平成２２年度(2010 年度)長野県数学会夏期研修会

データ分析における統計手法の基礎理論と実践

松本大学総合経営学部総合経営学科教授　林　昌孝　先生

平成 22 年 8 月 10 日　長野県総合教育センター

【講義1】 ツールとしての統計手法の基礎理論の講義

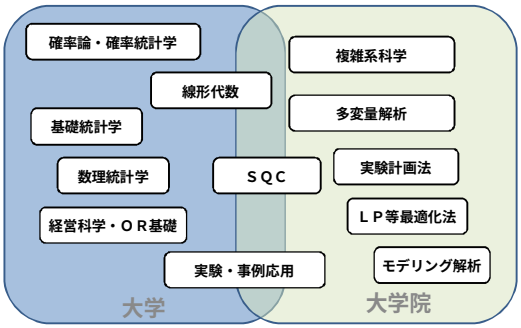
- 統計学のカリキュラム上の配置

1. ツールとして活用する学部･学科
   - ⇒ 経済･経営･工学部他
   - ⇒ 専門基礎の科目配置 (1～2年)
   - ⇒ 本日の内容！
2. 統計学を研究する学部･学科
   - ⇒ 理学部数理学科, 工学部計数工学
     情報システム工学など

- 松本大学のカリキュラム上の統計関連科目配置

学部	学科	1年	2年	3年･4年
総合経営学部	総合経営学科	情報処理ⅡⅢ	ﾃﾞｰﾀ分析Ⅰ	基礎統計学 社会調査論 ﾃﾞｰﾀ分析Ⅱ ﾏｰｹﾃｨﾝｸﾞ関連科目
	観光ﾎｽﾋﾟﾀﾘﾃｨ学科	情報処理ⅡⅢ		基礎統計学
人間健康学部	健康栄養学科	基礎統計学 情報処理ⅡⅢ		栄養統計学
	ｽﾎﾟｰﾂ健康学科	基礎統計学 情報処理ⅡⅢ		ﾏｰｹﾃｨﾝｸﾞ関連科目

- 高校以後の統計学分野



- 松本大学総合経営学部･人間健康学部
  基礎統計学のシラバス

目 的:「データ処理への興味」
「数学への苦手意識回避」

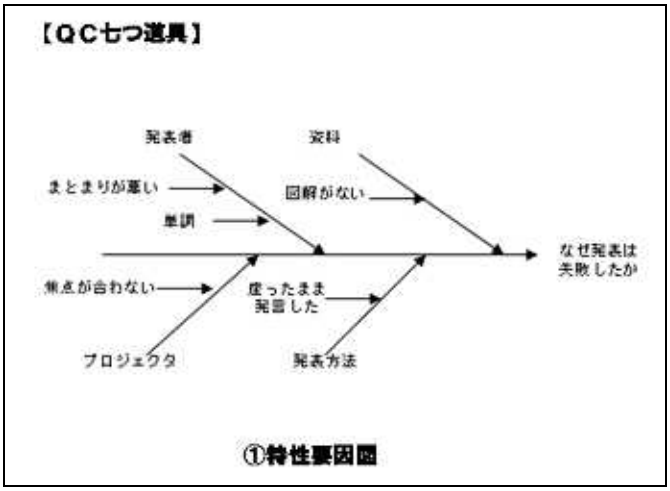
内 容: 1. データの整理方法 (記述統計)
QC7つ道具の紹介
2. データから母集団を推理する(推測統計)
正規分布 と t分布 が中心

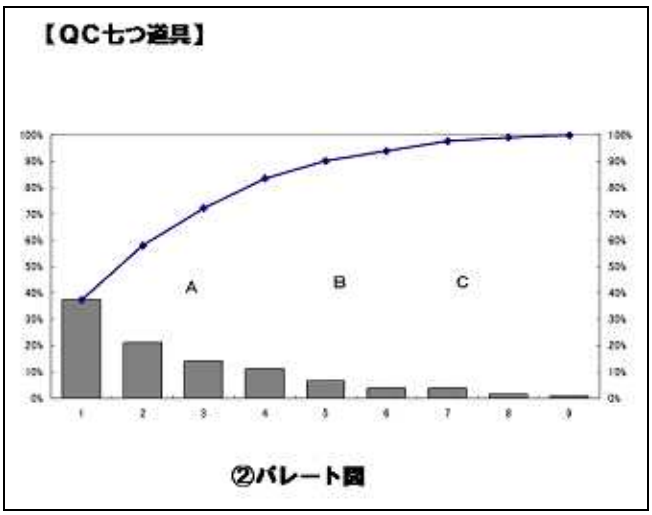
- 基礎統計学のシラバス(授業計画)

第1回	オリエンテーション　尺度とデータ	第9回	信頼係数と有意水準
第2回	量的データの整理	第10回	母集団と標本をめぐる定理
第3回	質的データの整理	第11回	母平均の推定
第4回	データを代表する値	第12回	t 分布を用いた母平均の推定
第5回	平均と標準偏差	第13回	仮説検定　基本手順と z 分布を用いた例
第6回	統計量の計算練習	第14回	t 分布を用いた仮説検定
第7回	正規分布とその他の分布	第15回	まとめ
第8回	数値表の使い方	第16回	試験

