



احتمال و آمار مهندسی

پروژه پایا ترم با زبان برنامه نویسی R



زمستان ۱۴۰۱

علیرضا محمدزاده

مدی جعفریان



مقدمه

یکی از مهمترین شاخص ها جهت بهره برداری از مخازن نفت و گاز، تخلخل و ویسکوزیته سیال می باشد. در این پروژه از این داده ها برای پنج مخزن استفاده شده است.

با داشتن تخلخل و ویسکوزیته که به ترتیب از خواص سنگ و سیال هستند، در کنار سایر مشخصات مانند تراوایی، میتوان مقادیری همچون دبی تولیدی از یک مخزن را بدست آورد.

شرح پروژه

یک میدان نفتی ممکن است از چند ناحیه بر اساس خواص پتروفیزیکی تشکیل شده باشد و شناخت خواص آن جهت اقدامات آتی و تعداد حلقه های چاه در هر ناحیه حائز اهمیت است.

داده ها

در یک میدان نفتی، مانند میدان نفتی اهواز، ۵ ناحیه A, B, C, D و E وجود دارد. از هر ناحیه با انجام عملیات مغزه گیری، خواص سنگ و سیال درون هر مغزه اندازه گیری شده است. برای شناخت مخزن و تعمیم این مقادیر به نحوی که کل مخزن را نمایندگی کند از مبحث آماری تحت عنوان Scale up استفاده می شود. مع الوصف، اولین گام بدست آوردن تخلخل و ویسکوزیته میانگین سنگ و سیال مخزن است.

در این پروژه، از مقادیر تخلخل و ویسکوزیته که برای هر نمونه با آزمایش های RCAL بدست آمده، استفاده شده است.

روش محاسبات

ابتدا داده های موجود با پسوند CSV. ذخیره شده و با دستور read.csv() فراخوانی میشود. نام فایل داده ها در سیستم APdata و در فایل R. ذخیره شده data_petrol می باشد. همچنین، با دستور str() نیز میتوان چند سطر اول از هر ستون را مشاهده کرد:

```
R 4.2.2 · D:/Alireza/E-Classes/401-1/statistics and Probabilities/R class 1401_2/AlirezaProject/
> data_petrol=read.csv("D:\\Alireza\\E-Classes\\401-1\\statistics and Probabilities\\R
class 1401_2\\AlirezaProject\\second data\\APdata.csv",header=TRUE)
> str(data_petrol)
'data.frame': 51 obs. of 4 variables:
 $ Grid      : chr  "A" "A" "A" "A" ...
 $ Porosity  : num  0.326 0.41 0.417 0.441 0.424 0.301 0.348 0.416 0.404 0.305 ...
 $ viscosity : num  0.97 1.31 1.16 0.9 1.15 0.98 1.04 1.12 1.16 1.11 ...
 $ Fluid.type: chr  "light oil" "heavy oil" "light oil" "light oil" ...
> |
```

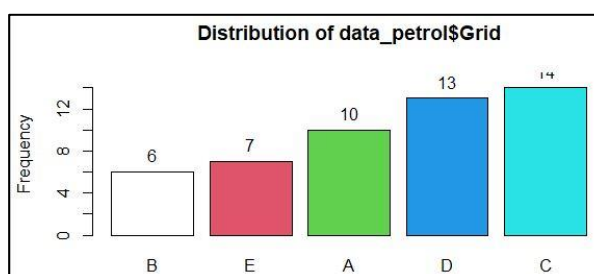
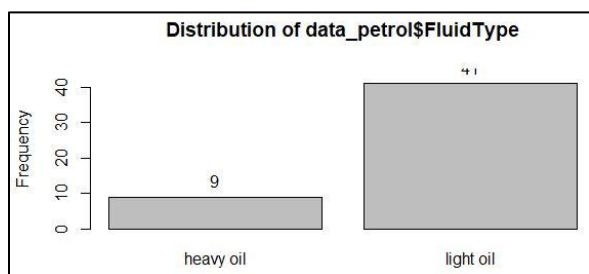
با استفاده از دستور head() و tail() مقادیر پنج ردیف اول و آخر نمایش داده می شوند:

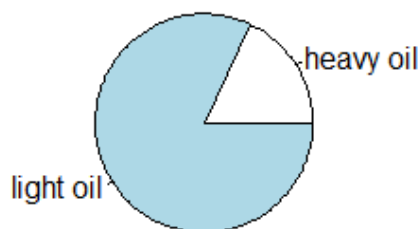
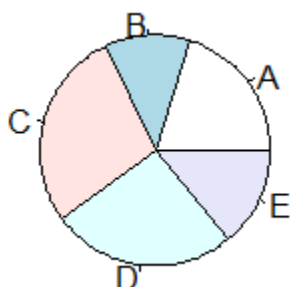


```
> head(data_petro1)
  Grid Porosity Viscosity Fluid.type
1    A    0.326     0.97 light oil
2    A    0.410     1.31 heavy oil
3    A    0.417     1.16 light oil
4    A    0.441     0.90 light oil
5    A    0.424     1.15 light oil
6    A    0.301     0.98 light oil
> tail(data_petro1,5)
  Grid Porosity Viscosity Fluid.type
46    E    0.415     1.11 light oil
47    E    0.445     1.02 light oil
48    E    0.400     1.23 light oil
49    E    0.439     1.20 light oil
50    E    0.404     0.97 light oil
> |
```

با استفاده از تابع `epiDisplay` جدول فراوانی برای تعداد داده های مربوط به هر ناحیه و همچنین تعداد موارد نفت سبک یا سنگین تولید می شوند:

```
> tab1(data_petro1$Grid, sort.group = "increasing", cum.percent = TRUE)
data_petro1$Grid :
      Frequency Percent Cum. percent
B              6      12          12
E              7      14          26
A             10      20          46
D             13      26          72
C             14      28         100
Total          50     100         100
> tab1(data_petro1$FluidType, sort.group = "increasing", cum.percent = TRUE)
data_petro1$FluidType :
      Frequency Percent Cum. percent
heavy oil      9      18          18
light oil     41      82         100
Total          50     100         100
> |
```





با استفاده از دستوری summary میتوان خلاصه ای از اطلاعات داده ها را مشاهده کرد:

Grid		Porosity	Viscosity	Fluidtype
Length: 50		Min. : 0.3010	Min. : 0.900	Length: 50
Class : character		1st Qu.: 0.3688	1st Qu.: 1.012	Class : character
Mode : character		Median : 0.4170	Median : 1.150	Mode : character
		Mean : 0.4022	Mean : 1.131	
		3rd Qu.: 0.4348	3rd Qu.: 1.200	
		Max. : 0.4480	Max. : 1.400	

