?
- Did we hire a higher proportion of white applicants than we did minority applicants (adverse impact)?
- Did turnover go down after we started using our new structured interview?

What Statistic to Use

- Frequencies
  - Chi Square
- Medians
  - Fisher's exact test
- Ranks
  - Mann-Whitney (Wilcoxon)
- Means
  - two groups: *t*-test
  - Analysis of Variance
    - more than two groups
    - more than one independent variable



Differences in Frequencies
Chi-Square

- Goodness of Fit
  - Does the observed frequency differ from the expected frequency
  - Example

	%	%
• Secretary	92	80
• Welder	20	25
• Supervisor	40	50

- Tests of Independence
  - Does the distribution for one group differ from that of another
  - Example

	Hired	Not
• Male	32	16
• Female	10	20

Differences in Frequencies
Standard Deviation Method

- Confidence interval approach
- Number of standard deviations approach

Confidence Interval Approach

1. Compute Standard Deviation

$$\sqrt{\frac{\text{female applicants}}{\text{total applicants}} \times \frac{\text{male applicants}}{\text{total applicants}} \times \text{total hired}}$$

2. Multiply standard deviation by 2
3. Compute expected number of females to be hired
   (female applicants/total applicants) x total hired
4. Compute confidence interval (expected ± 2 SD)
5. Determine if number of females hired falls within the confidence interval

Confidence Interval Example

1. Compute Standard Deviation

$$\sqrt{\frac{10}{50} \times \frac{40}{50} \times 20} = \sqrt{.20 \times .80 \times 20} = \sqrt{3.2} = 1.79$$

2. Multiply standard deviation by 2 = 1.79 * 2 = 3.58
3. Compute expected number of females to be hired
   (10 ÷ 50) x 20 = .2 x 20 = 4
4. Compute confidence interval (.42 ← 4 → 7.58)
5. Determine if number of females hired falls within the confidence interval

Number of SDs Approach

1. Compute Standard Deviation

$$\sqrt{\frac{\text{female applicants}}{\text{total applicants}} \times \frac{\text{male applicants}}{\text{total applicants}} \times \text{total hired}}$$

2. Compute expected number of females to be hired
   (female applicants/total applicants) x total hired
3. Subtract the actual number of women hired from the expected number of women to be hired
4. Divide the difference in step 3 by the standard deviation computed in step 1

Number of SDs Example

1. Compute Standard Deviation

$$\sqrt{\frac{10}{50} \times \frac{40}{50} \times 20} = \sqrt{.20 \times .80 \times 20} = \sqrt{3.2} = 1.79$$

2. Compute expected number of females to be hired
   (10 ÷ 50) x 20 = .2 x 20 = 4
3. Subtract the actual number of women hired from the expected number of women (4 – 2 = 2)
4. Divide the difference in step 3 by the standard deviation computed in step 1 (2 ÷ 1.79 = 1.12)

Let's Practice!

	White	Black
Number of applicants	544	133
Number hired	54	2
Selection ratio	.099	.015

.015/.099 = .15 < .80 (adverse impact)

Confidence Interval Answer

1. Compute Standard Deviation

$$\sqrt{\frac{133}{677} \text{ X } \frac{544}{677} \text{ X } 56} \qquad \sqrt{.196 \times .804 \times 56} \qquad \sqrt{8.83} = 2.97$$

2. Multiply standard deviation by 2 = 2.97 * 2 = 5.94
3. Compute expected number of Blacks to be hired
   (133/677) x 56 = .196 x 56 = 11
4. Compute confidence interval (5.04 ← 11 → 16.94)
5. Determine if number of Blacks hired (2) falls within the confidence interval

Number of SDs Answer

1. Compute Standard Deviation

$$\sqrt{\frac{133}{677} \text{ X } \frac{544}{677} \text{ X } 56} \qquad \sqrt{.196 \times .804 \times 56} \qquad \sqrt{8.83} = 2.97$$

2. Compute expected number of Blacks to be hired
   (133 ÷ 677) x 56 = .196 x 56 = 11
3. Subtract the actual number of Blacks hired from the expected number of Blacks (11 – 2 = 9)
4. Divide the difference in step 3 by the standard deviation computed in step 1 (9 ÷ 2.97 = 3.03)

