HS 267 -- Exam 1 -- March 9, 2006

Please write your name on the **back**. Don't forget!

**Part A:**

Short answer, multiple choice, and true or false questions. **No** use of calculators, notes, lab workbooks, cell phones, neighbors, brain implants, banned performance-enhancing substances, or other unspecified test-taking aides. Unless otherwise specified, each Part A answer is worth 1 point. **Good luck!**

1) Select a **synonym** for variance:
   a. Standard deviation
   b. Root mean square
   c. Mean square
   d. Inter-quartile range
2) Standard error is a measure of:
   a. Variance
   b. The mean's precision
   c. Mean square
   d. Validity
3) Why would you pool the variances if the F-ratio test is **not** significant?

   _____

   _____
4) Which of the following is **not** a measure of central location?
   a. Median
   b. Kurtosis
   c. Mean
   d. Mode
5) Name two methods of exploratory data analysis (EDA):
   a. _____
   b. _____
6) True or False? Both a stem-leaf plot and a box plot show **all** of the original data points.

   _____ If false, explain why: _____

   _____
7) True or False? 95% of normally distributed data are within $\pm$1 standard deviation of the mean.

   _____ If false, explain why: _____

   _____

8) Add the following data points to a stem-leaf plot with a common stem:

```
Group A: 74, 82, 85, 86, 86, 88, 89, 90, 91
Group B:     70, 80, 85, 85, 86, 88, 93
```

```
|7|
|7|
|8|
|8|
|9|
|9|
(x10)
```

9) Compare the means and variances between test scores for Group 1 and Group 2 for each figure below. Match the figure, labeled 1-4, with each of the following statements (a-d):

a. Same mean, same variance Fig. _____

b. Different means, same variance Fig. _____

c. Same mean, different variances Fig. _____

d. Different means, different variances Fig. _____

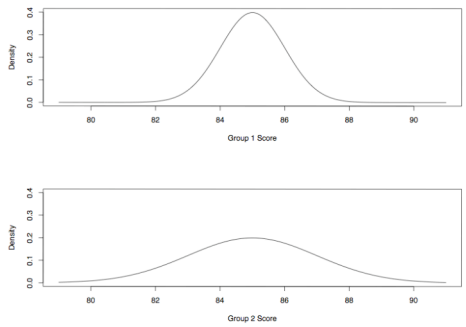



Fig. 1

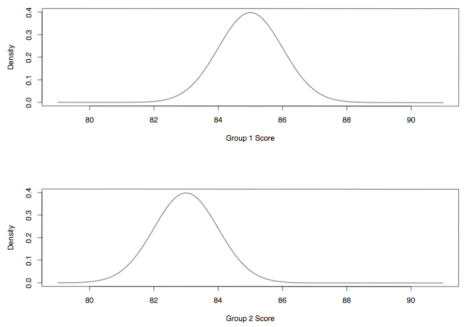



Fig. 2

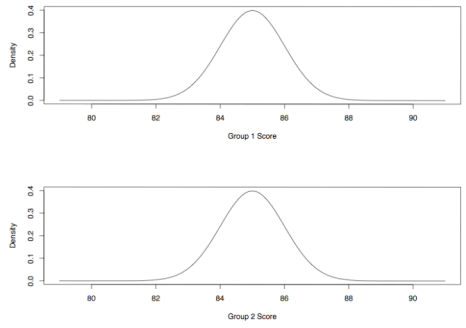



Fig. 3

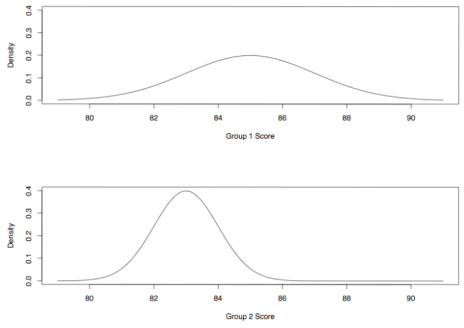



Fig. 4

10) True or False? Alpha is a measure of **Type I** error. _____

    If false, explain why: _____

11) Name two **validity** assumptions when comparing means using a t-test or ANOVA.