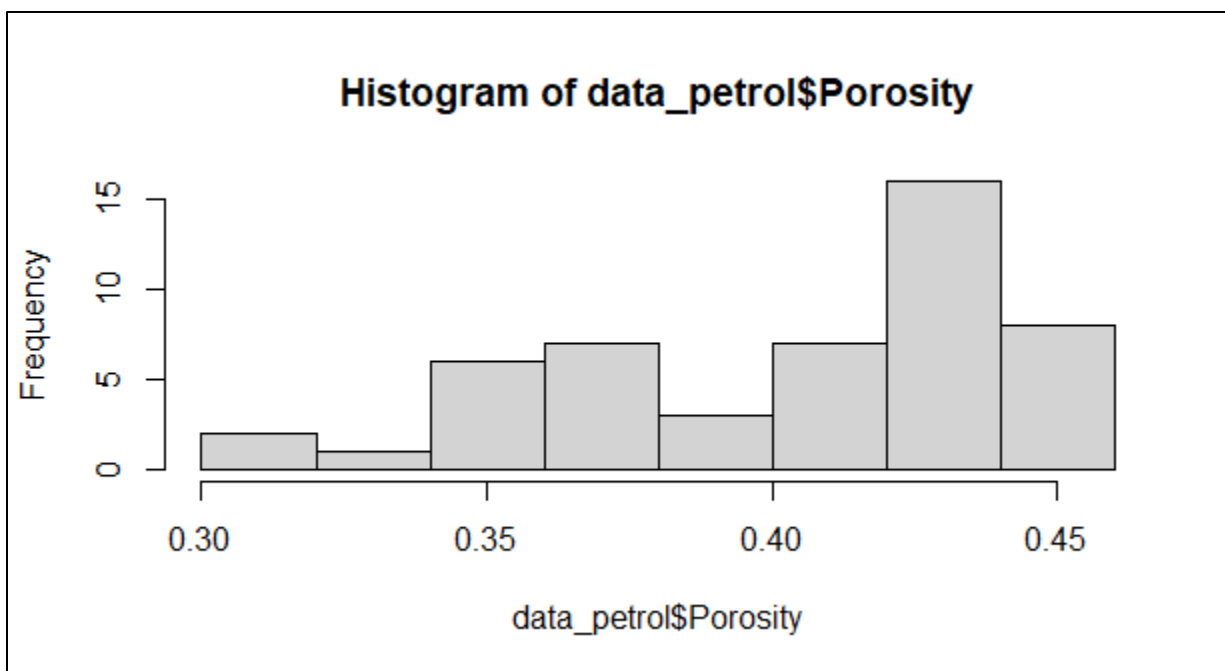
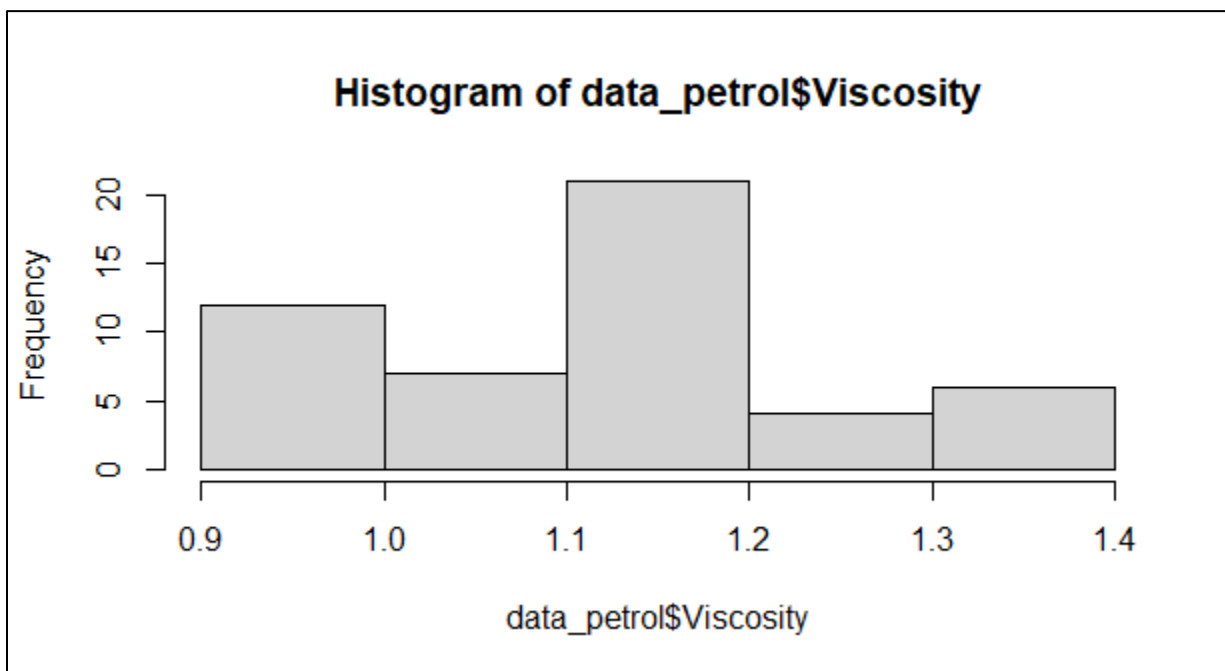
حال با استفاده از کتابخانه modeest و دستورهایی مربوطه، مقادیر مد، میانه، میانگین، دامنه، واریانس و انحراف معیار محاسبه

