Comparing Two Means

DISCOVERING STATISTICS USING R

ANDY FIELD

Aims

- *T*-tests:
  - Independent
  - Dependent (aka paired, matched)
- Rationale for the tests
  - Assumptions
- *T*-tests as a GLM
- Interpretation
- Reporting results

Experiments

- The simplest form of experiment that can be done is one with only one independent variable that is manipulated in only two ways and only one outcome is measured.
  - More often than not, the manipulation of the independent variable involves having an experimental condition and a control.
  - E.g., Is the movie *Scream 2* scarier than the original *Scream*? We could measure heart rates (which indicate anxiety) during both films and compare them.
- This situation can be analysed with a *T*-test

Independent T-test

- Independent *T*-test
  - Compares two means based on independent data.
  - E.g., data from different groups of people.
- Significance testing
  - Testing the significance of *Pearson's correlation coefficient*
  - Testing the significance of b_1 in regression.

Rationale for the *T*-test

$$t = \frac{\begin{array}{c}\text{observed difference}\\ \text{between sample means}\end{array} - \begin{array}{l}\text{expected difference}\\ \text{between population means}\\ \text{(if null hypothesis is true)}\end{array}}{\begin{array}{c}\text{estimate of the standard error of the difference between two}\\ \text{sample means}\end{array}}$$

Example

- Is arachnophobia (fear of spiders) specific to real spiders or is a picture enough?
- Participants
  - 24 arachnophobic individuals
- Manipulation
  - 12 participants were exposed to a picture of a spider (=group1)
  - 12 were exposed to a real spider (=group2)
- Outcome
  - Anxiety

workflow

- Summary statistics
- Graphical display
  - Bar charts
- Testing assumptions
- Perform the T-test

Summary statistics

```
> ddply(df, .(catvar), summarise,
    Nobs = sum(!is.na(contvar)),
    Nmiss = sum(is.na(contvar)),
    mean = mean(contvar, na.rm=TRUE),
    sd = sd(contvar, na.rm=TRUE),
    se   = sd/sqrt(Nobs),
    t = qt(0.975, Nobs-1),
    lower = mean - t*se,
    upper = mean + t*se)
```

Graphical display

- Bar charts
  - Used to display and compare the number, frequency or other measure (e.g. mean) for different discrete categories of data
  - Vertical bars: up to 7 categories
  - Horizontal bars
    - More than 7 categories
    - Long label names

Graphical display

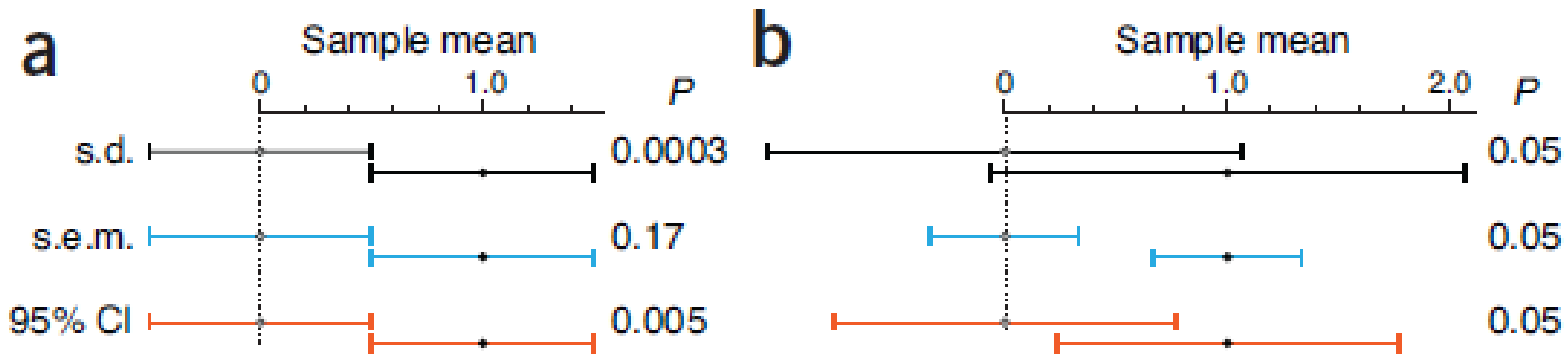
- Barchart with 95% CI

```
> plot4 <- ggplot(df, aes(catvar, contvar))
> plot4 +
  stat_summary(fun.data = mean_cl_normal, geom = "errorbar", width=0.2) +
  stat_summary(fun.y = mean, geom = "bar", fill = "slateblue", colour = "Black") +
  labs(x = “xlabel", y = “ylabel") +
  geom_jitter() +
  theme(legend.position="none", axis.title.x=element_blank())
```