Differences in Medians
Fisher's Exact Test

- A Fisher's Exact Test is used when determining whether the median salaries differ statistically for smaller groups
  - HR Equator automatically splits employees into those that are above the median and those that are below
- Generally, a Fisher's Exact Score of .025 or less is considered statistically significant
  - One-tail test is .05
  - Two-tail test is .025

Differences in Ranks
Mann-Whitney

- A statistical test that rank-orders salaries and looks for sex and race differences in average ranks
- Preferred by some statisticians and contractors over the *t*-test due to problems with outliers and assumptions about salary not being normally distributed
- Nothing wrong with Mann-Whitney but we prefer the *t*-test
  - Results will be consistent with and can be compared to regression results
  - Ranks don't capture salary differences between ranks

Example A

Salary	Sex	Rank
$28,000	M	1
$27,000	F	2
$26,000	F	3
$25,000	M	4
$24,000	F	5
$23,000	M	6
$22,000	M	7
$21,000	F	8
	Men	Women
Average Salary	$24,500	$24,500
Average Rank	4.5	4.5

Example B

Salary	Sex	Rank
$34,000	M	1
$26,100	F	2
$26,000	F	3
$25,900	M	4
$24,000	F	5
$23,900	M	6
$23,000	M	7
$17,000	F	8
	Men	Women
Average Salary	$26,700	$23,275
Average Rank	4.5	4.5

Aggregating Analyses

- Unit of Analysis
  - By hiring decision – determine adverse impact for each requisition
  - Aggregated across hiring requisitions, jobs, years, etc.
- Mantel-Haenszel Test
  - Computes an overall impact ratio across a set of subgroups.
  - Measures overall disparity controlling for subgroup membership (e.g., hiring period or job position).
  - Permits multiple divisions of the data without loss of statistical power.
- Breslow Day Test
  - Used to determine if the aggregated impact ratios have similar magnitude and can be combined into one analysis.

Example

Assembler

Sex	Selected	Not-Selected	Total
Women	66	91	157
Men	154	66	220
Total	220	157	377

Odds ratio .311 (3.2 times)
Impact ratio .601 (.42/.70)
Shortfall 25
# of Std Errors 5.42

Quality Control

Sex	Selected	Not-Selected	Total
Women	12	17	29
Men	27	5	32
Total	39	22	61

Odds ratio .131 (7.6 times)
Impact ratio .490 (.41/.84)
Shortfall 6
# of Std Errors 3.49

Shipping

Sex	Selected	Not-Selected	Total
Women	3	4	7
Men	14	7	21
Total	17	11	28

Odds ratio .375 (2.7 times)
Impact ratio .643 (.43/.67)
Shortfall 1
Fisher's Exact p-value .249 One-tailed

Mantel Haenszel
Common Impact Ratio .5848
Test of Independence $p < .0001$

Breslow-Day
p-value
Test for Homogeneity .3899

- **Mantel Haenszel test** shows overall statistically significant association between gender and selection status across three job groups.
- **Breslow-Day test** indicates disparity is homogeneous across jobs.

The t-test

Tests Differences in Means Between Two Groups

	Sex	
	Male	Female
Salary	$46,000	$43,000
	Race	
	Nonminority	Minority
Interview score	52.6	47.3

t-test

- Assumptions
  - Normal distribution
  - Equal variances in each group
- *Generally*, a *t*-test score of 2.00 or higher is considered statistically significant
- Size and significance of the *t*-value
  - Mean group differences in salary
  - Variability of salaries in each group
  - Sample size

t-value Needed for Significance

Degrees of Freedom	Significance Level (2-tailed)	
	.05	.01
10	2.228	3.169
15	2.131	2.947
20	2.086	2.845
30	2.042	2.750
40	2.021	2.704
60	2.000	2.660
120	1.980	2.617

Company A

Men	Women
$22,000	$18,000
$22,000	$18,000
$22,000	$18,000
$22,000	$18,000
$22,000	$18,000
Mean = $22,000	Mean = $18,000

Significant

Company B

Men	Women
$18,000	$10,000
$20,000	$16,000
$22,000	$18,000
$24,000	$20,000
$26,000	$26,000
Mean = $22,000	Mean = $18,000

