THE UNIVERSITY OF SYDNEY

FACULTY OF ARTS & SOCIAL SCIENCES

Business and Economic Statistics A

ECMT1010

Final Exam

CONFIDENTIAL

Family Name:
First Names:
Student ID No.:

June 2012 Time allowed: 120 minutes

Instructions

- Write your family name, first name(s) and SID on the exam paper (above), answer booklet, and also on the computer card provided.
- The whole exam is worth **50 marks**.
- Answer all 30 multiple choice questions in Part A by marking the computer card. Unanswered or incorrect answers are given a mark of zero.
- Answer all parts of both questions Part B in the booklet provided
- Do not take this examination paper from the room.
- Non-programmable calculators are permitted.
- Formulae and statistical tables are provided on pages 10-20.

PART A: MULTIPLE CHOICE QUESTIONS

Answer all 30 multiple choice questions

1. Consider the following stem and leaf plot:

Stem	Leaf
1	0, 2, 5, 7
2	2, 3, 4, 4
3	0, 4, 6, 6, 9
4	5, 8, 8, 9
5	2, 7, 8

Suppose that a frequency distribution was developed from this, and there were 5 classes (10-under 20, 20-under 30, etc.). What is the cumulative frequency for the 30-under 40 class interval?

A. 5
B. 9
C. 13
D. 14

2. Chebyshev's Theorem says that the number of values within 3 standard deviations of the mean will be _______.

A. at least 75%
B. at least 68%
C. at least 95%
D. at least 89%

3. A commuter travels many kilometres to work each morning. She has timed this trip 5 times during the last month. The time (in minutes) required to make this trip was 44, 39, 41, 35 and 41. The mean time required for this trip was 40 minutes. What is the variance for this sample data?

A. 8.8
B. 11
C. 0
D. 3

4. Meagan Davies manages a portfolio of 200 common stocks. Her staff classified the portfolio stocks by 'industry sector' and 'investment objective'.

Investment	Industry Sector			
Objective	Electronics	Airlines	Healthcare	Total
Growth	100	10	40	150
Income	20	20	10	50
Total	120	30	50	200

If a stock is selected randomly from Meagan's portfolio, P(Growth|Healthcare) = _____.

A. 0.25
B. 0.40
C. 0.20
D. 0.80

5. Adam Shapiro, Director of Human Resources, is exploring employee absenteeism at Plain Power Plant. Ten per cent of all plant employees work in the finishing department; 20% of all plant employees are absent excessively; and 7% of all plant employees work in the finishing department and are absent excessively. A plant employee is selected randomly; F is the event ‘works in the finishing department’; and A is the event ‘is absent excessively’. P(A|F) = _____________.
   A. 0.37
   B. 0.70
   C. 0.13
   D. 0.35

6. One hundred policyholders file claims with StressFree Insurance. Ten of these claims are fraudulent. Claims manager Emma Ropati randomly selects four of the ten claims for thorough investigation. If X represents the number of fraudulent claims in Emma’s sample, X has a _______________.
   A. normal distribution
   B. hypergeometric distribution, but may be approximated by a binomial
   C. binomial distribution, but may be approximated by a normal
   D. binomial distribution, but may be approximated by a Poisson

7. On Saturdays, cars arrive at Sandy Schmidt’s Scrub and Shine Car Wash at the rate of 6 cars per fifteen minute intervals. Using the Poisson distribution, the probability that five cars will arrive during the next five minute interval is _____________.
   A. 0.1008
   B. 0.0361
   C. 0.1339
   D. 0.1606

8. Suppose you are working with a data set that is normally distributed with a mean of 400 and a standard deviation of 20. Determine the value of x such that only 1% of the values are greater than x.
   A. 446.6
   B. 353.4
   C. 400.039
   D. 405

9. According to the Australian Department of Industry, Tourism and Resources (DITR), 8.6% of the total employment in NSW is related to manufactured exports. A sample of 200 employees in NSW is randomly selected. If X is the number of employees in the sample with jobs related to manufactured exports, then the standard deviation of X is _______________.
   A. 8.60
   B. 17.20
   C. 15.72
   D. 3.96

