

**STATISTICS DEPARTMENT
M.S. EXAMINATION**

**PART II
OPEN BOOK**

Tuesday, May 29, 2001

9:00 a.m. - 1:00 p.m.

Statistics Department Computer Lab, SC S152

Instructions: Complete problems *one, two, and three* and choose between problems *four and five*. Each problem counts 25 points. Unless otherwise noted, points are allocated approximately equally to lettered parts of a problem. Spend your time accordingly.

The web site address for data and program files for this exam is:

<http://www.telecom.csu Hayward.edu/~esuess/MSexam/MSexam.htm>

Begin each problem on a new page. Write the problem number and the page number in the specified locations at the top of each page. Also write your chosen ID code number on every page. Please write only within the black borderlines, leaving at least 1" margins on both sides, top and bottom of each page. Write on one side of the page only.

At the end of this part of the exam you will turn in your answer sheets, but you will keep the question sheets and your scratch paper.

You may use a computer to work any of the problems, but your answers must be handwritten on standard paper provided for the examination. Printers may *not* be used during the exam, and pages printed out by computer may *not* be submitted. As indicated, some problems have data files available on disk.

1. A company wants to improve the lifetimes of a kind of delicate electronic device it manufactures. Engineers have proposed two new variations on the manufacturing process Method B and Method C, hoping that one of them will be better than the present Method A. In one investigation of the methods, 15 devices made by each method were subjected to stress tests, and times to failure (in days) were recorded for each of the 45 devices tested. The data are shown below, and are available in the file LIFETEST.TXT.

	A	B	C
	0.72	0.41	5.60
	2.63	1.54	1.87
	3.95	0.39	2.99
	0.02	1.25	0.54
	0.04	1.40	0.53
	0.14	0.74	1.42
	1.86	1.21	1.64
	0.07	0.16	2.77
	0.25	1.36	0.36
	0.99	1.33	1.95
	1.38	0.44	1.54
	0.44	0.04	0.95
	0.67	0.97	0.62
	2.12	0.37	1.39
	1.80	1.22	2.42
Mean	1.139	0.855	1.773

- (a) For now, suppose that the usual assumptions for an ANOVA are satisfied. What is the name of the appropriate ANOVA design for this experiment? State the assumptions explicitly.
- (b) Perform the ANOVA of part (a). Show the ANOVA table. Do the methods differ significantly at the 5% level? 10% level?
- (c) It would be considered a major improvement if time to failure in the stress test could be increased by 1 day. How many observations on each of 3 groups would be required (instead of $n = 15$ here) to achieve a power of 80% for the ANOVA F-test. Use the 5% level of significance and assume that the population standard deviations are $\sigma = 1.1$ days.
- (d) Are the assumptions of the ANOVA model met? Give the name(s) and P-value(s) of the test(s) you use.
- (e) Past experience with similar test data has shown that it is sometimes appropriate to do a log or rank transform of the data. (That is either to replace each observed time with its logarithm, say base 10; or to replace each of the 45 observations with its rank, a number from 1 through 45.) Try both transformations and comment whether either is beneficial.
- (f) Recall that the purpose of this experiment is to determine whether either of the two new methods (B or C) gives significantly longer times to failure than current Method A. Based on your results in (a) - (e), or the results of any other statistical procedure(s) you deem appropriate, what is your answer to that question? State clearly what data and procedures lead to your conclusion.
- (g) Suppose that the company has four sites where an experiment such as the one described here can be performed (3 methods, 15 observations for each method). Suppose that the same three methods are tested at each site (giving a total of $4 \times 3 \times 15 = 180$ observations). Write the ANOVA model for this larger experiment.

2. Professor Leo Kahane, Department of Economics, collected the wage data partially shown below. Use the complete data set to answer questions a-g.

Wage	Education	Female	Married	Experience	Black/Hispanic
11	8	male	married	42	other
20.4	17	male	unmarried	3	other
9.1	13	male	unmarried	16	black/hispanic
13.75	14	male	married	21	other
24.98	16	male	married	18	other
7.7	13	female	unmarried	8	other
10	10	female	married	25	other
...
6.25	9	male	unmarried	30	other

Wage is 1991 adjusted thousands of dollars.

Education is years of schooling completed.

Experience is years of labor market experience.

- Write a SAS program to successfully read in the data from the text file named wages. Even if your program fails, paste the data into a data step and CONTINUE!
- Fit a model predicting wages based solely on years of education.
- Test whether education is a significant predictor of wages at the 5% level.
- Find a 95% confidence interval for the rate at which wages increase due to an increase of 1 year of education.
- Include years experience in the prediction model. Are education and years experience both significant predictors of wages?
- Create dummy variables for the other variables and find the "best" prediction model for these data that predicts wages. Cite your criterion for "best".
- Based on your model, what can you conclude about the importance of each of the variables in determining wages.

3. Consider a Markov chain X_0, X_1, X_2, \dots with state space $S = \{1, 2, 3, 4, 5, 6\}$, initial probability vector $\pi_0 = (1, 0, 0, 0, 0, 0)$, and transition probability matrix

$$P = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 2/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 3/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 4/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}.$$

- (a) Evaluate the probabilities $P\{X_2 = 1\}$ and $P\{X_{n+2} = 3 \mid X_n = 3\}$.

[Save time: Do only the *necessary* arithmetic.]

- (b) What does it mean for a chain to be irreducible? Is this chain irreducible?

What does it mean for a state to be persistent (recurrent)? Transient?

For each state in this chain, say whether it is persistent (recurrent) or transient.

What does it mean for a chain to be ergodic? Is this chain ergodic?

- (c) Carefully explain how to verify, using hand computations and a theorem about Markov chains, that the vector $\lambda = (1/30, 1/20, 1/12, 1/6, 1/2, 1/6)$ gives the *limiting* distribution of this chain.

Explain briefly how you could approximate λ by finding higher and higher powers of P on a computer.

- (d) Use the information in part (c) to evaluate the following limits (as $n \rightarrow \infty$):

$\lim P\{X_n = 5 \mid X_0 = 3\}$ and $\lim P\{X_n = 3\}$.

Also, evaluate μ_1 , the mean number of stages after the chain leaves state 1 until it next returns to state 1.

