2016 Swarthmore Honors Examination in Statistics

Weiwen Miao
Haverford College

May 2016

Instructions: The exam has four questions. Number the questions clearly on your answer sheets. You must make it clear how you arrived at your answer. Answers without any work may lose credit even if they are correct.

This is a closed-book three-hour exam. You may not refer to notes or textbooks. You may use a calculator that does not do algebra or calculus. Normal, t and F tables should be supplied with this exam.

1. Let $X_1, X_2, \ldots, X_n$ be a random sample from Uniform distribution $U(0, \theta)$.
   (a) Find the method of moment estimator and the maximum likelihood estimator for θ. Call them $\hat{\theta}_{MM}$ and $\hat{\theta}_{MLE}$, respectively.
   (b) Is $\hat{\theta}_{MM}$ unbiased?
   (c) Is $\hat{\theta}_{MM}$ consistent?
   (d) Use the Central Limit Theorem to find the limiting distribution of the sample mean $\bar{X}$.
   (e) Construct a $100(1-\alpha)\%$ confidence interval for θ based on $\hat{\theta}_{MM}$.
   (f) Is $\hat{\theta}_{MLE}$ unbiased?
   (g) Is $\hat{\theta}_{MLE}$ sufficient?
   (h) Construct a $100(1-\alpha)\%$ confidence interval for θ based on the $\hat{\theta}_{MLE}$.

(i) Consider the hypothesis test: $H_0 : \theta = 5$ v. $\theta > 5$ at 5% level of significance. Find the critical region of the test using the limiting distribution of $\bar{X}$. Call it method A.

(j) Construct another critical region based on the statistic $X_{\max}$. Call it Method B.

(k) For method A, find the power of the test when $\theta = 5.5$ for $n = 50$.

(l) For method B, find the power of the test when $\theta = 5.5$ for $n = 50$.

2. An estimator $\hat{\theta}_n$ is said to be squared-error consistent if $\lim_{n\to\infty} E\big[(\hat{\theta}_n - \theta)^2\big] = 0$.

   (a) Show that any squared-error consistent $\hat{\theta}_n$ is asymptotically unbiased. i.e. Show that $E(\hat{\theta}_n) \to \theta,\ n \to \infty$.

   (b) Show that the variance of any squared-error consistent $\hat{\theta}_n$ goes to 0. i.e. $Var(\hat{\theta}_n) \to 0,\ n \to \infty$.

   (c) Show that any squared-error consistent $\hat{\theta}_n$ is consistent. i.e. Show that for all $\epsilon > 0$, $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| < \epsilon) = 1$.

3. The case *United Black Firefighters Association v. City of Akron* concerned whether the promotional exam for the position of Fire Lieutenant had a disparate impact on African Americans. The promotional exam consisted of a written job knowledge test (70%) and an assessment center test (30%). The raw exam scores were transformed into scores on a scale of 0-100. Those who scored 70 or above passed the exam and were placed on an eligible list in rank order. The following table provides the means, the standard deviations, the numbers of applicants took the promotional exam (called "Sample Size" in the table) and the numbers of applicants who scored 70 or above, for both white and African American candidates.

	Mean	SD	Sample Size	# $\geq$ 70
Whites	81.33	7.15	99	89
African Americans	75.18	9.37	38	30

   (a) Do you think white candidates had significantly higher exam scores than the African American candidates? State the null and alternative hypotheses, find the test statistics and the critical value,

and then make conclusion in both statistical language and plan English. ($\alpha = 5\%$. To save calculation time, use the degrees of freedom 30.)

(b) Find the 95% confidence interval for the mean difference of the exam scores between white candidates and African American candidates. (Use 30 as the degrees of freedom)

(c) Find the 95% confidence interval for the ratio of the variances. (Use the nearest degrees of freedom available in the F table.)

(d) Do you think the exam scores for white candidates and the African American candidates have different variances? State the null and alternative hypotheses, and then *use the confidence interval you obtain from (c) to make conclusion.* ($\alpha = 5\%$.)

(e) Do you think white candidates had significantly different pass rates than the African American candidates? State the null and alternative hypotheses, find the p-value and then make conclusion in both statistical language and plan English. ($\alpha = 5\%$)

(f) Find the 95% confidence interval for the difference of the pass rates.

4. For the sake of this exam, suppose you are interested in the relationship between the scores of the written job knowledge test (variable name: WJKScore) and the scores of assessment center test (variable name: ACScore), and if the two variables are related, whether the relationship is the same for both races. The dummy variable RaceW = 1 if the candidate is white, RaceW=0 if the candidate is an African American. The following tables and graphs give the regression outputs, scatter plot and residual plots for three models:

Model #1: lm(formula= WJKScore $\sim$ ACScore)

Coefficients:

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	46.24183	5.57676	8.292	9.97e-14
ACScore	0.50347	0.07639	6.591	9.01e-10

Residual standard error: 8.178 on 135 degrees of freedom
Multiple R-squared: 0.2435, Adjusted R-squared: 0.2378
F-statistics: 43.44 on 1 and 135 DF, p-value: 9.014e-10

Model #2: lm(formula=WJKScore $\sim$ ACScore + RaceW)

Coefficients:

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	47.49756	5.58530	8.504	3.17e-14
ACScore	0.45805	0.08033	5.702	7.24e-08
RaceW	2.81476	1.64134	1.715	0.0887

