Learning Center's Middlebush Room 12
Statistics 1300 Friday, May 9, 2007
Final Exam Review 10:00 a.m.-12:00 p.m.

Note 1: In addition to these review problems, I would advise students to work through problems from the sample finals posted on Blackboard.
Note 2: There will be a Q&A session on Thursday, May 15 from 12:00 to 1:30 p.m. in Middlebush 211.

1. A student's final grade can sometimes be affected in a negative way by missing too many classes throughout a semester. To analyze the notion that the more days a student misses (x), the lower your final grade (y, in percentage), the data below were collected from a random sample of 10 students.

x	2	1	10	5	25	12	7	40	22	0
y	84	95	85	75	54	58	81	70	62	91

$\Sigma x = 124,\ \Sigma y = 755,\ \Sigma x^2 = 3032,\ \Sigma y^2 = 58797,\ \Sigma xy = 8265$

a) Find the equation of the least squares line.
b) What final grade percentage should be predicted for a student who misses 9 classes?
c) Find the sample correlation coefficient.
d) Assuming that, for each value of x, the corresponding y's have the same variance, find the estimate for this common variance.

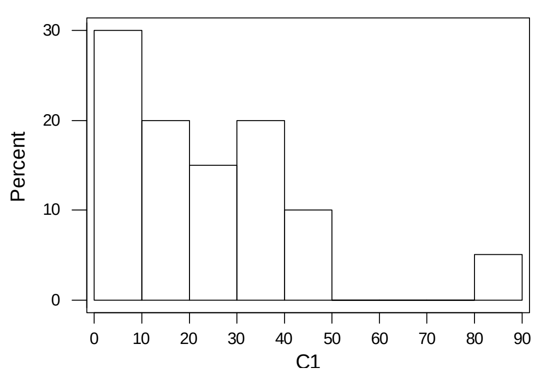
2. Six subjects were selected at random, and the age and systolic blood pressure for each was recorded. Their ages ranged from 43 to 70, and their blood pressures ranged from 120 to 152. Using this data, *Minitab* produced the regression analysis output that is given below.

Regression Analysis: Pressure versus Age

```
The regression equation is
Pressure = 81.0 + 0.964 Age

Predictor        Coef     SE Coef          T        P
Constant        81.05       13.88       5.84    0.004
Age            0.9644      0.2381       4.05    0.015

S = 5.641       R-Sq = 80.4%     R-Sq(adj) = 75.5%

Analysis of Variance
Source           DF          SS          MS         F        P
Regression        1      522.21      522.21     16.41    0.015
Residual Error    4      127.29       31.82
Total             5      649.50
```

a) If we test the hypothesis $H_0 : \beta_1 = 0$ against $H_A : \beta_1 \neq 0$, what is the *p*-value for the test?
b) Using $\alpha = .05$, what would we conclude for the test in (a)?
c) If we test the hypothesis that blood pressure tends to increase with age, what is the *p*-value for the test?

3. Use the relative frequency histogram at the right to answer the following questions. Read all relative frequencies to the nearest 5%.



a) If there are 80 observations in the data set, how many observations are between 30 and 50?
b) Find the median.
c) Based on the histogram, which should be larger, the mean or the median? Why?

4. Find the standard deviation of the following data set: {1, 4, 5, 9}.

5. Of the members of a particular Poker club, 60% play *Hold 'em* regularly, 55% play *Seven Card Stud* regularly, and 30% play both *Hold 'em* and *Seven Card Stud* regularly.
a) If members are selected at random, what is the probability that he or she will play either *Hold 'em* or *Seven Card Stud* regularly?
b) Are the events "plays *Hold 'em* regularly" and "plays *Seven Card Stud* regularly" independent? Explain.
c) Are the events "plays *Hold 'em* regularly" and "plays *Seven Card Stud* regularly" mutually exclusive? Explain.

6. Suppose that 20% of professional comedians quit performing by the time they are 40 years of age.
a) If 15 professional comedians are selected at random, what is the probability that at least 5 will quit by age 40?
b) If 16 professional comedians are selected at random, what is the probability that exactly 2 will quit by age 40?