10. Financial analyst Ben Taimana needs a sample of securities listed on the New Zealand Stock Exchange. He decides to select the sample from the *NZ Investing Journal*’s Composite Transactions, an alphabetical listing of all securities traded on the previous business day. The *NZ Investing Journal*’s Composite Transactions is ______________.
   A. a poll
   B. a frame
   C. Ben’s sampling distribution
   D. Ben’s target population

11. With ______ random sampling, there is homogeneity within a subgroup or stratum.
    A. judgmental
    B. simple
    C. cluster
    D. stratified

12. Penny Bauer, Chief Financial Officer for Harrison Haulage, suspects irregularities in the payroll system. If 10% of the 5000 payroll vouchers issued since 1 January 2005, have irregularities, the probability that Penny's random sample of 200 vouchers will have a sample proportion .06 and .14 is ___________.
    A. 0.4706
    B. 0.9706
    C. 0.0588
    D. 0.9412

13. Catherine Cho, Director of Marketing Research, needs a sample of Darwin households to participate in the testing of a new toothpaste. If 40% of the households in Darwin prefer the new toothpaste, the probability that Catherine's random sample of 300 households will have a sample proportion between 0.35 and 0.45 is ___________.
    A. 0.9232
    B. 0.0768
    C. 0.4616
    D. 0.0384

14. James Weepu, Human Resources Manager with Auckland First Bank (AFB), is reviewing the employee training programs of AFB branches. His staff randomly selected personnel files for 100 tellers in the Southern Region and determined that their mean training time was 25 hours. Assume that the population standard deviation is 5 hours. The 95% confidence interval for the population mean of training times is ________.
    A. 15.20 to 34.80
    B. 24.18 to 25.82
    C. 24.02 to 25.98
    D. 16.78 to 33.23

15. A random sample of 64 items is selected from a population of 400 items. The sample mean is 200. The population standard deviation is 48. From this data, a 90% confidence interval to estimate the population mean can be calculated as _______.
    A. 189.21 to 210.79
    B. 188.24 to 211.76
    C. 190.13 to 209.87
    D. 190.94 to 209.06

16. A researcher wants to determine the sample size necessary to adequately conduct a study to estimate the population mean to within 5 points. The range of population values is 80 and the researcher plans to use a 90% level of confidence. The sample size should be at least _______.
    A. 44
    B. 62
    C. 216
    D. 692

17. Restaurateur Daniel Valentine is evaluating the feasibility of opening a restaurant in Richmond. The Chamber of Commerce estimates that 'Richmond families, on the average, dine out at least 3 evenings per week'. Daniel plans to test this hypothesis at the 0.01 level of significance. His random sample of 81 Richmond families produced a mean of 2.7. Assuming that the population standard deviation is 0.9 evenings per week, the appropriate decision is ___________.
   A. do not reject the null hypothesis
   B. reject the null hypothesis
   C. reduce the sample size
   D. increase the sample size

18. When the rod shearing process at Newcastle Steel is 'in control' it produces rods with a mean length of 120 cm. Periodically, quality control inspectors select a random sample of 36 rods. If the mean length of sampled rods is too long or too short, the shearing process is shut down. The last sample showed a mean of 120.5 cm. The population standard deviation is 1.2 cm. Using $\alpha = 0.05$, the appropriate decision is __________.
   A. do not reject the null hypothesis and shut down the process
   B. do not reject the null hypothesis and do not shut down the process
   C. reject the null hypothesis and shut down the process
   D. reject the null hypothesis and do not shut down the process

19. When a researcher fails to reject a false null hypothesis, a ______ error has been committed.
   A. Type II error
   B. Type I error
   C. sampling error
   D. powerful error

