

統計学II

早稲田大学政治経済学術院

西郷 浩

本日の目標

- 誤差項に課せられた条件の検証
 - 残差プロットによる検証
- 実例
 - 1: コンピュータの修理時間
 - x : 修理を要する部品の個数
 - y : 修理に要した時間
 - 2: X線放射によるバクテリア生存数
 - x : 200キロボルトのX線の放射時間(6分単位)
 - y : バクテリア生存数

実例データの引用先

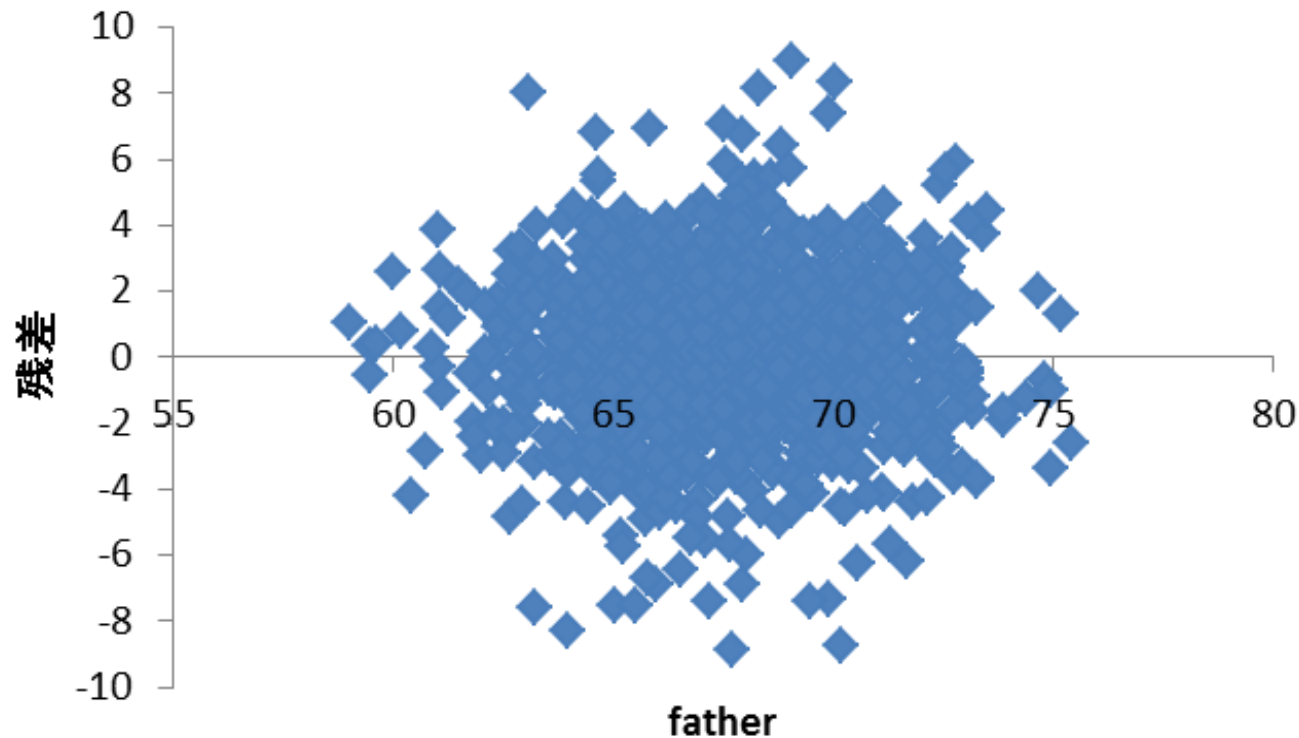
- S. チャタジー、B. プライス（佐和隆光・加納悟訳）『回帰分析の実際』新曜社 1981年
 - － 実例1:16ページ
 - － 実例2:35ページ

残差プロットによる検証(1)

- 誤差項に課せられた条件の検証
 - 諸仮定
 - 期待値が0である: $E(u_i) = 0$
 - 分散が一定である: $V(u_i) = \sigma^2$
 - 相互に関係をもたない: $Cov(u_i, u_j) = E(u_i u_j) = 0$
 - 正規分布にしたがう。
 - 説明変数 x と無関係である。
- 残差
 - 誤差項の代用
 - 誤差項(観察不能) $u_i = Y_i - \beta_0 - \beta_1 x_i$
 - 残差(観察可能) $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

残差プロットによる検証(2)