7. Find the expected value and variance of the probability distribution at the right.

x	1	2	3	4	5
p(x)	0.2	0.3	0.3	0.1	0.1

8. Scores from an old Statistics 250 exam (from winter semester 2003) are normally distributed with a mean of 78 and a standard deviation of 6.
a) What is the probability that a randomly selected student (who took that particular exam) will score above an 88?
b) How high a score would a person need to be in the **top** 15% on this particular exam?

9. In a random sample of 200 Mizzou football fans, 160 said they were already excited about next year's season of Tiger football.
a) Find a 95% Confidence Interval for the true proportion of Mizzou football fans that are excited about the upcoming football season.
b) To estimate the true proportion of Mizzou fans that are excited about the upcoming football season to within 0.04 with 95% confidence, what sample size is needed?

10. A random sample of 9 trips by a professor to the beautiful *Isle of Capri* casino yielded mean winnings of $22 with a standard deviation of $15.
a) Find a 90% confidence interval for the professor's average winnings.
b) Based on the 90% confidence interval, should we be confident that the professor wins more than $15 on an average visit? Explain.

11. A random sample of 100 brand new Firestone tires has 13 defective tires, and a random sample of 120 brand new Goodyear tires yields 9 that are defective.
a) Using $\alpha = 0.05$, conduct a test to determine if the proportion of defective tires is different for the two brands. What should you conclude?
b) Find the p-value of the test. Based on the *p*-value, what should you conclude?

12. To decide whether a new type of bumper performs better in low-speed crashes, 6 cars with the new bumpers and 6 cars with the old bumper design were each randomly selected and crashed into a concrete wall at a speed of 5 miles per hour. The cost of repairing the damages is given in the table below. Your assistant did not know whether you wanted to look at each sample separately or at paired differences, so he did calculations both ways.

New	127	168	143	165	122	139	$\bar{x} = 144$	$s = 19.06$
Old	154	135	132	171	153	149	$\bar{x} = 149$	$s = 14.21$
Difference	–27	33	11	–6	–31	–10	$\bar{x} = -5$	$s = 24.02$

a) Decide what kind of test would be appropriate, and carry it out at the .05 level of significance. You may assume equal variances.
b) Find the *p*-value for the test as accurately as possible. Would your conclusion in (a) have been different if the level of significance had been .01? Explain.

ANSWERS

1. (a) $\hat{y} = 84.6 - .734x$
(b) 78%
(c) $-.670$
(d) 123.66
2. (a) 0.015
(b) Reject H_0. There is evidence that the slope is not 0.
(c) .0075
3. (a) 24
(b) 20
(c) Mean > median. The distribution is skewed to the right.
4. 3.304
5. (a) 0.85
(b) Not independent;
$P(H \cap S) = 0.30 \neq P(H) \cdot P(S) = 0.33$
(c) Not mutually exclusive; $P(H \cap S) = 0.30 \neq 0$
6. (a) .164
(b) .211
7. $E(x) = 2.6$, $\sigma^2 = 1.44$
8. (a) .0475
(b) 84.24
9. (a) (0.745, 0.855)
(b) 385
10. (a) (12.7, 31.3)
(b) No, 15 is inside the confidence interval.
11. (a) $H_O: p_F - p_G = 0$, $H_A: p_F - p_G \neq 0$
Test statistic: $z = 0.96$
Rejection Region: $z < -1.96$ or $z > 1.96$
Conclusion: Do not reject H_o. There is not evidence that the proportion of defective tires differs for the two brands.
(b) *p*-value = .1770; same conclusion as (a)
12. (a) $H_O: \mu_1 - \mu_2 = 0$, $H_O: \mu_1 - \mu_2 < 0$
Test statistic: $t = -.515$
Rejection Region: $t < -1.812$
Conclusion: Do not reject H_o. There is not evidence that the new bumpers perform better.
(b) *p*-value > .10. If $\alpha = .01$, the *p*-value would be greater than α. We would not reject H_o.