    **Briefly** describe each assumption.

    a. _____

    _____

    b. _____

    _____

12) Name two **distributional** assumptions when comparing means using a t-test or ANOVA.

    **Briefly** describe each assumption.

    a. _____

    _____

    b. _____

    _____

13) Write the **alternative** hypothesis for ANOVA:

    _____

14) ANOVA compares means by partitioning the variance into:

    _____ and _____

15) The **mean square between** quantifies the variability of the groups around the

    _____

16) The **mean square within** quantifies the variability of the individuals around the

    _____

17) A **non-parametric** test that is analogous to ANOVA is called:

    _____

18) How many **post hoc** comparisons are possible for 3 groups?

    _____

19) Which of the following does **not** test for equality of central location?

    a. t-test
    b. Kruskal-Wallis
    c. F-ratio test
    d. ANOVA

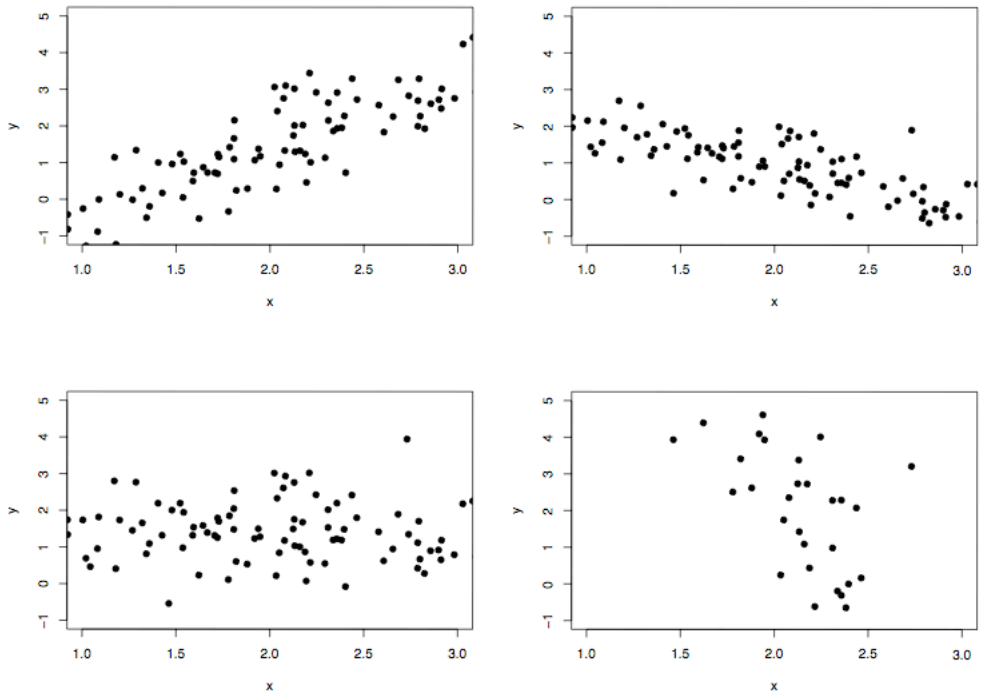
20) SJSU collected survey data from 256 current and former Health Sciences students. An initial analysis of the data focused on participant grade point average (GPA). The participants were divided into 5 groups by age at entry into the department. The surveys were hand-written and took a great deal of labor to enter into computer software. With limited resources, **two of the groups** (the oldest and youngest) were entered. A **t-test** was conducted to compare the GPAs of the two groups. The difference in GPA was found to be **highly significant**. If the Health Sciences department allocates additional resources to enter the rest of the groups, and all group GPAs are compared using **ANOVA**, **what** will the result be and **why**?

   a. Reject the null hypothesis _____
   b. Retain the null hypothesis _____
   c. Accept the null hypothesis _____
   d. Not enough information to conclude _____

21) A **negative** correlation means that:

   _____ values of variable X are associated with
   _____ values of variable Y.

22) What type of plot is shown below? _____




23) **Circle** the diagram above that shows variables with **positive** correlation.