Not Significant

Analysis of Variance

Tests differences in means when there

Are more than two groups

White	$23,121
African-American	$20,243
Hispanic	$21,176
Texan	$18,543

Is more than one independent variable

	White	Black	Total
Male	$28,100	$21,900	$25,000
Female	$24,000	$22,000	$23,000
Total	$26,050	$21,950	$24,000

Is an interaction between the two independent variables

Interpreting the Results of an ANOVA

	DF	SS	MS	F	p <
Sex	1	382106006	382106006	13.16	.0004
Race	1	42857538	42857538	1.48	.2260
Race * Sex	1	14079430	1079430	0.48	.4871
Error	174	5051526673	29031762		
Total	177	935490569647			

	White	Black	Total
Male	$45,008	$43,349	$44,621
Female	$41,556	$41,330	$41,505
Total	$43,874	$42,708	

Interpreting an ANOVA

- What is an F Ratio?
  - The between group variance divided by the within group variance
  - An F of 1.0 indicates that there are equal amounts of within and between groups variance
  - *t* is the square root of F
  - significance determine by size of F and sample size
- Sample Size Cautions
  - Sample size in each cell should be reasonable (at least 10)
  - Sample size in each cell should be about equal or at least proportional to the marginal totals

Multiple Comparisons Example

Employee Education	Performance Rating
GED	3.13
High school diploma	3.41
Associate's degree	4.26
Bachelor's degree	4.35
Master's degree	4.37

Correlation

What is a Correlation?

- Correlation is a statistic that determines the *relationship* between two variables
- Correlations do not infer causation
- The values of a correlation range from 0 to +1 and 0 to –1. The further the number is from zero, the greater the relationship between the two variables
- The + and – indicate the direction of the correlation

Correlation

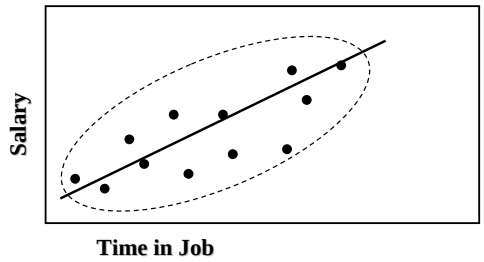
- To determine if we are even allowed to interpret a correlation coefficient, we must determine if that coefficient is statistically significant
- The significance level for a correlation is a function of two factors, the sample size and the size of the correlation coefficient

Sample Size	Smallest Significant Correlation
10	.63
20	.44
30	.36
40	.31
50	.27
60	.25
70	.22
100	.19

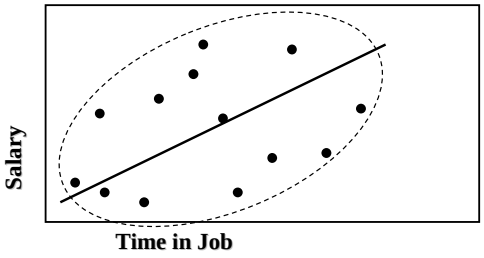
Interpreting Correlations

- Direction
  - positive
  - Negative
- Type of Relationship
  - Linear
  - Curvilinear
- Magnitude
  - distance from zero
  - comparison to norms
  - utility analysis

Correlation of .50



Correlation of .20



Example of a Positive Correlation

Performance Rating	12	13	14	15	16	17	18	19	20	21	22	23	24	25
6													x	
5									x			x		
4						A	B					x	x	
3				x		C	D		x	x	x			
2		x	x		x									
1	x													
	Test Score													

Example of a Negative Correlation

6	x													
5		x		x										
4					x	A	B		x	x				
3					x	C	D		x	x				
2				x								x		
1													x	
Performance Rating	12	13	14	15	16	17	18	19	20	21	22	23	24	25
	Test Score													

Example of Two Uncorrelated Variables

6	x													
5		x		x							x			
4						A	B		x	x				
3					x	C	D		x	x				
2		x		x								x		
1														
Performance Rating	12	13	14	15	16	17	18	19	20	21	22	23	24	25
	Test Score													