20. Auckland First Bank's policy requires consistent, standardised training of employees at all branches. Consequently, David Marshall, Human Resources Manager, orders a survey of mean employee training time in the Southern region (population 1) and the Northern region (population 2). His staff randomly selected personnel records of 81 employees from each region, and reported the following: $\bar{x}_1 = 30$ hours and $\bar{x}_2 = 27$ hours. Assume that $\sigma_1 = 6$, and $\sigma_2 = 6$. With a two-tail test and $\alpha = .05$, the appropriate decision is _______________.
   A. reject the null hypothesis $\mu_1 - \mu_2 = 0$
   B. accept the alternate hypothesis $\mu_1 - \mu_2 \leq 0$
   C. reject the null hypothesis $\mu_1 - \mu_2 \neq 0$
   D. do not reject the null hypothesis $\mu_1 - \mu_2 \geq 0$

21. Assume that two independent random samples of size 100 each are taken from a population that has a variance of 36. What is the probability that the difference in the sample means is less than 2?
   A. 0.4909
   B. 0.9909
   C. 0.0091
   D. 0.5091

22. A random sample of 36 items is taken from a population which has a population variance of 144. The resulting sample mean is 45. A random sample of 36 items is taken from a population which has a population variance of 121. The resulting sample mean is 49. Using this information, calculate a 98% confidence interval for the difference in means of these two populations.
   A. –10.99 to 2.99
   B. –10.32 to 2.32
   C. –8.46 to 0.46
   D. –9.32 to 1.32

23. Lucy Baker is analysing demographic characteristics of two television programs, *McLeod's Daughters* (population 1) and *60 Minutes* (population 2). Previous studies indicate no difference in the ages of the two audiences. (The mean age of each audience is the same.) Her staff randomly selected 100 people from each audience, and reported the following: $\bar{x}_1 = 43$ years and $\bar{x}_2 = 45$ years. Assume that $\sigma_1 = 5$ and $\sigma_2 = 8$. Assuming a two-tail test and $\alpha = .05$, the observed z-value is _________________.

A. –2.12
B. –2.25
C. –5.58
D. –15.38

24. In testing a hypothesis about two population means, if the t-distribution is used, we must assume _______.

A. the sample sizes are equal
B. the population means are the same
C. the standard deviations are not the same
D. both populations are normally distributed

25. A researcher wishes to determine the difference in two population means. To do this, she randomly samples 9 items from each population and calculates a 90% confidence interval. The sample from the first population produces a mean of 780 with a standard deviation of 240. The sample from the second population produces a mean of 890 with a standard deviation of 280. Assume that the values are normally distributed in each population. The point estimate for the difference in the means of these two populations is _______.

A. –110
B. 40
C. –40
D. 0

26. Peter Kennedy, a cost accountant at Platypus Plastics Ltd (PPL), is analysing the manufacturing costs of a moulded plastic telephone handset produced by PPL. Peter's independent variable is production lot size (in 1000s of units) and his dependent variable is the total cost of the lot (in $100s). Regression analysis of the data yielded the following tables.

	Coefficients	Standard Error	*t* Statistic	*P*-value
Intercept	3.996	1.161268	3.441065	0.004885
x	0.358	0.102397	3.496205	0.004413

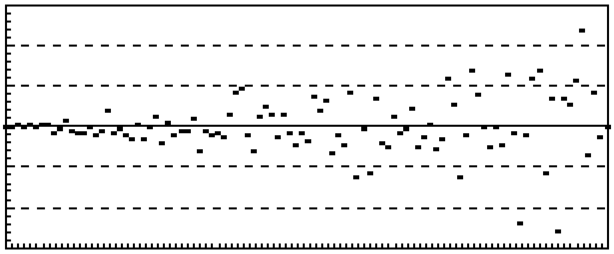
Source	df	SS	MS	*F*
Regression	1	9.858769	9.858769	12.22345
Residual	11	8.872	0.806545	
Total	12	18.73077		

$S_e = 0.898$
$R^2 = 0.526341$

Using $\alpha = 0.05$, Peter should _________________.

