

MAT 167: STATISTICS

TEST II

INSTRUCTOR: ANTHONY TANBAKUCHI

SPRING 2008

Name: _____

Computer / Seat Number: _____

No books, notes, or friends. **Show your work.** You may use the attached equation sheet, R, and a calculator. No other materials. If you choose to use R, write what you typed on the test or copy and paste your work into a word document labeling the question number it corresponds to. When you are done with the test print out the document. Be sure to save often on a memory stick just in case. Using any other program or having any other documents open on the computer will constitute cheating.

You have until the end of class to finish the exam, manage your time wisely.

If something is unclear quietly come up and ask me.

If the question is legitimate I will inform the whole class.

Express all final answers to 3 significant digits. Probabilities should be given as a decimal number unless a percent is requested. Circle final answers, ambiguous or multiple answers will not be accepted. Show steps where appropriate.

The exam consists of 13 questions for a total of 84 points on 9 pages.

This Exam is being given under the guidelines of our institution's **Code of Academic Ethics**. You are expected to respect those guidelines.

Points Earned: _____ out of 84 total points

Exam Score: _____

1. An experiment consists of randomly sampling 10 students at Pima Community College, recording their heights and computing their mean height. If we repeat the experiment over and over (assume simple random samples, no human errors, no bias) we observe that the sample mean height varies each time.
 - (a) (2 points) What is the name of the error that causes the mean to vary each time?
 - (b) (2 points) Explain how it is possible for the sample mean height to vary each time? What is going on?
 - (c) (2 points) In words, state what the population distribution represents in this experiment. (Be specific.)
 - (d) (2 points) In words, state what the sampling distribution represents in this experiment. (Be specific.)
2. (2 points) Under what conditions can we approximate a binomial distribution as a normal distribution?
3. (2 points) Which distribution (normal, binomial, both, or neither) would be appropriate for describing:

The distribution for the number of people who wear glasses in a random sample of 20 people where the probability an individual person wears glasses is 0.56.

4. In regards to \bar{x} and the Central Limit Theorem:

(a) (2 points) What are the two conditions under which the CLT applies?

(b) (2 points) If the conditions are met, what type of distribution will \bar{x} have?

5. Let x be a random variable with a normal distribution where $\mu = 45$ and $\sigma = 10$.

(a) (2 points) Make a meaningful sketch that represents $P(50 < x < 55)$.

(b) (2 points) Find $P(50 < x < 55)$.

6. The following questions regard hypothesis testing in general.
- (a) (2 points) When we conduct a hypothesis test, we assume something is true and calculate the probability of observing the sample data under this assumption. What do we assume is true?

 - (b) (2 points) If you are using a hypothesis test to make a decision where the effect of a Type I error may negatively effect human lives, should you increase or decrease α ?

 - (c) (2 points) If you failed to reject H_0 but H_0 is false. What type of error has occurred? (Type I or Type II)

 - (d) (2 points) What variable do we use to represent a Type II error with?

 - (e) (2 points) Two studies were conducted, study A had a power of 0.9 and study B had a power of 0.10. Which study would be more likely to support a true alternative hypothesis?
7. (2 points) Describe what a normal Q-Q plot should look like for a set of data that has a normal distribution.

8. Engineers must consider the breadths of male heads when designing motorcycle helmets. Men have head breadths that are normally distributed with a mean of 6.0 in and a standard deviation of 1.0 in (based on anthropometric survey data from Gordon, Churchill, et al.).
- (a) (2 points) If 1 man is randomly selected, find the probability that his head breadth is less than 6.2 in.
- (b) (2 points) If 100 men are randomly selected, find the probability that their mean head breadth is less than 6.2 in.
- (c) (2 points) ACME motorcycle company is making a new adjustable helmet. In reality, it is not economical to make a helmet that fits everyone. You must design a helmet that will fit all but largest 5% of male head breadths. What is the largest size male head breadth that your new helmet will fit?

9. (2 points) ACME helmet company needs to know the mean head breadth of women for a new helmet design. You conduct a study of 8 randomly selected women (via a simple random sample). Below is the data from the study.

