කැලණිය විශ්වවිද්‍යාලය - ශ්‍රී ලංකාව

දුරස්ථ සහ අඛණ්ඩ අධ්‍යාපන කේන්ද්‍රය

වාණිජ හා කළමනාකරණ අධ්‍යයන පීඨය

ව්‍යාපාර කළමනාකරණවේදී (සාමාන්‍ය) උපාධි දෙවන පරීක්ෂණය (බාහිර) - 2019

මැයි - 2023

BMGT E2045 - කළමනාකරණය සඳහා සංඛ්‍යානය

ප්‍රශ්න සංඛ්‍යාව : අටයි (08) කාලය: පැය 03 යි

ඕනෑම ප්‍රශ්න පහකට (05) පිළිතුරු සපයන්න.

ප්‍රශ්න අංක 01.

අ) සංඛ්‍යාන දත්ත සහ සංඛ්‍යානය නිර්වචනය කරන්න.

(ලකුණු 05)

ආ) සංඛ්‍යානය අධ්‍යයනය කිරීමට හේතු දෙකක් ලැයිස්තුගත කර ඒවා පැහැදිලි කරන්න.

(ලකුණු 05)

ඇ) සංඛ්‍යානය භාවිතා කල හැකි අවස්ථා දෙකක් ලැයිස්තුගත කර ඒවා පැහැදිලි කරන්න.

(ලකුණු 05)

ඈ) සංඛ්‍යාන විශේෂඥයින්ට අවශ්‍ය කුසලතා පැහැදිලි කරන්න.

(ලකුණු 05)

(මුළු ලකුණු 20)

ප්‍රශ්න අංක 02.

අ) මිනුම් මට්ටම් හතර නිර්වචනය කර එක් එක් මිනුම් මට්ටම සඳහා උදාහරණය බැගින් දෙන්න.

(ලකුණු 05)

ආ) පරාමිතියක් සහ සංඛ්‍යාන දත්තයක් අතර වෙනස කුමක්ද?

(ලකුණු 05)

ඇ) (i) ප්‍රවර්ග දත්ත, (ii) විවික්ත සංඛ්‍යාත්මක දත්ත සහ (iii) සන්තතික සංඛ්‍යාත්මක දත්ත සඳහා උදාහරණය බැගින් දෙන්න.

(ලකුණු 05)

ඇ) සසම්භාවී නියැදීමේ ක්‍රම තුනක් සහ සසම්භාවී නොවන නියැදීමේ ක්‍රම දෙකක් ලැයිස්තුගත කරන්න.
ඔබ ලැයිස්තුගත කර ඇති ක්‍රම පැහැදිලි කරන්න.

(ලකුණු 05)

(මුළු ලකුණු 20)

ප්‍රශ්න අංක 03.

අ) වඩාත් ඵලදායී තීරු ප්‍රස්ථාර නිර්මාණය කිරීමට ඔබට උපකාර වන මාර්ගෝපදේශ පහක් සඳහන් කරන්න.

(ලකුණු 05)

ආ) පහත දත්ත සමූහය සඳහා මධ්‍යනය, මධ්‍යස්තය සහ මාතය සොයන්න.

GPAs (සිසුන් 10) 1.96, 2.01, 2.25, 2.55, 2.95, 3.02, 3.04, 3.37, 3.51, 3.66

(ලකුණු 05)

ඇ) සහෝදරයන් දෙදෙනෙක් වන සුනිල් සහ නිමල් හට ගෝවා වගා කරන කුඹුරු ඇත. සුනිල් අතින් ගෝවා සිටුවන අතර ගෝවා අතර දුර ප්‍රවේශමෙන් පාලනය කිරීමට නිමල් යන්ත්‍රයක් භාවිතා කරයි. එක් එක් වගාකරුගේ ගෝවා ගෙඩි වල විෂ්කම්භය මනිනු ලැබේ. සුනිල්ගේ ඵලදාවේ ගෝවා ගෙඩි වල අඟල් 2.75 ක සම්මත අපගමනයක් සමඟ අඟල් 7.10 ක සාමාන්‍ය (මධ්‍යනය) විෂ්කම්භයක් ඇති අතර නිමල්ගේ ඵලදාවේ ගෝවා ගෙඩි වල සාමාන්‍ය විෂ්කම්භය අඟල් 6.85ක් වන අතර සම්මත අපගමනය අඟල් 0.60කි.