24) We'd like to examine the relationship between % of reduced-fee school lunches and % usage of bicycle helmets. For this study, specify the dependent and independent variables:

   a. Independent _____

   b. Dependent _____

25) Which of the following is **not** a possible value for a **correlation coefficient**?

   a. 1

   b. 2

   c. -1

   d. 0

26) Correlation requires one extra **distributional** assumption. **Name** this and **briefly** describe.

   _____

   _____

27) A calculation of **r = -0.5** generally indicates:

   a. An error: **r** can never be negative

   b. Strong negative correlation

   c. Weak negative correlation

   d. Moderate negative correlation

28) Assuming appropriate distributions, a calculation of $\mathbf{r^2 = 0.64}$ can be interpreted as:

   a. There is no correlation between X and Y

   b. 64% of the variability in Y can be statistically explained by X

   c. There is a strong negative correlation

   d. 0.64% of the variability in Y can be statistically explained by X

29) **BONUS QUESTION +1 Point** (OPTIONAL):

   Under what condition would an F statistic of 1.0 give you exactly 50% probability in both the left and right tails. (Hint: think about the degrees of freedom).

   _____

   _____

**Part B:**

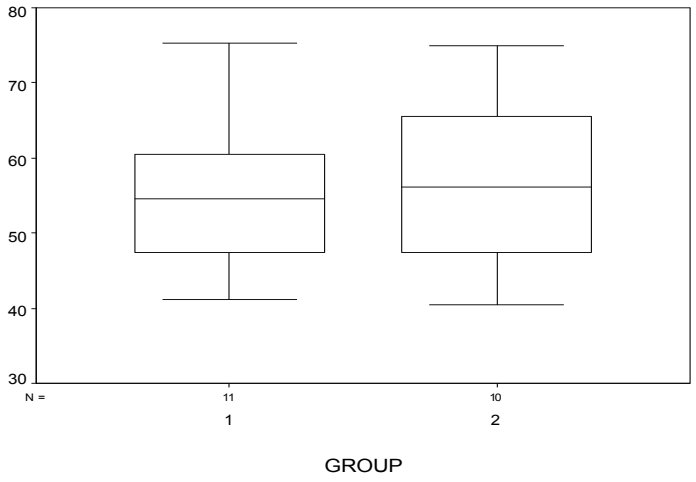
Please turn in Part A **before** working on Part B.

You can use your **Procedure Notebook** for Part B of the exam.

Write all your answers in your **blue book**, numbering each response. Good luck

**1) Comparing means and variances**

From the following side-by-side boxplot, **briefly** compare the two groups in terms of: central location, spread, range, potential outliers.




**2) t-Test**

From the following SPSS output (modified) of performing a **two-sided** independent t-test,

| | | | | | | |
|---|---|---|---|---|---|---|
| | t | df | | | | |
| Equal variances assumed | | 19 | .000 | | | |
| Equal variances not assumed | | 19.0 | | | 3 | |

A) Assuming equal variance, **draw** the *t* curve, l**abel** $t_{stat}$ , *df* , and landmarks, **shade** the area corresponding to *p* value region.

B) Give the estimated *p* value.

C) Do we reject or retain $H_o$ at alpha = 0.01 ?

**3) F-ratio Test**

From the following table,

| Group | Group size (n) | Mean | Standard deviation |
|---|---|---|---|
| 1 | 7 | 7.9 | 1.23 |
| 2 | 9 | 9.8 | 2.32 |

A) Calculate the $F_{stat}$

B) **Draw** the $F$ curve, l**abel** $F_{stat}$ , $df_N$ , $df_D$ , and landmarks, **shade** the area corresponding to $p$ value region.

C) Give the estimated $p$ value

D) With alpha = 0.05, do the groups have equal variances?

**4) ANOVA**

**Hypertension**

Some common strategies for treating hypertensive patients by nonpharmacologic methods include weight reduction and trying to get the patient to relax by meditation. Suppose these strategies are evaluated by randomizing hypertensive patients to 3 groups.

Table. Change in diastolic blood pressure (DBP) among hypertensive patients who receive different kinds of nonpharmacologic therapy.

| Group | Group size (n) | Mean (change in DBP – mmHg) | Standard deviation | Variance |
|---|---|---|---|---|
| 1 (Counseling for both weight reduction and meditation) | 20 | 8.6 | 6.2 | 38.4 |
| 2 (Counseling for weight reduction only) | 17 | 3.3 | 5.7 | 32.5 |
| 3 (Counseling for meditation only) | 15 | 2.1 | 7.0 | 49.0 |

Using the data presented in the table, conduct an ANOVA.