5.1, 5.7, 5.5, 6.4, 5.7, 5.9, 5.1, 6.2

Construct a 95% confidence interval for the mean head breadth size for women (**Assume σ is unknown.**)

10. You believe that the true mean head breadth for women is less than that of men (6.0 in). Using the same study data from the previous question of 8 randomly women (shown below again), conduct a hypothesis test to test your claim. Use a significance level of 0.1 and **assume σ is unknown** and women's head breadths are normally distributed.

5.1, 5.7, 5.5, 6.4, 5.7, 5.9, 5.1, 6.2

(a) (2 points) What type of hypothesis test will you use?

(b) (2 points) What are the test's requirements?

(c) (2 points) Are the requirements satisfied? **State how they are satisfied.**

- (d) (2 points) What are the hypothesis?
- (e) (2 points) What α will you use?
- (f) (2 points) Conduct the hypothesis test. What is the p -value?
- (g) (2 points) What is your formal decision?
- (h) (2 points) State your final conclusion in words.
- (i) (2 points) What is the *actual* probability of a Type I error for this study data?
- (j) (2 points) If the researcher had an α of 0.005 and failed reject H_0 , have we proven that the mean head breadth size of women is 6.0in?
11. Over the past 55 years, data from the National Oceanic and Atmospheric Administration (NOAA) indicates the the probability of precipitation¹ on a given day in Tucson is 0.146. (Use 365 days in a year.)
- (a) (2 points) Find the mean and standard deviation for the number of days per year with precipitation in Tucson.

¹Data from <http://www.wrcc.dri.edu/cgi-bin/cli1cd.pl?az23160>. Precipitation defined as 0.01 inches or more.

- (b) (2 points) What is the probability of 40 or fewer days of precipitation in Tucson annually?
- (c) (2 points) Would it be unusual to have 40 or fewer days of precipitation annually in Tucson? (Why.)
12. (2 points) You have been hired by the Tucson Weekly to estimate the proportion people who are in support of John McCain in Tucson. What random sample size should you use to estimate the proportion with a margin of error of 1%?
13. The Tucson Republican party claims that more than half the people who live in Tucson support John McCain.² They conduct a random sample of 100 people at the annual Pima Country Fair. You are hired by them to analyze the data and test their hypothesis.
- (a) (2 points) What type of hypothesis test will you use?

²The data in this question is fictitious and is not an endorsement of any candidate or party.

(b) (2 points) What are the test's requirements?

(c) (2 points) What are the hypothesis?

Using the data from the study, you run the analysis in R. Below is the output.

```
1-sample proportions test with continuity correction

data: 78 out of 100, null probability 0.5
X-squared = 30.25, df = 1, p-value = 1.899e-08
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.6995942 1.0000000
sample estimates:
      p
0.78
```

- (d) (2 points) What is the point estimate from the study for the proportion of people who support John McCain?
- (e) (2 points) What is the p -value.
- (f) (2 points) What is your formal decision?
- (g) (2 points) State your final conclusion in words based upon the analysis above.
- (h) (2 points) What is wrong with this study.

```
*****
End of exam. Reference sheets follow.
*****
```

Basic Statistics: Quick Reference & R Commands

by Anthony Tambakuchi. Version 1.8
<http://www.tambakuchi.com>
 ANTHONY@TANBAKUCHI.COM
 Get R at: <http://www.r-project.org>
 R commands: bold typewriter text

1 Misc R

To make a vector v store data: $x=c(x1, x2, \dots)$
 Get help on function: `?functionName`
 Get column of data from table:
`tableName$columnName`
 List all variables: `ls()`
 Delete all variables: `rm(list=ls())`

$$\begin{aligned} \sqrt{x} &= \text{sqrt}(x) & (1) \\ x^n &= x^n & (2) \\ n &= \text{length}(x) & (3) \\ T &= \text{table}(x) & (4) \end{aligned}$$

