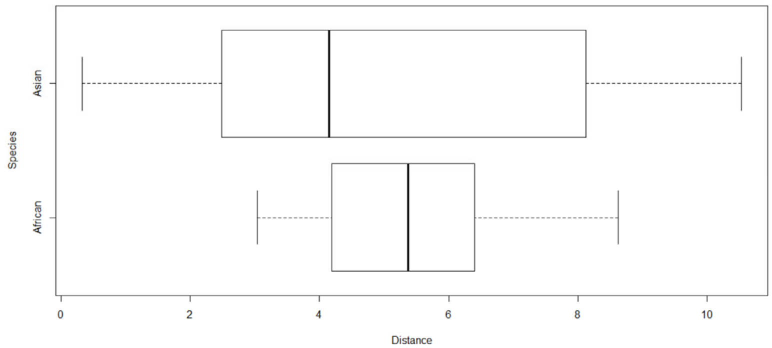
Stat 301 – Day 29
Two-sample *t* procedures (Ch. 4)

Example 1: Researchers Holdgate et al. (2016) studied walking behavior of elephants in North American zoos to see if there is a difference in average distance traveled by African and Asian elephants. They put GPS loggers on 33 African elephants and 23 Asian elephants and measured the distance (in kilometers) the elephants walked per day

```
> with(elephants, boxplot(Distance ~ Species, horizontal=T))
```



(a) Where there more Asian or African elephants in the study?

(b) Where there more Asian or African elephants that walked 4 km or less? 8.5 km or more?

(c) Were there any "outlier" elephants? Explain how you know.

(d) Which species tend to walk more? Which species showed more variability? Which distribution is more symmetric?

(e) Do you think the difference in means will be statistically significant?

(f) State appropriate null and alternative hypotheses to compare the walking distances for these two species.

(g) Does the central limit theorem apply here?

```
> with(elephants, iscamsummary(Distance, Species))
         n   Min    Q1 Median    Q3    Max  Mean     SD
African 33 3.040 4.190  5.370 6.400  8.620 5.399  1.473
Asian   23 0.330 2.485  4.150 8.125 10.530 5.300  3.410
```

(h) Calculate the *t*-test statistic. What does it tell you about the two-sided p-value?

(i) Approximate a 95% confidence interval. Interpret the interval in context.

R output:

```
with(elephants, t.test(Distance ~ Spec
ies, alt="two.sided", var.equal=FALSE)
)
        Welch Two Sample t-test

data:  Distance by Species
t = 0.13092, df = 27.773,
p-value = 0.8968
alternative hypothesis:
true difference in means is not equal
to 0
95 percent confidence interval:
 -1.449990  1.647908
```

JMP output

Analyze > Fit Y by X

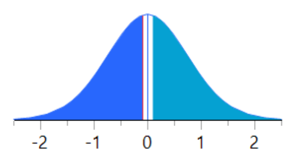
Hot spot > t Test

t Test

Asian-African

Assuming unequal variances

Difference	-0.0990	t Ratio	-0.13092
Std Err Dif	0.7559	DF	27.7734
Upper CL Dif	1.4500	Prob > |t|	0.8968
Lower CL Dif	-1.6479	Prob > t	0.5516
Confidence	0.95	Prob < t	0.4484



(j) Did/Should we use the pooled or unpooled standard error?

(k) How will the p-value and confidence interval change if our sample data turn out to be:

	Group 1 African		Group 2 Asian
n:	66	n:	46
mean, $\bar{x}$:	5.399	mean, $\bar{x}$:	5.300
sample sd, s:	1.462	sample sd, s:	3.372

(l) Have we proven that there is no difference in the walking distances between Asian and African elephants?