A. increase the sample size
B. suspend judgement
C. not reject H_0: $\beta_1 = 0$
D. reject H_0: $\beta_1 = 0$

27. The following residuals plot indicates _______________.



A. a nonlinear relation
B. a nonconstant error variance
C. the simple regression assumptions are met
D. the sample is biased

28. The following data is to be used to construct a regression model:

x	5	7	4	15	12	9
y	8	9	12	26	16	13

The regression equation is _______________.
A. $y = 2.16 + 1.36x$
B. $y = 1.36 + 2.16x$
C. $y = 0.68 + 0.57x$
D. $y = 0.57 + 0.68x$

29. A manager wishes to predict the annual cost (y) of a car based on the number of kilometres (x) driven. The following model was developed: $y = 1550 + .36x$. If a car is driven 30,000 km, the predicted cost is ______________.
A. 10,800
B. 12,350
C. 2630
D. 9250

30. The following data is to be used to construct a regression model:

x	10	9	5	4	8	9
y	2	4	9	9	3	3

The 90% confidence interval for the average value of y at $x = 8$ is ____.
A. 2.18 to 6.56
B. 7.85 to 10.93
C. 3.53 to 5.22
D. 5.76 to 10.51

PART B: SHORT ANSWER QUESTIONS

Answer all parts of both questions in booklet provided

Question 1 (10 MARKS)

The table below provides summary information about students in a class. The sex of each individual and the major is given.

	Male	Female	Total
Accounting	12	18	30
Finance	10	8	18
Other	26	26	52
Total	48	52	100

a. If a student is randomly selected from this group, what is the probability that the student is male? (2 MARKS)

0.48

b. If a student is randomly selected from this group, what is the probability that the student is a female who majors in accounting? (2 MARKS)

0.18

c. A student is randomly selected from this group, and it is found that the student is majoring in finance. What is the probability that the student is a male? (2 MARKS)

0.56

d. A student is randomly selected from this group, and it is found that the student is a male. What is the probability that the student is majoring in accounting? (2 MARKS)

0.25

e. A student is randomly selected from this group. Let A be the event that the student is an accounting major and let F be the event that the student is female. Are A and F independent and why or why not? (2 MARKS)

No, because P(A|F) does not equal P(A)

Question 2 (10 MARKS)

Below is Excel output from a regression analysis to predict *y* from *x*.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.437095
R Square	0.191052
Adjusted R Square	0.137122
Standard Error	5.524306
Observations	17

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	108.113	108.113	3.542602	0.079359
Residual	15	457.7694	30.51796		
Total	16	565.8824			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	48.91031	4.084939	11.97333	4.46E-09
x	-0.20012	0.106325	-1.88218	0.079359

a. What is the equation of the regression model? (2 MARKS)

$\hat{y}$ **= 48.91 – 0.2*x***

b. What is the meaning of the coefficient of *x*? (1 MARK)

A one unit change in *x* leads to a change in *y* of -0.2.

c. What is the result of the test of the slope of the regression model? (α = 0.10) (1 MARK)

The slope of the regression model is significantly different from zero at 10% (but not at 5%).

d. Comment on r^2 and the standard error of the estimate. (2 MARKS)

Approximately 19% of the variance in *y* is explained by the variance in *x*. The standard error of the estimate is 5.52 meaning that approximately 68% of data points are within +/- this distance from the predicted value of *y*.

e. Comment on the relationship of the *f* value to the t ratio (2 MARKS)

$t^2_{\alpha/2,n-2} = F_{\alpha,1,n-2}$. **As there is only one rhs variable this relation holds, i.e. 1.88^2=3.54.**

f. What sample size was used? (1 MARK)

17 observations

g. What is the value of the correlation coefficient? (1 MARK)