図1: 残差 vs father



資料:
Pearson の
データ

残差プロットによる検証(3)

図2: 残差のヒストグラム

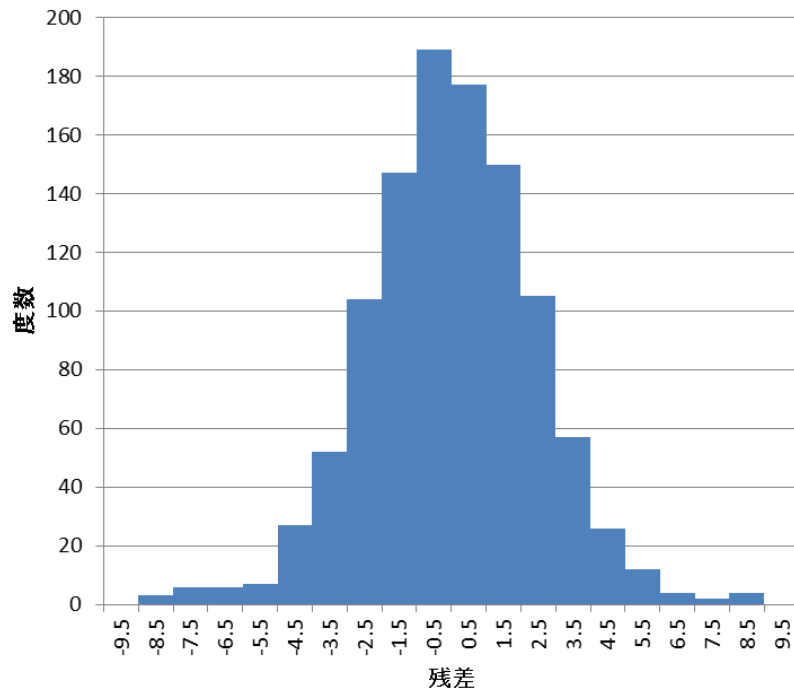
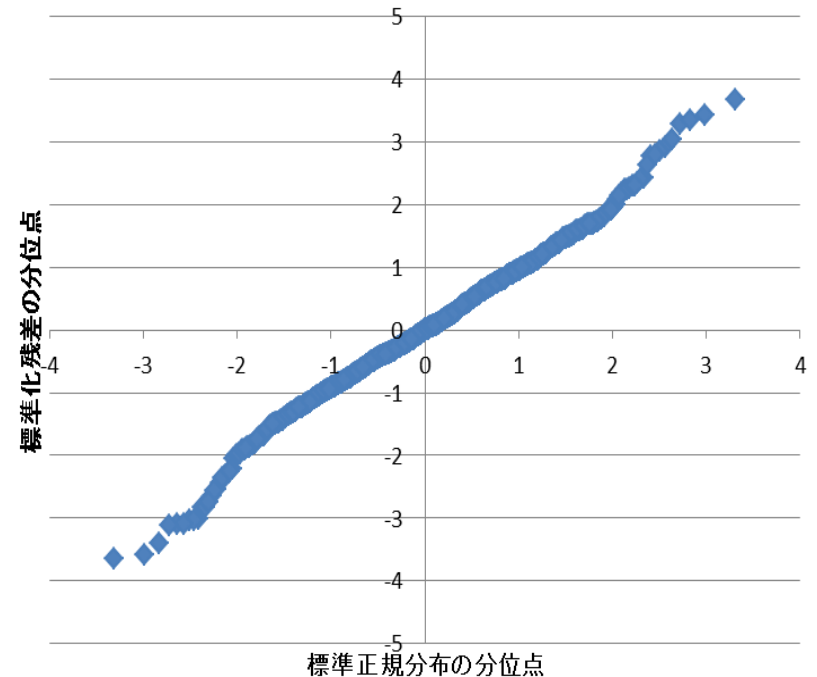