Error Bar Charts

- The bar (usually) shows the mean score
- The error bar sticks out from the bar like a whisker.
- Depending on what you want to show:
  - The confidence interval (usually 95%)
  - The standard deviation
  - The standard error (of the mean)

Error bars and statistical significance



Refs
Krzywinski, M., & Altman, N. (2013). Points of Significance 02: Error bars. Nature Methods, 10(10), 921–922. doi:10.1038/nmeth.2659
Cumming, G., Fidler, F., & Vaux, D. L. (2007). Error bars in experimental biology. The Journal of Cell Biology, 177(1), 7–11. doi:10.1083/jcb.200611141

Assumptions of the *t*-test

- The independent *T*-test is a *parametric test* based on the normal distribution. Therefore, it assumes:
  - Outcome var is continuous.
  - The sampling distribution is normal in each group.
- The independent *T*-test, because it is used to test different groups of people, also assumes:
  - Variances in these populations are roughly equal (*homogeneity of variance*).
  - Scores in different treatment conditions are independent (because they come from different people).

Testing normality assumption

- Histograms

> ggplot(**df**,aes(x=**contvar**)) +

geom_histogram(data=subset(**df**,**catvar** == '**level1name**'),fill = "red", alpha = 0.2, binwidth=4) +

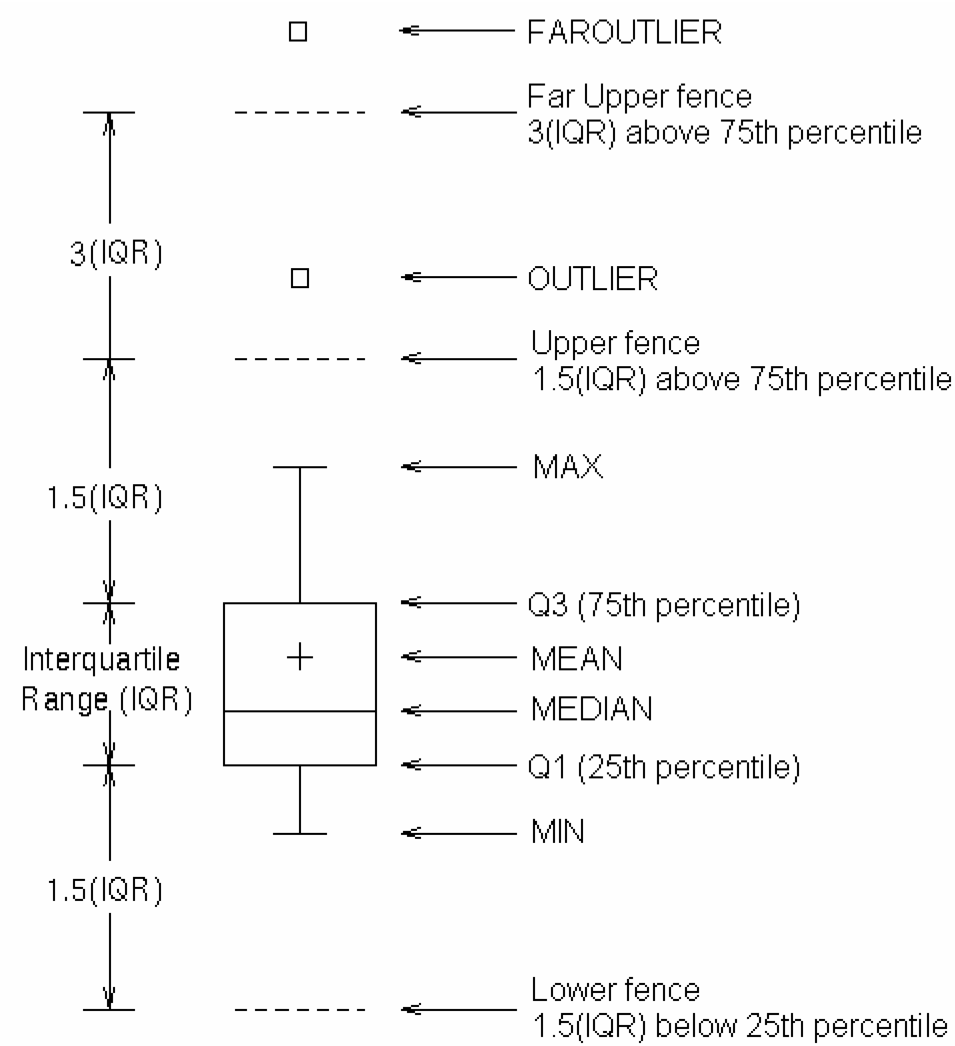
geom_histogram(data=subset(**df**,**catvar** == '**level2name**'),fill = "blue", alpha = 0.2, binwidth=4)