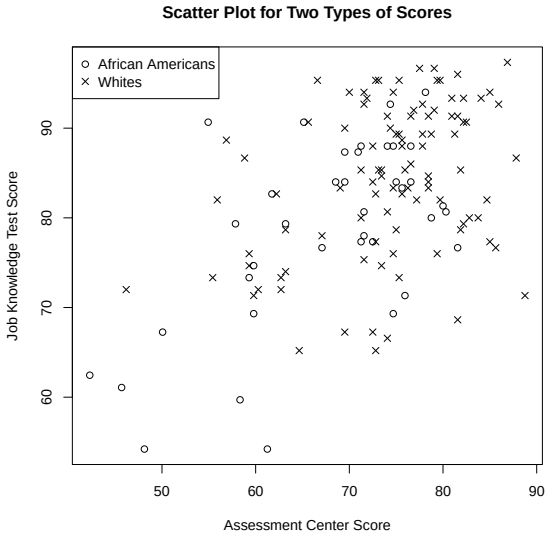
Residual standard error: 8.12 on 134 degrees of freedom
Multiple R-squared: 0.2597, Adjusted R-squared: 0.2486
F-statistic: 23.5 on 2 and 134 DF, p-value: 1.78e-09

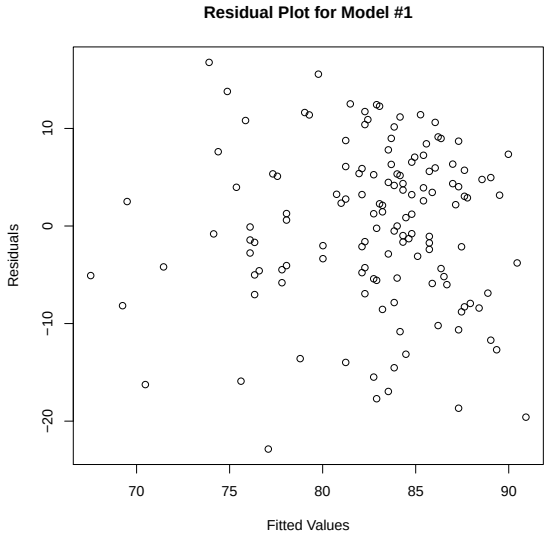
Model #3: lm(formula= WJKScore $\sim$ ACScore * RaceW)

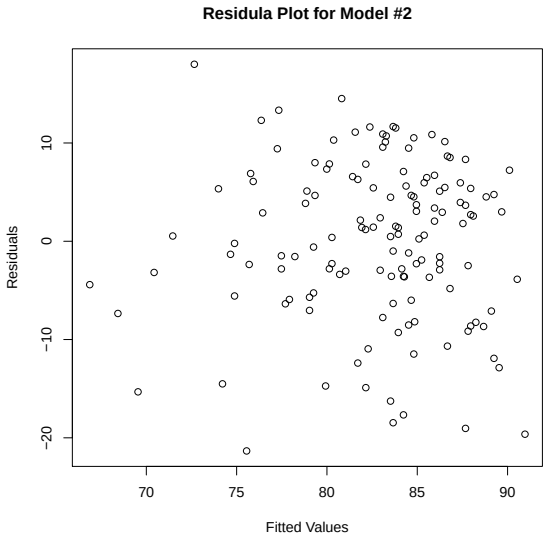
Coefficients:

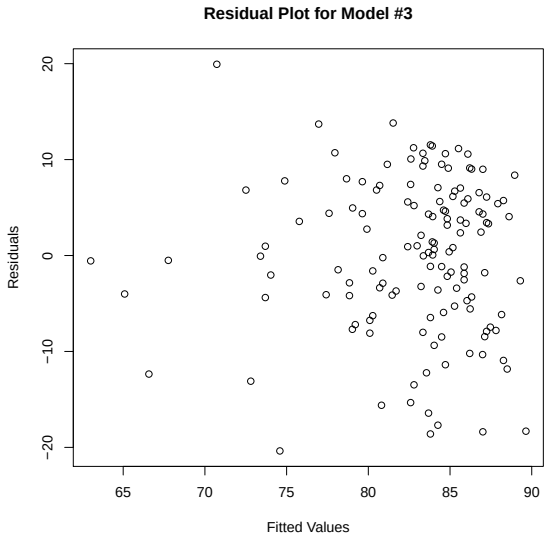
	Estimate	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	37.1856	8.9298	4.164	5.58e-05
ACScore	0.6107	0.1307	4.671	7.24e-06
RaceW	19.9105	11.6982	1.702	0.0911
SCScore:RaceW	-0.2439	0.1653	-1.476	0.1423

Residual standard error: 8.085 on 133 degrees of freedom
Multiple R-squared: 0.2716, Adjusted R-squared: 0.2552
F-statistics: 16.53 on 3 and 133 DF, p-value: 3.453e-09









(a) According to Model #1, what's the regression line between the written job knowledge test scores and assessment center test scores?

(b) According to Model #1, do you think the scores for written job knowledge test and assessment center test are related?

(c) According to Model #2, what's the regression line between the

written job knowledge test scores and assessment center test scores for white candidates? For African American candidates?

(d) According to model #2, do you think the regression lines are the same for white candidates and African American candidates? *Briefly* explain why.

(e) According to Model #3, what's the regression line between the written job knowledge test scores and assessment center test scores for white candidates? For African American candidates?

(f) Which model are you going to choose? *Briefly* explain why.

(g) The following is the ANOVA table for Model #1. Use the available information, find the values of the question marks "?" in the table.

Analysis of Variance Table

Response: WJKScore

	DF	Sum Sq	Mean Sq	F value	$Pr(>\|t\|)$
ACScore	?	?	?	?	?
Residuals	?	?	?		