නිමල් කියා සිටින්නේ තම යන්ත්‍ර භාවිතයෙන් සිටුවීමේ ක්‍රමය වඩා හොඳ බවයි. සුනිල් අවධාරණය කරන්නේ අතින් වගා කිරීම වඩා හොඳ බවයි.

සංඛ්‍යානයේ කේන්ද්‍රීය ප්‍රවණතා මිනුම් සහ අපකිරණ මිනුම් භාවිතා කරමින් ඉහත දක්වා ඇති තත්ත්වය පැහැදිලි කරන්න. තර්කයේ දෙපැත්තම සාධාරණීකරණය කිරීමට හේතුවක් සැපයීම සඳහා දත්ත භාවිතා කරන්න.

(ලකුණු 10)

(මුළු ලකුණු 20)

ප්‍රශ්න අංක 04.

ආ) සම්භාවිතාව තීරණය කිරීමේ ප්‍රධාන ප්‍රවේශ තුන කුමක්ද? ඒවා අතර ඇති වෙනස්කම් පැහැදිලි කරන්න.

(ලකුණු 05)

ආ) $P(A) = .40$, $P(B) = .50$, සහ $P(A \cap B) = .05$ ලෙස ලබා දී ඇත. එසේනම්

i. $P(A \mid B)$ සොයන්න

ii. මෙම ගැටලුවේදී A සහ B ස්වායත්තද? පැහැදිලි කරන්න.

(ලකුණු 05)

ඇ) මෙම සම්භව්‍යතා වගුව ව්‍යාපාර කළමනාකරණ සිසුන් 200කගේ ප්‍රධාන ක්ෂේත්‍ර විස්තර කරයි.

ප්‍රධාන ක්ෂේත්‍ර				
ස්ත්‍රී පුරුෂ භාවය	**ගිණුම්කරණය (A)**	**ආර්ථික විද්‍යාව (E)**	**සංඛ්‍යානය (S)**	**පේළි වල එකතුව**
ස්ත්‍රී (F)	44	30	24	98
පුරුෂ (M)	56	30	16	102
තීරු වල එකතුව	100	60	40	200

පහත සඳහන් එක් එක් සම්භාවිතාව සොයා ඒවා වචන වලින් අර්ථ නිරූපණය කරන්න.

i. $P(A)$

ii. $P(A \cap M)$

iii. $P(F \cap S)$

iv. $P(A/M)$

v. $P(F/S)$ **(ලකුණු 10)**

(මුළු ලකුණු 20)

ප්‍රශ්න අංක 05.

අ) සාමාන්‍යයෙන්, " හෙල්ත් ගාඩ්" රෝහලේ හදිසි කාමර රෝගීන්ගෙන් සියයට 20 කට සෞඛ්‍ය රක්ෂණයක් නොමැත. රෝගීන් හතර දෙනෙකුගේ අහඹු නියැදියකින්,

i. දෙදෙනෙකු රක්ෂණ රහිත වීමේ සම්භාවිතාව කුමක්ද?

ii. රෝගීන් 2කට වඩා අඩු සංඛ්‍යාවක් රක්ෂණය කර තිබීමේ සම්භාවිතාව කුමක්ද?