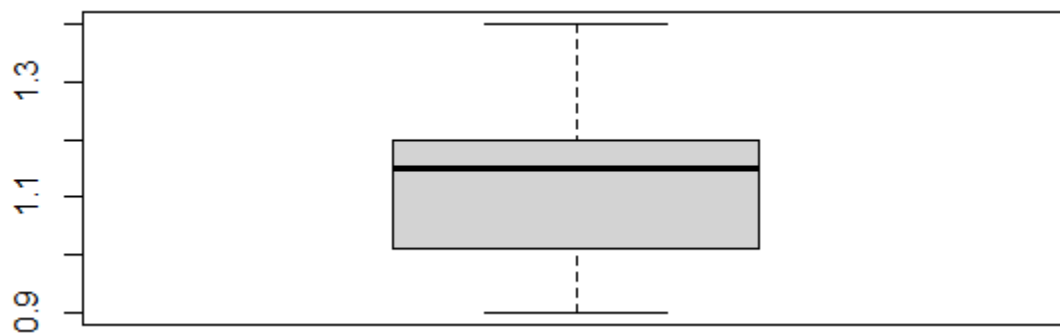
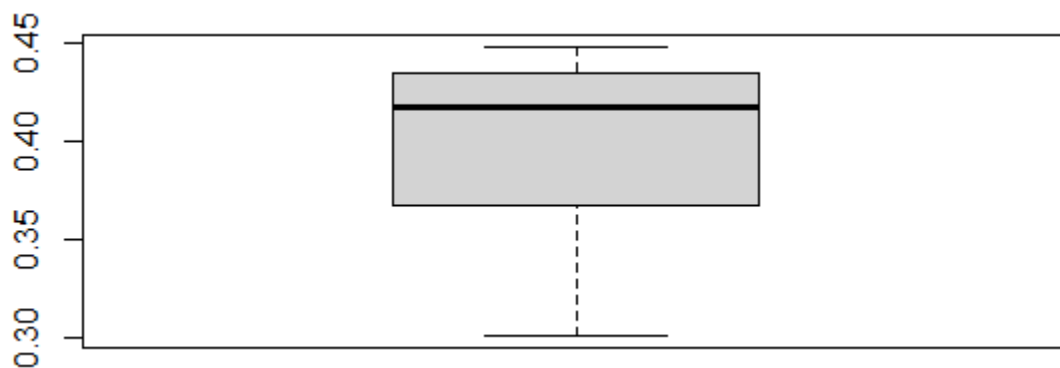
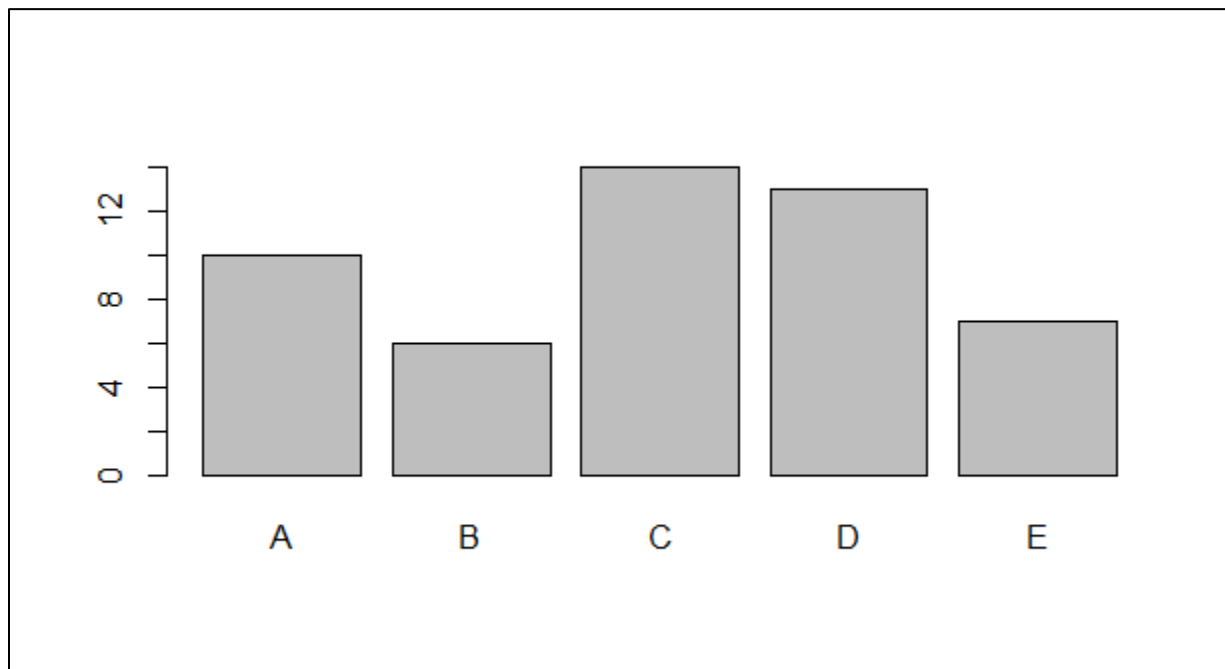
می شوند:

```
> #Range
> range(data_petro$Porosity)
[1] 0.301 0.448
> range(data_petro$Viscosity)
[1] 0.9 1.4
> #variance
> var(data_petro$Porosity)
[1] 0.001547117
> var(data_petro$Viscosity)
[1] 0.0183282
> #standard deviation
> sd(data_petro$Porosity)
[1] 0.0393334
> sd(data_petro$Viscosity)
[1] 0.1353817
> |
```

```
> library(modeest)
> mfv(data_petro$Porosity)
[1] 0.441 0.444
> mfv(data_petro$Viscosity)
[1] 1.2
> #median
> median(data_petro$Porosity)
[1] 0.417
> median(data_petro$Viscosity)
[1] 1.15
> #mean
> mean(data_petro$Porosity)
[1] 0.40216
> mean(data_petro$Viscosity)
[1] 1.1306
```