$r = \sqrt{r^2} = 0.437$

Formulae

Population mean (ungrouped)

$$\mu = \frac{\Sigma x}{N}$$

Sample mean (ungrouped)

$$\bar{x} = \frac{\Sigma x}{n}$$

Interquartile range

$$\text{IQR} = Q_3 - Q_1$$

Sum of deviations from the arithmetic mean is always zero

$$\Sigma (x - \mu) = 0$$

Population variance (ungrouped)

$$\sigma^2 = \frac{\Sigma (x - \mu)^2}{N}$$

Population standard deviation (ungrouped)

$$\sigma = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}}$$

Chebyshev's theorem

$$1 - \frac{1}{k^2}$$

Sample variance

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1}$$

Sample standard deviation

$$s = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n - 1}}$$

Computational formula for population variance and standard deviation

$$\sigma^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{N}}{N}$$

$$\sigma^2 = \frac{\Sigma x^2 - N\mu^2}{N}$$

Computational formula for sample variance and standard deviation

$$s^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n - 1}$$

$$s^2 = \frac{\Sigma x^2 - n(\bar{x})^2}{n - 1}$$

z score

$$z = \frac{x - \mu}{\sigma}$$

Coefficient of variation

$$\mathrm{CV} = \frac{\sigma}{\mu}(100)$$

Population grouped mean

$$\mu_{\text{grouped}} = \frac{\Sigma f M}{N}$$

Population median (grouped)

$$\text{Median}_{\text{grouped}} = L_M + \left[\frac{\frac{N}{2} - F}{f_M}\right] W$$

Sample median (grouped)

$$\text{Median}_{\text{grouped}} = L_M + \left[\frac{\frac{n}{2} - F}{f_M}\right] W$$

Population first quartile (grouped)

$$Q_1 = L_{q1} + \left[\frac{\frac{N}{4} - F_{q1}}{f_{q1}}\right] W$$

Sample first quartile (grouped)

$$Q_1 = L_{q1} + \left[\frac{\frac{n}{4} - F_{q1}}{f_{q1}}\right] W$$

Population third quartile (grouped)

$$Q_3 = L_{q3} + \left[\frac{\frac{3N}{4} - F_{q3}}{f_{q3}}\right] W$$

Sample third quartile (grouped)

$$Q_3 = L_{q3} + \left[\frac{\frac{3n}{4} - F_{q3}}{f_{q3}}\right] W$$

Population variance (grouped)

$$\sigma^2 = \frac{\Sigma f (M - \mu)^2}{N} = \frac{\Sigma f M^2 - \frac{(\Sigma f M)^2}{N}}{N}$$

Population standard deviation (grouped)

$$\sigma^2 = \frac{\Sigma f (M - \mu)^2}{N} = \frac{\Sigma f M^2 - \frac{(\Sigma f M)^2}{N}}{N}$$

Sample variance (grouped)

$$s^2 = \frac{\Sigma f (M - \overline{x})^2}{n - 1} = \frac{\Sigma f M^2 - \frac{(\Sigma f M)^2}{n}}{n - 1}$$

Sample standard deviation (grouped)

$$s^2 = \sqrt{\frac{\Sigma f (M - \bar{x})^2}{n - 1}} = \sqrt{\frac{\Sigma f M^2 - \frac{(\Sigma f M)^2}{n}}{n - 1}}$$

Pearsonian coefficient of skewness

$$S_k = \frac{3(\mu - M_d)}{\sigma}$$

Pearson's product-moment correlation coefficient

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2}} = \frac{\Sigma xy - \frac{(\Sigma x \Sigma y)}{n}}{\sqrt{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}}$$