4. An experiment was done to measure the effects of ozone, a component of smog. A group of twenty-two 70-day-old rats was kept in an environmental containing ozone for 7 days, and their weight gains (in grams) were recorded. Another group of twenty-three rats of similar age was kept in an ozone-free environment for a similar time, and their weight gains were recorded. Using the Splus program available at

<http://www.telecom.csuhayward.edu/~esuess/MEsp01ob.ssc>

analyze the data to determine the effect of ozone.

- (a) Describe the shapes of the histograms for the ozone and control groups.
- (b) What are the sample means for each group? What are the sample standard deviations?
- (c) Perform an independent two-sample t-test to see if there is a difference in the weight gains between the ozone group and control group. Before performing the test, check the assumptions of equal variance and normality.
- (d) What is the 95% independent two-sample t-confidence interval for $\mu_{oz} - \mu_{con}$?
- (e) Use the nonparametric bootstrap to estimate the difference in the means, to calculate a standard error estimate of the difference in the two means, and to calculate a 95% empirical bootstrap confidence interval.
- (f) Comment on the difference you observe between the 95% confidence intervals produced in parts (d) and (e).
- (g) Does ozone influence the rats' weight gains? If so, give an estimate of how much.

5. Table A: gives the birth weights (in pounds) of Poland China pigs in eight litters. The data come from *Statistical Methods* by George Snedecor.

- Construct the analysis of variance table. Test at the 0.10 level the hypothesis of no difference between mean weights in the eight litters. (See file pig.txt.)
- Suppose litters 1,3,4 were sired by one boar and the other five litters were sired by another boar. Is there significant difference between the mean weights in those two groups? Assume this is a planned comparison.
- Is there a difference between the mean weight for large litters (numbers 1,2,3,4) and small litters (numbers 5,6,7,8)? Assume this is a comparison suggested by the data.

Table A

1	2	3	4	5	6	7	8
2.0	3.5	3.3	3.2	2.6	3.1	2.6	2.5
2.8	2.8	3.6	3.3	2.6	2.9	2.2	2.4
3.3	3.2	2.6	3.2	2.9	3.1	2.2	3.0
3.2	3.5	3.1	2.9	2.0	2.5	2.5	1.5
4.4	2.3	3.2	3.3	2.0		1.2	
3.6	2.4	3.3	2.5	2.1		1.2	
1.9	2.0	2.9	2.6				
3.3	1.6	3.4	2.8				
2.8		3.2					
1.1		3.2					

Overall Grading: Total points will probably not be divisible by the number of parts, which is 7. To avoid fractional points, 1 more point each may be given to some earlier parts than to some later parts.

- (a) This is a one-way ANOVA (completely randomized design) with three levels of the fixed factor. Assumptions are that the data are normal and that the variances of the three groups are equal. Although it is not requested here, the formal model could be written as

$$Y_{ij} = \mu + \alpha_i + e_{ij}; \quad i = 1, 2, 3 \text{ (methods)}, j = 1, 2, \dots, 15 \text{ (replications)};$$

where $\sum \alpha_i = 0$ and the e_{ij} are independent, identically distributed as $N(0, \sigma^2)$.

Grading: No points for anything but the one-way model. Substantial penalty for not mentioning independence, normality, and equal variances. Some penalty for saying it's a random effects model, but none for omitting to say the one effect is fixed. (If wrong model, try to give partial credit below for consistent adherence to whatever model is incorrectly stated.)

- (b) In Minitab, we stacked the data into a column 'Life' with subscripts in 'Method'. The ANOVA tables follows, along with the Fisher LSD comparisons, which are useful in later parts. We also stored residuals for later use. Results are significant at the 10% level, but not the 5% level.

Grading: Heavy penalties for incorrect DF, MS, or F. Heavy penalty if interpretation is not consistent with P-value

Analysis of Variance for Life

Source	DF	SS	MS	F	P
Method	2	6.62	3.31	2.95	0.063
Error	42	47.05	1.12		
Total	44	53.67			

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	
1	15	1.139	1.146	(-----+-----+-----+-----+)
2	15	0.855	0.511	(-----+-----+-----+)
3	15	1.773	1.336	(-----+-----+-----+)
Pooled StDev =		1.058		0.60 1.20 1.80 2.40

Fisher's pairwise comparisons

Family error rate = 0.120
Individual error rate = 0.0500

Critical value = 2.018

Intervals for (column level mean) - (row level mean)

	1	2
2	-0.497 1.063	
3	-1.414 0.146	-1.697 -0.137

- (c) Minitab's power/ sample size procedure shows that we would need 25 observations per method, to have 80% power.

Grading: Heavy penalties for using incorrect α , σ or power. Because comparison of A with B and of A with C is the crux, it would be OK to use the sample size formula, with correct parameters, for two-sample t-tests.

One-way ANOVA

Sigma = 1.1 Alpha = 0.05 Number of Levels = 3

SS Means	Sample Size	Target Power	Actual Power	Maximum Difference
0.5	25	0.8000	0.8116	1

(d) A normal probability plot of the residuals shows a distinct curvature characteristic of nonnormal data. The P-value of the Anderson-Darling test is 1.1%. Bartlett's test for homoscedasticity performed on the original data (by Method) gives P-value 0.4%. Neither assumption seems to be met.

Grading: Roughly half credit each for some valid test of normality and some valid test of equal variances. Heavy penalty for doing Anderson-Darling, or other normality test, on the original data instead of the residuals. Partial credit on normality for rough description "curvature" in normal probability plot. (Bartlett would usually be done on the data, but for this model results should be the same if done on the residuals; no penalty.) There is really no way to check whether groups are independent. Control charts, runs tests, etc. might be used to check for independence within each group. But this is not expected; bonus points against errors in this part if discussed or correctly performed.

(e) With log-transformed data, the P-value is 5.6% for the F-test in the ANOVA table; 2% for Anderson-Darling on residuals, and 1.7% for Bartlett on the log-transformed data (by Method). With rank-transformed data, the P-value is 4.7% for the F-test, 11.3% for Anderson-Darling on residuals, and 19.5% for Bartlett on ranks (by Method). The rank-transformed data meet the ANOVA assumptions better and give a slightly smaller P-value for the F-test.