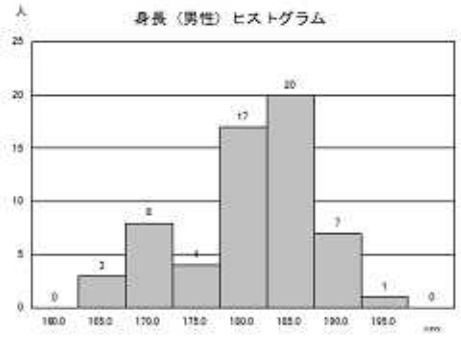
- QC7つ道具(データの整理方法の代表)

手法	概要
①特性要因図	特性（結果）と特性に影響をおよぼしている要因（原因）との関連を体系的に整理した図。問題点を要因別に分析できる図形の形状から「魚の骨」とも呼ばれる。
②パレート図	データを項目別に分類して集計し、数値の大きいものから順に並べた棒グラフとその累積比率を折れ線グラフに、1つの図で表したもの。パレート図を使った分析手法のひとつにＡＢＣ分析がある。
③ヒストグラム	データを複数の区間に分けて、各区間に入るデータの度数を表したグラフ。データの分布状態を把握するのに用いる。
④散布図	2種類のデータをＸ軸・Ｙ軸にプロットしたグラフで、2種類のデータの関連性を把握するのに用いる。
⑤チェックシート	データの収集や確認漏れを防ぐ目的で、チェック項目を分類した表や図のこと。
⑥層別	収集したデータや調査結果などを関連する項目に分類すること。通常は分けた項目ごとに他のＱＣ七つ道具を利用して問題の分析をおこなう。
⑦管理図	品質や肯定を管理する目的で作成する折れ線グラフで、品質に関するデータが特定の範囲内に収まっているかどうかで問題が発生しているかどうかを判断する。



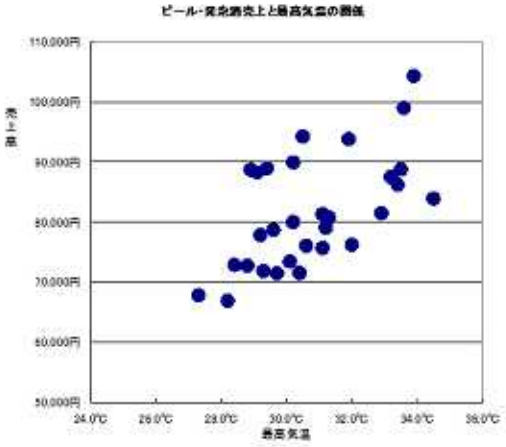


【QC七つ道具】



③ヒストグラム

【QC七つ道具】



④散布図

➢ クロス集計

あるコンビニの売上データ（POS）のレコード例

購入品目コード	数量	価格	顧客性別	顧客年齢	購入時間

性別と購入品目の2項目について集計した表(例)

性別	A品目	B品目	C品目	D品目
男	60	337	134	61
女	31	236	106	42

➢ 基礎統計学で伝えたいこと＝論理性

（統計的検定の重要性）

自然科学では、

理論の証明方法として

① 演繹的方法○　② 帰納的方法○

→ 真・理論

社会科学では、

演繹的な理論構築は難しい

最適化×（不可能）　満足化△　最善化△

学説を過去にさかのぼり、調べまくり、全てを網羅するような議論が必要とされる

→ 学説・理論

論理的に証明できる唯一の方法＝仮説検定！

➢ 目標とする問題（1）

単3乾電池を製造している会社がある。従来から、単3乾電池の寿命は 180 時間と公表している。今、この会社が新しい製造技術を導入し、その結果、新製品の寿命は 180 時間より伸びたと考えられる。

そこで、このことを確かめるために

「新製品の寿命は 180 時間より長い」

という仮説をたてて検証したい。

そのために標本を 20 本抽出して、寿命を試験した結果、平均＝198 時間、標本標準偏差 s ＝15 時間であった。

この仮説は正しいのだろうか？

「答えは、危険率5%で有意に長くなった。つまり、仮説は正しい。」

➢ 目標とする問題（2）

あるコンビニで顧客データを調べたところ、男女間で購入していった品物に差がありそうなことが判明した。そこで、顧客属性データから、性別と購入品別の購入数を引き出して、購入品と性別に差があるか調べることにした。

その結果は、以下のとおりである。

性別	A品目	B品目	C品目	D品目
男	60	337	134	61
女	31	236	106	42

男女間によって購入品目に差があると判断できるのか？

「答えは、有意水準（危険率）5%で男女間に差がない。」

➢ 数理統計学の分野では、

・平均値・分散・標準偏差の定義と計算

E[X]　V[X]　SD[X]