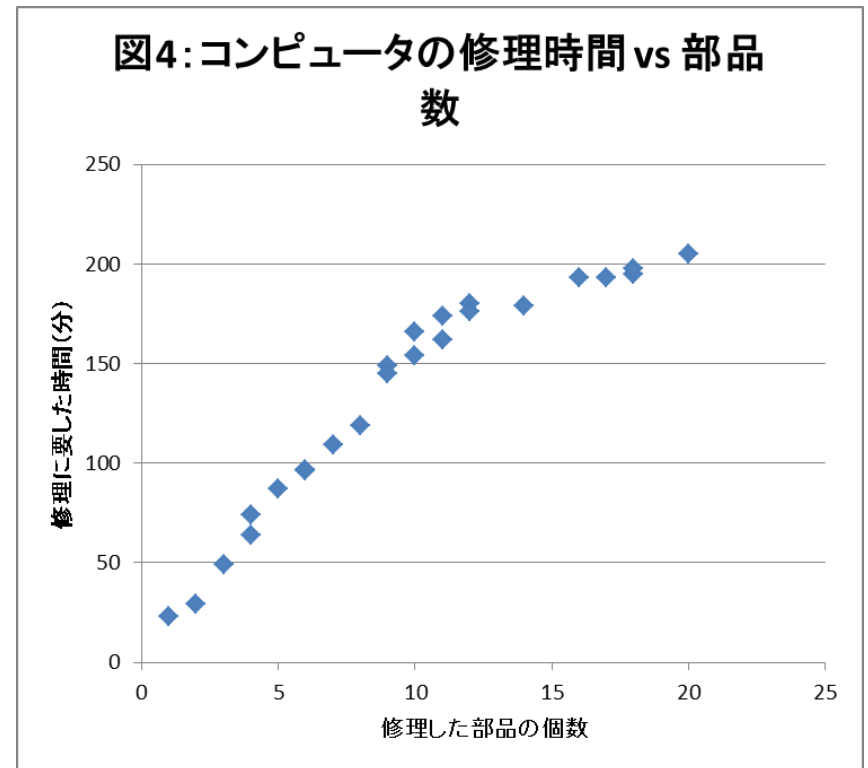
図3: 標準化残差の正規QQプロット



資料: Pearson のデータ

実例1:コンピュータの修理時間

- コンピュータの修理
 - x : 要修理部品数
 - y : 修理時間(分)
- 散布図
 - 強い相関
 - 非線形性あり?



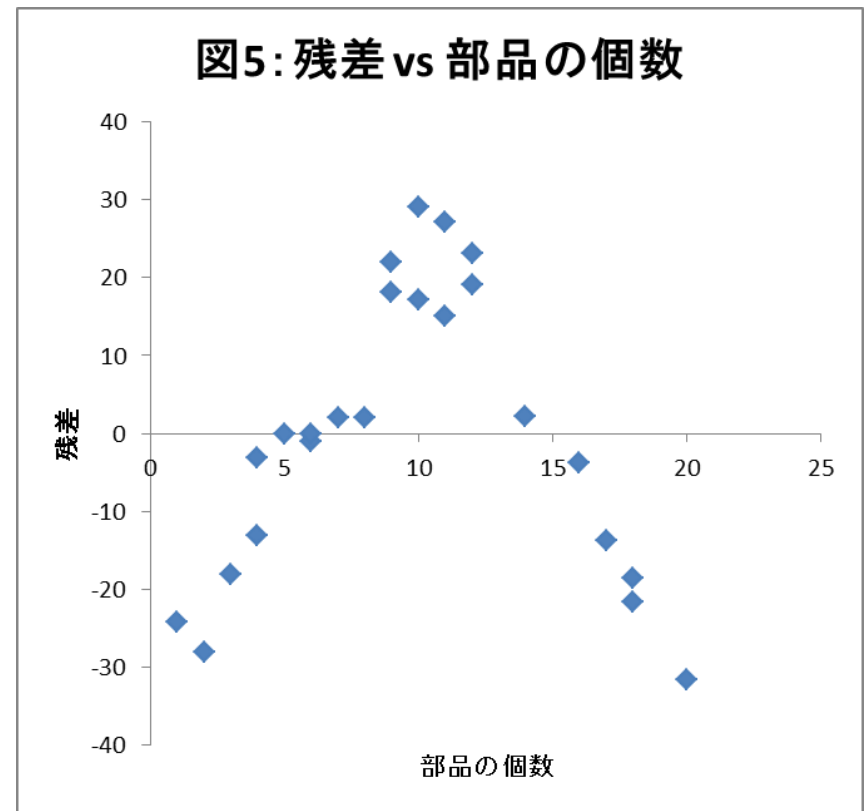
実例1:コンピュータの修理時間

- 計算結果

概要								
回帰統計								
重相関 R	0.946885							
重決定 R2	0.896592							
補正 R2	0.891891							
標準誤差	18.75343							
観測数	24							
分散分析表								
	自由度	変動	分散	割された分散	有意 F			
回帰	1	67084.79	67084.79	190.7492	2.56E-12			
残差	22	7737.206	351.6912					
合計	23	74822						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	37.21273	7.985252	4.660182	0.00012	20.65233	53.77313	20.65233	53.77313
部品の個数	9.969504	0.721842	13.8112	2.56E-12	8.472495	11.46651	8.472495	11.46651