Grading: Somewhat more than 1/2 credit for doing *either one* of the transformations correctly all the way through.

(f) Fisher's LSD comparisons seem appropriate for comparing B with A and C with A. (Other comparison methods OK.) Only in the case of the log-transformed data does LSD find a significant difference between A and one of the new methods (specifically, between A and C). *There is evidence to suggest, but by no means to confirm, that Method C may be superior to Method A.* As seen in part

(c) we do not really have enough data here reliably to detect differences of a size deemed important. The words "delicate" and "stressed" in the description of the problem were intended to give a clue that these devices may be failing because of "accidents" rather than by "wearing out," and hence may have exponentially distributed life times. If so, the log transformation may be defensible on theoretical grounds. (Bear in mind, however, that differences in means of logged lifetimes correspond to ratios different from 1 of geometric means of actual lifetimes.)

Grading: On purpose, sort of an open-ended and vague question. Credit strongly based on reasonable conclusions based on specific work done above. Penalty for making strong claims based on P-values around 5%, or making a strong distinction between 5% and 6% levels.

(g) This is a standard two-factor ANOVA (both factors fixed, crossed) with three levels of Method and four of Site. Interaction is supported. Model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}; \quad i = 1, 2, 3 \text{ (methods)}; \quad j = 1, 2, 3, 4 \text{ (sites)}; \quad k = 1, \dots, 15 \text{ (replications)}.$$

Restrictions: $\sum_i \alpha_i = \sum_j \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$. Distribution: e_{ijk} i.i.d. $N(0, \sigma^2)$

Grading: No points if two factors not mentioned. Heavy penalty for not giving the model, since it was asked for. Heavy penalty for turning replications into a third effect, or for inventing bogus nesting. Some penalty if ranges of subscripts not given, interaction omitted, any factors declared random, restrictions omitted, or distributional information missing.

Source: Data were randomly generated as follows: All three methods exponential, A and B with mean 1 and C with mean 2. Modeled after a real situation with proprietary data.

Open book # 2

Solution to the regression problem:

- a. SAS program. Many versions will work.

```
data new;
  infile 'i:\courswork\stat\pctumbo\exam questions spring 2001\reg.txt'
  delimiter='09'x MISSOVER;
  input wage educ gender $ marital $ exper blackhis $;
  proc print;run;
```

- b. The SAS System 21:55 Tuesday, April 24, 2001 4

The REG Procedure
Model: MODEL1
Dependent Variable: wage

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	368.23327	368.23327	15.19	0.0003
Error	43	1042.28064	24.23908		
Corrected Total	44	1410.51391			

Root MSE	4.92332	R-Square	0.2611
Dependent Mean	8.65444	Adj R-Sq	0.2439
Coef Var	56.88777		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-5.12727	3.61126	-1.42	0.1629
educ	1	1.08423	0.27817	3.90	0.0003

Notice that the F ratio is highly significant. The equation is
 $wage = -5.127 + 1.08423 \cdot education$.

- c. Since the above model is significant, education must be a significant predictor of wage.
 $P = 0.0003$

- d. Using the estimate of the slope and its standard error from the print out, at 95% ci for the coefficient of education in the equation is $1.08 \pm 1.96 \cdot 0.278$ or $1.08 \pm .54$.

The SAS System 21:55 Tuesday, April 24, 2001 5

The REG Procedure
Model: MODEL1
Dependent Variable: wage

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	432.04007	216.02004	9.27	0.0005

Error	42	978.47384	23.29700
Corrected Total	44	1410.51391	

Root MSE	4.82670	R-Square	0.3063
Dependent mean	8.65444	Adj R-Sq	0.2733
Coeff var	55.77130		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-9.40101	4.38214	-2.15	0.0378
educ	1	1.26301	0.29333	4.31	<.0001
exper	1	0.11009	0.06652	1.65	0.1054

Although education remains a significant predictor, experience is not significant since $p = .1054$ is $> .05$.

f.

The REG Procedure
Model: MODEL1
Dependent variable: wage

C(p) Selection Method

Number in Model	C(p)	R-Square	Variables in Model
4	4.2049	0.4025	educ female blkhis exper
3	4.4751	0.3679	educ female blkhis
2	4.9743	0.3298	educ blkhis
3	5.1474	0.3577	educ female exper
3	5.3201	0.3550	educ blkhis exper
5	6.0000	0.4056	educ female married blkhis exper
4	6.1234	0.3733	educ female married blkhis
2	6.5170	0.3063	educ exper
4	6.6818	0.3647	educ female married exper
3	6.8570	0.3316	educ married blkhis
2	6.9094	0.3003	educ female
4	7.2806	0.3556	educ married blkhis exper
1	7.4852	0.2611	educ
3	8.0581	0.3133	educ female married
3	8.3332	0.3091	educ married exper
2	9.0350	0.2679	educ married
2	16.2444	0.1580	female blkhis
3	17.7226	0.1660	female married blkhis
3	18.2442	0.1581	female blkhis exper
1	18.3738	0.0951	blkhis
4	19.7179	0.1661	female married blkhis exper
2	20.2079	0.0976	married blkhis
2	20.2496	0.0970	blkhis exper
1	20.2583	0.0664	female
2	21.0168	0.0853	female married
3	22.0528	0.1000	married blkhis exper
2	22.0849	0.0690	female exper
3	22.9234	0.0867	female married exper
1	23.9768	0.0097	married

1	24.6088	0.0001	exper
2	25.9764	0.0097	married exper

Using the CP criterion several models could be useful, but the first in this list is a model containing education, a black hispanic indicator, a female indicator, and work experience.