Classical method of assigning probabilities

$$P(\mathrm{E}) = \frac{n_e}{N}$$

Range of possible probabilities

$$0 \leq P(E) \leq 1$$

Probability of relative frequency of occurrence

$$P(\mathrm{E}) = \frac{x}{N}$$

Mutually exclusive events X and Y

$$P(X \cap Y) = 0$$

Independent events X and Y

$$P(X|Y) = P(X) \text{ and } P(Y|X) = P(Y)$$

Probability of the complement of A

$$P(\mathrm{A}') = 1 - P(\mathrm{A})$$

Counting rule

$$mn$$

Sampling from a population with replacement

$$N^n$$

Combination formula

$$\binom{N}{n} = \frac{N!}{n!(N - n)!}$$

General law of addition

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$

Special law of addition

$$P(X \cup Y) = P(X) + P(Y)$$

General law of multiplication

$$P(X \cap Y) = P(X)P(Y|X) = P(Y)P(X|Y)$$

Special law of multiplication

$$P(X \cap Y) = P(X)P(Y)$$

Law of conditional probability

$$P\left(X\middle|Y\right) = \frac{P(X \cap Y)}{P(Y)} = \frac{P(X)P(Y|X)}{P(Y)}$$

Bayes' rule

$$P\left(X_i\middle|Y\right) = \frac{P(X_i)P(Y|X_i)}{P(X_1)P(Y|X_1) + P(X_2)P(Y|X_2) + \ldots + P(X_n)P(Y|X_n)}$$

Mean or expected value of a discrete distribution

$$\mu = E(X) = \sum_{all\ x} x p_X(x)$$

Variance of a discrete distribution

$$\sigma^2 = \sum\left[(x-\mu)^2 p(x)\right]$$

Standard deviation of a discrete distribution

$$\sigma = \sqrt{\sum\left[(x-\mu)^2 p(x)\right]}$$

Binomial formula

$$p(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

Mean of a binomial distribution

$$\mu = np$$

Standard deviation of a binomial distribution

$$\sigma = \sqrt{npq}$$

Poisson formula

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Hypergeometric formula

$$p(x) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}}$$

Probability density function of a uniform distribution

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{for all other values} \end{cases}$$

Mean and standard deviation of a uniform distribution

$$\mu = \frac{a+b}{2} \quad \sigma = \frac{b-a}{\sqrt{12}}$$

Probabilities for a uniform distribution

$$P(x_1 < X < x_2) = \frac{x_2 - x_1}{b-a}$$

Probability density function of the normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{1}{2}\right)\left[\frac{x-\mu}{\sigma}\right]^2}$$

Standardisation

$$z = \frac{x - \mu}{\sigma} \quad (\sigma \neq 0)$$

Mean and standard deviation of Bin(n, p) distribution

$$\mu = np \quad \text{and} \quad \sigma = \sqrt{npq}$$

Exponential probability density function

$$f(x) = \lambda e^{-\lambda x}$$

Probabilities of the right tail of the exponential distribution

$$P(X \geq x_0) = e^{-\lambda x_0}$$

Determining the value of k

$$k = \frac{N}{n}$$

z formula for sample means

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

z formula for sample means when there is a finite population

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

Sample proportion

$$\hat{p} = \frac{X}{n}$$

z formula for sample proportions

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

100 (1 − α)% confidence interval to estimate μ

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Confidence interval to estimate μ using the finite population correction factor

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Confidence interval to estimate μ: population standard deviation unknown

$$\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$\text{df} = n - 1$$

Confidential interval to estimate p

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Sample size when estimating μ

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{E} = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$$

Sample size when estimating p

$$n = \frac{z_{\alpha/2}^2 pq}{E^2}$$

z test for a single mean

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Formula to test hypotheses about μ with a finite population

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

t test for μ

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$\text{df} = n - 1$$

z test of a population proportion

$$z = \frac{\bar{p} - p}{\sqrt{\frac{pq}{n}}}$$

z formula for the difference in two sample mean s (independent samples and population variances known)

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Confidence interval to estimate μ1 – μ2

$$(\bar{x}_1 - \bar{x}_2) - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

t formula to test the difference in means assuming $\sigma_1^2 = \sigma_2^2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\text{df} = n_1 + n_2 - 2$$

t formula to test the difference in means

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{df} = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Confidence interval to estimate μ1 − μ2 assuming $\sigma_1^2 = \sigma_2^2$