2 Descriptive Statistics

2.1 NUMERICAL

Let $x=c(x1, x2, x3, \dots)$

$$\begin{aligned} \text{total} &= \sum_{i=1}^n x_i = \text{sum}(x) & (5) \\ \text{min} &= \text{min}(x) & (6) \\ \text{max} &= \text{max}(x) & (7) \\ \text{six number summary} &: \text{summary}(x) & (8) \\ \bar{\mu} &= \frac{\sum x_i}{N} = \text{mean}(x) & (9) \\ \bar{x} &= \frac{\sum x_i}{n} = \text{mean}(x) & (10) \\ \tilde{x} &= P_{50} = \text{median}(x) & (11) \end{aligned}$$

$$\begin{aligned} \sigma &= \sqrt{\frac{\sum (x_i - \mu)^2}{N}} & (12) \\ s &= \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \text{sd}(x) & (13) \\ CV &= \frac{\sigma}{\mu} = \frac{s}{\bar{x}} & (14) \end{aligned}$$

2.2 RELATIVE STANDING

$$z = \frac{x - \bar{x}}{\sigma} = \frac{x - \bar{x}}{s} \quad (15)$$

Percentiles:

$$\begin{aligned} P_k &= x_{(k)} \quad (\text{sorted } x) \\ k &= \frac{i-0.5}{n} \cdot 100\% & (16) \end{aligned}$$

To find x_i given P_k , i is:

- $L = (k/100)n$
- if L is an integer: $i = L + 0.5$; otherwise $i = L$ and round up.

2.3 VISUAL

All plots have optional arguments:

- `main=""` sets title
- `xlab=""`, `ylab=""` sets x/y-axis label
- `type="p"` for point plot
- `type="l"` for line plot
- `type="b"` for both points and lines

Ex: plot(x, y, type="b", main="My Plot")

Plot Types:

- `hist(x)` histogram
- `stem(x)` stem & leaf
- `boxplot(x)` box plot
- `plot(T)` bar plot, `T=table(x)`
- `plot(x, y)` scatter plot, x, y are ordered vectors
- `plot(t, y)` time series plot, t, y are ordered vectors
- `curve(expr, xmin, xmax)` plot expr involving x

2.4 ASSESSING NORMALITY

Q-Q plot: `qqnorm(x)`; `qqline(x)`

3 Probability

Number of successes x with n possible outcomes. (Don't double count!)

$$\begin{aligned} P(A) &= \frac{x}{n} & (17) \\ P(\bar{A}) &= 1 - P(A) & (18) \\ P(A \text{ or } B) &= P(A) + P(B) - P(A \text{ and } B) & (19) \\ P(A \text{ or } B) &= P(A) + P(B) \quad \text{if } A, B \text{ mut. excl.} & (20) \\ P(A \text{ and } B) &= P(A) \cdot P(B|A) & (21) \\ P(A \text{ and } B) &= P(A) \cdot P(B) \quad \text{if } A, B \text{ independent} & (22) \\ n! &= n(n-1) \cdots 1 = \text{factorial}(n) & (23) \\ n!_k &= \frac{n!}{(n-k)!} \quad \text{Perm. no elem. alike} & (24) \\ &= \frac{n!}{n_1! n_2! \cdots n_k!} \quad \text{Perm. } n_1 \text{ alike, } \dots & (25) \\ nC_k &= \frac{n!}{(n-k)!k!} = \text{choose}(n, k) & (26) \end{aligned}$$

4 Discrete Random Variables

$$\begin{aligned} P(x_i) &: \text{probability distribution} & (27) \\ E = \mu &= \sum x_i \cdot P(x_i) & (28) \\ \sigma &= \sqrt{\sum (x_i - \mu)^2 \cdot P(x_i)} & (29) \end{aligned}$$

4.1 BINOMIAL DISTRIBUTION

$$\begin{aligned} \mu &= n \cdot p & (30) \\ \sigma &= \sqrt{n \cdot p \cdot q} & (31) \\ P(x) &= {}^n C_x p^x q^{n-x} = \text{dbinom}(x, n, p) & (32) \end{aligned}$$

4.2 POISSON DISTRIBUTION

$$P(x) = \frac{\mu^x \cdot e^{-\mu}}{x!} = \text{dpois}(x, \mu) \quad (33)$$

5 Continuous random variables

CDF $F(x)$ gives area to the left of x , $F^{-1}(p)$ expects p is area to the left.