・分布と確率変数

・離散型の確率分布　と　連続型の確率分布

（二項分布・ポワソン分布）（正規分布・指数分布）

・正規分布・t分布・カイ2乗分布・F分布

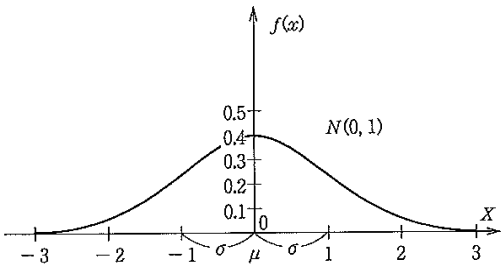
と展開する

➢ 連続型の確率分布の代表 = 正規分布

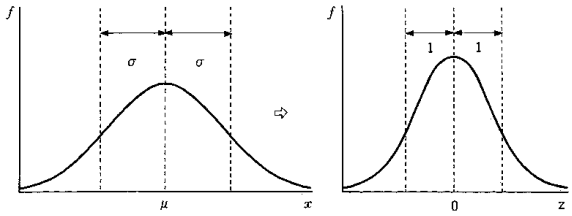
連続的な確率変数 X が確率密度関数

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

をもつとき，X は正規分布 $N(\mu,\sigma^2)$ に従うという。

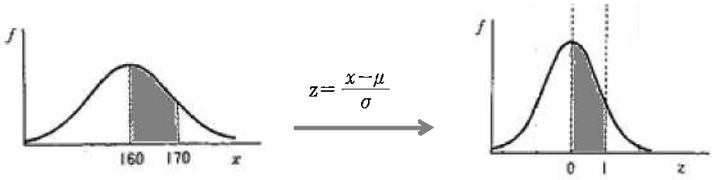


➢ 正規分布における確率の計算では、
N（μ，σ²)で計算せずに、N（0，1²)の標準正規分布に変換して計算する。



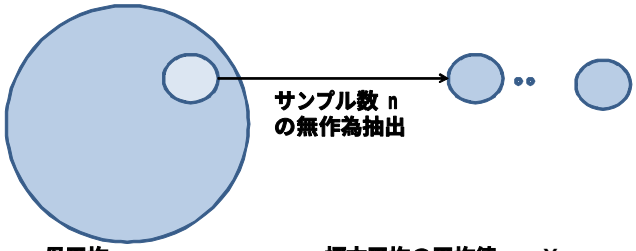
標準化(変数変換)　$z=\frac{x-\mu}{\sigma}$

➢ （具体例として）
16～20 歳の女性の身長は，N（160,10²）の正規分布をする。この分布の上で身長が 160cm 以上 170cm 以下である確率はいくらであろうか。言い換えれば、P（160≦X≦170）を求めよということである。



P（160≦X≦170）＝P（0≦Z≦1）
となるから、
Z＝1.00 のときの黒塗面積を正規分布表より求めて 0.34134 を得る。
したがって、
P（160≦X≦170）＝P（0≦Z≦1）＝0.34134
となる。

➢ 母集団　　標本(ｻﾝﾌﾟﾙ)



母平均　μ
母標準偏差　σ

標本平均の平均値　X
標本平均の標準偏差　s

両者の関係　X＝μ
s＝σ/√n

➢ 母平均　μ
母標準偏差　σ

標本平均　X
標本標準偏差　s

両者の関係　μ＝X
σ＝s/√n

母集団が正規分布に従う場合
N（μ，σ²）　→ Xは正規分布に従う
母集団が正規分布に従わない
→ nが大きくなるにしたがって
Xは正規分布に従う
（中心極限定理）

➢ 母平均μを推定する場合、
① σが分かっているとき（既知）は、
Xは、N（X，（σ/√n）²)に従う
② σが分かっていないとき（未知）は、
Xは、n を変数(自由度）に持つt分布に従う。
t分布は、N分布よりなだらかな分布で、
n→∞ で t分布＝N分布 となる分布である。

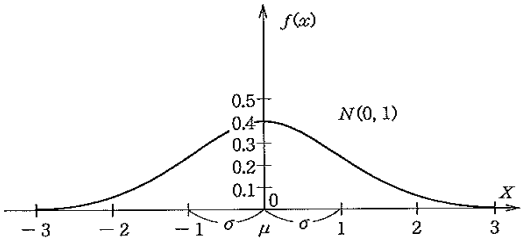
以下、中心的な分布について紹介する。

➢ 正規分布

連続的な確率変数 X が確率密度関数

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

をもつとき，X は正規分布 $N(\mu,\sigma^2)$ に従うという。



➢ t分布

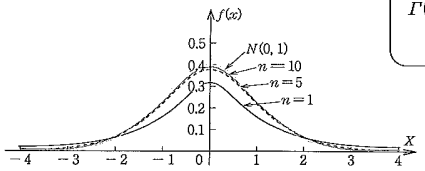
連続的な確率変数 X が確率密度関数

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (n=1,2,3,\cdots)$$

をもつとき，X は自由度 n の t 分布に従うという。

Γ 関数

$$\Gamma(p)=\int_0^\infty x^{p-1}e^{-x}dx \quad (p>0)$$



➢ 正規分布は、平均値 μ の 推定･検定に利用される。

σ が既知の場合 → 正規分布を使用

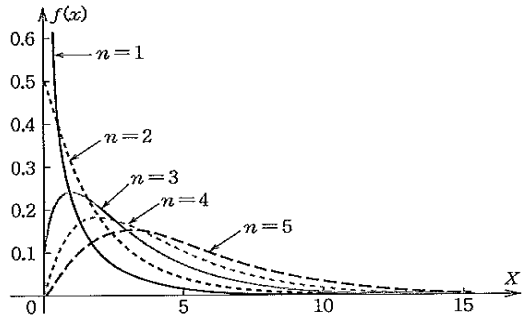
σ が未知の場合 → t分布を使用 → データ数(n)の増加 → t分布と正規分布は一致