The SAS System 21:55 Tuesday, April 24, 2001 9

The REG Procedure
Model: MODEL1
Dependent variable: wage

Analysis of variance

Source	DF	Sum of Squares	Mean Square	F value	Pr > F
Model	4	567.72968	141.93242	6.74	0.0003
Error	40	842.78423	21.06961		
Corrected Total	44	1410.51391			

Root MSE	4.59016	R-Square	0.4025
Dependent Mean	8.65444	Adj R-Sq	0.3427
Coeff Var	53.03823		

Parameter Estimates

variable	DF	Parameter Estimate	Standard Error	t-value	Pr > t
Intercept	1	-5.79887	4.41281	-1.31	0.1963
educ	1	1.14697	0.28353	4.05	0.0002
female	1	-2.48446	1.39358	-1.78	0.0822
blkhis	1	-3.05441	1.76283	-1.73	0.0909
exper	1	0.09957	0.06543	1.52	0.1359

While education is still the only significant factor in this model, the CP statistics suggests that this is the best model to use.

- g. If we rely on the model above rather than traditional significance, we note that being female reduces ones wages by 2.48 thousand dollars, and being black or Hispanic reduces you wages by just over \$3000, while work experience increases salary about \$90 for each year worked. But education increases salary over \$1000 for each year of school completed up to grade 12 which is all that is included in this data.

We have NOT looked at assessment factors and we have not tried to see how well the model does on a hold out sample.

Openbook # 3

Answers:

(a) $P\{X_2 = 1\} = P\{X_2 = 1 \mid X_0 = 1\} = (1/6)^2 + (1/6)^2$. More formally, element 1 of $\pi_0 P^2$.
 $P\{X_{n+2} = 3 \mid X_n = 3\} = 10/36 = 5/6$. This is element (3, 3) of P^2 .

(b) A chain is irreducible if it has only one class of intercommunicating states; that is, in some number of stages each state can lead to any other states. This chain is irreducible because the path $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 1$ is possible. Another way to look at it is that all elements in P^2 are positive, so any state can be reached from any other state in 2 stages.

A state is persistent if it is visited infinitely often over the long run. It is transient if it will surely be visited for a last time. In a finite irreducible chain all states are persistent.

A chain is ergodic if it has a limiting distribution. A finite irreducible chain is ergodic if it is aperiodic of period d . (A state is periodic if the X -process can return to that state only at multiples of $d > 1$ stages. Here the process can return to any state in either 2 or 3 stages; there can be no periodicity because the greatest common divisor of 2 and 3 is 1.

(c) It is easy to verify that $\lambda P = \lambda$. Thus λ gives the steady-state (stationary) distribution of the chain. For an ergodic chain, the steady-state distribution is also the limiting distribution.

Each row in higher powers of P approaches λ . The approach is geometrically fast and errors decrease monotonically. Stop when you have the required number of decimal places of accuracy. In this case, entries in P^{32} are accurate to 5 or more decimal places.

(d) Hence, $\lim P\{X_n = 5 \mid X_0 = 3\} = \lambda_5 = 1/2$ and $\lim P\{X_n = 3\} = \lambda_3 = 1/12$.

Also, the average return time for a state is the reciprocal of the limiting probability that the process will be in that state: $\mu_1 = 1/\lambda_1 = 30$.

Note: Here is an algorithm for simulating this chain with a fair die:

For stage 0; because of the initial vector, we begin in state 1.

For stage 1 roll the die; go to the state indicated on the die.

For subsequent stages roll the die:

If now in any state other than 6, go to the state corresponding to the maximum number seen on the die since the last visit to 6.

If now in state 6, go to the state indicated on the die (which becomes the new "maximum").

oxygen levels ~~not~~ μ

(H) a) Ozone: non normal
Control: skewed right, outlier left.

b) Ozone: $\bar{X} = 11.01$, $s = 19.02$
Control: $\bar{X} = 22.43$, $s = 10.78$

c) equal variance? No
normality? No

$H_0: \mu_{O_2} = \mu_{con}$ $t = -2.492$ $p\text{-value} = .0166$
 $H_A: \neq$

Reject H_0 , There is
a difference

d) 95% CI $(-20.657, -2.177)$

e) $\bar{\theta}^* = -11.4$, $s_{\theta} = 4.563$, 95% Bootstrap CI
 $(-20.16, -2.228)$

f) The Bootstrap CI is slightly more
narrow.

g) Yes, There is an approximate decrease
of -11.4 grams.

```

# ozone data from rats

rat.oz <- c(10.1, 6.1, 20.4, 7.3, 14.3, 15.5, -9.9, 6.8, 28.2, 17.9, -12.9, 14.0, 6.6, 12.1,
15.7,
39.9, -15.9, 54.6, -14.7, 44.1, -9.0, -9.0)

rat.con <- c(41.0, 38.4, 24.9, 25.9, 21.9, 18.3, 13.1, 27.3, 28.5, -16.9, 17.4, 21.8, 15.4,
27.4,
19.2, 22.4, 17.7, 26.0, 29.4, 21.4, 22.7, 26.0, 26.6)

# histograms

hist(rat.oz, xlim=c(-20,60))
hist(rat.con, xlim=c(-20,60))

# means

mean(rat.oz)
mean(rat.con)

mean(rat.oz) - mean(rat.con)

# standard deviations

sqrt(var(rat.oz))
sqrt(var(rat.con))

# check normality

qqnorm(rat.oz)
qqnorm(rat.con)

# independent two sample test and confidence interval

t.test(rat.oz, rat.con)

# nonparametric bootstrap confidence interval

B <- 10000      # number of bootstrap samples

rat.oz.n <- length(rat.oz)  # rat.oz sample size
rat.con.n <- length(rat.con) # rat.con sample size

mean.diff.boot <- numeric(B) # vector for bootstrap differences in the means

for(i in 1:B){
  oz.samp <- sample(rat.oz, rep=T, size=rat.oz.n) # bootstrap sample from rat.oz
  con.samp <- sample(rat.con, rep=T, size=rat.con.n) # bootstrap sample from rat.con
  mean.diff.boot[i] <- mean(oz.samp) - mean(con.samp) # calculation of the bootstrap
  differences in the means
}

mean(mean.diff.boot)      # bootstrap estimate of the difference in the means

sqrt(var(mean.diff.boot)) # bootstrap standard error of the difference in the means

hist(mean.diff.boot)      # bootstrap distribution of the difference in the means

quantile(mean.diff.boot, c(0.025, 0.975)) # calculation of the empirical bootstrap
confidence interval