حال نمودارهای جعبه ای و هیستوگرام مربوطه رسم میشوند:







```
> #part 1.4
> #covariance
> cov(data_petro1$Porosity, data_petro1$viscosity )
[1] 0.001208269
> #correaltion coeffient
> cor(data_petro1$Porosity, data_petro1$viscosity )
[1] 0.2269041
```

حال میخواهیم یک فاصله اطمینان ۹۵ درصدی برای میانگین تخلخل محاسبه کنیم:

می دانیم که فاصله اطمینان برای میانگین در بازه ی زیر تعریف می شود:

$$\mu \in \left(\bar{x} - t_{1-\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\frac{\alpha}{2}}(n-1) \frac{s}{\sqrt{n}} \right)$$

لذا مانند زیر، با تعریف متغیر هایی این فاصله را بدست می آوریم:

```
> #part 2.1 & 2.2
> sample.mean=mean(data_petro1$Porosity)
> print(sample.mean)
[1] 0.40216
> sample.n=length(data_petro1$Porosity)
> sample.sd=sd(data_petro1$Porosity)
> sample.se=sample.sd/sqrt(sample.n)
> print(sample.se)
[1] 0.005562583
> alpha= 0.05
> degrees.freedom=sample.n-1
> t.score=qt(p=1-(alpha/2), df=degrees.freedom, lower.tail = TRUE)
> margin.error=t.score*sample.se
> L= sample.mean - margin.error
> U= sample.mean+margin.error
> print((c(L, U)))
[1] 0.3909816 0.4133384
```

لذا بازه ی مورد نظر (0.3909816, 0.4133384) می باشد.