Figure 4.5
Example of a Curvilinear Relationship

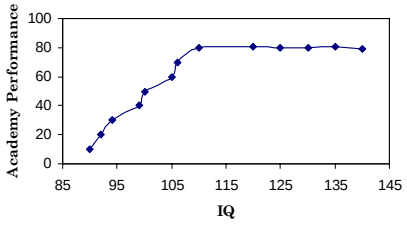
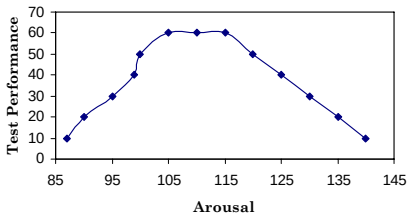


Figure 4.6
Example of the "Inverted U"



Interpreting Correlations

- Magnitude
- Comparisons to norms
- Utility

Typical Corrected Validity Coefficients for Selection Techniques

Method	Validity	Method	Validity
Structured Interview	.57	Experience	.27
Cognitive ability	.53	References	.29
Biodata	.51	Conscientiousness	.24
Job knowledge	.48	Unstructured interviews	.20
Work samples	.48	Emotional stability	.15
Assessment centers	.38	Interest inventories	.10
Situational judgment	.34	Extraversion	.09
Integrity tests	.34	Handwriting analysis	.09
College grades	.32	Projective personality tests	.00

Average Correlations with Salary

Merit Variable	Average Correlation
Experience	.24
Age	.22
Time in Company	.21
Time in Grade	.17
Education	.10
Performance	.08

Utility Analysis
Taylor-Russell Tables

- Estimates the percentage of future employees that will be successful
- Three components
  - Validity
  - Baserate (successful employees ÷ total employees)
  - Selection ratio (hired ÷ applicants)

Utility Analysis
Brogden-Cronbach-Glesser

- Utility = (n) (t) (r) (SD$) (m) – cost of testing
  - N = number of applicants hired each year
  - T = average tenure
  - r = test validity
  - SD$ = standard deviation of performance in dollars (40% of salary)
  - *m* = standardized selection ratio
  - Cost of testing = # applicants * cost of the test

Standardized Selection Ratio

SR	m
1.00	.00
.90	.20
.80	.35
.70	.50
.60	.64
.50	.80
.40	.97
.30	1.17
.20	1.40
.10	1.76
.05	2.08

Exercise 1

You have 500 applicants and 250 job openings. You are considering using a structured interview to select employees. The uncorrected validity of the structured interview is .30 and costs $15 per person to administer. You have 900 current employees, 560 of which are satisfactory. The salary for the position is $40,000 and the typical employee stays for 3 years. Currently you are using a test with a validity of .20 that costs $25 per applicant to administer and score.

Exercise 1 Questions

- Based on norms, how good is your structured interview?
- Using the Taylor-Russell Tables, what percent of future employees will be successful if you adopt the structured interview?
- Using the utility formula, how much money will the structured interview save the company over doing nothing at all?
- How about compared to the test currently being used?

Exercise 2

You have 300 applicants and 100 job openings. The uncorrected validity of your proposed integrity test is .40 and costs $30 per person to administer. You have 700 current employees, 600 of which are satisfactory. The salary for the position is $30,000 and the typical employee stays for 2 years. Currently you are using a test with a validity of .10 that costs $35 per applicant to administer and score.

Exercise 2 Questions

- Based on norms, how good is your integrity test?
- Using the Taylor-Russell Tables, what percent of future employees will be successful if you adopt the integrity test?
- Using the utility formula, how much money will the integrity test save the company over doing nothing at all?
- How about compared to the test currently being used?

Interpreting Correlations

- Types of Correlation
  - Pearson
  - Spearman rank order
  - Point biserial

Regression Analysis

The Concept of Regression

- Advantages
  - Enables prediction
  - Allows combinations of small correlations
  - Accounts for overlap of variables
- Two common types in HR
  - Least squares
    - Used when evaluating interval or ratio data
    - Salary, ratings of job performance
  - Logistic (Logit)
    - Used in evaluating dichotomous decisions
    - Hire/not hire, interview/not interview, promote/not promote
- Two main strategies
  - Stepwise
  - Hierarchical