$$\begin{aligned} f(x) &: \text{probability density} & (34) \\ E = \mu &= \int_{-\infty}^{\infty} x \cdot f(x) dx & (35) \\ \sigma &= \sqrt{\int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) dx} & (36) \\ F(x) &: \text{cumulative prob. density (CDF)} & (37) \\ F^{-1}(p) &: \text{inv. cumulative prob. density} & (38) \\ F(x) &= \int_{-\infty}^x f(x') dx' & (39) \\ p = P(x < x') &= F(x') & (40) \\ x' &= F^{-1}(p) & (41) \\ p = P(x > a) &= 1 - F(a) & (42) \\ p = P(a < x < b) &= F(b) - F(a) & (43) \end{aligned}$$

5.1 UNIFORM DISTRIBUTION

$$\begin{aligned} p &= P(a < u') = F(u') \\ &= \text{punif}(u', \text{min}=0, \text{max}=1) & (44) \\ u' &= F^{-1}(p) = \text{qunif}(p, \text{min}=0, \text{max}=1) & (45) \end{aligned}$$

5.2 NORMAL DISTRIBUTION

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} & (46) \\ p &= P(z < z') = F(z') = \text{pnorm}(z') & (47) \\ z' &= F^{-1}(p) = \text{qnorm}(p) & (48) \\ p &= P(x < x') = F(x') \\ &= \text{pnorm}(x', \text{mean}=\mu, \text{sd}=\sigma) & (49) \\ x' &= F^{-1}(p) \\ &= \text{qnorm}(p, \text{mean}=\mu, \text{sd}=\sigma) & (50) \end{aligned}$$

5.3 t-DISTRIBUTION

$$\begin{aligned} p &= P(t < t') = F(t') = \text{pt}(t', \text{df}) & (51) \\ t' &= F^{-1}(p) = \text{qt}(p, \text{df}) & (52) \end{aligned}$$

5.4 χ^2 -DISTRIBUTION

$$\begin{aligned} p &= P(\chi^2 < \chi'^2) = F(\chi'^2) \\ &= \text{pchisq}(\chi'^2, \text{df}) & (53) \\ \chi'^2 &= F^{-1}(p) = \text{qchisq}(p, \text{df}) & (54) \end{aligned}$$

5.5 F-DISTRIBUTION

$$\begin{aligned} p &= P(F < F') = F(F') \\ &= \text{pf}(F', \text{df1}, \text{df2}) & (55) \\ F' &= F^{-1}(p) = \text{qf}(p, \text{df1}, \text{df2}) & (56) \end{aligned}$$

6 Sampling distributions

$$\mu_{\bar{x}} = \mu \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (57)$$

$$\mu_{\bar{p}} = p \quad \sigma_{\bar{p}} = \sqrt{\frac{pq}{n}} \quad (58)$$

7 Estimation

7.1 CONFIDENCE INTERVALS

$$\text{proportion: } \hat{p} \pm E, \quad E = z_{\alpha/2} \cdot \sigma_{\bar{p}} \quad (59)$$

$$\text{mean } (\sigma \text{ known): } \bar{x} \pm E, \quad E = z_{\alpha/2} \cdot \sigma_{\bar{x}} \quad (60)$$

$$\text{mean } (\sigma \text{ unknown, use } s): \bar{x} \pm E, \quad E = t_{\alpha/2} \cdot \sigma_{\bar{x}} \quad (61)$$

$$\begin{aligned} df &= n - 1 \\ \text{variance: } \frac{(n-1)s^2}{\chi^2_{\alpha/2}} &< \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} & (62) \\ df &= n - 1 \end{aligned}$$

$$2 \text{ proportions: } \Delta \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}} \quad (63)$$

$$2 \text{ means (indep): } \Delta \bar{x} \pm t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad (64)$$

$$\begin{aligned} df &\approx \min(n_1 - 1, n_2 - 1) \\ \text{matched pairs: } \bar{d} \pm z_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}, \quad d_i &= x_i - y_i, & (65) \\ df &= n - 1 \end{aligned}$$