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\text{df} = n_1 + n_2 - 2$$

t formula to test the difference in two dependent populations

$$t = \frac{\bar{d} - D}{\frac{s_d}{\sqrt{n}}}$$

$$\text{df} = n - 1$$

Formula for $\bar{d}$

$$\bar{d} = \frac{\sum d}{n}$$

Formula for sd

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}} = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n - 1}}$$

Confidence interval formula to estimate the difference in related populations, D

$$\bar{d} - t\frac{s_d}{\sqrt{n}} \le D \le \bar{d} + t\frac{s_d}{\sqrt{n}}$$

$$\text{df} = n - 1$$

z formula for the difference in two population proportions

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

z formula for testing the difference in two population proportions

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{(\overline{pq})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$ and $\bar{q} = 1 - \bar{p}$

Confidence interval to estimate p1 − p2

$$\hat{p}_1 - \hat{p}_2 - z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$
$$\leq (p_1 - p_2) \leq (\hat{p}_1 - \hat{p}_2) + z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Equation of the simple regression line

$$\hat{y} = b_0 + b_1 x$$

Slope of the regression line

$$b_1 = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{\sum xy - n\bar{x}\,\bar{y}}{\sum x^2 - n\bar{x}^2} = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

Sum of squares

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$
$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$
$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n}$$

Alternative formula for slope

$$b_1 = \frac{SS_{xy}}{SS_{xx}}$$

y intercept of the regression line

$$b_0 = \bar{y} - b_1\bar{x} = \frac{\Sigma y}{n} - b_1\frac{\Sigma x}{n}$$

Sum of squares of error (SSE)

$$\text{SSE} = \Sigma\,(y - \hat{y})^2$$

Computational formula for SSE

$$\text{SSE} = \Sigma y^2 - b_0\,\Sigma y - b_1\,\Sigma xy$$

Standard error of the estimate (se)

$$s_e = \sqrt{\frac{\text{SSE}}{n-2}}$$

Coefficient of determination

$$r^2 = 1 - \frac{\text{SSE}}{\text{SS}_{yy}} = 1 - \frac{\text{SSE}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}$$

Computational formula for r2

$$r^2 = \frac{b_1^2 \text{SS}_{xx}}{\text{SS}_{yy}}$$

t test of slope

$$t = \frac{b_1 - \beta_1}{s_b}$$

$$s_b = \frac{s_e}{\sqrt{SS_{xx}}}$$

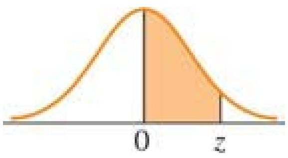
Confidence interval to estimate E(yx) for a given value of x

$$\hat{y} \pm t_{\alpha/2,n-2} s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}}}$$

Prediction interval to estimate y for a given value of x

$$\hat{y} \pm t_{\alpha/2,n-2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SS_{xx}}}$$

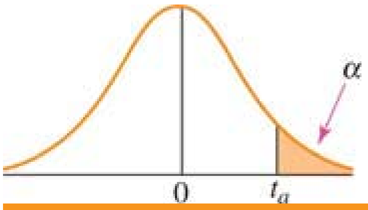
Areas of the standard normal distribution



The entries in this table are the probabilities that a standard normal random variable is between 0 and *z* (the shaded area).

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998									
4.0	.49997									
4.5	.499997									
5.0	.4999997									
6.0	.499999999									

Critical values from the t-distribution



df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
	Values of α for one-tailed test and α/2 for two-tailed test					
1	3.078	6.314	12.706	31.821	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.328
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.894
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
70	1.294	1.667	1.994	2.381	2.648	3.211
80	1.292	1.664	1.990	2.374	2.639	3.195
90	1.291	1.662	1.987	2.368	2.632	3.183
100	1.290	1.660	1.984	2.364	2.626	3.174
150	1.287	1.655	1.976	2.351	2.609	3.145
200	1.286	1.653	1.972	2.345	2.601	3.131
∞	1.282	1.645	1.960	2.326	2.576	3.090