Regression Formula

$$Y = a + (b1)\,(x1) + (b2)\,(x2)$$

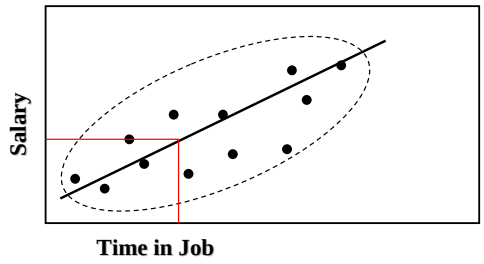
Y = predicted criterion score

a = constant (intercept)

b = weight (slope)

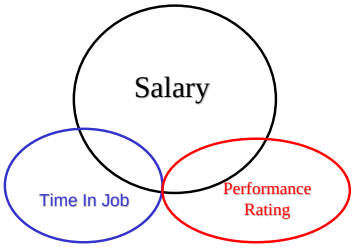
x = score on the predictor

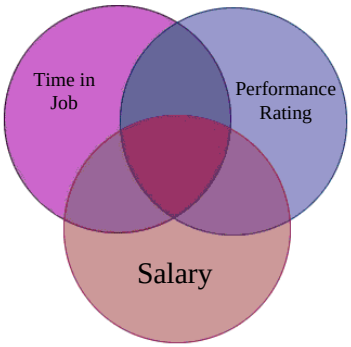
Predicting Salary from Time in Job



Considerations

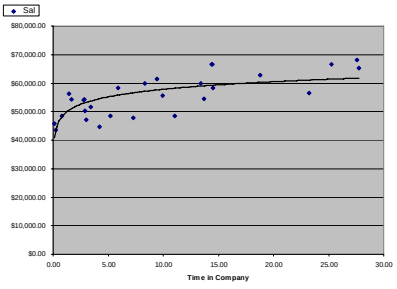
- Number of employees in the analysis
  - Total number (30 is probably the minimum)
  - Number of employees in each group (5 is the minimum)
- Employee-to-variable ratio
  - Minimum is a ratio of 5:1
  - 10:1 is more comfortable
- Missing variables
- Inclusion of nonsignificant variables
- Multicollinearity
- Balance of simplicity versus statistical power

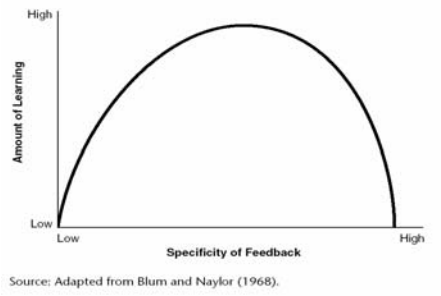




Types of Relationships

- Linear
  - Each increment in the merit variable results in an increase in salary
- Yes/No
  - Use dummy codes
    - College degree (0=no, 1=yes)
    - Manager (0=no, 1=yes)
- Nonlinear
  - Relationship levels off or slows
  - Relationship changes direction (e.g., inverted U)
- Interactions





Source: Adapted from Blum and Naylor (1968).

Interactions

- Relationship between a merit variable and salary "depends on" or "interacts with" another variable

	College Degree		
Manager	No	Yes	
No	$22,000	$22,000	$22,000
Yes	$26,000	$38,000	$32,000
	$24,000	$30,000	

Interactions

- You multiply the two variables in regression to account for the interaction

Employee	Manager	Bachelor's Degree	Interaction
Hill, E.D.	1	1	1
Kilmeade, Brian	1	0	0
Doocy, Steve	0	1	0
Chetry, Kiran	0	0	0
Wright, Kelly	1	1	1
Hopkins, Page	0	0	0
Green, Lauren	1	1	1
Camerota, Alisyn	0	1	0

Interpreting Regression Results

Variable	Regression Weight	r^2	R^2	F	P<
Constant	3.67				
IQ	0.10	.151	.151	3.69	.05
Interview	.59	.036	.187	3.69	.05
Model				9.57	.001

Performance = 3.67 + (.10)(IQ) + (.59)(Interview)

Regression Statistics

How Well Do Our Merit Variables Explain Pay?

Multiple R	.73
R-square	.53
Adjusted R-square	.51
Standard error	$5,159
Observations	119

Is Our R^2 Statistically Significant?

ANOVA	df	SS	MS	F	Sig of F
Regression	5	3,434,436,542	686,887,308	25.80	.01
Residual	113	3,008,138,683	26,620,696		
Total	118	6,442,575,225			

Which Variables Significantly Explain Pay?