➢ χ^2 分布

連続的な確率変数 X が確率密度関数

$$f(x)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}} \quad (x>0\,;\ n=1,2,3,\cdots)$$

をもつとき，X は自由度 n の χ^2 分布に従うという。



➢ 確率変数 $X_1,\cdots,X_n$ が互いに独立で，すべて $N(0,1)$ に従うとき

$$X_1^2+\cdots+X_n^2$$

は自由度 n の χ^2 分布に従う。

t 分布と χ^2 分布の関係:

X, Y は互いに独立な確率変数で，それぞれ $N(0,1)$ と自由度 1 の χ^2 分布に従うとき，$T=\dfrac{X}{\sqrt{\frac{Y}{n}}}$ は自由度 n の t 分布に従う。

分散（ばらつき）の推定・検定に利用される。

➢ 主な定理

【定理１】
確率変数 X が自由度nのカイ二乗分布に従うとき，
E［X］＝n，　V［X］＝ 2nである。

《 説明 》
nが大き＜なれば平均，分散ともに大き＜なり，確率密度関数 f(x)のグラフの山が平た＜なってい＜ことを示している。

【定理２】
確率変数 X が N(0,1) に従うとき，確率変数 X^2 は自由度 1 の カイ二乗 分布に従う。

《 説明 》 0 < a < b とするとき、

$$P(a<X^2\leqq b)=P(-\sqrt{b}<X\leqq-\sqrt{a})+P(\sqrt{a}<X\leqq\sqrt{b})$$
$$=\int_{-\sqrt{b}}^{-\sqrt{a}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx+\int_{\sqrt{a}}^{\sqrt{b}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$$

被積分関数は偶関数なので第 1 項，第 2 項とも同じ値となり，$x=y^2$ と変数変換すると

$$=\frac{1}{\sqrt{2\pi}}\int_a^b\frac{1}{\sqrt{y}}e^{-\frac{y}{2}}dy$$
$$=\int_a^b\frac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)}y^{\frac{1}{2}-1}e^{-\frac{y}{2}}dy$$

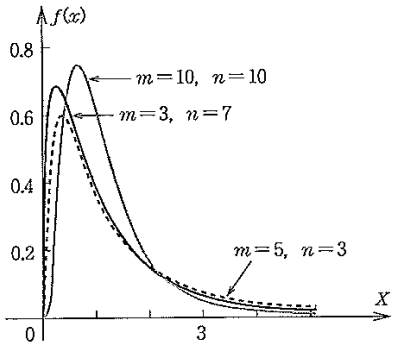
この被積分関数は自由度1のカイ二乗分布の確率密度関数である。これより、X^2 は自由度 1 のカイ二乗分布に従うことが示せる。

➢ F 分布

連続的な確率変数 X が，確率密度関数

$$f(x)=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{m}{n}\right)^{\frac{m}{2}}x^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}$$
$$(x>0\,;\ m,n=1,2,3,\cdots)$$

をもつとき，X は自由度 (m,n) の F 分布 $F(m,n)$ に従うという。



➢ 確率変数 X，Y が互いに独立で、それぞれ自由度（ m ，n)のχ²分布に従っているとき，

$$F=\frac{\frac{X}{m}}{\frac{Y}{n}}$$

は、自由度（ m ， n)のF 分布に従う。

この性質によりF 分布は 2 組のデータの分散の比較に利用される(等分散の検定)。

➢ 目標とする問題(1)

単3乾電池を製造している会社がある。従来から、単3乾電池の寿命は 180 時間と公表している。今、この会社が新しい製造技術を導入し、その結果、新製品の寿命は 180 時間より伸びたと考えられる。

そこで、このことを確かめるために

「新製品の寿命は 180 時間より長い」

という仮説をたてて検証したい。

そのために標本を 20 本抽出して、寿命を試験した結果、平均＝198 時間、標本標準偏差 s ＝15 時間であった。

この仮説は正しいのだろうか？

解法の基本原理(t分布を用いた仮説検定)

① 仮説「新製品の寿命は180時間である」とする仮説をたてる。

帰無仮説： μ＝180

対立仮説： μ＞180

② 製品の寿命(母平均)を有意水準5%(信頼度95%)で推定して、 その推定値に標本観測値である、198時間が入っていれば仮説は正しいと判断する。

母平均の推定

母平均 μ は、標本の平均値(X)の分布が従う、t分布(平均値＝180　標本標準偏差＝15　自由度＝20－1＝19　有意水準5%)

t＝(X－μ)/(s/√n)の変数変換より、

母平均を推定すると、

X＝μ＋t×(s/√n)

＝180＋1.729×15÷(√20)＝185.79‥となる。

つまり、母平均は 185.79 以下が有意水準 5%の推定値となっている。

③ 標本データは、198 時間であるから推定値内からは外れる (推定値外の領域を棄却域という)

したがって、帰無仮説(μ=180)は棄却されて、対立仮説が採択される。つまり、「寿命は 180 時間より長くなった」が、結論となる。

➢ 目標とする問題(2)

あるコンビニで顧客データを調べたところ、男女間で購入していった品物に差がありそうなことが判明した。そこで、顧客属性データから、性別と購入品別の購入数を引き出して、購入品と性別に差があるか調べることにした。