(ලකුණු 10)

ආ) අධ්‍යයනයකින් හෙළි වූයේ සමාගමක සේවකයින්ගේ සාමාන්‍ය මාසික වැටුප් රු. 160 000 මධ්‍යනයකින් සහ රු. 25 000 සම්මත අපගමනයකින් ප්‍රමතව ව්‍යාප්ත වී ඇති බවයි. පහත ප්‍රශ්නවලට පිළිතුරු දී ඔබේ පිළිතුර අර්ථකථනය කරන්න.

i. සේවකයින්ගෙන් කොපමණ ප්‍රතිශතයක් රු. 100 000 ට වඩා අඩුවෙන් උපයයිද ?

ii. සේවකයින්ගෙන් කොපමණ ප්‍රතිශතයක් රු. 100 000 සහ රු. 150 000 අතර උපයයිද ?

iii. ඉහලම වැටුප් ලබන සේවකයින් 5% අවම වැටුප කුමක්ද?

(ලකුණු 10)

(මුළු ලකුණු 20)

ප්‍රශ්න අංක 06.

Pizza Hut හි හිමිකරුවෙකු තම නිෂ්පාදනයේ මිල සහ වෙළඳ ප්‍රචාරණ පිරිවැය මත පදනම්ව එහි විකුණුම් රටාව ගැන සැලකිලිමත් වේ. ඔහුගේ සති 30ක සමීක්ෂණ ප්‍රතිඵල මෙසේය.

(මෙම දත්ත උපකල්පිත දත්ත බව සලකන්න)

ඔබට එක්සෙල් මගින් ලබා ගත් ප්‍රතිඵලය පහත දී ඇති අතර, මෙම ප්‍රතිඵලය මත පදනම්ව ලබා දී ඇති ප්‍රශ්නවලට පිළිතුරු සපයන්න.

අලෙවි (ඒකක)	මිල රුපියල්	වෙළඳ දැන්වීම පිරිවැය රුපියල්	අලෙවි (ඒකක)	මිල රුපියල්	වෙළඳ දැන්වීම පිරිවැය රුපියල්	අලෙවි (ඒකක)	මිල රුපියල්	වෙළඳ දැන්වීම පිරිවැය රුපියල්
35000	250	330	34000	220	350	34000	310	300
46000	150	320	30000	290	320	43000	300	450
35000	300	300	44000	190	350	36000	275	320
43000	200	450	45000	80	150	38000	250	400
35000	288	320	30000	200	270	42500	150	300
38000	250	400	43000	160	400	47000	140	370
43000	150	300	35000	280	330	35000	300	300
47000	140	370	38000	240	400	34000	220	350
45000	200	350	43000	100	310	30000	290	320
49000	150	400	47000	150	370	44000	190	350

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.826344357
R Square	0.682844997
Adjusted R Square	0.659352034
Standard Error	3347.280027
Observations	30

ANOVA

	df	SS	MS	F	Significance F
Regression	2	6.51E+08	3.26E+08	29.06594	1.84998E-07
Residual	27	3.03E+08	11204284		
Total	29	9.54E+08			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	40829.10051	3929.586	10.39018	6.21E-11	32766.25586	48891.95	32766.26	48891.95
price	-67.05839693	9.19349	-7.29412	7.6E-08	-85.9218807	-48.1949	-85.9219	-48.1949
Advertising cost	38.07995751	10.90152	3.493086	0.001663	15.71188201	60.44803	15.71188	60.44803

i. අනුසීහුම් කරන ලද ප්‍රතිපායන රේඛාවේ සමීකරණය සඳහන් කරන්න.

ii. ඇස්තමේන්තුගත සංගුණක අර්ථකථනය කරන්න.

iii. මෙම ප්‍රතිපායන රේඛාවේ බෑවුම ගැන ඔබේ නිගමනය කුමක්ද?

iv. ද්වි අත් සඳහා නිදහසේ මට්ටම් (DF) සඳහන් කරන්න

v. බෑවුම සඳහා සියයට 95ක විශ්‍රම්භ සීමා අර්ථ දක්වන්න.

vi. R^2 හි අර්ථ නිරූපණය කුමක්ද?

vii. මෙම අවස්ථාව සඳහා සුදුසු වෛකල්පිත කල්පිතයක් ලියන්න.

viii. ඒකකයක මිල රු. 500 ක් නම් සහ වෙළඳ ප්‍රචාරණ පිරිවැය සතියකට රු. 450 ක් නම් සතියක් සඳහා Pizza අලෙවිය පුරෝකථනය කරන්න?

ix. ආකෘතියේ සමස්ත වැදගත්කම තහවුරු කරන්නේ කෙසේද?

x. අනුසීහුම් කරන ලද ප්‍රතිපායන රේඛාව ඔබේ වචන වලින් විස්තර කරන්න.