7.2 CI CRITICAL VALUES (TWO SIDED)

$$z_{\alpha/2} = F_z^{-1}(1 - \alpha/2) = \text{qnorm}(1 - \alpha/2) \quad (66)$$

$$t_{\alpha/2} = F_t^{-1}(1 - \alpha/2) = \text{qt}(1 - \alpha/2, \text{df}) \quad (67)$$

$$\chi^2_{\alpha/2} = F_{\chi^2}^{-1}(\alpha/2) = \text{qchisq}(\alpha/2, \text{df}) \quad (68)$$

$$\chi^2_{1-\alpha/2} = F_{\chi^2}^{-1}(1 - \alpha/2) = \text{qchisq}(1 - \alpha/2, \text{df}) \quad (69)$$

7.3 REQUIRED SAMPLE SIZE

$$\begin{aligned} \text{proportion: } n &= \hat{p}\hat{q} \left(\frac{z_{\alpha/2}}{E}\right)^2, & (70) \\ (\hat{p} = \hat{q} = 0.5 \text{ if unknown}) \end{aligned}$$

$$\text{mean: } n = \left(\frac{z_{\alpha/2} \cdot \hat{\sigma}}{E}\right)^2 \quad (71)$$

8 Hypothesis Tests

Test statistic and R function (when available) are listed for each.

Optional arguments for hypothesis tests:

alternative="two_sided" can be:
"two_sided", "less", "greater"

conf.level=0.95 constructs a 95% confidence interval. Standard CI only when alternative="two_sided".

Optional arguments for power calculations & Type II error:

alternative="two_sided" can be:
"two_sided" or "one_sided"

sig.level=0.05 sets the significance level α .

8.1 1-SAMPLE PROPORTION

$H_0: p = p_0$

prop.test(x, n, p=p₀, alternative="two_sided")

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0q_0/n}} \quad (72)$$

8.2 1-SAMPLE MEAN (σ KNOWN)

$H_0: \mu = \mu_0$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad (73)$$

8.3 1-SAMPLE MEAN (σ UNKNOWN)

$H_0: \mu = \mu_0$

t.test(x, mu= μ_0 , alternative="two_sided")

Where **x** is a vector of sample data.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \quad df = n - 1 \quad (74)$$

Required Sample size:

power.t.test(delta= h , sd= σ , sig.level= α , power=1 - β , type="one.sample", alternative="two_sided")

8.4 2-SAMPLE PROPORTION TEST

$H_0: p_1 = p_2$ or equivalently $H_0: \Delta p = 0$

prop.test(x, n, alternative="two_sided")

where: **x=c(x₁, x₂)** and **n=c(n₁, n₂)**

$$z = \frac{\hat{\Delta p} - \Delta p_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}, \quad \Delta \hat{p} = \hat{p}_1 - \hat{p}_2 \quad (75)$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}, \quad \hat{q} = 1 - \hat{p} \quad (76)$$

Required Sample size:

power.prop.test(p1= p_1 , p2= p_2 , power=1 - β , sig.level= α , alternative="two_sided")

8.5 2-SAMPLE MEAN TEST

$H_0: \mu_1 = \mu_2$ or equivalently $H_0: \Delta \mu = 0$

t.test(x1, x2, alternative="two_sided")

where: **x1** and **x2** are vectors of sample 1 and sample 2 data.

$$t = \frac{\Delta \bar{x} - \Delta \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad df \approx \min(n_1 - 1, n_2 - 1), \quad \Delta \bar{x} = \bar{x}_1 - \bar{x}_2 \quad (77)$$

Required Sample size:

power.t.test(delta= h , sd= σ , sig.level= α , power=1 - β , type="two.sample", alternative="two_sided")

8.6 2-SAMPLE MATCHED PAIRS TEST

$H_0: \mu_d = 0$

t.test(x, y, paired=TRUE, alternative="two_sided")

where: **x** and **y** are ordered vectors of sample 1 and sample 2 data.

$$t = \frac{\bar{d} - \mu_{d0}}{s_d/\sqrt{n}}, \quad d_i = x_i - y_i, \quad df = n - 1 \quad (78)$$

Required Sample size:

power.t.test(delta= h , sd= σ , sig.level= α , power=1 - β , type="paired", alternative="two_sided")

8.7 TEST OF HOMOGENEITY, TEST OF INDEPENDENCE

$H_0: p_1 = p_2 = \dots = p_k$ (homogeneity)

$H_0: X$ and Y are independent (independence)

chisq.test(D)

Enter table: **D=matrix(c1, c2, ...)**, where **c1, c2, ...** are column data vectors.