その結果は、以下のとおりである。

性別	A品目	B品目	C品目	D品目
男	60	337	134	61
女	31	236	106	42

男女間によって購入品目に差があると判断できるのか？

解法の基本原理(カイ二乗分布を用いたバラツキ(分散)の検定)

① 男女グループとカテゴリー(品目)との関係は期待値に比べてバラツキが大きい場合には、差があると判断して、バラツキが小さい場合には差がないと判断する。

② 元データ

性別	A品目	B品目	C品目	D品目	合計
男	60	337	134	61	592
女	31	236	106	42	415
総計	91	573	240	103	1007

男女に差がないとして期待される期待値(エクセル上で計算)

性別	A品目	B品目	C品目	D品目	合計
男	53.4	336.8	141.0	60.5	592
女	37.5	236.1	98.9	42.4	415
総計	91	573	240	103	1007

この比率で品目別総計を比例配分する。

＝91×(592/1007)

＝91×(415/1007)

③ カイ二乗の値を計算する。(ピアソンの近似値法)

男女別品目ごとに

(元データのセルの値 － セルの期待値)の2乗÷セルの期待値

を計算して、総合計を求める。

男・A品目＝(60-53)^2/53=0.79

性別	A品目	B品目	C品目	D品目	合計
男	0.79	0.00	0.35	0.00	1.15
女	1.12	0.00	0.50	0.00	1.64
総計	1.91	0.00	0.86	0.00	2.79

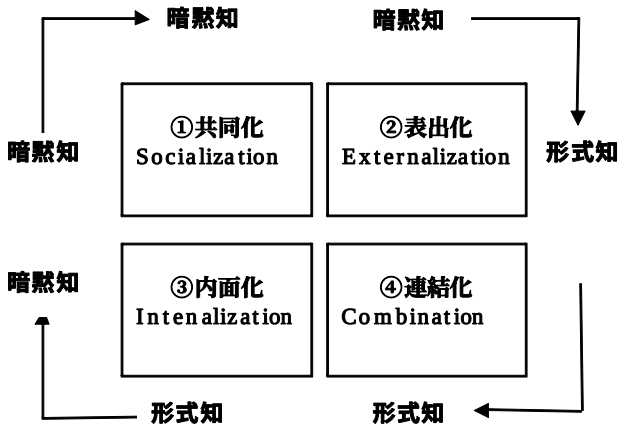
自由度＝（グループ数ー1）×（カテゴリ数ー1）
＝（2-1）（4-1）=3
有意水準5％のカイ二乗の値は、7.815　となる。
（エクセルでは、=chiinv(0.05,3)=7.815）

④総合計の　2.79　は、グループ間に差があるとされる7.815より小さいため、　このデータは、ままあること・偶然に生じるバラツキであったと判断される。つまり、有意水準 5%で「男女間では、差がない」が結論となる。

➢　**（研究レポートの中での応用事例）**
ビジネスゲームにおけるナレッジ・マネジメント
（2005/1　松本大学研究紀要）

ナレッジ・マネジメント
（1995 野中ら知識経営の生みの親）
・知的作業の生産性向上を支援する技術
・個人が経験し蓄積した知識(個人知)を組織内で共有して組織知(企業知)として活用して、社員の能力・スキルアップ・組織の総合力を高める手法

知識創造理論
(1) 知識：「形式知」　明確な言語・数字・図表で表現される
「暗黙知」　メンタル・モデル（信念や世界観など）やコツやノウハウ
(2) 人間の創造的活動において、両者は互いに作用し合い、形式知は暗黙知へ、暗黙知は形式知へ互いに成り変わる。
(3) 組織の知は、異なったタイプの知識（暗黙知と形式知）そして異なった内容の知識を持った個人が相互に作用し合うことによって創られる。



ナレッジ・マネジメントの基礎理論（SECIモデル）

①「共同化」：個人の暗黙知（思い）を共通体験をつうじて**互いに共感**し合う
②「表出化」：その暗黙知から明示的な言葉や図で表現された**形式知**として**コンセプト**などを創造する
③「連結化」：形式知と形式知を組合せて**体系的な形式知**を創造する
④「内面化」：**実体験**を通じてその体系的な形式知を身に付け**暗黙知として体化**する

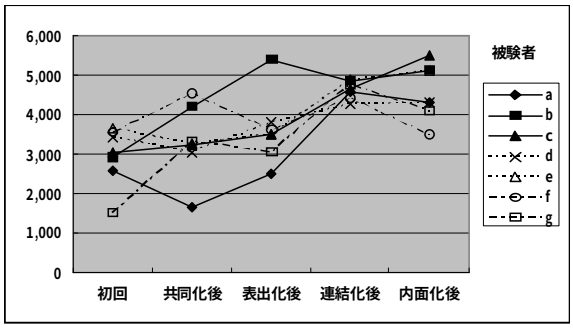
組織の知は、この四つのモードをめぐるダイナミックな**スパイラル**によって創られる。**