(එකකට ලකුණු 2 බැගින්)

(මුළු ලකුණු 20)

ප්‍රශ්න අංක 07.

අ) මධ්‍ය සීමා ප්‍රමේය විස්තර කරන්න.

(ලකුණු 05)

ආ) මෝටර් රථ අලෙවිකරුවෙකු පාරිභෝගික තෘප්තිමත්භාවය පිළිබද සමීක්ෂණයක් පවත්වයි. ප්‍රතිචාර දැක්වූවන් 250 දෙනෙකු සඳහා ආන්තික දෝෂය (එනම්, 95% විශ්‍රම්භ මට්ටම හා $\pi = .50$ උපකල්පනය කරමින්) සොයන්න.

(ලකුණු 05)

ඇ) ජංගම දුරකථනයක් හිමි ශ්‍රී ලාංකිකයන් 4,581ක් යොදාගෙන කරන ලද සමීක්ෂණයකින් හෙළි වූයේ සියයට 58ක් තම ජංගම දුරකථන සපයන්නාගේ ආවරණය පිළිබඳව සෑහීමකට පත්වන බවයි.

i. මෙය අහඹු නියැදියක් යැයි උපකල්පනය කරමින්, තෘප්තිමත් ශ්‍රී ලංකා ජංගම දුරකථන හිමිකරුවන්ගේ සත්‍ය අනුපාතය සඳහා සියයට 90 ක විශ්‍රම්භ සීමා ගොඩනගන්න.

ii. විශ්‍රම්භ සීමා මෙතරම් පටු වන්නේ ඇයි?

(ලකුණු 10)

(මුළු ලකුණු 20)

ප්‍රශ්න අංක 08.

අ) i) කල්පිතයක් පරීක්ෂා කිරීමේ පියවර ලැයිස්තුගත කරන්න.

ii) අප්‍රතික්ෂේය කල්පිතය සහ වෛකල්පිත කල්පිතය අතර වෙනස පැහැදිලි කරන්න.

iii) "H_0 පිළිගන්න" වෙනුවට "H_0 ප්‍රතික්ෂේප කිරීමට අසමත්" යැයි අප කියන්නේ ඇයි?

iv) පළමු ප්‍රරූප දෝෂය සහ දෙවන ප්‍රරූප දෝෂය නිර්වචනය කරන්න.

v) වම් අත් පරීක්ෂාව, ද්වි අත් පරීක්ෂාව සහ දකුණු අත් පරීක්ෂාව අතර වෙනස පැහැදිලි කරන්න.

(ලකුණු 10)

ආ) වසර ගණනාවක් පුරා විශ්වවිද්‍යාලයක් සොයාගෙන ඇත්තේ විභාගයට පෙනී සිටි සියලුම සිසුන් අතරින් සමත් වන අනුපාතය .70ක් බවයි. නව විෂය මාලාවකින් සහ ඉගැන්වීමේ ක්‍රමවලින් පසුව, සිසුන් සමත් වීමේ අනුපාතය සැලකිය යුතු ලෙස වෙනස් වී ඇත්දැයි පරීක්ෂා කිරීමට විශ්වවිද්‍යාලයට අවශ්‍ය වී ඇත. විෂයමාලා වෙනස් වීමෙන් පසු සිසුන් 1200 ක් විභාගයට පෙනී සිටි අතර ඉන් 888 ක් සමත් විය. විෂයමාලාව සහ ඉගැන්වීමේ ක්‍රමය සංශෝධනය කිරීමෙන් පසු විභාගය සමත් අනුපාතය වෙනස් වී ඇති බවට ප්‍රකාශ කිරීමට $\alpha = .05$ මට්ටමේ දී ප්‍රමාණවත් සාක්ෂි තිබේද? කල්පිත පරීක්ෂා මගින් ඔබේ පිළිතුර සනාථ කරන්න.

