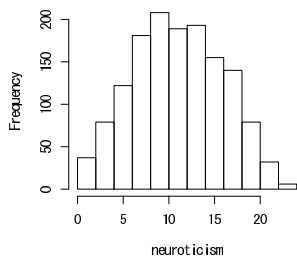


各変数の分布

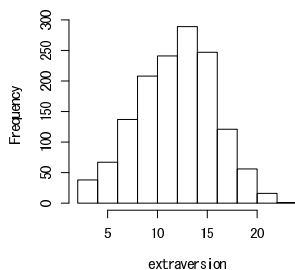
分析をする上で、まずやらなければならないのは、個々の変数の分布を見ることである。そこで次の枠内のコードにより、量的な変数についてはヒストグラムと正規確率プロットを作成し、要因型の変数についてはカテゴリ別度数を得る。

```
it12ans-2007.R(1)
library(car)
attach(Cowles)
layout(matrix(1:4,2,2))
hist(neuroticism)
qqnorm(neuroticism)
hist(extraversion)
qqnorm(extraversion)
layout(1)
table(sex)
table(volunteer)
```

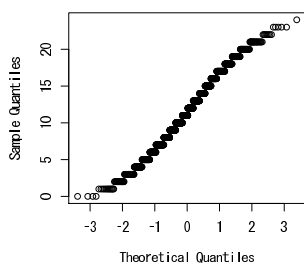
Histogram of neuroticism



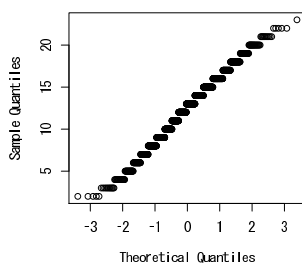
Histogram of extraversion



Normal Q-Q Plot



Normal Q-Q Plot

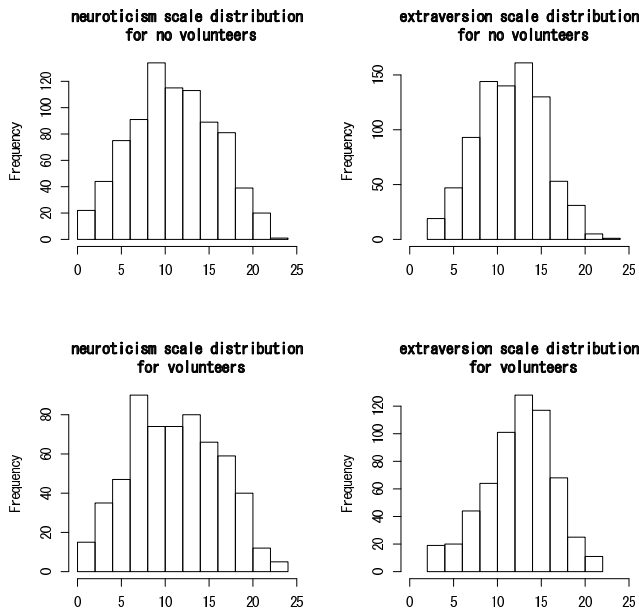


図をみるかぎり、神経症尺度も外向性尺度も、ほぼ正規分布しているように見える。性別は女性 780 名、男性 641 名と女性の方がやや多いが極端な差はない。ボランティアとなるかどうかは、no が 824 人、yes が 597 人であり、no がやや多いが yes の回答者数も十分多い。従ってカテゴリの併合や変数変換は必要ない。

個別の独立変数と従属変数の関係

次に、独立変数 1 つずつ、従属変数との関係を見るため、次の枠内のコードを入力する。

```
it12ans-2007.R(2)
layout(matrix(1:4,2,2))
hist(subset(neuroticism,volunteer=="no"),xlim=c(0,25),main="neuroticism scale distribution\n for no volunteers",xlab="")
hist(subset(neuroticism,volunteer=="yes"),xlim=c(0,25),main="neuroticism scale distribution\n for volunteers",xlab="")
hist(subset(extraversion,volunteer=="no"),xlim=c(0,25),main="extraversion scale distribution\n for no volunteers",xlab="")
hist(subset(extraversion,volunteer=="yes"),xlim=c(0,25),main="extraversion scale distribution\n for volunteers",xlab="")
layout(1)
wilcox.test(neuroticism~volunteer)
wilcox.test(extraversion~volunteer)
print(SV <- table(sex,volunteer))
fisher.test(SV)
```



神経症尺度の分布はボランティアになるかどうかで差がないようにみえるが、外向性尺度の分布はボランティアになる人の方が高い方に偏っているようにみえる。ウィルコクソンの順位和検定の結果でも、神経症尺度にはボランティアになるかどうかで有意差がない ($p = 0.73$) が、外向性尺度には有意差があった ($p < 0.001$)。性別とボランティアになるかどうかをクロス集計した結果は次の通り。

性別 \	ボランティアになるか	
	いいえ	はい
女性	431	349
男性	393	248

フィッシャーの正確な確率は 0.023 となり、ボランティアになるかどうかと性別には 5% 水準で有意な関連がある。ボランティアにならないことの女性の男性に対するオッズ比は 0.78 (95% 信頼区間 [0.62,0.97]) であり、女性は男性よりもボランティアになりやすいと考えられた。

ロジスティック回帰分析

以上の解析より、ボランティアになるかどうかは神経症尺度とは関係がなく、外向性尺度及び性別と関係があるので、次のコードにより、これら 2 つを独立変数とし、ボランティアになるかどうかを従属変数とするロジスティック回帰分析を行う。念のため 3 つ全部を独立変数としたモデルも立てて尤度比検定もしておく。

```
it12ans-2007.R(3)
res <- glm(volunteer ~ sex + extraversion, family=binomial)
summary(res)
exp(coef(res))
exp(confint(res))
resx <- glm(volunteer ~ ., data=Cowles, family=binomial)
print(lambda <- -2*(logLik(res)-logLik(resx)))
1-pchisq(lambda,1)
detach(Cowles)
```

ロジスティック回帰分析の結果は下表のようにまとめられ、心理学研究のボランティアになることには、外向性が高いことと性別が女性であることが寄与しており、外向性の影響を調整しても男性は女性の 0.78 倍しかボランティアを引き受けないことがわかった*1。なお、尤度比検定の結果、有意確率は $p = 0.575$ であり独立変数 3 つのモデルと独立変数 2 つのモデルの間に最大尤度の有意差は認められないので、独立変数 2 つのモデルで十分である。

表. 心理学研究のボランティアを引き受ける要因の分析

独立変数	オッズ比	95% 信頼区間		p 値
		下限	上限	
男性 (女性)	0.78	0.63	0.97	0.024

Nagelkerke の R^2 : 0.025, AIC: 1912.4, D_{null} : 1933.5 (自由度 1420), D : 1906.4 (自由度 11418)

カッコ内はリファレンスカテゴリ。外向性尺度を共変量として調整した (係数 0.0655, $p < 0.001$)。

*1 ただし、実は外向性の影響を考慮しないでクロス集計により求めたオッズ比とほとんど同じ値である。このことからわかるのは、ボランティアになるにあたって、外向性の影響と性別の影響はほぼ独立しているということである