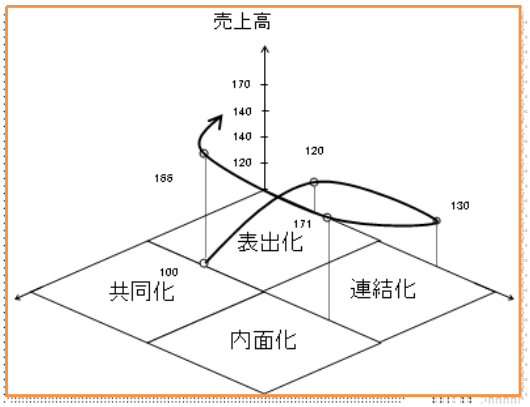
図に表して表示（研究目的）

コンビニ経営のネットワークシミレーション教育システム（ザ・コンビニ for フォレスト）を用いて、実験・考察を行った。
＜ザ・コンビニ for フォレスト＞とは、
・コンビニ経営のネットワークシミレーション教育システム
・どこでどんなものを売ればいいか、商品を売る立場、買い物をする立場で考えながら、互いに売り上げを競いながらのシミュレーションゲーム
・変化する季節、経過する時間の中でその都度判断することが必要で、ひとりひとりの知識や経験を総動員して臨むゲームです
・小学校（社会科）の「買い物の体験」、中学校（公民）の「商品と価格」用に開発

〔実験結果（売上高）〕

被験者名	初回	共同化後	表出化後	連結化後	内面化後
a	2,602	1,664	2,502	4,585	4,301
b	2,900	4,199	5,399	4,857	5,145
c	3,049	3,235	3,519	4,650	5,522
d	3,438	3,048	3,817	4,279	4,290
e	3,670	3,251	3,522	4,866	5,098
f	3,559	4,526	3,615	4,422	3,502
g	1,492	3,320	3,070	4,771	4,101
平均	2,959	3,320	3,635	4,633	4,566





知識創造過程のスパイラル（初回を 100）

共同化後の平均値の差の検定（t検定）では、連結化・内面化で5％有意となった。

➢ (学生のアンケート調査での応用事例)

学生のリゾートに対するイメージと
連泊に対する意識調査

(2009 卒業研究　松本大学総合経営学部観光ホスピタリティ学科)

研究目的:

近年のリゾートホテルを利用する若年旅行者の減少が続いている。

特に、

①若者の旅行の条件は安さだけではない

②過去の旅行先が、旅行のイメージに大いに関係する

などの点について、アンケート調査をもとに整理検討して、若年旅行者の増加対策を考察する。

アンケートの概要:

・これまでの旅行経験 (行先・宿泊数・修学旅行内容など)

・リゾートのイメージ調査 (連想する言葉、イメージ、場所、色など)

・連泊で重視する点・考え方

以上、記入式・選択式を混ぜて20数項目のアンケート調査

アンケート調査の規模:

対象:松本大学総合経営学部所属の学生　1～3年

配布数:586枚

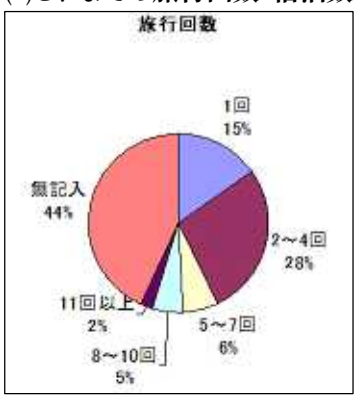
回答数:359枚 (回答率　約61.3%)

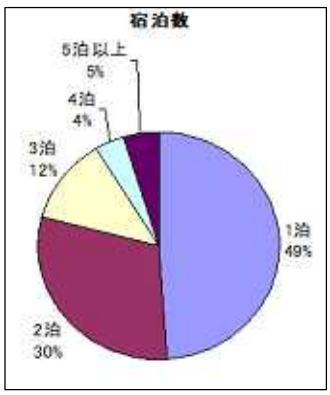
内訳　男子:248人 (69.1%)

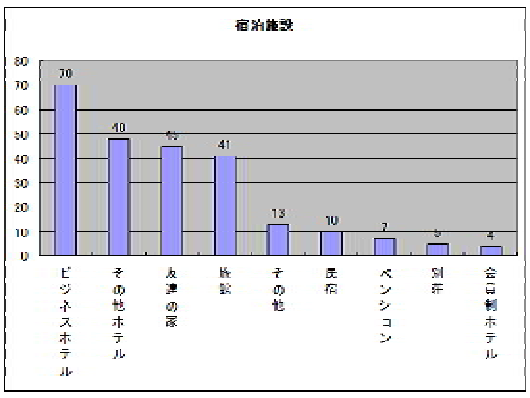
女子:111人 (30.9%)

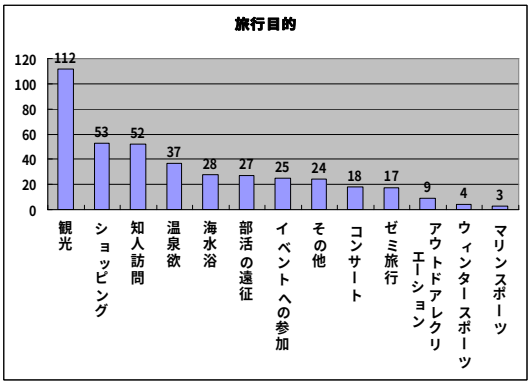
調査結果 (抜粋)

