THE UNIVERSITY OF MANITOBA

DATE: December 11, 2006 (9:00-11:00 A.M.)
PAPER NO: 167
COURSE NO.: STAT 2000
EXAMINATION: Basic Statistical Analysis II

FINAL EXAMINATION

TIME: 2 hours
EXAMINER(S): Thavaneswaran & Zhang

INSTRUCTIONS

I. You have been provided with:

a) the examination paper
b) a multiple choice answer sheet
c) selected formula sheet
d) a blank sheet of paper at the end of examination paper
e) a booklet of tables

II. The total number of marks possible is 50.

III. a) This examination is worth 50 marks and consists of 32 multiple-choice questions (of equal value). It is suggested that you first complete the questions on the examination paper by choosing the **BEST** answer out of five in each case; then transfer your answers to the multiple choice answer sheet by blackening the appropriate space with an "HB" or "F" pencil. Only one space should be blackened; otherwise, the question will be marked wrong. There is no penalty for guessing; therefore, all questions should be attempted.

b) Do **NOT** blacken spaces with a pen.

IV. At the end of the examination period, **turn in** (i) your multiple choice answer sheet, and (ii) the booklet of tables. Be sure to write your **NAME and STUDENT NUMBER, and your instructor's name on your MULTIPLE CHOICE ANSWER SHEET. Carefully blacken the spaces for your student number.**

V. Calculators are permitted; however, graphing, programmable calculators, electronic dictionaries, and cell phones are **NOT** permitted.

1. Suppose that we wish to estimate the proportion of Canadians who actually understand the Constitution of Canada. What is the approximate number of Canadians who need to be sampled so that the 95% confidence interval has a width of two percentage points?

   a) about 500
   b) about 1,000
   c) about 2,500
   d) about 5,000
   e) about 10,000

2. The following table gives the number of wins for each of the first four post positions at Assiniboine Downs for 80 races during the 2005 horseracing season.

Post Position	**No. of Wins**
1	24
2	17
3	19
4	20

   For testing the hypothesis that the probability of winning is the same for all four post positions, the calculated value of the test statistic is:

   a) 26.00 b) 1.25 c) 1.30 d) 0.40 e) 20.00

3. A candy manufacturer sells packages containing 10 gumdrops. The gumdrops are made in many different colours and the colours of gumdrops vary from one package to another. We take a sample of 200 packages and count the number of red gumdrops in each one. Packages can contain anywhere from zero to 10 red gumdrops. We will conduct a chi-square goodness-of-fit test to determine if the number of red gumdrops per package follows a binomial distribution. The parameter p is unknown and must be estimated from the sample data. Assuming all expected cell counts are at least five, the degrees of freedom for the test is:

   a) 8 b) 9 c) 10 d) 198 e) 199

Questions 4 and 5 refer to the following:

Popular wisdom is that eating presweetened cereal tends to increase the number of dental caries (cavities) in children. A random sample of children was (with parental consent) entered into a study and followed for several years. Each child was classified as a sweetened cereal lover or a nonsweetened cereal lover. At the end of the study, the amount of tooth damage was measured. Here is the summary data:

Group	**N**	**Mean**	**S.D.**
Sugar bombed	10	6.41	5.0
No Sugar	15	5.20	15.0

4. An approximate 95% confidence interval for the difference in the mean tooth damage is:

a) $(6.41 - 5.20) \pm 2.26\sqrt{\frac{5}{10}+\frac{15}{15}}$

b) $(6.41 - 5.20) \pm 2.26\sqrt{\frac{25}{10}+\frac{225}{15}}$

c) $(6.41 - 5.20) \pm 1.96\sqrt{\frac{25}{10}+\frac{225}{15}}$

d) $(6.41 - 5.20) \pm 2.07\sqrt{\frac{146}{10}+\frac{146}{15}}$

e) $(6.41 - 5.20) \pm 1.96\sqrt{\frac{146}{10}+\frac{146}{15}}$

5. In order to test the hypothesis that eating presweetened cereal tends to increase cavities:

a) The P-value is greater than 0.2 and conclude that presweetened cereal tends to increase cavities.
b) The P-value is greater than 0.1 and conclude that presweetened cereal does not tend to increase cavities.
c) The P-value is less than 0.05 and conclude that presweetened cereal tends to increase cavities.
d) The P-value is less than 0.05 and conclude that presweetened cereal does not tend to increase cavities.
e) The P-value is between 0.05 and 0.1 and conclude that presweetened cereal tends to increase cavities.