A) List $H_o$ and $H_1$

B) Sum of Squares Between ($SS_B$) =

C) $df_B$ =

D) Mean Square Between ($s^2_B$) =

E) Sum of Squares Within ($SS_W$) =

F) $df_W$ =

G) Mean Square Within ($s^2_W$) =

H) $F_{stat}$ =

I) The $p$ value for ANOVA = 0.007. What's your conclusion and interpretation?

**5) Post-hoc**

The problem above is further analyzed with LSD post-hoc to see which group means are different. From the following SPSS output,

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 0 | | | | |
| | 3 | | | | | |
| 2 | 1 | | | | | 8 |
| | 3 | 0 | | | | 4 |
| 3 | 1 | | | | 8 | |
| | 2 | | 0 | | | |

A) Test 1: group 1 vs. group 2, $p$ value =

B) Test 2: group 1 vs. group 3, $p$ value =

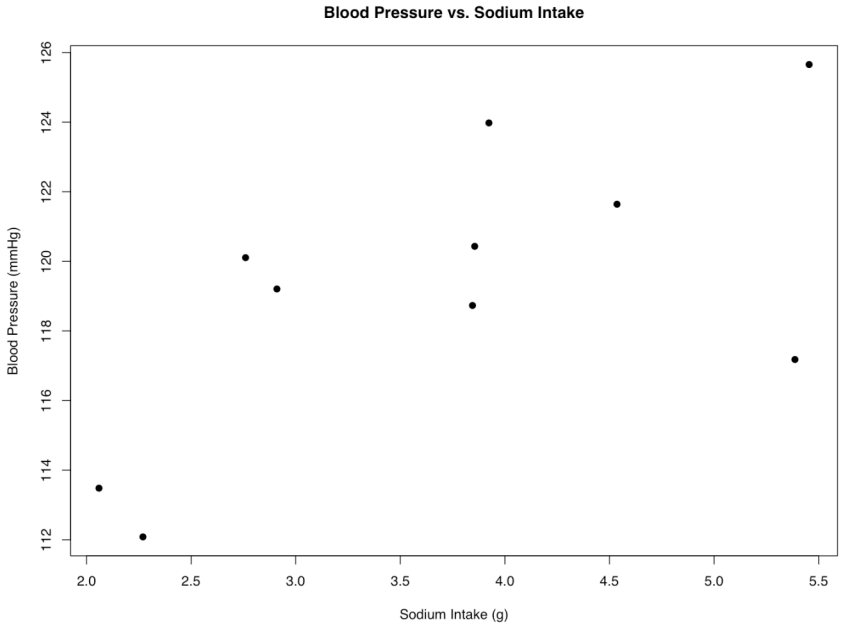
C) Test 3: group 2 vs. group 3, $p$ value =

D) Which groups are significantly different at alpha = 0.05 ?

E) Using Bonferroni's adjustment for Test 2: group 1 vs. group 3, $p_{Bonf}$ value =

**6) Correlation**

A study of the relationship between sodium intake and blood pressure was conducted. Measurements of sodium intake (g) and blood pressure (mmHg) were taken for each of 10 individuals. Below is a plot of the measurements.




Descriptive statistics for this data set include:

| N = 10 | Sodium Intake (g) | Blood Pressure (mmHg) |
|---|---|---|
| Mean | 3.70 | 119.25 |
| Standard Deviation | 1.20 | 4.22 |
| Variance | 1.44 | 17.82 |

Let **X** be sodium intake and **Y** be blood pressure. Using the descriptive statistics given:

A) Calculate $SS_{XX}$

B) Calculate $SS_{YY}$

C) Given $SS_{XY} = 29.85$, correlation coefficient $r =$

D) Testing for the significance of this correlation coefficient $r$, we obtained a $p$ value = 0.04
Provide an interpretation of the correlation coefficient **and** the result of the significance test.

**YOU ARE DONE WITH THIS EXAM**

**Have a relaxing weekend !**