Testing normality assumption

- Formal normality test

```
> by(df$contvar, df$catvar, shapiro.test)
```

Boxplots (Box-Whisker Diagrams)



Testing normality assumption

- Boxplots

> plot <- ggplot(**df**, aes(**catvar**, **contvar**))

> plot + geom_boxplot(fill="slateblue") +

labs(x = “**catvarlabel**", y = “**contvarlabel**") +

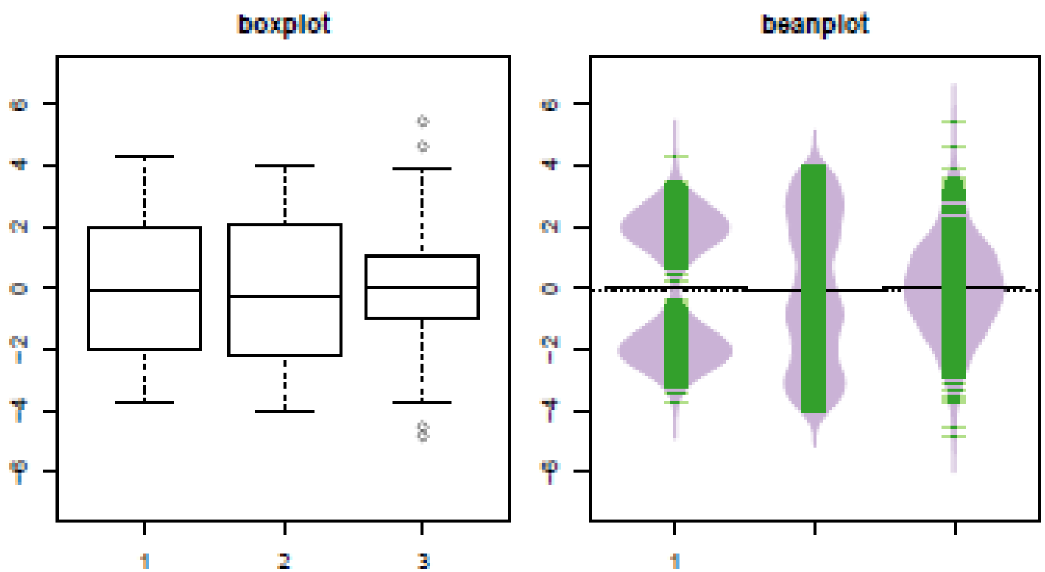
stat_summary(fun.y=mean, geom="point", shape=20, size=5, color="red", fill="red") +

geom_jitter() +

theme(legend.position="none", axis.title.x=element_blank())

Testing normality assumption

- Bean plot {beanplot}
  - Alternative to boxplot
  - Shows individual observations
  - Shows the shape of the distribution
  - Draws a line at the average



Testing normality assumption

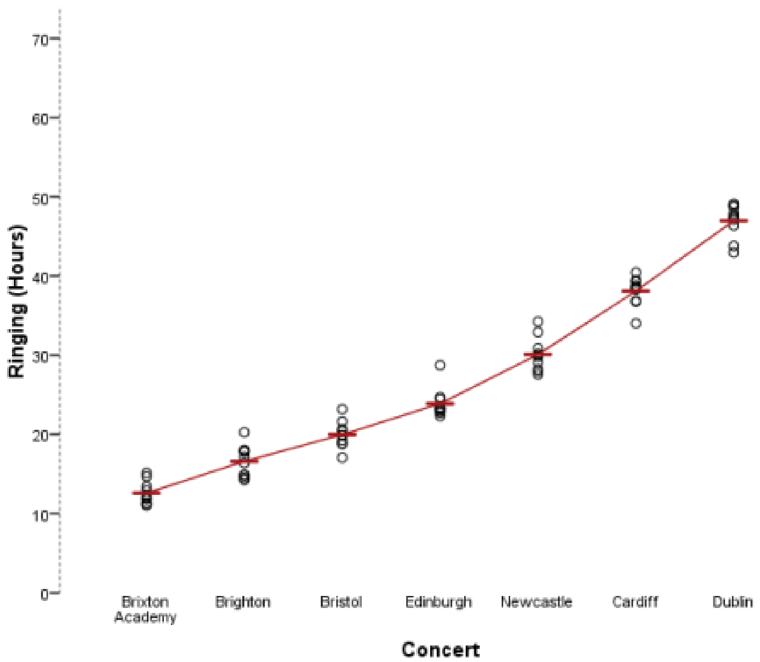
- Beanplot {beanplot}
  - beanplot(**contvar** ~ **catvar**, data=**df**, col="#CAB2D6")