حال میخواهیم یک فاصله اطمینان ۹۵ درصدی برای واریانس محاسبه کنیم. می دانیم:

$$\sigma^2 \in \left(\frac{(n-1)s^2}{X_{1-\frac{\alpha}{2}}^2(n-1)}, \frac{(n-1)s^2}{X_{\frac{\alpha}{2}}^2(n-1)} \right)$$



لذا با توجه به حد بالا (L) و حد پایین (U) به صورت زیر این فاصله را بدست می آوریم:

```
> #part 2.2
> sample.var = var(data_petro1$Porosity)
> print(sample.var)
[1] 0.001547117
> sample.n1 = length(data_petro1$Porosity)
> alpha = 0.05
> degrees.freedom = sample.n1 - 1
> L = qchisq(1-(alpha/2),df=degrees.freedom,lower.tail = TRUE)
> U = qchisq(alpha/2,df=degrees.freedom,lower.tail = TRUE)
> l = (degrees.freedom - sample.var)/L
> u = (degrees.freedom - sample.var)/U
> print(c(l,u))
[1] 0.6977609 1.5527993
```

بنابراین فاصله اطمینان ۹۵ درصدی برای واریانس داده های مربوط به تخلخل (0.6977609, 1.5527993) می باشد.

در بخش بعدی، می خواهیم آزمون زیر را برای میانگین با سطح معناداری ۰.۰۵ انجام دهیم:

$$H_0: \mu = 10, H_1: \mu > 10$$

سطح بحرانی عبارت است از:

$$T = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{1-\alpha}(n-1)$$

بدین ترتیب خواهیم داشت:

```
> #part 2.3
> alpha=0.05
> sample.mean=mean(data_petro1$Porosity)
> sample.n=length(data_petro1$Porosity)
> sample.sd=sd(data_petro1$Porosity)
> sample.se=sample.sd/sqrt(sample.n)
> print(sample.se)
[1] 0.005562583
> t.stat=(sample.mean-10)/sample.se
> degrees.freedom=sample.n-1
> t.score=qt(p=1-alpha, df=degrees.freedom, lower.tail = TRUE)
> t.stat
[1] -1725.428
> t.score
[1] 1.676551
```




چون $t.stat > t.score$ نشده است، پس نمیتوان فرض H_0 را رد کرد.

با توجه به نمودارهای دایره ای، میتوان مشخص کرد که سهم تعداد نمونه های مغزه گیری شده از ناحیه ی C بیشتر بوده و همچنین بیشتر نمونه گیری ها نشان دهنده سبک بودن سیال مخزن است.

اگر فرضا معیار این باشد که نفت با ویسکوزیته 1.25 cp به بالا نفت سنگین محسوب شود، میانگین نمونه ها که 1.1306 cp است نشان میدهد که نمونه ی نفت احتمالا سبک است.

همچنین تخلخل میانگین 0.040216 است که تخلخل بسیار بالایی است.

با توجه به کم بودن مقدار واریانس برای تخلخل و ویسکوزیته، میتوان نتیجه گرفت که احتمالا دقت اندازه گیری برای این مقادیر، دقت مناسبی بوده است.

مقدار کوواریانس در این پروژه نمیتواند مشخص کننده چیزی باشد چرا که اساسا ویسکوزیته و تخلخل هیچ رابطه ای با یکدیگر ندارند.