6. An experiment was done to look at the effect of the tranquilizer, *T*, on people who are depressed. Nine patients were used in a clinical trial. The measurements were taken at the first and second visits after the start of therapy (which included the use of *T*).
Note: Lower measurements on the scale are better.

The data are presented as follows:

Patient	**1st Visit**	**2nd Visit**
1	1.83	.88
2	.5	.65
3	1.62	.6
4	2.48	2.05
5	1.68	1.06
6	1.88	1.29
7	1.55	1.06
8	3.06	3.14
9	1.30	1.29

Using signed rank, a researcher wants to test the hypothesis that the tranquilizer *T* (and the therapy) help alienate depression, using $\alpha = 0.025$. The observed value of the test statistic and the critical value are respectively:

a) 40, 40 b) 40, 5 c) 45, 5 d) 5, 40 e) 5, 45

7. A group of 10 men were given a special diet for two weeks to test for weight loss in pounds. The observed data are:

Man	Weight Before Diet	Weight After Diet
1	181	178
2	171	172
3	190	185
4	187	184
5	210	201
6	202	201
7	166	160
8	173	168
9	183	180
10	184	179

To determine if the data provide sufficient evidence to indicate that the special diet leads to a weight loss, the appropriate test procedure is either:

a) two sample t-test or Wilcoxon Rank Sum test
b) paired t-test or Wilcoxon Signed Rank test
c) paired t-test or Wilcoxon Rank Sum test
d) two sample t-test or Sign test
e) two sample t-test or paired t-test

8. Two presidential candidates, *candidate A* and *candidate B*, were rated by 18 randomly picked voters on a scale of 0 to 5 as follows:

Candidate A:	0	1	5	3	5	4	1	4	3	4	1	4	0	5	4	3	4	5
Candidate B:	4	4	3	0	0	2	3	2	2	0	2	1	0	2	2	1	2	1

The sign test has been used to test the null hypothesis that there is no significant preference for one candidate over the other. The P-value of the test is:

a) 0.0490 b) 0.0245 c) 0.9755 d) 0.0364 e) 0.0182

9. An insurance company issues a policy on a small 'motor bike' under the following conditions: the replacement cost ($6,000) will be paid for a total loss. If it is not a total loss, but the damage is more than $3,000, then $2,500 will be paid. Nothing will be paid for damage costing $3,000 or less, and of course, nothing is paid out if there is no damage. The company estimates the probability of the first three events in a year as 0.05, 0.20 and 0.35 respectively. The amount the company should charge if it wishes to make a profit of $50 above the expected amount paid out in a year is:

a) $500 b) $550 c) $450 d) $800 e) $850

10. A basketball player is successful on 40% of his three-point shots. Each shot can be assumed to be independent of any other shot. If the player takes 150 three-point shots in one season, what is the approximate probability that he makes at least 68 of them?

a) 0.8944 b) 0.5000 c) 0.3944 d) 0.1056 e) 0.6056

Questions 11 and 12 refer to the following:

In a randomized comparative experiment, researchers compared five white rats poisoned with DDT with a control group of seven unpoisoned rats. Electrical measurements of nerve activity are the main clue to the nature of DDT poisoning. When a nerve is stimulated, its electrical impulse shows a sharp spike followed by a much smaller second spike. The experiment found that the second spike is different in rats fed with DDT than in normal rats. The researchers measured the height of the second spike as a percent of the first spike when a nerve in the rat's leg was stimulated.

The data for the DDT group are:

32.3 20.3 10.5 18.6 28.5

The control group data are:

24.6 39.2 18.6 33.2 38.7 25.2 28.5

It is difficult to assess normality from such small samples, so we will use a nonparametric test to assess whether DDT affects nerve response. At 5% level of significance, is there sufficient evidence to indicate that the median spikes are different? Let the value of the rank sum test statistic based on DDT group be T_{DDT}.