	Coefficient	Standard Error	*t*-test	p	Beta
Intercept	$34,746	$3,312.71	10.49	.00	0.00
TiG	$715.20	99.65	7.18	.00	0.51
TiJ	$842.96	331.79	1.92	.07	0.13
Age	$239.60	72.91	3.29	.00	0.23
TiC	$1,507.69	489.57	3.08	.00	0.20
Performance	$3,964.96	883.66	4.49	.00	0.30

What is the Formula to Estimate Salary?

	Coefficient	Standard Error	*t*-test	p	Beta
Intercept	$34,746	$3,312.71	10.49	.00	0.00
TiG	$715.20	99.65	7.18	.00	0.51
TiJ	$842.96	331.79	1.92	.07	0.13
Age	$239.60	72.91	3.29	.00	0.23
TiC	$1,507.69	489.57	3.08	.00	0.20
Performance	$3,964.96	883.66	4.49	.00	0.30

Regression Formula

$$Y = a + (b1)\,(x1) + (b2)\,(x2)$$

Y = predicted salary
a = constant (intercept)
b = weight (slope)
x = score on the merit variable

Salary = $34,746 + ($715*TiG) + ($843*TiJ) + ($240*Age) + ($1,508*TiC) + ($3,965*Performance)

What should Jane be earning?

Salary = $34,746 + ($715*2) + ($843*1) + ($240*32) + ($1508*5) + ($3,965*3)

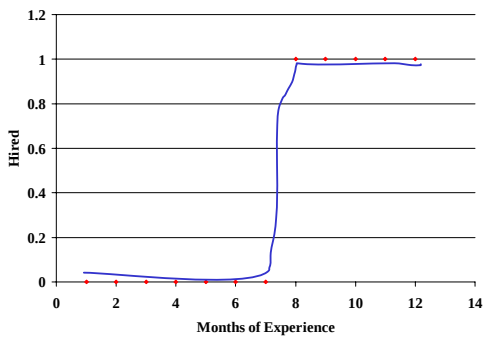
Salary = $34,746 + ($1,430) + ($843) + ($7,640) + ($7,540) + ($11,895)

Predicted Salary = $64,094

Logistic Regression

Basics of Logistic Regression

- Used when the variable being predicted is dichotomous (e.g., not hired/hired) rather than continuous (e.g., salary)
- Primarily used in OFCCP enforcement efforts for such decisions as
  - Contacting applicants
  - Interviewing
  - Hiring
  - Promoting
  - Firing
- Rule of thumb is to have at least 50 employees for every predictor variable in the equation
- Predictor variables can be dichotomous or continuous



Example

- You believe that your selection decisions are based on
  - Years of previous experience
  - The number of months spent on last job
  - Completing the application correctly
- You want to determine
  - Are we really using these three variables
  - After accounting for these three variables, is there still an effect for sex

Sample Logistic Regression Output

Logistic Regression Results for Applicants Considered for Position

Predictor Variables	Dependant Variables: (MODEL 1) Offers for Hire B	SE	Exp(B)	(MODEL 2) Applicants Interviewed B	SE	Exp(B)	(MODEL 3) Offers for Hire Among Those Interviewed B	SE	Exp(B)
Female	-.961	.242 **	.383	-.985	.200 **	.373	-.378	.324	.685
Complete Application	-.483	.291	.617	-.002	-.249	.998	-.865	.427*	.421
Tenure	.959	.223**	2.608	.794	.184**	2.213	.660	.291*	1.934
Previous Experience	1.627	.216**	5.089	1.480	.185**	4.391	.794	.317*	2.212
Total N	729			729			270		
Model Chi-Square	104.37 **			133.9**			16.1*		
% Correct Predictions	78.8			70.0			61.2		
Nagelkerke-R^2	.209			.232			.078		

*Significant at the .05 level
**Significant at the .001 level.

Interpretation

- **Model 1: Males were 2.6 times more likely than females to be offered a position, controlling for tenure, experience, and a complete application.**
- **Model 2: Males were 2.7 times more likely than females to be offered an interview.**
- **Model 3: Among those interviewed, males were no more likely than females to be offered a position.**

QUESTIONS