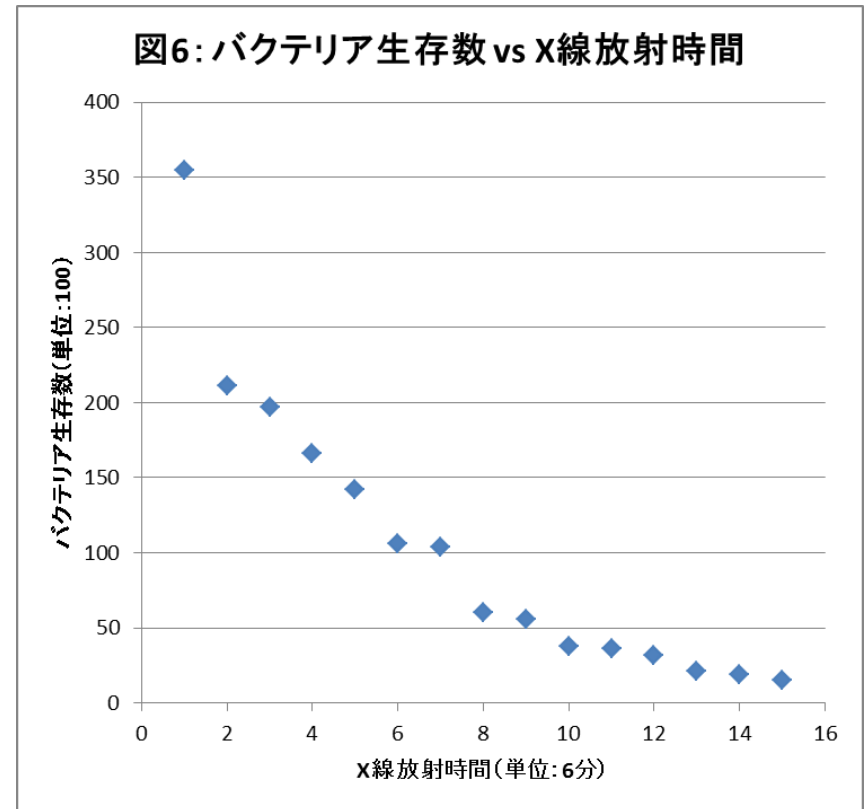
実例1:コンピュータの修理時間

- 残差プロットの様子
 - 非線形性が認められる。
 - 線形回帰モデルは不十分である。
- なぜ、線形性が成り立たないかを確認する必要あり。
 - 修理部品数が10を超えると何かある？



実例2：X線放射によるバクテリア生存数

- X線放射によるバクテリアの生存数
 - x : 200キロボルトのX線の放射時間(単位: 6分)
 - y : バクテリア生存数(単位: 100)
- 散布図
 - 非線形性あり?