(1)これまでの旅行回数・宿泊数など









(2)リゾートときいてイメージする観光地

イメージする観光地・多かった意見			
ハワイ	182	北海道	11
沖縄	81	ディズニーリゾート	11
グアム	71	サイパン	10
軽井沢	49	伊豆	8
ドバイ	15	タヒチ	7
ラスベガス	12	海	7
熱海	12	星野リゾート 池の平らホテル	4

(3)リゾートから連想・イメージするもの

リゾートのイメージ					
海・砂浜	182	温泉	18	別荘	4
豪華	70	プール	15	観光	5
南国	56	楽しい	15	飛行機	3
のんびり	46	美しい	10	島	2
外国	42	ディズニー	7	ロマンティック	2
ホテル	40	食	6	人ごみ	2
自然(山・森・川)	25	冬	6	危険・拉致	2
自然(太陽・空)	16	ショッピング	5	その他	10

(4)リゾートからイメージする色

イメージする色					
青	210	赤	6	緑茶色	1
水色	35	エメラルド	5	青緑	1
白	34	ピンク	3	トロピカル	1
緑	22	黄緑	2	灰色	1
オレンジ	14	スカイブルー	2	肌色	1
金	13	オーシャンブルー	2	茜色	1
レインボー	11	無色・透明	2	銀	1

(5) リゾートと聞いてイメージする観光地と
修学旅行先 (沖縄) の関係性 (検定例の一例として)

	(観光地としてどこをイメージするか?)	
	沖縄をイメージする	沖縄以外をイメージする
A:沖縄に修学旅行に行ったひと	60 (人)	162 (人)
B:沖縄以外に修学旅行に行ったひと	22 (人)	115 (人)

・A/B のグループ間には差がある
（5%有意・カイ 2 乗検定）
・つまり、沖縄に修学旅行に行った人の方が沖縄をよりリゾートとして強いイメージをもつようである。

（6）リゾートにする若者のイメージは?
（検定済みのもの）
①「山」ではなく「海」のイメージ
②「高級」「リッチ」の認識
③名前に左右されやすいリゾート意識
④過去の経験や記憶がイメージ構築に大きく関与

提言

①平日学生向けプランの作成
②豪華なイメージをおさえた広報・告知
③海以外のリゾートの特徴を強調するイメージ構築が必要
④学生が選べる内容のプランを増やす

【講義 2】 データ分析の実際としての講義と演習（参加者各自が PC を使用）
統計解析の演習として、スーパー銭湯の調査データ（Excel ファイル）を使用し、次のことを演習する。

① クロス集計表を作成し、グラフ化する。
② χ^2 検定を行う。

① 元のデータは、ある 1 日のスーパー銭湯の来客調査の一覧である。(図1)
このデータをクロス集計する。
（1） ピボットテーブルの作成
分単位の来店時間を時間単位に修正しておく。
(a) 来店時間の列を選択し、セルの書式設定の表示形式
→ユーザー定義 種類に h"時" と入力する。
(b) 挿入→ピボットテーブル→ピボットテーブルの作成
→新規ワークシート を確認し O.K.
(c) Sheet1 に入り、ピボットテーブルのフィールドリストの
NO, 来店時間,性別にチェックを入れる。
(d) 行ラベルを来店時間、列ラベルを性別、Σ 値を NO と
する。
(e) 値フィールドの設定でデータの個数を選択する。
作成したピボットテーブルを別の場所にコピーし、整理する。(図2)
（2） (図2) のデータから、男女別の時間帯別の積み上げ棒グラフを
作成する。
データ範囲を選択後、挿入→縦棒→積み上げ縦棒 を選択する。
グラフが完成する。(図3)

no	来店時間	性別	年齢	種類
1	6時00分	1	5	1
2	6時00分	1	5	1
3	6時00分	1	4	1
4	6時00分	1	5	1
5	6時00分	2	5	1
6	6時00分	2	4	1
7	6時00分	2	4	1
8	6時00分	2	4	2
9	6時00分	1	1	2
10	6時00分	1	4	2
11	6時00分	2	5	2
12	6時00分	1	4	1
13	6時00分	2	4	1
14	6時00分	2	4	1
15	6時00分	2	4	1
16	6時05分	1	5	1
17	6時05分	1	4	3
18	6時05分	1	4	3
⋮	⋮	⋮	⋮	⋮
1512	23時40分	2	4	1
1513	23時40分	2	3	3
1514	23時40分	2	3	3
1515	23時45分	1	3	3
1516	23時45分	2	3	3