(ලකුණු 10)

(මුළු ලකුණු 20)

FORMULAE

Summary Measures

Sample Mean

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

Sample Standard Deviation

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

Probability Rules

- **Complement rule**

  $P(A^C) = 1 - P(A)$

- **Addition rule**

  General: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

  For independent events:

  $P(A \text{ or } B) = P(A) + P(B) - P(A)P(B)$

  For mutually exclusive events: $P(A \text{ or } B) = P(A) + P(B)$

- **Multiplication rule**

  General: $P(A \text{ and } B) = P(A)P(B \mid A)$

  For independent events: $P(A \text{ and } B) = P(A)P(B)$

  For mutually exclusive events: $P(A \text{ and } B) = 0$

- **Conditional Probability**

  General: $P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)}$

  For independent events: $P(A \mid B) = P(A)$

  For mutually exclusive events: $P(A \mid B) = 0$

Discrete Random Variables

Mean

$$E(X) = \mu = \sum x_i p_i = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$$

Standard Deviation

$$s.d.(X) = \sigma = \sqrt{\sum (x_i - \mu)^2 p_i} = \sqrt{\sum \left(x_i^2 p_i\right) - \mu^2}$$

Binomial Random Variables

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $\dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$

Mean

$$E(X) = \mu_X = np$$

Standard Deviation

$$s.d.(X) = \sigma_X = \sqrt{np(1-p)}$$

Normal Random Variables

- $z - \text{score} = \dfrac{\text{observation} - \text{mean}}{\text{standard deviation}} = \dfrac{x - \mu}{\sigma}$
- Percentile: $x = z\sigma + \mu$
- If X has the $N(\mu, \sigma)$ distribution, then the variable

  $Z = \dfrac{X - \mu}{\sigma}$ has the $N(0,1)$ distribution.

Normal Approximation to the Binomial Distribution

If X has the $B(n, p)$ distribution and the sample size n is large enough (namely $np \geq 10$ and $n(1-p) \geq 10$),

then X is approximately $N\left(np, \sqrt{np(1-p)}\right)$.

Sample Proportions

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

Mean

$$E(\hat{p}) = \mu_{\hat{p}} = p$$

Standard Deviation

$$\text{s.d.}(\hat{p}) = \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Sampling Distribution of $\hat{p}$

If the sample size n is large enough (namely, $np \geq 10$ and $n(1-p) \geq 10$)

then $\hat{p}$ is *approximately* $N\left(p, \sqrt{\dfrac{p(1-p)}{n}}\right)$.

Sample Means

Mean

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

Standard Deviation

$$s.d.(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Sampling Distribution of $\bar{X}$

If X has the $N(\mu, \sigma)$ distribution, then $\bar{X}$ is

$$N(\mu_{\bar{X}}, \sigma_{\bar{X}}) \Leftrightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

If X follows *any* distribution with mean μ and standard deviation σ and n is large,

then $\bar{X}$ is *approximately* $N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

This last result is **Central Limit Theorem**

Population Proportion

Parameter p

Statistic $\hat{p}$

Standard Error

$$\text{s.e.}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Confidence Interval

$$\hat{p} \pm z^* \text{s.e.}(\hat{p})$$

Conservative Confidence Interval

$$\hat{p} \pm \frac{z^*}{2\sqrt{n}}$$

Large-Sample z-Test

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Sample Size

$$n = \left(\frac{z^*}{2m}\right)^2$$

Two Population Proportions

Parameter $p_1 - p_2$

Statistic $\hat{p}_1 - \hat{p}_2$

Standard Error

$$\text{s.e.}(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Confidence Interval