実例2：X線放射によるバクテリア生存数

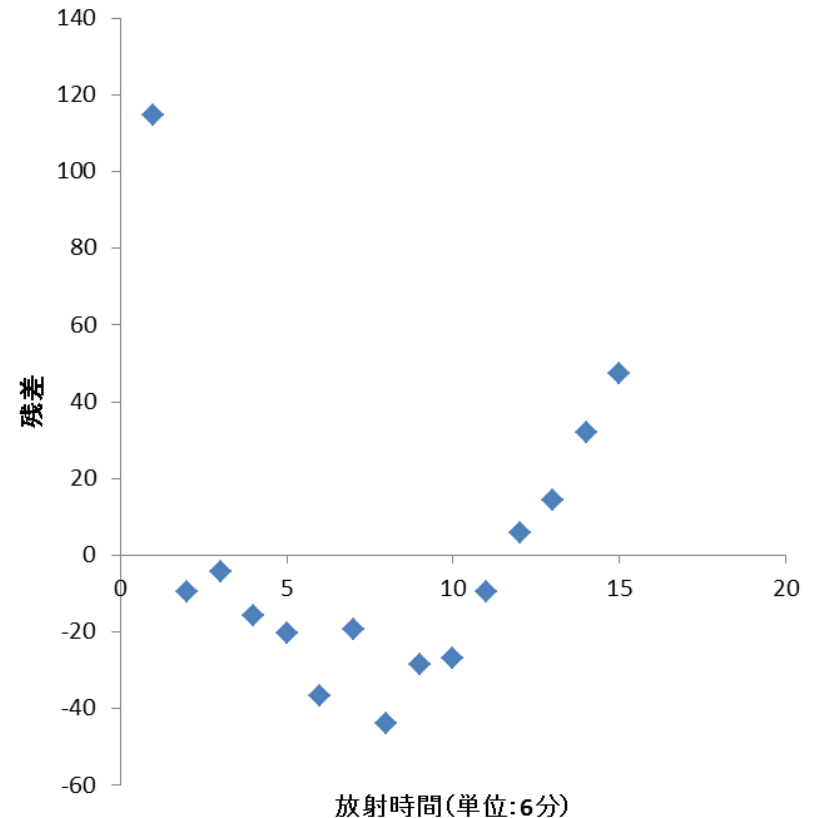
• 計算結果

概要								
回帰統計								
重相関 R	0.907422							
重決定 R2	0.823415							
補正 R2	0.809832							
標準誤差	41.83243							
観測数	15							
分散分析表								
	自由度	変動	分散	割された分散	有意 F			
回帰	1	106080.4	106080.4	60.61901	3.01E-06			
残差	13	22749.38	1749.952					
合計	14	128829.7						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	259.581	22.72999	11.4202	3.78E-08	210.4758	308.6861	210.4758	308.6861
放射時間(	-19.4643	2.499966	-7.78582	3.01E-06	-24.8651	-14.0634	-24.8651	-14.0634

実例2：X線放射によるバクテリア生存数

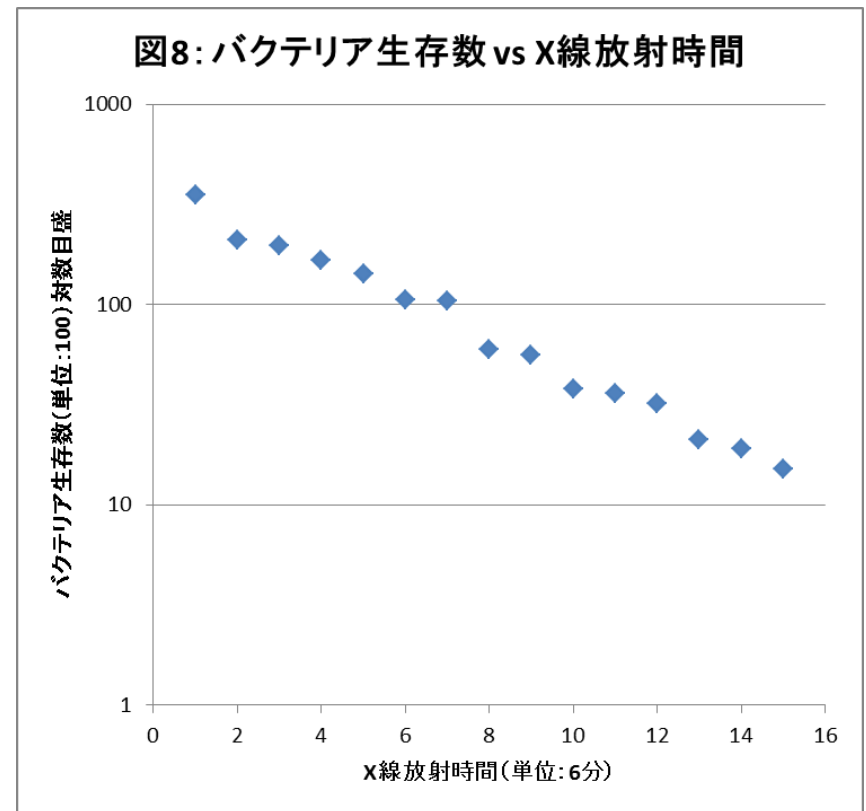
- 散布図による検証
 - 非線形が認められる。
 - 解決策あり？
- 理論仮説
 - シングルヒット仮説
 - バクテリアに単一の生命の中心があり、そこにX線が当たればバクテリアは死滅する。

図7：残差 vs 放射時間（単位：6分）



実例2：X線放射によるバクテリア生存数

- 仮定：
 - SH仮説が正しい。
 - バクテリアが一様に分布している。
- 結論：
 - 単位時間ごとに一定の割合でバクテリアが死滅する。
 - 「一定の数」ではない。
 - 縦軸を対数変換すると直線関係になるはず。