```

```

> # ozone data from rats
rat.oz <- c(10.1, 6.1, 20.4, 7.3, 14.3, 15.5, -9.9, 6.8, 28.2, 17.9, -12.9, 14, 6.6, 12.1,
  15.7, 39.9, -15.9, 54.6, -14.7, 44.1, -9, -9)
> rat.con <- c(41, 38.4, 24.9, 25.9, 21.9, 18.3, 13.1, 27.3, 28.5, -16.9, 17.4, 21.8, 15.4,
  27.4, 19.2, 22.4, 17.7, 26, 29.4, 21.4, 22.7, 26, 26.6) # histograms
> hist(rat.oz, xlim = c(-20, 60))
> hist(rat.con, xlim = c(-20, 60)) # means
> mean(rat.oz)
[1] 11.01
> mean(rat.con)
[1] 22.43
> mean(rat.oz) - mean(rat.con) # standard deviations
[1] -11.42
> sqrt(var(rat.oz))
[1] 19.02
> sqrt(var(rat.con)) # check normality
[1] 10.78
> qqnorm(rat.oz)
> qqnorm(rat.con) # independent two sample test and confidence interval
> t.test(rat.oz, rat.con) # nonparametric bootstrap confidence interval

```

Standard Two-Sample t-Test

```

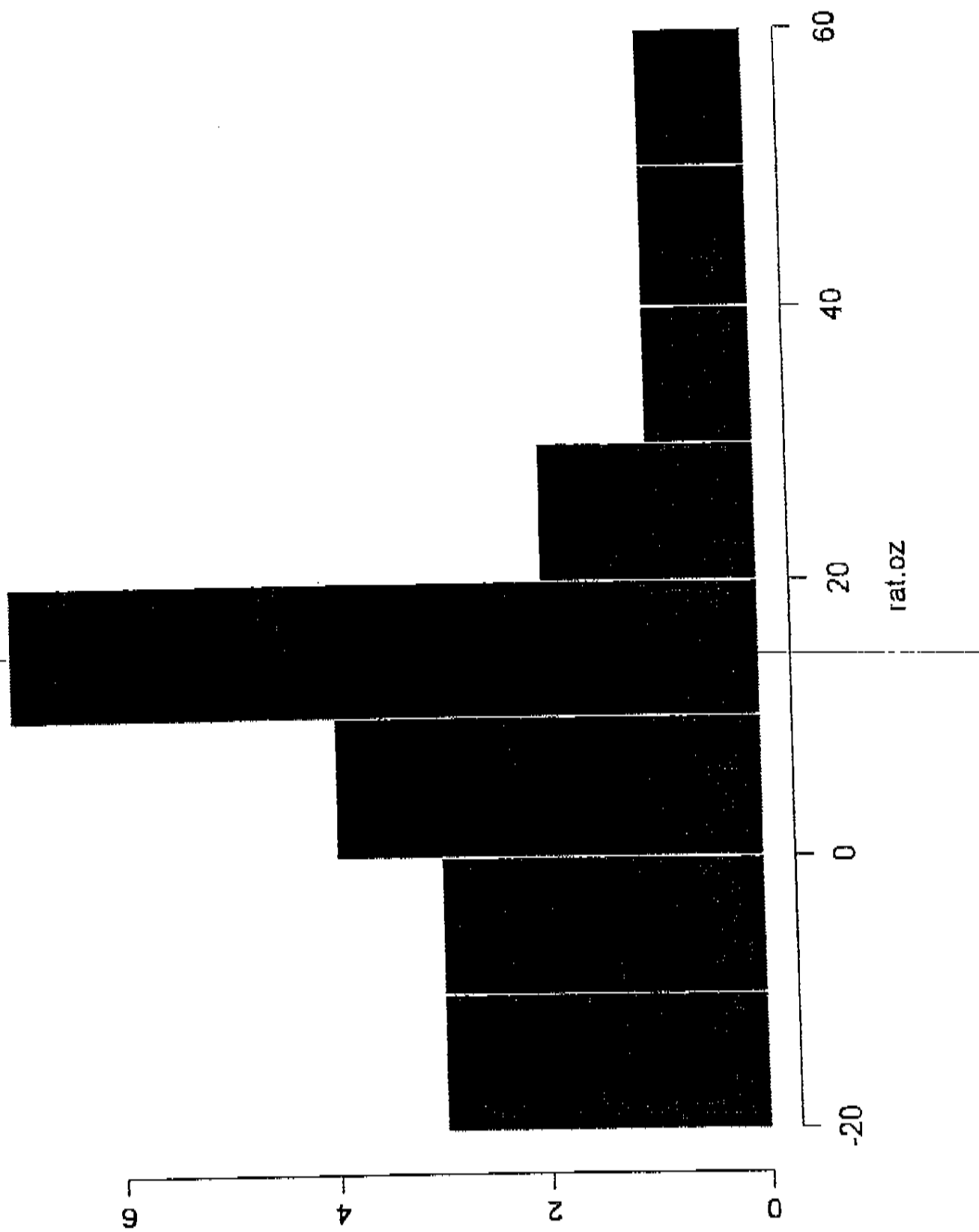
data: rat.oz and rat.con
t = -2.492, df = 43, p-value = 0.0166
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -20.657 -2.177
sample estimates:
mean of x mean of y
 11.01    22.43

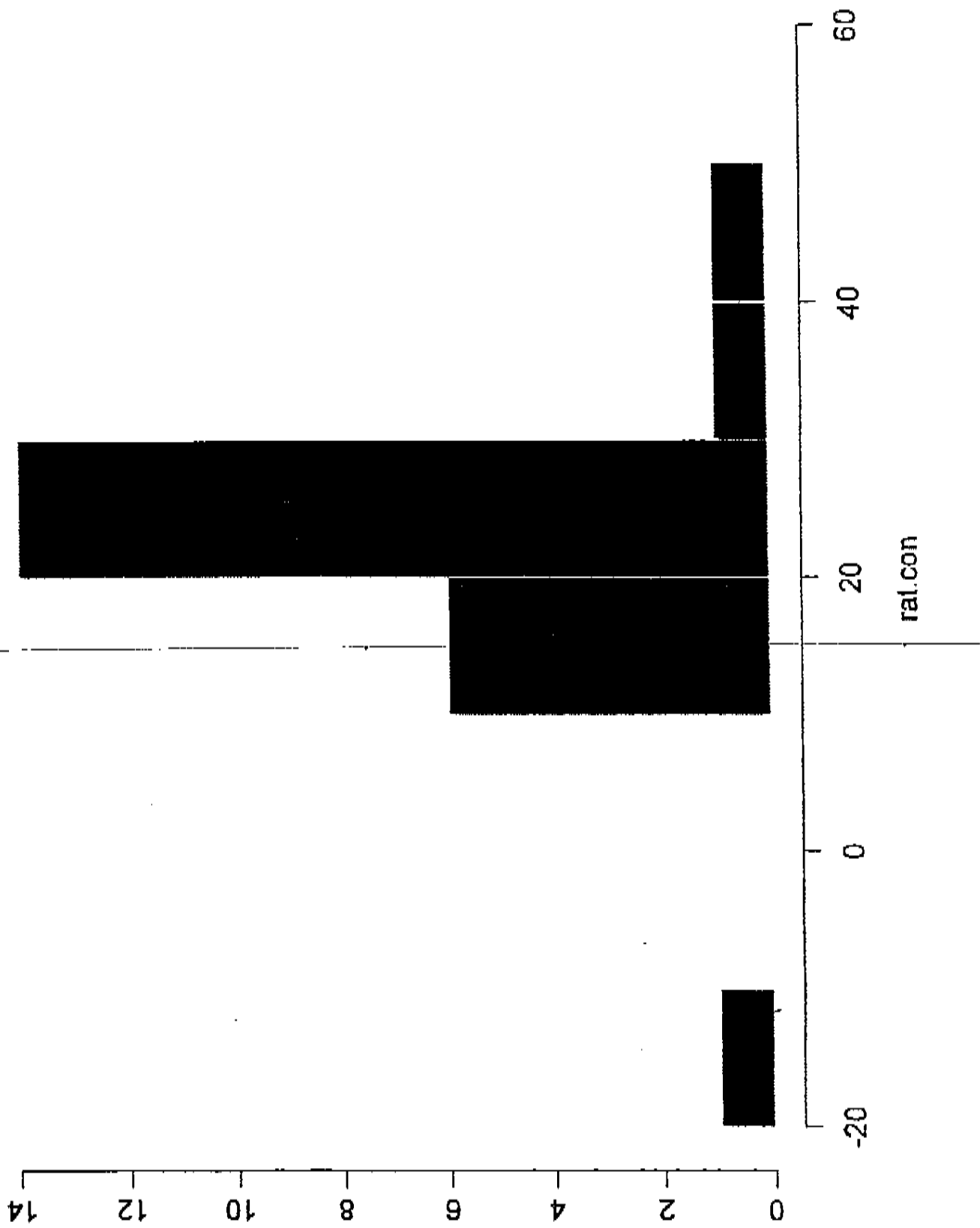
```