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \text{s.e.}(\hat{p}_1 - \hat{p}_2)$$

Large-Sample z-Test

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

where $\hat{p} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

Population Mean

Parameter μ

Statistic $\bar{x}$

Standard Error

$$\text{s.e.}(\bar{x}) = \frac{s}{\sqrt{n}}$$

Confidence Interval

$\bar{x} \pm t^* \text{s.e.}(\bar{x})$ df $= n - 1$

Paired Confidence Interval

$\bar{d} \pm t^* \text{s.e.}(\bar{d})$ df $= n - 1$

One-Sample t-Test

$t = \dfrac{\bar{x} - \mu_0}{\text{s.e.}(\bar{x})} = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ df $= n - 1$

Paired t-Test

$t = \dfrac{\bar{d} - 0}{\text{s.e.}(\bar{d})} = \dfrac{\bar{d}}{s_d/\sqrt{n}}$ df $= n - 1$

Two Population Means

General

Parameter $\mu_1 - \mu_2$

Statistic $\bar{x}_1 - \bar{x}_2$

Standard Error

$$\text{s.e.}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Confidence Interval

$(\bar{x}_1 - \bar{x}_2) \pm t^*(\text{s.e.}(\bar{x}_1 - \bar{x}_2))$ df $= \min(n_1 - 1, n_2 - 1)$

Two-Sample t-Test

$t = \dfrac{\bar{x}_1 - \bar{x}_2 - 0}{\text{s.e.}(\bar{x}_1 - \bar{x}_2)} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ df $= \min(n_1 - 1, n_2 - 1)$

Pooled

Parameter $\mu_1 - \mu_2$

Statistic $\bar{x}_1 - \bar{x}_2$

Standard Error

$$\text{pooled s.e.}(\bar{x}_1 - \bar{x}_2) = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

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

Confidence Interval

$(\bar{x}_1 - \bar{x}_2) \pm t^*(\text{pooled s.e.}(\bar{x}_1 - \bar{x}_2))$ df $= n_1 + n_2 - 2$

Pooled Two-Sample t-Test

$t = \dfrac{\bar{x}_1 - \bar{x}_2 - 0}{\text{pooled s.e.}(\bar{x}_1 - \bar{x}_2)} = \dfrac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ df $= n_1 + n_2 - 2$

One-Way ANOVA

$$\text{SS Groups} = \text{SSG} = \sum_{\text{groups}} n_i(\bar{x}_i - \bar{x})^2$$

$$\text{SS Error} = \text{SSE} = \sum_{\text{groups}} (n_i - 1)s_i^2$$

$$\text{SS Total} = \text{SSTO} = \sum_{\text{values}} (x_{ij} - \bar{x})^2$$

$$\text{MS Groups} = \text{MSG} = \frac{\text{SSG}}{k-1}$$

$$\text{MS Error} = \text{MSE} = s_p^2 = \frac{\text{SSE}}{N-k}$$

$$F = \frac{\text{MS Groups}}{\text{MS Error}}$$

Confidence Interval $\bar{x}_i \pm t^* \dfrac{s_p}{\sqrt{n_i}}$ df $= N - k$

ANOVA Table

Source	SS	DF	MS	F
Groups	SS Groups	$k-1$	MS Groups	F
Error	SS Error	$N-k$	MS Error	
Total	SSTO	$N-1$		

Under H_0, the F statistic follows an $F(k-1, N-k)$ distribution.

