

مجموعه تمرین های پنجم

۱. در مجموعه داده های **emissions** یک داده پرت در ستون داده انتشار گازهای گلخانه ای (CO_2) وجود دارد. با استفاده از **identify** این مقدار را پیدا کنید و سپس نمودار را بدون این نقطه رسم کنید.

۲. مجموعه داده های ساده **chips** شامل اطلاعاتی درباره ضخامت تراشه های کامپیوتری می باشد. داده ها برای دو نوع تراشه وجود دارد، هر کدام در ۴ مکان مختلف اندازه گیری می شوند. یک **boxplot** کنار هم از ضخامت برای هر مکان اندازه گیری رسم کنید. (این شکل باید شامل ۸ **boxplot** در کنار هم باشد). آیا میانگین ها شبیه هم می باشند؟ واریانس ها چگونه؟

۳. داده های ساده **chicken**، حاوی وزن مرغ هایی هستند که ۱ یا ۳ جیره (سهم) غذایی مختلف داده می شوند. یک **boxplot** از تمام ۳ جیره ایجاد کنید. آیا به نظر می رسد تفاوت در میانگین وجود دارد؟

۴. مجموعه داده های ساده **WeightData** شامل اطلاعاتی درباره وزن کودکان 0-144 ماه می باشد. یک نمودار **boxplot** کنار هم از وزن تقسیم بر سن در سالهای مختلف رسم کنید. چه نوع روندی را می بینید؟

(متغیر **age** براساس ماه است. برای تبدیل به سال می توانید با استفاده از دستور **cut** به صورت زیر انجام دهید:

```
> age.yr = cut(age,seq(0,144,by=12),labels=0:11)
```

فرض بر این است که مجموعه داده **attach** شده است.)

۵. مجموعه داده های ساده **carbon** شامل اطلاعاتی درباره مقادیر مونوکسید کربن در ۳ مکان مختلف صنعتی است. داده ها دارای دو متغیر: مونوکسید کربن (**a carbon monoxide reading**) و یک متغیر عامل برای پیگیری سایت (**track of the site**) هستند. یک نمودار **boxplot** از سطوح مونوکسید برای هر سایت ایجاد کنید. آیا تفاوت معنی داری وجود دارد؟

۶. برای مجموعه داده های، نوزادان (**babies**)، نمودار جفت (**pairs(babies)**) را برای بررسی روابط بین متغیرها رسم کنید. کدام متغیرها دارای رابطه خطی هستند؟ برای متغیرهای وزن تولد (**birthweight**) و حاملگی (**gestation**)، یک نمودار پراکندگی با استفاده از کاراکترهای مختلف (**pch**) بسته به مقدار عامل سیگار کشیدن رسم کنید.