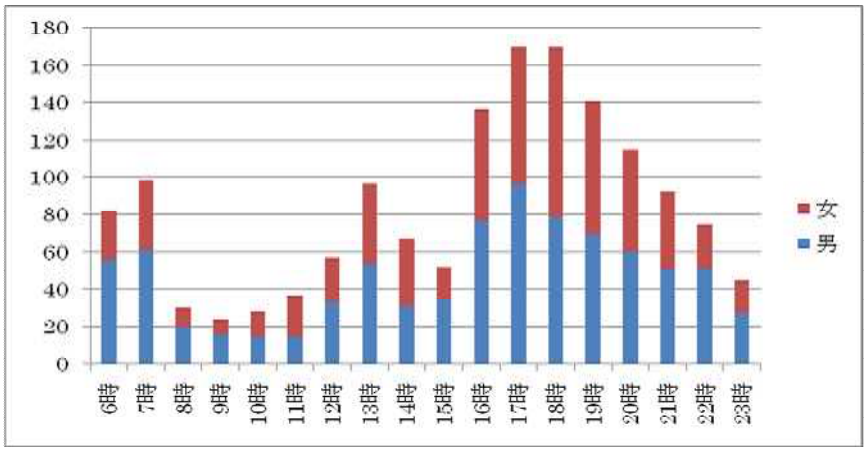
（図１）

カテゴリーの内容

性別	
男	1
女	2

年齢	
幼児(0-5)	1
未成年(6-19)	2
青年(20-35)	3
中年(35-59)	4
老人(60-)	5

種類	
一人	1
家族	2
グループ	3

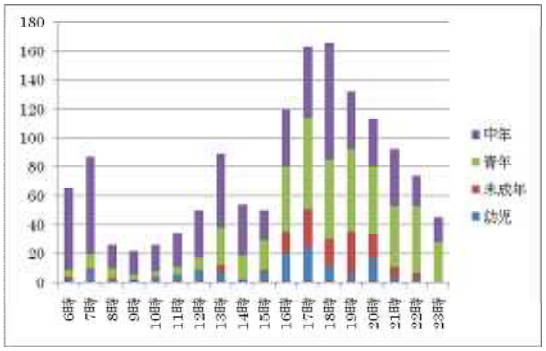


（図３）

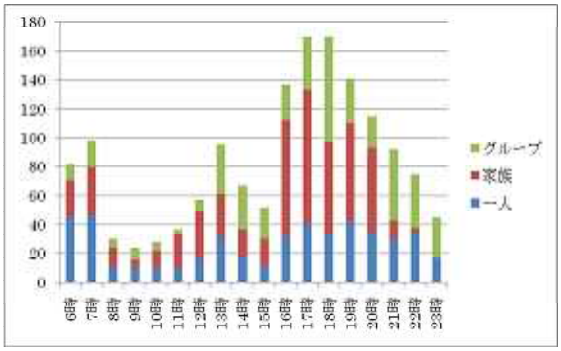
時間	男	女
6時	56	26
7時	61	37
8時	20	10
9時	16	8
10時	15	13
11時	15	22
12時	33	24
13時	54	42
14時	31	36
15時	34	18
16時	77	60
17時	96	74
18時	79	91
19時	70	71
20時	60	55
21時	51	41
22時	51	24
23時	27	18

（図２）

同様にして、年齢別の時間帯別の積み上げ棒グラフ（図4）、種類別の時間帯別の積み上げ棒グラフ（図5）を作成する。



（図４）



（図５）

② 時間帯によって男女間に差があるか検定する。

帰無仮説は、男女のグループ間には差がない。
対立仮説は、グループ間には差がある。

⑴ 簡単にするために、午前と午後の入りに男女間の差があるか検討する。
データを、6 時から 3 時までを午前、4 時から 23 時までを午後として集計する。（図6）

観測値

時間	午前	午後	合計
男	335	511	846
女	236	434	670
合計	571	945	1516

理論値の計算

時間	午前	午後	合計
男	319	527	846
女	252	418	670
合計	571	945	1516

観測値のＸ二乗値の計算

時間	午前	午後	合計
男	0.839	0.507	1.347
女	1.060	0.640	1.700
合計	1.899	1.148	3.047

（図６）

理論値の計算
（例）午前の男
576×846÷1516=319

χ^2値の計算
（例）午前の男
$(319-571)^2 \div 319=0.839$

χ^2値の合計が 3.047
この値が、CHIINV(有意水準、自由度)の中ならば、帰無仮説を採択し、
この値が、CHIINV(有意水準、自由度)の外ならば、対立仮説を採択する。

CHIINV(0.05,1)＝3.841459 ＞ 3.047 となって、棄却域にはないので、有意差はなし。
帰無仮説が採択される。
つまり、男女グループ間には、午前と午後の比率の差はないといえる。

(2)　発展させ、時間帯によって男女間の差があるか検討する。（図7）

観測値

時間	6時	7時	8時	9時	10時		21時	22時	23時	合計
男	56	61	20	16	15		51	51	27	846
女	26	37	10	8	13	・・・	41	24	18	670
合計	82	98	30	24	28		92	75	45	1516

期待値

時間	6時	7時	8時	9時	10時		21時	22時	23時	合計
男	46	55	17	13	16		51	42	25	846
女	36	43	13	11	12	・・・	41	33	20	670
合計	82	98	30	24	28		92	75	45	1516

観測値のX二乗の計算

時間	6時	7時	8時	9時	10時		21時	22時	23時	合計
男	2.292	0.728	0.634	0.507	0.025		0.002	1.999	0.142	13.771
女	2.893	0.920	0.801	0.641	0.032	・・・	0.003	2.524	0.179	17.389
合計	5.185	1.648	1.435	1.148	0.057		0.005	4.523	0.321	31.160

（図７）

自由度＝（グループ数－1）（カテゴリー数－1）　＝　1×17　＝　17

χ^2値＝CHIINV(0.05,17)　＝　27.58711　＜　31.160

となり、データのＸ二乗値の観測地は、棄却域にあるから、有意差がある。対立仮説が採択される。
つまり、男女間のグループには差がある（時間帯の比率は異なる）といえる。