Regression

Linear Regression Model

Population Version:

Mean: $\mu_Y(x) = E(Y) = \beta_0 + \beta_1 x$

Individual: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

where ε_i is $N(0, \sigma)$

Sample Version:

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

Individual: $y_i = b_0 + b_1 x_i + e_i$

Standard Error of the Sample Slope

$$\text{s.e.}(b_1) = \frac{s}{\sqrt{S_{XX}}} = \frac{s}{\sqrt{\sum (x-\bar{x})^2}}$$

Confidence Interval for β_1

$b_1 \pm t^* \text{s.e.}(b_1)$ df $= n-2$

t-Test for β_1

To test $H_0 : \beta_1 = 0$

$t = \dfrac{b_1 - 0}{\text{s.e.}(b_1)}$ df $= n-2$

or $F = \dfrac{MSREG}{MSE}$ df $= 1, n-2$

Parameter Estimators

$$b_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2} = \frac{\sum (x-\bar{x})y}{\sum (x-\bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Confidence Interval for the Mean Response

$\hat{y} \pm t^* \text{s.e.(fit)}$ df $= n-2$

where $\text{s.e.(fit)} = s\sqrt{\dfrac{1}{n} + \dfrac{(x-\bar{x})^2}{S_{XX}}}$

Residuals

$e = y - \hat{y} =$ observed y − predicted y

Prediction Interval for an Individual Response

$\hat{y} \pm t^* \text{s.e.(pred)}$ df $= n-2$

where $\text{s.e.(pred)} = \sqrt{s^2 + (\text{s.e.(fit)})^2}$

Correlation and its square

$$r = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}$$

$$r^2 = \frac{SSTO - SSE}{SSTO} = \frac{SSREG}{SSTO}$$

where $SSTO = S_{YY} = \sum (y-\bar{y})^2$

Standard Error of the Sample Intercept

$$\text{s.e.}(b_0) = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}}$$

Confidence Interval for β_0

$b_0 \pm t^* \text{s.e.}(b_0)$ df $= n-2$

Estimate of σ

$$s = \sqrt{MSE} = \sqrt{\frac{SSE}{n-2}} \quad \text{where } SSE = \sum (y-\hat{y})^2 = \sum e^2$$

t-Test for β_0

To test $H_0 : \beta_0 = 0$

$t = \dfrac{b_0 - 0}{\text{s.e.}(b_0)}$ df $= n-2$

Chi-Square Tests

Test of Independence & Test of Homogeneity

Expected Count

$$E = \text{expected} = \frac{\text{row total} \times \text{column total}}{\text{total } n}$$

Test Statistic

$$X^2 = \sum \frac{(O-E)^2}{E} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

df $= (r-1)(c-1)$

Test for Goodness of Fit

Expected Count

$E_i = \text{expected} = np_{i0}$

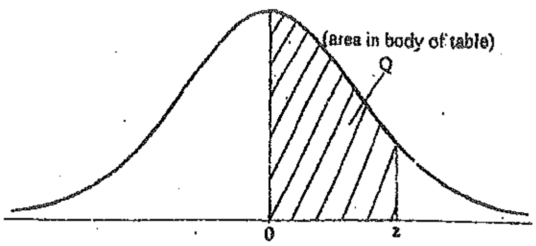
Test Statistic

$$X^2 = \sum \frac{(O-E)^2}{E} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

df $= k-1$

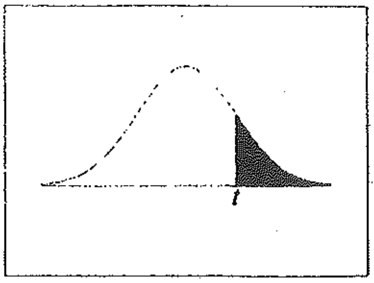
If Y follows a $\chi^2(df)$ distribution, then $E(Y) = df$ and $\text{Var}(Y) = 2(df)$.

Table I AREAS UNDER THE STANDARD NORMAL CURVE



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

t-Distribution Table



The shaded area is equal to α for $t = t_\alpha$.

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
32	1.309	1.694	2.037	2.449	2.738
34	1.307	1.691	2.032	2.441	2.728
36	1.306	1.688	2.028	2.434	2.719
38	1.304	1.686	2.024	2.429	2.712
∞	1.282	1.645	1.960	2.326	2.576