11. The 5% rejection region for the test is:

a) $T_{DDT} \leq 34$ or $T_{DDT} \geq 57$

b) $T_{DDT} \leq 33$ or $T_{DDT} \geq 58$

c) $T_{DDT} \leq 21$

d) $T_{DDT} \geq 44$

e) $T_{DDT} \leq 20$ or $T_{DDT} \geq 45$

12. The P-value of the test and conclusion are:

a) P-value = 0.101 and there is evidence that the median spikes are different.
b) P-value = 0.101 and there is evidence that the median spikes are the same.
c) P-value = 0.202 and there is evidence that the median spikes are the same.
d) P-value = 0.899 and there is evidence that the median spikes are the same.
e) P-value = 0.202 and there is evidence that the median spikes are different.

13. A quality control chart has been maintained for a certain measurable characteristic of items taken from a conveyor belt at one point in an assembly line. The measurements are:

68.2 71.6 69.3 71.3 70.3 65.0 63.6 68 64.7

Using the signed-rank test, we wish to test the hypothesis that the median of the population from which these measurements are taken is greater than 68. The values of the test statistic T^+ and the P-value are respectively:

a) 17, 0.5273 b) 36, 0.4727 c) 19, 0.4727 d) 18.5, 0.4727 e) 18.5, 0.5273

14. Water samples from five sites on a river before and two years after an anti-pollution program was started gave the following results. The numbers represent scores for combined pollution measures, high scores indicating greater pollution. The data are presented as follows:

Site:	01	02	03	04	05
Initial Score:	97	65	75	63	90
After Two Years:	92	60	97	43	81

To test whether the anti-pollution program is effective in reducing the pollution, a nonparametric procedure, observed value of the test statistic and the P-value are respectively:

a) Rank-sum procedure, $T_I = 25$, P-value > 0.11
b) Rank-sum procedure, $T_I = 25$, P-value > 0.222
c) Signed-rank procedure, $T^+ = 10$, P-value = 0.3125
d) Signed-rank procedure, $T^+ = 10$, P-value = 0.25
e) Sign test, Y = 4, P-value = 0.3750

15. A manager at a convenience store and gas station tabulates the following facts:

- 60% of all customers purchase gasoline
- 40% of all customers purchase cigarettes
- 24% of all customers buy **both** gasoline **and** cigarettes
- 14% of all customers buy **both** gasoline **and** a slurpee
- 68% of all customers buy **either** gasoline **or** a slurpee

What is the probability that a customer buys neither gasoline nor cigarettes?

a) 0.06 b) 0.24 c) 0.28 d) 0.76 e) 0.54

Questions 16 to 18 refer to the following situation:

Unknown to Yuri, his employer keeps track of his productivity. Productivity is measured in terms of market value (in dollars) of the items produced by Yuri and is recorded on 18 randomly selected days (see below). The employer is interested in whether Yuri's productivity is not the same for all the different days of the week.

Yuri's Productivity

Monday	***Tuesday***	***Wednesday***	***Thursday***	***Friday***
143	162	160	138	110
128	136	132	168	130
110	144	180	120	135
	158	160		
		138		

THE UNIVERSITY OF MANITOBA

Assume that Yuri's productivity on any weekday is normally distributed with the same variance and that his productivity on one day does not influence his productivity on another day. Let μ_M, μ_T, μ_W, μ_R, and μ_F be the mean productivity for Monday, Tuesday, Wednesday, Thursday, and Friday respectively. Using the *JMPintro* output for the data:

Analysis of Variance

Source	*DF*	*Sum of Squares*	*Mean Square*	*F-Ratio*
Model		2517.1111		
Error				
Total		6517.1111		

16. The hypotheses to be tested are:

a) $H_o: \mu_M = \mu_T = \mu_W = \mu_R = \mu_F$ versus $H_a: \mu_M \neq \mu_T \neq \mu_W \neq \mu_R \neq \mu_F$
b) $H_o: \mu_M = \mu_T = \mu_W = \mu_R = \mu_F$ versus H_a: Not all means are equal
c) $H_o: \bar{x}_M = \bar{x}_T = \bar{x}_W = \bar{x}_R = \bar{x}_F$ versus $H_a: \bar{x}_M \neq \bar{x}_T \neq \bar{x}_W \neq \bar{x}_R \neq \bar{x}_F$
d) $H_o: \bar{x}_M = \bar{x}_T = \bar{x}_W = \bar{x}_R = \bar{x}_F$ versus H_a: Not all means are equal
e) $H_o: \mu_M = \mu_T = \mu_W = \mu_R = \mu_F$ versus H_a: The means are all different

