Math 192 Exam 1 Solutions
Spring 2009 - Brad Hartlaub

1 a. The distribution of Pete's batting averages is skewed to the left (slightly) with a mean of about .300.

b. Pete Rose had 4256 hits and 14,053 at bats during his career. Thus, a 95% confidence interval for his true batting average is $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ or $\frac{4256}{14,053} \pm 1.96\sqrt{\frac{4256/14,053\,(1-4256/14,053)}{14,053}}$ (.302853)

MTB: Stat > Basic Stats > 1-Prop (.295256, .310450)

Thus, we are 95% confident that Pete's true hitting ability is between .295 and .310.

c. The 5 number summary for OBP is

Min = .316, Q_1 = .348, Median = .382, Q_3 = .398, Max = .430

d. Total # of hits = 4256 (MTB: Calc > Column Stats)

e. The quadratic model provides a much better fit. Only 15.8% of the variability in the # of doubles is explained using the linear model, but the quadratic model explains 68.7% of this variability. The estimated model is

$$\widehat{2B} = -871638 + 883.6\ \text{Season} - .2239\ \text{Season}^2$$

The residual plots show no obvious departs for normality and constant variance. The typical size of the error is s = 7.83794

Math 192 Exam 1 Solutions
Spring 2009 - Brad Hartlaub

1 f. Let's assume that the Reds played approximately 26 games in a typical month and Pete came to the plate 4 times in each of these games. Thus, $n = 26 \times 4 = 104$. In 1972, $\hat{p}_{hit} = \frac{198}{645} = .307$ A reasonable Binomial model for the # of hits in a typical month is $b(n=104, p=.307)$

g Pete Rose walked 106 times during the 1974 season. The baseball season is approximately 6 months long so use a Poisson model with mean $106/6 = 17.6667$

h. Let X = # of at bats to get on base for the 1st time

$X \sim Geo(.334)$

$P(X=3) = .148148 = (1-.334)^2(.334)$

i Let Y = # of at bats to get on base 3 times

$Y \sim Negbin(.334)$

$P(Y < 10) = P(Y \leq 9) = .624459$

Cumulative (MTB with Negbin(.334), r=3)

Math 192 Exam 1 Solutions
Spring 2009 - Brad Hartlaub

#2 a. As expected, the AL tends to have more hits (higher mean 1492.9 > 1441.9, median 1454.5 > 1391.8, and quartiles 1454.5 > 1391.8 and 1492 > 1442) than the NL because of the DH. The standard deviation is slightly higher for the AL, 71.6 > 67.9, but the IQR for the AL is slightly lower, 83.5 < 92. There is one outlier for the AL, the Oakland A's with 1318 hits.

b. The distribution of the number of doubles (2B) is clearly skewed to the right, so the median and IQR should be used to describe the center and spread.

c. Yes, the scatterplot shows a clear positive association and the value of the correlation coefficient is $r = .68$

d. Yes, there is one unusually large residual (98.1008) for the Minnesota twins, but otherwise the residuals plots show no problems with the assumptions of normality and constant variance. The regression equation is

$$\hat{R} = -369.4 + 2697\,SLG,$$

which explains 81.6% of the variability in runs scored. The size of the typical error is $s = 29.877$ runs.

Math 192 Exam 1 Solutions
Spring 2009 - Brad Hartlaub

#3 a. The standardized z scores are $z_{era} = \frac{ERA - 4.3167}{.4493}$
The team with the lowest z-score is Toronto (-1.83974).
Toronto's team ERA is 1.84 std deviations below the MLB average.
The team with the highest z-score is Texas (2.34418).
The team ERA for Texas is 2.34 std deviations above the MLB average

b. Calc > Calculator
$PCT = W/GP$
Other than Wins (r=1) and losses (r=-1), the variable with the strongest relationship to a team's winning percentage is Saves (SV), r = .749.

c. The least squares line is $\hat{PCT} = .9408 - .1021\ ERA$

d. A one unit change in a team's ERA would result on average in a .1021 decrease in winning percentage

e. The lowest residual is -.101545 (San Diego) and the highest residual is .09525 (Texas). However, these residuals are not unusual. The standardized residual for Texas is 2.10153

f. The fitted value for Baltimore is
$$\hat{Y} = .9408 - .1021(5.13) = .416910$$
The residual is $.422360 - .416910 = .005450$

g. Step 1: Create RPG
Calc > Calculator
$RPG = R/GP$
Step 2: Fit Models
Student results will vary. They must comment on the fit of the model (R^2 and residual plots).

Simple linear reg with ERA $r^2 = 97.8\%$