実例2：X線放射によるバクテリア生存数

• 計算結果

概要								
回帰統計								
重相関 R	0.994162							
重決定 R2	0.988359							
補正 R2	0.987463							
標準誤差	0.047779							
観測数	15							
分散分析表								
	自由度	変動	分散	割された分散	有意 F			
回帰	1	2.519604	2.519604	1103.702	5.86E-14			
残差	13	0.029677	0.002283					
合計	14	2.549282						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	2.594111	0.025961	99.92222	3.79E-20	2.538025	2.650197	2.538025	2.650197
放射時間(	-0.09486	0.002855	-33.222	5.86E-14	-0.10103	-0.08869	-0.10103	-0.08869

実例2: X線放射によるバクテリア生存数

図9: 残差 vs 放射時間(単位:6分)

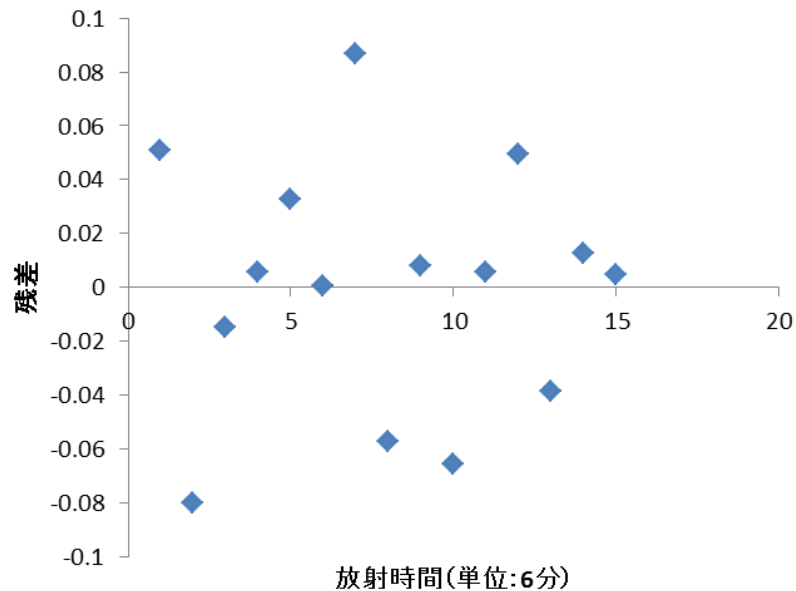
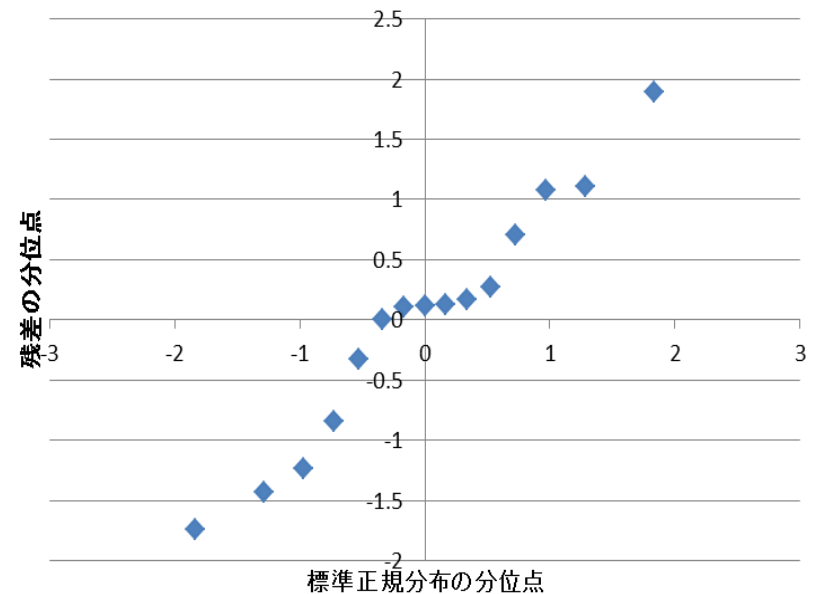


図10: 残差の正規QQプロット



実例2：X線放射によるバクテリア生存数

- 最終判断

- 生存数を対数変換した後に線形回帰モデルを当てはめるのが適切である。
 - シングルヒット仮説という理論的な支持がある。
- 6分間のX線放射によって、約10%のバクテリアが死滅する。
 - 死滅率に関する信頼係数95%信頼区間:[8.9%, 10.1%]
- 残差を検証した結果、誤差項に課せられた条件はおおよそ成り立っている。