17. The estimate of the common population variance is:

a) 307.7 b) 4,000 c) 2517.1 d) 629.3 e) 6517.1

18. The P-value for the testing the hypothesis above is:

a) less than 0.01
b) between 0.01 and 0.025
c) between 0.025 and 0.05
d) more than 0.05
e) more than 2.045

Questions 19 to 22 refer to the following situation:

A health club manager wants to determine whether the number of days per week that college students spend exercising depends on gender. A random sample of 275 college students is selected and the results are classified as shown in the following contingency table.

Notes:
1. Expected values for **some cells** are enclosed in brackets.
2. Cell χ^2 values are given in the third row of **some cells**.

Gender	**Days per week spent exercising**				**Total**
	0 – 1	**2 - 3**	**4 – 5**	**6 - 7**	
Male	40 (?) ***1.20***	53 (55.00)	26 (?) ***0.24***	6 (7.73) ***0.39***	**125**
Female	34 (40.36)	68 (66.00) ***0.06***	37 (?) ***0.20***	11 (9.27)	**150**
Total	**74**	**121**	**63**	**17**	**275**

19. Under the appropriate null hypothesis, the expected frequency for the cell corresponding to **female** and **4 - 5** days per week spent exercising is:

a) 28.64 b) 44.36 c) 34.36 d) 9.27 e) 7.73

20. The critical value of the test at the 0.01 level of significance is:

a) 9.348 b) 9.489 c) 11.345 d) 6.251 e) 7.815

21. The calculated value of the test statistic is:

a) 2.65 b) 7.69 c) 3.48 d) 2.96 e) none of these

22. The null hypothesis to be tested is:

a) The number of days spent exercising per week depends on gender.
b) The number of days spent exercising per week is independent of gender.
c) The proportion of days per week spent exercising associating with gender is different for all college students.
d) The probability of selecting the number of days spent exercising per week is related with the gender.
e) The number of days per week spent exercising for every cell depends on the gender.

23. Suppose flaws (cracks, chips, specks, etc.) occur in the surface of glass with density of three per square metre. What is the probability of there being exactly four flaws on a sheet of glass of area 0.5 square metre?

a) 0.047 b) 0.168 c) 0.981 d) 0.815 e) 0.647

24. The distribution of phone calls arriving in one-minute periods at a switchboard is assumed to be Poisson with a parameter λ. During 100 periods, the following distribution was obtained:

No. of Calls:	0	1	2	3	4 or more
Frequency:	30	43	21	6	0

An estimate for λ based on this data set is:

a) 1.00 b) 1.03 c) 1.04 d) 1.33 e) 1.37

Questions 25 and 26 refer to the following situation:

Nitric oxide is one component of the pollution emitted by automobiles. Two different control devices are to be compared by equipping 10 cars with *device I* and 7 cars with *device II*. The data was analyzed with *JMPintro* and the output follows:

Type	**N**	**Mean**	**Std. Dev.**	**Std. Error**
I	10	1.0160	0.0377	0.0119
II	7	0.9942	0.0350	0.0132

Variances	**T**	**DF**	**Prob > |T|**
Unequal	1.2173	13.7	0.2441
Equal	1.2004	15.0	0.2486

25. We wish to test if the mean level of nitric oxide from *device I* is greater than that of *device II*. The null and alternate hypothesis are:

a) $H_o: \mu_1 - \mu_2 = 0 \quad H_a: \mu_1 - \mu_2 \neq 0$
b) $H_o: \bar{X}_1 - \bar{X}_2 = 0 \quad H_a: \bar{X}_1 - \bar{X}_2 < 0$
c) $H_o: \mu_1 - \mu_2 = 0 \quad H_a: \mu_1 - \mu_2 > 0$
d) $H_o: \bar{X}_1 - \bar{X}_2 = 0 \quad H_a: \bar{X}_1 - \bar{X}_2 > 0$
e) $H_o: \mu_1 - \mu_2 = 0 \quad H_a: \mu_1 - \mu_2 < 0$