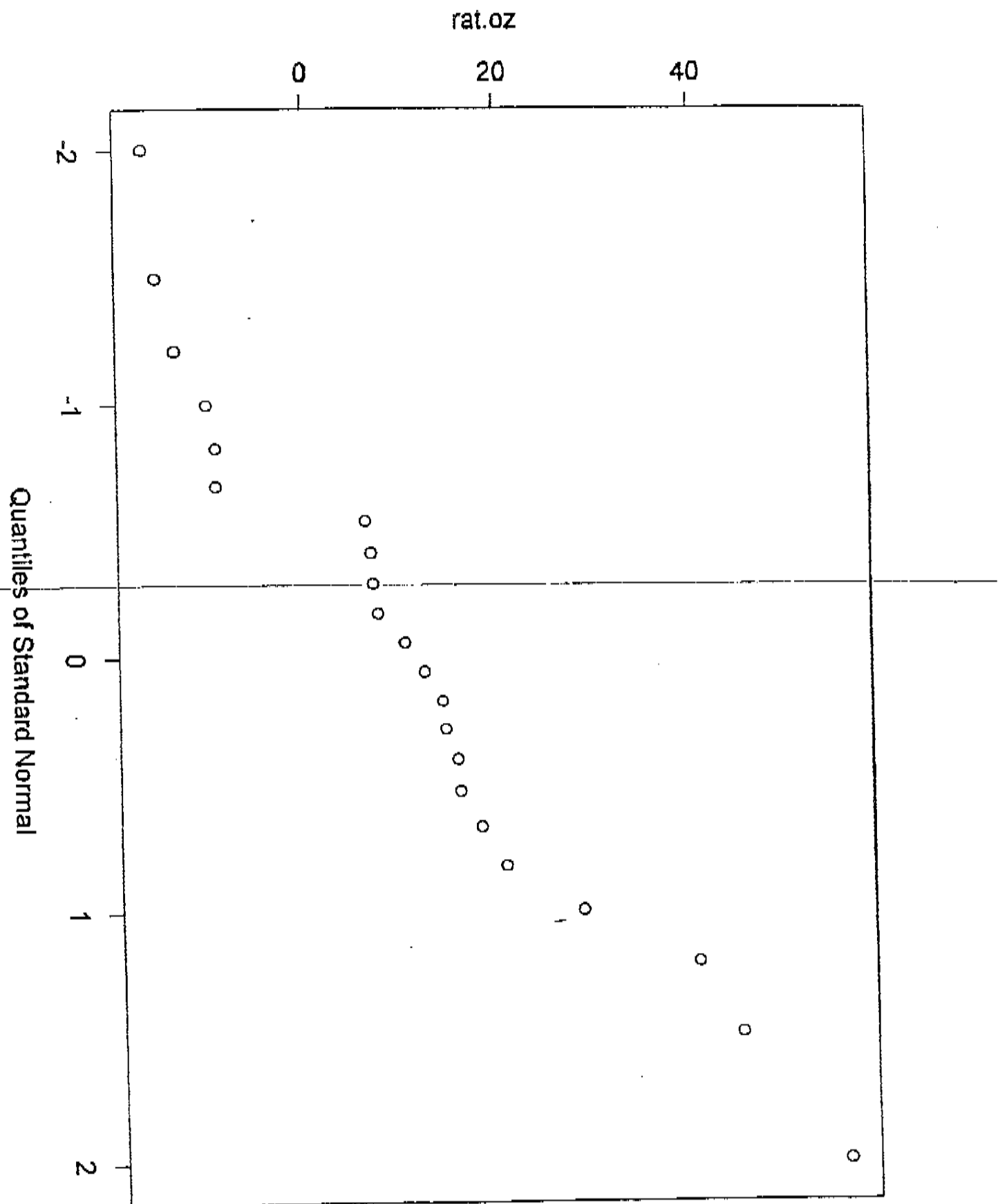
```

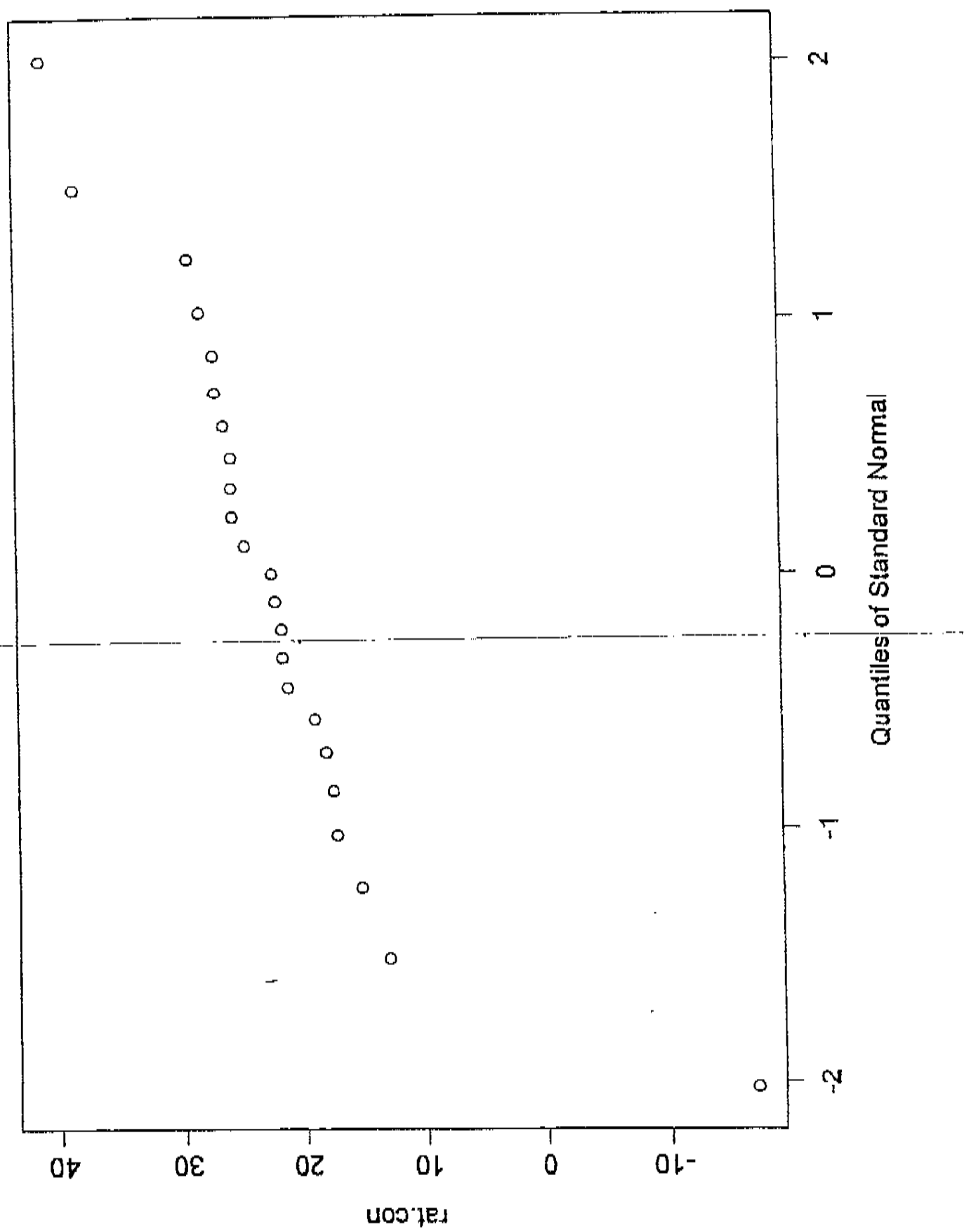
> B <- 10000 # number of bootstrap samples
> rat.oz.n <- length(rat.oz) # rat.oz sample size
> rat.con.n <- length(rat.con) # rat.con sample size
> mean.diff.boot <- numeric(B) # vector for bootstrap differences in the means
> for(i in 1:B) {
  oz.samp <- sample(rat.oz, rep = T, size = rat.oz.n) # bootstrap sample from rat.oz
  con.samp <- sample(rat.con, rep = T, size = rat.con.n)
  # bootstrap sample from rat.con
  mean.diff.boot[i] <- mean(oz.samp) - mean(con.samp)