Assessing Homogeneity of Variance

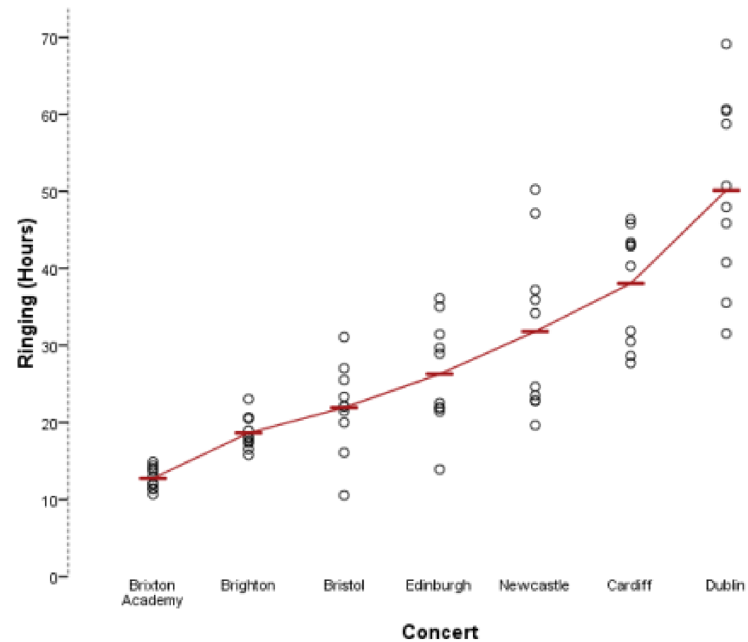
- Graphically
- Two-sample tests
  - Folded form F-test
- Two or more samples
  - Levene's test
  - Brown-Forsythe

Ref: Zhang Shuqiang:
Fourteen Homogeneity of Variance Tests: When and How to use them

Homogeneity of Variance



Homogeneous



Heterogeneous

Levene's test

- Two or more samples
- Based on
  - Absolute deviations from the group mean
  - Squared deviations from the group mean
- Assumption:
  - normality
  - Sample sizes need not to be equal
- H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$
- R function

  > leveneTest(…center="mean")

Output for Levene's Test

```
> leveneTest(rexam$exam, rexam$uni)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  1  2.0886 0.1516
      98

> leveneTest(rexam$numeracy, rexam$uni)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  1   5.366 0.02262 *
      98
```

Testing equality of variance

- F test to compare the variances of two samples from normal populations

  > var.test()

- Levene's test (robust test)

  > leveneTest()

The Independent *T*-test (equal variances)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Unequal variance T-test

- Equal variance T-test
  - Pooled variance
- Unequal variance T-test
  - Welch approximation

The Independent *T*-test Using **R**

- To do a *T*-test we use the function *t.test()*.

  > t.test(**contvar** ~ **catvar**, data = **df**, paired=FALSE, var.equal=TRUE)

Output from the Independent *t*-test

```
        Welch Two Sample t-test

data:  Anxiety by Group

t = -1.6813, df = 21.385, p-value = 0.1072

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:
 -15.648641   1.648641

sample estimates:
    mean in group Picture mean in group Real Spider
                       40                          47
```

34.09538 45.90462

Reporting the Results

- On **average**, participants experienced greater anxiety from real spiders (*mean* = 47, *95%CI*= [40;54]), than from pictures of spiders (*mean* = 40, *95%CI*=[34;46]). This difference was not significant at the 5% significance level (p =0.10, $n_1 = n_2$ =12)