26. The rejection region ($\alpha = .05$) and the p-value for testing the hypotheses in question #25 are respectively:

a) reject if $t_{obs} > 1.7709$; p-value = .2441
b) reject if $t_{obs} > 1.7530$; p-value = .2486
c) reject if $t_{obs} > 1.7530$; p-value = .1243
d) reject if $t_{obs} > 1.7709$; p-value = .1220
e) reject if $t_{obs} > 2.1314$; p-value = .1243

27. An SRS of size 100 of a certain popular model car in 2005 found that 20 had a minor defect in the brakes. An SRS of 400 of the same model in 2006 found that 50 had the same minor defect in the brakes. Let p_1 and p_2 denote the true proportions of all cars in the 2005 and 2006 model years, respectively, that actually contain this brake defect. A 90% confidence interval for the true difference in proportions p_1-p_2 is:

a) $-.075 \pm 1.645\sqrt{\frac{(.2)(.8)}{100} + \frac{(.125)(.875)}{400}}$

b) $.075 \pm 1.96\sqrt{\frac{(.2)(.8)}{100} + \frac{(.125)(.875)}{400}}$

c) $.075 \pm 1.645\sqrt{(.14)(.86)\left(\frac{1}{100} + \frac{1}{400}\right)}$

d) $.075 \pm 1.645\sqrt{\frac{(.2)(.8)}{100} + \frac{(.125)(.875)}{400}}$

e) $-.075 \pm 1.96\sqrt{(.14)(.86)\left(\frac{1}{100} + \frac{1}{400}\right)}$

28. An SRS of 38 companies were measured with respect to net sales and net profit. The sample correlation between these variables was r = 0.3. In testing H_0: $\rho = 0$ vs. H_a: $\rho > 0$ at $\alpha = .05$, the value of the test statistic and its approximate p-value are, respectively:

a) 0.3; p-value larger than 0.1
b) 1.887; p-value between .025 and .05
c) 1.887; p-value between 1.633 and 2.028
d) 2.151; p-value between .015 and .025
e) 2.151; p-value between 2.031 and 2.2595

29. A jar containing 120 flies was sprayed with an insecticide of *Brand A* and it was found that 95 of the flies were killed. When another jar containing 145 flies of the same type was sprayed with *Brand B*, 124 flies were killed. Do the two brands differ in their effectiveness? The value of the appropriate test statistic for testing H_0: $p_1 = p_2$ is:

a) $\left(\frac{95}{120}-\frac{124}{145}\right)\Big/\sqrt{\frac{95}{120}\cdot\frac{25}{120}+\frac{124}{145}\cdot\frac{21}{145}}$

b) $\left(\frac{95}{120}-\frac{124}{145}\right)\Big/\sqrt{\frac{95}{120}\cdot\frac{25}{120}+\frac{124}{145}\cdot\frac{21}{145}\cdot\frac{1}{145}}$

c) $\left(\frac{95}{120}-\frac{124}{145}\right)\Big/\sqrt{\frac{219}{265}\cdot\frac{1}{120}+\frac{219}{265}\cdot\frac{1}{145}}$

d) $\left(\frac{95}{120}-\frac{124}{145}\right)\Big/\sqrt{\frac{219}{265}\cdot\frac{46}{265}}$

e) $\left(\frac{95}{120}-\frac{124}{145}\right)\Big/\sqrt{\frac{219}{265}\cdot\frac{46}{265}\left(\frac{1}{120}+\frac{1}{145}\right)}$