  # calculation of the bootstrap differences in the means
}
> mean(mean.diff.boot) # bootstrap estimate of the difference in the means
[1] -11.4
> sqrt(var(mean.diff.boot)) # bootstrap standard error of the difference in the means
[1] 4.563
> hist(mean.diff.boot) # bootstrap distribution of the difference in the means
> quantile(mean.diff.boot, c(0.025, 0.975))
# calculation of the empirical bootstrap confidence interval
 2.5% 97.5%
-20.16 -2.228

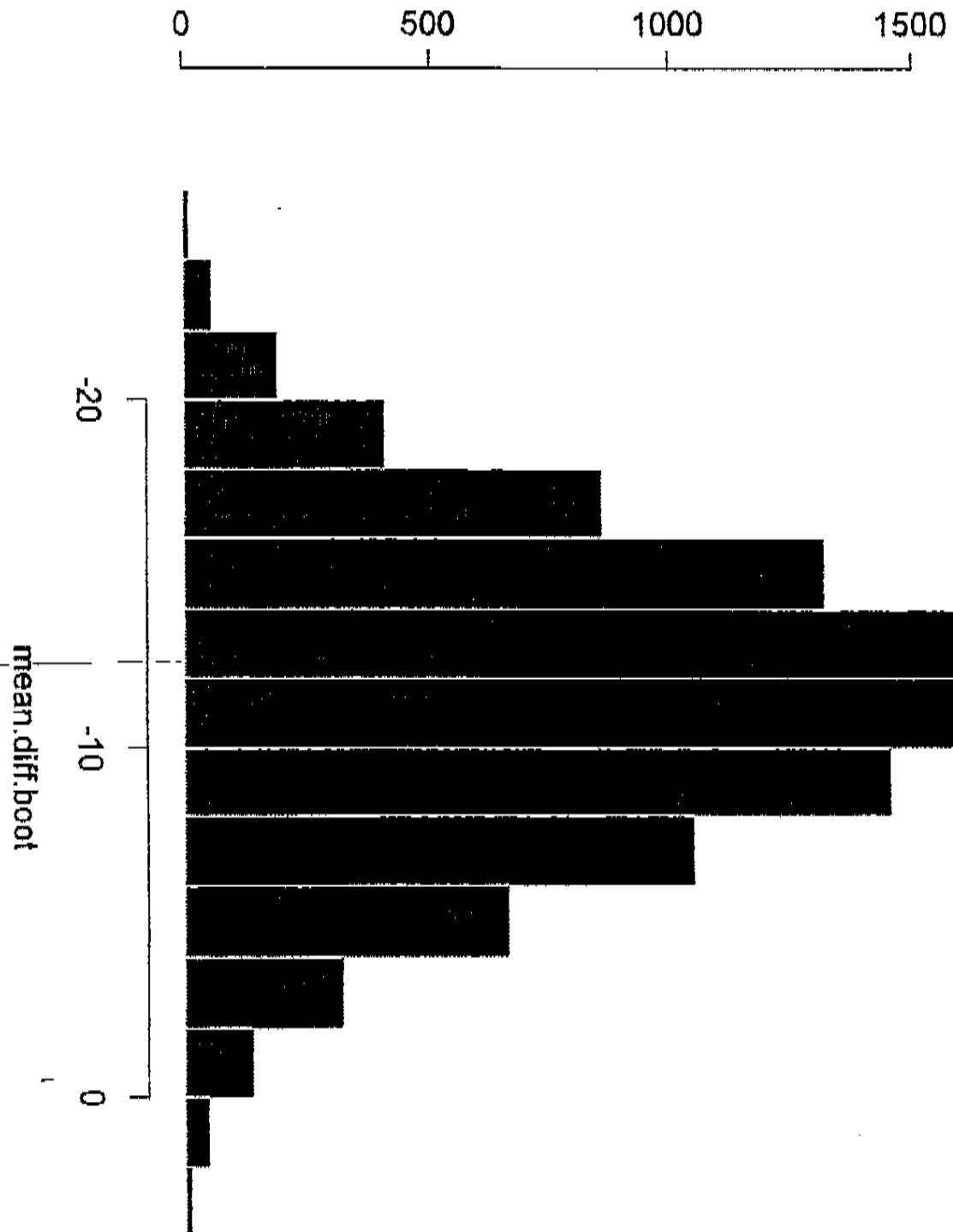
```