The *T*-test as a GLM

$$A_i = b_0 + b_1 G_i + \varepsilon_i$$

$$\text{anxiety}_i = b_0 + b_1 \text{group}_i + \varepsilon_i$$

Dummy coding

- Provides a way of using categorical predictors in linear regression
- Uses zeros and ones to convey group membership
- For k groups, we can create k-1 dummies
- For 2 groups, only one dummy variable
  - X=1 when an observation belongs to group 2 and 0 otherwise

Picture Group

- X=0
- Intercept = mean of baseline group=picture

$$E(anxiety|group = picture) = b_0 + b_1 0$$
$$E(anxiety|group = picture) = b_0 = \bar{x}_{picture}$$

Real Spider Group

- X= 1
- b_1 = Difference between means

$$E(anxiety|group = real) = b_0 + b_1 1$$
$$E(anxiety|group = real) = b_0 + b_1 = \bar{x}_{real}$$
$$b_1 = \bar{x}_{real} - \bar{x}_{picture}$$

Output from a Regression

```
Call:
lm(formula = Anxiety ~ Group, data = spiderLong)

Residuals:
   Min     1Q Median     3Q    Max
 -17.0   -8.5    1.5    8.0   18.0

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        40.000      2.944  13.587 3.53e-12 ***
GroupReal Spider    7.000      4.163   1.681    0.107
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.2 on 22 degrees of freedom
Multiple R-squared: 0.1139,    Adjusted R-squared: 0.07359
F-statistic: 2.827 on 1 and 22 DF,  p-value: 0.1068
```

Dependent T-test

- Compares two means based on related data.
  - E.g., Data from the same people measured at different times.
  - Data from ‘matched’ samples.

The Dependent *T*-test

$$t = \frac{\bar{D} - \mu_D}{\frac{s_D}{\sqrt{n}}}$$

Assumptions dependent *T*-test

- The dependent *T*-test is a *parametric test* based on the normal distribution. Therefore, it assumes:
  – The sampling distribution is normally distributed. In the dependent *T*-test this means that the sampling distribution of the *differences* between scores should be normal, not the scores themselves.
  – Continuous outcome.

Example

- Is arachnophobia (fear of spiders) specific to real spiders or is a picture enough?
- Participants
  - 12 spider phobic individuals
- Manipulation
  - Each participant was exposed to a real spider and a picture of the same spider at two points in time
- Outcome
  - Anxiety

The Dependent *T*-test Using **R**

- To do a dependent *T*-test we again use the function *t.test()* but this time include the option *paired = TRUE*.
  - > t.test(**contvar** ~ **catvar**, data = **df**, paired=TRUE)

Dependent *T*-test Output

```
        Paired t-test
data:  spiderWide$real and spiderWide$picture
t = 2.4725, df = 11, p-value = 0.03098

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:
  0.7687815 13.2312185

sample estimates:
mean of the differences
```

Reporting the Results

- On average, participants experienced significantly greater anxiety from real spiders (*mean* = 47.00, *95%CI*= [40;54])) than from pictures of spiders (*mean* = 40.00, *95%CI*=[34;46]) (p = 0.03, n=12)

One-sample T-test

- used to determine whether a sample of observations could have been generated by a process with a specific mean

When Assumptions are Broken

- One-sample T-test
  - Wilcoxon one-sample signed rank test
- Independent *T*-test
  - Mann–Whitney test
  - Wilcoxon rank-sum test
- Dependent *T*-test
  - Wilcoxon signed-rank test
- Robust tests

DEMO T-TESTS

Open the program Ch9_means.R

Exercises

- Task 1
  - The **sleep** dataset available in R (package datasets) shows the effect of two soporific drugs. Both drugs were administered to each person. The outcome variable is the increase/decrease in hours of sleep compared to control. Analyse the data with the appropriate T-test. Interpret the results and verify the assumptions.

Exercises

- Task 2
  - In this example taken from SAS, a stimulus was being examined to determine its effect on systolic blood pressure. Twelve men participated in the study. Each man’s systolic blood pressure was measured both before and after the stimulus was applied. The following statements input the data:
    - Group<-gl(2, 12, labels = c("SBPbefore", "SBPafter"))
    - pressure=c(120,140,126,124,128,130,130,140,126,118,135,127,128,132,118,131,125,132,131,141,129,127,137,135)
    - pressureLong=data.frame(Group,pressure)
  - Analyse the data with the appropriate T-test.