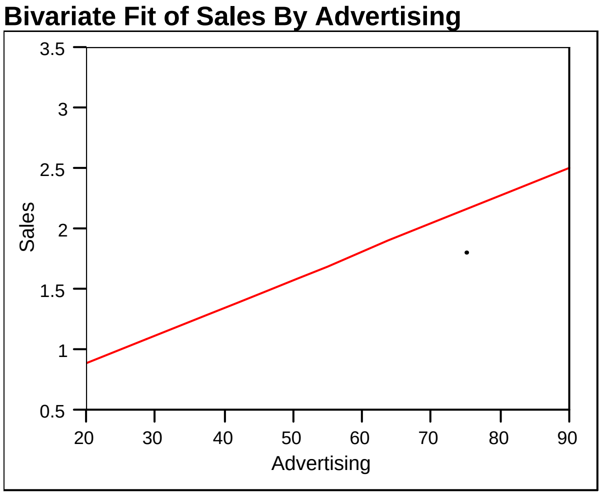
Questions 30 to 32 are based on the following:

The following data refer to the amount spent (in thousands of dollars) on advertising and the volume of sales (in millions of dollars) from 1976 to 1983.

Year	**Advertising X** ***(thousands of $)***	**Sales Y** ***(millions of $)***
1976	30	0.8
1977	35	1.1
1978	50	1.3
1979	40	1.8
1980	60	1.2
1981	80	2.2
1982	60	3.1
1983	75	1.8

For the regression of volume of sales during a year on the amount spent on advertising during that year, a partially completed *JMPintro* output is given below:

Bivariate Fit of Sales By Advertising



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.230	1.230	2.873
Error	6	2.56	0.428	
C. Total	7	3.79		

Parameter Estimates

Term	Estimate	Std Error	t-Ratio	Prob>|t|
Intercept	0.429	0.763	0.56	0.5941
Advertising	0.023	0.01	1.70	0.1410

30. The fitted line is:

a) $\hat{Y} = 0.429 + 0.023\ X$
b) $\hat{Y} = 0.429 + 0.763\ X$
c) $\hat{Y} = 0.023 + 0.014\ X$
d) $\hat{Y} = 0.023 + 0.429\ X$
e) none of the above

31. An approximate 95% confidence interval for the slope is:

a) 0.429 ± 0.763
b) 0.023 ± 0.014
c) 0.023 ± 0.034
d) 0.023 ± 0.024
e) 0.429 ± 1.867

32. It is suspected that sales and the amount spent on advertising are positively correlated. An appropriate null and alternative hypotheses, the test statistic, and the P-value are respectively:

a) $H_0: \rho = 0$ vs. $H_a: \rho \neq 0$, $t_{obs} = 0.56$, $P-\text{value} = 0.5941$
b) $H_0: \rho = 0$ vs. $H_0: \rho \neq 0$, $t_{obs} = 1.70$, $P-\text{value} = 0.1410$
c) $H_o: r = 0$ vs. $H_a: r > 0$, $t_{obs} = 1.70$, $P-\text{value} = 0.2820$
d) $H_0: \rho = 0$ vs. $H_a: \rho > 0$, $t_{obs} = 1.70$, $P-\text{value} = 0.0705$
e) $H_o: r = 0$ vs. $H_a: r > 0$, $t_{obs} = 1.70$, $P-\text{value} = 0.0705$

Selected Formulae for STAT 2000

1. $SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$ with df = smaller of $n_1 - 1$ and $n_2 - 1$

2. $SE(\bar{x}_1 - \bar{x}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ with df = $n_1 + n_2 - 2$ if $\sigma_1^2 = \sigma_2^2$

   where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

3. $\mathrm{SSG} = \sum_{i=1}^{I} n_i\left(\bar{X}_i - \bar{X}\right)^2$

4. Poisson Distribution $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$, k = 0, 1, 2, …

5. $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

6. $s_b = \frac{s_e}{\sqrt{\sum (x_i - \bar{x})^2}}$, $s_e = \sqrt{MSE}$

7. $SE_{\hat{\mu}} = s_e\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x - \bar{x})^2}}$

8. $SE_{\hat{y}} = s_e\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$

9. $SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ if $p_1 = p_2$ where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

   $SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ if $p_1 \neq p_2$

Blank Paper

STAT 2000 Final Examination

Answer Keys

1. e
2. c
3. b
4. b
5. b
6. a
7. b
8. a
9. e
10. d
11. e
12. c
13. d
14. c
15. b
16. b
17. a
18. d
19. c
20. c
21. c
22. b
23. a
24. b
25. c
26. c
27. d
28. b
29. e
30. a
31. d
32. d