Or generate table: **D=table(x1, x2)**, where **x1, x2** are ordered vectors of raw categorical data.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}, \quad df = (\text{num rows} - 1)(\text{num cols} - 1) \quad (79)$$

$$E_i = \frac{(\text{row total})(\text{column total})}{(\text{grand total})} = np_i \quad (80)$$

For 2×2 contingency tables, you can use the Fisher Exact Test:

fisher.test(D, alternative="greater")

(must specify alternative as greater)

9 Linear Regression

9.1 LINEAR CORRELATION

$H_0: \rho = 0$

cor.test(x, y)

where: **x** and **y** are ordered vectors.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}, \quad t = \frac{r-0}{\sqrt{\frac{1-r^2}{n-2}}} \quad df = n - 2 \quad (81)$$

9.2 MODELS IN R

MODEL TYPE	EQUATION	R MODEL
linear 1 indep var	$y = b_0 + b_1x_1$	$y \sim x_1$
... 0 intercept	$y = 0 + b_1x_1$	$y \sim 0 + x_1$
linear 2 indep vars	$y = b_0 + b_1x_1 + b_2x_2$	$y \sim x_1 + x_2$
... interaction	$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2$	$y \sim x_1 + x_2 + x_1 * x_2$
polynomial	$y = b_0 + b_1x_1 + b_2x_1^2$	$y \sim x_1 + 1(x_1^2)^2$

9.3 REGRESSION

Simple linear regression steps:

1. Make sure there is a significant linear correlation.
2. **results=lm(y~x)** Linear regression of y on x vectors
3. **results** View the results
4. **plot(x, y); abline(results)** Plot regression line on data
5. **plot(x, results\$residuals)** Plot residuals

$$y = b_0 + b_1x_1 \quad (82)$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (83)$$

$$b_0 = \bar{y} - b_1\bar{x} \quad (84)$$

9.4 PREDICTION INTERVALS

To predict y when $x = 5$ and show the 95% prediction interval with regression model in results:

predict(results, newdata=data.frame(x=5), int="pred")

10 ANOVA

10.1 ONE WAY ANOVA

1. **results=aoov(depVarColName~indepVarColName, data=tableName)** Run ANOVA with data in **tableName**, factor data in **indepVarColName** column, and response data in **depVarColName** column.
 2. **summary(results)** Summarize results
 3. **boxplot(depVarColName~indepVarColName, data=tableName)** Boxplot of levels for factor
- To find required sample size and power see **power.anova.test(...)**

11 Loading and using external data and tables

11.1 LOADING EXCEL DATA

1. Export your table as a CSV file (comma separated file) from Excel.
2. Import your table into **MyTable** in R using:
MyTable=read.csv(File.choose())

11.2 LOADING AN .RDATA FILE

You can either double click on the .Rdata file or use the menu:

- Windows: **File—Load Workspace...**
- Mac: **Workspace—Load Workspace File...**

11.3 USING TABLES OF DATA

1. To see all the available variables type: **ls()**
2. To see what's inside a variable, type its name.
3. If the variable **tableName** is a table, you can also type **names(tableName)** to see the column names or type **head(tableName)** to see the first few rows of data.
4. To access a column of data type **tableName\$columnName**

An example demonstrating how to get the women's height data and find the mean:

```
> ls() # See what variables are defined
[1] "women" "x"
> head(women) # Look at the first few entries
height weight
1 58 115
2 59 117
3 60 120
> names(women) # Just get the column names
[1] "height" "weight"
> women$height # Display the height data
[1] 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72
> mean(women$height) # Find the mean of the heights
[1] 65
```