Openbook #5

Solution to open book version:

a.

Test of $\mu = 0$ vs $\mu \neq 0$

Variable	N	Mean	StDev	SE Mean
C1	10	0.660	1.858	0.587

Variable	95.0% CI	T	P
C1	(-0.669, 1.989)	1.12	0.290

One-way ANOVA: C1, C2, C3, C4, C5, C6, C7, C8

Analysis of Variance

Source	DF	SS	MS	F	P
Factor	7	7.489	1.070	2.99	0.011
Error	48	17.165	0.358		
Total	55	24.654			

Individual 95% CIs For Mean
Based on Pooled StDev

Level	N	Mean	StDev
C1	10	2.8400	0.9536
C2	8	2.6625	0.7050
C3	10	3.1800	0.2741
C4	8	2.9750	0.3196
C5	6	2.3667	0.3830
C6	4	2.9000	0.2828
C7	6	1.9833	0.6274
C8	4	2.3500	0.6245

Pooled StDev = 0.5980

1.80 2.40 3.00 3.60

According to this MINITAB output, There is a significant difference at 5% and at 10% between the means of the 8 litters. That is, we reject the hypothesis that all eight litter means are equal.

- b. This is a planned comparison so we can use one of our degrees of freedom to test whether the means of the two sires are equal. If the sample sizes were the same in each group, we would form the comparison as $(\mu_1 + \mu_3 + \mu_4)/3 - (\mu_2 + \mu_5 + \mu_7 + \mu_8)/4$ and test this comparison as to whether it is zero. However, the sample sizes are different and we have to compute a weighted mean comparison. The standard error of the comparison $\Sigma \lambda_i$ is $\sqrt{(s^2 \Sigma (\lambda_i^2 / n_i))}$. So using MSE as s^2 , we have $\sqrt{(0.358(1/9/10 + 1/16/8 + 1/9/10 + 1/9/8 + 1/16/6 + 1/16/4 + 1/16/6 + 1/16/4))} = 1.86$ as the standard error for the comparison $1/3(8.995) - 1/4(13.56) = -0.3924$. so the t-statistics with error df is $-0.3924/1.86$ or $t = -2.109$ with 48 degrees of freedom. This is significant at 5%.
- c. Since this comparison is a post hoc test, we need to use Scheffe's test. We compare the comparison divided by its standard error with $\sqrt{(8-1)F_{7,48}}$ or 3.93. The linear comparison is: $1/4(2.8400 + 2.6625 + 3.1800 + 2.9750) - 1/4(2.3667 + 2.9000 + 1.9833 + 2.3500)$ and the se is $\sqrt{(0.358(1/16/10 + 1/16/8 + 1/16/10 + 1/16/8 + 1/16/6 + 1/16/4 + 1/16/6 + 1/16/4))} = .47$. Hence the ratio of comparison to se is 1.09 which is smaller than 3.93 and hence the null hypothesis of no difference is not rejected.